A Photo-Crosslinking Platform for the Construction of Chemically Modified and Biofunctional RNA Molecules

Brendan Largey
Linacre College, University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2023

Declaration

I, Brendan Largey, declare that the thesis titled “A Photo-Crosslinking Platform for the Construction of Chemically Modified and Biofunctional RNA Molecules” and the related research presented in this thesis were performed and written by the author under the supervision of Prof. Tom Brown at the University of Oxford. The research is original and has not been submitted in full or in part for any other degree at this or any other university. Contributions from colleagues are appropriately acknowledged.

Acknowledgements

I would like to thank Prof. Tom Brown for the opportunity to conduct research on a fascinating area of science, for his guidance, and for sharing his passion for the research process. I would also like to thank Prof. Afaf El-Sagheer for her guidance and encouragement, and for teaching me many fundamental techniques of nucleic acid chemistry. I would like to thank both for fostering a research environment that encourages innovation and collaboration, which I am lucky to have been a part of.

With respect to the work presented in this thesis, I would like to thank Alexandra Bisia for her invaluable advice on mammalian cell culture and related techniques, Dr. Alice Kennett for synthesizing a phosphoramidite monomer described in Chapter 4, and Sebastian Gołojuch for his advice and help with RNA purification. All oligonucleotide synthesis performed with the Applied Biosystems 394 was done by Prof. Afaf El-Sagheer.

I am grateful for all my friends and colleagues in the Brown group, past and present, whom I have had the pleasure of sharing this experience with.

I would like to thank Linacre College and the EPA Cephalosporin Trust for funding my studies and research at Oxford.

Lastly, I would like to thank my friends and family who have been with me and encouraged me throughout this process, whether they be in Oxford or New York. To my parents, thank you for your unfaltering support throughout my academic journey, tested as it may have been by the fateful washing machine incident. Your love has made all things possible. Θα ήθελα να ευχαριστήσω πολύ την Αλεξάνδρα που μοιράστηκε μαζί μου το ταξίδι.

Abstract

Several key developments in medicine have been made by exploiting the varied biological roles of RNA, as highlighted by the recent applications of mRNA vaccines for SARS-CoV-2 and RNA-guided Cas9 gene editing. These technologies are further enhanced by advances in nucleic acid chemistry that allow for highly targeted and extensive modification of oligonucleotides to improve or alter their functions. Despite the new opportunities presented by RNA technologies, several key challenges remain in their application. These include limitations to the length of chemically synthesized oligonucleotides and a lack of adequate methods to control the spatial-temporal activity of oligonucleotide-based medicines. The work described in this thesis attempts to address these issues with carbazole-derived nucleosides which undergo site-specific reversible crosslinking when irradiated with UV light. Crosslinking nucleosides are first applied to construct Cas9 guide RNAs (gRNAs) from shorter starting RNAs. This approach provides a simple and cost-effective strategy to construct individual gRNAs or libraries and is demonstrated to be compatible with extensive chemical modification. Crosslinked gRNAs are functional both *in vitro* and in cells, with comparable activity to full-length control gRNAs. Several methods of conjugating large DNA cargos to gRNAs are then explored which are compatible with this crosslinking strategy. DNA cargos encoding expressible gene cassettes are applied to attempt to report the transcriptional status of target sites as directed by dCas9. Lastly, the reversible crosslinking of carbazole nucleosides is explored as a platform to control the translation of exogenous mRNAs in cells. Short carbazole-oligonucleotide blockers are applied to inhibit the translation of mRNA in a crosslinking-dependent and sequence-specific manner. Upon reversing the crosslink by UV irradiation, translation is initiated. The blockers are optimized to maximize the dynamic range of activation and are then applied to novel mRNA vectors that cannot be controlled with existing light-regulation strategies. The work of this thesis underscores the extensive versatility of carbazole nucleosides in the manipulation of synthetic RNAs.

Abbreviations

8FV-cap	8-Fluorenylvinyl cap
8NV-cap	8-Naphthylvinyl cap
8ST-cap	8-Styryl cap
a.a.	Amino acids
APS	Ammonium persulfate
ASO	Antisense Oligonucleotide
ATP	Adenosine Triphosphate
Bac	6-Bromo-7-aminoethoxycoumarin-4-ylmethoxycarbonyl
BG	O^6-Benzylguanine
bGH	Bovine growth hormone
Bhc-diazo	6-Bromo-4-diazomethyl-7-hydroxycoumarin
BSA	Bovine serum albumin
bz	Benzoyl
c-hcrRNA	Circular hybrid crRNA
Cas	CRISPR associated protein
CAR-T	Chimeric antigen receptor (T-cells)
CE	2-Cyanoethyl
circRNA	Circular mRNA
CMV	Cytomegalovirus
^{CNV}K	3-Cyanovinylcarbazole nucleoside
Cpf1	CRISPR-Cas12a
CPG	Controlled pore glass
CpG	Cytosine Guanosine pair
CRISPR	Clustered regularly interspersed palindromic repeats
crRNA	CRISPR RNA
CuAAC	Cu(I)-catalyzed azide–alkyne cycloaddition
DBCO	Dibenzocyclooctyne
dCas9	Catalyically inactivce Cas9
DEACM	[7-(diethylamino)coumarin-4-yl]-methyl
dgRNA	Dual guide RNA
DMEM	Dulbecco's modified Eagle medium
DMNB	3,4-Dimethoxy-2-nitrobenzyl
DMNPE	4,5-Dimethoxy-2-nitrophenylethyl
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT3a	DNA methyltransferase 3 alpha
DPBS	Dulbecco's phosphate-buffered saline
ds	Double stranded
DSB	Double-strand break
EDTA	Ethylenediaminetetraacetic acid
eIF4E	Eukaryotic translation initiation factor 4E
EMSA	Electromobility shift assay
EMX1	Empty spiracles homeobox 1
FAM	6-Carboxyfluorescein
FBS	Fetal bovine serum
FSC	Forward scatter
GalNAc	N-Acetylgalactosamine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GFP1-10	GFP peptide chains 1-10
GFP11	GFP peptide chain 11
GIM	EGFP-IRES-mCherry mRNA
gRNA	Guide RNA
HDR	Homology-directed repair

HEG	Hexaethyleneglycol
HEK	Human embryonic kidney cells
hgRNA	Hybrid gRNA
HPLC	High performance liquid chromatography
hsgRNA	Hybrid sgRNA
HUH	Histidine–hydrophobic residue–histidine motif
ib	Isobutyryl
IEDDA	Inverse-electron-demand Diels–Alder reaction
InDel	Insertion deletion
IRES	Internal ribosome entry site
IVT	*In vitro* transcription
KRAB	Krüppel-associated box
l-hcrRNA	Linear hybrid crRNA
LED	Light emitting diode
M	2′OMe
MEPK	4-Methylpyranocarbazole
METTL	Methyltransferase-like protein
MIG	mCherry-IRES-EGFP mRNA
mMe-2PA	2-Meta-methyl-phenylazo cap
MPACE	2′OMe 3′thioPACE
MPS	2′OMe phosphorothioate
MS	Mass spectrometry
N	Unspecified base
nCas9	Cas9 nickase
NHEJ	Non-homologous end joining
NHS	N-Hydroxysuccinimide
NMI	N-Methylimidazole
nn	Nearest neighbor
NPM	6-Nitropiperonyl-methyl
nt	Nucleotides
ODN	Oligodeoxyribonucleotide
ORF	Open reading frame
ORN	Oligoribonucleotide
OXN	Oligodeoxyribonucleotide or Oligoribonucleotide
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PC	Photo-cleavable linker
^{PC}K	Pyranocarbazole nucleoside
PCR	Polymerase chain reaction
PE	Prime editor
pegRNA	Prime editor guide RNA
PNA	Peptide nucleic acid
preQ$_1$	Pre-queuosine$_1$
R	Purine Base
RIG-1	Retinoic acid-inducible gene 1
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
ROS	Radical oxygen species
RT	Room temperature
RTase	Reverse transcriptase
saRNA	Self-amplifying RNA
SDS	Sodium dodecyl sulfate
sgRNA	Single guide RNA
siRNA	Short interfering RNA
SpyCas9	Streptococcus pyogenes Cas9
ss	Single stranded
SSC	Side scatter
SV40	Simian Virus 40
T2A	Thosea asigna virus 2A self cleaving peptide
tac	Tert-butylphenoxyacetyl

TBE	Tris Borate EDTA Buffer
TBDMS	Tert-butyldimethylsilyl
TC	Thionocarbamate
TCA	Trichloroacetic acid
TEAA	Triethylammonium acetate buffer
TEAB	Triethylammonium bicarbonate buffer
TEMED	Tetramethylethylenediamine
TET1	Ten-eleven translocation methylcytosine dioxygenase
TGT	Transglycosylase
THF	Tetrahydrofuran
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
TIDE	Tracking of Indels by DEcomposition
T_M	Melting temperature
tracrRNA	Trans-activating CRISPR RNA
Tris	Tris(Hydroxymethyl)aminomethane
TSR	Transcriptional state reporter
UPLC	Ultra-performance liquid chromatography
UTR	Untranslated region
WST-8	2-(2-Methoxy-4-nitrophenyl)-3-(4-nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium sodium salt
Y	Pyrimidine Base

Contents

1

Introduction

1.1 Chemical Synthesis of Oligonucleotides

The ability to chemically synthesize DNA and RNA molecules in high yield and purity is the bedrock of modern biological sciences and a testament to the power of organic chemistry. Oligonucleotide synthesis requires the stepwise joining of chemically diverse DNA or RNA monomers at near quantitative yield. Since its introduction by Caruthers in 1981 [1], phosphoramidite chemistry has been the gold standard for accomplishing this. Paired with advancements in solid-phase chemistry [2, 3], this approach yielded the classic 4-step synthesis cycle [4], which facilitated the automation and widespread accessibility of oligonucleotide synthesis.

The availability of low-cost custom oligonucleotides through chemical synthesis is a necessity for critical technologies such as PCR [5], DNA sequencing [6], and gene editing [7], which have become fundamental to biological sciences. The abundance of synthetic oligonucleotides and the highly predictable nature of their interactions make new

technologies viable outside of traditional biological applications, such as the assembly of DNA nanomaterials [8]. These principles are also relevant to biomedical applications of synthetic oligonucleotides, which are increasingly used in new classes of diagnostic and therapeutic tools [9]. Oligonucleotide-based technologies are further enhanced by developments in nucleic acid chemistry, which make it routinely possible to modify the phosphate linkages, sugars, and bases of oligonucleotides to alter their properties [10].

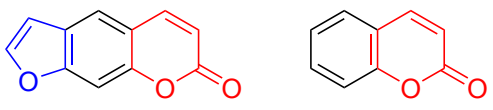
The explosion of nucleic acid-based technologies has created many opportunities and challenges which require new ways of working. In this chapter, I will introduce the 3-cyanovinylcarbazole nucleotide, a photo-crosslinking agent which can be used to manipulate the behavior and physical properties of oligonucleotides. I will then discuss two therapeutic applications of synthetic RNAs which could be improved with crosslinking agents: CRISPR-Cas9 gene editing and synthetic mRNAs. The research goals of my thesis will then be presented as they relate to crosslinking and these two applications.

1.2 Photo-Crosslinking

1.2.1) Crosslinking Stabilizes Oligonucleotide Duplexes

Interactions between oligonucleotides are largely defined by non-covalent Watson-Crick base pairing. One of the several benefits of oligonucleotide class drugs and nanomaterials is that base pairing is highly predictable, making it simple to design interactions between oligonucleotides. However, the strength of interaction between two oligonucleotides is strongly influenced by temperature, ion concentration, and length of oligonucleotide duplex. Accordingly, several oligonucleotide applications benefit from the formation of covalent interactions within a duplex to stabilize its interactions.

Crosslinking is a widely implemented technique to stabilize DNA and RNA duplexes by covalent-bond formation. Several classes of crosslinking agents have been developed that can crosslink to naturally occurring bases spontaneously or when catalyzed by ROS or UV light [11]. Of these strategies, photo-crosslinking is the most popular as it can be induced with a highly controllable and bio-orthogonal signal: light. Photo-crosslinkers such as psoralen **(Figure 1.1)** can produce inter-strand DNA crosslinks *via* the [2 + 2] cycloaddition reaction when irradiated with UV-A (365 nm) light. Crosslinking occurs with the C5-C6 bond of pyrimidine bases *via* either the pyrone or furan photoreactive sites of psoralen [12, 13]. The resulting crosslink can be photo-split by irradiation with shorter 254 nm light [14]. Psoralen can be applied to crosslink oligonucleotide duplexes either as a free agent, in which case it will indiscriminately target TA or TU sequences, or it can be incorporated into oligonucleotide sequences to allow for sequence-specific photo-crosslinking. Despite the convenience of photo-catalyzed and reversible crosslinking, psoralen-derived nucleosides are limited by their sequence requirement and the photodamage they cause to DNA [15]. The reverse crosslinking wavelength (254 nm) is particularly problematic as it can catalyze the formation of pyrimidine dimers [16]. Other popular crosslinking agents, such as the structurally related coumarin **(Figure 1.1)**, have similar limitations [11].

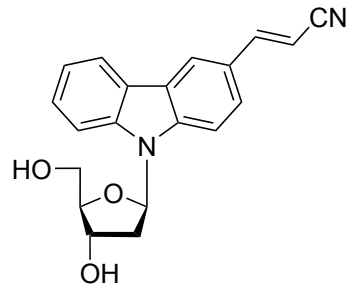


psoralen coumarin

Figure 1.1: Chemical structures of photo-crosslinking agents psoralen (left) and coumarin (right). Photo-reactive pyrone and furan groups are colored red and blue respectively.

1.2.2) 3-Cyanovinylcarbazole Nucleotide

To improve the versatility of photo-crosslinking agents, Fujimoto and colleagues described the synthesis of the 3-cyanovinylcarbazole nucleoside (^{CNV}K)**(Figure 1.2)**[17]. Similar to psoralen and coumarin, ^{CNV}Ks can crosslink to the C5-C6 bond of pyrimidine bases *via* the [2 + 2] cycloaddition reaction catalyzed by 365 nm radiation **(Figure 1.3a)**. However, the ^{CNV}K/pyrimidine crosslink can be photo-split with 312 nm light without causing DNA damage [17, 18]. ^{CNV}K crosslinking is also incredibly site-specific: when incorporated in duplexes, the ^{CNV}K nucleotide will selectively crosslink pyrimidine bases in the n+1 position on the opposite strand **(Figure 1.3b)**. Importantly, this reaction does not require a specific sequence other than an available pyrimidine, and the rate of crosslinking is largely unaffected by the identity of neighboring bases [17]. This feature makes ^{CNV}K much more versatile and easier to implement than the sequence-restricted psoralen crosslinking.



3-cyanovinylcarbazole nucleoside

Figure 1.2: Chemical structure of the 3-cyanovinylcarbazole nucleoside.

^{CNV}Ks are capable of crosslinking to the three most commonly occurring pyrimidine bases: thymine, cytosine, and uracil [17, 18], but not at equal rates. A side-by-side analysis of ^{CNV}K crosslinking in a 9 bp DNA duplex revealed that the rate of ^{CNV}K/T crosslinking was 26.9 fold higher than ^{CNV}K/C crosslinking [19]. This discrepancy is likely due in part to the

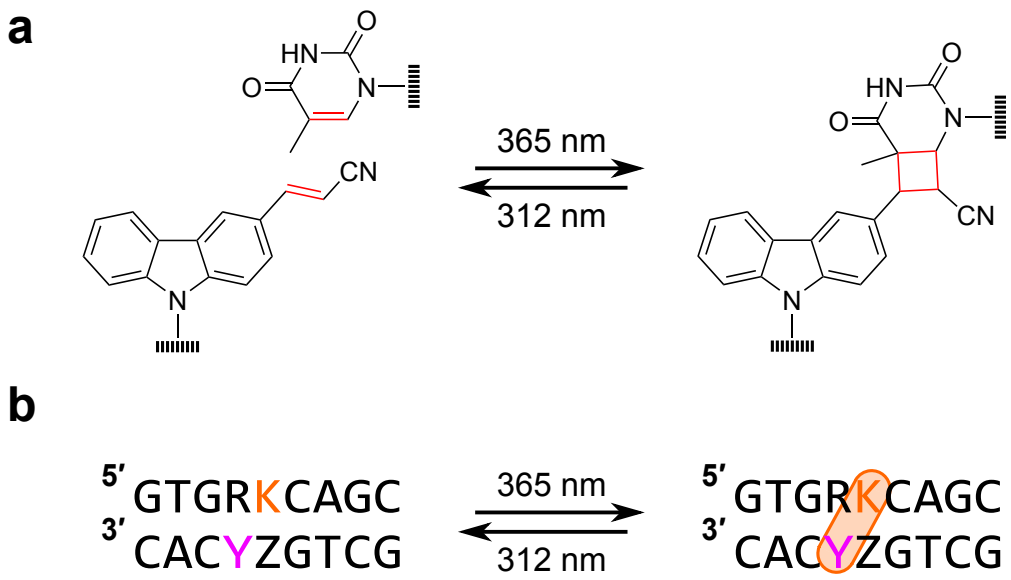


Figure 1.3: Reversible photo-crosslinking of a 3-cyanovinylcarbazole and thymine. **a.)** When irradiated with 365 nm light, the vinyl group of the ^{CNV}K reacts with C5-C6 bond of thymine via a [2 + 2] cycloaddition to form a cyclobutane ring. Crosslinking can be reversed with 312 nm irradiation. **b.)** A ^{CNV}K residue (K) can crosslink to pyrimidine residues (Y) in the n+1 position of hybridized oligonucleotide strands. Crosslink is indicated by orange circle between K and Y.

increased hydrogen bonding at target C residues, as substitution of its complementary base with non-canonical purine bases or carbon linkers dramatically increases the rate of crosslinking [20]. The ^{CNV}K/T crosslinking rate in dsDNA duplexes is also 2.2 fold greater than the ^{CNV}K/U crosslinking rate in comparable heteroduplexes [19]. However, a comparison of ^{CNV}K and psoralen crosslinking to uracil in a 25 bp heteroduplex revealed that the rate of ^{CNV}K crosslinking is nearly 20-fold greater than psoralen [21], making the ^{CNV}K residue more appropriate for use in RNAs. In short heteroduplexes, ^{CNV}K/U pairs have been demonstrated to crosslink to 94% completion within 1 s of irradiation with a 366 nm 1,600 mW/cm^2 LED [18].

Molecular modeling studies suggest that the crosslinking between ^{CNV}K and target pyrimidines is facilitated by base stacking, positioning the vinyl group of the ^{CNV}K in the proximity of the C5-C6 bond of the pyrimidine [22]. The vinyl group itself is capable of

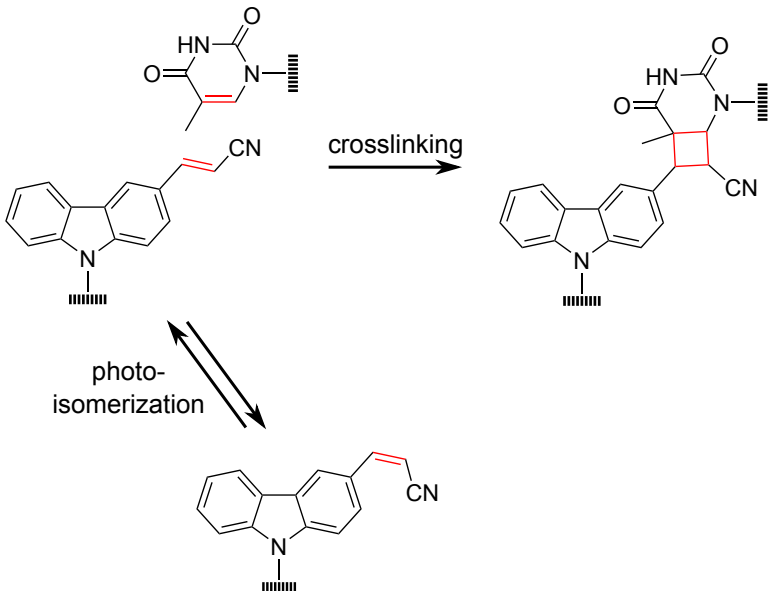


Figure 1.4: ^{CNV}K photo-crosslinking and photo-isomerization. When irradiated with 365 nm light, the ^{CNV}K cyanovinyl group can undergo *trans-cis* isomerization. The *cis* isomer does not participate in crosslinking.

underdoing *trans-cis* photoisomerization when irradiated with 365 nm light. NMR analysis indicates that only the *trans* isomer participates in the crosslinking reaction. However, the high reaction rates of crosslinking and isomerization favor the *trans* isomer over the *cis*, which has no impact on the overall reaction **(Figure 1.4)** [23]. It is interesting to note that while ^{CNV}K crosslinking is most effective in duplexes, it has also been applied to stabilize the structures of DNA triplexes [24] and quadruplexes [25] *in vitro*.

Melting curve analysis has demonstrated that the incorporation of ^{CNV}K residues in an oligonucleotide duplex has a negative effect on duplex stability. Upon crosslinking, this effect is reversed, massively increasing the T_M of crosslinked duplexes [17, 18, 26]. In addition to increased stability in high temperatures, the introduction of crosslinks has also been demonstrated to improve duplex resistance to phosphodiesterase in some scenarios [27].

1.3 Applications of 3-Cyanovinylcarbazole Nucleotides

1.3.1) Application of ^{CNV}Ks in DNA Nanostructures

One of the primary motivations for the development of oligonucleotide crosslinking agents was to convert the non-covalent interactions of duplexes into covalent ones. To this end, the ^{CNV}K nucleoside and other crosslinking agents have been widely applied in the assembly of DNA nanostructures. Introducing ^{CNV}K crosslinks into 2D DNA lattices has been demonstrated to increase the heat resistance of these structures by 40 °C [28, 29], massively improving the range of conditions within which DNA nanostructures can exist. This effect has been applied as the basis for a self-replication and selection strategy for DNA nanostructure evolution [30]. Because ^{CNV}K crosslinking is largely unaffected by buffer and does not require a co-factor, it can also be applied in a wider array of environments than other crosslinking agents [31].

^{CNV}K-oligonucleotides, when conjugated to molecular cargo, can be used to organize and fix the cargo into patterns based on oligonucleotide interactions. Our group demonstrated this with the construction of gold nanoparticle-conjugated DNA crosslinked into lattices [32, 33] which were stable under denaturing conditions. Similarly, lipid-conjugated ^{CNV}K-oligonucleotides have been applied to manipulate liposomes *in vitro* through crosslinking, creating behaviors such as liposome aggregation [34] and endosome formation [35].

1.3.2) Application of ^{CNV}Ks in Oligonucleotide Probes

The robust and activatable covalent interactions formed between ^{CNV}K and target bases have been widely applied to improve the activity of probes to identify and capture oligonucleotides. As a general principle, the covalent interactions formed through ^{CNV}K crosslinks can be used to replace stretches of non-covalent base pairing which would

otherwise be required for duplex stability. This allows the length of oligonucleotide probes to be reduced, which in turn increases their specificity [36].

Crosslinking between oligonucleotide duplexes and ^{CNV}K-probes favors hybridization with the probe, which encourages strand displacement and can force branch migration to occur [37]. This phenomenon has been applied to design DNA duplex invading probes whose activity can be controlled with light [38, 39], as opposed to traditional PNA-based duplex invading probes.

When conjugated to fluorescent moieties, ^{CNV}K probes can be used to identify oligonucleotide targets of interest. This has been applied to fluorescence *in situ* hybridization labeling of RNA, which improved the labeling of highly structured RNAs and reduces the loss of signal after washing steps [40]. The impact of ^{CNV}K crosslinking on duplex hybridization has also been applied to improve the target discrimination and signal of molecular beacon probes [23, 41].

1.3.3) ^{CNV}K Crosslinking Allows High-Precision Spatial Targeting

Because ^{CNV}K crosslinking is controlled by light, regulating the spatial activation of ^{CNV}Ks is limited only by the resolution of the light source. Advances in photolithography have made targeted irradiation with light less of a barrier to research, making photo-crosslinking an attractive method for capturing or functionalizing oligonucleotides with μm-scale resolution. This is underscored by two recently described ^{CNV}K applications: Action-PAINT [42], a method for simultaneous super-resolution imaging and labeling of biological materials, and Light-Seq [43], which applies cell-specific illumination to capture oligonucleotides for RNA-seq in complex tissues.

1.3.4) ^{CNV}K Translation Inhibiting Oligonucleotides

The majority of techniques described earlier are performed *in vitro* or *ex vivo*. However, one of the largest advantages of the ^{CNV}K compared to psoralen- or coumarin-derived compounds is its rapid crosslinking rate, which generally means that less UV light is required for crosslinking. Because of this, ^{CNV}K crosslinking is much more amenable to application in cells than other photo-crosslinking agents. This is especially relevant for therapeutic oligonucleotides such as siRNA and ASOs, whose activity could theoretically be regulated through photo-crosslinking. This point was first raised by Shigeno *et al*., who demonstrated that 15-28 nt ^{CNV}K-ASOs could crosslink to full-length human K-Ras mRNA and inhibit its translation *in vitro* [44]. Importantly, the inhibitory effect was crosslink-dependent, likely because non-crosslinked ^{CNV}K residue would otherwise destabilize ASO interactions.

This work was followed up by Sakamoto *et al*., who applied phosphorothioate ^{CNV}K-ASOs to regulate the expression of GFP in a GFP-stable HeLa cell line [45]. After UV irradiation, the IC_{50} value of each ^{CNV}K-ASO was lower than their non-^{CNV}K control, suggesting that the stable crosslinking interaction improves their inhibitory effect. Importantly, no UV-mediated damage to DNA was observed throughout these experiments. A direct comparison of ASO-mediated gene silencing in the same cell line showed that a ^{CNV}K-ODN was much more effective at inhibiting GFP expression after 10 s 365 UV irradiation than an isosequential psoralen-ODN (77±10% and 7±6 inhibition respectively) [21]. Although it has yet to be demonstrated, a chemically stable ^{CNV}K-ASO could theoretically be regenerated *in vivo* due to its photo-splitting capabilities.

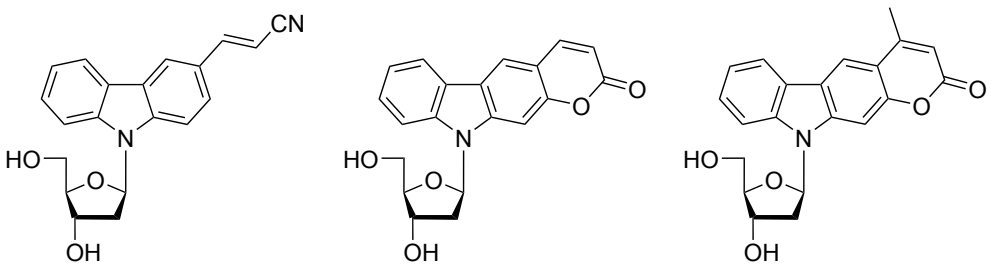
1.3.5) Dynamic Regulation of Functions by Reversible Crosslinking

Despite the varied uses for ^{CNV}K nucleosides, few applications make full use of its *reversible* photo-crosslinking, as referenced above. Some examples were discussed previously, such as the acquisition and recovery of oligonucleotides using ^{CNV}K-probes [36]. In these instances, recovery (i.e., reverse crosslinking) can be performed outside of the biological or biochemical system in question, which allows for more flexible reverse crosslinking conditions and UV irradiation durations. The reversible crosslinking of ^{CNV}Ks has been applied to short DNA structures which will adopt a hairpin structure when crosslinked and a linear structure when non-crosslinked. This has been used as the basis of light-controlled DNAzymes [46] and DNA logic chips [47], both of which can be activated through reverse crosslinking. However, these structures have only been applied *in vitro*; ^{CNV}K reverse crosslinking and related strategies remain largely unexplored in cells.

1.4) Carbazole Nucleotides Present New Opportunities

1.4.1) Alternative Carbazole Crosslinking Agents

In the past few years, Fujimoto and colleagues have developed variants of the ^{CNV}K moiety which function via a similar [2+2] cycloaddition mechanism. One success was the synthesis



3-cyanovinylcarbazole (^{CNV}K) | pyranocarbazole (^{PC}K) | 4-methylpyranocarbazole (^{MEP}K)

Figure 1.5: Chemical structures of photo-crosslinking 3-cyanovinylcarbazole and pyranocarbazole nucleosides.

of the pyranocarbazole nucleoside (^{PC}K)**(Figure 1.5)**, which can be selectively crosslinked to adjacent bases with visible light irradiation (between 400-450 nm) [48]. Similar to the ^{CNV}K moiety, ^{PC}K can crosslink with pyrimidine bases with a preference for thymine over cytosine residues (90 and 20% crosslinked respectively in a 9 nt duplex, irradiated for 10 s 400 nm).

The crosslinking reaction between ^{PC}K and thymine similarly produces a photo-adjunct linked by a cyclobutane ring **(Figure 1.6)** [48]. ^{PC}K crosslinks can be photo-split by 312 nm irradiation, but up to 65% photo-splitting can also be achieved with 330 nm irradiation. The increased photoactivity of ^{PC}K compared to ^{CNV}K are hugely exciting for biological applications of photo-crosslinking nucleosides, as these higher wavelengths are generally less toxic to cells [16].

Despite its favorable photo-active properties, the ^{PC}K nucleoside has yet to be used in any biological or biochemical application; this is partially due to the incredibly low yield of its published synthesis route. For this reason, Mihara and Fujimoto published the synthesis of a 4-methylpyranocarbazole nucleoside (^{MEP}K)**(Figure 1.5)** three years after their work on ^{PC}Ks [49]. Methylation of the pyranocarbazole suppressed by-product formation during synthesis,

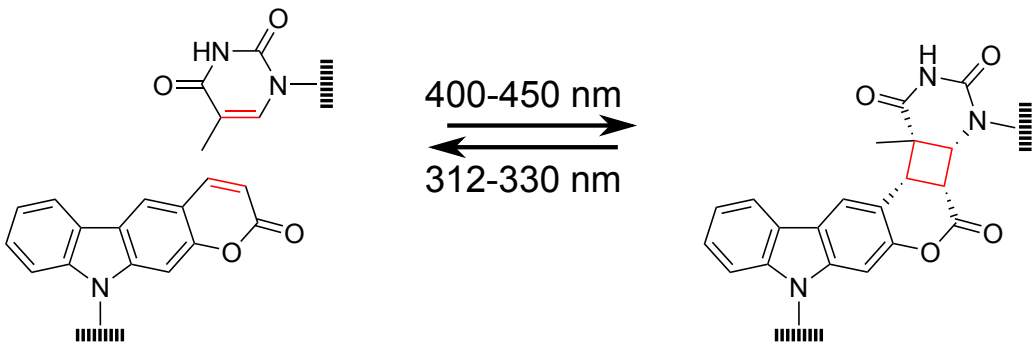


Figure 1.6: Reversible photo-crosslinking of a pyranocarbazole and thymine. When irradiated with visible light, the ^{PC}K pyrone group reacts with C5-C6 bond of thymine via a [2 + 2] cycloaddition to form a cyclobutane ring. Crosslinking can be reversed with 312-330 nm irradiation.

increasing the synthetic yield of ^{MEP}K compared to ^{PC}K. ^{MEP}K can also be crosslinked at 400 nm and photo-split at 330 nm. Inconsistencies in reporting make it difficult to discern whether ^{MEP}K has a comparable photo-splitting rate to ^{PC}K, however. Notably, ^{MEP}K is entirely unable to crosslink cytosine residues, making it selective for thymine bases [49].

Substitution of the rigid deoxyribose sugar in either ^{CNV}K or ^{PC}X with a more flexible D-threoninol linker improves the photo-crosslinking rate of each moiety, [19, 50]. The improved crosslinking rate was attributed to the increased flexibility of the linker, which minimizes the entropic loss during duplex formation caused by the carbazole group. Inhibitory 3-cyanovinylcarbazole threoninol oligonucleotides have also been applied inhibitory oligonucleotides to regulate mRNA expression [21].

1.4.2) The Untapped Potential of Carbazole Nucleotides.

The ^{CNV}K nucleotide is a powerful photo-crosslinking agent which can be applied to manipulate the behavior of oligonucleotides. Accordingly, it has been applied to control or improve the activity of oligonucleotides in many applications, such as DNA nanostructures and oligonucleotide probes. Owing to their ease of use, ^{CNV}Ks have even been applied in DNA-encoded chemical libraries [51] and cDNA display assays [52], where its function pertains to information storage and tagging as opposed to oligonucleotide manipulation. Despite these varied uses, ^{CNV}Ks have not found widespread application in *in vivo* systems apart from improving ASO activity. Moreover, the dynamic nature of its reversible crosslinking has not been widely implemented. The recent development of novel carbazole nucleosides with faster crosslinking rates and higher photo-splitting wavelengths might see wider use, but this has yet to be the case. Therefore, the central question of this thesis was

whether these crosslinking agents could be used to address some of the challenges currently facing the field of therapeutic oligonucleotides.

1.5) CRISPR Cas9

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein 9 (Cas9) is an RNA-guided DNA endonuclease. When provided with a short guide RNA (gRNA), Cas9 can identify and catalyze the cleavage of matching DNA sequences. The target sequence is determined by a 20 nt "spacer" sequence encoded at the 5′ terminus of the gRNA. Potential target sequences are only limited by a protospacer adjacent motif (PAM), a short sequence that is required for DNA interrogation by Cas9. Several Cas9 orthologs have been identified and coopted for biotechnological applications, but the most commonly used variant is the 1388 a.a. *Streptococcus pyogenes* Cas9 (SpyCas9), which has a minimally restrictive "NGG" PAM sequence required 3′ of putative target sites [53].

The natural Cas9 gRNA is composed of two separate RNAs: a 42 nt CRISPR RNA (crRNA) which contains the 20-nt spacer sequence, and an 89 nt trans-activating crRNA (tracrRNA)

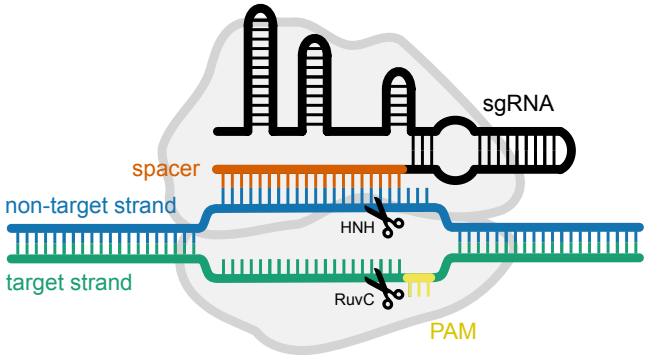


Figure 1.7: Simplified diagram of Cas9-mediated DNA cleavage. When complexed with a gRNA (black, in this case, a sgRNA), Cas9 (grey) will interrogate DNA substrates for NGG PAM motifs (yellow). Recognition of the PAM motif will trigger strand invasion: the sgRNA spacer sequence (orange) will hybridize with the non-target DNA strand (blue), which is complementary to the PAM-bearing target strand (green). Sequence homology between the spacer sequence and target will trigger dsDNA cleavage: the Cas9 NHN nuclease domain cleaves the non-target strand, and the RuvC nuclease domain cleaves the target strand.

which partially hybridizes with the crRNA and provides essential structural motifs. Importantly, the crRNA/tracrRNA dual gRNA (dgRNA) can be synthesized as one molecule to yield a single gRNA (sgRNA). The Cas9/gRNA ribonucleoprotein (Cas9-RNP) cleaves the DNA target strand (PAM-containing strand) and non-target strand (complementary to spacer sequence) between positions 3 and 4 upstream of the PAM **(Figure 1.7)**, yielding blunt-ended DNA fragments. The catalytic HNH and RuvC domains of Cas9 can be inactivated through single point mutations (H840A and D10A respectively) to yield single-stranded Cas9 nickases (nCas9) or catalytically inactive Cas9 (dCas9), which still retains its DNA binding capabilities [54].

Originally components of the bacterial acquired immunity system, Cas9 and related nucleases have since been repurposed for the targeted cleavage of DNA [54, 55] and to enact genome edits [56, 57]. Owing to the ease with which Cas9 can be programmed through simple manipulation of its gRNA, Cas9-derived technologies are significantly more accessible and popular than preceding gene-editing tools such as meganucleases [58], ZFNs [59], and TALENs [60]. The discovery of Cas9, its role in bacterial immunity, related Cas nucleases, and its repurposing for gene editing have been extensively reviewed elsewhere [61-63].

1.6) Cas9 Genome Targeting and Cleavage

The RNA-guided cleavage of DNA by Cas9 is a stepwise process that involves (1) Cas9 binding to gRNA, (2) Cas9-RNP recognition of PAMs on substrate DNA (3) strand invasion and Cas9/gRNA/DNA complex formation resulting in (4) cleavage of DNA.

1.6.1) Cas9/gRNA Complex Formation

The crystal structure of SpyCas9 in the absence of gRNA (apo-Cas9) reveals that the protein has a bi-lobe architecture, consisting of a target recognition (REC) lobe and a nuclease (NUC) lobe **(Figure 1.8a)** [64]. The HNH and RuvC nuclease domains which catalyze target cleavage are located within the NUC lobe. The C-terminal domain (CTD), which is critical for PAM recognition, is also encompassed within the NUC lobe.

The key effects of Cas9/sgRNA complex formation are best illustrated by the crystal structures of Cas9-RNP **(Figure 1.8b)**[65]. In the absence of a gRNA, apo-Cas9 has a

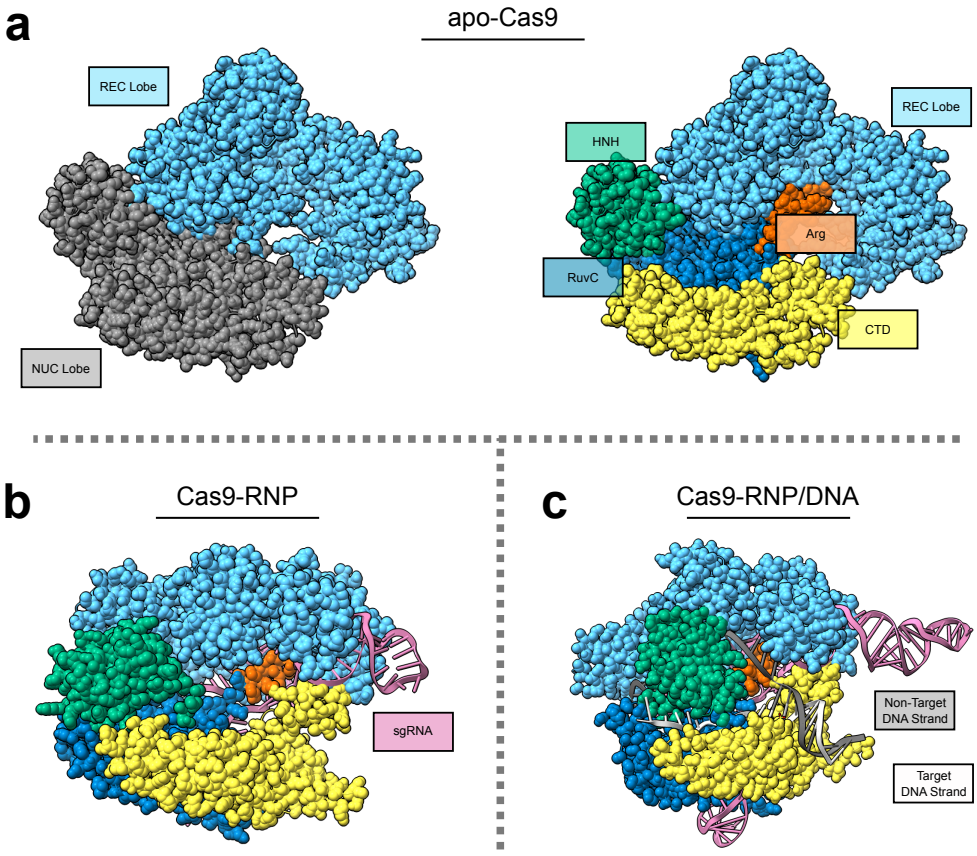


Figure 1.8: Structures of Cas9 in various conformations. apo-Cas9 **a.)** adopts the inactive conformation (PDB: 4CMP). Upon binding a gRNA substrate, **b.)** Cas9 enters into a DNA-recognition competent conformation (PDB: 4ZT0). The Cas9-RNP complex undergoes a final conformational change upon DNA target binding, **c.)** enabling Cas9 to form an active conformation for DNA target cleavage (PDB: 5F9R).

disordered CTD which prevents PAM recognition and target cleavage. Instead, apo-Cas9 will indiscriminately bind and release substrate DNA until it complexes with a gRNA [66]. Upon Cas9/gRNA complexing, a prominent conformational change occurs in the REC lobe, shifting it into a functional state. Additionally, the PAM-interacting module of the CTD is repositioned for DNA interaction. The large structural changes which occur when Cas9 complexes with a gRNA stand in contrast to the formation of the Cas9-RNP/DNA complex **(Figure 1.8c)**, which undergoes a much smaller conformation change, and confirms the gRNA loading step as a key regulator of Cas9 enzymatic function [64].

1.6.2) PAM Recognition

Regardless of the homology between the gRNA spacer sequence and target DNA, Cas9 is unable to cleave DNA in the absence of the PAM. In endogenous systems, this mechanism is crucial for Cas9's ability to discriminate between self (CRISPR array) and non-self (viral) DNA [67]. Consequentially, any mutations in the PAM of viral targets will prevent recognition by Cas9 and improve virus/plasmid fitness [68].

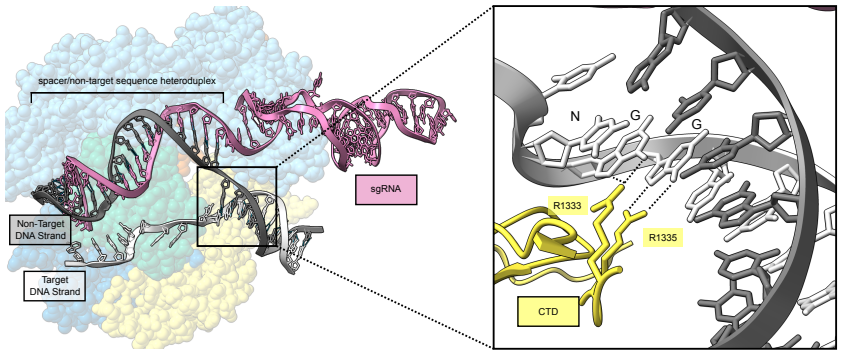


Figure 1.9: PAM recognition and gRNA/non-target DNA heteroduplex formation by Cas9-RNP. gRNA (pink), non-target strand (grey), and target strand (white) are overlaid against Cas9 protein where they would appear within the structure. The key residues for PAM recognition are R1333 and R1335 within the CTD domain (yellow), which forms hydrogen bonds with the two guanine residues in the NGG PAM motif (PDB: 5F9R).

Single-molecule experiments have demonstrated that DNA target search is initiated by Cas9 probing substrate DNA for suitable PAMs. Target recognition occurs through random three-dimensional collisions in which Cas9 rapidly dissociates from substrates lacking an appropriate PAM [66].

Comparison between the structure of Cas9-RNP and Cas9-RNP bound to a PAM duplex demonstrates that the GG dinucleotide of the SpyCas9 PAM is read by two key residues (R1333 and R1335) in the CTD through hydrogen bond interactions with the substrate major groove [69] **(Figure 1.9)**. The CTD makes additional hydrogen-bonding interactions with the backbone of the PAM-containing nontarget strand, however no direct contacts are been observed between Cas9 and the nucleotides complementary to the PAM. Accordingly, mismatched bases and mutations are tolerated in the target strand [54, 66].

1.6.3) RNA Strand Invasion

After PAM recognition, Cas9 residues K1107 and S1109 interact with the phosphate group immediately 5′ of the PAM on the non-target strand to form a phosphate lock loop [69]. This interaction creates a kink-turn in the target strand upstream of the PAM, which causes local DNA melting and facilitates gRNA invasion and hybridization to the non-target DNA strand [64]. The target DNA strand is displaced, and its PAM-proximal region is stabilized by base stacking onto the PAM duplex [70], resulting in R-loop formation [71]. Without the local melting caused by PAM recognition, the first nucleobase of the target DNA sequence cannot be readily flipped and accessed by the gRNA, and as such the gRNA would rarely be able to bind substrate DNA and initiate strand invasion.

1.6.4) Target DNA Cleavage

Formation of the gRNA/DNA hybrid is the final event required for substrate cleavage. The RNA-DNA hybrid lies in the central channel between the REC and NUC lobes and is recognized by Cas9 in a sequence-independent manner [72], indicating that Cas9 recognizes the geometry of the RNA/DNA hybrid as opposed to its nucleobases.

However, analysis of Cas9-RNP/DNA interactions by ChIP-seq suggests that DNA binding events are much more promiscuous than DNA cleavage. Although Cas9-RNP is capable of binding many off-target sequences, only a fraction of these will be cleaved [73-75], suggesting that DNA binding alone is not sufficient to trigger DNA cleavage. Rather, extensive base-pairing between gRNA and DNA is required to trigger DNA cleavage activity [76]. Specifically, the interactions at the 14th through 17th bases of the spacer sequence are critical for stabilizing the active conformation of Cas9 [77].

Comparisons of Cas9-RNP with and without a DNA substrate **(Figure 1.8 b,c)** reveal that further conformational changes occur in Cas9 upon target binding. The HNH nuclease domain moves toward the target DNA strand and is believed to position its active site within 10 Å from the scissile target phosphate. The HNH and RuvC domains communicate through two hinge linkers. The reorientation of the HNH domain causes a rearrangement of the linkers, which in turn causes the displacement of the non-target DNA strand into proximity of the RuvC active site. Both strands of the substrate DNA are then simultaneously cleaved between the 3rd and 4th bases upstream of the PAM in a metal-ion-dependent manner, generating a blunt-ended double-strand break (DSB) [70].

1.7) gRNA Structure and Interactions

The Cas9 gRNA can be divided into 6 elements on the basis of each part's structure or function. At the 5′ terminus is the 20-nt spacer sequence, which specifies DNA target and participates in target hybridization. 3′ of the spacer sequence is the repeat-antirepeat duplex. In a dgRNA, the crRNA "repeat" sequence hybridizes with the tracrRNA "anti-repeat", forming the duplex through which the two RNAs interact. In a sgRNA, the repeat-antirepeat duplex is covalently linked by a 4 nt tetraloop. Downstream of this duplex are stem-loop 1, a short linker sequence, stem-loop 2, and stem-loop 3 **(Figure 1.10)**.

Cas9 forms extensive direct contacts with the repeat-antirepeat duplex and stem-loop 1, as well as the linker between stem-loops 1 and 2. In contrast, Cas9 makes significantly fewer contacts with stem-loops 2 and 3 [70, 72]. While *in vitro* assays have demonstrated that sgRNAs missing the linker and stem-loops 2 and 3 retain their ability to trigger Cas9-mediated DNA cleavage [54], functional studies have shown that these elements are required for robust Cas9 activity *in vivo* [72]. These observations suggest that the repeat-antirepeat duplex and stem-loop 1 are requisite for Cas9-RNP formation, whereas the other elements

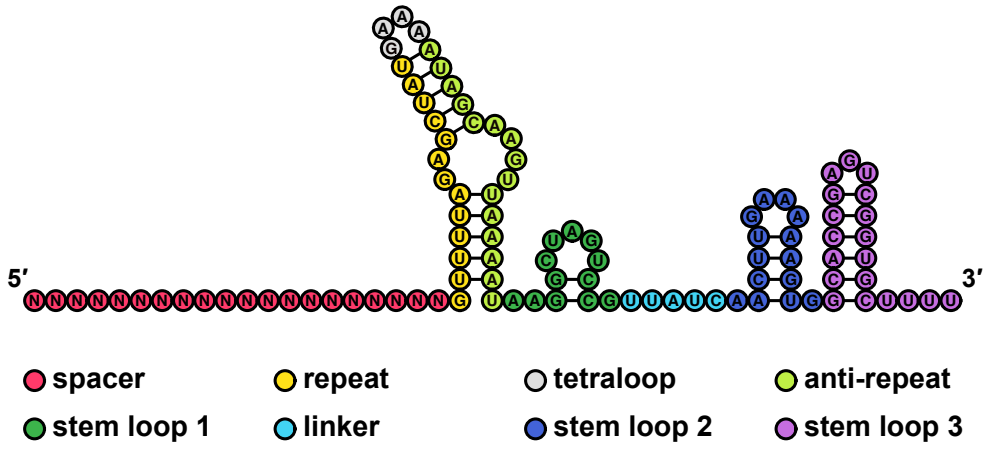


Figure 1.10: Secondary structure of SpyCas9 sgRNA. The variable spacer sequence (red) is represented with 20 'N' nucleotides.

may not be required, but do stabilize Cas9/gRNA interactions to promote active complex formation and increase catalytic efficiency [56, 78-80].

The 20-nt spacer sequence can be further divided into two contiguous halves: the "seed" region, which hybridizes with the 10-12 bases immediately upstream of the PAM in the non-target strand, and the remaining "non-seed" which hybridizes to the PAM-distal sequence. During target recognition, the A-form structure of the seed-region heteroduplex forms many direct contacts with Cas9. The negatively-charged backbone of the seed-region duplex is accommodated in a tight positively-charged channel formed between the NUC and REC lobes. Mismatches between the spacer seed region and DNA substrate will perturb the A-form helix structure, inhibiting further Cas9 binding to the substrate. In contrast, the heteroduplex formed by the non-seed region and DNA substrate has more conformational flexibility, and therefore mismatches between the spacer and substrate are more tolerated in this region. Consequently, off-target Cas9 binding and cleavage events are more commonly observed in sequences that mismatch in the non-seed region. [65, 75, 81, 82]. Truncated gRNAs with 17-18 nt long spacer sequences can interestingly retain DNA cleavage activity while greatly reducing off-target effects [83].

In the early days of Cas9 reprogramming and gene editing, several variations of the SpyCas9 dgRNA and sgRNA were used for initial DNA cleavage and gene editing experiments, to which much of the knowledge on the differential behavior of truncated gRNAs *in vitro* and *in vivo* can be attributed. The most commonly used sgRNA for effective gene editing *in vivo* was first described by George Church's group [57], and is pictured in **(Figure 1.10)**. It is interesting to note that due to the variable nature of the spacer sequence, non-canonical interactions can form between the spacer sequence and structural elements of the gRNA,

which hinders Cas9 activity [84, 85]. Because of this, attempts have been made to redesign the sgRNA stem-loops to remove unwanted internal gRNA interactions [86]. Despite this issue, the Church sgRNA is appropriate for most applications and remains the most popular gRNA sequence to this day.

1.8) Cas9 Genome Editing Strategies

Cas9 and related nucleases provide an unprecedented means to target individual loci within complex genomes. Accordingly, several strategies have been developed to leverage their targeting capabilities to produce desired genome edits. The most basic of these are Cas9 mediated gene knockouts, in which Cas9 is used to remove genetic information at desired loci. While this is not traditionally considered to be gene editing, it is a useful and versatile technique for genome manipulation. True gene editing with Cas9 can be achieved by providing a donor DNA sequence to template the repair of double-strand breaks (DSB) and introduce new information through homology-directed repair (HDR). While HDR-mediated gene editing is a powerful technique for installing sequences ranging from a single nucleotide to whole genes in size, it is inefficient and prone to mutagenesis [87]. Attempts to improve gene editing by fusing DNA-manipulating proteins to catalytically inactive Cas9 have yielded several impactful technologies, including Cas9 base editors [88, 89], transposases [90], and recombinases [91, 92]. Most of these have been superseded by the recently described prime editing [93], which will be discussed in more depth below.

1.8.1) Gene Knockouts with Cas9 Nuclease

Cas9-mediated DNA targeting can be used to introduce novel material into genomes through several pathways. Typically, these processes are initiated by the Cas9-catalyzed formation of DSBs in genomic DNA. DSBs are hugely cytotoxic and lead to the activation of the cell's

endogenous repair machinery [94]. Several factors influence which pathway is used by a cell [95]; in fully differentiated mammalian cells, DSBs are largely repaired through the error-prone non-homologous end joining (NHEJ) pathway (reviewed in [96]). Accurate repair through NHEJ regenerates the Cas9 target sequence, which then remains a substrate for cleavage. NHEJ can also result in the insertion or deletion (InDel) of nucleotides around the break site, which prevents recognition by Cas9 **(Figure 1.11)**[97].

If Cas9 is targeted to an open reading frame, InDel mutations produced through NHEJ frequently result in frameshift mutations that invalidate protein function, a method known as "gene knockout". While this may seem like a random and undesirable effect, the ability to create targeted gene knockouts with Cas9 is hugely useful as a research tool and is widely used to investigate the function of genes [98-100] and engineer plant lines [101-103]. This

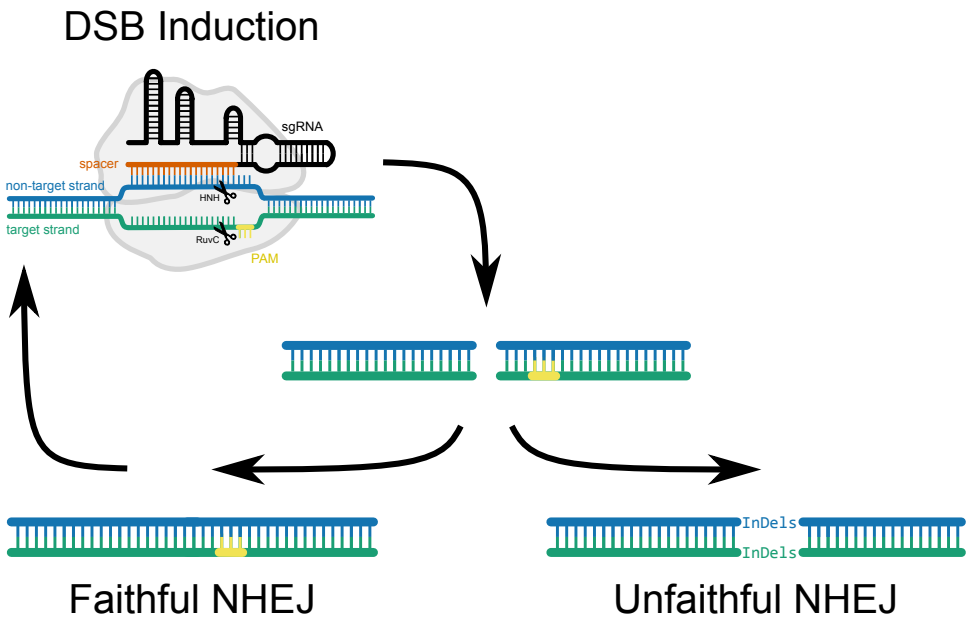


Figure 1.11: Simplified diagram of Cas9-mediated gene knockout. Cas9-RNP recognizes target DNA and creates a DSB in genomic DNA. DSBs are commonly resolved through NHEJ, which can faithfully or unfaithfully repair the damage. Faithful repair of a DSB by NHEJ restores the original Cas9-RNP target, which can be recognized and cleaved by Cas9 again. Unfaithful repair of DSB results in the formation of InDel mutations at the site of cleavage, destroying the target sequence.

capacity is further enhanced by the ease with which Cas9 gene knockouts can be multiplexed by introducing multiple gRNAs to cells at once [56, 57, 104].

1.8.2) Gene Editing with Cas9 Nuclease

The resolution of DSBs by NHEJ can lead to a loss of genomic information and the accumulation of mutations over time. DSBs can also be repaired through the more faithful homology-directed repair (HDR) pathway, in which a homologous DNA sequence to the DSB break locus is used to template repair [105]. HDR has been leveraged to introduce new genetic material into human cells for nearly three decades [106], but the accuracy and flexibility with which Cas9 can generate DSBs have now made it a routine technique. To perform Cas9 gene editing, a custom DNA sequence is prepared that contains an insert of interest flanked by sequences homologous to either side of the Cas9-induced DSB. Optimal donor template design for HDR varies by application, but ssDNA, PCR products, and plasmids are all appropriate to use [107]. HDR-mediated Cas9 gene editing can install a variety of genome edits, including point mutations, precise insertions and deletions, and integration of genome-sized sequences **(Figure 1.12)** [108].

Cas9 gene editing is routinely used in the laboratory setting and has already been translated into therapeutics [109], such as the development of FDA-approved CAR-T cell therapies for hematologic cancer s [110]. However, in addition to the ubiquitous risk of off-target DNA cleavage events creating undesired mutations, the efficiency of Cas9 gene editing is hindered by two factors: HDR is generally active only in dividing cells and requires genes expressed predominately in the S and G2 cell-cycle phases [111-113], and is therefore a rare event in fully differentiated cells, which are generally averse to gene editing. Secondly, nuclease-

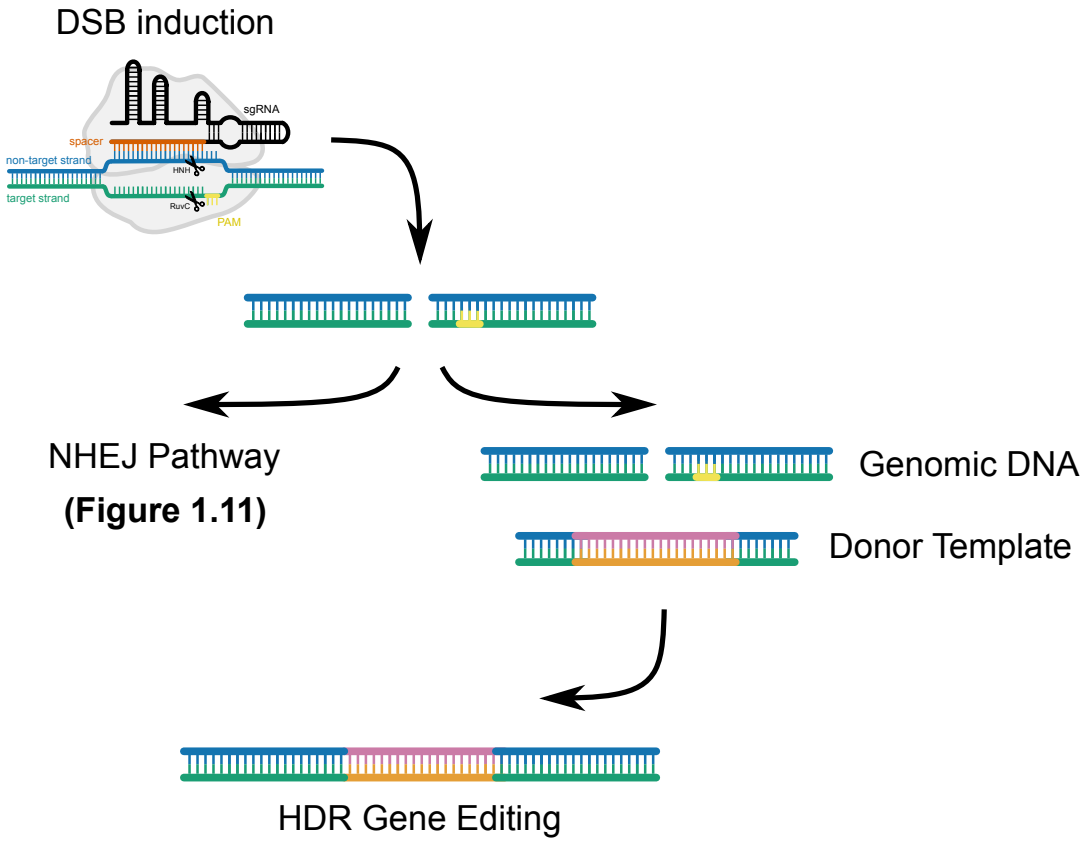


Figure 1.12: Simplified diagram of Cas9-mediate gene editing. Cas9-RNP recognizes target DNA and creates a DSB in genomic DNA. In some circumstances, DSBs can be resolved through HDR pathways. By supplying a donor template with homology arms (blue and green) to the DSB site, new genetic material (pink and orange) can be incorporated into the genome during repair. Successful repair strategies will remove the original target sequence to prevent further cleavage by Cas9.

induced DSBs are more frequently repaired by NHEJ than the desired HDR pathway, which raises the likely risk of generating InDel mutations in place of successful edits [114].

The efficiency of gene editing can be increased with Cas9 fusion proteins which bear effectors to inhibit NHEJ [115] or enhance HDR [116], however these strategies may be difficult to implement *in vivo*. nCas9s are also widely applied to create staggered single-stranded DNA breaks which can be used to stimulate HDR while minimizing the risk of InDel formation or large undesired chromosome deletions [117, 118].

1.8.3) Gene Editing with Prime Editors

Developed as a platform for versatile gene editing without requiring DSB formation and HDR, prime editing enables all types of targeted DNA substitutions, deletions, and small insertions without generating DSBs. Prime editors (PEs) are fusions of nCas9 (H840A) and Moloney murine leukemia virus reverse transcriptase (RTase). After cleavage of PAM-containing DNA target strands by nCas9, the RTase domain can directly transcribe new information at the 3′ terminus of the nick. The desired edit is templated by a prime editor guide RNA (pegRNA): a sgRNA with a critical 3′ extension that hybridizes to the nicked strand and templates reverse transcription. After synthesis, the edited strand displaces the adjacent 5′ strand of genomic DNA through flap interconversion. Excision of the displaced flap followed by ligation results in a heteroduplex in which one genomic strand contains the edit. DNA repair or replication copies the edit to the complementary strand and makes the prime edit permanent **(Figure 1.13)** [93].

Through this mechanism of directly rewriting target DNA, prime editing offers incredible versatility, editing purity, and DNA targeting specific in comparison to traditional Cas9-based strategies. Importantly, prime editing requires three checkpoints of complementary base interactions to function: the nCas9-DNA targeting, hybridization of pegRNA to nicked 3′ termini of target DNA, and hybridization of the newly edited strand to the target locus [119]. These checkpoints, coupled with the nonrequirement of DSBs in prime editing, contribute to an unusually low off-target mutagenesis rate which has been observed across mammalian cells [120], organoids [121, 122], mouse embryos [123, 124], and plants [125].

The higher editing rates and lower off-target mutagenesis of prime editing make it an attractive alternative to Cas9 gene editing. Currently, prime editing can only efficiently

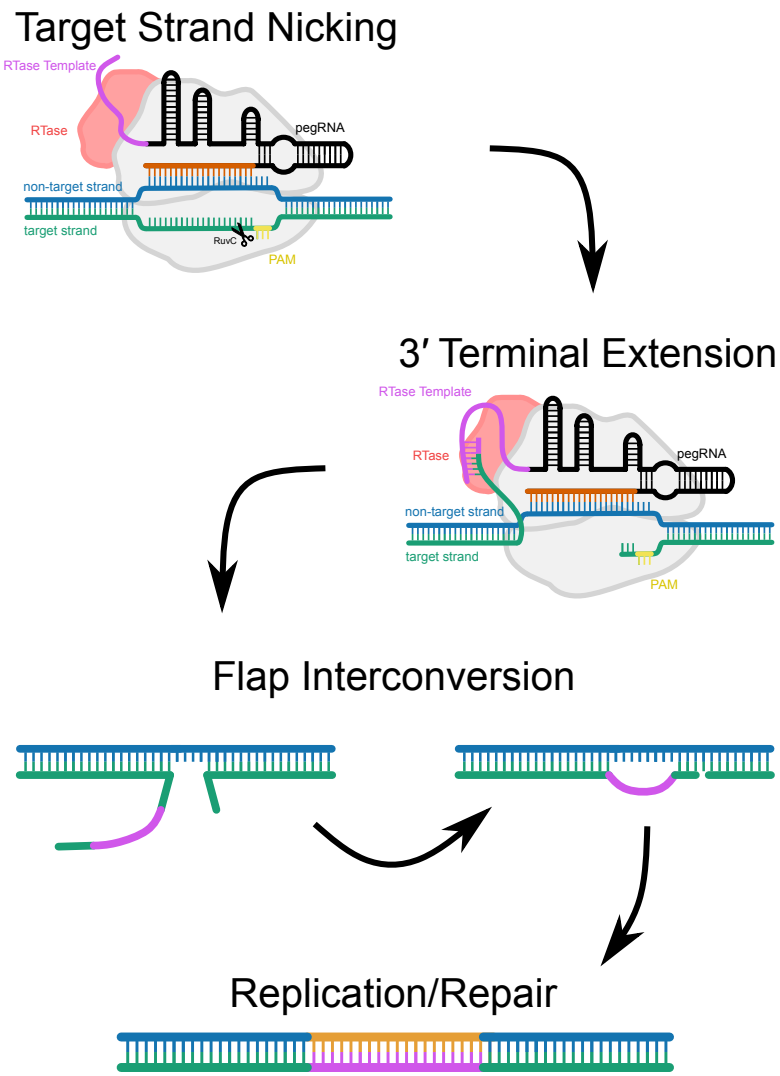


Figure 1.13: Simplified diagram of prime editing. A PE composed of nCas9 (grey) and RTase (red) cleaves the target strand (green) of substrate DNA. The 3′ extension of the pegRNA (pink) can hybridize with the target strand and template the synthesis of new DNA from the nicked strand's 3′ terminal. After extension of the 3′ terminal, flap interconversion occurs and the 5′ terminal flap is preferentially degraded. Normal replication or repair of the edited locus stably incorporates the edit.

generate genomic insertions of up to ~100 bp without the assistance of recombinases [108]. From a therapeutic perspective, more than 90% of known pathogenic mutations in the human genome are smaller than 30 bp [126], which is well within the remit of prime editing. However, Cas9 gene editing is still the tool of choice for generating large insertions for the time being. Attempts have been made to address other limitations of the prime editing

system, such as inconsistent editing rates between genomic loci and organisms, through protein engineering [127-129]. The pegRNA itself also poses challenges to prime editing, as its 3′ extension is more susceptible to nucleases than traditional Cas9 sgRNAs and can generate non-functional PE-RNPs [130]. Additionally, parameters for the design of efficient pegRNAs vary widely by target and insertion sequence, making them cumbersome to optimize [131]. The large size of pegRNAs (>120 nt) largely precludes chemical synthesis as a method to streamline this process.

1.9) Therapeutic mRNAs

1.9.1) Synthetic mRNAs as a Drug Platform

The rapid design and development of two SARS-CoV-2 mRNA vaccines realized decades of aspirations for a viable mRNA drug platform [132]. The impressive safety profile of these vaccines [133, 134] has renewed interest in mRNA to treat conventionally undruggable targets. The basic principles of mRNA design, production, translation and decay are reviewed in [135]. As a drug platform, mRNAs are easy to redesign for different targets and can be produced at scale in cell-free systems in a single reaction [136]. Accordingly, mRNA vaccines against infectious diseases such as influenza, Zika, HIV, and rabies [137], as well as some cancers [138] are currently being explored.

The rapid development of mRNA vaccines and their acceptable safety profile underscores the potential for new generations of mRNA therapeutics. mRNA vectors can be used to theoretically produce any protein in transfected cells and will produce an amplified signal through translation relative to protein-based therapeutics. Compared to DNA-based drugs, mRNA vectors have a relatively high transfection efficiency and low toxicity because they do not need to reach the cell nucleus to be functional. mRNA additionally precludes any risk of insertional mutagenesis [139, 140].

A wide array of therapeutic mRNA applications are being examined, including endogenous protein replacement, antibody expression, and gene editing. While mRNA therapeutics is a relatively unexplored field compared to mRNA vaccines, two therapeutics have already yielded promising safety and efficacy profiles from clinical studies: VEGF protein replacement for heart failure and a Cas9 gene therapy for amyloidosis [141-143]. Synthetic mRNA is additionally an effective vector for Cas9 [144] and prime [145] gene editing.

1.9.2) Enhancing mRNA Expression *in vivo*

One of the key challenges in the development of mRNA therapeutics is increasing the amount of protein produced from an mRNA to therapeutic levels. mRNA-based vaccines require minimal levels of protein expression to function as the immune system markedly amplifies antigenic signal, whereas mRNA protein replacement or gene therapies require as much as a 1,000-fold increase in protein to reach therapeutic thresholds [141].

Several strategies exist to improve the translational yield of mRNAs. Key features of the mRNA, such as the 5′ cap, 5′ and 3′ untranslated regions (UTRs), and poly(A) tail can each

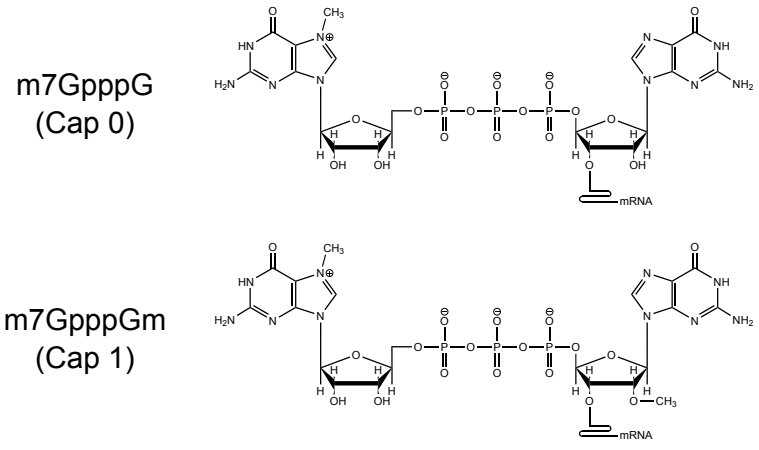


Figure 1.14: Structures of m7GpppG (cap 0) and m7GpppGm (cap 1) 5′ mRNA caps.

be optimized and tailored for target cells to increase stability and translational capacity of mRNA [146-150]. As an example, the Pfizer SARS-CoV-2 mRNA vaccines were produced with the less immunogenic cap 1 5′ cap structure using recently described CleanCap® technology [150] in place of the more common cap 0 structure **(Figure 1.14)**.

The incorporation of certain chemically modified nucleosides also markedly increases protein production from synthetic mRNA *in vivo*. Methylpseudouridine and other uridine analogs are particularly effective at reducing immune sensing by Toll-like receptors *in vivo* and thereby improving protein expression [151, 152]. Precise chemical modification of synthetic mRNAs remains a challenge however, and typically results in partial and heterogeneous modification.

Therapeutic mRNAs can also be designed to prolong the duration of protein expression **(Figure 1.15)**. Self-amplifying mRNAs (saRNAs) derived from RNA alphaviruses can be designed to amplify *in cis* genes of interest, extending their expression [153]. Broadly,

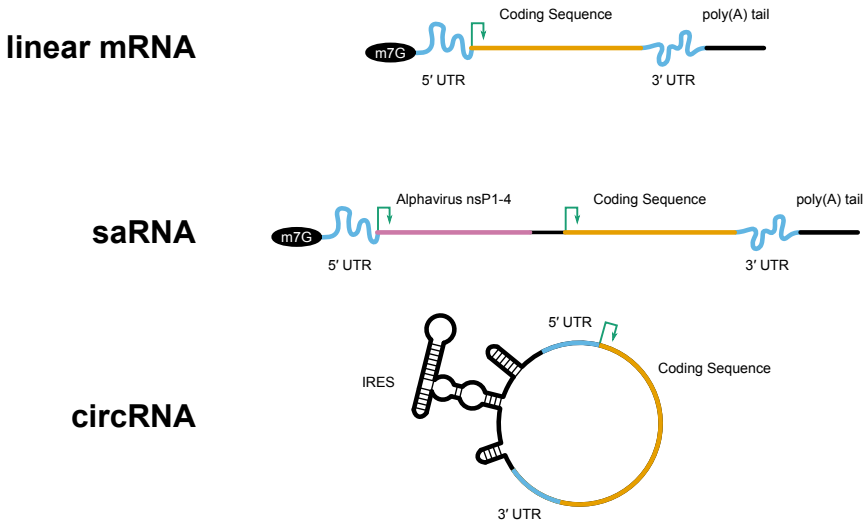


Figure 1.15: Diagram of three different mRNA vectors. Some key features of conventional linear mRNA can be optimized to improve its translation. SaRNAs encode an alphavirus RNA polymerase and necessary features to replicate the coding sequence of interest. CircRNA contains an IRES to initiate translation in place of a 5′ cap.

saRNA strategies require ~10-fold less RNA to generate the equivalent protein expression of conventional mRNAs [154, 155]. Another alternative to linear mRNA is circular mRNA (circRNA). CircRNA can be readily produced by flanking an mRNA with self-splicing introns derived from the T4 phage thymidine synthase gene; co-transcriptional back-splicing during *in vitro* synthesis circularizes the mRNA and excises the introns [156]. CircRNA is resistant to exonucleases and reduces immune recognition, greatly increasing its lifespan *in vivo* [157, 158]. The introduction of an internal ribosomal entry site (IRES) to initiate translation within circRNAs circumvents the need for 5′ capping and poly(A) tailing, making them cheaper and less cumbersome to produce than linear mRNAs. Recently engineered features (IRES, UTRs) for circRNAs are making them an increasingly relevant platform for RNA delivery [159].

1.10) Aims and Objectives

Bolstered by advancements in nucleic acid chemistry, oligonucleotides are increasingly being used in applications outside of their traditional role as information carrier molecules. This paradigm shift is embodied by two groundbreaking biomedical tools: RNA-directed gene editors and synthetic mRNAs. The length of Cas9 sgRNAs precludes chemical synthesis, making them difficult to chemically modify and prepare at scale. Synthetic mRNA is an incredibly versatile tool for therapeutics and research, but there are few adequate methods to control its translation with high spatial-temporal precision. This especially true of novel mRNA vectors that yield higher protein expression than linear mRNA. The under-utilized ^{CNV}K nucleoside can be applied in novel approaches to construct and control the activity of these biomolecules.

The specific aims of this work are:

Chapter 2 – Develop a ^{CNV}K-based approach for preparing chemically modified Cas9 sgRNAs that is compatible with library construction.

Chapter 3 – Apply the findings of Chapter 2 to prepare gRNA/DNA conjugates to expand the functions of Cas9.

Chapter 4 – Explore the reversible crosslinking of carbazole nucleotides as a platform for controlling mRNA translation with light.

2

^{CNV}K-sgRNAs

2.1 Background

2.1.1) Improving Cas9 Activity with Modified gRNAs

Cas9 and other CRISPR nucleases are widely applied to manipulate genomes, and their capacity for gene editing has been further expanded with engineered variants (Section 1.8). The crux of all Cas9 technologies is the simplicity with which they can be reprogrammed to new genomic targets by altering the 20 nt spacer sequence of the gRNA. The ease with which gRNAs can be designed and synthesized has been leveraged to produce extensive Cas9 gRNA libraries that have been applied for high-throughput genetic screens and strain engineering at a scale that would not have been possible more than a decade ago.

While engineered variants of Cas9 protein have yielded several impactful technologies, the gRNA itself can also be modified to improve or give new functionality to Cas9-RNP. The defined chemical nature of gRNAs makes them amenable to manipulation with the suite of tools produced through nucleic acid chemistry. Here, I will discuss the impact of synthetic

modifications to Cas9 gRNAs in their capacity to improve *in vivo* gene editing, as well as functional groups that have been introduced to the gRNA to give Cas9-RNP entirely new functions. I will then describe the currently available methods of preparing chemically modified gRNAs, as well as their shortcomings. To accommodate the demand and scale of chemically modified gRNAs, we are proposing a new ^{CNV}K-based platform for the construction of biofunctional, structured RNAs. I will describe our intended strategy for the construction of Cas9 sgRNAs with ^{CNV}K nucleotides, as well as the advantages this strategy could offer compared to existing methods of sgRNA preparation.

2.1.2) Modification of gRNA Backbone Improves Cas9 Activity

As discussed in Section 1.1, chemical modifications can easily be incorporated into synthetic oligonucleotides and have been widely explored to improve the pharmacokinetic properties of therapeutic oligonucleotides. Modifications of oligonucleotides are based on one or a combination of the following: modification of phosphate linkage between nucleotides, ribose sugar modifications, and nucleobase modifications [160]. Any application of Cas9 is fundamentally driven by Watson-Crick interactions between the gRNA and target DNA, and is subject to the stability and chemical properties of the gRNA. Therefore, many of the modifications designed to influence these parameters in therapeutic oligonucleotides are directly relevant to Cas9 gRNAs.

The goal of most gRNA modification strategies is to improve Cas9 gene editing, typically through increasing the intracellular stability of gRNA. A summary of the modifications discussed and their relevance for Cas9 gene editing is listed in **Table 2.1**. The potential of modified gRNAs was first demonstrated in 2015 by Hendel *et al*., who replaced 3 nt at the 5′ and 3′ termini of synthetic sgRNAs with 2′OMe residues in combination with either

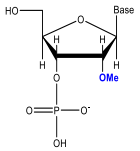
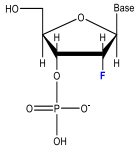
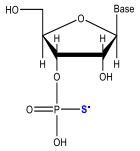
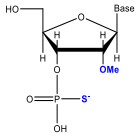
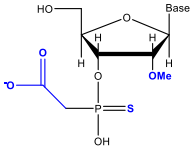
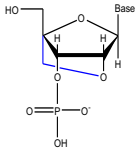
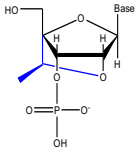
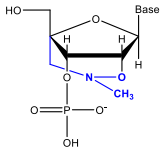
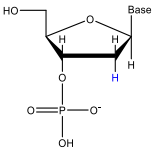
phosphate, phosphorothioate, or 3′thioPACE linkages (M, MPS, MPACE respectively) [161]. Not only were these modifications tolerated by the Cas9 enzyme, but they also greatly increased the overall stability of the sgRNA. Because of this, Hendel *et al.* were able to achieve gene knockouts at several loci by co-transfecting Cas9-mRNA and modified-sgRNAs in multiple cell lines, including primary human T cells and $CD34^+$ hematopoietic stem and progenitor cells. Prior to this work, Cas9 and gRNA were typically delivered to cells as a plasmid or RNP complex because unmodified naked gRNA is rapidly degraded *in vivo*. Therefore, this relatively simple modification strategy made an entirely novel method for Cas9 delivery feasible. Owing to their simplicity but relative effectiveness, MPS and MPACE modification of the gRNA's terminal 3 nt remains a standard for Cas9 editing [162-164], and have even been applied for *in vivo* gene editing with lipid nanoparticle delivery vectors [165].

On a similar tack, Rahdar *et al.* reported that full replacement of the crRNA with phosphorothioate (PS) linkages in a dgRNA system improved activity four-fold and that this could be further improved by incorporating 2′OMe residues at each terminus [166]. A more systematic approach to modifying sgRNAs was carried out by Yin *et al.*, who used crystallography data to identify every contact between Cas9 and the sgRNA to determine which positions in the sgRNA could be modified [162]. Using this data, Yin *et al.* were able to design an enhanced sgRNA with an optimized combination of 2′OMe and 2′F substitution and phosphorothioate (PS) linkages which had higher gene editing activity than the terminally modified sgRNAs described by Hendel *et al.* [161]. With a single injection of lipid nanoparticles containing their enhanced sgRNA and Cas9 mRNA, Yin *et al*., achieved >80% editing of a target gene in mouse livers [162]. A similar strategy was adopted for the design of modified dgRNAs by Mir *et al*. [163], who arrived at a different pattern of

modification than Yin *et al.* but were also able to demonstrate that the incorporation of 2′OMe and 2′F residues with PS linkages improves gRNA stability and activity. Mir *et al.* also took their design a step further and demonstrated that modified dgRNAs also improve HDR gene editing rates. These different strategies broadly demonstrate that stabilizing modifications to the gRNA improve Cas9 editing outcomes.

In addition to the improved stability and editing efficiencies offered by modified gRNAs, several groups have also explored modification as a method to improve on-target/off-target discrimination. The risk posed by off-target DNA cleavage and mutagenesis is one of the greatest barriers to therapeutic applications of Cas9. When incorporated into certain positions of the gRNA spacer sequence, 2′-O-methyl-3′-phosphonoacetate modifications have been demonstrated to reduce off-target editing while maintaining high on-target activity [167]. Others have introduced modifications in the gRNA spacer sequence which were originally designed to improve target discrimination in translation inhibitory oligonucleotides, such as bridged and locked nucleic acids [168]. It is additionally possible to substitute larger segments of sgRNA or dgRNA positions with DNA in place of RNA [169-172]. Because DNA residues are less reactive during synthesis than RNA, chimeric gRNAs are relatively cheaper and easier to make than a full-RNA gRNA. When incorporated into the gRNA spacer sequence, DNA has also been shown to reduce off-target Cas9-mediated cleavage, likely due to the weaker interaction between the DNA target and gRNA compared to a full RNA/DNA interaction [10,11]. However, this modification strategy typically reduces the on-target activity as well because of the loss of hydrogen bonding between RNA nucleotides and Cas9 protein.

Table 2.1: Some relevant backbone and sugar modifications for SpyCas9-gRNAs and their effect on Cas9 activity. Table adapted from [174, 175].

Modification		Effect on Cas9 Activity	Mod. Location	References
M		Improved genome editing efficiency [161, 166]	sgRNA termini [161] crRNA termini [166]	Hendel [161], Rahdar [166]
F		Improved genome editing efficiency when incorporated with other modifications (M, PS, MS) [162,163,166,172]	crRNA [163, 166,172] sgRNA [162]	Yin [162], Mir [163], Rahdar [166], O'Reilly [172]
PS		Full PS modification of crRNA improves genome editing efficiency [166] PS tolerated in highly modified sgRNAs [162]	crRNA [166] sgRNA termini [162]	Rahdar [166], Yin [162]
MPS		Improved genome editing efficiency [161,162,163,164,165,166] Reduced toxicity [164]	terminal residues [161,164,166] dgRNA [163, 172] sgRNA [162,165]	Hendel [161], Yin [162], Mir [163], Basilia [164], Finn [165], Rahdar [166], O'Reilly [172]
MPACE		Improved genome editing efficiency [161]	sgRNA termini [161]	Hendel [161]
LNA		Up to 6 LNA residues are tolerated in the spacer sequence and reduce off-target gene editing [168]	spacer sequence [168]	Cromwell [168]
S-(c)ET		Improved genome editing efficiency in dgRNA system when incorporated with F, M modifications [166]	spacer sequence [166]	Rahdar [166]
BNANC		Up to 6 BNACN residues can be incorporated into spacer sequence, improving genome editing efficiency and broadly reducing off-target activity [168]	spacer sequence [168]	Cromwell [168]
DNA		DNA modifications are tolerated in crRNA, tracrRNA [169,171,172] DNA substitutions reduces off-target activity [169,172]	spacer sequence [171], crRNA [169][172]	Yin [169], Kartje [171], O'Reilly [172]

Any oligonucleotide-based drugs can stimulate a deleterious interferon response via cellular pattern recognition receptors in the cytoplasm or endosomes [173], and Cas9-gRNAs are no exception. Certain chemical modifications allow oligonucleotides to evade immune detection however, improving their efficacy and reducing their toxicity. Schubert *et al.* demonstrated that modification of gRNA with 2′OMe residues eliminates the type 1 interferon response in human peripheral blood mononuclear cells [176].

While significant progress has been made to improve Cas9 activity through gRNA modification, these improvements are likely only scratching the surface of the potential offered by nucleic acid chemistry. This point was illustrated by O'Reilly *et al*., who demonstrated that modifications as exotic as 2′5′ linked RNA are tolerated in the Cas9 spacer sequence [172]. Many of the modifications that have been introduced to the Cas9 gRNA are also relevant for other CRISPR proteins such as Cpf1 [177, 178], but these have yet to be widely explored. Modification of gRNAs for newer Cas9-based technologies, such as base editing and prime editing, have also not been explored extensively apart from terminal nucleotide modification [179, 180].

2.1.3) Chemical Modifications for Expanded Cas9 Functionality

The examples discussed above highlight the impact of modified phosphate linkages and ribose sugars in gRNAs as they relate to Cas9 editing efficiency. However, oligonucleotide modification is not limited to replacing existing features with structural analogs; modification strategies also make it possible to give Cas9 entirely new functions.

One example of this is the generation of "activatable" or "toggleable" Cas9, whose gene-editing activity can be controlled with external stimuli. Limiting the duration of Cas9 activity

can help to reduce off-target cleavage events in cells. Additionally, these strategies can also be applied to limit gene editing to cells of interest, further reducing the risk associated with Cas9 gene editing. Several strategies have been developed to manipulate the spatiotemporal activity of Cas9 with stimuli by modifying the Cas9 protein or its delivery vesicle [181]. However, these methods tend to suffer from leaky activation and generally increase the already large size of Cas9 protein, making it challenging to deliver to cells. These toggleable properties can alternatively instead be produced using chemically modified gRNAs, which are generally smaller and easier to prepare: Zhou *et al.* and Moroz-Omori *et al*. both developed gRNA with photo-caged nucleobases to block spacer sequence and target DNA hybridization, preventing Cas9 function [182, 183]. By pulsing UV light, these photo-caging groups can be removed, restoring Cas9 activity. Both groups applied their strategy in zebrafish embryos and were able to demonstrate a high degree of control over Cas9 activity. Carlson-Stevermer *et al*. were able to create the opposite behavior, whereby UV irradiation cleaves photo-sensitive linkers in the sgRNA, destroying its structure and function [184]. In addition to photo-controllable Cas9, ligand-responsive caged gRNAs have also been developed, whose interactions with Cas9 are inhibited until induction with a water-soluble phosphine [185].

Another application of modified bases has been the development of Cas9 and Cpf1 gRNAs which can detect natural polymorphisms or degeneracies in target DNA. By synthesizing gRNAs with “universal bases”, such as inosine, that base pair to any of the canonical DNA nucleotides, *Krysler* et al. enabled simultaneous targeting of polymorphic sequences with selective degeneracy at the site of universal base incorporation [186].

In addition to altering the dynamics of DNA target recognition, modified gRNAs have also been applied to improve Cas9 HDR gene editing efficiency. The rate of editing through HDR can be increased by improving transfection efficiency of the DNA donor to the edited cells [187] and localizing the DNA donor to the immediate vicinity of the Cas9-induced DSB. Lee *et al.* attempted to improve both of these factors by conjugating the crRNA of a dgRNA system to a DNA donor repair template [188]. By click-ligating the two oligonucleotides together, their system would ensure an equal delivery of gRNA and repair template, as well as localizing the repair template to the site of Cas9 cleavage. By transfecting their conjugated gRNAs into HEK cells with cationic lipid nanoparticles, Lee *et al.* demonstrated a three-fold improvement in HDR rates compared to a non-conjugated control [188].

Because of their unprecedented ability to interrogate genomic content, Cas9 and other CRISPR nucleases are widely applied to visualize the subcellular locations and interactions of genes, chromatin, and RNA [189]. In these applications, fluorescently labeled dCas9-RNP complexes are used to bind to targets without cleaving them. The field of Cas9-cellular imaging is broad and has been extensively reviewed elsewhere [190]. It is worth noting though that gRNAs are more amenable to synthetic modification than the Cas9 nuclease itself and are a useful platform for incorporating organic dyes into the Cas9-RNP. Chemically modified sgRNAs have been applied to visualize cellular dynamics in both fixed [191] and live cells [192].

2.1.4) Methods of gRNA Synthesis

Cas9 gRNAs are largely produced with one of three methods: transcription from a plasmid *in vivo*, *in vitro* transcription (IVT), or RNA phosphoramidite synthesis. While delivery of

gRNA through plasmid is relatively cheap and effective, this method has several drawbacks: The delivery of Cas9 and gRNA by plasmid has been demonstrated to be more stressful to cells and increase off-target effects compared to Cas9-RNP delivery, and raises the risk of insertional mutagenesis [193]. Additionally, gRNA expression within cells precludes the introduction of chemical modifications, making their stability a concern with *in vivo* applications [194].

IVT is widely applied as a simple and cheap approach to preparing gRNAs: by preparing a linear DNA template under the control of a T7 RNA polymerase promoter, large amounts of sgRNA can be prepared within a day using commercially available transcription kits. IVT preparation of gRNAs removes the cloning steps required for plasmid transcription and overall expedites the Cas9 gene editing workflow [195]. After preparation, the gRNAs can be complexed with recombinant Cas9 protein for transfection and effective editing of primary cell lines [196]. sgRNAs produced through IVT (IVT-sgRNA) are not without their disadvantages, however. T7 RNA polymerase reactions can produce significant amounts of 3′ and even 5′ sequence heterogeneity, which could reduce sgRNA activity and increase off-target effects [197, 198]. IVT-sgRNAs can also be recognized by cellular sensors for viral RNA and trigger immune responses and cell death [199]. Retinoic acid-inducible gene 1 (RIG-1) can identify the 5′ triphosphate left on IVT-sgRNAs after transcription and trigger the type 1 interferon response [200]. The immune response can be ameliorated through enzymatic removal of the 5′ triphosphate prior to transfection [199, 200], but adds additional steps to the RNA preparation workflow. In addition to this, site-specific modification of IVT-produced RNAs remains a challenge [201, 202], making this technique difficult to pair with the benefits of modified Cas9 gRNAs.

Only chemically synthesized sgRNA molecules can be effectively modified in a site-specific manner. They also benefit from reduced sequence heterogeneity and lack the 5′ triphosphate group that triggers cell immunity. However, the occurrence of side reactions during RNA chemical synthesis makes it challenging and expensive to synthesize a full-length sgRNA, especially if synthetic modifications are included in the sequence. The gradual accumulation of failure sequences is especially detrimental for sgRNA synthesis, as mutations in the 5′ spacer sequence can cause dangerous off-target effects and are difficult to purify away from the correct sequence. These barriers have not prevented some groups from designing and implementing chemically modified sgRNAs, [161, 162]. However, these challenges do make the technology less accessible to other researchers. Additionally, while one modified sgRNA can be prepared with some effort and cost, preparing a sgRNA library of any reasonable size would be prohibitively expensive.

As an alternative approach for preparing libraries, some groups have instead used dgRNAs as a platform for modification [163, 167-169]: because the dgRNA is composed of an approximately 40 nt long crRNA and 80 nt long tracrRNA, the two RNAs can be synthesized individually and modified more easily than a full-length sgRNA. The tracrRNA itself only needs to be prepared once, while the shorter, variable crRNA would be remade with new sequences or modifications. While this is a useful approach for testing modifications, in practice dgRNAs have lower activity than sgRNAs and increase the complexity of transfections.

2.1.5) RNA Ligations for sgRNA Preparation

Click-chemistry has been applied to marry the efficiency of sgRNAs with the flexibility and cost of dgRNAs. This was first explored by He *et al.* [203], who in a communications paper

demonstrated that a 3′-azide crRNA and a 5′-hexyne tracrRNA could be click ligated with the copper-catalyzed Huisgen 1,3-dipolar cycloaddition. The repeat-antirepeat duplex of the resulting sgRNA was joined by a carbon linker with triazole linkage, as shown in **Figure 2.1a**. He *et al.* reported that their triazole-sgRNAs had comparable activity to a non-ligated dgRNA control and a full-length synthetic sgRNA in a U2OS cell mutagenesis assay. It should be noted that He *et al.* only performed one biological/biochemical assay of their triazole-sgRNA, and the T7E1 assay they used to assess editing is semi-quantitative and suffers from high background signal [204]. While this work was notable for beginning to explore chemical approaches to constructing sgRNAs, their triazole-sgRNA had neither improved *in vivo* activity nor did it have a lower synthetic burden than their non-click ligated control.

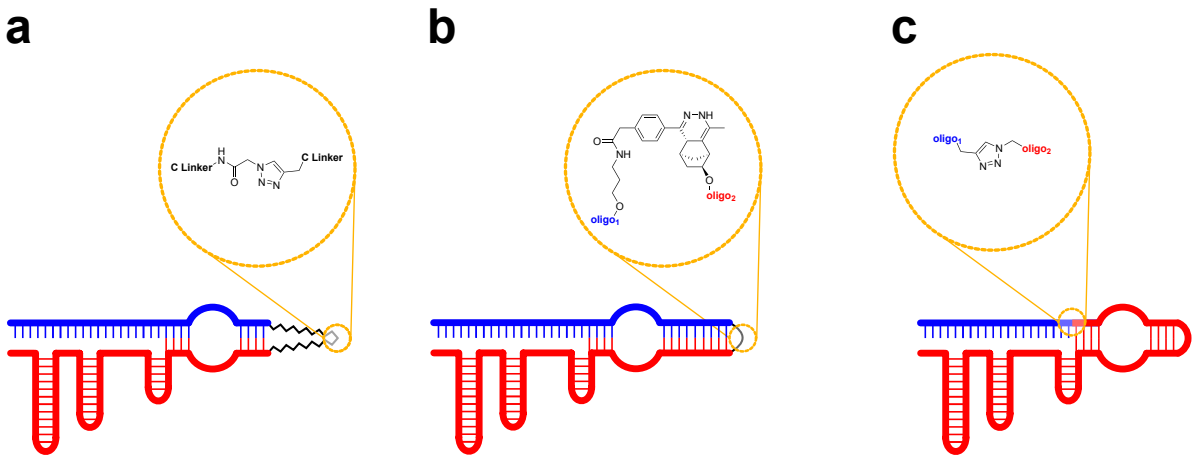


Figure 2.1: Diagram of click-ligation strategies for sgRNA construction. **a.)** Strategy described by He *et al.* (2016) [203], in which a 3′-azide crRNA (blue) and a 5′-hexyne tracrRNA (red) are ligated with the copper-catalyzed Huisgen 1,3-dipolar cycloaddition. The resulting triazole linkage (circled) is separated from each RNA strand by a C4 linker. **b.)** Strategy described by Chen *et al.* (2022) [205], who ligated tetrazine-crRNAs (blue) and norbornene-tracrRNAs (red) using the inverse-electron-demand Diels–Alder (IEDDA) reaction. The crRNA and tracrRNA are both lengthened compared to a typical dgRNA system to improve ligation and Cas9 efficiency. **c.)** Strategy described by Taemaitree *et al.* (2019) [206], whereby a 20 nt alkyne RNA containing the Cas9 spacer sequence (blue) is ligated to a 79 nt structural azide RNA (red). The triazole linkage replaces a natural phosphate linkage in the sgRNA.

A similar approach to constructing sgRNAs was described in 2022 by Chen *et al.*, who ligated tetrazine-crRNAs and norbornene-tracrRNAs using the inverse-electron-demand Diels–Alder (IEDDA) reaction (**Figure 2.1b**) [205]. After increasing the length of the crRNA/tracrRNA repeat-antirepeat stem and optimizing the length of their linker, Chen *et al*. designed a functional ligated sgRNA. Unlike the design used in [203], Chen *et al.* demonstrated that their tetrazine-sgRNA had higher InDel forming activity in cells than a dgRNA of comparable length at 4 different target loci. However, the magnitude of improved activity varied by target.

Our group made a key innovation in the field of click-ligated sgRNAs: By studying published crystallographic data of the interactions between Cas9 and the gRNA [72], we were able to identify a phosphate linkage immediately 3′ of the spacer sequence in the sgRNA which interacted minimally with Cas9 protein and was suitable for replacement with a triazole. This position was leveraged to generate a new triazole-sgRNA, in which a 20 nt alkyne-RNA was ligated to a 79 nt azide-RNA (**Figure 2.1c**) [206]. The synthetic burden of RNA synthesis is reduced by splitting the sgRNA into the variable DNA/genome-targeting 20-mer, produced on-demand and in high purity, and the fixed Cas9-binding 79-mer, produced cost-effectively in large-scale. This strategy provides an excellent platform for modification of the sgRNA spacer sequence for improved pharmacokinetics and lends itself well to multiplexed sgRNA synthesis. This was demonstrated by comparing the activity of a pooled-click ligation of 6 triazole-sgRNAs to IVT-sgRNA controls, which were found to have comparable *in vitro* activity. We additionally showed that triazole-sgRNAs have comparable editing efficiency to IVT-sgRNAs *in vivo* and that the placement of the triazole did not cause a significant difference in the frequency of off-target effects [206].

2.1.6) A Photo-Crosslinking Platform for sgRNA Construction

Chemical modification of gRNAs has dramatically improved the efficiency of Cas9 gene editing and has enabled new functions that would be difficult to achieve through manipulation of the protein alone. However, the only reliable method for installing site-specific modifications into gRNAs is chemical synthesis, which is not adequate for producing full-length sgRNAs at scale. This arrangement creates a bottleneck in the development of Cas9 techniques, as researchers must choose between the high costs of synthetic sgRNAs or the reduced efficiency of dgRNAs. Robust platforms are therefore needed to generate gRNA libraries and to continue exploring the potential of CRISPR proteins. Previous work in our group has shown that RNA ligations can be used to construct full-length sgRNAs, highlighting the potential of novel chemical strategies to address these issues.

As an alternative to click chemistry, we envisage that the ^{CNV}K nucleoside could form the basis of a crosslinking platform to construct large structural RNA molecules like Cas9 sgRNAs. ^{CNV}Ks can be used to covalently link two hybridized oligonucleotides, locking their interaction in place even under denaturing conditions [32]. Any region of dsRNA in a large structural RNA could be used as a site for ^{CNV}K incorporation and crosslinked, allowing a single RNA to be split into smaller constituent parts while still covalently bound together as if they were a single molecule **(Figure 2.2)**. Because the ^{CNV}K-RNA would be constructed *in vitro*, crosslinking conditions can be optimized to favor interactions between constituent RNA strands that would otherwise be unstable *in vivo*.

The naturally occurring stem-loop structures in the Cas9 gRNA are ideal candidates for replacement with a ^{CNV}K crosslink. The gRNA can already be expressed as a split dgRNA. The interaction between the two strands can therefore be stabilized by inserting a ^{CNV}K

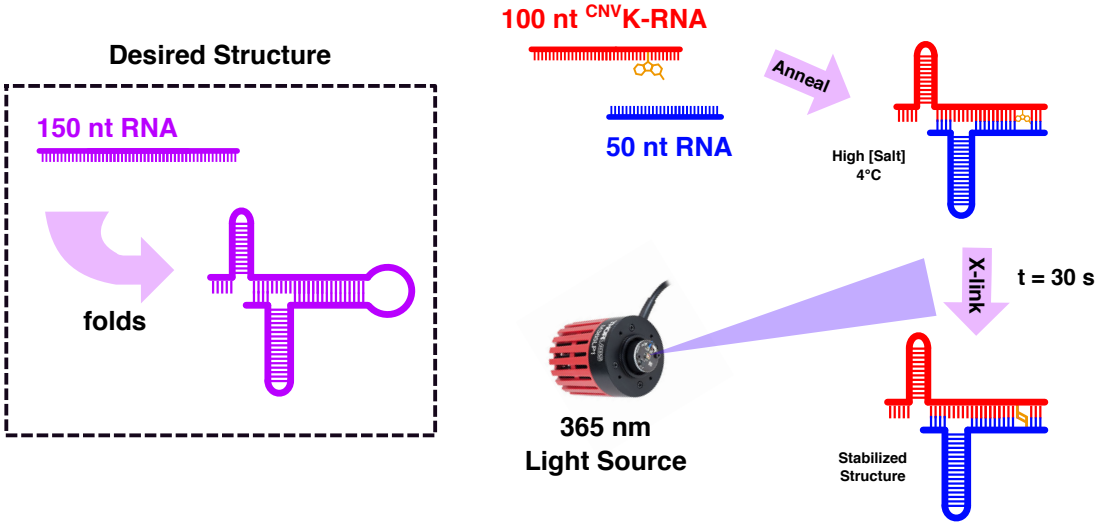


Figure 2.2: General schematic of structured RNA construction by CNVK crosslinking. A hypothetical 150 nt RNA is shown in purple which folds to form three stem-loops. The central stem-loop is used as a point of division to split the RNA into two smaller components: A CNVK-bearing 100 nt fragment and a smaller 50 nt fragment. The two RNA fragments can be annealed in favorable *in vitro* conditions, such as low temperature and high salt concentrations. The annealed structure is then crosslinked with a 30 s pulse of UV light, fixing the structure into place.

crosslink, which would covalently link the two RNAs and staple their structures into a fixed interaction. From a synthesis perspective, the CNVK residue will be incorporated into the half of the sgRNA which has the larger synthetic burden and can then be produced at scale; the shorter half of the design, which bears the sgRNA spacer sequence, can then be multiplexed to produce libraries of potential targets and modifications. Because the RNAs would be crosslinked, both strands could potentially be shorter than a conventional dgRNA as they would not rely on Watson-Crick interactions for stability. Because CNVK crosslinks are stable in denaturing conditions, a CNVK-sgRNA could easily be prepared *in vitro* and then purified before further *in vitro* or *in vivo* application.

While click-ligation has been demonstrated to be an effective method for preparing chemically synthesized sgRNAs, a CNVK crosslinking approach could offer some advantages over the former. Click-ligation of two RNAs requires both strands to be chemically modified. In contrast, CNVK nucleotides can be incorporated during solid-phase RNA

synthesis and do not require partner-strand modifications for crosslinking. While click-ligations are known for their speed, ^{CNV}K photo-crosslinking is also a rapid reaction, taking as little as 30 s to run to completion. Unlike click-ligations however, crosslinking can be performed in physiological conditions and does not require additional cofactors or ligands. Our intended design for a ^{CNV}K-sgRNA would leave each terminus of the constituent RNA exposed, whereas a ligation-based strategy will seal two termini together. While this may seem detrimental to the stability of the ^{CNV}K-sgRNA, this can be easily overcome with some of the modifications described earlier [161, 163, 164, 166]. More importantly, these additional termini become prime sites for further modification during synthesis.

On a similar note, by forgoing click-chemistry to construct synthetic sgRNAs, it becomes easier to use click reactions to further functionalize the resulting sgRNA. This could be essential for constructing larger gRNAs, such as prime editor guide RNA (pegRNAs), and for functionalizing gRNAs with other biomacromolecules. Both applications will be discussed further in this chapter and the next.

While it is outside the scope of this project, ^{CNV}Ks could also potentially give RNAs toggleable activity because of the reversible nature of their crosslinking: The disruptive effect ^{CNV}Ks have on duplex stability prior to crosslinking [17] could potentially be used to break apart a crosslinked RNA structure once the crosslink is photo-split, rendering the structure non-functional. The ability to inducibly inactivate Cas9 gene editing is a highly desirable trait [184] as it can be used to limit off-target Cas9 gene edits and control other Cas-based technologies.

We envisage that a ^{CNV}K platform for sgRNA construction could be relevant for preparing any structured and biofunctional RNA, and will open new avenues for preparing libraries of synthetically modified RNAs. Because the interactions between Cas9 and gRNA are highly specific and structure-dependent, replicating a functional sgRNA stapled by crosslinks will be a critical proof-of-concept for this method.

2.2 Objectives

Modern nucleic acid chemistry offers a suite of tools for improving and expanding the activity of biologically relevant oligonucleotides, but several barriers limit their application; most notable is the difficulty of synthesizing RNAs >100 nt in length. In this chapter, the photo-crosslinking ^{CNV}K nucleotide will be explored as a novel platform for constructing large biofunctional RNAs. As a case study, we will attempt to construct the Cas9 sgRNA by crosslinking smaller synthetic RNA precursors, which will reduce the synthetic burden of the sgRNA and open it up to further chemical modification and library generation. However, ^{CNV}K crosslinking in dsRNA and short duplexes is not well characterized; therefore our first objective will be to explore the limits of efficient ^{CNV}K crosslinking in these duplexes. We will then attempt to incorporate ^{CNV}K crosslinks into a dgRNA system and determine whether it has a deleterious effect on Cas9 DNA cleavage activity. The design of our system will be optimized, and then applied to the construction of chemically modified sgRNAs and sgRNA libraries. Lastly, we will describe how ^{CNV}Ks could be used to prepare modified pegRNAs and outline the steps that have been taken to implement this.

2.3 Characterizing ^{CNV}K Crosslinking

2.3.1) ^{CNV}K Crosslinking in RNA/RNA Duplexes

As discussed in Section 1.3, the ^{CNV}K nucleotide has been implemented in a wide range of applications: ^{CNV}K oligonucleotides have been designed to inhibit gene expression, capture biological molecules, regulate nanostructures, and stabilize structures for further study. Despite this variety of applications, most ^{CNV}K oligonucleotides are made with a similar design: the ^{CNV}K will invariably be incorporated in the center of large sequences so that the destabilizing effect they have prior to crosslinking can be mitigated by extensive Watson-

Crick interactions on either side of the position. Because of this, it was initially unclear whether any structure in the sgRNA structure would be suitable for replacement with a ^{CNV}K residue, as no duplex in the original sgRNA structure is longer than 6 bp [54]. Additionally, although several publications have described the crosslinking of synthetic ^{CNV}K oligonucleotides to RNA, to our knowledge ^{CNV}K crosslinking has never been described in an RNA/RNA duplex. Given that RNA duplexes predominately adopt A-form helices compared to B-form DNA helices, it should not be taken for granted that ^{CNV}K crosslinking will proceed as effectively in RNA homoduplexes as in DNA duplexes. Because a ^{CNV}K residue and its target pyrimidine are not base-paired but are staggered by n-1 in the duplex, the geometry of the duplex could influence the interaction of the two nucleotides and restrict the crosslinking reaction.

The SpyCas9 sgRNA contains 4 dsRNA structures that could potentially be split and crosslinked to form the basis of a ^{CNV}K-sgRNA platform. However, the current lack of knowledge on ^{CNV}K behavior in short duplexes makes it difficult to determine at which sites it would be viable to incorporate a crosslink. More fundamentally, we also needed to determine whether crosslinking a dsRNA duplex was possible. To this end, we prepared two 17 nt oligonucleotide sequences that would anneal to form a short duplex. The top strand contained a ^{CNV}K residue and was synthesized as either a DNA or RNA molecule (^{CNV}K-ODN and ^{CNV}K-ORN respectively). Similarly, the bottom strand was also synthesized as both a DNA and RNA molecule (comp-ODN and comp-ORN respectively), so that the 4 oligonucleotides could be annealed to form any type of DNA/RNA homoduplex or heteroduplex (**Figure 2.3a**). To test whether the nature of an oligonucleotide duplex impacts ^{CNV}K crosslinking efficiency, we prepared 1 M solutions of each duplex in 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, which were subsequently annealed and crosslinked with a 365

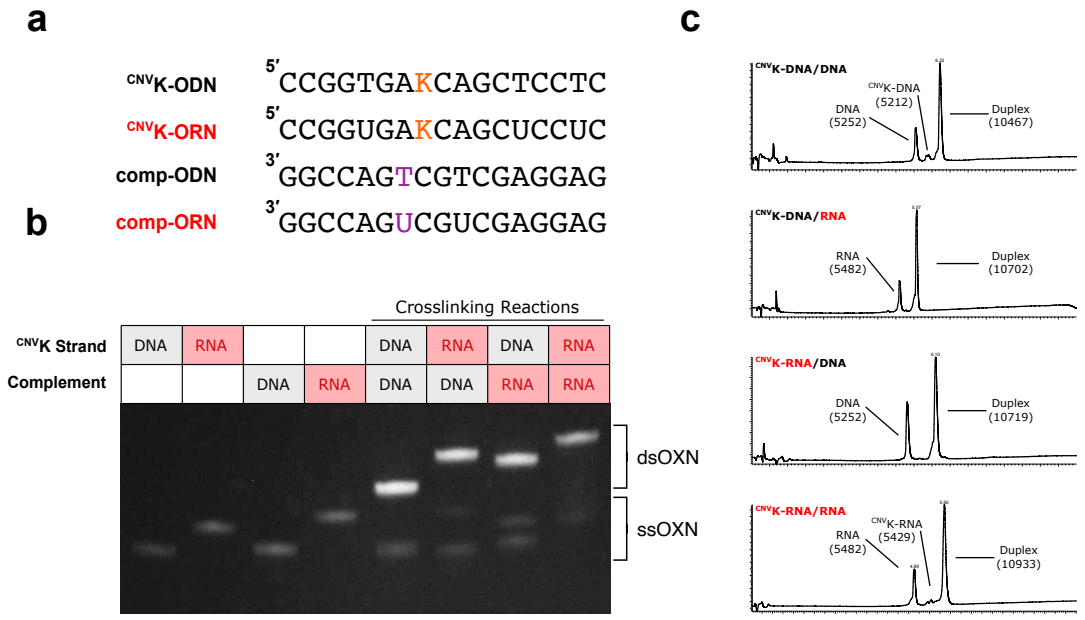


Figure 2.3a: Sequences of short oligonucleotides used to form DNA/RNA homoduplexes and heteroduplexes. The CNVK residue is demarcated with an orange 'K' in appropriate strands. The crosslinking target pyrimidine base in the complementary strand is colored purple. **Figure 2.3b:** Denaturing PAGE analysis of short CNVK-duplex crosslinking reactions. The species of oligonucleotide duplex used in each reaction is described above its respective lane. A blank box indicates no oligonucleotide. Brackets to the right of the gel highlight single-stranded (ssOXN) and double-stranded (dsOXN) material. **Figure 2.3c:** HPLC-MS analysis of CNVK duplex crosslinking reactions. The species of CNVK and complementary oligonucleotide used in each reaction are described above respective plot. Peaks are identified with their m/z ratio.

nm UV LED for 1 min at 4°C. The crosslinking reactions were then analyzed by denaturing PAGE and visualized by SYBR Gold staining **(Figure 2.3b)**. In each reaction, we observed the formation of a larger product compared to the ssOXN starting material, which likely corresponds to the crosslinked duplex. These new products were also significantly brighter than the ssOXNs, which could be explained by the stronger interaction of the intercalating staining dye with dsOXNs compared to ssOXNs. Each crosslinking reaction produced one product with similar efficiency regardless of the type of oligonucleotide duplex, suggesting that RNA/RNA duplexes crosslink with equal efficiency to DNA homoduplexes or heteroduplexes. It is also interesting to note that despite being the same length, each ssOXN and dsOXN migrated to different positions within the gel. This migration pattern can be

described by the m/z differences between RNA and DNA, as well as uracil and thymine, which causes RNA strands to migrate more slowly than comparable DNA strands.

We next repeated each crosslinking reaction as described above with a 1.5x excess of non-^{CNV}K strand and analyzed these by HPLC-MS **(Figure 2.3c)**. In each reaction, we were able to observe the formation of crosslinked duplex and the almost total consumption of the ^{CNV}K-strand. We did not observe a significant difference in crosslinking yields between the different duplex species, which further demonstrates that ^{CNV}K crosslinking is viable for RNA/RNA duplexes. Additionally, these results track well with our observation from the previous gel-based experiment and confirm that PAGE is a reliable method for analyzing crosslinking. Importantly, the mass of each crosslink product corresponded only to the appropriate ^{CNV}K-OXN/comp-OXN heterodimers, as opposed to ^{CNV}K-OXN homodimers, confirming the correct product was formed.

2.3.2) ^{CNV}K Crosslinking in Minimal Duplexes

After confirming that ^{CNV}K crosslinking in RNA homoduplexes is possible, we next wanted to investigate the effects of duplex stability on crosslinking efficiency. The dsRNA structures in a Cas9 sgRNA vary in size, but none are as long as the 17-mer duplex we studied above. If dsRNA crosslinking is still viable in short RNA duplexes, we would have fewer constraints on the final design of our ^{CNV}K-sgRNA.

First, we wanted to explore the requirements for duplex length using the ^{CNV}K-ODN we had already prepared. To test the impact of duplex length on crosslinking, we synthesized a set of 5′ truncated comp-ODN sequences, as shown in **Figure 2.4a**. These truncated comp-ODNs can hybridize with ^{CNV}K-ODN, but reduce the amount of duplex 3′ of the ^{CNV}K nucleotide.

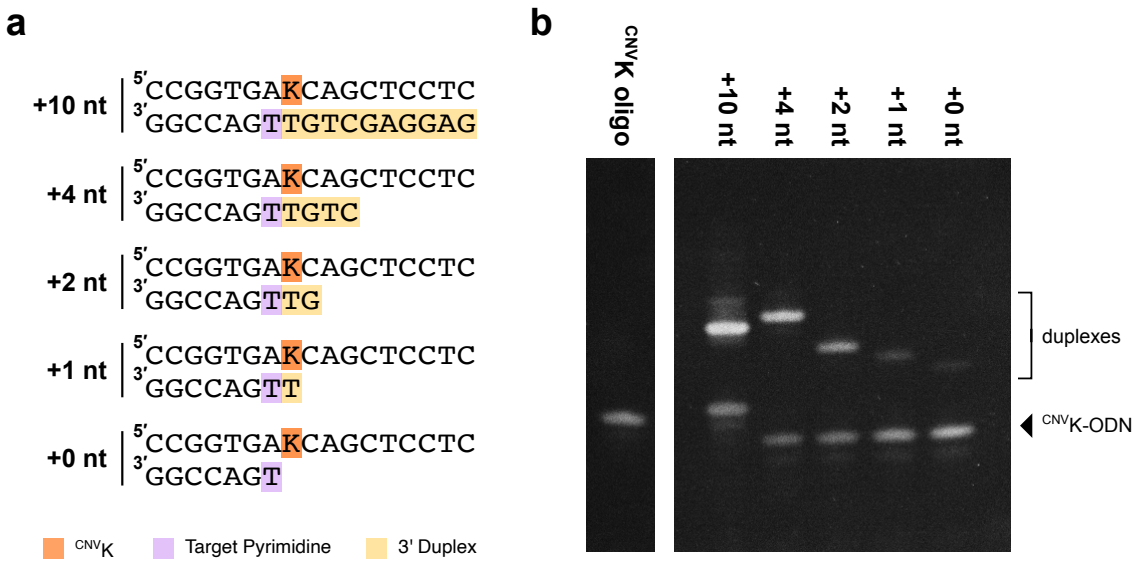


Figure 2.4a: Sequences of truncated duplexes used to study the effect of duplex length on CNVK crosslinking efficiency. The amount of base-paired nucleotides 5′ of the target pyrimidine is described to the left of each duplex. Each duplex is comprised of CNVK-ODN and comp-ODN, comp-ODN^{-6}, comp-ODN^{-8}, comp-ODN^{-9}, or comp-ODN^{-10} (+10, +4, +2, +1, and +0 nt duplex respectively). The CNVK residue is highlighted in orange, the target thymidine in purple, and every "excess" base 5′ of the target thymidine in yellow. **Figure 2.4b:** Denaturing PAGE analysis of crosslinking reactions with truncated DNA duplexes. Text above the gel indicated whether a lane corresponds to the CNVK-ODN by itself or the species of duplex used in crosslinking reaction. The bracket and arrow to the right of the gel highlight whether bands correspond to CNVK-ODN or crosslinked duplex. In lane +10 nt, comp-ODN is visible but not CNVK-ODN. In all other lanes, the complementary ODN is not visible.

The conserved duplex interaction 3′ of the target thymidine base in the truncated comp-ODNs had a TM of 8.8 °C as estimated by nn model, and therefore duplex formation should be possible for each sequence after annealing to 4°C.

Each truncated sequence was annealed to and crosslinked with CNVK-ODN as described in Section 3.1 and was then analyzed by denaturing PAGE and visualized with SYBR Gold staining **(Figure 2.4b)**. It should be noted that the percentage of acrylamide used in the gel was only high enough to effectively resolve full-length, non-truncated ODNs. We observed that each truncated duplex was capable of crosslinking to some extent, although there was a steep drop-off in crosslinking efficiency as the size of the duplex decreased. Other than the

reaction between full-length comp-ODN/^{CNV}K-ODN, no crosslinking reactions entirely consumed the ^{CNV}K-ODN strand despite the slight excess of complementary strand. Interestingly, the +4 nt duplex crosslinking product migrated to a higher position in the gel than the +10 nt duplex, despite having less mass overall. We think this effect is likely caused by the denaturation of the +4 duplex, creating a branched structure that would migrate slower through a gel than an annealed duplex. The full-length duplex may have been stable enough to stay annealed and could have migrated farther through the gel relative to its length.

This experiment demonstrated that ^{CNV}K crosslinking is possible but unfavorable in very short dsRNA duplexes. We were therefore interested in determining whether the crosslinking rate could be improved by increasing the stability of each duplex by altering reaction conditions. To this end, we repeated the crosslinking reactions as described above with a 1:1 ratio of comp-ODN/^{CNV}K-ODN but increased concentrations of Mg^{2+} ions (10 mM, 50 mM, 100 mM, and 500 mM). After analyzing the crosslinking reactions by denaturing PAGE, we attempted to quantify the crosslinking efficiency by measuring the fluorescence intensity of crosslinked and non-crosslinked ^{CNV}K-ODN (F_{dsDNA} and F_{ssDNA} respectively) using FIJI. To reflect the increased fluorescence of dsDNA relative to ssDNA, the percent of crosslinked DNA was approximated with the following equation:

$$\% \text{ Crosslinked} = \frac{\frac{F_{dsDNA}}{2}}{\frac{F_{dsDNA}}{2} + F_{ssDNA}} \times 100$$

The crosslinking rates for each truncated duplex are shown in **Figure 2.5a**. Increased Mg^{2+} did seem to mildly improve the crosslinking rates of the +2 nt duplex, but was unable to force the reaction to completion. No effect on crosslinking rate was observed for the remaining duplexes. Because the duplexes were short in length, it is possible that the original Mg^{2+}

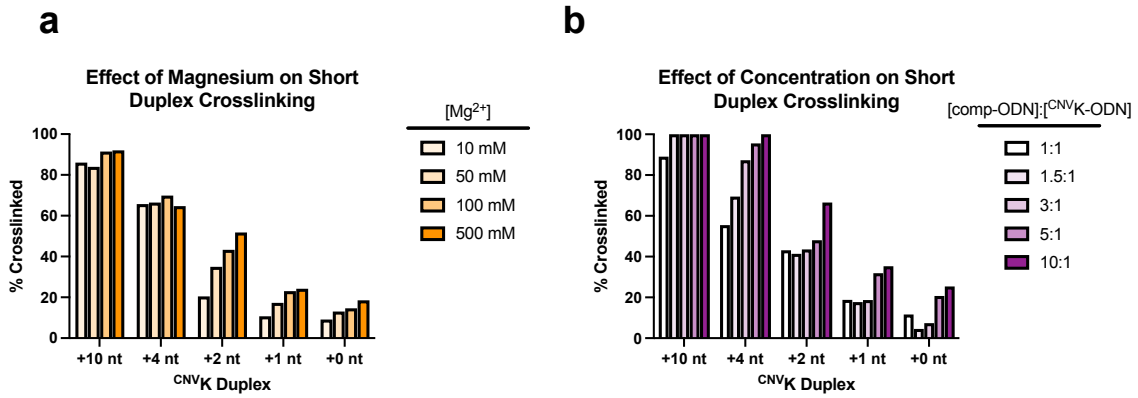


Figure 2.5: FIJI analysis of truncated DNA duplex crosslinking with increasing concentrations of **(a)** magnesium ions or **(b)** complementary ODN strand. Y-axis corresponds to the intensity of the crosslinked duplex band compared to excess ssDNA CNVK-ODN.

concentration was sufficient to saturate the phosphate backbones of each oligonucleotide, which could explain why no increase in crosslinking was observed for most duplexes.

We next tested whether increasing the concentration of comp-ODN strand relative to CNVK-ODN would improve duplex formation and crosslinking rates. The crosslinking reactions were repeated with a 1, 1.5, 3, 5, and 10 x molar excess of truncated comp-ODNs. Reactions were once again analyzed by denaturing PAGE and FIJI as described above. Crosslinking rates for each truncated duplex are displayed in **Figure 2.5b**. An overall trend emerged that increasing the excess of comp-ODN strand improved crosslinking rates. In the case of the full-length duplex, even the slightest excess of comp-ODN was able to drive the crosslinking reaction to completion. A 10x excess of comp-ODN^{-6} was sufficient to drive the crosslinking of the +4 nt duplex to completion, which stands in contrast to our experiment with increased Mg^{+2} concentration. A 10x excess of comp-ODN also increased the crosslinking of the +2, +1, and +0 duplexes, but did not result in the total consumption of the CNVK-ODN strand. Overall, this demonstrates that the unfavorable interaction between a short CNVK-

oligonucleotide and complement can be overcome to a degree with large excesses of one strand. However, these gains are limited in minimal duplexes.

2.4 ^{CNV}K-crRNAs

2.4.1) Design of a ^{CNV}K-crRNA/sgRNA Crosslinking Pair

The results of the previous section demonstrated that ^{CNV}K crosslinking in dsRNA duplexes is possible and helped to define the limits of crosslinking in short duplexes. Our next goal was to implement this information to construct a synthetic sgRNA through crosslinking.

Based on the results of the previous section and existing literature, we had a few parameters to consider in our ^{CNV}K-sgRNA design. One of our main concerns was that the introduction of a ^{CNV}K residue might perturb interactions between the sgRNA and Cas9 protein. To date, there is no crystallographic or other substantial structural information on the impact of ^{CNV}K crosslinks on oligonucleotide duplex structure. Based on the destabilizing effect ^{CNV}Ks have on duplex T_M prior to crosslinking, it is likely that their introduction could perturb the dsRNA structure. Therefore, we wanted to incorporate the ^{CNV}K residue into the sgRNA structure outside of the sgRNA's core interactions with the Cas9 protein [72]. Secondly, the Cas9 sgRNA has four stem-loop structures that could each potentially be split and replaced with a ^{CNV}K crosslink (Section 1.7). Of the four, it would be most optimal to replace the repeat-antirepeat duplex, which is the most proximal stem-loop to the gRNA spacer sequence. This would minimize the size of the sgRNA "variable" component, making it easier to prepare libraries with our platform.

Based on the results of our truncated-duplex crosslinking experiment, none of the existing stem-loops in the sgRNA structure appeared long enough for effective crosslinking. Therefore, a final consideration was that any stem-loop would invariably need to be

lengthened. Fortunately, the sgRNA structure tolerates extensions to any of its stem-loops, a property which has previously been used to introduce functional aptamers to sgRNAs [207-209]. Extending the repeat-antirepeat stem-loop has additionally been reported to improve Cas9 cleavage rates [210].

With these parameters in mind, we developed our first design for a CNVK-sgRNA **(Figure 2.6)**. We split the sgRNA at the repeat-antirepeat stem-loop and extended this structure by 10 bp, resulting in an uninterrupted dsRNA duplex of 14 nt to crosslink. This also neatly split the sgRNA into "crRNA" and "tracrRNA" halves, to which they will be referred from this point. The extended repeat-antirepeat structure was designed with the assistance of the Vienna RNA websuite [211] to avoid generating unintended stable interactions with other elements of the sgRNA. The extended duplex should not interact with Cas9, and the nearest functional interaction between Cas9 and gRNA is formed by Arg340 (Cas9) and the

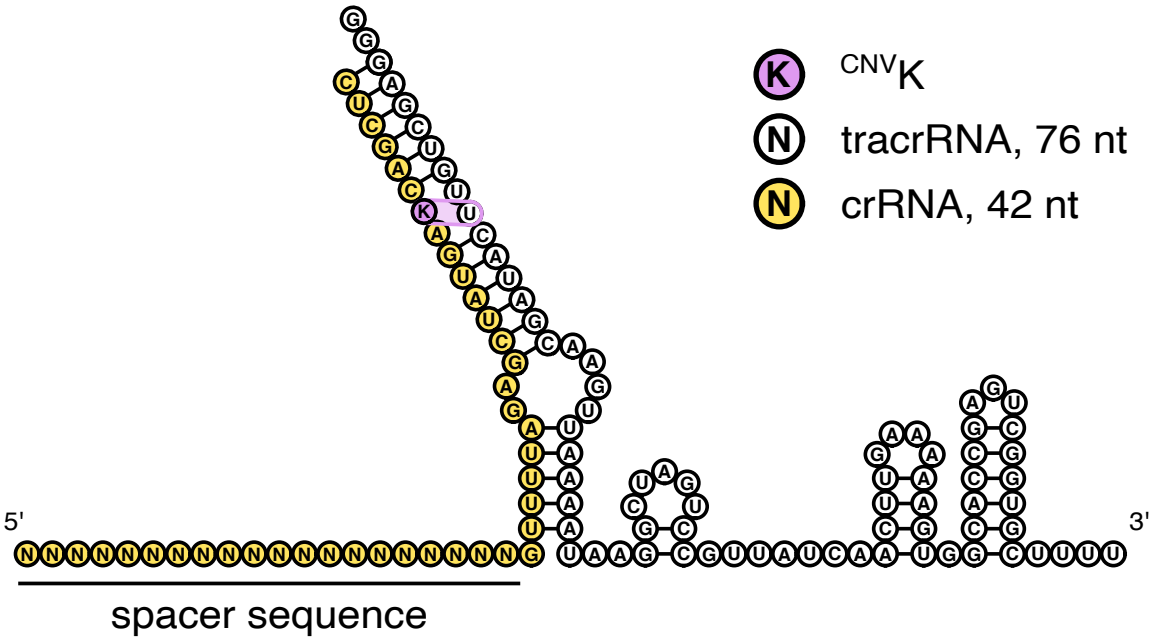


Figure 2.6: Diagram of CNVK-crRNA/tracrRNA interaction. The 76 nt -tracrRNA (white) hybridizes to the 42 nt crRNA (yellow) through the anti/direct repeat stem. The CNVK residue (K) is highlighted, and the crosslinking interaction it forms with target uracil base is circled in purple. The 20 nt variable spacer sequence in the crRNA is underlined.

phosphate linkage between C40 and A41 (gRNA), 7 nt away from the crosslink site [70, 72]. Lastly, the ^{CNV}K residue itself was incorporated 6 nt from the 3′ end of crRNA, which should provide a sufficient amount of dsRNA duplex to stabilize its interactions.

Between the two RNA strands that compose our ^{CNV}K-sgRNA, we chose to incorporate the ^{CNV}K residue into the crRNA for this round of design. This allowed us to prepare the larger tracrRNA component by IVT of a DNA template, which reduced the costs of RNA synthesis during this development phase. In our final designs, the ^{CNV}K nucleotide will be incorporated into the tracrRNA so that the variable crRNA component can be synthesized without needing to include it. For the time being, the 5′ end of the tracrRNA therefore begins with a G trinucleotide, which is required for efficient transcription during IVT reactions. For the ^{CNV}K-sgRNA target sequence, we decided to use an established spacer sequence for the human Empty Spiracles Homeobox 1 (EMX1) gene (GAGUCCGAGCAGAAGAAGAA), with which our group has previous experience [206].

2.4.2) ^{CNV}K-crRNA/tracrRNA Crosslinking

After designing the ^{CNV}K-crRNA/tracrRNA pair, our next goal was to determine if the two RNAs could be efficiently crosslinked to form a ^{CNV}K-sgRNA. A simple schematic showing the intended reaction is shown in **Figure 2.7a**. Informed by the work of the previous section, we prepared a solution with ^{CNV}K-crRNA (1 μM) and a slight excess of the non-^{CNV}K tracrRNA (1.5 μM) in 100 mM NaCl, 10 mM $MgCl^{2+}$, and 10 mM Tris-HCL, which was annealed and crosslinked as described in Section 6.8.1. The crosslinking reaction was then analyzed by denaturing PAGE alongside both reactant RNAs and a non-irradiated control **(Figure 2.7b)**. Excitingly, we observed the formation of a new product in the crosslinking

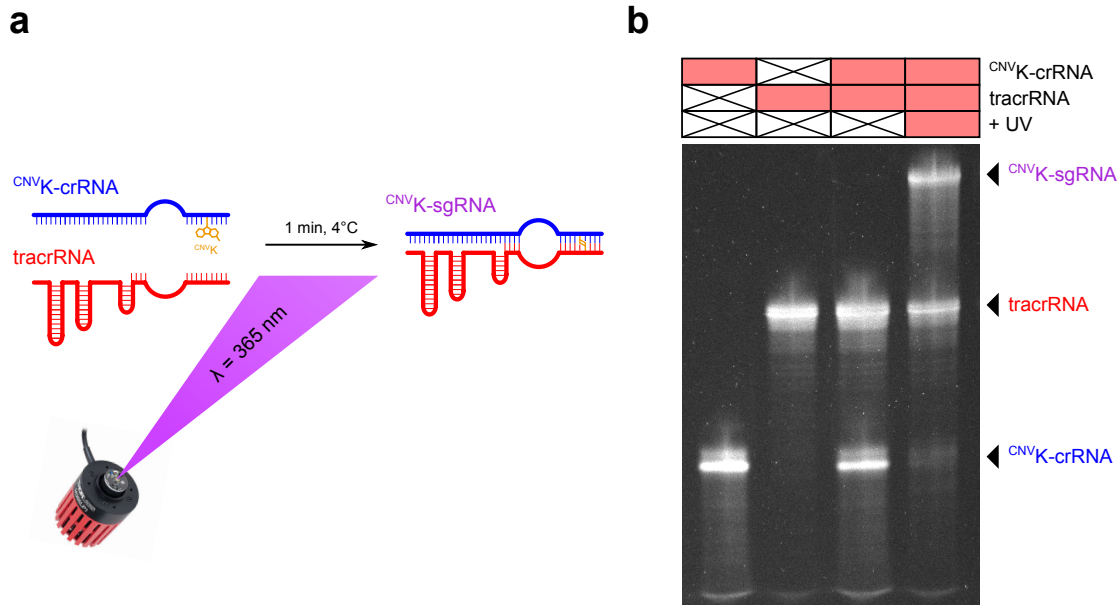


Figure 2.7a: Schematic of CNVK-crRNA/tracrRNA crosslinking. The crRNA (blue) and tracrRNA (red) hybridize through their anti/direct stem interaction. This interaction brings the CNVK nucleotide into the proximity of its target pyrimidine base, and crosslinking stabilizes the interaction between the two RNAs. Reaction is catalyzed by 1 min of 365 nm irradiation at 4 °C. **Figure 2.7b:** Denaturing PAGE analysis of CNVK-crRNA/tracrRNA crosslinking reaction. The boxes above the gel indicate the presence (red) or absence (x) of crRNA, tracrRNA, and 365 UV light in each reaction. Arrows to the right of the gel highlight the bands corresponding to CNVK-crRNA, tracrRNA, and CNVK-sgRNA.

reaction which was absent from the non-crosslinked control. This product was subsequently purified and then analyzed by HPLC-MS, which confirmed its identity as the expected CNVK-sgRNA (Appendix 1). The fact that only one main product was observed in the crosslinking reactions suggests that the repeat-antirepeat interaction is specific and robust enough to prevent side reactions from occurring.

We previously observed that increasing the *relative* concentration of non-CNVK strand during CNVK crosslinking had a positive impact on efficiency. We were also interested in exploring whether increasing the concentration of both constituent oligonucleotides would impact the reaction dynamics. Until this point, we performed all crosslinking reactions with a 1 µM concentration of CNVK-RNA. This was purposefully done to avoid the formation of a crRNA/crRNA duplex or misaligned crRNA/tracrRNA duplex, which could be promoted in

highly concentrated reactions and erroneously crosslinked. However, concentrated CNVK-sgRNA crosslinking reactions are required to prepare material for further assays at a reasonable scale. We therefore repeated our crosslinking reaction at a 5 µL scale with either 1, 5, 10, 50, or 100 µM CNVK-crRNA and a 1.5x molar excess of tracrRNA. After crosslinking, the equivalent of 5 pmol of CNVK-crRNA from each reaction was analyzed by denaturing PAGE **(Figure 2.8)**. Despite increasing the concentration of RNA 100-fold, we did not observe unexpected crosslink products by gel or a significant decrease in crosslinking efficiency. This demonstrates that the dsRNA interaction between our crRNA/tracrRNA is robust enough to hybridize in a wide range of concentrations, which is also convenient for preparing CNVK-sgRNAs at either analytical or preparative scale.

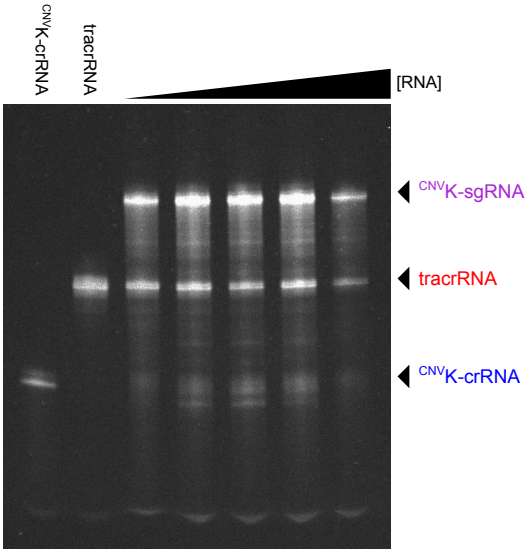


Figure 2.8: Denaturing PAGE analysis of CNVK-crRNA/tracrRNA crosslinking reactions performed with 1, 5, 10, 50, and 100 µM CNVK-crRNA concentration. Black arrow above the gel indicates increasing concentration of RNA in the five reactions. CNVK-crRNA and tracrRNA have been run on the left side of the gel and labeled accordingly for reference. Arrows to the right of the gel highlight the bands corresponding to CNVK-crRNA, tracrRNA, and CNVK-sgRNA.

2.4.3) Alternative CNVK-crRNA/tracrRNA Stem Designs

Our first design for a CNVK-crRNA/tracrRNA generated a functional pair that could crosslink to completion with only a small excess of non-CNVK strand. We were interested in optimizing our reaction to minimize the amount of excess RNA required for complete crosslinking, which could make it possible to prepare and apply CNVK-sgRNAs without an intermediate purification step. We considered that the overall crosslinking efficiency could

be improved by increasing the stability of the crRNA/tracrRNA interaction, which could easily be accomplished by lengthening the repeat-antirepeat duplex. However, this would increase the synthetic burden of both strands and likely reduce the activity of a ^{CNV}K-sgRNA relative to a non-^{CNV}K dgRNA.

As an alternative to increasing the length of repeat-antirepeat, we wanted to explore chemical modification of the crRNA to strengthen the crRNA/tracrRNA interaction. To this end, we substituted the 3′ terminal 10 nt of ^{CNV}K-crRNA with 2′OMe residues, which would both strengthen its interaction with the tracrRNA and allow us to explore whether the ^{CNV}K-sgRNA platform is compatible with chemical modification. The resulting ^{CNV}K-2′OMe-crRNA **(Figure 2.9a)** was crosslinked in a 1:1 molar ratio with tracrRNA in a 1 μM solution as described earlier and was subsequently analyzed by denaturing PAGE **(Figure 2.9b**). We encouragingly observed the formation of only one crosslink product, similar to our first design. However, the addition of the 2′OMe residues was not sufficient to drive the reaction to completion and a significant excess of both ^{CNV}K-2′OMe-crRNA and tracrRNA remained. While the 2′OMe modifications used here did not significantly improve the crosslinking efficiency, they did not hinder it either, suggesting that the ^{CNV}K-sgRNA is a viable platform for chemical modification.

Because modification of the ^{CNV}K-crRNA did not improve crosslinking efficiency, we considered that the length of repeat-antirepeat duplex interaction in our original design was already more than sufficient to template the ^{CNV}K interaction. Because of this, we wanted to determine if it would be possible to reduce the length of the repeat-antirepeat duplex while maintaining high crosslinking efficiency. Reducing the size of either RNA would reduce the synthetic burden of the entire design, making our method of sgRNA construction cheaper and

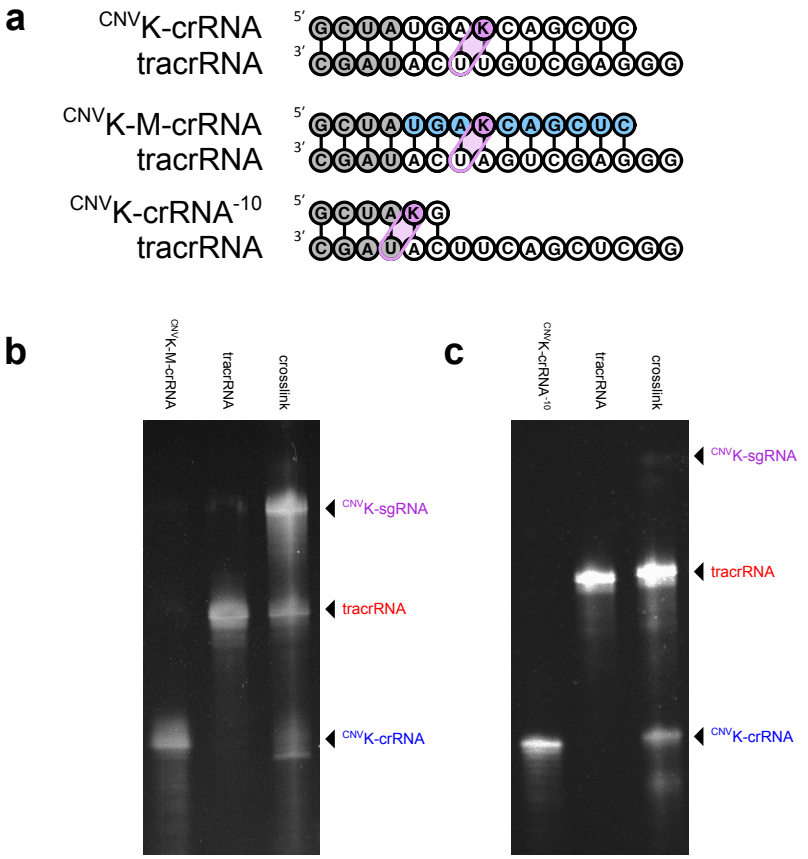


Figure 2.9a: Diagram showing the repeat-antirepeat stem interaction between tracrRNA and CNVK-crRNA, CNVK-M-crRNA, or CNVK-crRNA-10. The CNVK residue (K) is highlighted, and the crosslinking interaction it forms with the target uracil base is circled in purple. Nucleotides replaced with a 2′OMe modified residue are colored blue. Nucleotides shown in grey are conserved between all gRNAs. **Figure 2.9b,c:** Denaturing PAGE analysis of CNVK-crRNA/tracrRNA crosslinking reaction performed with CNVK-M-crRNA **(b)** or CNVK-crRNA-10. Text above the gel indicates whether lanes correspond to crRNA, tracrRNA, or the crosslinking reaction with both oligonucleotides. Arrows to the right of the gel highlight the bands corresponding to CNVK-crRNA, tracrRNA, and CNVK-sgRNA.

more accessible. The results from Section 2.3.2 suggest that there is a minimum duplex size requirement for efficient crosslinking, however it is possible that the other interactions formed between crRNA and tracrRNA could promote hybridization and crosslinking. We therefore synthesized CNVK-crRNA with a 10 nt 3′ truncation (CNVK-crRNA-10). This truncation drastically reduced the size of the repeat-antirepeat duplex formed with the tracrRNA, as illustrated in **Figure 2.9a**. We attempted to crosslink CNVK-crRNA-10 and tracrRNA using the same conditions described for CNVK-2′OMe-crRNA crosslinking. A denaturing PAGE analysis of this reaction is shown in **Figure 2.9c**. Crosslinking and sgRNA

formation with ^{CNV}K-crRNA^{-10} was incredibly diminished compared to either ^{CNV}K-crRNA or ^{CNV}K-2′OMe-crRNA. The ^{CNV}K-crRNA^{-10} additionally appeared to form a self-crosslinked product which was running lower than the original ^{CNV}K-crRNA^{-10} in the gel. Because the results of this reaction did not outperform the crosslinking efficiency observed in minimal duplexes shown earlier **(Figure 2.4b)**, we decided against attempting to further optimize this reaction.

2.4.4) Functional Characterization of ^{CNV}K-sgRNA

The experimental work in this section resulted in the development of two functional crRNAs (^{CNV}K-crRNAs, ^{CNV}K-2′OMe-crRNA) which can crosslink with a partner tracrRNA to produce a single ^{CNV}K-sgRNA product. Our next aim was to perform functional assays with these crosslink products to determine if they were compatible with Cas9 protein. To this end, we prepared two primers to PCR amplify a 622 bp DNA fragment which contained the EMX1 spacer sequence target (Appendix 4) for an *in vitro* Cas9 DNA digest assay. Cleavage of the PCR amplicon with Cas9-RNP^{EMX1} will yield 314 and 308 bp fragments, which were intentionally designed to comigrate on gel to make post-analysis measurement easier. We next prepared an EMX1 sgRNA generated through IVT (IVT-sgRNA^{EMX1}) to use as a positive control in digestion reactions. To determine if the ^{CNV}K crosslink has a positive impact on Ca9s/gRNA activity, we also synthesized two non-^{CNV}K crRNAs which were otherwise isosequential to our ^{CNV}K-crRNAs (crRNA^{EMX1}, 2′OMe-crRNA^{EMX1}). These could be annealed with the tracrRNA to form a non-crosslinked, dual-guide RNA control (dgRNA^{EMX1} and 2′OMe-dgRNA^{EMX1}).

While designing our *in vitro* digest assay, we were aware of the common practice of applying excessive quantities of Cas9-RNP to DNA substrate to account for low, almost single

enzymatic turnover of Cas9. To accurately compare the activity of gRNAs however, we wanted to avoid a scenario in which excess Cas9-RNP saturates the DNA substrate and causes total cleavage regardless of its catalytic rate. To account for this, we prepared titrated *in vitro* digest reactions with recombinant Cas9 protein and one of IVT-sgRNA, CNVK-sgRNA, 2′OMe-CNVK-sgRNA, dgRNA, or 2′OMe-dgRNA. A stock solution of each Cas9-RNP was prepared, diluted in serial, and then applied to 100 nm of EMX1 PCR product as described in Section 6.10.1. The reactions were subsequently separated on agarose gel and analyzed with FIJI to quantify DNA cleavage rates **(Figure 2.10)**.

Each Cas9-gRNA complex was capable of catalyzing DNA cleavage, confirming that our CNVK-sgRNAs were functional constructs. Furthermore, at every concentration of Cas9-RNP, the CNVK-sgRNAs outperformed their non-crosslinked dgRNA counterparts, suggesting that crosslinking improves gRNA activity in this design. The CNVK-sgRNAs additionally had comparable activity to the IVT-sgRNA positive control. There was no amount of Cas9-RNP

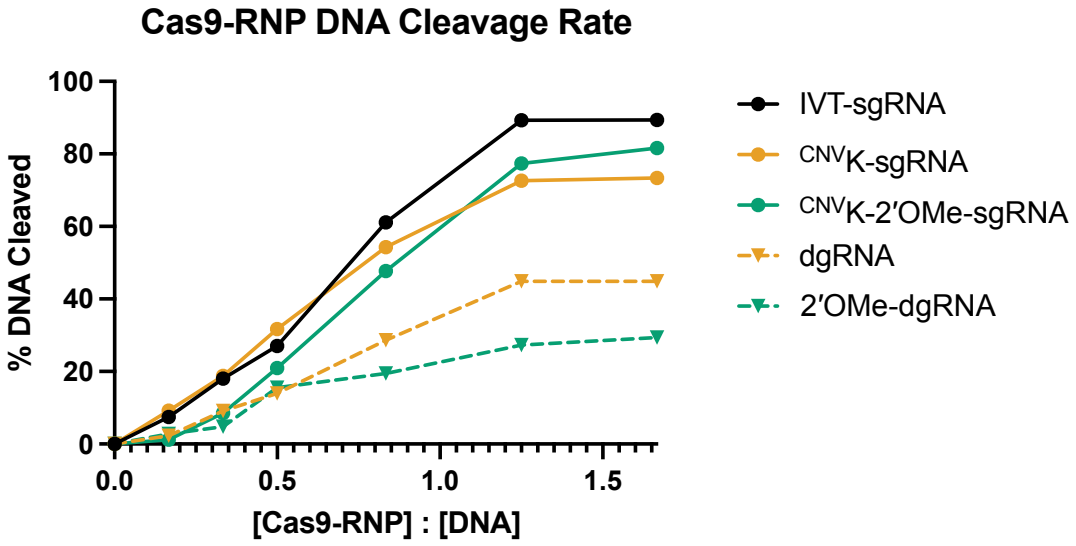


Figure 2.10: EMX1 PCR product cleavage efficiency by Cas9-RNP at varying concentrations of Cas9-RNP. X-axis corresponds to the relative molar concentration of Cas9-RNP compared to the 3 nM DNA substrate. Y-axis corresponds to the percent of DNA cleaved after 1 h incubation as determined with FIJI. n = 1.

in this experiment that was capable of totally cleaving the EMX1 PCR product. This suggests that performing *in vitro* digest reactions with an equimolar amount of Cas9-RNP and DNA substrate in the future will be appropriate.

As a final test of this ^{CNV}K-sgRNA design, we attempted a gene knockout of EMX1 to determine if the crosslink would also be tolerated in cells, where the relatively lower concentration of Cas9-RNP could influence the dynamics of ^{CNV}K-sgRNA and Cas9 interaction. For gene editing in mammalian cells, there are several viable formats of Cas9 source (stably-expressing cell line, plasmid, mRNA, or protein) and delivery method (lipofection, electroporation). For our purposes, electroporation of pre-incubated Cas9-RNP complexes seemed the most appropriate and had the added benefit of being a widely used and relevant strategy for preparing gene-edited cell lines. To this end, HEK293 cells were electroporated with Cas9 protein which was incubated with either IVT-sgRNA or ^{CNV}K-sgRNA as described in Section 6.11.2. Transfected cells were then left to recover for 72 hours before harvesting, genome extraction, and InDel quantification.

Several assays have been developed since the advent of Cas9 gene editing for detecting InDel mutations to estimate or quantify gene knockouts [212]. We decided to employ TIDE (Tracking of Indels by DEcomposition) [204], a powerful but cost-effective strategy that accurately quantifies editing efficacy and identifies the predominant types of InDels in a targeted pool of cells. TIDE compares Sanger sequencing chromatograms of edited populations to an unedited control and reconstructs the spectrum of InDel mutations that could account for aberrant sequence reads observed downstream of the Cas9 cleavage site [204]. This method is widely accepted [213] and can generate high-quality population-level data from Sanger sequencing of PCR products.

We therefore amplified the EMX1 target locus from the genomic extracts of our edited populations and a negative control population using the earlier described primers. These PCR amplicons were purified, Sanger sequenced, and analyzed by TIDE. The InDel spectrums generated with TIDE analysis for IVT-sgRNA and CNVK-sgRNA edited populations are shown in **Figure 2.11a** and **Figure 2.11b** respectively. Electroporation of both Cas9-RNP species resulted in the generation of InDel mutations at the EMX1 target site, indicating that both gRNAs are functional. While the pattern of InDel mutations varies slightly between the two populations, the overall editing efficiency of IVT-sgRNA and CNVK-sgRNA are remarkably similar (41.3% and 43.8% respectively). This experiment was only

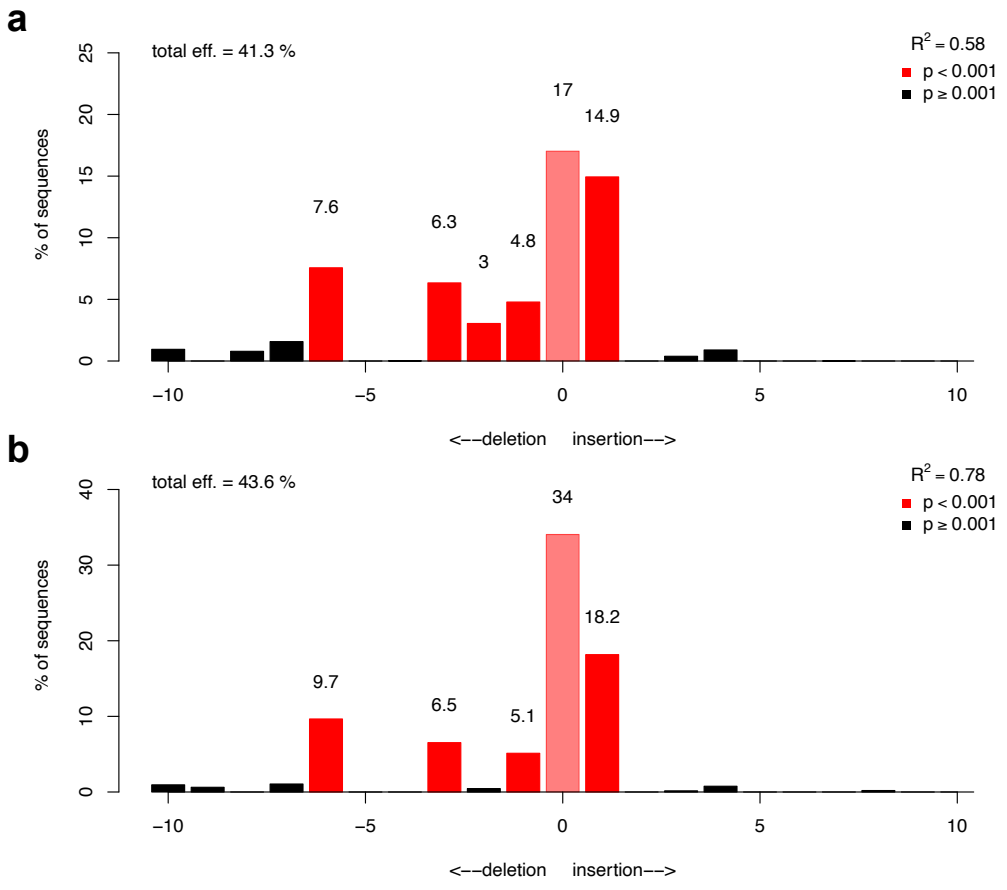


Figure 2.11: TIDE analysis of InDels generated by Cas9-RNP cleavage of EMX1 target in HEK293 cells. Sanger sequence chromatograms of cells electroporated with **(a)** Cas9/IVT-sgRNA and **(b)** Cas9/CNVK-sgRNA were compared to a non-transfected control. Plots describe the generation of insertions, deletions, and their size around the Cas9 target sequence. The total InDel generation efficiency is described in the top left of each plot. Deconvolution settings: alignment window (100-267), decomposition window (295-581), indel size (10).

performed to determine if our first ^{CNV}K-sgRNA design was functional in cells and was not replicated, so we cannot generalize the activity of ^{CNV}K-sgRNAs from it. However, it is promising that the two gRNAs have comparable editing activity in this experiment. This demonstrates that our ^{CNV}K-sgRNA construction method, in addition to robustly and rapidly generating crosslinked material, also produces gRNAs with high levels of activity.

2.5 ^{CNV}K-tracrRNA

2.5.1) Design of a crRNA/^{CNV}K-tracrRNA Pair

The work in the previous section resulted in the design of a functional ^{CNV}K-sgRNA, which could be effectively crosslinked in a 1 min UV-induced reaction. A combination of *in vitro* and cell-based assays suggests that ^{CNV}K-sgRNAs have comparable activity with IVT-sgRNAs. While we were optimizing the construction of ^{CNV}K-sgRNAs, it was sensible to incorporate the ^{CNV}K residue into the crRNA strand, which is shorter and easier to synthesize than the tracrRNA strand. Now that we had a functional design, our new goal was to synthesize a ^{CNV}K-tracrRNA strand instead. Because the tracrRNA sequence is constant between different genomic targets, a single ^{CNV}K-tracrRNA could be used as a photo-crosslinking platform for any compatible crRNA regardless of their spacer sequence, facilitating sgRNA library production.

To this end, we synthesized a 74 nt long tracrRNA which retained the extended stem region we introduced to our previous ^{CNV}K-sgRNA pair. In this case however, a ^{CNV}K residue was incorporated into position 67 of the sequence, and an adenine residue was introduced to position 68 to account for the uracil residue in the crRNA strand which would be required for crosslinking. A schematic showing the sequence of the new ^{CNV}K-tracrRNA and its interactions with a compatible crRNA is shown in **Figure 2.12**.

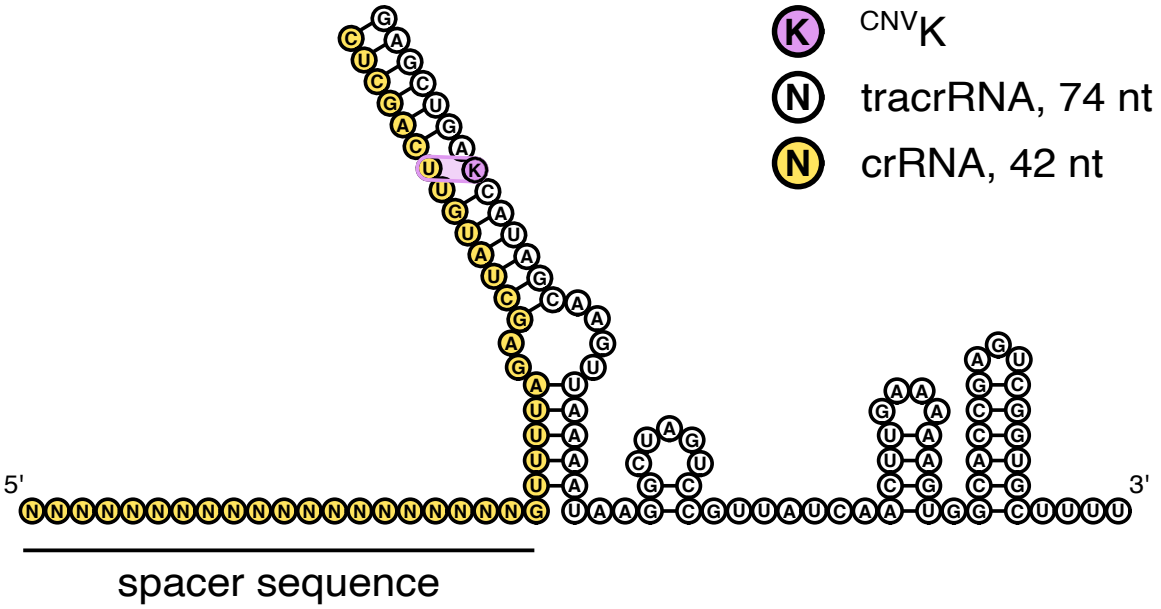


Figure 2.12: Diagram of crRNA/CNVK-tracrRNA interaction. The 74 nt CNVK-tracrRNA (white) hybridizes to the 42 nt crRNA (yellow) through the anti/direct repeat stem. The CNVK residue (K) is highlighted, and the crosslinking interaction it forms with target uracil base is circled in purple. The 20 nt variable spacer sequence in the crRNA is underlined.

Prior to synthesizing the CNVK-tracrRNA described above, we had attempted to create a CNVK-tracrRNA with a similar sequence and CNVK nucleotide 6 nt from the 5′ termini (CNVK-tracrRNA^{9700}). The incorporation of the CNVK residue during synthesis generated a large quantity of failure sequences of similar lengths to CNVK-tracrRNA^{9700}. We attempted to purify the correct CNVK-tracrRNA^{9700} product from its failure sequences, but a significant amount of heterogeneity remained after one round of RP-HPLC and two PAGE purifications. While designing the CNVK-tracrRNA shown in **Figure 2.12**, we therefore decided to cap the sequence with a 5′ C8 chain modification to increase the hydrophobicity of the full-length product, making it easier to purify by RP-HPLC. We additionally added a 3′ propanol modification to reduce the susceptibility of the tracrRNA to RNA exonucleases. The resulting RNA was easily purified with one round of RP-HPLC, and the difference in post-purification quality between CNVK-tracrRNA and CNVK-tracrRNA^{9700} is shown in **Figure 2.13**.

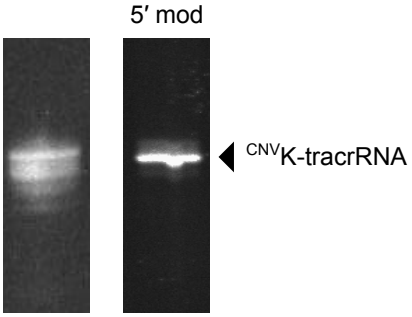


Figure 2.13: Denaturing PAGE analysis of post-purification ^{CNV}K-tracrRNAs. The ^{CNV}K-tracrRNA with 5′ hydrophobic modification for enhanced purification is labeled. Arrow to the right of the gel highlights the ^{CNV}K-tracrRNA band.

After preparing ^{CNV}K-tracrRNA, we synthesized a compatible crRNA to test the crosslinking efficiency of our new ^{CNV}K-tracrRNA. We decided to continue using the EMX1 spacer sequence from our previous designs to simplify the design process. Uracil residues were incorporated into positions 35 and 36 of the crRNA (as shown in **Figure 2.12**) to account for the tracrRNA's ^{CNV}K residue, but the sequence was otherwise isosequential with ^{CNV}K-crRNA from the previous section.

After purifying the new crRNA^{EMX1}, we prepared a 1 M solution of ^{CNV}K-tracrRNA and crRNA^{EMX1} and crosslinked them as previously described. The crosslinking reaction was subsequently analyzed by denaturing PAGE, as shown in **Figure 2.14**. We observed the formation of only one product after the crosslinking reaction, consistent with our previous experience with crRNA/tracrRNA crosslinking. Given that the sequences of the two ^{CNV}K-sgRNA designs differ in only two positions, this result was not unexpected. However, the larger tracrRNA sequence is more likely to form monomeric or homodimeric structures than the crRNA in the conditions used during crosslinking. If this was the case, it could have been possible for a ^{CNV}K crosslink to stabilize a tracrRNA/tracrRNA interaction, but this was encouragingly not observed.

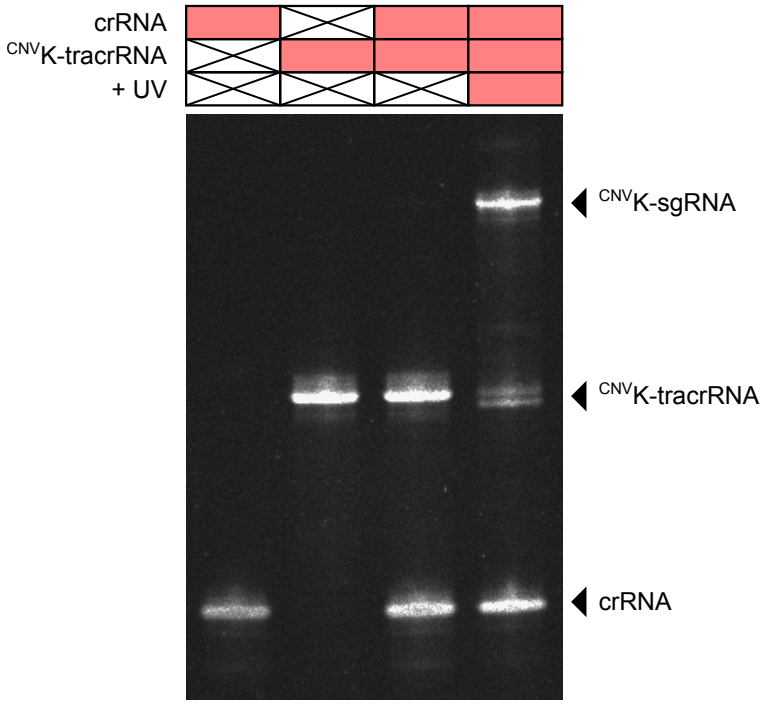


Figure 2.14: Denaturing PAGE analysis of crRNA^{EMX1}/^{CNV}K-tracrRNA crosslinking reaction. The boxes above the gel indicate the presence (red) or absence (x) of crRNA, tracrRNA, and 365 UV light in each reaction. Arrows to the right of the gel highlight the bands corresponding to crRNA, ^{CNV}K-tracrRNA, and ^{CNV}K-sgRNA.

As with our previous designs, crosslinking equimolar quantities of crRNA^{EMX1} and ^{CNV}K-tracrRNA was not sufficient to drive the reaction to completion. In addition to unconsumed ^{CNV}K-tracrRNA, we also observed by gel the formation of a second smaller band beneath the main ^{CNV}K-tracrRNA fraction (**Figure 2.14**). This is likely the product of UV-catalyzed ^{CNV}K *trans-cis* isomerization (Section 1.2.2), as the two isomers can subtly change the migration of ^{CNV}K-oligonucleotides [23]. The non-reactive *cis* ^{CNV}K stereoisomer does not hinder crosslinking reactions as it rapidly photo-isomerizes back to the functional *trans* stereoisomer, and only appears when there are excess non-crosslinked ^{CNV}K nucleotides in a reaction [23]. We therefore repeated this crosslinking reaction with a 5x molar excess of the crRNA^{EMX1} to drive the crosslinking reaction to completion. This reaction was then analyzed and purified by RP-HPLC (**Figure 2.15a**), from which we identified and collected two fractions. To confirm their identity, these two fractions were analyzed on denaturing PAGE

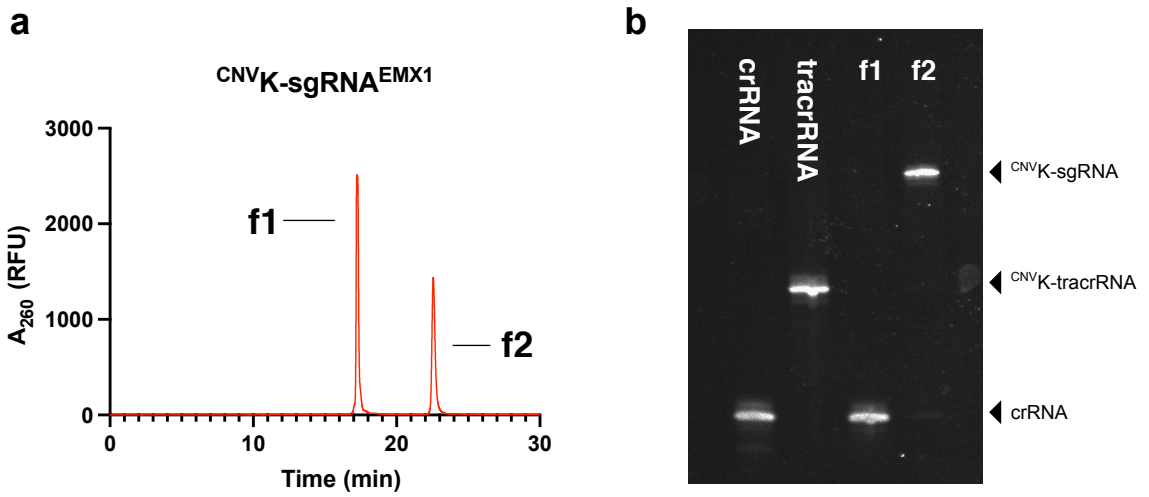


Figure 2.15a: RP-HPLC analysis and purification of crRNA^{EMX1}/^{CNV}K-tracrRNA crosslinking reaction. Purification was performed with an Agilent HPLC according to Section 6.3. Two fractions were collected from this purification, identified with either 'f1' or 'f2'. **Figure 2.15b**: Denaturing PAGE analysis of RP-HPLC purified crRNA^{EMX1}/^{CNV}K-tracrRNA crosslinking reaction from Figure 5.4a. Lanes are labeled with corresponding reactant RNA (crRNA or ^{CNV}K-tracrRNA) or fraction from HPLC purification. Arrows to the right of the gel highlight the bands corresponding to crRNA, ^{CNV}K-tracrRNA, and ^{CNV}K-sgRNA.

alongside the reactant oligonucleotides (**Figure 2.15b**). Fraction 1 corresponded to the excess crRNA^{EMX1} used in the crosslinking reaction, which we were able to collect and reuse in other reactions. The larger product in fraction 2 was confirmed to be successfully crosslinked ^{CNV}K-sgRNA by HPLC-MS (Appendix 1). By HPLC, we observed that the ^{CNV}K-tracrRNA was entirely consumed in this reaction.

After confirming that our new crRNA/ ^{CNV}K-tracrRNA pair could be efficiently crosslinked and purified, we wanted to determine whether the dsRNA stem region interaction between the two could be shortened. We had already explored this unsuccessfully during our ^{CNV}K-crRNA design process (Section 2.4.3), and therefore decided to reduce the current design by a more conservative 5 bp instead (**Figure 2.16a**). This was accomplished by synthesizing a truncated crRNA molecule (crRNA$^{EMX1, -5}$) to avoid resynthesizing the larger ^{CNV}K-tracrRNA. We knew from our previous experiments of crosslinking short duplexes (Section

2.3.2) that the reduced dsRNA interaction would be unfavorable, so we attempted to compensate for this by annealing crRNA$^{EMX1, -5}$ in 1.5, 3, 5, and 10 x excess of ^{CNV}K-tracrRNA with either 10 mM or 50 mM Mg^{2+}. Crosslinking reactions were carried out as previously described and were analyzed by denaturing PAGE (**Figure 2.16b**). Unlike the crosslinking reaction between full-length ^{CNV}K-tracrRNA/crRNA pair, which produces a single distinct product, the crosslinking reaction with truncated crRNA$^{EMX1, -5}$ produced several products, none of which were clearly the correct ^{CNV}K-sgRNA. Increasing the concentration of the crRNA strand or Mg^{2+} resulted in the formation of more side products, and under no conditions was the ^{CNV}K-tracrRNA entirely consumed.

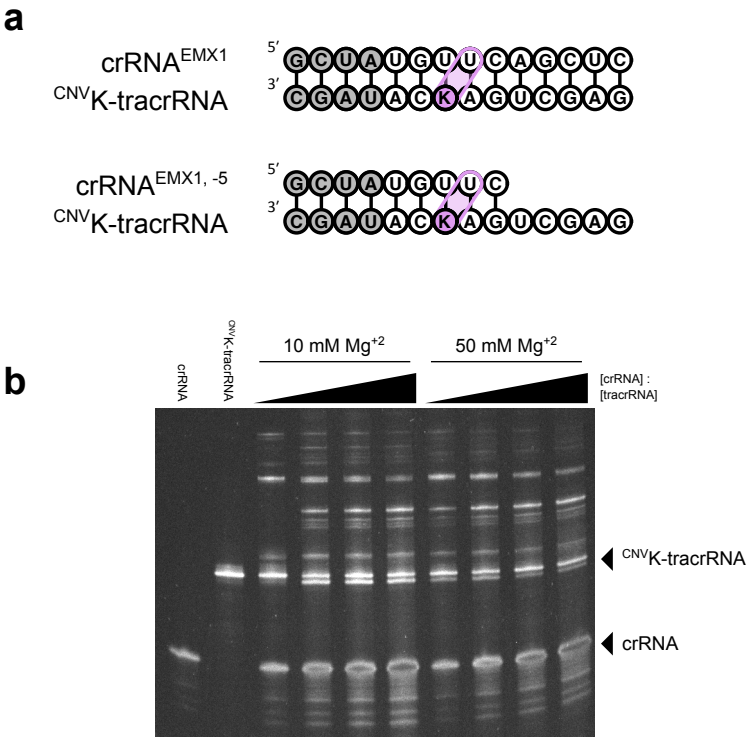


Figure 2.16a: Diagram showing the anti/direct repeat stem interaction between ^{CNV}K-tracrRNA and crRNA^{EMX1} or crRNA$^{EMX1, -5}$. The ^{CNV}K residue (K) is highlighted, and the crosslinking interaction it forms with the target uracil base is circled in purple. Nucleotides shown in grey are conserved between all gRNAs. **Figure 2.16b**: Denaturing PAGE analysis of crRNA$^{EMX1, -5}$/^{CNV}K-tracrRNA crosslinking reactions. The non-crosslinked reactant RNA strands have been run on the left side of the gel and are labeled appropriately. Crosslinking reactions were carried out with either 10 or 50 mM Mg^{+2}, as indicated above respective lanes. The black triangles above each lane indicate an increasing relative concentration of crRNA$^{EMX1, -5}$ (1.5, 3, 5, 10x excess) compared to ^{CNV}K-tracrRNA. Arrows to the right of the gel highlight the bands corresponding to crRNA and ^{CNV}K-tracrRNA.

A ^{CNV}K-platform for sgRNA preparation should be high-yielding, robust, and easy to purify, which this reaction did not appear to be. As such, we did not attempt to further optimize this reaction with crRNA$^{EMX1, -5}$ or purify and identify the products generated by this reaction. All our results so far suggest that a dsRNA duplex of ~12 bp needs to be maintained for efficient crosslinking. Although slightly longer than our truncated designs, the full-length crRNA/^{CNV}K-tracrRNA pair reacts quickly to form a single product that is easily purifiable. We therefore decided to continue working with this design as our platform for sgRNA preparation.

2.5.2) Functional Characterization of crRNA/^{CNV}K-tracrRNA ^{CNV}K-sgRNAs

As earlier, our next aim was to perform functional assays with the new ^{CNV}K-sgRNA to confirm that this design was also compatible with Cas9. To accomplish this, we repeated the *in vitro* DNA digest assay of a 622 bp EMX1 PCR product. Cas9 protein was incubated with either IVT-sgRNA, ^{CNV}K-sgRNA, or non-crosslinked crRNA^{EMX1}/^{CNV}K-tracrRNA (^{CNV}K-dgRNA). Because the sequence of our new crRNA^{EMX1}/^{CNV}K-tracrRNA pair varied from the design of the previous section in only two positions, we decided to reuse the dgRNA prepared in Section 2.4.4 as a non-^{CNV}K dgRNA control.

The Cas9-RNPs were applied to EMX1 PCR product in equimolar amounts and were then incubated for 2 h at 37°C, after which the reactions were analyzed by agarose gel electrophoresis (**Figure 2.17**) and quantified by FIJI analysis. Excitingly, we found that our new ^{CNV}K-sgRNA cleaved the target DNA at a slightly higher rate than the IVT-sgRNA. This was the first instance when one of our ^{CNV}K-sgRNA designs performed as well as a sgRNA in an *in vitro* assay, which we believe is likely due to the improved purification technique of our ^{CNV}K-tracRNA. As in previous experiments, the ^{CNV}K-sgRNA had a higher

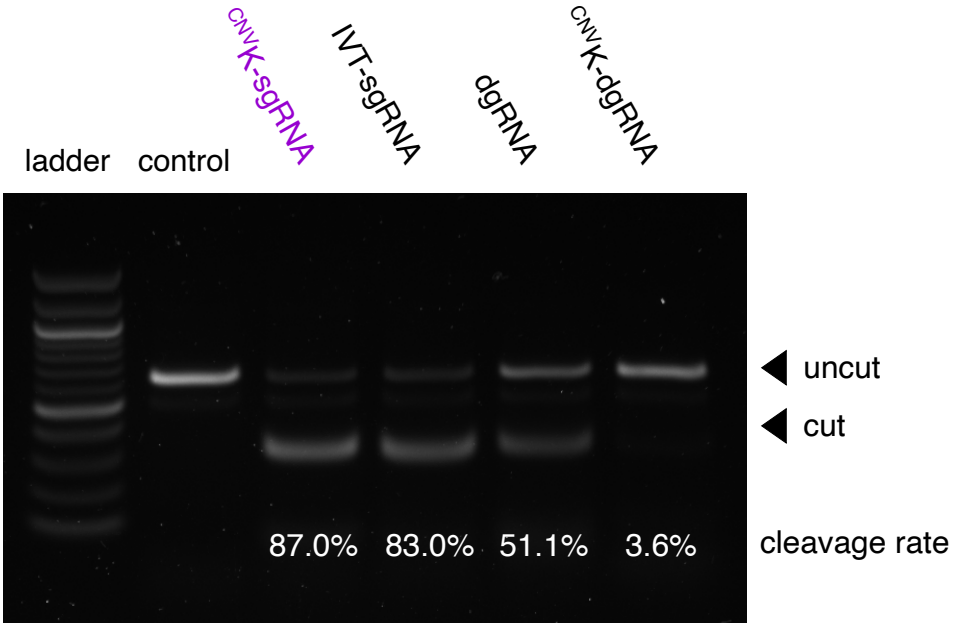


Figure 2.17: Agarose gel analysis of Cas9 *in vitro* DNA digest assay. From left to right, lanes correspond to NEB 100 bp DNA ladder, undigested EMX1 PCR product, and digestion reactions with either CNVK-sgRNA, IVT-sgRNA, dgRNA, and CNVK-dgRNA. Arrows to the right of the gel highlight bands corresponding to 'cut' and 'uncut' DNA. The cleavage rate, as determined with FIJI, is listed below each respective lane.

cleavage rate than the dgRNA control; this demonstrates that crosslink is responsible for the higher activity of the CNVK-sgRNA, as opposed to the elongate crRNA/tracrRNA stem region that both CNVK-sgRNA and dgRNA share. The influence of the CNVK crosslink is further demonstrated by the almost negligible degree of cutting catalyzed by the non-crosslinked CNVK-dgRNA, likely due to the destabilizing effect non-crosslinked CNVK residues have on oligonucleotide duplexes.

After confirming that our new CNVK-sgRNA design was functional with an *in vitro* assay, we next wanted to apply it in cells to determine if it would have a comparable DNA cleavage rate to IVT-sgRNAs. Cas9-RNP prepared with either IVT-sgRNA, CNVK-sgRNA, or dgRNA were electroporated in triplicate into HEK293 cells, which were then incubated for 48 h

before harvesting and genotyping. As earlier, we wanted to quantify the amount of InDels generated from Cas9-RNP cleavage with TIDE. To improve the quality of analysis, a new pair of PCR primers (Appendix 4) was designed to amplify a 700 bp region of EMX1 which would better match TIDE's deconvolution parameters. The Cas9 cleavage site was located 200 bp from the forward primer, leaving a larger portion of the sequence available for the sequence decomposition window.

TIDE analysis of EMX1 knockouts with these three guide RNAs (**Figure 2.18**) revealed that the generation of InDel mutations tracks very well with their *in vitro* cutting activity. Approximately 70% of PCR amplicons contained an InDel from populations that were edited with either IVT-sgRNA or ^{CNV}K-sgRNA, and we did not observe a significant difference between the two populations (unpaired T-test, P = 0.6626). Here again, the dgRNA had reduced gene editing efficiency compared to the two sgRNAs, showing that the introduction of a ^{CNV}K crosslink improves the activity of a dgRNA. Although the tracrRNA component of

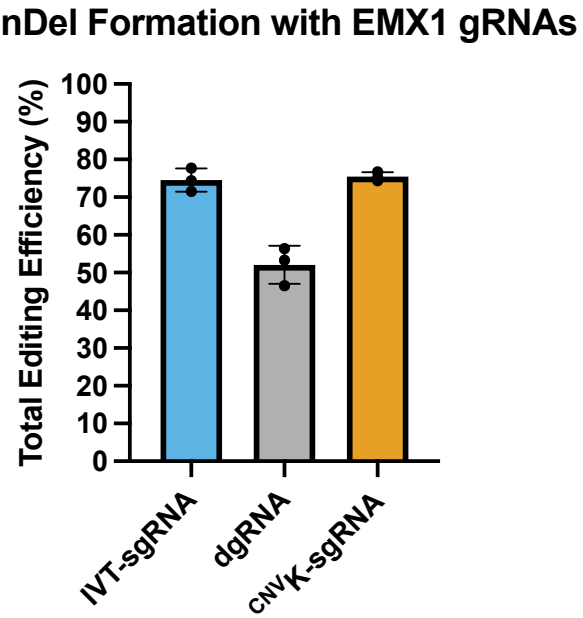


Figure 2.18: TIDE analysis of EMX1 InDel formation in HEK293 cells with Cas9-RNP. Bar represents mean total editing efficiency as determined by TIDE (n = 3, error bars = SD).

the ^{CNV}K-sgRNA had 5′ and 3′ end modifications which should increase its intracellular stability, these modifications did not translate to an improvement in relative cutting activity between *in vitro* assays and cellular experiments. This is likely due to the Cas9 protein occluding gRNAs from nucleases when transfected as an RNP complex, making the benefits of nuclease resistance modifications less pronounced compared to naked RNA transfections.

Taken together, these results demonstrate that the introduction of a ^{CNV}K crosslink improves the activity of gRNAs, to the point where there are as functional as IVT-sgRNAs. The increase in activity is likely caused by the ^{CNV}K crosslink encouraging the correct folding of the sgRNA while it is incubated with Cas9 or in cells, and the exact 1:1 ratio of crRNA/tracrRNA they provide after purification. Part of the success can be attributed to the design of our ^{CNV}K-tracrRNA, which had modifications to improve the purification of the tracrRNA, and subsequently ^{CNV}K-sgRNA. Because we transferred the ^{CNV}K residue to the tracrRNA, we now had a generalized platform for preparing ^{CNV}K-sgRNAs which we wanted to further explore.

2.6 Chemically Modified ^{CNV}K-sgRNAs

2.6.1) Chemically Modified ^{CNV}K-tracrRNA for Increased Stability

One of the main aims of this project was to create a platform for preparing highly modified RNAs which would be otherwise difficult to synthesize. As such, our next goal was to prepare a chemically modified ^{CNV}K-sgRNA to demonstrate the compatibility of our design with some of the modifications previously shown to improve Cas9 activity. As discussed earlier, several gRNA modification strategies have been implemented with varying benefits. We decided to adopt the modification pattern described by the Anderson group in 2017 [162]. Their modified gRNAs were designed by analyzing the interactions between gRNA and

Cas9, so that any 2′OH group or phosphate linkage that did not interact with Cas9 was replaced with a more stable analog. Because of this, their gRNAs are some of the most heavily modified designs in the field and would therefore be one of the most ambitious to replicate with our crosslinking approach.

We re-synthesized the CNVK-tracrRNA designed in the previous section with the modification pattern described by He *et al.* [162] where appropriate. A diagram of the 2′OMe and PS substitutions (MPS) introduced to the new CNVK-MPS-tracrRNA is shown in **Figure 2.19**. Similar to the design of our original CNVK-tracrRNA, a C12 amine modification was added at the 5′ end of the sequence to enhance purification. The 5′ C12 amine modification gives the CNVK-tracrRNA additional functionality, as it can be further modified with NHS-ester labeling chemistry to create CNVK-sgRNA conjugates. An MPS-tracrRNA without CNVK was also prepared to use in dgRNA controls.

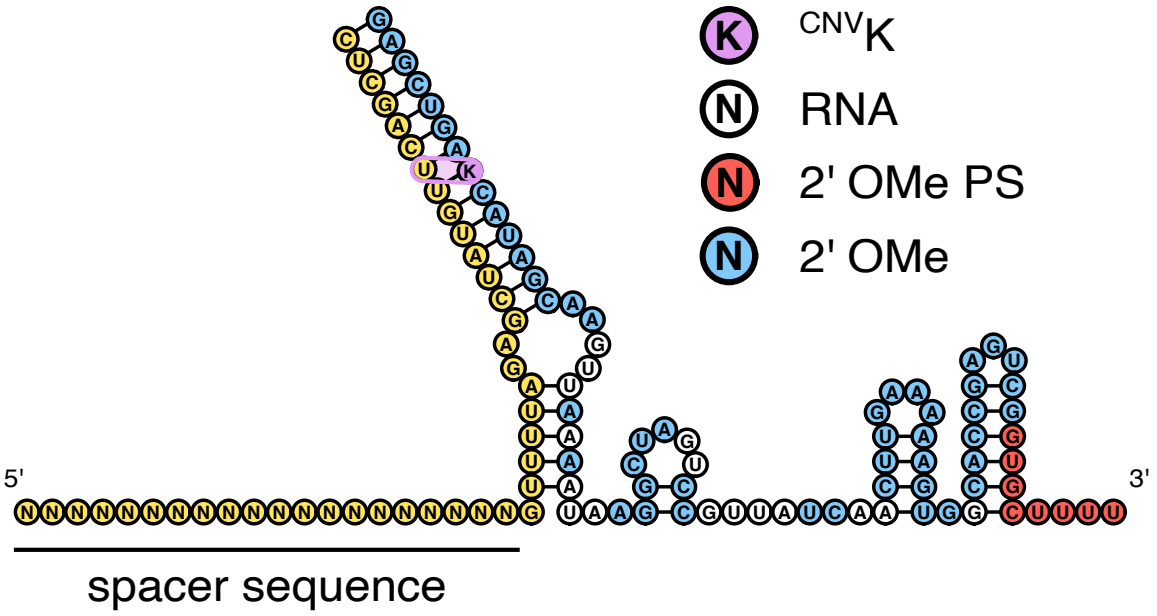


Figure 2.19: Diagram of chemical modification strategy for CNVK-MPS-tracrRNA. Nucleotides replaced with a 2′OMe modified residue are colored blue. 2′OMe residues with a phosphorothioate linkage (PS) are colored red. The CNVK residue (K) is colored purple. Unmodified RNA nucleotides are colorless. The partner crRNA strand, which is unmodified, is colored yellow. The variable spacer sequence is underlined.

To test the crosslinking activity of ^{CNV}K-MPS-tracrRNA, we performed a crosslinking reaction between a 50 µM solution of MPS-^{CNV}K-tracrRNA and a 5x molar excess of crRNA^{EMX1}. The reaction was purified by RP-HPLC to isolate the crosslink product, as shown in **Figure 2.20**. Two fractions were purified from this reaction: the excess crRNA^{EMX1} and the crosslink product between crRNA^{EMX1} and MPS-^{CNV}K-tracrRNA. The identities of these two fractions were confirmed by HPLC-MS (Appendix 1). An additional peak was eluted after the crosslinking product **(Figure 2.20)**, but we were unable to isolate a sufficient amount of this fraction to identify it by MS. The 2′OMe and PS modifications could change the strength of the interactions the ^{CNV}K-MPS-tracrRNA forms, this third fraction may correspond to a non-specific interaction formed between the crRNA/tracrRNA or tracrRNA/tracrRNA. However, it is also possible that this extra fraction could correspond to a non-denatured ^{CNV}K-MPS-sgRNA whose secondary structure is causing it to elute later during RP-HPLC. In either case, we calculated that the yield of the correctly crosslinked

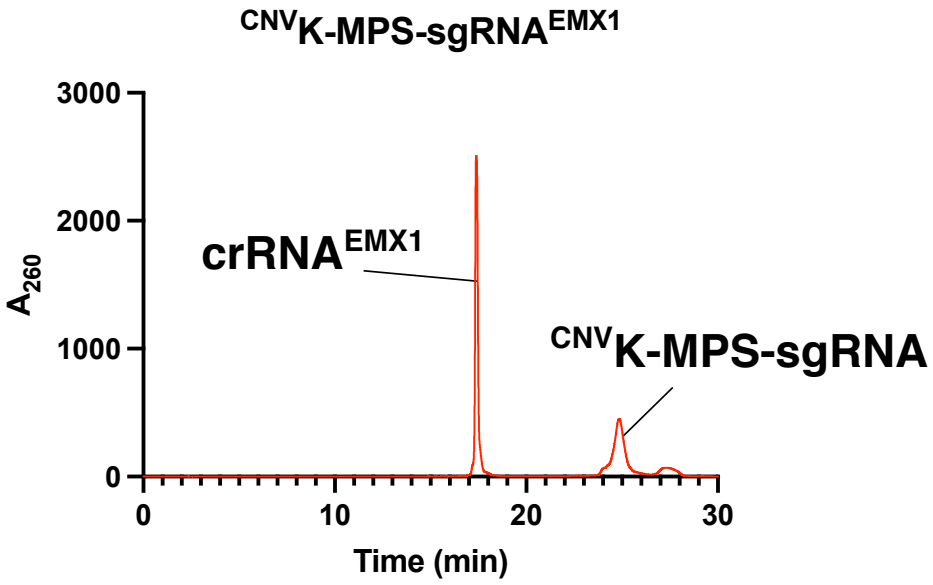


Figure 2.20: RP-HPLC purification and analysis of crRNA^{EMX1}/^{CNV}K-MPS-tracrRNA crosslinking reaction. Purification was performed with an Agilent HPLC according to Section 6.3. Two fractions were collected from this purification and analyzed by HPLC-MS, confirming their identities as excess crRNA^{EMX1} or ^{CNV}K-MPS-sgRNA^{EMX1} (labeled respectively). A third product was eluted between 27-28 min which was unable to be identified with HPLC-MS.

CNVK-MPS-sgRNA fraction was 86.5% based on the UV intensity of each peak, which was an acceptable amount to proceed with.

2.6.2) Functional Characterization of Modified CNVK-sgRNAs

After purifying CNVK-MPS-sgRNA, we next wanted to conduct functional assays to confirm that the modified CNVK-sgRNA had at least comparable activity to our non-modified CNVK-sgRNA. We therefore repeated the *in vitro* DNA digest assay of EMX1 PCR we previously described in **Figure 2.10** and **Figure 2.17**. Cas9 was incubated with either an IVT-sgRNA, dgRNA, CNVK-sgRNA, MPS-dgRNA, or CNVK-MPS-sgRNA. The resulting Cas9-RNPs were incubated with an equimolar amount of EMX1 PCR product for 1 h analysis by agarose gel **(Figure 2.21)**. In this instance, both CNVK-sgRNA and CNVK-MPS-sgRNA had digestion rates slightly lower than the IVT-sgRNA, but were otherwise comparable to each other. Additionally, both CNVK-sgRNAs also had significantly higher cleavage rates than their respective non-CNVK dgRNA controls.

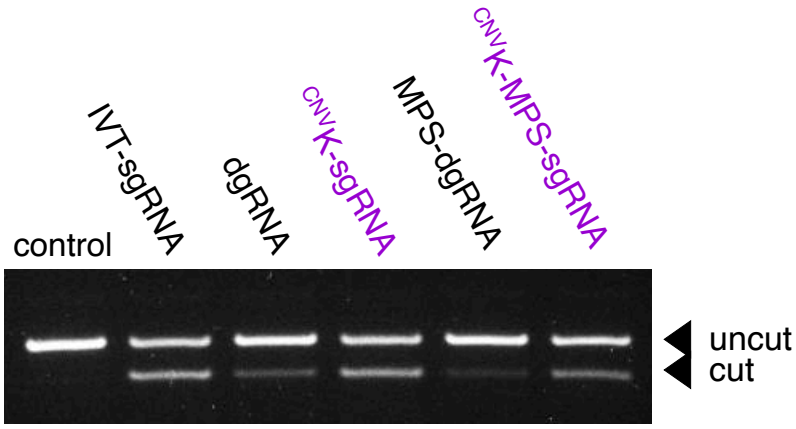


Figure 2.21: Agarose gel analysis of Cas9 *in vitro* DNA digest assay of EMX1 PCR product. From left to right, lanes correspond to undigested EMX1 PCR product, and digestion reactions with either IVT-sgRNA, dgRNA, CNVK-sgRNA, MPS-dgRNA, or CNVK-MPS-sgRNA. Arrows to the right of the gel highlight bands corresponding to 'cut' and 'uncut' DNA.

We next compared the ability of ^{CNV}K-sgRNA and ^{CNV}K-MPS-sgRNA to generate InDel mutations in cells. Cas9-RNP prepared with either ^{CNV}K-sgRNA or MPS-CNVK-sgRNA was used for an EMX1 knockout in HEK293 cells (Section 7.11.2). The EMX1 locus was then amplified with the PCR primers previously described in Section 2.5.2 for TIDE analysis. Analysis of EMX1 knockouts revealed that the two ^{CNV}K-sgRNAs generated similar amounts of InDel mutations (**Figure 2.22**). Although the 2′OMe and PS substitutions typically enhance Cas9 editing, the modifications we incorporated in our MPS-^{CNV}K-tracrRNA only confer increased stability to the resulting sgRNA, which is not as relevant when the sgRNA is delivered to cells in complex with Cas9. Therefore, we were not anticipating a higher degree of InDel generation. However, the results of this EMX1 knockout experiment, along with our *in vitro* assay, confirms that we successfully produced a functionally modified sgRNA. These results demonstrate that our ^{CNV}K-crosslinking approach to constructing biofunctional RNAs is robust enough to handle extensive chemical modification.

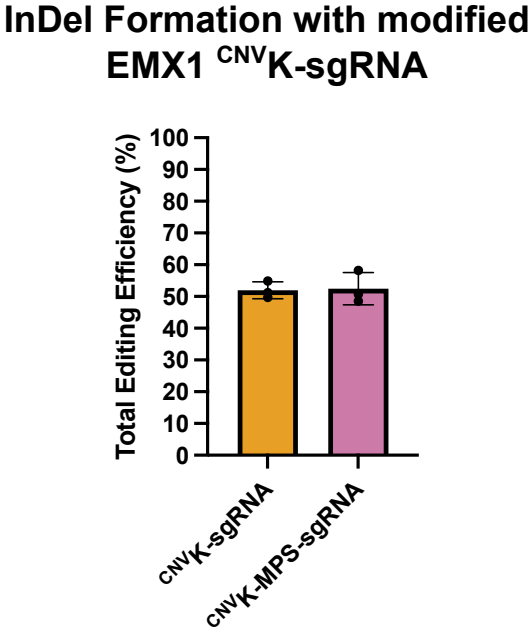


Figure 2.22: TIDE analysis of EMX1 InDel formation in HEK293 cells from Cas9-RNP. Bar represents mean total editing efficiency as determined by TIDE (n = 3, error bars = SD).

2.6.3) Preparation of Cas9 mRNA

After preparing and characterizing a ^{CNV}K-MPS-tracrRNA, the next step for this project would have been the preparation of a fully modified crRNA complement, which I unfortunately did not have enough time to complete. Based on the results of our work with this modified ^{CNV}K-MPS-tracrRNA and the partially 2′OMe crRNA described in Section 2.4.3, it is likely that full modification of the ^{CNV}K-sgRNA would be tolerated well.

One of the several benefits of working with modified sgRNAs is that they can be co-delivered *in vivo* with Cas9 mRNA and can survive nuclease degradation until the mRNA is translated to produce protein [161]. Any further gene knockout experiments with fully modified ^{CNV}K-sgRNAs would be conducted by co-delivering gRNA and Cas9 mRNA to highlight this property. To this end, we began to prepare a Cas9-mRNA while we were working on the experiments described above. A full description of the logic and workflow for preparing mRNA is described in Chapter 4 and Section 6.7 respectively. In short, we prepared a template for IVT of Cas9 mRNA by cloning a Cas9 insert into the pRSET vector as described in (Appendix 5). The Cas9 protein was linked by the T2A self-cleaving peptide [214] to EGFP so that the expression of Cas9 could be tracked by fluorescence, without increasing the size of the protein. The resulting mRNA from the IVT reaction was column purified and appropriately modified for eukaryotic translation (Section 6.7).

To confirm the mRNA was functional, we electroporated 1.5 µg of prepared Cas9-T2A-EGFP mRNA per 1 x 10^5 HEK293 cells. Cells were analyzed by flow cytometry 20 h after electroporation to determine the percent of EGFP+ cells. Representative histograms of EGFP fluorescence are shown in **Figure 2.23**, which demonstrate that approximately 75% of cells were positive for EGFP fluorescence after transfection. Downstream genes from the T2A

peptide are expressed with approximately 20% of the intensity as upstream genes [214], so Cas9 is likely more abundant in transfected cells than the fluorescence alone would indicate. We next attempted to co-transfect Cas9 mRNA as described above with varying quantities of IVT-sgRNA^{EMX1}, but were unable to observe InDel generation by TIDE any amount of sgRNA (data not shown), likely due to rapid intracellular degradation of IVT-sgRNA. This experiment will be continued after I prepare a fully chemically modified ^{CNV}K-sgRNA.

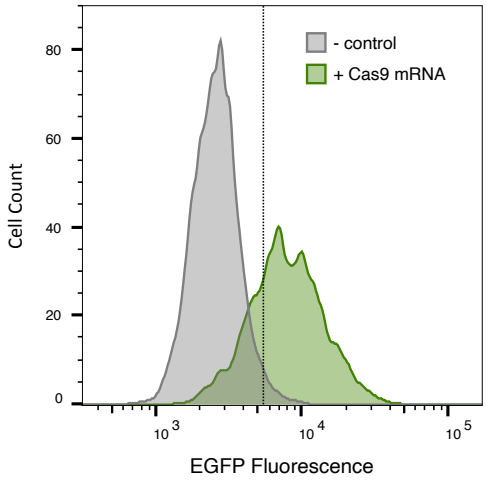


Figure 2.23: EGFP fluorescence of Cas9-T2A-mCherry mRNA transfected HEK293 cells. EGFP fluorescence was measured with 488 nm laser with 525/40 BP filter. A non-transfected control population (grey) was used to set gates for EGFP +/- cells. 74.2% of transfected cells (green) are over the threshold for EGFP fluorescence.

2.7 Preparation of ^{CNV}K-sgRNA Libraries

2.7.1) crRNA Spacer Library

As a final display of the potential of our ^{CNV}K-sgRNA platform, we wanted to prepare a sgRNA library for multiple genomic targets. This would demonstrate that the ^{CNV}K-tracrRNA can tolerate sequence variation within the crRNA without disrupting the base pairing in the repeat-antirepeat duplex required for crosslinking. To this end, we selected four commonly used sgRNA targets with therapeutic relevance: CCR5 [215], HBB [216], IL2RG [167], and VEGFA [167]. These targets have diverse spacer sequences to challenge our Watson-Crick-based crosslinking platform and have each been validated in previous

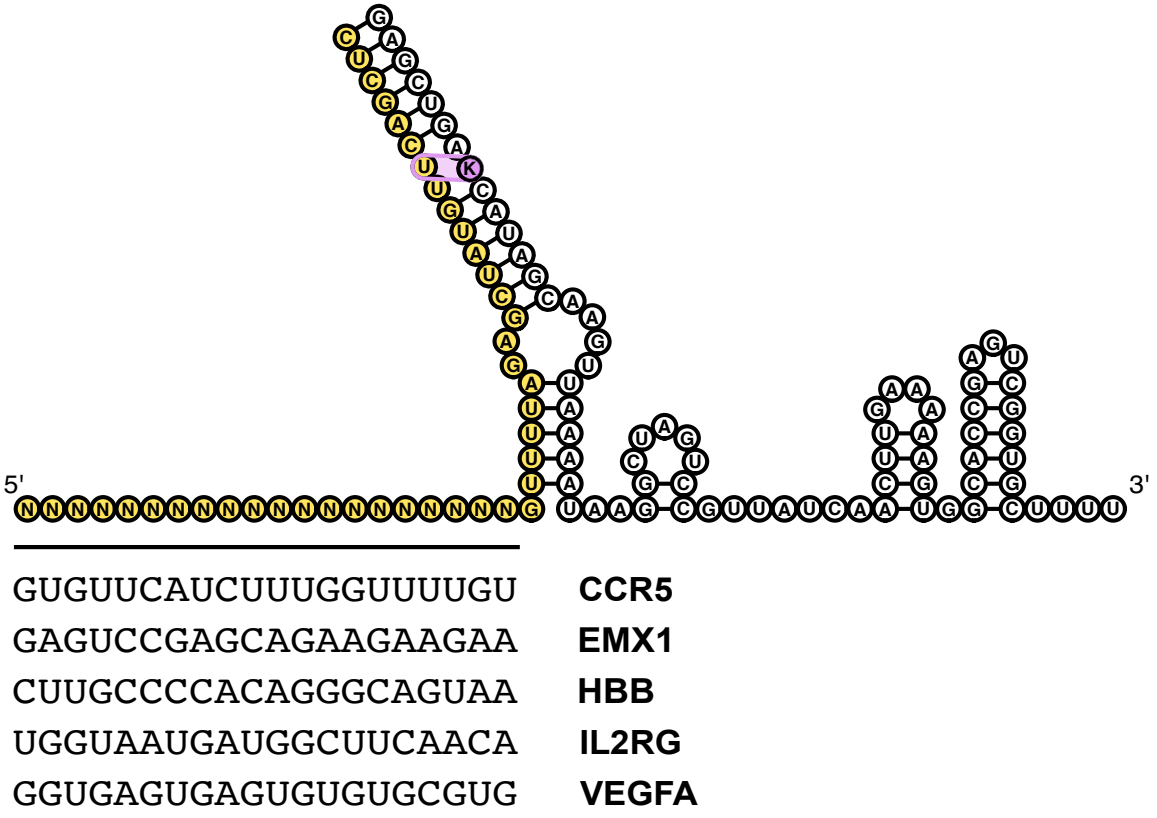


Figure 2.24: Diagram of crRNA/CNVK-tracrRNA structure with spacer sequence library. The 20 nt 'N' variable spacer sequence of the crRNA strand (yellow) is substituted with one of the 5 sequences shown below, yielding crRNA^{CCR5}, crRNA^{EMX1}, crRNA^{HBB}, crRNA^{IL2RG}, and crRNA^{VEGFA}.

studies. A new crRNA compatible with the CNVK-tracrRNA was synthesized for each target, varying only in their 20 nt spacer sequence. Including our original crRNA^{EMX}1, this provided us with 5 potential CNVK-sgRNAs with which to compare crosslinking efficiency and sgRNA activity. Each spacer sequence is shown in **Figure 2.24** next to the conserved CNVK-sgRNA structure.

Crosslinking reactions were prepared with 50 μM CNVK-tracrRNA and a 5x molar excess of each crRNA. Reactions were subsequently analyzed and purified by RP-HPLC. An overlaid graph showing the UV absorbance and retention time during HPLC of each crosslinking reaction is shown in **Figure 2.25a**. As is consistent with our previous results, we observed two peaks in each reaction: one which corresponded to unreacted crRNA, and the other to

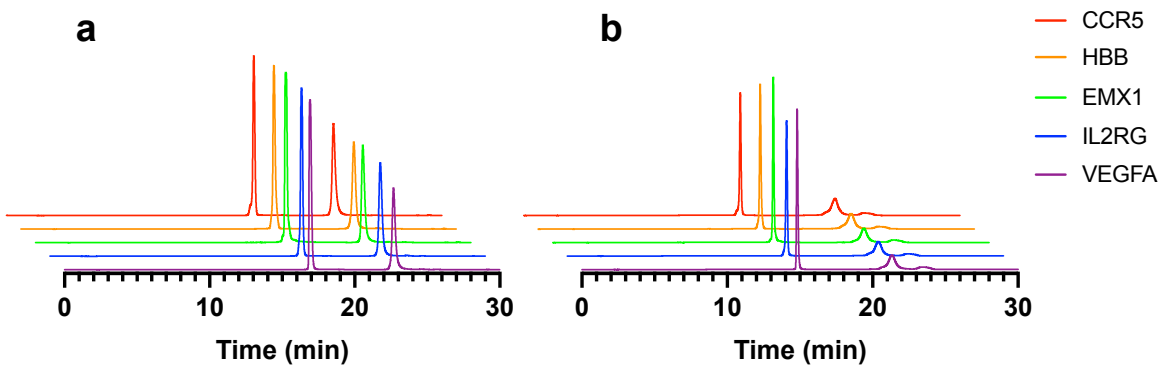


Figure 2.25: RP-HPLC purification and analysis of crosslinking reactions between crRNA library and **(a)** CNVK-tracrRNA or **(b)** CNVK-MPS-tracrRNA. Y-axis corresponds to A_{260}. Absorbance plots are colored according to the crRNA used in crosslinking reaction and are staggered for increased clarity.

CNVK-sgRNA. We observed the total consumption of CNVK-tracrRNA and no side products in each reaction regardless of spacer sequence, which suggests that the crosslinking reaction is not sensitive to sequence variety at these positions. We repeated each reaction using MPS-CNVK-tracrRNA to determine if the new spacer sequences would also be compatible with the heavily modified tracrRNA. In this instance, a 10x molar excess of crRNA was used as we previously observed a possible side product in our reactions with MPS-CNVK-tracrRNA and $crRNA^{EMX1}$ (**Figure 2.20**). The HPLC analysis of each reaction is shown in **Figure 2.25b**. As with the unmodified CNVK-sgRNA library, we did not observe a difference in crRNA/MPS-CNVK-tracrRNA crosslinking reactions between the different crRNA species. Despite increasing the relative concentration of crRNA in this reaction compared to the one in Figure 6.2, we did not observe a change in the yield of the reaction, which further leads us to believe that the product eluting after the CNVK-MPS-sgRNA is due to secondary structure of the sgRNA. Note that the peaks corresponding to crosslink product appear smaller in **Figure 2.25b** compared to **Figure 2.25a** as a greater excess of crRNA is present in the reaction. To conserve MPS-CNVK-tracrRNA for future experiments, these reactions were

performed at an analytical scale, and all further work in this section was done with the unmodified CNVK-tracrRNA.

After purification by HPLC, CNVK-sgRNAs were also analyzed by denaturing PAGE **(Figure 2.26)**. Each isolated crosslink product produced one major band, again confirming that the crRNA/tracrRNA interactions are not disrupted by sequence variation at the spacer sequence. However, we also observed that a small fraction of crRNA and CNVK-tracrRNA is visible by gel in each isolated product. Based on the conditions of the RP-HPLC purification and previous analyses, we do not believe that non-crosslinked product was co-purified with CNVK-sgRNAs. Therefore, we think it is possible that this non-crosslinked material is being generated either by the baseline normalization wavelength used during HPLC (which is close to the CNVK reverse crosslinking wavelength) or decomposition during long-term storage.

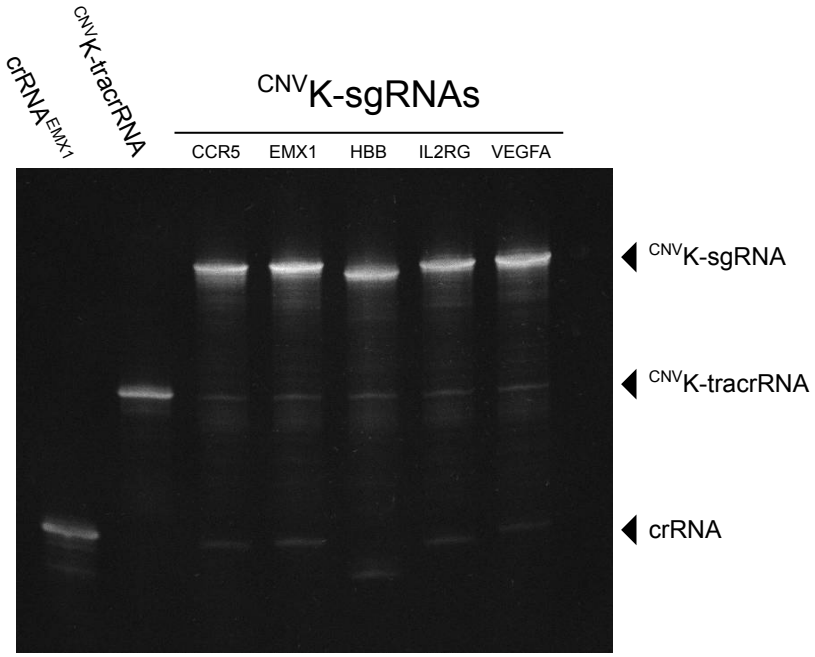


Figure 2.26: Denaturing PAGE analysis of RP-HPLC purified CNVK-sgRNA library prepared in Figure 7.2a. The purified CNVK-sgRNAs are labeled by the spacer sequence of their constituent crRNA. crRNAEMX1 and CNVK-tracrRNA are run on the left of the gel and labeled accordingly for reference. Arrows to the right of the gel highlight the bands corresponding to crRNA, CNVK-tracrRNA, and CNVK-sgRNA.

Further experiments are needed to determine if either of these possibilities are correct, and will ultimately contribute to improving current knowledge on ^{CNV}K handling.

2.7.2) *In vitro* Characterization of ^{CNV}K-sgRNA Library

To compare the activity of the 5 ^{CNV}K-sgRNAs in our minimal library, we prepared an IVT-sgRNA template for each spacer sequence with the common scaffold also used for IVT-sgRNA^{EMX1}. We also synthesized an unmodified, non-^{CNV}K tracrRNA to prepare control dgRNAs. Lastly, a set of TIDE-compatible primers was designed for each target sequence to amplify approximately 700 bp around the target locus (Appendix 4). These primers were also used to prepare PCR amplicons for *in vitro* digestion assays. For each gene target, Cas9 protein was incubated with an equimolar amount of IVT-sgRNA, dgRNA, or ^{CNV}K-sgRNA, which were then used to digest respective amplicons as described earlier. The agarose gel analysis of each digestion reaction is shown in **Figure 2.27**.

The digestion reactions revealed that each ^{CNV}K-sgRNA is able to initiate Cas9 DNA cleavage, demonstrating that our ^{CNV}K crosslinking platform can be used to produce a functional library. In the case of three targets (HBB, EMX1, VEGFA) the rate of ^{CNV}K-sgRNA DNA cleavage observed was comparable to the activity of the IVT-sgRNA. For the other two targets (CCR5, IL2RG), ^{CNV}K-sgRNAs were functional but had cleavage rates comparable to the dgRNA.

The overall DNA cleavage rates varied between targets, with some targets cleaved at higher rates than others. The limited cleavage is not unexpected, given that we designed this *in vitro* digest assay to avoid total digestion of the DNA substrate (Section 2.4.4). Out of the 5 targets, the CCR5 amplicon was cleaved to the greatest extent. This could be explained by

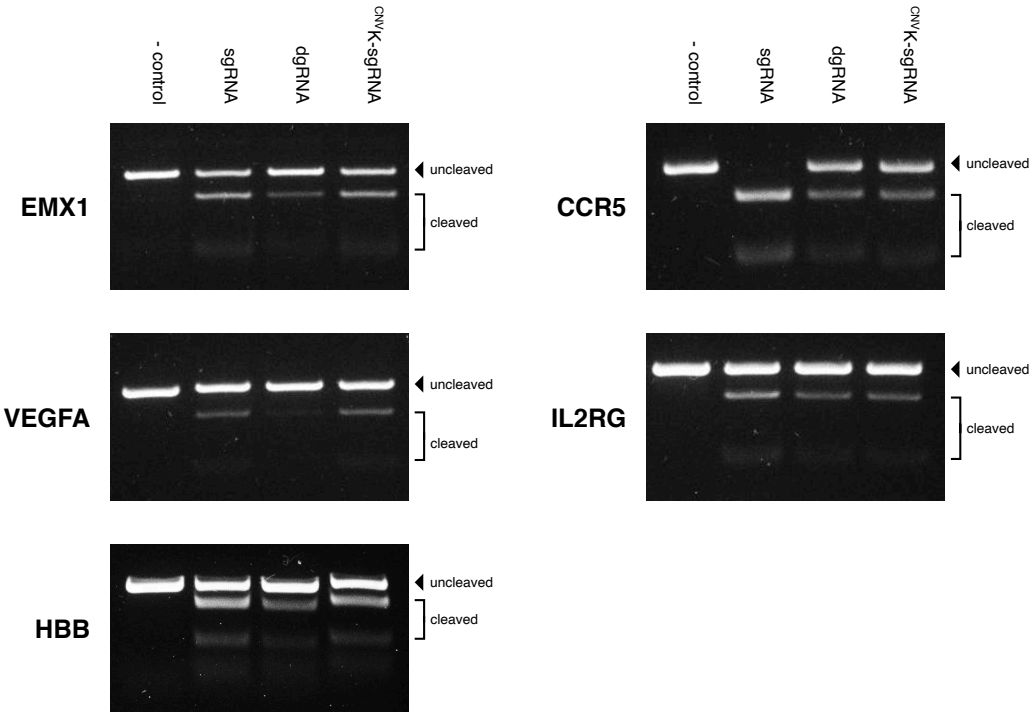


Figure 2.27: Agarose gel analyses of Cas9 *in vitro* DNA digest assays using gRNA libraries. Gels are separated by target/spacer sequence (CCR5, EMX1, HBB, IL2RG, VEGFA). In each gel, lanes are organized from left to right with undigested PCR product, and digestion reaction with IVT-sgRNA, dgRNA, or CNVK-sgRNA. Arrow to the right of the gel highlights undigested product, and bracket identifies the two cleavage product bands.

the low T_M of its target sequence (50.6 °C, nn model), making it more amenable to local melting by Cas9. The duplex which was cleaved the least, VEGFA, had one of the highest local T_M (59.66 °C , nn model) in contrast.

Although this experiment demonstrated effective library preparation, we were disappointed to find that CNVK-sgRNA activity varied with respect to IVT-sgRNA activity depending on the spacer sequence. We attempted to troubleshoot the digests with CCR5 and VEGFA gRNAs by increasing the length of incubation, applying excess gRNA to ensure Cas9 was saturated with RNA, and resynthesizing library members, but none of these conditions significantly impacted the CNVK-sgRNA activity rate (data not shown). We noticed that in some click-sgRNA strategies [205], the relative activity of triazole-sgRNAs also fluctuates with respect

to dgRNA activity between different spacer sequences. Therefore, it could be possible that the degree that chemical conjugation improves dgRNA activity has some natural variability based on spacer sequence. We also considered that this inconsistency could be caused by the trace amounts of non-crosslinked material in each ^{CNV}K-sgRNA we observed in **Figure 2.26**

Regardless of the reason, this experiment demonstrates that our ^{CNV}K platform can be used to prepare functional libraries of synthetic sgRNAs that at a minimum match the activity of dgRNAs. Additionally, during the *in vitro* assays we performed, gRNAs are highly concentrated and are handled in conditions that favor RNA hybridization during Cas9-RNP formation. It is more than likely that ^{CNV}K-sgRNAs would outperform dgRNAs in cellular conditions where the Cas9 and gRNA are relatively dilute, such as co-delivery of gRNA and Cas9-mRNA. Any continuation of this research would involve cell-based assays to develop a more accurate picture of ^{CNV}K-sgRNA activity.

2.8 Conclusion and Future Works

Over the course of this project, we were able to enhance knowledge of ^{CNV}K crosslinking properties, incorporate a ^{CNV}K residue into an RNA to replicate a natural stem-loop structure, produce functional ^{CNV}K-sgRNAs, and use these as a platform for library preparation. ^{CNV}K-sgRNAs can be prepared in as little as 30 min, requiring only a short annealing program and a pulse of 365 nm light. *In vitro* and cell-based assays demonstrated that ^{CNV}K-sgRNAs have comparable activity to sgRNAs produced through enzymatic methods. Furthermore, our ^{CNV}K platform for sgRNA construction makes the incorporation of synthetic modifications more permissible, which we used to prepare an extensively MPS-modified ^{CNV}K-tracrRNA. Further work is needed to prepare a modified partner crRNA, which could be used to explore the effects of chemical modification on the spacer sequence.

The current benchmark in the field of synthetic RNA ligations is click-chemistry, which has also been used to prepare sgRNAs [203, 205, 206]. ^{CNV}K-based construction has some advantages however, in that it requires only one modification and does not require cofactors or organic solvents. Because ^{CNV}K crosslinking is controlled by Watson-Crick interactions, it should be possible to construct RNAs with multiple crosslinks more easily than it could be prepared with multiple click ligations. To this end, our next goal for this project is to prepare prime editor guide RNAs (pegRNAs) with ^{CNV}K crosslinking. The versatility of prime editors (PEs) has arguably made this technology as disruptive as Cas9 gene editing originally was; PEs have already been applied for the study and treatment of disease, functional screens of genetic variants, model animal generation, gene tagging, and crop engineering [217].

The preparation of pegRNAs could be significantly improved with ^{CNV}K crosslinking. PegRNAs have two variable domains at both 5′ and 3′ termini; by preparing a scaffold that encompasses the conserved region, pegRNA libraries could easily be prepared by exchanging their spacer sequence or RTase template sequence. Applying libraries of pegRNAs for editing has the potential to generate a huge amount of specific genetic diversity at a single locus [218, 219]. In addition to simplified library preparation, this would also open up pegRNAs to chemical modification, which has yet to be widely explored but could be hugely impactful for this technology. Modification of the pegRNA 3′ extension in particular could help address several key issues in prime editing, such as the degradation of the RTase template sequence.

Synthetic pegRNAs can be constructed with either ^{CNV}K crosslinking or a mix of ^{CNV}K crosslinking and click-chemistry. We have begun some work on these options but decided against including these relatively new and sparse results in this report. Taken together, our

work on ^{CNV}K-sgRNAs and ^{CNV}K-pegRNAs should demonstrate that crosslinking is a viable strategy for constructing biofunctional RNAs, and that it can supplement or even replace click-chemistry in some areas of RNA handling.

3

Hybrid gRNAs

3.1 Background

3.1.1) Eukaryotic Transcription is Regulated from a Distance

Eukaryotic gene expression is the product of RNAPol II transcriptional activity, which can initiate the transcription of pre-mRNA and some non-coding RNAs at specialized core promoters. A key step in transcription initiation is the assembly of the pre-initiation complex at core promoters, a complex of six general transcription factors (GTFs) which recruit and activate RNAPol II [220]. Recruitment of the GTFs to the core promoter is in turn controlled by a confluence of *cis* and *trans*-acting regulatory elements, reviewed in [221, 222]. A principal component of transcription regulation is distal enhancer sequences, which pool around gene promoters to recruit protein transcription factors (TFs), among other effects (reviewed in [223]).

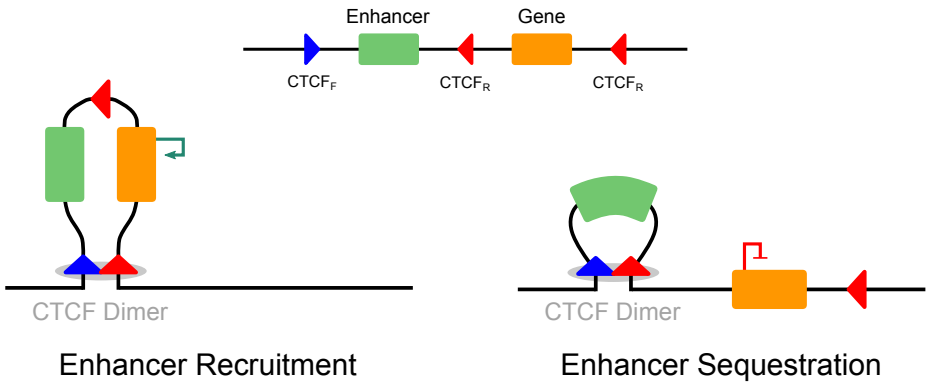


Figure 3.1: Simplified diagram of the impact of chromatin looping on gene regulation. The primary sequence of a hypothetical stretch of chromatin is displayed: An enhancer sequence (green) and gene of interest (vermillion) are interspersed with the CTCF recognition sequence in either forward (blue) or reverse (red) orientation. Alternative chromatin loops stabilized by a CTCF dimer can either recruit or sequester the enhancer sequence from the gene of interest, which will up and down regulate gene expression respectively.

Interactions between the core promoter and enhancer sequences, which can even be located on entirely different chromosomes, are mediated by extensive chromatin looping. Chromatin loops colocalize regulatory elements by bringing genomic regions which are distant in primary structure close together within 3D space [224]. The distribution of chromatin loops across the genome is not random but is rather another highly dynamic [225] and key element of gene regulation [226]. Looping is largely controlled by cohesin, a ring-like protein structure through which chromatin loops are extruded, and the architectural protein CTCF that binds defined recognition sequences and anchors chromatin looping [227].

The primary unit of interphase chromatin folding is topologically associated domains (TADs): chromatin loops which are defined by the preferential interaction of genetic elements *within* the loop [228, 229]. Alternate TAD groupings can sequester or bring enhancer sequences into contact with gene promoters **(Figure 3.1)**, leading to differential

gene expression outcomes [226]. The regulatory importance of TADs is highlighted by genetic diseases which are caused by perturbations of TAD boundaries [230].

3.1.2 dCas9 Transcriptional State Reporters

The discovery of Cas9 gene editing has yielded several tools which can cleave DNA targets and introduce new genetic material ranging from 1 nt to kilobases of sequence insertion (Section 1.8). The crux of Cas9's utility is not only its nuclease functions, but also its unprecedented and programmable ability to identify genomic loci. Accordingly, several impactful tools have been developed from catalytically dead dCas9, which do not have nuclease activity but retain genome interrogating functions. One example of this are the fluorescent dCas9 tools that have been widely applied for live-cell genomic imaging, as discussed in Section 2.1.3. To date, a full suite of Cas9/effector fusion proteins has also been developed to mimic or influence transcriptional effectors. These include programmable transcription factors (dCas9-VP64), transcription repressors (dCas9-KRAB), CpG methylation (dCas9-DNMT3a) and demethylation enzymes (dCas9-TET1), as well as multiple histone-modifying enzymes [231-235].

Despite the diverse ways Cas9 can manipulate genetic material, it has not been widely applied to study the behavior of genes in live cells, apart from visualizing their sub-cellular location [234]. In other words, although dCas9 tools can affect the transcriptional environment around a gene, they cannot provide direct information on the status of that environment nor the expression of the gene in question. In the past decade, there have been major advances in the study of the transcriptome owing to improvements in techniques such as single-cell RNA seq [236, 237]. However, nearly all techniques used to study the dynamic

expression of a gene in *live cells* still rely on reporter knock-ins, which are work-intensive and difficult to scale and multiplex.

Therefore, the central question of this chapter was whether dCas9 could be applied to report the transcriptional activity and environment at a genomic locus. We reasoned these two factors would be critical to this goal: This first is the relatively long residency time of dCas9 on DNA substrates [238], making it possible to use dCas9 as a platform to pin conjugated molecules to a genomic locus. The second is the role of chromatin looping in eukaryotic gene activation, as well as the *trans*-acting nature of eukaryotic protein transcription factors. We hypothesize that if a short, minimally-expressed "DNA cargo" is pinned to a genomic locus of interest, the specific transcriptional environment around that locus could either enhance or repress the cargo. Provided that the DNA cargo is only transcribed in highly permissive environments, cargo expression could be read out as a measure of transcription at that specific locus. By conjugating this DNA cargo to dCas9-RNP, it could be programmed to interrogate different loci in a highly multiplexable fashion **(Figure 3.2)**.

It is important to note that due to the multiple overlapping mechanisms that regulate the processing and translation of eukaryotic mRNA (reviewed in [239]), expression of the DNA cargo would not directly correspond to gene expression of the target locus. However, dCas9-DNA cargo conjugates could provide information about the transcriptome, in this regard acting as a transcriptional state reporter (TSR).

We envisage at least two potential uses for dCas9-TSRs: the first is as a research tool to quickly study the expression and environment around a gene of interest. In this regard, dCas9-TSRs would encode a fluorescent reporter peptide to provide a readable signal of gene

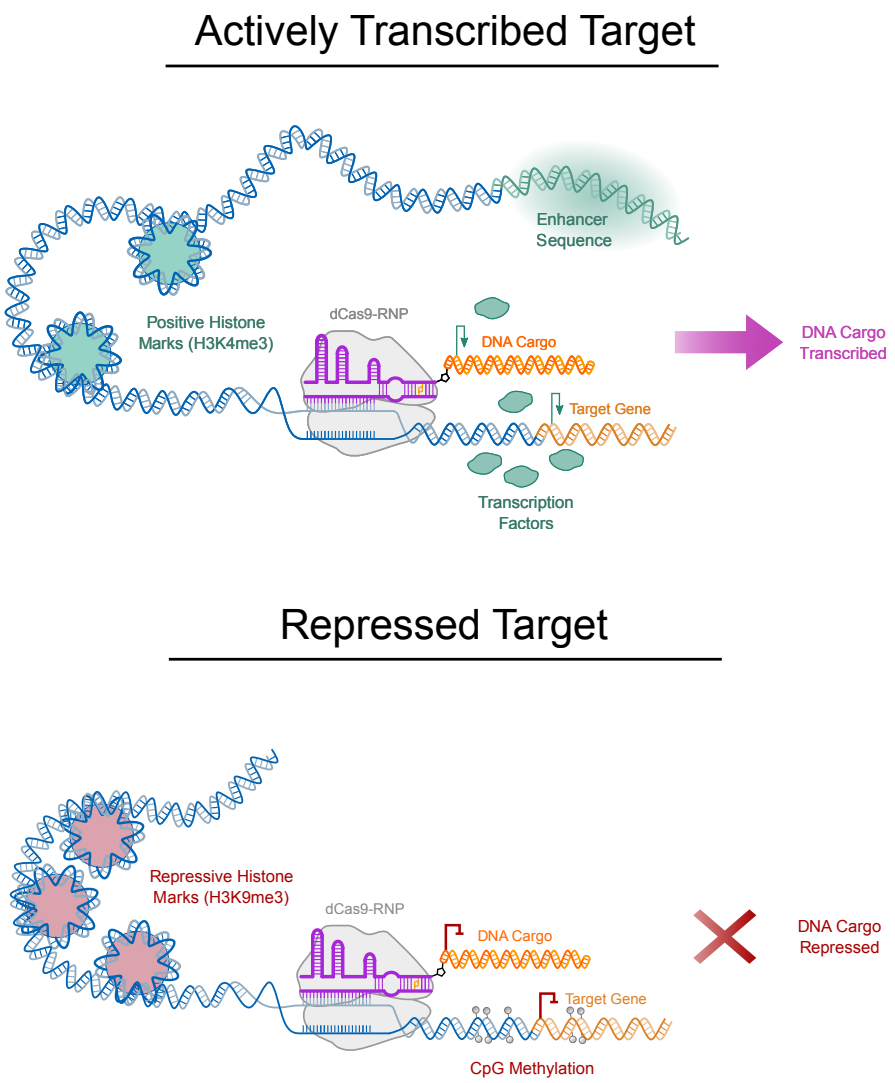


Figure 3.2: Diagram of dCas9/DNA cargo transcription at differentially expressed targets. When targeted to an actively expressed gene, enhancer sequences, active histone marks, and other positive regulatory factors recruit TFs to the target. The presence of these elements and TFs promotes the transcription of the dCas9 DNA cargo. When targeted to a repressed target, the repressive environment around the target prevents the DNA cargo from being transcribed.

expression; TSRs would not require the generation of genetically modified cell lines and could be easily reprogrammed by exchanging spacer sequences, making the much easier to implement than traditional reporters. Secondly, we believe that dCas9 TSRs could have potential therapeutic applications as they could be used to distinguish the transcriptome profiles of healthy and diseased cells. This is especially relevant where the expression of a set of genes demarcates disease phenotypes, such as oncogenes in cancerous cells. The short

cytotoxic peptide soricidin [31] can be encoded in oncogene-targeting TSRs to be expressed in and selectively kill cancer cells.

3.1.3 Cas9-RNP/DNA Conjugation Methods

Establishing a reliable method for the conjugation of DNA cargos to Cas9-RNP complexes will be critical to preparing TSRs. Several conjugation approaches have previously been reported, and are typically used to improve HDR rates in gene editing by physically linking DNA repair templates to Cas9-RNP. One of these methods was previously described in Section 2.1.3 in the context of chemically modified gRNAs: Lee *et al.* [188] demonstrated it was possible to click-ligate a 5′ DBCO-crRNA to an 86 nt 3′ azido-DNA repair template, and then form a functional gRNA after hybridizing the new crRNA/DNA conjugate with a tracrRNA **(Figure 3.3a)**. By conjugating the gRNA and donor DNA strand, their system ensured an equal delivery of gRNA and repair template and localized the repair template to the site of Cas9 cleavage. As previously discussed, this conjugation strategy resulted in a three-fold improvement in HDR rates in HEK293 compared to non-conjugated controls [188]. Unfortunately, the modifications and location of click-ligation used in this strategy precluded synthesizing this gRNA as a sgRNA or a triazole-sgRNA respectively, forcing the authors to employ the less effective dgRNA system.

The ability to conjugate full-length sgRNAs and donor DNAs would combine high DSB forming activity with an improved HDR rate, an objective which has traditionally been difficult to achieve with chemical modification strategies. Carlson-Stevermer *et al.* (2017) were able to navigate this challenge by incorporating the S1m aptamer into the anti-direct stem-loop of the sgRNA [209]. The S1m aptamer interacts strongly with streptavidin [240], which allowed them to link sgRNAs to biotinylated DNAs entirely through non-covalent

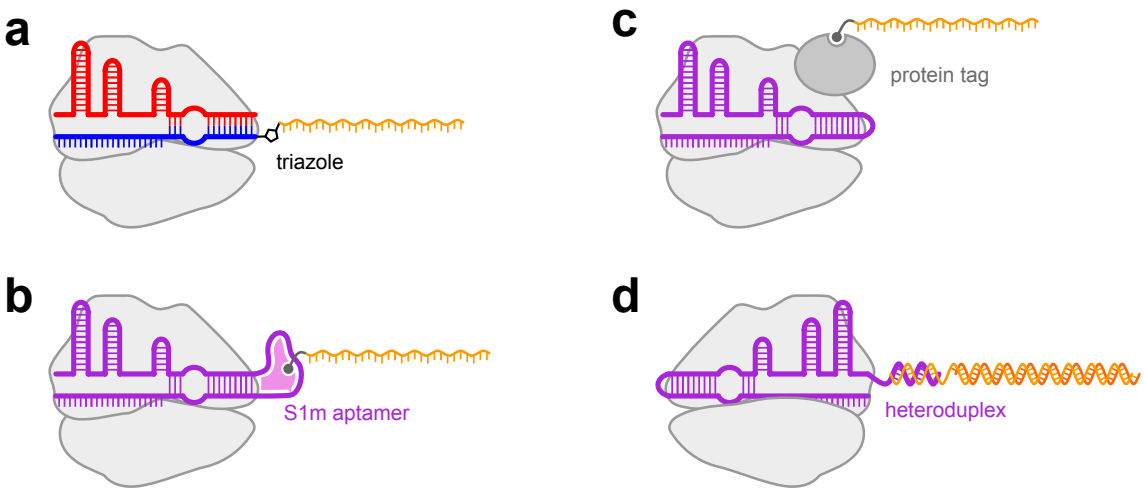


Figure 3.3: Alternative strategies for conjugating a DNA cargo (orange) to either Cas9 (grey) or the Cas9 dgRNA or sgRNA (red/blue and purple, respectively). Conjugation is achieved either by **(a)** chemical ligation of DNA cargo to crRNA, **(b)** recruitment of biotinylated DNA through streptavidin-binding S1M aptamer, **(c)** a protein tag which is capable of binding to modified or unmodified DNA, and **(d)** a region of heteroduplex between the DNA cargo and the 3′ end of an sgRNA.

interactions with a streptavidin intermediate **(Figure 3.3b)**. Carlson-Stevermer *et al*. reported between a 3-10 fold increase in HDR events at various loci in hPSCs using their system, termed "S1mplex", highlighting the potential that sgRNA conjugate systems have. However, the S1mplex design increases the length of the sgRNA by approximately 90 nt, removing the possibility of chemical modification to further improve gRNA function. Additionally, the S1m aptamer protrudes greatly from the Cas9 nuclease, making it more susceptible to nuclease degradation. This weakness, partnered with the non-covalent interaction between sgRNA and DNA makes it unclear if S1mplex would be as effective for HDR gene editing *in vivo* as it is in single cell culture.

In place of direct gRNA/DNA conjugation, it is also possible to conjugate donor DNAs to the Cas9 protein through covalent and non-covalent interactions. These methods generally take advantage of established techniques for post-translation labeling of proteins via recombinant protein tags **(Figure 3.3c)**. One of the first instances of this approach was the production of

Cas9 with a C-terminal avidin tag which could bind biotin-labeled donor DNA strands [241]. The interaction recruits donor DNA to the site of Cas9 cleavage, improving HDR outcomes in both HEK293T cells and mouse zygotes. Savic *et al*. [242] describe a similar approach using SNAP-tag technology, which allows covalent binding of O^6-benzylguanine (BG)-labeled molecules to SNAP-tag domains [243]. By preparing recombinant Cas9 with a C-terminal Snap-tag, Savic *et al* could covalently link a BG-labeled DNA donor strand to Cas9 by incubating the two components for 60 min at 30 °C. When applied in HEK293 cells, this system was able to achieve between a 7-20 fold increase in gene correction of several endogenous targets with no observed increase in off-target mutagenesis.

The two methods above can be applied to rapidly prepare Cas9-DNA conjugates, but require a modified DNA substrate for binding by fusion domains. To remove this requisite, the Gordon group has engineered HUH endonuclease protein tags that can form robust covalent bonds with ssDNA strands, requiring only a short 13 nt recognition sequence [244]. Formation of a phosphotyrosine bond between ssDNA and HUH endonucleases can occur within minutes at room temperature. By adding a C-terminal HUH tag to Cas9, they were also able to produce Cas9-DNA conjugates with improved HDR rates in HEK293 and U2OS culture without the need for chemical modification of the ssDNA [245].

Each of these methods was able to improve rates of HDR editing in cells without modifying the Cas9 gRNA, eliminating the challenges associated with modifying full-length sgRNAs. For their convenience, however, Cas9-DNA conjugation strategies have some common flaws. The additional handling steps required for conjugation increase the risk of partially denaturing Cas9 and reducing its activity, which is not as great a concern when handling RNA at RT in comparison. Because all these strategies require Cas9-DNA conjugation

before transfection, each one requires transfection of Cas9-RNP and precludes the use of alternative Cas9 deliver strategies, such as the highly effective mRNA co-delivery.

Each of the conjugation methods described above was designed to improve HDR rates of Cas9 editing. In contrast, Xu *et al*. attempted to produce dCas9 transcriptional activators by attaching engineered transcriptional enhancers [246] to the Cas9 gRNA [247], which they hypothesized would promote transcription at target loci. To accomplish this, they extended the 3′ end of the sgRNA to create a "capture sequence" which could anneal to a single-stranded overhang in their DNA cargo **(Figure 3.3d)**. This method is problematic, as the extensive RNA/DNA duplex it forms should be degraded instantly in the nucleus by RNAse H. However, they did report that this system, when paired with an existing CRISPRa fusion protein, could upregulate the expression of several endogenous human genes, including pluripotency factors. This method has not been published anywhere else, so it remains to be validated whether DNA/RNA hybrids connected through a stretch of heteroduplex make a viable activation system. Regardless, it does highlight that gRNA/DNA hybrids can be used for more applications than HDR gene editing alone.

3.1.4) Hybrid Single Guide RNA TSRs

In the previous chapter, the ^{CNV}K nucleoside was explored as a platform to construct sgRNAs. Our ^{CNV}K approach conferred several benefits to synthesizing complex RNAs, including a simplified construction of highly modified RNAs and complex RNA libraries. An additional benefit of our crosslinking approach is that it freed up the termini of our ^{CNV}K-sgRNA for further functionalization. This was considered in the design of our ^{CNV}K-tracrRNA, which included a 3′ terminal amine to provide a site for modification with azide or alkyne residues, among other reasons. Click-chemistry and ^{CNV}K crosslinking can

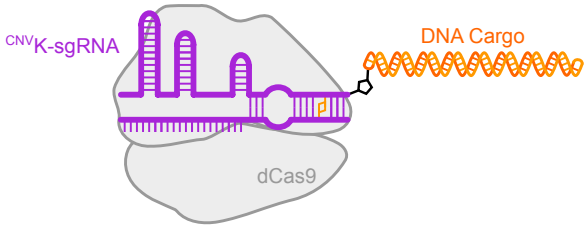


Figure 3.4: Proposed hsgRNA based on the work of Chapter 3. The CNVK-sgRNA (purple) would be covalently attached to a dsDNA cargo (orange) via a triazole linkage. Catalytically inactive dCas9 (grey) directs the dsDNA cargo to specified target sequence without causing DNA cleavage.

complement one another to construct complex oligonucleotides which would be difficult to produce with either method alone. We were interested in demonstrating this strategy as a new approach to constructing DNA/gRNA molecules, dubbed hybrid-gRNAs (hgRNAs)**(Figure 3.4)**.

Unlike the DNA/gRNA conjugation strategies described by Carlson-Stevermer *et al.* [209] and Xu *et al.* [247], our hgRNAs would be linked only through covalent interactions, improving their stability *in vivo*. Likewise, our method would retain the flexibility of the dgRNA/DNA click-ligation strategy described by Lee *et al.* [188] but would benefit from the higher activity of sgRNAs which would be generated by CNVK crosslinking. Where possible, we also aim to produce DNA cargos with no exposed termini to reduce their susceptibility to nuclease degradation, which no prior group has attempted while constructing DNA/Cas9-RNP conjugates. Ultimately, ligation of the DNA cargo to the tracrRNA portion of the CNVK-sgRNA will allow libraries of hgRNAs to be easily prepared by swapping out the crRNA component, facilitating TSR multiplexing.

3.2 Objectives

Cas9-derived technologies can serve as highly precise tools to manipulate genes, chromatin, and transcription dynamics. By physically linking a DNA cargo to Cas9 gRNAs, the DNA can be directed to genomic loci by Cas9 with a high degree of accuracy. These hybrid gRNAs (hgRNAs) have been applied to improve rates of HDR editing and to affect target gene expression. The goal of this chapter was to prepare a new class of hgRNAs, transcriptional state reporters (TSRs), which can provide a readable signal based on the transcription dynamics occurring at loci of interest. TSRs could potentially be applied to study gene expression dynamics in live cells without requiring reporter knock-ins, making TSRs more scalable and less work-intensive than comparable methods. Our first objective was to design a suitable DNA cargo to use in a TSR system. After this, we began to explore methods of ligating long dsDNA molecules to various single-stranded gRNA components which would be compatible with our ^{CNV}K-sgRNA platform. After exploring various constructions, we attempted to implement our most successful approach in cells to determine if hgRNAs could be used as TSRs.

Our intention for this project was to use our ^{CNV}K-sgRNA platform to simplify the construction of hgRNAs. However, most of the work done in this chapter took place before or during the work of the ^{CNV}K-sgRNA chapter. The hgRNAs presented in this chapter were therefore prepared without the amine-functionalized ^{CNV}K-tracrRNA described in the previous chapter.

3.3 Transcriptional State Reporter System

3.3.1) GFP1-10/11 Complementation Assay

To determine whether it is possible to read the transcriptional state of endogenous genes with hgRNAs, we first needed to design a suitable dsDNA "cargo" to use in cells. We wanted to minimize the length of the cargo so that it would not impede dCas9 diffusion within the nucleus, and to reduce the toxicity of the DNA upon transfection [248]. In this respect, the previously mentioned soricidin would have been ideal for hgRNA cargo owing to its short size. However, soricidin would be a poor reporter for initial experiments as its expression would be difficult to distinguish from stress-induced cell death. A conventional protein reporter such as GFP would provide an easier-to-interpret fluorescent readout, but its large size (238 a.a.) makes it an encumbrance to encode in a DNA cargo. Because of this, we decided to pursue a split-GFP strategy, in which separate fragments of GFP are expressed in a cell and only become fluorescent upon self-assembly. Using this approach, a smaller fragment of GFP could be encoded into a hgRNA DNA cargo, and the remainder of the protein could be supplied by plasmid. The GFP1-10/11 pair, engineered by splitting superfolder GFP between its 10^{th} and 11^{th} β-strands [249], is an ideal candidate as the pair was specifically designed for untemplated self-assembly in cells. Additionally, GFP11 (17 a.a.) remains among the smallest split fluorescent protein fragment described to date [250], making it suitable to code on a hgRNA.

To complete the GFP1-10/11 assay, a stable source of GFP1-10 expression is also required. To accomplish this, the GFP1-10 sequence was cloned into a plasmid which could be transfected into cells in advance of the hsgRNA-Cas9 RNP. The Epstein-Barr virus-derived EBNA-1 episomal vector was selected for the plasmid backbone as it is reported to provide reliable expression (~70% of cells) of protein up to a week after initial transfection [251].

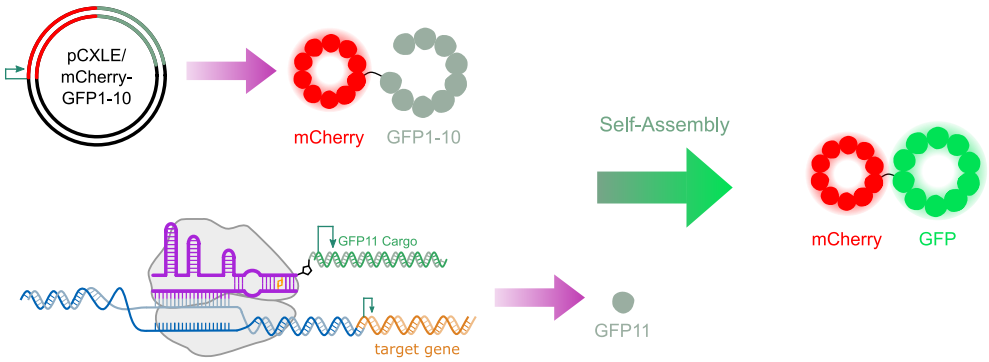


Figure 3.5: GFP1-10/11 complementation assay. A plasmid encoding an mCherry/GFP peptides 1-10 fusion is transfected into cells. dCas9 brings an hsgRNA into the proximity of a gene of interest. The positive transcriptive environment around the target gene allows the expression of the hsgRNA's GFP11 cargo DNA. GFP peptide chains 1-10 and 11 spontaneously assembly when expressed in a cell together, yielding the correctly folded fluorescent protein.

GFP1-10 was cloned into this vector downstream of mCherry so that the expression of the plasmid could be detected even in the absence of GFP11. This yielded plasmid pCXLE-mCherry-GFP1-10, constructed as described in Appendix 5. A diagram of the intended GFP complementation assay between plasmid-supplied GFP1-10 and hsgRNA-supplied GFP11 is shown in **Figure 3.5**.

In order to determine if our GFP1-10 expression vector was functional, HEK293 cells were co-transfected with pCXLE-mCherry-GFP1-10 and a plasmid expressing GFP11 (pcDNA Zip-GFP11) as described in Section 6.11.5. Two days post-transfection, cells were analyzed by fluorescence microscopy **(Figure 3.6)**. Cells transfected with pCXLE-mCherry-GFP1-10 alone were positive for mCherry fluorescence but not GFP. Co-transfection with pcDNA Zip-GFP11 generated GFP fluorescence in some cells and overlapped entirely with mCherry fluorescence. This confirmed that the complementation assay was functional and that mCherry could be used to track the presence of pCXLE-mCherry-GFP1-10. During this assay, we noticed that the morphology of transfected HEK293 cells was abnormal and that

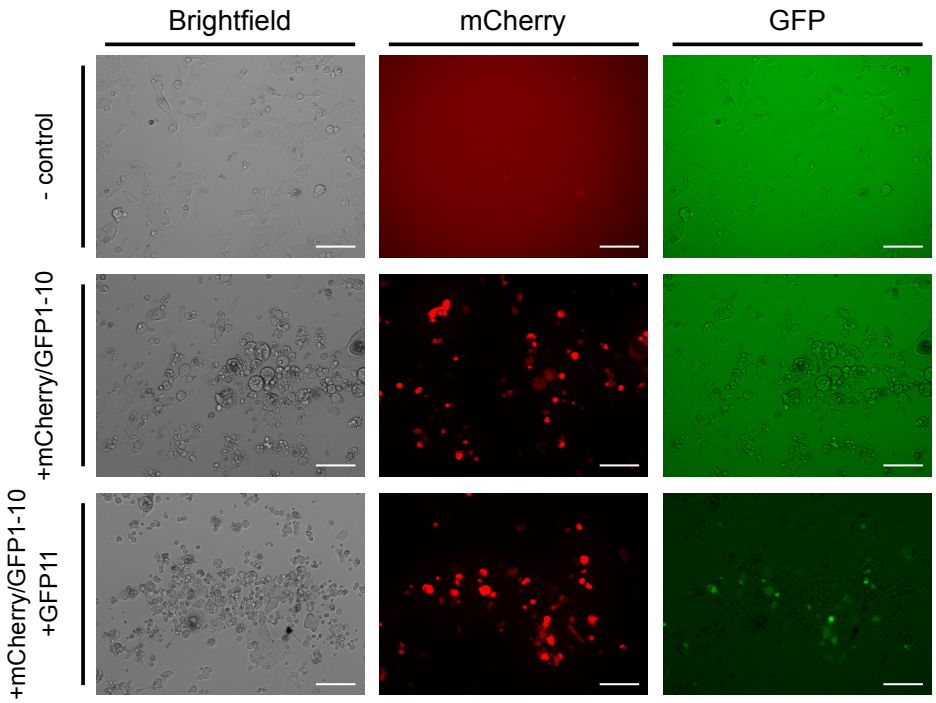


Figure 3.6: Light-microscopy analysis of HEK293 cells. Cells were either untreated, transfected with pCXLE-mCherry-GFP-10, or pCXLE-mCherry-GFP-10 and pcDNA Zip-GFP11 as indicated to the left of each row. Each population was imaged in brightfield, Tx red (mCherry) and GFP channels. Only the population transfected with pCXLE-mCherry-GFP-10 and pcDNA Zip-GFP11 exhibit GFP fluorescence. Scale bar = 100 µm.

many cells only loosely adhered to the plate, which was not observed in the non-transfected population. There is a lack of peer-reviewed literature regarding the adhesion of HEK293 cells but is generally agreed that the cell line is only semi- or loosely adherent, and that this can be further exacerbated by transfection conditions. In future experiments, this issue was corrected by applying Poly-D-Lysine to tissue culture plates prior to cell incubation and transfection, which provided a matrix to which HEK293 cells could more strongly adhere.

3.3.2) Design of a Minimal hgRNA DNA Cargo

After preparing a stable GFP1-10 expression vector, our next goal was to prepare a minimal-length expression cassette for GFP11 which could be used as the hgRNA TSR cargo. The

expression cassette needed to be capable of being transcribed, but only under the influence of a strong transcription-enhancing environment. To this end, we prepared a 403 bp sequence containing GFP11 under the control of a truncated CMV promoter (pCMVd1) to initiate transcription, and the SV40 polyadenylation sequence for efficient mRNA processing **(Figure 3.7a)**. The GFP11 coding sequence was flanked by multiple type IIS restriction sites so that alternate promoters and terminators can be cloned into the cargo if needed. The full sequence of the resulting DNA cargo (Cargo v1) can be found in Appendix 2.

To confirm that Cargo v1 could be expressed under *optimal* conditions, the sequence was cloned into the pCMV vector downstream of the CMV enhancer **(Figure 3.7d)**. The resulting pCMV-GFP11-v1 plasmid was co-transfected with pCXLE-mCherry-GFP1-10 into HEK293 cells but did not produce GFP fluorescence (data not shown). Because fluorescence was not observed while Cargo v1 was under the influence of a strong enhancer in *cis*, we reasoned that the hgRNA would not be responsive to *trans*-enhancer interactions in cells. It was likely that the truncated CMVd1 promoter and lack of 5′ UTR in Cargo v1 were hindering the expression of GFP11, so we replaced these elements by cloning the full-length CMV promoter into pCMV-Cargo-v1, generating GFP11 Cargo v2 **(Figure 3.7b)**.To

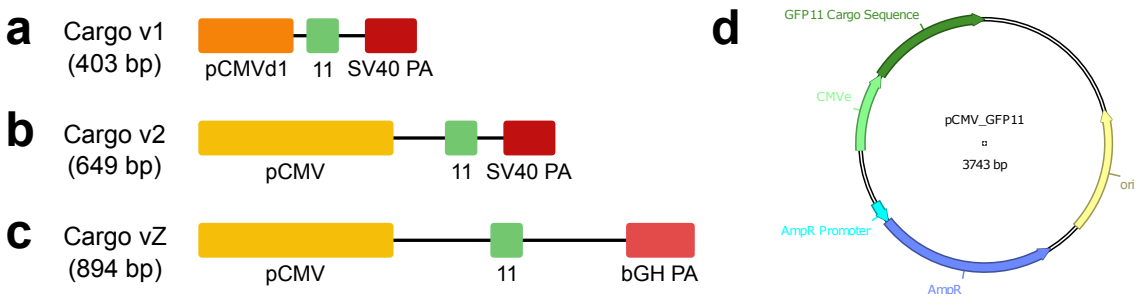


Figure 3.7: Simplified diagrams showing the approximate size of GFP11 cargos **(a)** v1, **(b)** v2, and **(c)** vZ. Promoter sequences (p) are colored in vermillion and yellow, the GFP11 coding sequence (11) in green, and polyadenylation sequences (PA) in red. **(d)** The pCMV vector used for cloning and enhancing the expression of each cargo in cell assays.

determine if the new design improved GFP11 expression, HEK293 cells were co-transfected with pCXLE-mCherry-GFP1-10 and pCMV-GFP11-v1, pCMV-GFP11-v2, or pcDNA Zip-GFP11. The core GFP11 expression cassette of pcDNA Zip-GFP11 **(Figure 3.7c)** was also under the control of the CMV enhancer. Two days after transfection, cells were analyzed by flow cytometry for GFP fluorescence. Histograms displaying GFP fluorescence of co-transfected populations overlaid against cells transfected with pCXLE-mCherry-GFP1-10 alone are shown in **Figure 3.8**. As with our microscopy experiments, we found that co-transfection of pCXLE-mCherry-GFP1-10 and pcDNA Zip-GFP11 produced a strong GFP fluorescence signal; in this case, approximately 31% of cells were GFP positive. However, co-transfections performed with either pCMV-GFP11-v1 or pCMV-GFP11-v2 did not increase the number of GFP-positive cells compared to the mCherry-GFP1-10 control.

Our goal for GFP11 cargo design was to produce a minimally expressed sequence with the smallest length possible. The design of Cargo v1 and v2, which had extensively reduced 5′ and 3′ UTRs, appeared to be too short to be effectively transcribed. For these reasons, we

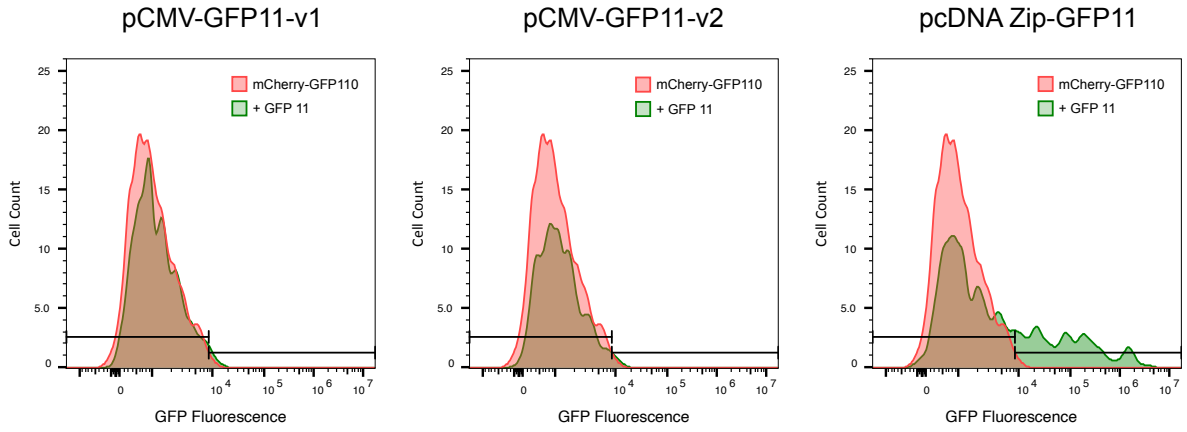


Figure 3.8: EGFP fluorescence of HEK293 cells transfected with pCXLE-mCherry-GFP-10 and one of pCMV-GFP11-v1, pCMV-GFP11-v2, or pcDNA Zip-GFP11 (as listed above histograms). Each histogram is overlaid with the GFP fluorescence of a control pCXLE-mCherry-GFP-10 transfected population (red). This population was used to set a threshold for GFP positive cells, indicated by the horizontal bars on each histogram. The GFP + rate of the three transfection conditions was 2.19, 2.06, and 31.3% respectively.

decided to adopt the GFP11 expression cassette in pcDNA Zip-GFP11 as the DNA cargo, which generated a much stronger GFP signal under the influence of a *cis* enhancer and was only about twice as long as our original design.

3.4 hgRNA Construction and Characterization

3.4.1) Construction and Characterization of hdgRNAs

While optimizing the design of the GFP11 DNA cargo, we also began to explore different DNA cargo/gRNA ligation strategies. We had not finished the design of our amine-functionalized ^{CNV}K-tracrRNA (Section 2.5) at the time of this work, but we knew a successful approach to covalently linking cargo/sgRNA would likely require a combination of photo-crosslinking and click-ligation. To this end, we synthesized a 3′ azido-modified crRNA which targeted a locus upstream of the human GAPDH core promoter (azide-crRNA^{GAPDH}, spacer sequence = AUUAUCAGGUCCAGGCUACA). GAPDH is widely used as a control in gene expression assays owing to its constitutive expression in most cell lines, which also makes it a suitable target for our first hgRNA. Apart from the spacer sequence, azide-crRNA^{GAPDH} was isosequential to the previously described EMX1 control crRNA (Section 2.4), making it compatible with the tracrRNA used in that section.

We next considered how to click-ligate azide-crRNA^{GAPDH} to one of the much larger GFP11 cargos. Previous work I undertook during my rotation project with the Brown group at the beginning of my DPhil demonstrated that untemplated click ligations between short RNAs and large dsDNAs are highly unfavorable. This is likely due to the large electrostatic effects between the two molecules preventing interaction between the azide and alkyne residues, and the fact that untemplated reactions are effectively at much lower substrate concentrations than templated versions. As mentioned earlier, we were also interested in sealing the DNA

cargo's termini to prevent exonuclease digestion. To accommodate both factors, we designed an approach of click-ligating azide-crRNA^{GAPDH} to a short DNA hairpin, which could then be ligated to a linear DNA cargo to form a circular-DNA/crRNA hybrid (c-crRNA) **(Figure 3.9a)**.

The key to this approach was the two BsaI restriction sites incorporated into our GFP11 cargo. Because BsaI and other type IIS restriction endonucleases cleave DNA outside of their restriction sites, it is possible to generate "custom" 4 nt DNA overhangs through digestion. BsaI digestion of pCMV-GFP11-v1 released the GFP11 cargo with unique GGAC

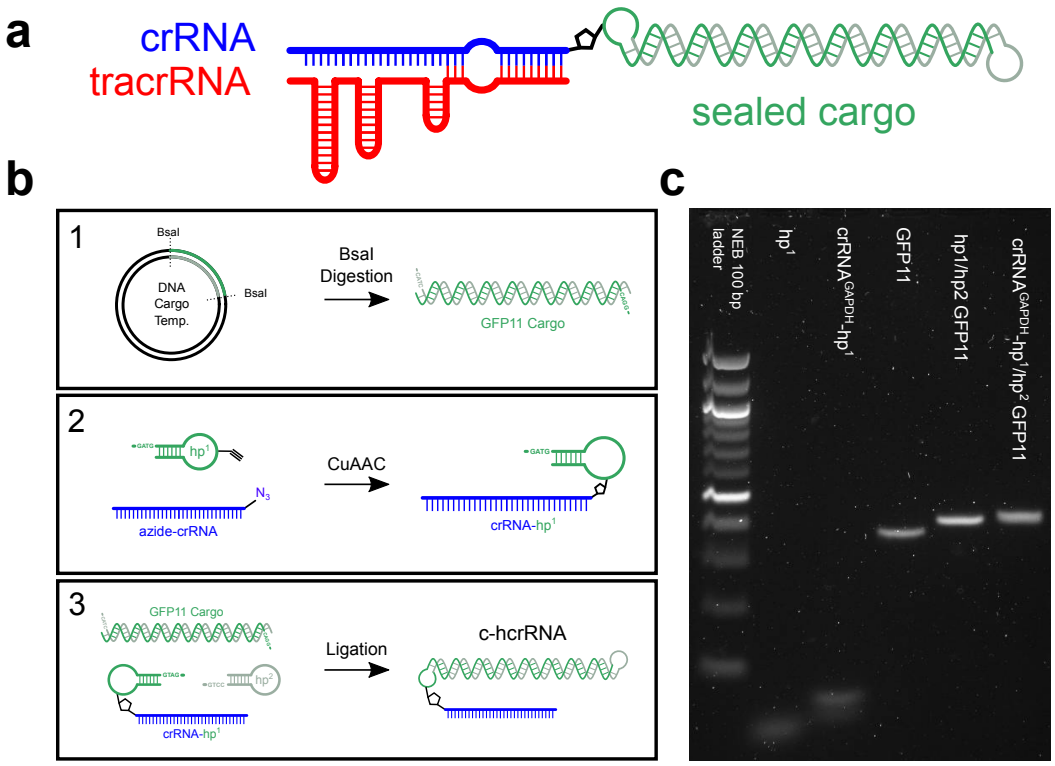


Figure 3.9a: Diagram of a circular-hybrid dual gRNA (c-hdgRNA). The crRNA (blue) is linked by a triazole (black) to a circular DNA cargo (green). The crRNA interacts with the tracrRNA (red) through Watson-Crick base pairing. **Figure 3.9b**: Diagram of circular-hybrid crRNA (c-hcrRNA) construction. **(1)** GFP11 cargo is digested via BsaI restriction sites and then purified, leaving unique CATC and CAGG 5′ overhangs. **(2)** Click ligation of a 3′ azide-crRNA to the internal alkyne residue of hp^1 yields the RNA/DNA hybrid crRNA-hp^1, linked by triazole. **(3)** Enzymatic ligation of crRNA-hp^1 and hp^2 to the previously digested DNA cargo via their 5′ overhangs produces the c-hcrRNA. **Figure 3.9c:** Agarose gel analysis of c-hcrRNA construction. From left to right, samples analyzed with a 100 bp ladder, hp1, hp1 after click ligation with crRNA^{GAPDH}, GFP11-v1 cargo after BsaI digestion, ligation of GFP11 cargo with hp^1 and hp^2, and the ligation of GFP11 cargo with crRNA^{GAPHD}-hp^1 and hp^2.

and CATC 5′ overhangs **(Figure 3.9b, 1)**. Azide-crRNA^{GAPDH} was click-ligated to a DNA hairpin (hp^1) which had a C8-alkyne-modified dU residue in the center of a T_{20} loop **(Figure 3.9b, 2)**. The resulting crRNA^{GAPDH}-hp^1 ligation product was then purified by denaturing PAGE and confirmed by HPLC-MS (Appendix 2). Finally, the linear GFP11 cargo could be enzymatically ligated with crRNA^{GAPDH}-hp^1 and an unmodified hp^2. Hp^1 and hp^2 self-anneal to form GATG and GTCC 5′ overhangs respectively that are complementary to the ends of the GFP11 cargo, ensuring correct ligation **(Figure 3.9b, 3)**. The resulting c-hcrRNA^{GAPDH} ligation product was purified by denaturing PAGE.

To confirm the results of the ligation reaction, purified c-hcrRNA^{GAPDH} was analyzed by agarose gel alongside hp^1, crRNA^{GAPDH}-hp^1, linear GFP11, and hp^1/hp^2 ligated GFP11 **(Figure 3.9c)**. By gel, we observed an increase in the size of the click-ligated crRNA^{GAPDH}-hp^1 compared to the precursor hp^1. This was also the case for crRNA^{GAPDH}-hp^1/hp^2 ligated GFP11, which migrated slower through the gel than hp^1/hp^2 ligated GFP11 and linear GFP11, suggesting that the ligation reaction was successful.

Encouraged by these results, we were interested in determining whether crRNA/DNA hybrids could be used in other biochemical reactions. To this end, we designed a second method of hgRNA construction in which a crRNA/PCR primer hybrid is used to amplify a cargo of interest, producing a linear hybrid crRNA (l-hcrRNA)**(Figure 3.10a)**. The DNA cargo of l-hdgRNAs would not have sealed termini from nucleases like c-hdgRNAs, and the high temperatures of PCR might cause degradation of the crRNA component. However, l-hcrRNA construction would be much simpler to implement than c-hcrRNA, given that l-hcrRNA construction will require fewer purification steps and minimal amounts of plasmid.

This would provide a convenient method for producing large amounts of RNA/dsDNA conjugates which could potentially be used in more generalized applications.

To construct l-hcrRNAs, azide-crRNA^{GAPDH} was click-ligated to a 5′ alkyne modified GFP11 forward (GFP11$_F$) PCR primer **(Figure 3.10b, 1)**. The ligation product was isolated by denaturing PAGE and confirmed with HPLC-MS (Appendix 2). The crRNA^{GAPHD}-GFP11$_F$ primer was then used to amplify cargo from pCMV-GFP11-v1 with a partner GFP11$_R$ primer **(Figure 3.10b, 2)**. PCR reactions performed with unmodified GFP11$_F$/GFP11$_R$ and crRNA^{GAPHD}-GFP11$_F$/GFP11$_R$ were analyzed and compared by agarose gel electrophoresis **(Figure 3.10c)**., demonstrating that the addition of the crRNA did not impact the overall reaction. Details of the PCR reaction can be found in Appendix 4.

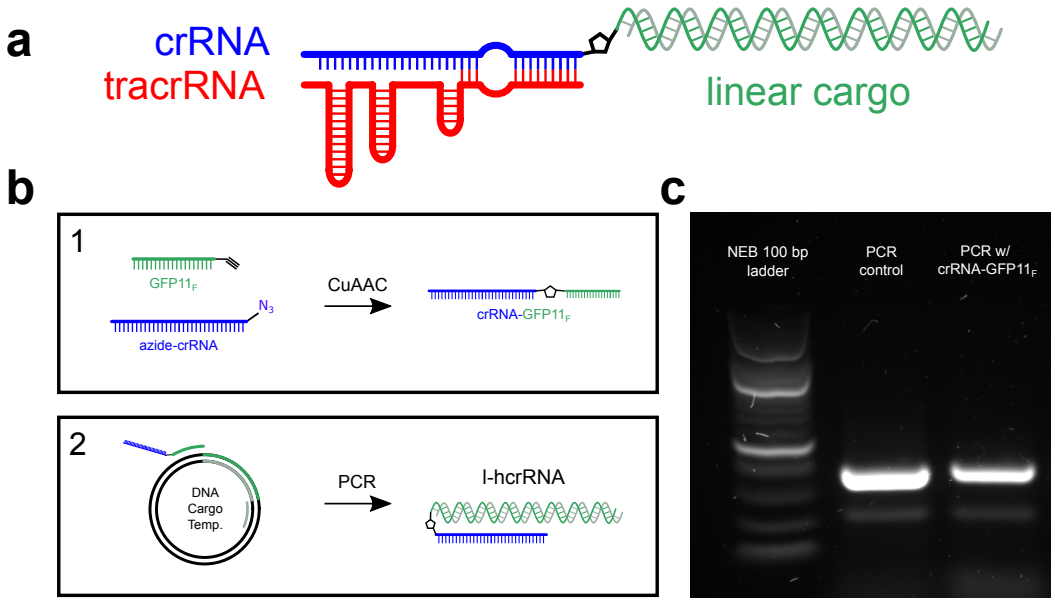


Figure 3.10a: Diagram of a linear-hybrid gual gRNA (l-hdgRNA). The crRNA (blue) is linked by a triazole (black) to a linear DNA cargo (green). The crRNA interacts with the tracrRNA (red) through Watson-Crick base pairing. **Figure 3.10b:** Diagram of linear-hybrid crRNA (;-hcrRNA) construction. **(1)** A 3′ azide-crRNA is click-ligated to a 5′ alkyne-GFP11 forward primer, yielding a DNA/RNA hybrid crRNA-GFP11$_F$ primer. **(2)** crRNA-GFP11$_F$ is used with GFP11$_R$ primer to PCR amplify cargo DNA, resulting in a GFP11 cargo linked by triazole to a crRNA. **Figure 3.10c:** Agarose gel analysis of GFP11 PCR reactions. PCR was carried out with pCVM-GFP11-v1 and either GFP11$_F$/GFP11$_R$ (PCR control) or crRNA-GFP11$_F$/GFP11$_R$. Gel lanes are labeled accordingly. Both reactions produced products of the expected size ~400 bp.

Both hcrRNA construction methods appeared to yield full-length DNA products of the expected size. However, it was difficult to determine by gel-based techniques whether the conjugated crRNA was degraded, and the crRNA/DNA hybrid was too large to analyze by HPLC-MS. To determine if these two methods produced functional hcrRNAs, we therefore tested the activity of each hcrRNA with an *in vitro* DNA binding electromobility shift assay (EMSA). A fluorescently labeled 56 bp DNA duplex that contained the GAPDH target sequence was prepared. dCas9-RNP would then be applied to the DNA substrate, whose migration was then tracked by native PAGE to determine whether a stable dCas9/gRNA/DNA complex would form. For this assay, dgRNAs were formed by annealing c-hcrRNA^{GAPDH}, l-hcrRNA^{GAPDH}, or crRNA^{GAPDH} with an equimolar amount of tracrRNA. An IVT-sgRNA^{GAPDH} was also prepared to use as a positive control.

An EMSA was performed with these 4 gRNA species as described in Section 6.10.2. The native-PAGE analysis of the DNA binding reactions, visualized by FAM fluorescence, is shown in **Figure 3.11**. Unbound DNA substrate migrates towards the bottom of the gel (Lane 1), and DNA bound non-specifically by dCas9 remains in the wells of the gel (Lane 2). Incubating the substrate DNA with either unmodified sgRNA or dgRNA causes the formation of the dCas9/gRNA/DNA complex (Lanes 3 and 4). C-hdgRNA (Lanes 6 and 7) and l-hdgRNA (Lanes 8 and 9) were also capable of forming dCas9/gRNA/DNA complex, indicating that they retain at least some gRNA function. However, both were less active than their unmodified dgRNA counterpart. C-hdgRNA binding activity was notably greater than l-hdgRNA activity, suggesting that the PCR-based approach to l-hdgRNA did impact the viability of the click-ligated crRNA. Importantly, incubating the target substrate and dCas9-RNP in the presence of non-ligated GFP11 PCR product (Lane 5) does not decrease dCas9/gRNA/DNA complex formation. This demonstrates that the DNA cargo itself does

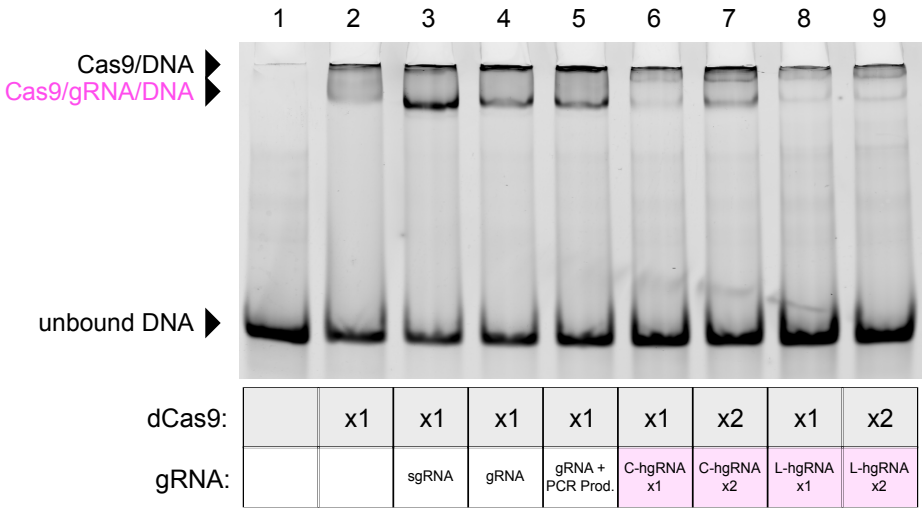


Figure 3.11: EMSA of FAM-labeled DNA substrate binding by dCas9-hdgRNAs. An *in vitro* binding assay was prepared with 1 pmol of a FAM-labeled 60-mer (from GAPDH) and 1 or 2 pmol of dCas9 and a guide RNA (as described below gel) in 20 µL of 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl2, and 100 µg/ml BSA. Reactions were incubated for 30 min at 37°C before being loading onto an 8.5% Native PAGE and run at 200 V for 2 h. Gel was then imaged with an Amersham Typhoon imager using the Cy2 channel. Arrows to the left of gel indicate the migration position of free DNA duplex, non-specific complex between DNA duplex and dCas9, and DNA Duplex-RNP complex.

not inhibit complex formation. These results suggest that the reduced activity of c-hdgRNA and l-hdgRNA is caused by crRNA degradation or steric effects between the click-ligated PCR product and dCas9 protein.

3.4.2) hsgRNAs Prepared by Enzymatic Ligation

Overall, the Cas9/dgRNA-RNP species were less effective at binding the substrate DNA than their sgRNA counterpart. Given that sgRNAs generally have higher activity than dgRNAs, this result was not unexpected. However, this also led us to believe that it would be disadvantageous to test our hdgRNA-GFP11 constructs in cells before we were able to apply our ^{CNV}K crosslinking platform to convert them to hsgRNAs. To compensate for this, we decided to use an enzymatic approach to construct IVT-sgRNA/GFP11 hybrids **(Figure**

3.12a) to begin testing the transcriptional reporter assay. To accomplish this, we designed a set of PCR primers to amplify the functional GFP11-vZ cargo. The GFP11-vZ$_F$ primer contained a 5′ 20 nt extension separated by a HEG linker, which left a significant amount of ssDNA substrate after PCR **(Figure 3.12b, 1)**. The 3′ end of the IVT-sgRNA^{GAPDH} was ligated to this ssDNA extension, templated by a 40 nt DNA splint **(Figure 3.12b, 2)**, and was then purified by denaturing PAGE. The difference in size between the 102 nt ssRNA sgRNA and 894 bp dsDNA GFP11 made it difficult to discern whether the ligation had occurred. To account for this, GFP11-vZ was digested with HindIII after steps 1 and 2 of this scheme,

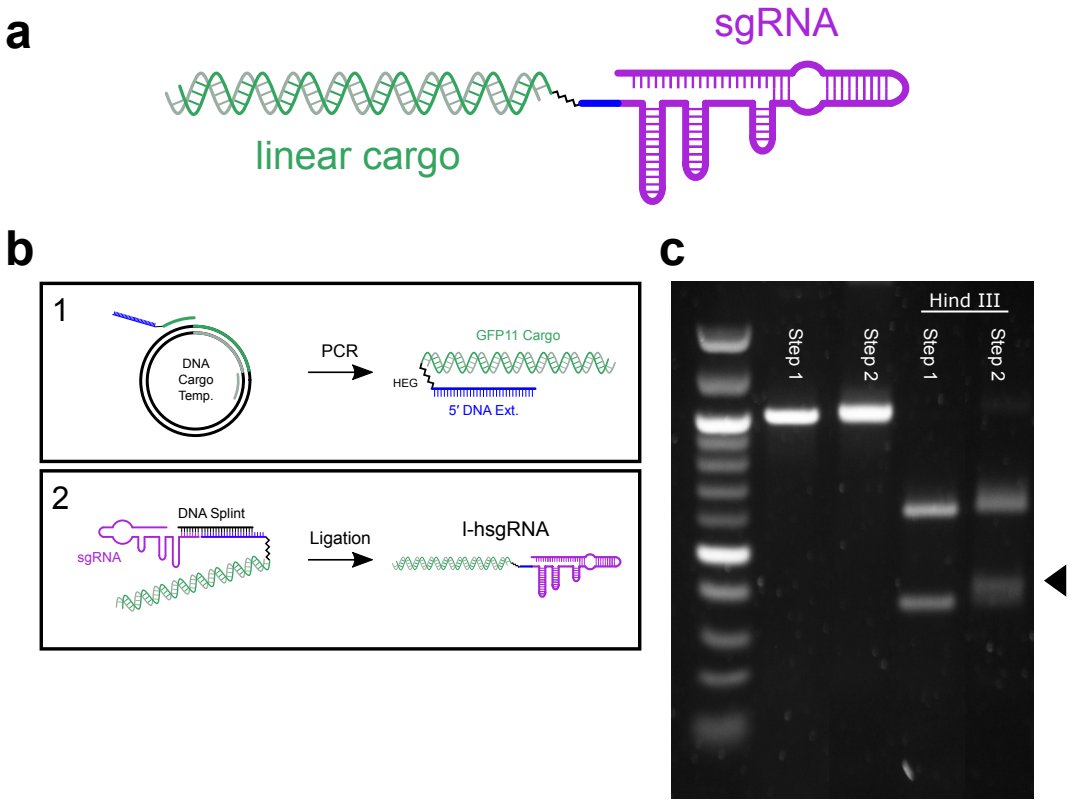


Figure 3.12a: Diagram of a linear-hybrid single gRNA (l-hsgRNA). An sgRNA (purple) has been enzymatically ligated to a short ssDNA sequence (blue), which is connected to a linear PCR product through a HEG linker. **Figure 3.12b:** Diagram of l-hsgRNA construction. **(1)** A DNA primer with a 5′ extension separated by a HEG linker is used to PCR amplify the GFP11 DNA cargo, generating a full-length GFP11 amplicon (green) with a 5′ ssDNA extension (blue). **(2)** An IVT-sgRNA is annealed to the 5′ ssDNA extension via a DNA splint, and is then enzymatically ligated to the PCR product. Purification by denaturing PAGE isolates the l-hsgRNA ligation product. **Figure 3.12c:** Agarose gel analysis of the two steps of l-hsgRNA construction. Lanes are labeled according to step in construction, PCR (1) or ligation (2). Samples on the right of the gel were digested with HindIII, which cleaves the GFP11 PCR product into two halves. Black arrow to the right of the gel indicates the 5′ fragment, which increases in size after the ligation step, corresponding to the addition of the sgRNA. An NEB 100 bp ladder was run on the left of the gel.

which cleaved the DNA 367 bp from its 5′ end and made the ligation event easier to observe **(Figure 3.12c)**. We observed an increase in the size of the 5′ fragment of GFP11-vZ after the ligation step **(2)**, suggesting that the IVT-sgRNA was successfully attached to the PCR product.

After preparing l-hsgRNA^{GAPDH}, we next wanted to determine if it could be transcribed in cells. To accomplish this, 1x10^4 HEK293 cells were plated in a 96-well plate and left to incubate for 48 h. At this point, each population was transfected with pCXLE-mCherry-GFP1-10 and left to incubate for another 24 h to produce a high GFP1-10 background prior to hsgRNA transfection. After this incubation period, cells were transfected with either 0.125, 0.25, or 0.5 pmol of dCas9/l-hsgRNA^{GAPDH} RNP (Section 6.11.3). An additional population was transfected with 0.5 pmol of GFP11-vZ PCR product to gauge the background signal generated by non-targeted DNA cargos. Cells were harvested and analyzed by flow cytometry 24 h after Cas9-RNP transfection. Flow events were gated for mCherry + cells to remove non-transfected cells, taking advantage of the mCherry-GFP1-10 fusion protein. Representative histograms of GFP fluorescence for each transfection condition are shown in **Figure 3.13**. The threshold for GFP +/- cells was set using a control pCXLE-mCherry-GFP1-10 transfected population.

Unfortunately, we observed that cells transfected with GFP11-vZ PCR product produced a very high GFP signal; 46% of cells were positive for GFP fluorescence, which is comparable to our best result from l-hsgRNA^{GAPDH} transfections. Transfection of 0.5 pmol of l-hsgRNA^{GAPDH} produced the smallest GFP signal out of the three dilutions used, which was likely due to the significant amount of cell death we observed after transfecting this quantity of RNP. The other two quantities of l-hsgRNA^{GAPDH} transfection produced a readable

fluorescent signal without significant cell death; however, the background expression of our DNA cargo was too high to discern if there was increased transcriptional activity due to target locus transcription. We were surprised by the intensity of fluorescence produced by GFP11-vZ given that the linear PCR fragment did not have an enhancer sequence and would have been highly susceptible to nuclease degradation. Additionally, we could not find an abundance of previous literature on the efficacy of gene expression through transfected PCR products. In either case, the DNA cargo was expressed too highly in cells to be suitable for a transcriptional reader assay.

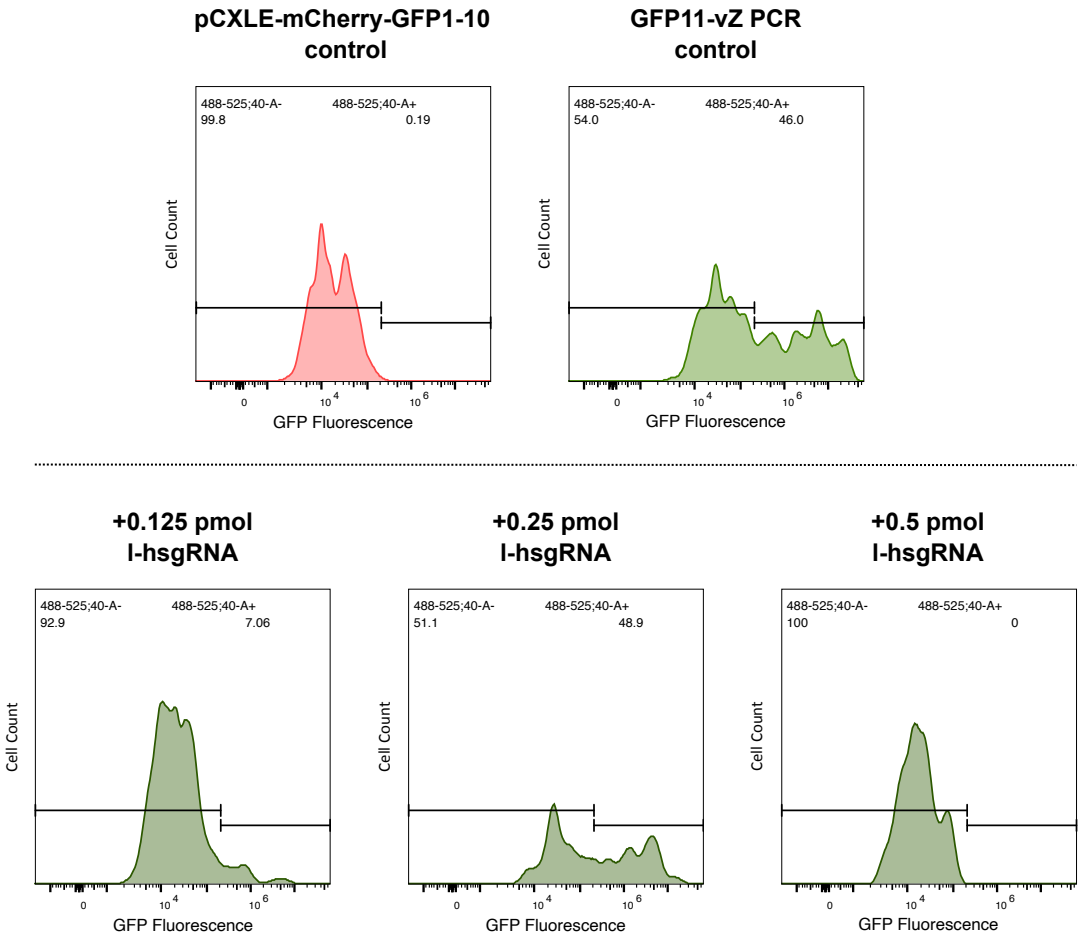


Figure 3.13: EGFP fluorescence of HEK293 cells transfected with pCXLE-mCherry-GFP-10 and increasing amounts of l-hsgRNA^{GAPDH}. Cells were transfected with pCXLE-mCherry-GFP1-10 and either 0.5 pmol GFP11 PCR product or a varying quantity of dCas9/l-hsgRNA^{GAPDH}. One population was transfected with only pCXLE-mCherry-GFP1-10 to use as a GFP - control (shown in red). The threshold for GFP +/- cells were determine using this control, as shown by the brackets on each histogram. The percent of GFP + and - cells is listed at the top of each histogram.

3.5 Conclusion and Future Works

The development of robust techniques to interrogate the dynamic behavior of gene expression will further our understanding of complex mechanisms that control cell behavior. The original intention of this project was to explore new strategies for constructing hgRNAs, with an ultimate goal of producing TSRs from these molecules. Of the three methods we explored, we found that the c-hcrRNA construction method was quite effective and had the bonus of creating a "sealed" DNA cargo with no exposed termini. Unfortunately, we found that the *in vitro* DNA binding activity of c-hdgRNAs suffered compared to dgRNA controls. It is possible that this could be ameliorated by implementing the c-hcrRNA construction with our ^{CNV}K-sgRNA platform, which could be a good direction for future work.

When we implemented our l-hsgRNA TSRs in cells, we found that the background expression of our DNA cargo was too high to discern increased transcriptional activity due to target locus transcription. To continue this project, we would have needed to redesign a DNA expression cassette with a significantly lower background signal and a very dynamic expression range. Because of the amount of resources this would have required, I decided to focus on the work in my other two chapters instead. That said, we were intrigued by the amount of GFP signal a linear, enhancer-less PCR product could generate. The DNA hairpins we developed in this project could be applied to this material to produce a sealed PCR product. An alternative direction for the work in this project could be determining if these sealed PCR products can be a suitable vector for gene expression in eukaryotic cells, which could have applications in other areas.

4

Photo-Regulated mRNA Translation

4.1 Background

4.1.1) Photo-Regulation of mRNA Translation

As discussed in Section 1.9, there has been an explosion of interest in mRNA-based technologies since the widespread use of SARS-CoV-2 mRNA vaccines [141]. Along with other recent advances in this field, synthetic mRNA have become an incredibly relevant and promising technology for the development of novel vaccines [252] and therapies [135], and maintains their status as versatile and powerful research tools.

Biological systems are characterized by the inherent complexity of timing, dynamics, and localization of regulatory events. Because of this, methods of precisely controlling the spatiotemporal activation of gene expression are highly desirable. As mRNA is transiently expressed in cells and is immediately available to cells for use upon transfection, controlling its expression can produce bursts of gene expression in a localized and time-dependent

manner. Several strategies for inducing gene expression have been developed, such as ligand or protein-protein interaction-based systems [253]. While these methods are appropriate for controlling the behavior of large populations, they are unable to control the behavior of individual cells with high spatial or temporal precision and have little therapeutic relevance.

Conversely, light has been proven to be an excellent stimulus for conditional gene expression: it is non-invasive, can be applied with high precision, and acts orthogonally to cellular signals [254]. Additionally, recent advances in biomedical optics have led to the development of increasingly precise light-based diagnostic and therapeutic tools, as reviewed in [255]. Several strategies to regulate the translation of exogenous mRNA have been developed from photo-caging and photo-isomerizing oligonucleotide modifications [256]. Most aim to silence the expression of mRNA until a light signal is applied, creating a burst of translation in a highly controllable manner. In this background section, I will weigh existing methods of photo-regulating the translation of endogenous and exogenous mRNA as well as their short-comings, and then discuss our designs for photo-crosslinking carbazole nucleotides to form the basis of an mRNA photo-regulation system.

4.1.2) Photo-Regulation of Endogenous mRNAs with Oligonucleotides

Traditionally, the most accessible method to photo-regulate mRNA translation was accomplished by controlling the activity of short translation-inhibitory oligonucleotides. Because short oligonucleotides are relatively simple to modify, it is easier to incorporate photo-responsive moieties into siRNA or ASOs compared to full-length mRNA. An early example of a photo-controllable inhibitory oligonucleotide was the preparation of an anti-GFP siRNA whose phosphate backbone was indiscriminately labeled with a 4,5-dimethoxy-2-nitrophenylethyl (DMNPE) photo-labile group which blocked the interaction of the siRNA

and RISC complex [257]. Cleaving the DMNPE group with 355 nm light irradiation partially restored the interaction and triggered a drop in GFP expression in HeLa cells.

Since then, more sophisticated methods of inhibiting siRNA/ASO activity have been developed along with more sensitive photo-active compounds. Some common strategies for light-induced activation of inhibitory oligonucleotides are **a.)** linking the inhibitory oligonucleotide to a complementary strand by photo-cleavable linker [258, 259], **b.)** directly photo-caging nucleobases to prevent hybridization [260-262], **c.)** circularizing of the oligonucleotide to prevent target hybridization until it is linearized by photo-cleavage [263, 264], and **d.)** incorporating a large, sterically demanding group to the end of the oligonucleotide by photo-cleavable linker [265-267] **(Figure 4.1)**. Some efforts have also gone into developing toggleable inhibitory oligonucleotides that can switch between an active and inactive state using photo-isomerizing groups [268-270].

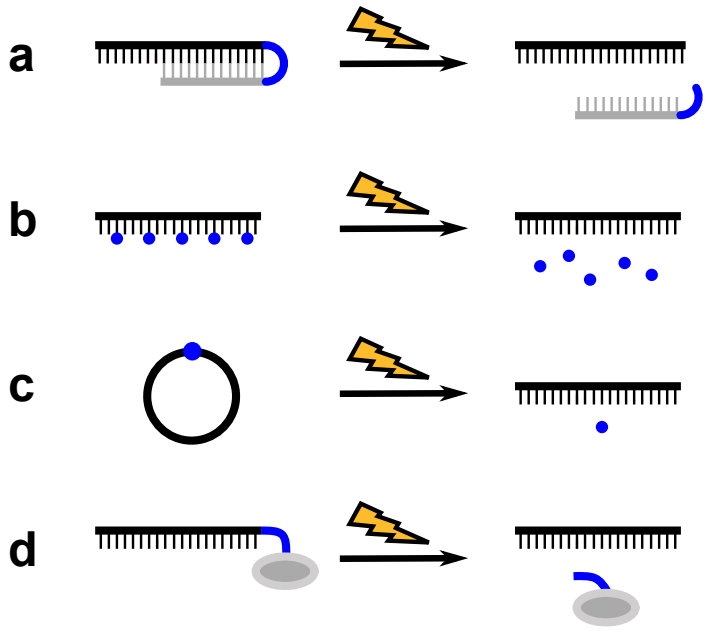


Figure 4.1: Simplified diagram of the photo-activatable translation inhibitory oligonucleotides. Some common strategies to activate mRNA inhibition include **(a)** linking the oligonucleotide to a complementary strand by photo-cleavable linker, **(b)** directly photo-caging bases, **(c)** circularizing the oligonucleotide to prevent hybridization with target, or **(d)** conjugating the oligonucleotide to a sterically hindering group to inhibit interactions with degradation machinery. Upon photo-irradiation, the blocking element is cleaved from the oligonucleotide, restoring function.

There is currently a robust body of photo-controllable translation-inhibiting oligonucleotide tools. These tools have been applied to control developmental and disease-relevant genes in single-cell culture, invertebrates, and mice [256]. While these tools have demonstrated the power and precision of light-controlled translation, they are largely used to regulate endogenous mRNA. As a tool to control exogenous mRNA, it would be ineffective and unreliable to deliver both an mRNA of interest and a sufficient excess of inhibitory oligonucleotide to control the former's expression. Additionally, many of the synthetic modifications used to control these short oligonucleotides are difficult to implement in full-length mRNA, which contributes to the disparity of tools to silence mRNA translation than to activate it.

4.1.3) Photo-Regulation of mRNAs with Small Molecules

Because of the challenges associated with modifying full-length mRNA, another approach to photo-regulate translation has been the development of photo-sensitive secondary systems which in turn regulate the translation of mRNA. These range in complexity, but typically involve introducing a photo-sensitive small molecule into the organism of interest, which can then affect the translation of endogenous or exogenous mRNA.

One dynamic approach has been the development of aptamers that change conformation upon binding to light-responsive ligands [271]. In particular, the engineered riboswitch were-1 will selectively bind the *trans* isomer of a stiff-stilbene 100-fold stronger than the *cis* isomer and change confirmation to occlude downstream sequences [272]. Were-1 has been incorporated into mRNA sequences to prevent access to ribosome-binding sites when in complex with *trans* stiff-stillbene isomers, reducing translation. Rotstan *et al.* implemented

this riboswitch to control the expression of luciferase in *E. coli* and could demonstrate the *trans-cis* photo-isomerization caused approximately a 4-fold reduction in translation after 1 hour.

Another instance of photo-regulation through secondary systems is the work done by Eichert *et al.*, who incorporated photo-caged amino acids into RNA-modification enzymes by expanding the genetic code of *E. coli* [273]. These cages inhibited efficient capping of unprocessed RNA with methylated guanosine, limiting their translation. By reconstituting capped enzymes with *in vitro* transcribed RNA, they were able to demonstrate *in vitro* a convoluted but effective method for activating translation and controlling the activity of RNA-modifying enzymes.

Both examples illustrate creative approaches to regulating translation and do so without directly modifying mRNA. However, they both struggled to fully inhibit the translation of mRNA and displayed a relatively low dynamic range of translation compared to other photo-regulation strategies. Although these two-part systems circumvent the difficulties of modifying long RNAs, they place extra demands on experimental design and preclude their use as therapeutics.

4.1.4) Photo-Regulation of mRNA by Post-Transcriptional Modification

Photo-regulation of mRNA through inhibitory oligonucleotides and secondary systems has been able to produce light-induced inhibition and activation of mRNA, but neither approach is easy to implement to control exogenous mRNA. Photo-regulation of translation through direct mRNA modification provides the most precise means to control mRNA behavior. The first report to describe photo-regulation of exogenous mRNA was published in 2001 by the

Okamoto group [274]. They were able to photo-cage *in vitro* transcribed mRNA **(Figure 4.2a)** by reacting it with a novel caging agent, 6-bromo-4-diazomethyl-7-hydroxycoumarin (Bhc-diazo), in DMSO. Bhc-diazo is presumed to react with free, non-ionized phosphoric acids in the RNA backbone to produce 6-bromo-7-hydroxycoumarin-4-ylmethyl esters of the phosphates **(Figure 4.2b)**. While it did provide direct control of mRNA behavior, this method had several flaws and was not widely implemented. Notably, although the Bhc caging could achieve robust mRNA silencing, it had incredibly poor rates of activation after irradiation and the mRNA suffered from site indiscriminate caging.. The caging conditions could also cause mRNA degradation, making a chemical approach to photo-caging less than ideal.

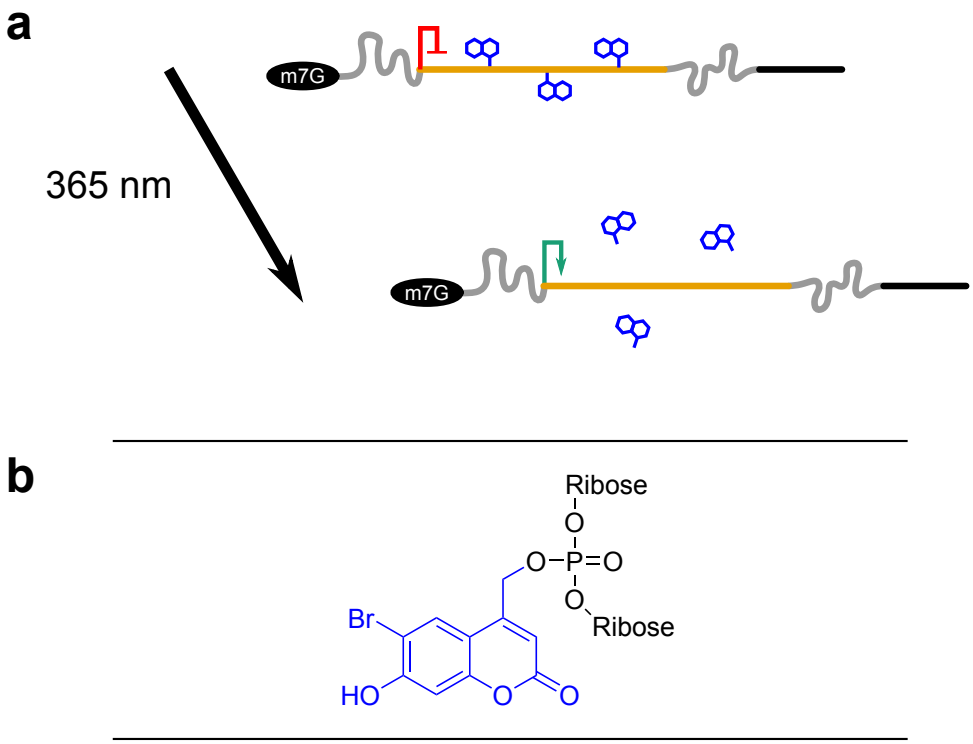


Figure 4.2: mRNA photo-regulation strategy described in [274]. **(a)** photo-caging groups (blue) placed on the mRNA phosphate backbone prevent ribosome progression and translation. Irradiation with 365 nm light cleaves the caging groups, permitting translation. **(b)** Chemical structure of Bhc caging group used in this strategy. The photoactive portion of the structure is highlighted in blue, and in this case will be cleaved from the phosphate backbone upon 365 nm irradiation.

In the past few years, several enzymatic approaches to modifying mRNA have emerged as an alternative to chemical methods and have been applied to introduce photo-sensitive moieties in gentler conditions. To this extent, several groups have begun to explore a more targeted approach to caging mRNA by repurposing nucleobase-modifying enzymes. In 2015, the Deveraj group demonstrated that *E. coli* tRNA guanine transglycosylase (TGT) could accept some modified pre-queuosine$_1$ (preQ1) substrates to modify guanine residue [275]. By inserting the 17-nt TGT recognition motif into RNAs, the introduction of modifications could be controlled in a site-specific matter. The Deveraj group went on to apply this system to modify mRNA with preQ1 linked to biotin by a photocleavable 6-bromo-7-aminoethoxycoumarin-4-ylmethoxycarbonyl (Bac) linker. By introducing three TGT recognition tags upstream of the start codon and introducing this modification, they were able to reduce the translation of an *in vitro* transcribed mRNA by 90% [276]. Upon 405 nm irradiation, they also report removing up to 90% of photo-caging groups **(Figure 4.3a)**. This work was followed up with the development of a second photo-caged preQ1 nucleoside, which used a 488 nm sensitive photo-linker, 7-(diethylamino)coumarin-4-yl]-methyl (DEACM) [277]. By labeling *in vitro* transcribed mCherry and GFP mRNA with Bac **(Figure 4.3b)** and DEACM **(Figure 4.3c)** respectively, they were able to demonstrate multiplexed photoactivation of two mRNA species within a single cell. The orthogonal photoactivation wavelengths of each cap allowed them to be activated sequentially from higher to lower wavelength. The system does have some limitations, however: separate elements within a single mRNA cannot be labeled individually with the two photo-caging groups, since TGT would label tags indiscriminately. Additionally, it is unclear whether the DEACM cage improves upon the translation silencing rate of Bac cages.

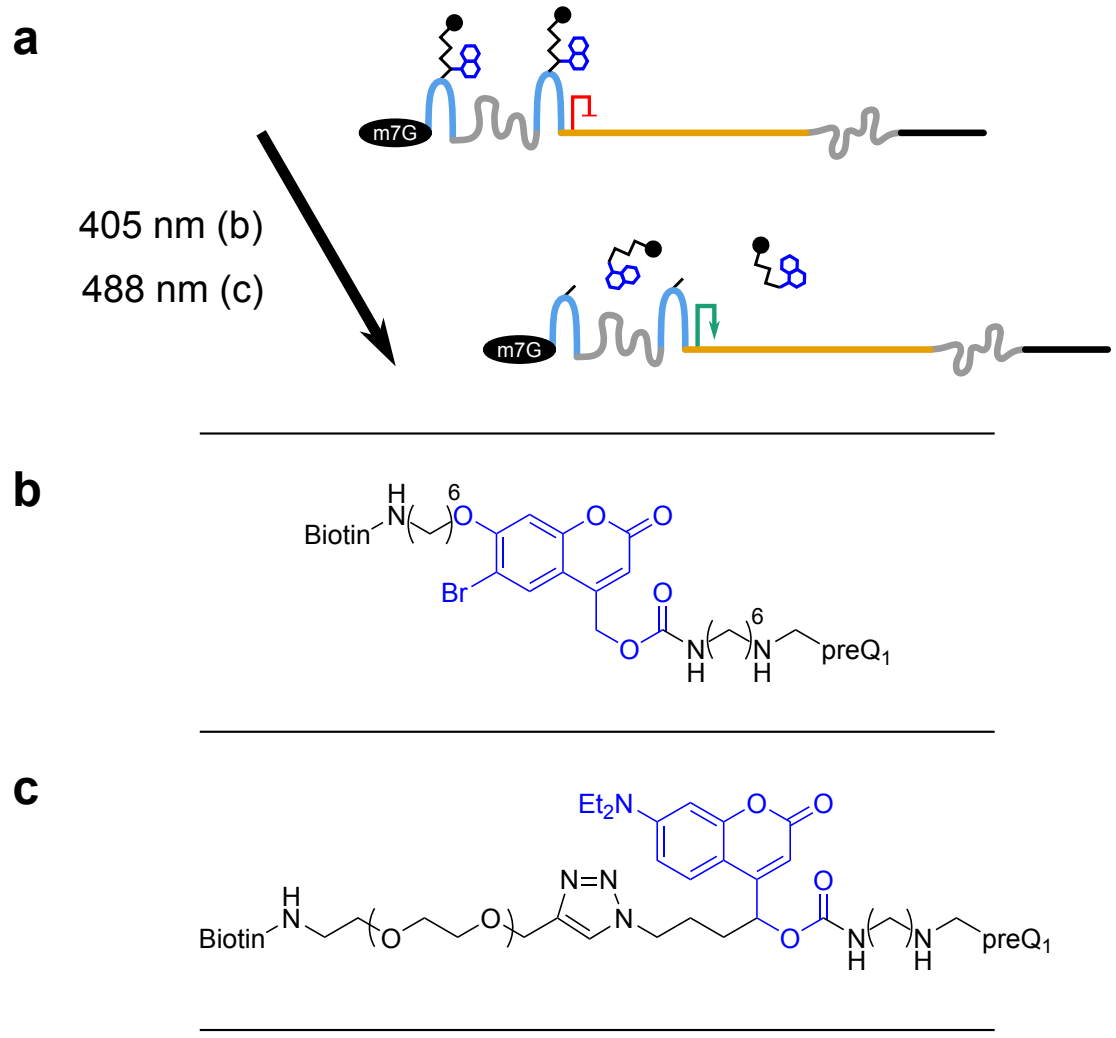


Figure 4.3: mRNA photo-regulation strategy described in [276,277]. **(a)** $PreQ_1$ bases linked with a biotin group via a photocleavable linker (blue) are introduced into TGT recognition tags (light blue) placed upstream of the mRNA ORF (vermillion). The biotin groups prevent ribosome progression through the mRNA and identification of the start codon. Irradiation with 405 nm or 488 nm light cleaves the biotin linker, permitting translation. The chemical structures of the **(b)** Bac and **(c)** DEACM photo-cleavable linkers are colored in blue. Upon irradiation with 405 nm or 488 nm light respectively, the linker will be cleaved, removing the biotin group.

It is possible that other RNA-modifying enzymes could be used in the future to produce photo-regulated mRNA. A recent publication attempted to repurpose two RNA methyltransferases, METTL3-METTL14 and METTL16, to introduce synthetic modifications onto adenine residues [278]. Both methyltransferases preferentially label adenine at consensus sequences, but these are short (6 and 9 nt respectively) and less restrictive than TGT tags. METTL3-METTL14 was successfully used to modify adenine

with *ortho*-nitrobenzyl and 6-nitropiperonyl photo-caging groups. However, the caging efficiency at consensus sequences was only 30 and 29% respectively, and neither modification was demonstrated in a translation assay. This does highlight the potential that RNA modifying enzymes offer as an alternative to chemical modification of RNA, and that they may provide relevant photo-regulation strategies in the future.

4.1.5) Photo-Regulation of mRNA by 5′ Cap Modification

Several groups have explored modifying the 5′ mRNA cap to introduce photo-sensitive entities which are meant to block the interaction between mRNA and eIF4E, the critical cap-binding enzyme required for assembly of the 80s ribosome on mRNAs (reviewed in [279]). One advantage of this route as opposed to the enzymatic modification of nucleobases is that modified caps can be incorporated into mRNA co-transcriptionally, reducing the number of steps required for photo-caging. The first attempt at regulating translation through photo-responsive caps was described by Ogasawara and Maeda in 2011 [280], who synthesized three photo-isomerizing 5′ caps: 8-styryl cap (8ST-cap), 8-naphthylvinyl cap (8NV-cap), and 8-fluorenylvinyl cap (8FV-cap). Each cap undergoes *cis*-*trans* photoisomerization when irradiated with 410 nm and 310 nm light respectively. When the cap is in the *trans*-state, Ogasawara proposes that the interaction between eIF4E and the 5′ cap is blocked, preventing ribosome recruitment to the mRNA and initiation of translation **(Figure 4.4a)**. Using an *in vitro* translation assay, they were able to demonstrate that the *trans* conformation of each cap did repress translation compared to the *cis*-state, although none were able to totally silence translation [280].

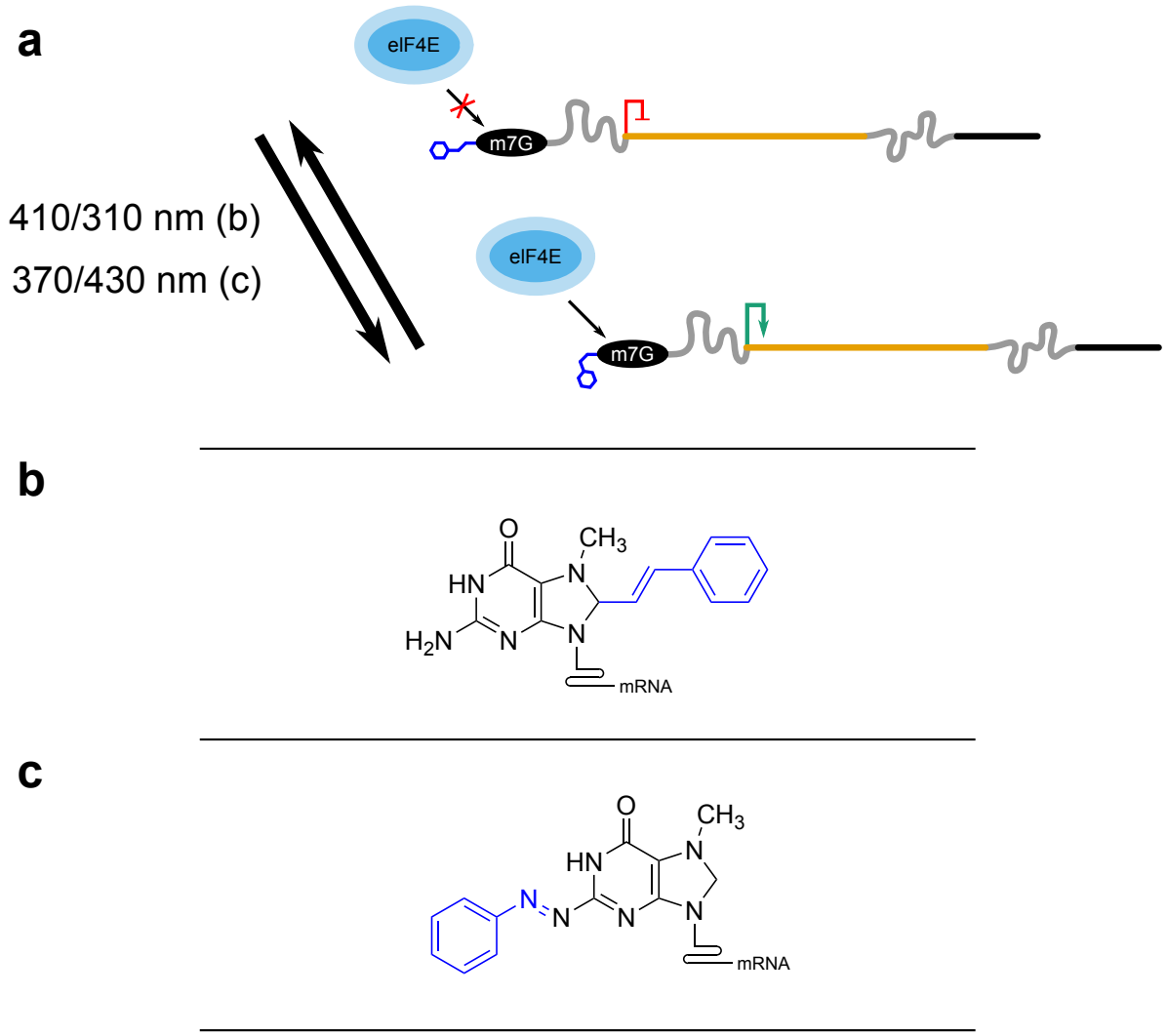


Figure 4.4: mRNA photo-regulation strategy described in [280-282]. **(a)** A chemically modification (blue) on the mRNA 5′ cap blocks interactions with eIF4E, preventing assembly of the ribosome loading complex on the mRNA. Upon photo-isomerization from *trans* to *cis* (410 or 370 nm), the modification ceases the steric blockage of eIF4E, allowing ribosome loading and translation. Isomerization to the *trans* state (310 nm or 430 nm) resumes blocking. The chemical structures of the **(b)** 8ST and **(c)** mMe-2PA photoisomerization groups attached to a 7-methylguanosine cap are colored in blue.

In 2014, Ogasawara followed up by publishing the first instance of light-regulated translation *in vivo* with a cap modification [281]. Curiously though, he only used the 8ST-cap **(Figure 4.4b)**, which was shown to inhibit translation by only 25% in its *trans*-state in their previous publication [280]. Here, Ogasawara reports that the 8ST-cap *trans*-state represses translation to a level comparable to his negative control, but does not offer an explanation for why the cap behaved so differently between the publications. That being said, here he reports being

able to induce a 12.7-fold increase of *in vivo* mRNA translation upon photo-irradiation and applies the technology to induce differentiation of rat phaeochromocytoma PC12 cells.

This strategy was improved upon by the development of a new generation of caps that underwent *cis-trans* photoisomerization at 370 nm and 430 nm respectively [282], higher wavelengths with reduced toxicity compared to the previous caps. The cap with the most distinct on/off states, 2-meta-methyl-phenylazo cap (mMe-2PA-cap)**(Figure 4.4c)**, yielded a 7.1-fold increase of *in vivo* protein expression after photoisomerization. mMe-2PA-cap was then applied to control the ectopic expression of Squint protein in zebrafish embryos in pulses, which caused the formation of a second anterior-posterior axis. The key development of this technology was the ability to reversibly activate translation, as opposed to photo-uncaging strategies which irreversibly activate translation. However, these caps are still limited by a leaky translation in their off-state: the *trans*-state mMe-2PA-cap permitted 10.7% of the protein expression as positive controls in these experiments [282]. An additional weakness of these studies [281, 282] is that all *in vivo* data was assayed by microscopy and may be more susceptible to bias compared to plate reader or flow cytometry-based assays.

An alternative strategy of disrupting the eIF4E/mRNA cap interactions is to protect the cap with photo-cleavable groups, as has recently been explored by the Rentmeister group [283, 284] **(Figure 4.5a)**. Although activation through photo-cleavage is irreversible compared to the reversible strategy described by Ogasawara, they argue that the total removal of modifications on 5′-cap is beneficial for translation as artifacts could impede eIF4E interactions. They reported the synthesis of cap 0 analogs caged with 3,4-dimethoxy-2-

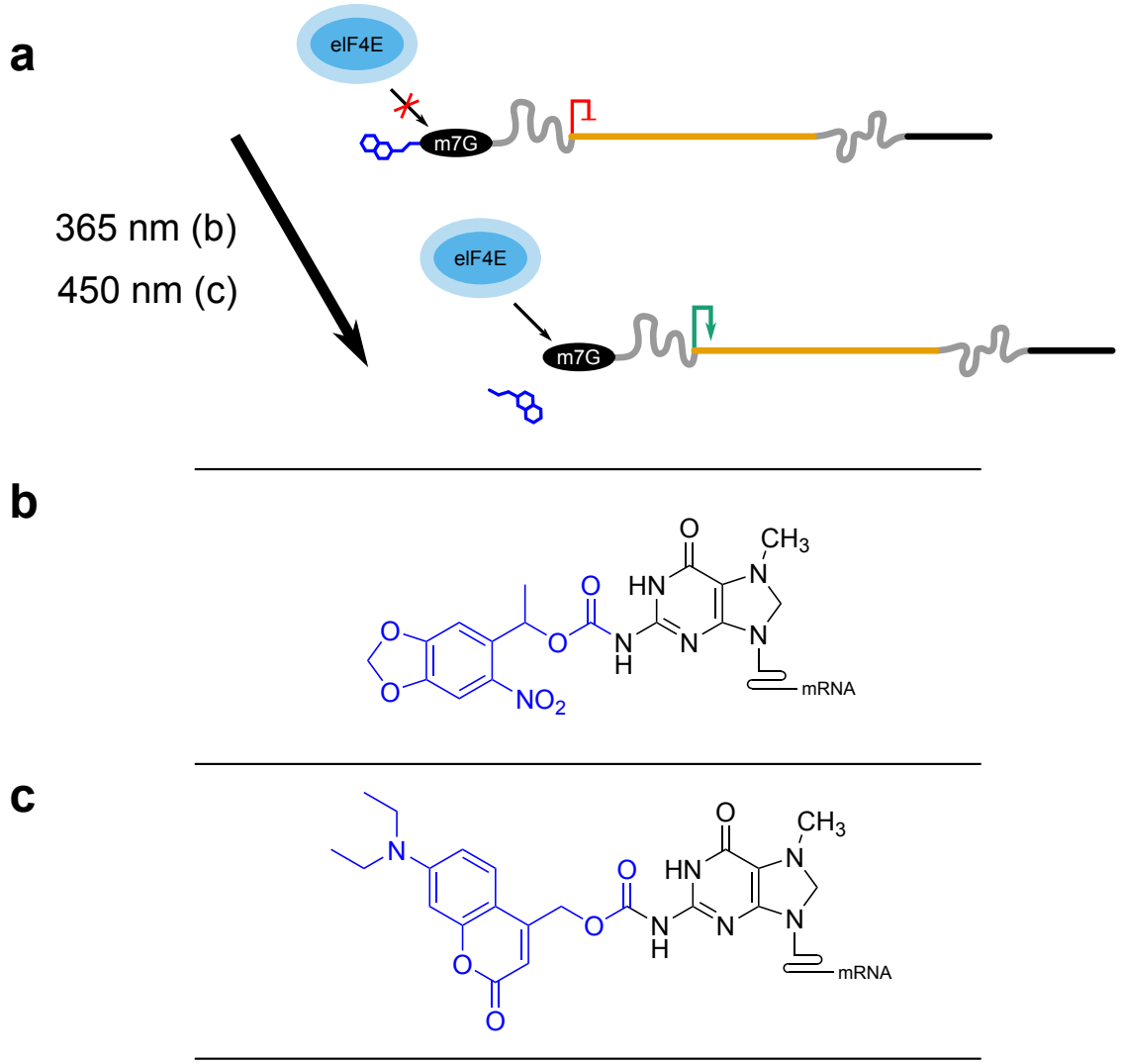


Figure 4.5: mRNA photo-regulation strategy described in [283,284]. **(a)** A chemically modification (blue) on the mRNA 5′ cap blocks interactions with eIF4E, preventing assembly of the ribosome loading complex on the mRNA. Upon photo-cleavage (365 or 450 nm) of the modification, eIF4E can interact with the cap to initiate ribosome loading and translation. The chemical structures of **(b)** NPM and **(c)** DEACM photo-caging groups attached to a 7-methylguanosine cap are colored in blue. Upon irradiation with 365 and 450 nm light respectively, the caging group is removed, leaving the natural cap 0 structure intact.

nitrobenzyl (DMNB), 6-nitropiperonyl-methyl (NPM) **(Figure 4.5b)** [283], and DEACM **(Figure 4.5c)** [284] at the N^2 position of the terminal guanosine. Like the work done by the Ogasawara group, a large advantage of these cap analogs is that they can be incorporated into mRNA strands co-transcriptionally. The improvement over the previous method is that NPM and DEACM caged caps can be used for multiplexed gene activation: because DEACM is

cleaved by higher-wavelength light than NPM, it is possible to sequentially activate mRNA translation by irradiating cells with 450 and 365 nm light respectively [284].

While this work demonstrates a convenient method for capping and controlling mRNA, it does have some shortcomings. Notably, their single-cell analysis of mRNA activation, which was set with a very generous threshold for mRNA expression "negative" cells, revealed that 18.9% of cells transfected with NPM-caged mRNA were positive for expression; activating translation with irradiation barely doubled the number of positive cells [283]. The authors attribute this leaky expression to partial uncaging of mRNA during the relatively longer course of this experiment. However, one must question the usefulness of a caging group which spontaneously uncages faster than the mRNA it is linked to is degraded in cells. Additionally, while their demonstration of multiplexed activation was impressive, the wavelengths used for activation were lower than those previously used by the Deveraj group [277], which should increase the toxicity of UV treatment.

4.1.6) ^{CNV}K Crosslinking Can Inhibit Translation

As discussed, several methods have been explored to photo-regulating the translation of exogenous mRNA, all of which have been developed from photo-caging or photo-isomerizing entities. To our knowledge, no group has attempted to control the translation of exogenous mRNA with photo-crosslinking groups, however. Interstrand crosslinks in mRNA terminate the translation of peptides at the location of the crosslink [285], and several crosslinking agents have been demonstrated to be capable of inducibly preventing mRNA translation when incorporated into translation-inhibitory oligonucleotides [21, 44, 45, 285-289]. While these agents have been applied to silence the expression of endogenous mRNA,

no group has taken the extra step to either activate translation after crosslink-induced silencing or to control the activity of an exogenous mRNA.

Because ^{CNV}Ks can be reversibly crosslinked, we believe they would be an ideal candidate to both inhibit and then activate mRNA translation. Yoshimura *et al.* [18] previously showed that ^{CNV}K ODNs can crosslink to ORNs with single-nucleotide specificity, making them ideal for targeting RNA in a sequence-specific manner. They additionally demonstrated that ^{CNV}K ODNs could crosslink to full-length mRNA and repress their translation *in vitro* [44]. It was noted however that a relatively high concentration of ^{CNV}K ODN was required to crosslink the target mRNA, as the ^{CNV}K residue itself does not participate in hydrogen bonding between the ODN and mRNA.

Translation inhibitory ^{CNV}K ODNs were first applied *in vivo* by the Fujimoto group to knock down the expression of GFP in a GFP-stable HeLa cell line [45]. From their data, it appears that they struggled to accurately represent mRNA inhibition after crosslinking, as they did not employ a method to destroy existing protein or prevent translation of newly synthesized mRNA. However, they were able to demonstrate a 40% reduction in GFP expression 24 h after photo-crosslinking and show that the abundance of functional mRNA decreased by 60% immediately after irradiation. The Fujimoto group would later use this assay to demonstrate that ASOs containing ^{CNV}Ks were more effective than unmodified-ASOs and psoralen-ASOs at repressing translation [21]. Taken together, these results show that ^{CNV}K ODNs can be used to rapidly crosslink and inhibit the translation of mRNA, although the degree of silencing is influenced by the strength of the mRNA/ODN interaction.

4.1.7) A Carbazole Nucleoside Platform for mRNA Photo-Regulation

Current methods of photo-regulating mRNA translation are inadequate for the demands of research. Existing strategies of photo-regulation either suffer from ineffective silencing of mRNA, unnecessary restrictions on RNA design and workflows, or a level of complexity that makes them irrelevant for most research and therapeutic needs. This is especially the case for the photo-regulation of exogenous mRNA, as every method examined in this section produces significant amounts of leaky expression, which is disastrous for precisely controlling when and where an mRNA is expressed. This is especially relevant when controlling the activity of novel high-expression mRNA vectors, as even a small percentage of leaky expression will translate into a large protein signal.

The reversible photo-crosslinking ^{CNV}K nucleosides can form the basis of a photo-regulation system that improves upon these shortcomings. We envisage that a short blocker ^{CNV}K-oligonucleotide which is complementary to mRNAs of interest can be used to regulate their translation: mRNA can easily be "primed" with a ^{CNV}K-blocker by crosslinking with 365 nm light *in vitro*. By targeting the blocker to the open-reading frame (ORF), it will impede ribosome progression and prevent the translation of full-length peptides. The ^{CNV}K crosslink can be reversed with a pulse of 312 nm light after the mRNA has been introduced to cells, activating translation **(Figure 4.6)**.

A photo-crosslinking approach to translation regulation would have several advantages over existing methods. Most photo-caging and photo-isomerizing-based methods attempt to regulate the recruitment of ribosomes to mRNA but are unable to inhibit translation from ribosomes that do access the mRNA, and therefore suffer from leaky expression. By using ^{CNV}Ks to prevent ribosomes from reading through the full-length mRNA, our method should

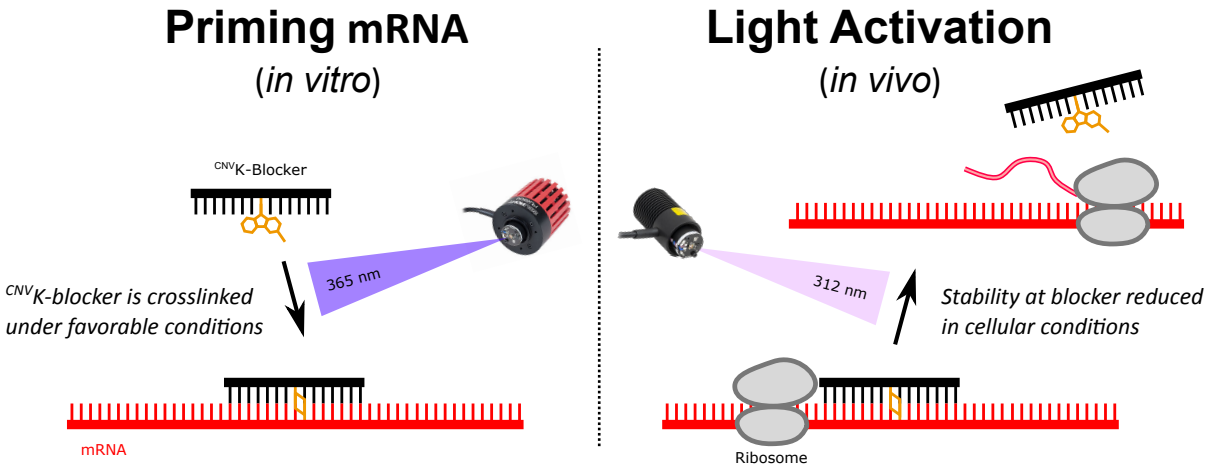


Figure 4.6: CNVK-blocker strategy for mRNA regulation. *In vitro*, a short CNVK-blocker oligonucleotide is crosslinked to a synthetic mRNA of interest with 365 nm light. Conditions during crosslinking will be such that hybridization and annealing between blocker and mRNA are favorable. After the primed mRNA is transfected in cells, 312 nm radiation can be applied in a targeted manner to selectively reverse the blocker/mRNA crosslink. The higher temperature and reduced blocker concentration *in vivo* will favor reverse crosslinking and blocker melting, allowing ribosome progression and translation. Thus, 312 nm light can be applied to activate translation.

provide a more robust silencing of gene expression, which is critical to limiting the effect of mRNA to desired targets. Our method will not require RNA tags or other sequence features that enzymatic-based approaches do [275, 277, 278], which means it can be implemented for any existing mRNA of interest and removes constraints on the design of new mRNAs. Likewise, because our method does not rely on cap modification, it allows greater flexibility in 5′ cap selection and optimization than cap-based photo-regulation strategies [280-284].

In addition to functioning independently of caps and recognition sequences, CNVK-blockers will also be compatible with a wider variety of mRNA topologies than existing methods. As discussed in Section 1.9.2, there is growing interest in circRNA vectors to improve the stability and expression of therapeutic mRNAs. circRNA is inherently incompatible with cap-based methods of regulation, and its synthesis can be impeded by highly-structured motifs [159], making tag-based labeling unideal. CNVK-blockers will provide a new platform to study and control circRNA.

Because ^{CNV}K-blockers will crosslink mRNA in a sequence-specific manner, they can also be used to regulate individual cistrons of a complex mRNA independent of their order. This makes our strategy a powerful method for regulating the expression of polycistronic and self-amplifying mRNA **(Figure 4.7)**.

Chemical modifications of oligonucleotides have greatly expanded the functional capacities and stability of DNA/RNA-based tools. However, many of these advances cannot be applied to mRNAs because their size precludes chemical synthesis, and current methods of enzymatically introducing modifications into mRNA are severely inadequate [201]. Because ^{CNV}K-blockers will be prepared through phosphoramidite synthesis, it would be trivial to modify them either during synthesis or post-synthetically. This can allow ^{CNV}K-blockers to act as a simple platform for attaching synthetic modifications onto *in vitro* transcribed mRNA

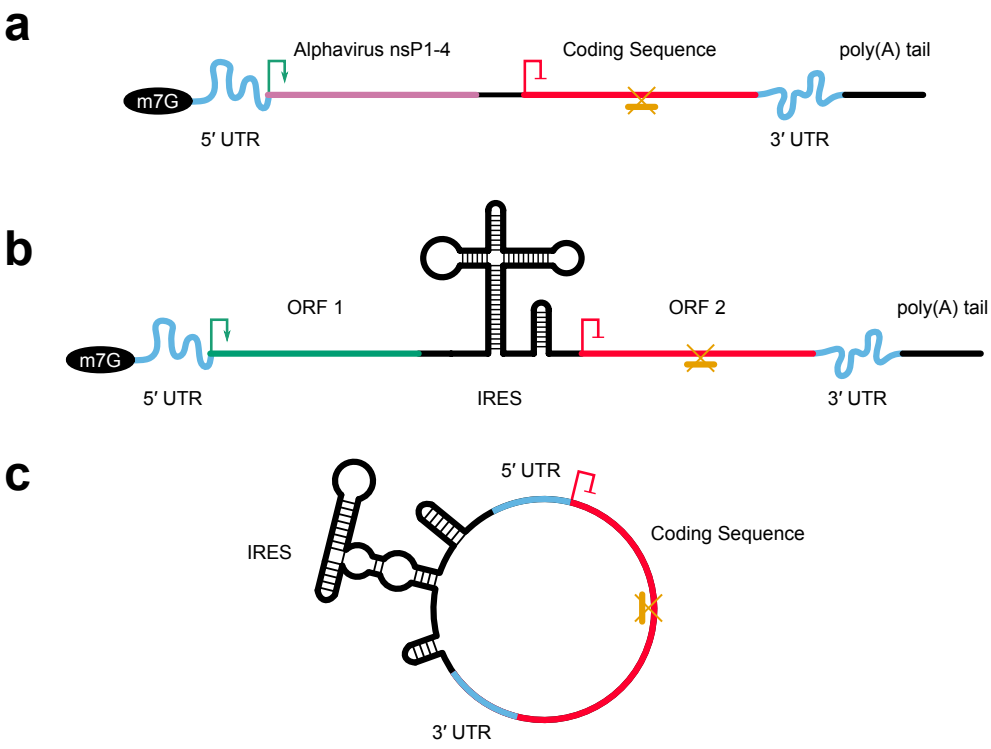


Figure 4.7: ^{CNV}K-blocker regulation of exotic mRNA species. ^{CNV}K-blockers (vermillion) can be applied to selectively silence ORFs of interest (red). This can be used to regulate **(a)** the translation of replicable coding sequences of saRNAs **(b)** individual cistrons in a polycistronic mRNA or **(c)** the entirety of a circRNA.

strands. Fluorophores, affinity tags for purification, and cell-specific uptake ligands **(Figure 4.8)** could all be tagged onto mRNA to improve its functional profile prior to photo-activation.

As a final note, regulating mRNA translation with ^{CNV}K-blockers is an incredibly accessible technology and easy to implement. One of our initial aims was to synthesize these blockers from commercially available monomers so that any other group could reasonably order similar oligonucleotides from a DNA synthesis company or acquire the material for themselves. Because the crosslinking reaction is light-catalyzed, our method does not require the purchase or synthesis of additional enzymes or chemical reagents, unlike other strategies. The crosslinking reaction can occur in physiologically compatible buffers, which removes the need for post-crosslinking purification prior to transfecting primed mRNA into cells. Photo-regulation with ^{CNV}K-blockers could therefore fit seamlessly into existing mRNA workflows with minimal effort.

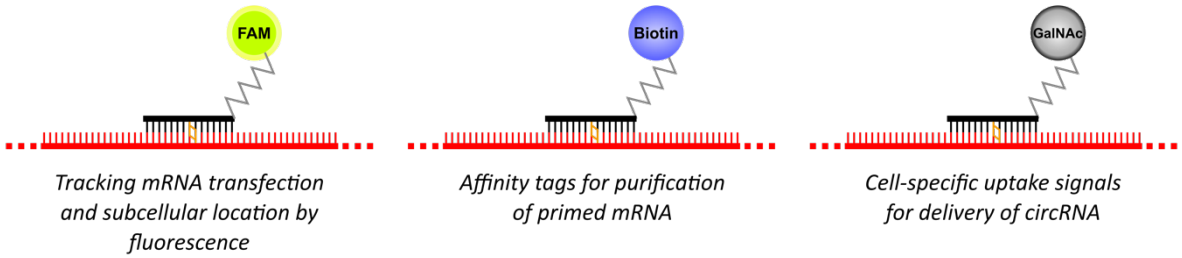


Figure 4.8: ^{CNV}K-blockers can be used to label mRNA with synthetic modifications. From left to right, the blocker is used to label mRNA with a fluorophore (FAM), an affinity tag for purification (Biotin), and a cell-specific uptake signal (GalNAc).

4.2 Objectives

In this chapter, the reversible photo-crosslinking of ^{CNV}Ks and other carbazole nucleosides will be investigated as the basis of a photo-regulated translation system. ^{CNV}K-oligonucleotides have previously been demonstrated to inhibit translation when crosslinked to mRNA, albeit to varied degrees of success. Our goal for this project is to prime *in vitro* transcribed mRNA with short ^{CNV}K "blockers" before transfecting them into cells so that translation will be inhibited until a light signal reverses the crosslink to activate translation. We envisage that ^{CNV}K blockers will be simpler to implement in mRNA workflows than existing photo-regulation methods and can also provide a platform to label mRNA with modifications. Our first aim is to determine whether ^{CNV}K-blockers can be used to totally silence translation from exogenous mRNAs. We will then attempt to initiate translation by photo-splitting the crosslink in cells. After characterizing and optimizing photo-regulation with ^{CNV}K-blockers, we will then apply blockers to mRNAs that have previously been incompatible with photo-regulation, such as circRNA and polycistronic mRNA.

4.3 ^{CNV}K-Induced mRNA Silencing

4.3.1) Preparation and Expression of IRES-mCherry mRNA

To assess whether ^{CNV}K-blockers can totally inhibit the translation of mRNA, we first needed to design an mRNA species whose translation could be measured with a simple assay. Several strategies exist to prepare *in vitro* transcribed mRNA and assess their translation with either *in vitro* or *in vivo* assays. I decided to use mCherry as a reporter for mRNA expression because it is well-established and does not require additional reagents to function, unlike luminescence-based assays. In place of a 7-methylguanylate cap to initiate translation, I decided instead to incorporate the encephalomyocarditis virus Internal Ribosome Entry Site (IRES)[290] upstream of the mCherry gene to initiate translation. The IRES sequence was

chosen because **1.)** it increased the complexity of the mRNA sequence, making it a better test of the Watson-Crick interaction driven ^{CNV}K-blocker strategy **2.)** can be used to prepare polycistronic and circular mRNA **3.)** simplified the mRNA preparation workflow and **4.)** was a cheaper alternative to commercially available capping systems.

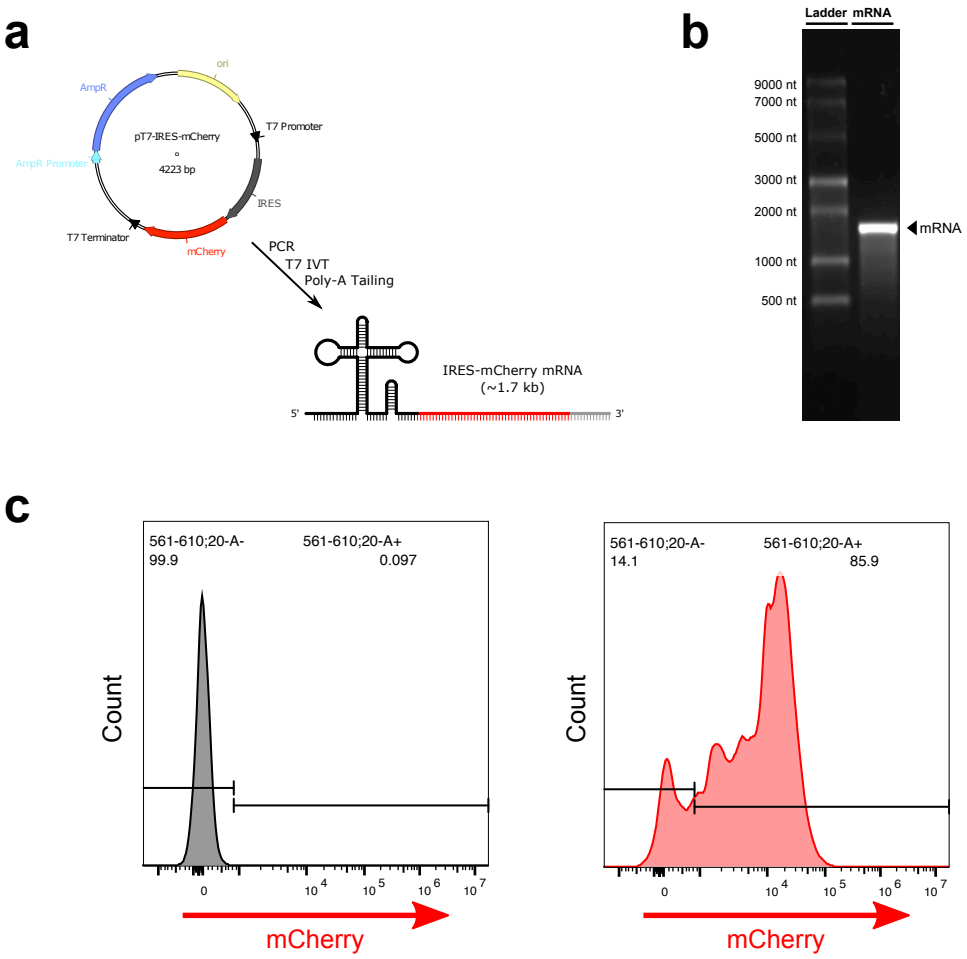


Figure 4.9a: Schematic of workflow for producing IRES-mCherry mRNA. pRSET-IRES-mCherry plasmid is PCR amplified with primers $pRSET_F$ and $pRSET_R$. The resulting PCR product is used to template an overnight T7 transcription reaction. mRNA product is purified from the reactions and is incubated with T4 polynucleotide kinase to add a ~100 nt long poly-A tail. A simplified diagram of the final mRNA is displayed. **Figure 4.9b:** Analysis of pre-tailed mRNA by agarose gel. A single band is observed on the gel from the mRNA transcription reaction. The expected size of the mRNA is 1646 nt, which is confirmed by the position of the band. Gel visualized by SYBR-Gold fluorescence. **Figure 4.9c:** HEK293 mCherry fluorescence in a control population (left, black) and IRES-mCherry transfected cells (right, red). A threshold for mCherry+ cells was set using the control population. Over 85% of transfected cells were positive for mCherry expression. Histograms produced with FlowJo.

A plasmid template for IRES-mCherry mRNA was constructed in the pRSET vector as described in Section 6.6. The plasmid was PCR amplified with two primers we designed for the pRSET vector (Appendix 4) to prepare a template for *in vitro* transcription, which was then used to generate full-length mRNA as described in Section 6.7 **(Figure 4.9a)**. The final mRNA was 1646 nt long **(Figure 4.9b)** with a poly-A tail of at least 100 nt added post-transcriptionally. To confirm that the mRNA was functional, it was transfected into HEK293 cells by lipofection, which were then observed by flow cytometry 24 h after transfection **(Figure 4.9c)**. The transfected population showed a marked increase in mCherry fluorescence which was easily distinguished from the control population, with >85% of transfected cells positive for mCherry expression.

4.3.2) Design of 1st Generation Translation Blockers

After preparing a functional mRNA, we next began to design the first generation of ^{CNV}K translational blockers to regulate the translation of mCherry. Some key considerations we had for the design of blockers were that **1.)** the blocker should inhibit translation only while it is crosslinked. Therefore, the interaction between blocker and mRNA should not be so stable that it sterically inhibits translation without being crosslinked **2.)** the blocker targets the open-reading frame (ORF) of the mRNA so that full-length peptides cannot be synthesized. Additionally, the ribosome should not be impeded from binding to the mRNA and beginning translation, so that its momentum can help displace non-crosslinked blockers, and **3.)** the blocker should avoid activating siRNA/ASO mRNA degradation mechanisms, and should produce a limited immunological response upon transfection. For these reasons, we designed a 15 nt blocker which was complementary to the stretch of sequence 70-84 nt bases from the start codon of the mCherry gene. The blocker was synthesized both as a DNA (B1.1) and 2′OMe RNA (B1.2) sequence (Appendix 3). While B1.1 was likely to trigger RNAse H

degradation if used *in vivo*, we wanted to use it to compare the effects that different oligonucleotide species would have on blocker/mRNA interactions.

To confirm that the blockers were functional, we prepared a 3′ FAM labeled DNA which was isosequential to residues 60-99 of the mCherry gene (FAM-ODN$_{60\text{-}99}$), which encompassed the target of the blockers. FAM-ODN$_{60\text{-}99}$ was mixed with equimolar amounts of B1.1 and B1.2, which were then annealed and crosslinked. The crosslinking reactions were analyzed by denaturing PAGE alongside non-crosslinked controls **(Figure 4.10a)**. No change in the gel migration of the FAM-ODN$_{60\text{-}99}$ is observed when irradiated with 365 nm light on its own, as expected. Both blockers were able to cause a shift in FAM-ODN$_{60\text{-}99}$ gel migration after 365 nm irradiation, indicating that both were functional crosslinkers. The difference observed in the migration of the crosslink products is likely due to the greater size and anionic charge of B1.2 compared to B1.1.

After determining that both ^{CNV}K blockers can crosslink to their target sequence, we next wanted to confirm that crosslinking still occurred with the more complex and structured mRNA target. We therefore prepared 5′ FAM-labelled versions of both blockers (B1.1F, B1.2F) to make it easier to observe whether they crosslinked to the significantly larger mRNA. Increasing amounts of the labeled blockers were annealed and crosslinked to 0.75 pmol of IRES-mCherry mRNA, which was then analyzed by agarose gel electrophoresis and visualized by FAM fluorescence **(Figure 4.10b)**. Blocker B1.2F behaved similarly to the previous experiment and appears to have saturated the mRNA target at relatively low concentrations. In contrast, Blocker B1.1F did not crosslink to the mRNA target at any concentration. Given that 2′OMe RNA/RNA hybridization is more favorable than DNA/RNA interactions, we did expect to observe a difference in the crosslinking rates of the

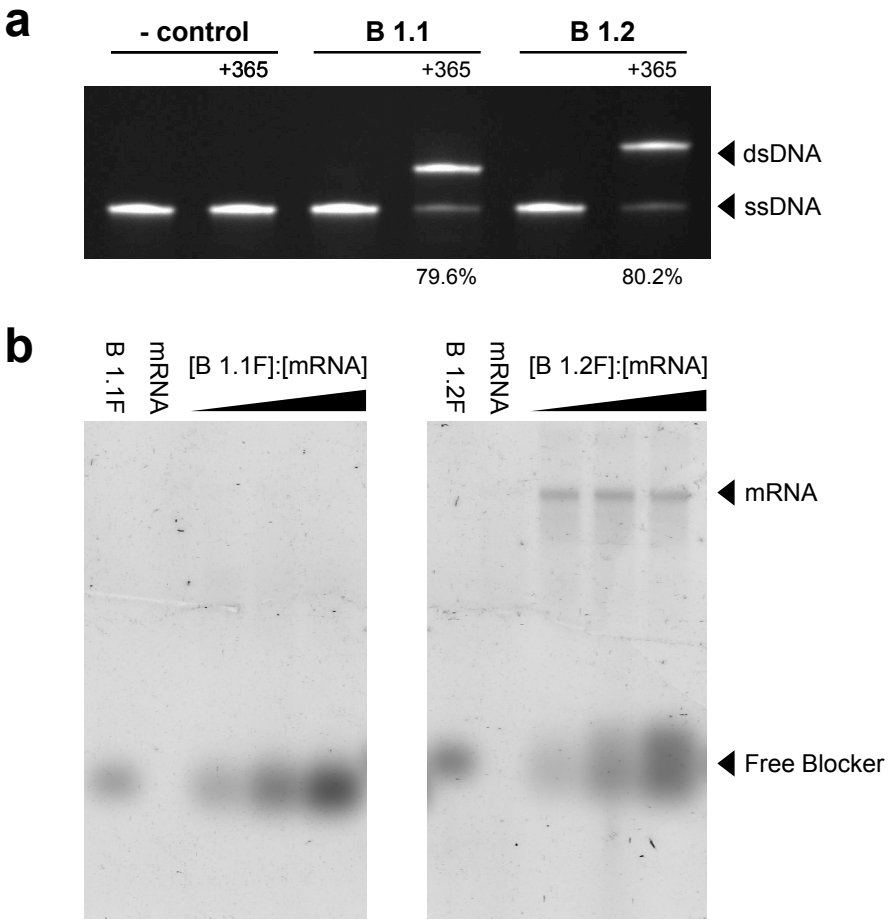


Figure 4.10a: Denaturing-PAGE of crosslinking reaction, imaged by FAM-fluorescence. Titles above the gel indicate whether samples were prepared with the FAM-$ODN_{60\text{-}99}$ by itself (- control) or with B1.1 and B1.2. Samples were treated with 365 nm to catalyze ^{CNV}K crosslinking as indicated. Arrows to the right of the gel highlight ssDNA FAM-$ODN_{60\text{-}99}$, and crosslinked (dsDNA) FAM-$ODN_{60\text{-}99}$. Crosslinking efficiency as calculated by Fiji analysis is displayed beneath relevant lanes. **Figure 4.10b:** Agarose gel of blocker/IRES-mCherry mRNA crosslinking reactions, imaged by FAM-fluorescence. 0.75 pmol mRNA was crosslinked with x2, x5 and x10 excess of B1.1F (left) and B1.2F (right), as indicated by the increasing bars above the gels. 0.75 pmol of B1.1F, B1.2F, and mRNA were run alongside the crosslinking reactions as a negative control. Arrows to the right of the gel indicate the now fluorescently labeled mRNA, or free excess blocker.

two blockers; however, it appears that this length of ^{CNV}K-DNA blocker is unable to crosslink to mRNA at all. In addition to the reported destabilizing effect ^{CNV}K residues have on duplex T_M, B1.1 is likely unable to outcompete mRNA secondary structures that could be forming near the target sequence. For this reason, and because of the increased intracellular stability they offer, we decided to use 2′OMe RNA for each ^{CNV}K-blocker moving forward.

4.3.3) Short 2′OMe ^{CNV}K-Blocker Can Inhibit Translation in Cells

After confirming that B1.2F can crosslink to IRES-mCherry mRNA, we next wanted to determine how effective B1.2F would be at silencing the expression of the mRNA. To accomplish this, two pools of IRES-mCherry mRNA were prepared: one with a 5x molar excess of Blocker 1.2F, and one without. Both pools were annealed and then irradiated with 365 nm light for 30 s. Aliquots were taken after both steps, which were transfected into HEK293 cells. Cells were analyzed by flow cytometry 24 h after transfection. Using a mock-transfected population as a negative control, cells were plotted based on their mCherry and FAM fluorescence **(Figure 4.11)**.

In each population that was transfected with unblocked mRNA, approximately 82% of cells were positive for mCherry expression, which indicates that transfection rates were high and

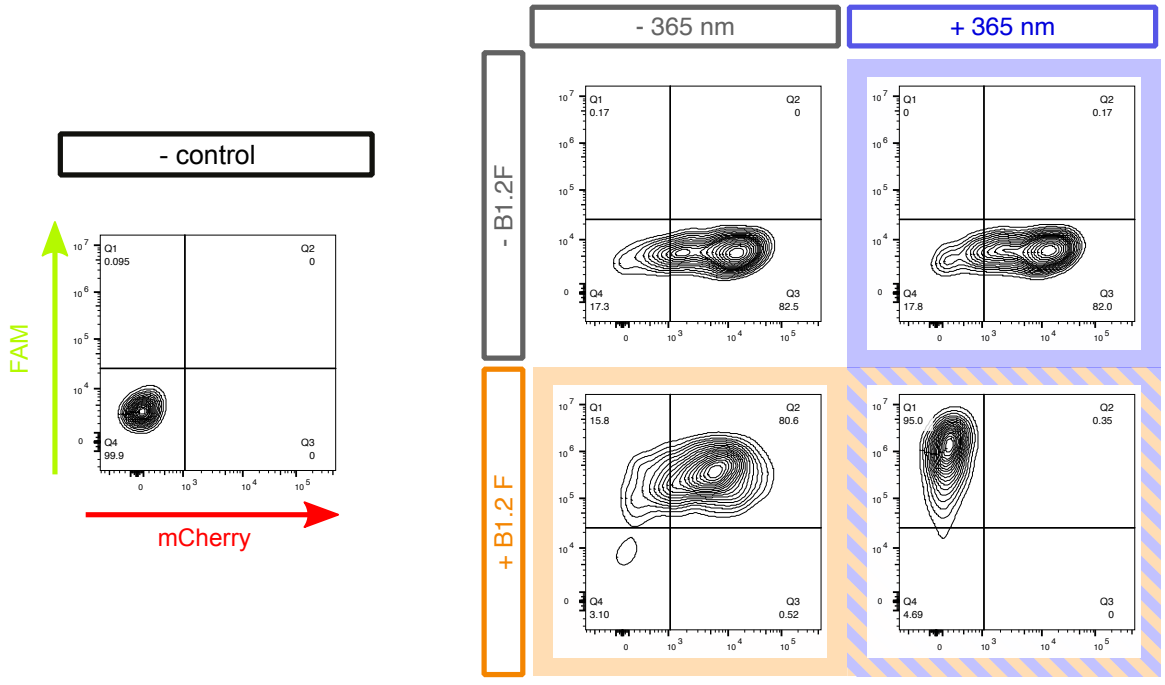
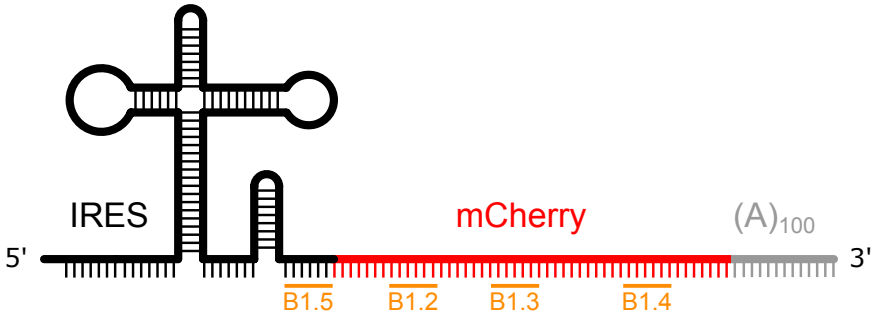


Figure 4.11: Flow-cytometry analysis of HEK293 cells transfected with IRES-mCherry mRNA and B1.2F. The x and y-axes of each plot correspond to mCherry and FAM fluorescence respectively. Each plot is split into quadrants to quantify mCherry +/- and FAM +/- as the % of total population. Thresholds for mCherry + and FAM + were set using a negative control population. The text boxes above and to the left of the flow-plots indicate which conditions IRES-mCherry mRNA were treated with before transfection (B1.2F +/-, 365 nm irradiation +/-). A mock-transfected population (- control) is shown to the left.

consistent between these populations. We observed no difference in mCherry expression between the B1.2F-free transfected populations, regardless of UV irradiation (-blocker/-UV, -blocker/+UV). This finding shows that the conditions used to crosslink ^{CNV}K blockers do not significantly damage or impair the translation of mRNA, and that any translation repression we observe will be caused by the blocker. The combination of B1.2F and 365 nm irradiation (+blocker/+UV) almost entirely eliminated the mCherry positive population in transfected cells, reducing it to 0.35% of the total population. This is a much more robust silencing of expression than had previously been reported with inhibitory ^{CNV}K-oligonucleotides and demonstrated to us that B1.2F could be used to produce an translational off-state. Importantly, the majority of cells in the +blocker/-UV transfected populations are both FAM and mCherry positive (80.6%), which indicates that mRNA expression is occurring in cells that also contain B1.2F. This demonstrates that the nature of translation inhibition is dependent on ^{CNV}K crosslinking, as opposed to blocker B1.2F repressing mRNA translation through a steric or siRNA effect. This result not only confirms the essential role of the ^{CNV}K residue in our blocker design; it also supports our hypothesis that translation of the mRNA can be initiated once the crosslink is reversed.

4.3.4) ^{CNV}K-Blockers Inhibit Translation Independent of ORF Target

Blocker 1.2 was extremely effective at inhibiting translation in a crosslink-dependent manner. We next wanted to determine if this was an effect of the particular target sequence we selected, or if our method could be expanded to other target sites on the mRNA. We therefore designed a set of 3 additional 2′OMe RNA ^{CNV}K blockers: two of the blockers (B1.3 and B1.4) targeted sequences within the mCherry ORF at varying distances from the mCherry start codon, and one (B1.5) targeted a sequence -16 to -30 nt upstream of the start codon **(Figure 4.12)**. Each new blocker was designed to have a similar T_M and length to



		Start	Stop
B1.2	GUUCACGGA ^{CNV}K CCCUC	70	84
B1.3	GGGGA ^{CNV}K GUCGGCGGG	241	255
B1.4	CUGCUUGA ^{CNV}K CUCGCCCUUC	486	504
B1.5	UGUGGCCA ^{CNV}K AUUAUC	-16	-30

Figure 4.12: A diagram showing the relative position of each 1st generation blocker on IRES-mCherry mRNA, with the sequence of each blocker below the diagram. The start and stop position of each blocker's target sequence is listed relative to the mCherry start codon.

B1.2. Because B1.3 and B1.4 target sequences more distal in mCherry ORF than B1.2, it is possible that the synthesis of some long peptides might occur from mRNA primed with these blockers. However, because all 11 β-strands of mCherry are required to generate fluorescence [291], we did not believe that these longer peptides would contribute to fluorescent signal.

IRES-mCherry mRNA was crosslinked with a x10 molar excess of each blocker and was then transfected into HEK293 cells. These were observed by flow cytometry after 24 h as described in the previous section. The average fluorescence of each population was normalized to a positive control, as shown in **Figure 4.13**. Impressively, B1.2, B1.3, and B1.4 were all able to repress mCherry translation by greater than 99% compared to the

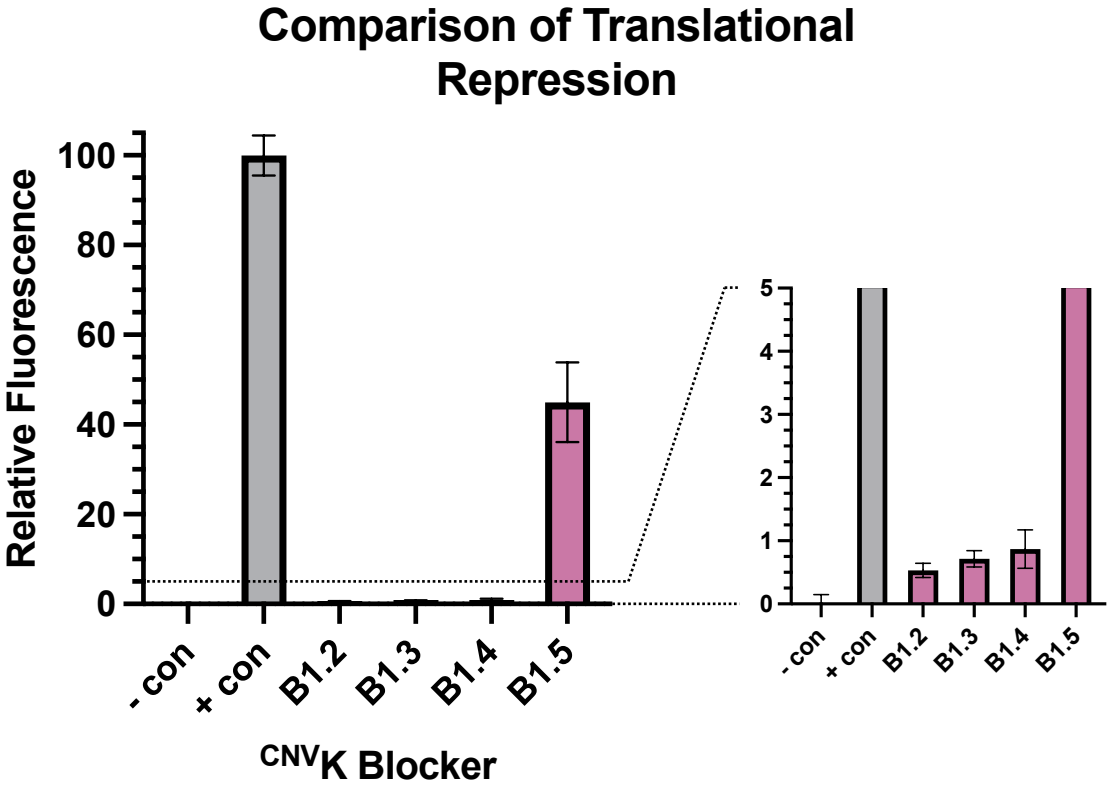


Figure 4.13: Relative mCherry fluorescence from HEK293 cells transfected with IRES-mCherry mRNA primed with 1st generation blockers. The background fluorescence was removed by subtracting the average fluorescence of mock-transfected cells from each population with propagation of error. The average fluorescence was then normalized to the positive control positive with propagation of error, which was set to 100%. A zoomed-in plot is shown in the upper-right corner. Values reflect the average fluorescence as determined by flow-cytometry of n=3 transfection replicates, error bars = SD.

positive control. This not only demonstrates a robust off-state but suggests that our strategy can be expanded to other sequences.

B1.5 was an outlier however and was only able to repress translation by 55.0%. This result was not unexpected however, as B1.5 targets a stretch of sequence outside of the ORF. It is possible that B1.5 can disrupt ribosomes from accessing the start codon, but it cannot block the translation of full-length peptides like B1.2-4 can. However, it is equally possible that B1.5's weak effect on translation could be the result of poor crosslinking rates between B1.5 and the mRNA. To determine which of these was correct, we prepared FAM-labeled B1.5F so that we could visualize by gel whether the blocker could crosslink to IRES-mCherry

mRNA. B1.5F appears to have saturated the target mRNA and showed a crosslinking profile similar to B1.2F **(Figure 4.14)**. These results indicate that 2′OMe RNA ^{CNV}K blockers can target a wide variety of sequences, but the most effective repression of translation will be generated from those which block the ORF.

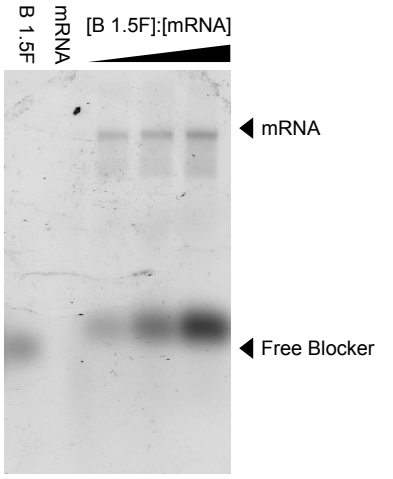


Figure 4.14: Agarose gel of blocker B1.5F/IRES-mCherry mRNA crosslinking reactions, imaged by FAM-fluorescence. 0.75 pmol mRNA was crosslinked with x2, x5 and x10 excess of B1.5F as indicated by the increasing bars above the gels. 0.75 pmol of B1.5F and mRNA were run alongside the crosslinking reactions as a negative control. Arrows to the right of the gel indicate the now fluorescently labeled mRNA, or free excess blocker.

4.3.5) Comparison of Photo-Regulation Methods

The results of this section demonstrated that ^{CNV}K-blockers could be designed to almost entirely silence the translation of mRNA. We lastly wanted to compare the rate of gene silencing we achieved with our method compared to some benchmarks in the literature.

When we began to compare our silencing results to those published in photo-caged cap methods, we noticed that many publications normalized their data to questionable negative controls. For example, the photo-switchable caps published by Shinzo Ogasawara frequently use cells transfected with non-capped but functional mRNA as a negative control [280-282], as opposed to mock-transfection, scrambled mRNA, or non-functional mRNA. Although 5′ caps are required for the effective translation and stability of mRNA, some expression of

mRNA can occur through cap-independent translation [292]. To demonstrate the impact this could have on data presentation, I prepared uncapped mCherry mRNA from a non-IRES vector to transfect into HEK293 cells. 24 h after transfection, cells were analyzed by flow cytometry and compared to mock-transfected cells **(Figure 4.15)**. While the uncapped mRNA caused only a moderate increase in fluorescence, it is notable that once a poly-A tail was added to the mRNA, a nearly 7-fold increase in fluorescence was observed. This demonstrates that even a slight increase in the stability of mRNA can lead to a large degree of cap-independent translation. While it is not clear from their papers whether a tailed or untailed mRNA was used, either transfection condition would have increased the

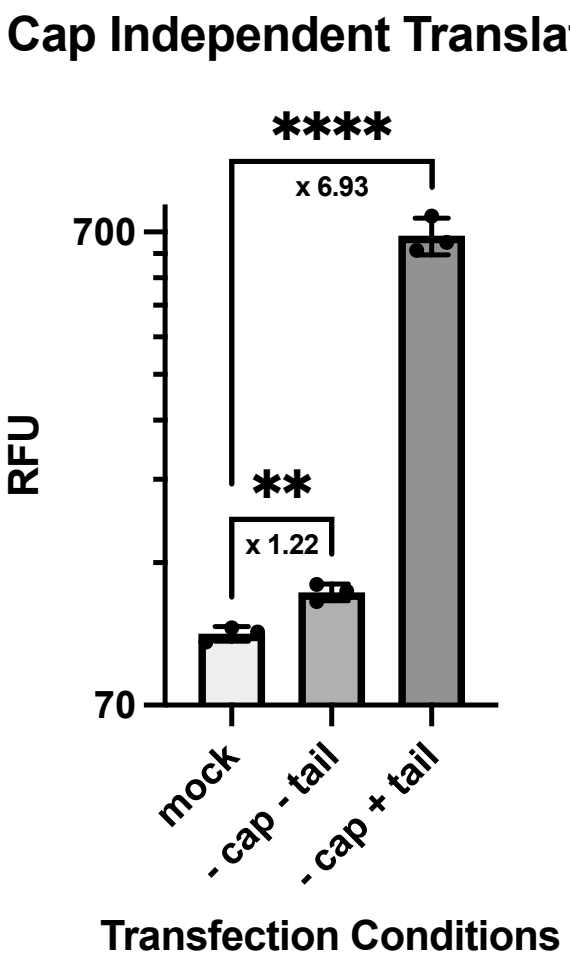


Figure 4.15: Relative mCherry fluorescence from HEK293 cells transfected with uncapped IRES-mCherry mRNA. Bars represent the average, non-normalized fluorescence of n=3 transfection replicates as determined by flow-cytometry. Brackets between bars indicate whether those values are significantly different as determined by unpaired t-test (** < 0.05, ** < 0.0001). Error bars = SD.

fluorescence of their negative control and made their "silencing" appear more robust in comparison.

More egregious than this however, the Rentmeister group [283, 284] uses an ApppG-capped mRNA as a negative control for their photo-caged caps. ApppG does not interact as strongly with eIF4E compared to cap-0 or cap-1, but it does provide a stabilizing effect to mRNA and has previously been shown to provide 12 ±3% of the level of translation of cap-0 mRNA [293]. Their choice of "negative" control and single-cell data which demonstrates high levels of leaky expression (Section 4.1.5) suggests that their photo-caged caps are not very effective at inhibiting translation.

The results of our experiment with blockers B1.2-5 suggested that a more effective translational repression can be achieved through blocking ribosome progression through the ORF than by inhibiting the recruitment of ribosomes to the start codon. We similarly hypothesize that cap-based methods of photo-regulation will always struggle to fully inhibit translation because not all translation occurs in a cap-dependent manner. Although we do not have a direct comparison between our silencing strategies and others, published data indicate that our method may produce the strongest translational off-state out of current activation standards.

4.4 Translation Activation by ^{CNV}K Photo-Splitting

4.4.1) Photo-Regulation of mRNA Translation by B1.2

The results of the previous section established that short ^{CNV}K blockers provide robust silencing of mRNA translation. Using a strategy of annealing and crosslinking ^{CNV}K

blockers, mRNA could be primed in as little as 30 min and transfected into cells without a purification step. Our data suggested that translation could be initiated by removing the ^{CNV}K blocker, which we now wanted to confirm by photo-splitting the blocker with 312 nm light.

To test whether translation could be resumed in a light-dependent manner, we began by photo-splitting primed mRNA *in vitro* before attempting it in cells. To this end, IRES-mCherry mRNA was pooled with a 10x molar excess of B1.2 and then annealed **(1)**. After annealing, the blocker/mRNA mixture was crosslinked with 30 s 365 nm light at 4 °C **(2)**, and was then irradiated with 60 s 310 nm light at 37 °C to reverse the crosslink **(3)**. Aliquots were taken after steps 1, 2, and 3 to track the photo-regulation of the mRNA, as shown in **Figure 4.16a**.

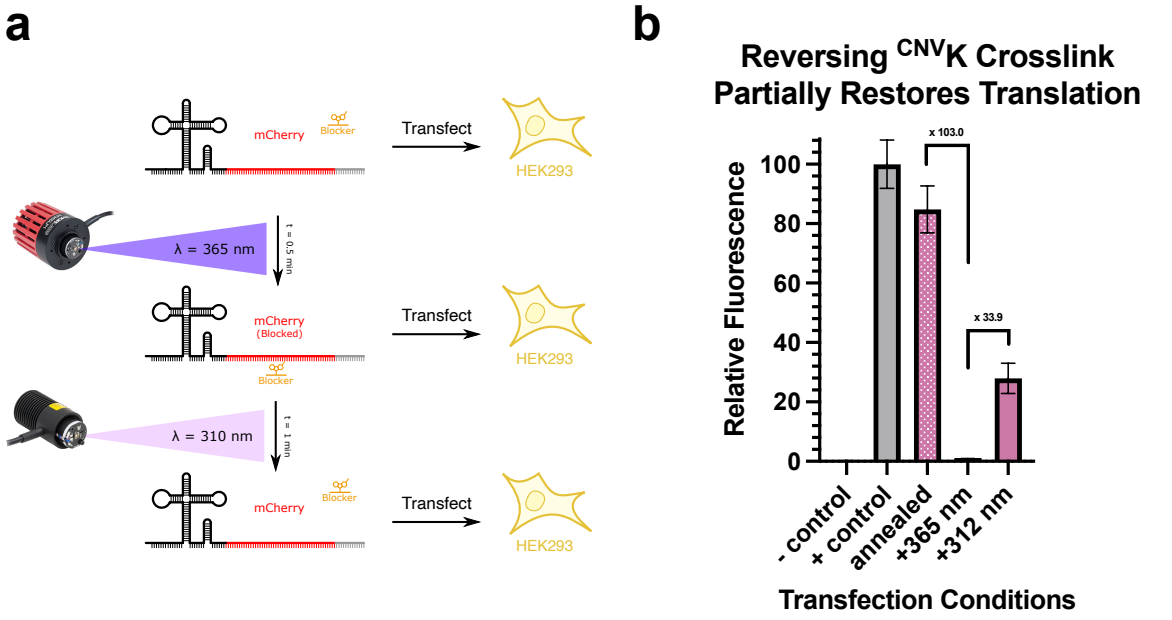


Figure 4.16a: Schematic of *in vitro* photo-regulation of IRES-mCherry mRNA. mRNA is annealed with B1.2, primed by irradiation with 365 nm light, and then activated by irradiation with 310 nm light. At each step, an aliquot is taken from the reaction to be lipofected into HEK293 cells. **Figure 4.16b:** Relative mCherry fluorescence from HEK293 cells transfected with IRES-mCherry mRNA photo-regulated with B1.2. The background fluorescence was removed by subtracting the average fluorescence of mock-transfected cells from each population with propagation of error. The average fluorescence was then normalized to the positive control positive with propagation of error, which was set to 100%. Brackets indicate the fold-change in relative fluorescence between samples. Values reflect the average fluorescence as determined by flow-cytometry of n=3 transfection replicates, error bars = SD.

The aliquoted samples were transfected into HEK293 cells by lipofection along with an untreated IRES-mCherry positive control. The normalized average fluorescence of each transfected population is shown in **Figure 4.16b**, as determined by flow cytometry. As previously shown, priming IRES-mCherry mRNA with blocker B1.2 represses translation by greater than 99%, representing a 103.0-fold drop in fluorescence. Excitingly, irradiating the primed mRNA with 312 nm caused a 33.9-fold increase in fluorescence, providing the first evidence that the reversible nature of ^{CNV}K crosslinking can be applied to optically regulate translation. The recovered fluorescence was disappointingly only 27.8% as intense as the positive control, however, which is less than we were aiming to achieve. Because it was reported that the ^{CNV}K can be photo-split to completion within 60 s of 312 nm irradiation without forming byproducts or damaging the crosslinked base [17], we had anticipated higher rates of gene activation.

This experiment was repeated, but we consistently observed mRNA activation rates between 20-30%. We considered that lower-than-expected activation rates could be caused partially by UV-induced damage of the mRNA. However, when we compared the translation rates of our positive control and mRNA irradiated with 60 s 310 nm light, we did not observe a notable difference in cell fluorescence.

4.4.2) Attempts to Improve ^{CNV}K Photo-Splitting Rate

Because we did not find that our photo-regulation conditions were damaging the IRES-mCherry mRNA, we suspected that the poor activation rates we observed might be due to incomplete photo-splitting of B1.2 from the mRNA. To quickly test this idea, we prepared a FAM-labelled DNA (comp-ODN^{FAM}) which was complementary to the short ^{CNV}K-ODN which was previously used in Section 2.3.1. The two oligonucleotides were annealed,

crosslinked with 30 s 365 nm light, and then photo-split with 60 s 312 nm light similar to how the mRNA was handled in the section above. The resulting mixtures were analyzed by denaturing PAGE **(Figure 4.17a)** and HPLC-MS **(Figure 4.17b)** to determine their photo-splitting rates and whether unexpected byproducts were forming during the reaction.

Analysis of the crosslinking reactions by PAGE **(Figure 4.17a)** showed that almost all comp-ODN^{FAM} was crosslinked by 365 nm light. 60 s of 312 nm irradiation only photo-split approximately 20% of comp-ODN^{FAM}, leaving the majority crosslinked. This was consistent with the rate of mRNA activation we observed in cells, in which 365 nm light could almost entirely silence expression, but 312 nm irradiation only restored 28% of expression.

By MS, we were also able to observe the formation of the photo-crosslink product after 365 nm irradiation, which almost entirely consumed ^{CNV}K-ODN. After irradiating the sample with 310 nm irradiation, the fraction corresponding to ^{CNV}K-ODN was restored and the intensity of the photo-crosslink product decreased. As with PAGE analysis, we observed that the majority of the crosslink product was still present after 310 nm irradiation. We did not identify aberrant masses in the peaks corresponding to either starting product after photo-splitting, which is consistent with previous reports that reversing ^{CNV}K crosslinks regenerates the target base without causing damage [17].

Taken together, these results suggest that our translation activation rates were poor because the photo-splitting reaction did not proceed to completion, as opposed to UV-induced DNA and mRNA damage. Therefore, we performed a time-course photo-splitting assay to determine how long the reaction took to reach completion under conditions which would be

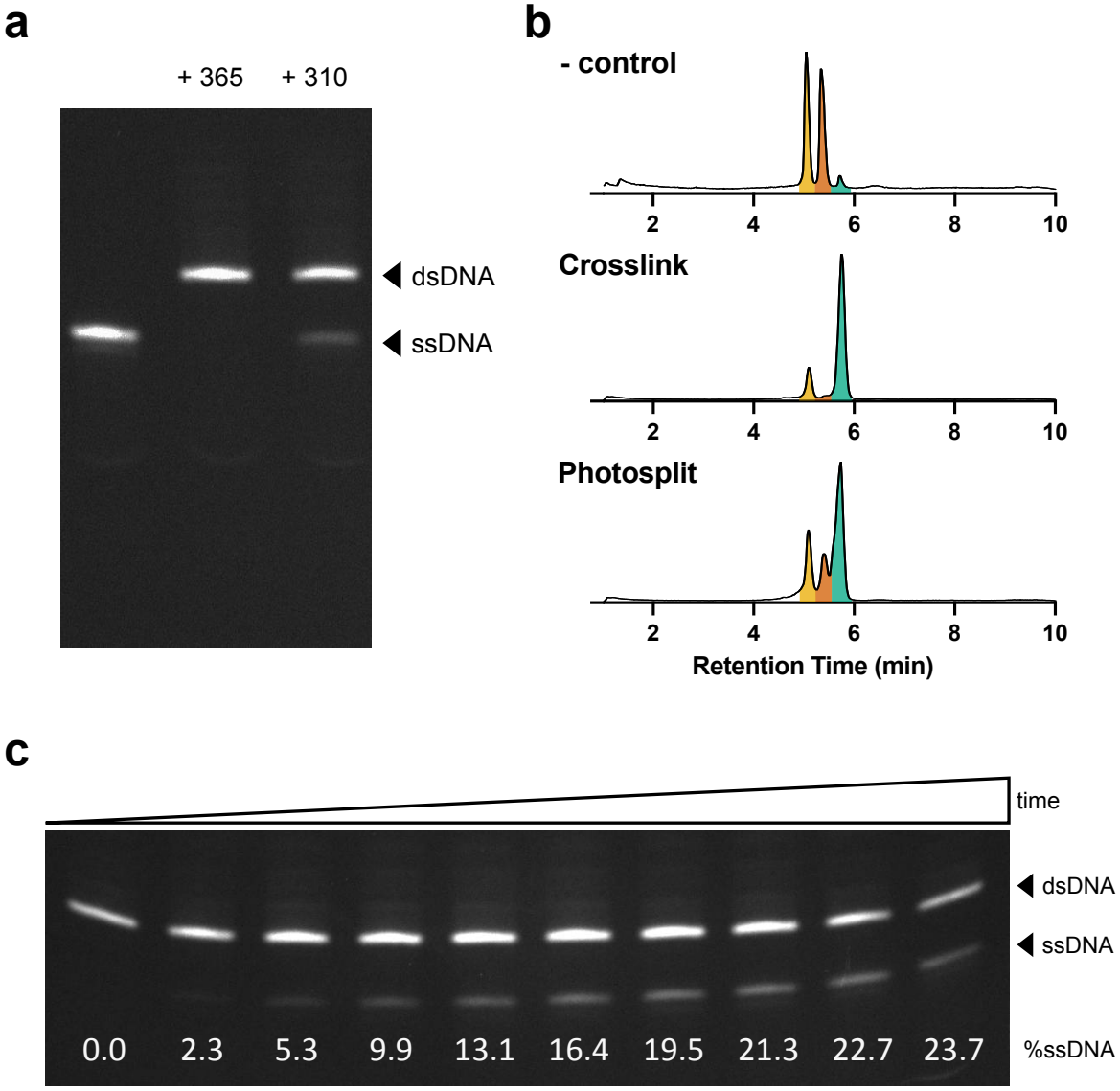


Figure 4.17a: Denaturing PAGE analysis of ^{CNV}K-ODN and comp-ODN^{FAM} crosslinking reactions. Oligonucleotide samples were annealed, crosslinked with 365 nm light, and photo-split with 310 nm light as indicated above each lane of the gel. Arrows to the right of the gel highlight ssDNA and crosslinked (dsDNA) comp-ODN^{FAM}. Gel visualized by FAM fluorescence. **Figure 4.17b:** HPLC-MS analysis of photo-crosslinking and photo-splitting reactions. Peaks corresponding to comp-ODN^{FAM} (yellow), ^{CNV}K-ODN (vermillion), and the photo-crosslink product (green) are highlighted. **Figure 4.17c:** Denaturing PAGE analysis of ^{CNV}K photo-splitting time course. Sample was irradiated with 310 nm light for 0, 5, 10, 20, 30, 60, 120, 180, 240, 300 s, as indicated by the increasing bar above the graph. The amount of ssDNA was calculated and written below each lane as the % of total fluorescence. Arrows to the right of the gel highlight ssDNA comp-ODN^{FAM}, and crosslinked (dsDNA) comp-ODN^{FAM}. Gel visualized by FAM fluorescence.

relevant for cell culture. Equimolar quantities of ^{CNV}K-ODN and comp-ODN^{FAM} were mixed, crosslinked, and then irradiated with 310 nm light for up to 5 min. Aliquots were taken at various time points and analyzed by denaturing PAGE **(Figure 4.17c)**. After 5 min

of irradiation, we were only able to achieve 23.7% photo-splitting of the crosslink product. Additionally, the rate of photo-splitting appeared to decrease over time. Even if the photo-splitting reaction could be driven to completion with increased time, the duration of irradiation would not be suitable for light-regulation *in vivo*.

We next considered that our UV source could be the reason why the reaction was not proceeding to completion. The LED we used in these experiments had a nominal wavelength and maximum intensity at 310 nm, whereas the reverse crosslinking reaction was reported to occur at 312 nm. Additionally, UV light sources vary in terms of the range of wavelengths they can produce: Our LED produces between 300 nm to 325 nm light before the intensity fell below 10% of the maximum output. The light source used in the original ^{CNV}K publication [17] might have had a broader range and covered wavelengths that also catalyzed the photo-splitting reaction. To test this possibility, we attempted to photo-split the ^{CNV}K-ODN/comp-ODN^{FAM} crosslink product using two additional UV sources: a gel-imaging transilluminator and a handheld UV lamp for TLC detection. Neither light source improved the rate of photo-splitting compared to the LED **(Figure 4.18a)**, and the gel imager even seemed to cause the formation of extra bands in the gel. Since there was no improvement in photo-splitting, we decided to continue using our original light source.

Even though we had examined the reverse crosslinking reaction by MS, we also thought it could be possible that the ^{CNV}K was being damaged by ROS generated by UV irradiation. As the reaction progressed, more ROS could be generated, which might explain why the rate of reverse crosslinking gradually decreases. To test if this was the case, we repeated the photo splitting of ^{CNV}K-DNA/comp-ODN^{FAM} with the inclusion of 10 mM ascorbic acid to reduce

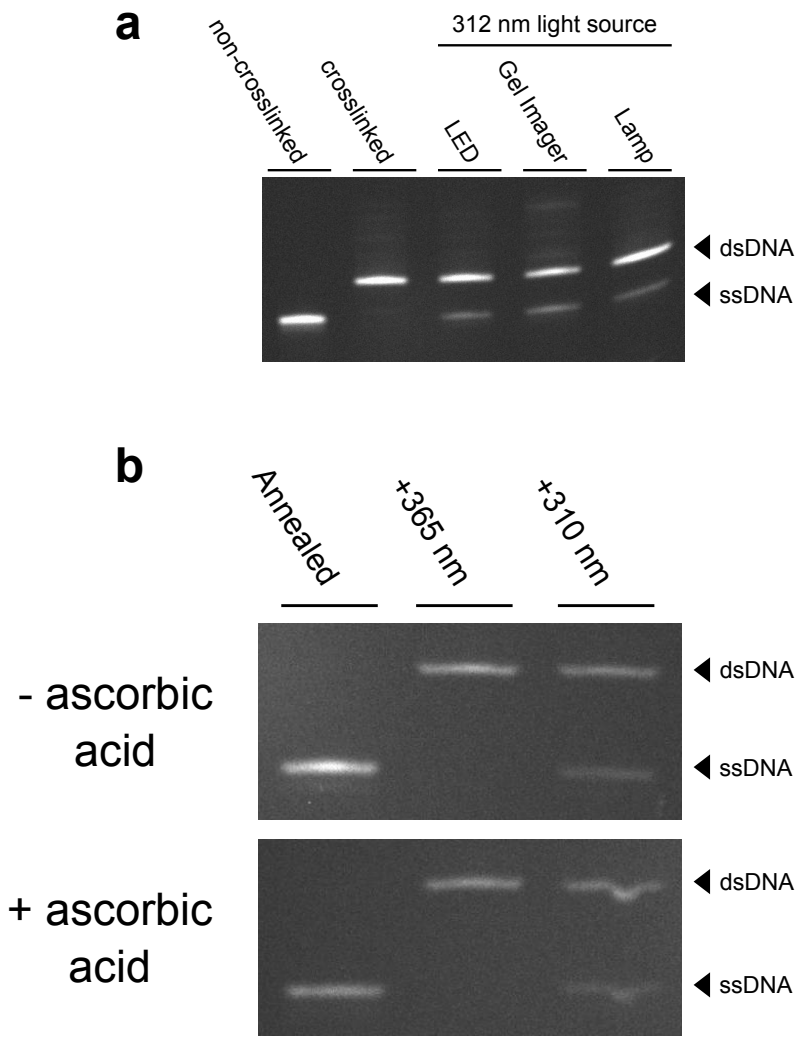


Figure 4.18a: Denaturing PAGE analysis of ^{CNV}K photo-splitting with alternative light-sources. A crosslinked ^{CNV}K-ODN/comp-ODN^{FAM} duplex was irradiated for 1 min with a 310 nm LED, gel imager transilluminator, or handheld UV lamp fitted with a 312 nm light bulb. The light source used for each sample is listed above the gel, in addition to the - control and non-crosslinked control. Arrows to the right of the gel highlight ssDNA comp-ODN^{FAM}, and crosslinked (dsDNA) comp-ODN^{FAM}. Gel visualized by FAM fluorescence. **Figure 4.18b:** Denaturing PAGE analysis of ^{CNV}K photo-splitting supplement with ascorbic acid. A crosslinked ^{CNV}K-ODN/comp-ODN^{FAM} duplex was irradiated for 1 min with a 310 nm LED light source with or without the addition of 10 mM ascorbic acid to the reaction, as listed next to the gels. Arrows to the right of the gel highlight ssDNA comp-ODN^{FAM}, and crosslinked (dsDNA) comp-ODN^{FAM}. Gel visualized by FAM fluorescence.

any ROS that could be formed. After reverse crosslinking, the oligonucleotides were analyzed by PAGE and compared to a control reaction **(Figure 4.18b)**. Again, no difference was observed reverse crosslinking rate.

4.4.3) ^{CNV}K Photo-Splitting Rate Varies by Duplex Stability

While these experiments with ^{CNV}K-ODN/comp-ODN^{FAM} did not identify why the photo-splitting reaction appears to stall, it did help us understand the cause of our poor mRNA photo-activation rates. The original ^{CNV}K publication did not describe the conditions used in the photo-splitting reaction, other than their light source and solvent [17]. However, in a more recent publication describing the synthesis of the ^{MEP}K [49], we noticed that the authors report, apparently erroneously, as having tested the photo splitting of ^{MEP}K at both 25 °C and 60 °C in different parts of their paper. Because of this, we decided to test whether denaturing conditions could improve the ^{CNV}K reverse crosslinking rate.

As earlier, a pool of ^{CNV}K-ODN and comp-ODN^{FAM} was crosslinked and separated into aliquots. The crosslinked aliquots were irradiated with 310 nm light for 3 min at 37 or 65 °C, in either 100% H_2O or 50% H_2O 50% formamide (v/v). Analysis by PAGE indicated that 22.7% of the crosslinked duplex was photo-split at 37 °C in water, consistent with our previous results. We observed that reversing the crosslink in denaturing conditions improved photo-splitting, but both increased temperature and 50% formamide was required to drive the reaction to near-completion **(Figure 4.19)**.

Because strong denaturing conditions were required to totally photo-split the ^{CNV}K crosslink, we considered that 310 nm light might be capable of catalyzing both the forward and reverse photo-crosslinking reactions. In this case, two oligonucleotides might immediately crosslink after they had been photo-split if they were still hybridized. To test this, two pools of ^{CNV}K-ODN/comp-ODN^{FAM} duplex were prepared, but one was crosslinked and the other was not. The pools were then irradiated at 37 °C with 310 nm light for up to 5 min to track the rate of

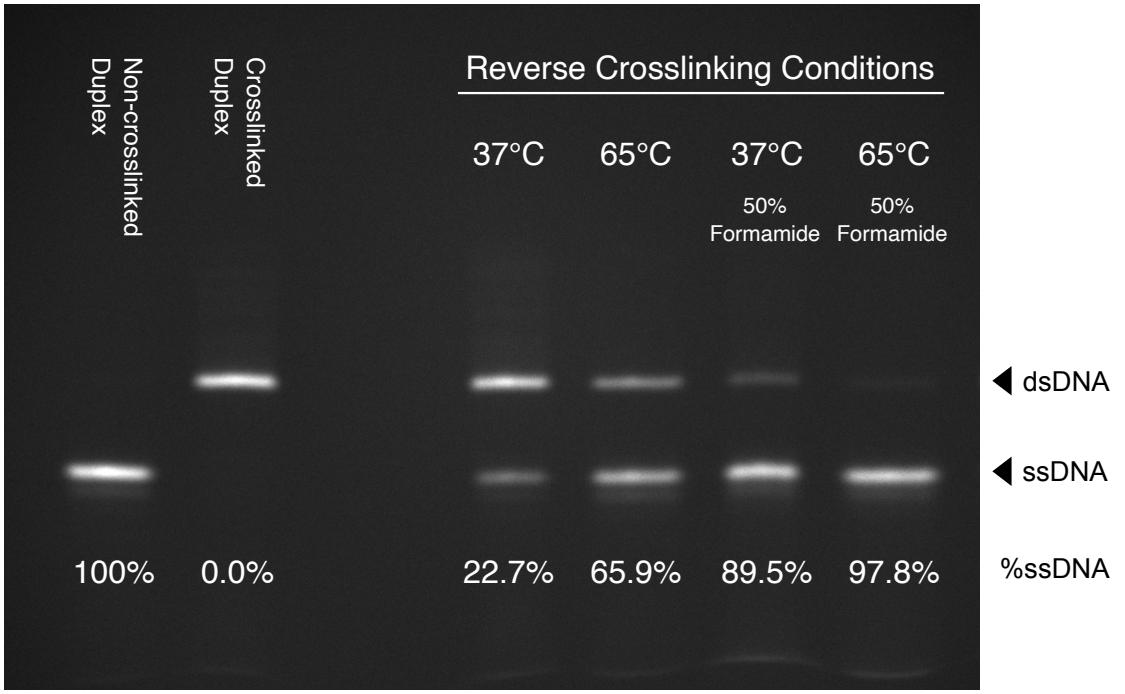


Figure 4.19: Denaturing PAGE analysis of ^{CNV}K photo-splitting under various denaturing conditions. A crosslinked ^{CNV}K-ODN/comp-ODN^{FAM} duplex was irradiated for 1 min with a 310 nm LED light source. The solvent and temperature used during photo-splitting is listed above the respective lane. The amount of ssDNA in each reaction was calculated and written below each lane as the % of total fluorescence. Arrows to the right of the gel highlight ssDNA comp-ODN^{FAM}, and crosslinked (dsDNA) comp-ODN^{FAM}.

photo-splitting (reverse) and crosslinking (forward) of the duplexes respectively. Time points from each reaction were run on denaturing PAGE, visualized by FAM fluorescence, and were then quantified by Fiji **(Figure 4.20)**. Consistent with our previous experiments, we found that the reverse crosslinking reaction proceeded from 0% to approximately 21% single-stranded species in 30 s. Interestingly, we observed that the forward crosslinking reaction was also being catalyzed by 310 nm light, which also proceeded until only 21% of single-stranded material remained. Not only did this confirm that 310 nm can catalyze both reactions, but it also demonstrated that a crosslinking equilibrium is reached after 30 s under these conditions.

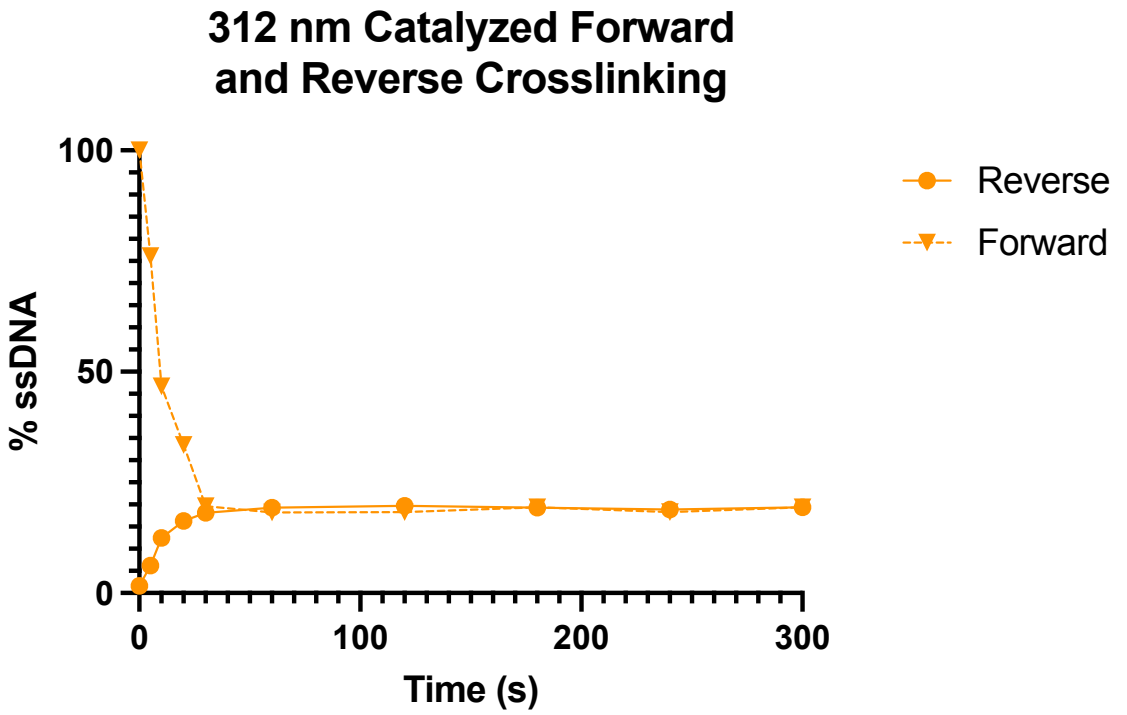


Figure 4.20: ^{CNV}K crosslinking/photo-splitting rates when catalyzed by 310 nm light. The reverse reaction (circles) tracks the photo-splitting of a crosslinked ^{CNV}K-ODN/comp-ODN^{FAM} duplex. The forward reaction (triangles) tracks the crosslinking of non-crosslinked ^{CNV}K-ODN/comp-ODN^{FAM}. The % ssDNA was determined by running each time point on denaturing PAGE, visualizing comp-ODN^{FAM}, and subsequent analysis with Fiji.

4.4.4) UV Toxicity Study of HEK293 Cells

Even though we were only partially able to reverse the ^{CNV}K crosslink under physiological conditions, we decided to proceed with cell experiments to determine if ^{CNV}K residues could be used to activate translation at least to the extent we observed *in vitro*. Additionally, we considered that the photo-splitting reaction may be more favorable in cells as there would be a relatively lower concentration of blocker compared to *in vitro* conditions.

To begin, we first performed a cell viability assay to determine what length of UV exposure HEK293 cells could tolerate before losing viability. For all cell-based UV experiments, measures were taken to minimize UV spillover between separate populations during irradiation as described in Section 6.11.6. Cells were seeded into wells, and groups of wells

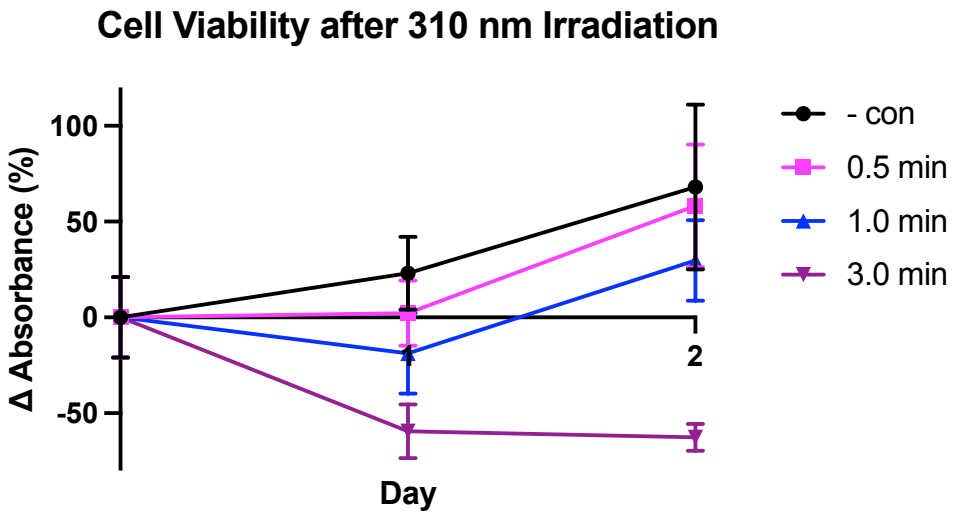


Figure 4.21: WST-8 cell viability assay of HEK293 cells irradiated with 310 nm light for different lengths of time. Absorbance from formazan was normalized to Day 0 negative control cells, which was set to 1, with propagation of error. Values reflect the average change in absorbance as determined by n=3 biological replicates, error bars = SD.

were irradiated with 312 nm light for 0, 30, 60, or 180 s. Viability assessments were performed using the WST-8 assay immediately before irradiation or 1- and 2-days post-irradiation **(Figure 4.21)**. The viability of populations irradiated for 30 s did not decrease compared to the Day 0 value and was not statistically different from the control population at any time point (Day 1 two-tailed unpaired P value = 0.2246, Day 2 two-tailed unpaired P value 0.7661). Cell irradiated for 60 s had an initial drop in viability at Day 1 but then recovered, and similarly were not statistically different from the control population at any time point (Day 1 two-tailed unpaired P value = 0.0602, Day 2 two-tailed unpaired P value 0.2357). Cells which were irradiated for 180 s showed an immediate and unreversed drop in viability compared to the control population. This demonstrated the toxicity of this wavelength over prolonged periods but indicated that HEK293 cells can tolerate a degree of UV irradiation, ideally for 30 s or less.

4.4.5) ^{CNV}K Photo-Regulation of mRNA in Cells

Through our previous experiments, we established that photo-splitting a ^{CNV}K from an mRNA allows the mRNA to be translated, that the ^{CNV}K photo-splitting reaction reaches an equilibrium after 30 s of 310 nm irradiation at 37 °C, and that HEK293 cells can tolerate 310 nm irradiation for at least up to 1 min. Our next goal was to combine this information by transfecting B1.2 primed mRNA into cells and activating their translation in cells through UV light control.

To accomplish this, HEK293 cells were seeded into a 96-well plate and were incubated for 2 days. At this point, IRES-mCherry mRNA primed with x10 excess B1.2 was transfected into the cells alongside an unblocked control. 4 h-post transfection, cell culture media was changed, and cells were irradiated with 312 nm light for either 0, 30, or 60 s. Cells were analyzed by flow cytometry after 24 h **(Figure 4.22)**.

As is consistent with previous experiments, we found that primed IRES-mCherry mRNA was less than 1% as fluorescent as the positive control population (0.79%, S.D. 0.11%). The transfected populations irradiated with 312 nm light for 30 and 60 s both showed an increase in fluorescence (9.48%, S.D. 0.82%, and 16.70% S.D. 1.10% respectively) as expected. The expression of activated mRNA was slightly less than the value we achieved when crosslinking was reversed *in vitro* before transfection (Section 4.4.1). This difference was most likely caused by the 4-hour delay between transfection and activation, as the positive control would have been translated during this time while the primed mRNA would not. Additionally, it is also likely that the growth media and cellular components scattered and

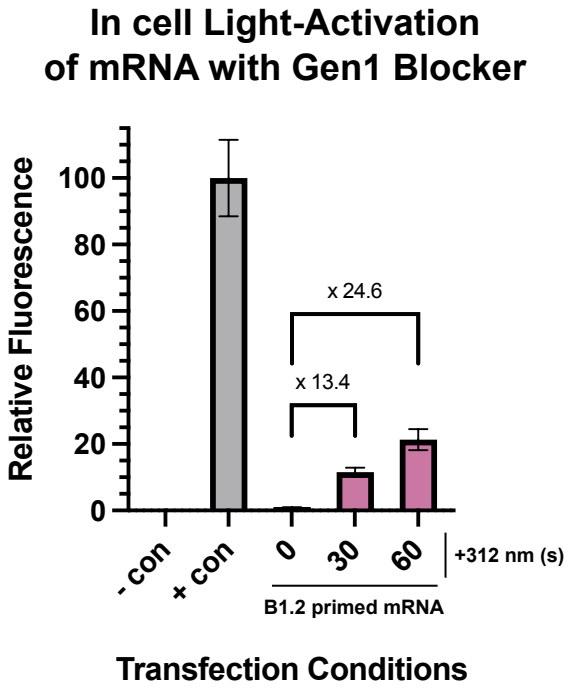


Figure 4.22: Relative mCherry fluorescence from HEK293 cells after *in vivo* mRNA activation. The background fluorescence was removed by subtracting the average fluorescence of mock-transfected cells from each population with propagation of error. The average fluorescence was then normalized to the positive control positive with propagation of error. Transfection conditions are listed as - control, + control, or the duration of 310 nm light irradiation for populations transfected with B1.2 primed mRNA. Brackets indicate the fold-change in relative fluorescence between samples. Values reflect the average fluorescence as determined by flow-cytometry of n=3 transfection replicates, error bars = SD.

absorbed some of the incident UV light. This effect could explain why mRNA expression was higher after 60 s irradiation than 30 s, even though *in vitro* data showed that reverse-crosslinking reaches an equilibrium after 30 s.

Although we only observed a moderate increase in mRNA expression after 310 nm irradiation, these results indicate that a ^{CNV}K-blocker strategy for photo-regulated mRNA translation is fundamentally possible. The ability of this approach to silence the expression of mRNA is particularly robust, which is critical for precisely regulating the spatial-temporal dynamics of translation. The relatively low rate of mRNA activation and the duration of irradiation to achieve it was less than ideal however, so our next goal for the project was to improve the rate of activation.

4.5 Improvement of ^{CNV}K-Blocker Design

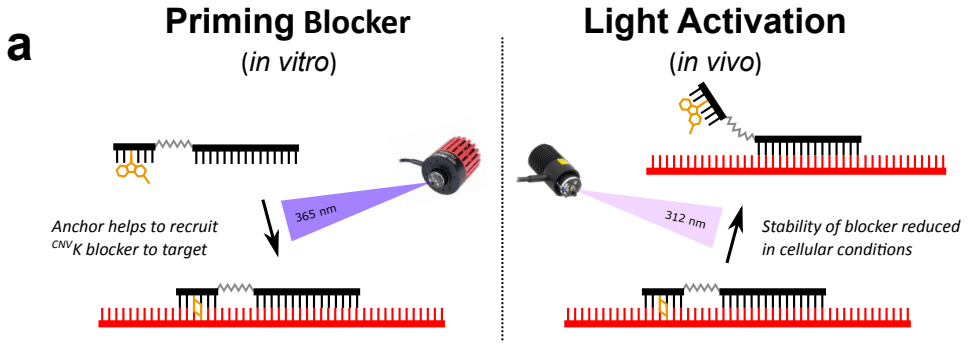
4.5.1) 2nd Generation ^{CNV}K-Blockers

As demonstrated in Section 4.4.3, 310 nm light is capable of catalyzing both ^{CNV}K crosslinking (forward) and photo-splitting (reverse) under physiologically relevant conditions. The ratio of forward/reverse crosslinking is influenced at least partially by the stability of the ^{CNV}K oligonucleotide duplex. Although denaturing conditions can improve the ^{CNV}K photo-splitting rate, these cannot be applied *in vivo*. Therefore, the only viable option to improve the activation of ^{CNV}K-primed mRNA in cells is to change the design of the blocker. The first-generation blockers are very effective at silencing translation, but this property is detrimental to reversing the crosslink. We hypothesized that blockers designed to interact less strongly with the mRNA target could have a lower forward-crosslinking rate, and therefore arrive at a more favorable photo-splitting equilibrium when irradiated with 310 nm light. Because the mRNA is primed with blockers *in vitro*, the reduced T_M of these blockers can be compensated for by adjusting conditions during crosslinking.

The simplest way to reduce the stability of the ^{CNV}K Blocker/mRNA interaction is to shorten the length of the blocker. However, our previous experience with the DNA B1.1 suggests that if the interaction is too weak, the blocker will be unable to overcome the RNA secondary structure to crosslink to its target. To balance these competing interests, we designed the 2nd generation of 2′OMe ^{CNV}K blockers as two parts connected by a C12 linker: The first half of the design is a 15 nt "anchor", and the other is the shorter "blocker" component which contains the ^{CNV}K residues. The anchor is approximately the same length as the 1st generation blockers and should have a high affinity for the target site, which will cause RNA structure disruption and increase the local concentration of the blocker. The blocker component is between 6-11 nt long, giving it a greatly reduced local T_M compared to the 1st

generation blockers. By decreasing the local T_M of the blocker component, we aimed to make the photo-splitting rate more favorable in cells. Importantly, the anchor and blocker's target sequences in the mRNA are separated by 5 nt; this, along with the carbon linker between the two segments, should prevent duplex propagation between the two halves of the blocker and insulate the T_M of the blocker half from the anchor half.

A simple diagram showing the dynamics of 2nd generation blockers is shown in **Figure 4.23a**. Although the blocker component interacts weakly with its target mRNA, the anchor should recruit it to the target site. Conditions during the *in vitro* mRNA priming, such as temperature, salt concentration, and excess blocker, can be optimized to force the priming reaction to completion. After the mRNA is introduced into cells, the increased temperature and reduced concentration of blocker should favor the reverse crosslinking reaction. Upon irradiation with 310 nm light, the blocker will be photo-split from the mRNA and allow ribosome progression.



b

	15 nt Anchor	Blocker	Size
B2.1	GGAGTCCTGGGTCAC	---CCA ^{CNV}K GC	6
B2.2	CCTGCAGGGAGGAGT	---GTCA ^{CNV}K GGT	8
B2.3	GTCCTGCAGGGAGGA	---GGGTCA ^{CNV}K GGTC	11

Figure 4.23a: Diagram of mRNA photo-regulation by 2nd generation blockers. **Figure 4.23b:** Sequences of 2nd generation ^{CNV}K-blockers. The C12 linker between the anchor and blocker components is represented by dashes. The size of the blocker component is listed to the right of the sequence.

We designed three 2nd generation blockers, each with a 15 nt anchor and a "blocker" component of varying length (B2.1, B2.2, and B2.3)**(Figure 4.23b)**, which should produce a range of crosslinking and photo-splitting rates. The target site selected for each blocker was within 323-360 nt from the IRES-mCherry mRNA start codon, as this region was predicted to have minimal secondary structure with RNAfold [211].

4.5.2) 2nd Generation Blocker *in vitro* Photo-Regulation

Similar to the experiment described in Section 4.4.1, IRES-mCherry mRNA was mixed with a x10 molar excess of each 2nd generation blocker, which was then annealed, irradiated with 365 nm light for 30 s, and then 312 nm light for 1 min. An aliquot was removed after each step and transfected into HEK293 cells, which were analyzed by flow cytometry after 24 h to determine levels of translation. An additional sample was prepared with B1.2 to compare the activity of the new blockers to the first generation. mCherry fluorescence was normalized to a positive control as reported in **Figure 4.24**.

One of the most apparent results from this experiment is that none of the 2nd generation blockers inhibit translation as effectively as B1.2 after crosslinking (+365 nm). The degree of inhibition achieved by each 2nd generation blocker correlated with the length of their "blocker" component. B2.3 was the most effective 2nd generation blocker, but was only able to repress translation to 7.2% of the positive control compared with B1.2, which silenced expression to 1.3% in this instance. Given that the blocker component of B2.3 was only 4 nt shorter than the entirety of B1.2, we did not expect such a dramatic difference in translation silencing and considered that it could have additionally been an effect of the new target site we selected.

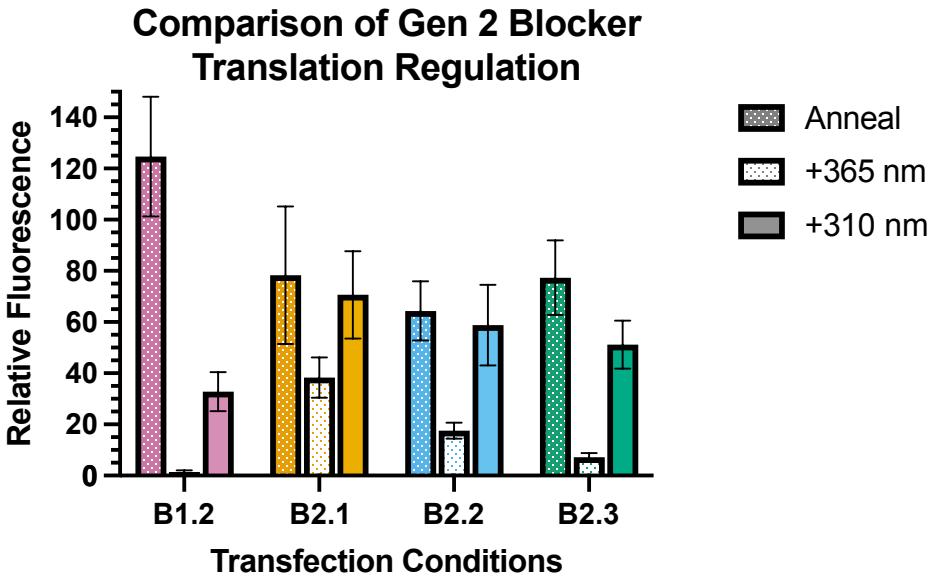


Figure 4.24: Relative mCherry fluorescence from 2nd generation ^{CNV}K-blocker photo-regulation. The background fluorescence of each cell population was removed by subtracting the average fluorescence of mock-transfected cells with propagation of error. The average fluorescence was then normalized to the positive control with propagation of error, which was set to 100%. Cell populations are grouped by which blocker was used for photo-regulation and are subdivided by photo-crosslinking state (annealed, crosslinked (+365 nm), and photo-split (+310 nm)).

A second finding is that even without 365 nm light to induce crosslinking, mRNA annealed with 2nd generation blockers experienced decreased levels of translation compared to the positive control. This effect was consistent across repeats of this experiment and was never observed with the 1st generation blockers. This effect is likely caused by the anchor component of the 2nd generation blockers, which is an unmodified 15 nt stretch of 2′OMe. We suspect that the anchor is long enough that the hybrid it forms with target mRNA is sufficiently stable to impede ribosome progression through a steric effect. Although several 1st generation blockers are the same length, the ^{CNV}K residue has a destabilizing effect on duplexes which might help to either melt the blocker from the mRNA or allow ribosomes to displace the blocker.

The design of B2.1, B2.2, and B2.3 therefore failed in two ways: none of the blockers can induce a translation-off state, but they also inhibit translation in the absence of crosslinking,

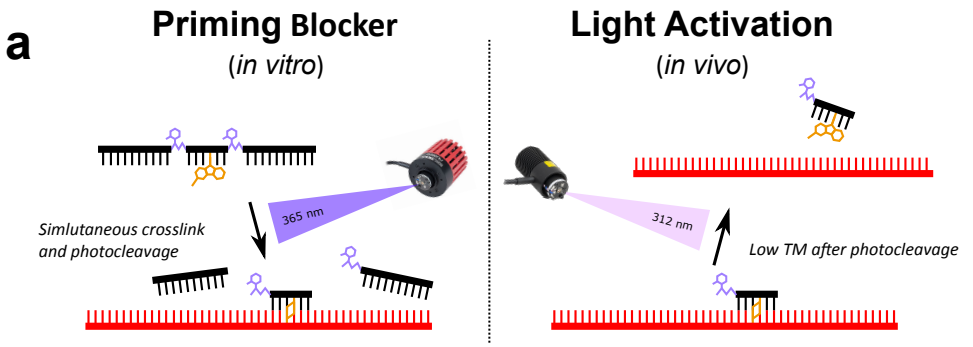
which places a ceiling on mRNA activation. However, it did appear that the 2nd generation design improved translation activation rates after 310 nm irradiation: mRNA expression after 310 nm irradiation is dependent on the proportion of blockers which are photo-split from the mRNA. The "annealed" stage of blocker regulation is the maximum achievable mRNA expression, and is equivalent to a condition in which all blockers are photo-split. In this case, the fold-difference of fluorescence between the 312 nm irradiation and annealed stages should be proportional to the percent of photo-splitting achieved. For B1.2, the fold-difference in fluorescence between the 312 nm and annealed stages was 3.80, which is consistent with the activation rates we observed in the previous section. For the 2nd generation blockers however, these values were 1.11 (B2.1), 1.09 (B2.2), and 1.51 (B2.3), which would indicate that a larger proportion of these blockers are photo-split after 310 nm irradiation. Unpaired t-tests did not identify a significant difference in mRNA expression between the "anneal" and "312 nm" states of the 2nd generation blockers (0.6974, 0.6495, and 0.0589 for B2.1, B2.2, and B2.3 respectively). Although the 2nd generation blockers were ineffective overall, this result suggests that our strategy of destabilizing the interaction between blocker and mRNA could be improving photo-splitting rate.

4.5.3) 3rd Generation ^{CNV}K-Blockers

Encouraged by the results of photo-regulation with 2nd generation blockers, we wanted to improve these blockers to determine whether ^{CNV}K photo-splitting rates could be improved through oligonucleotide design alone. Two issues emerged during our cell-based experiment: none of the 2nd generation blockers could effectively inhibit translation, indicating that we needed to strengthen the interaction between blocker and mRNA. At the same time, the blockers were having a steric effect that impeded ribosome progression regardless of crosslinking, which argues for a weaker interaction.

To balance these two factors, the ideal design for our blockers would have a high affinity for mRNA until crosslinking, at which point the interaction would become less favorable. We realized this could be achieved by incorporating photo-cleavable linkers into the blocker design. The T_M of a short ^{CNV}K-blocker can be increased by linking it to anchors, similar to the design of the 2nd generation blockers. Then, either during or after photo-crosslinking, the linkers could be photo-cleaved to reduce the blocker's T_M, which will improve its photo-splitting rate.

We therefore arrived at the design of 3rd generation ^{CNV}K-blockers **(Figure 4.25a)**, which were composed of a core ^{CNV}K-blocker 8-10 nt in length, flanked by either two (B3.1, B3.2) or one (B3.3) photo-cleavable anchor. The anchors of each 3rd generation were kept below 15 nt in length to avoid the steric effect we observed with 2nd generation blockers **(Figure 4.25b)**. The linkers between each component are a commercially available photo-cleavable



b

	Blocker
B3.1	UCGAACUCGU-PC-CGUUCA ^{CNV}K GGA-PC-CUCCAUGUGC
B3.2	UGGCCGUC-PC-GGA ^{CNV}K CCCA-PC-UGUGCAC
B3.3	CCGUUCACGGAG-PC-UCCA ^{CNV}K GUGC

Figure 4.25a: Diagram of mRNA photo-regulation by 3rd generation blockers. **Figure 4.25b:** Sequences of 3rd generation blockers. The core "blocker" which will remain crosslinked to target mRNA after photo-cleavage is highlighted in orange. Photocleavable linkers are indicated with "PC". Unhighlighted text corresponds to the anchors which help recruit the blockers to target site.

(PC) spacer: a C3 chain conjugated to an *ortho*-nitrobenzyl group, which is cleavable at 365 nm. This conveniently allowed us to simultaneously induce crosslinking and photo-cleavage, minimizing the number of steps involved with priming mRNA and the UV light sources required to complete them. Each 3rd generation target is within 63-98 nt from the IRES-mCherry mRNA start codon, similar to B1.2.

4.5.4) 3rd Generation Blocker Photo-Crosslinking and Photo-Cleavage

Because the activity of the 3rd generation blockers is more dynamic than our previous designs, we performed a simple gel-based assay to confirm whether it was capable of crosslinking to its intended target. To this end, 10 pmol of FAM-ODN$_{60\text{-}99}$ was mixed with equimolar amounts of each 3rd generation blocker, which were then annealed and crosslinked for 30 s. The crosslinking reactions and non-crosslinked controls were analyzed by denaturing PAGE and visualized by FAM fluorescence **(Figure 4.26a)**. The results of this

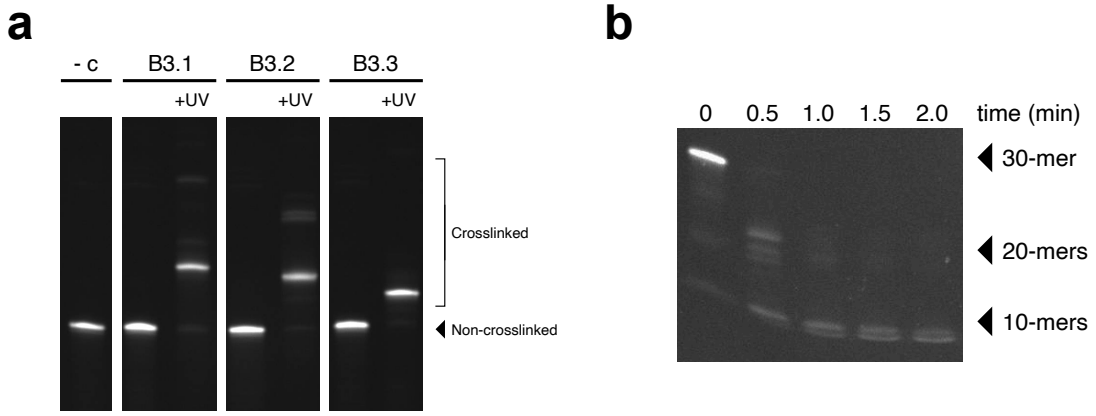


Figure 4.26a: Denaturing PAGE analysis of 3rd generation blockers crosslinked to FAM-ODN$_{60\text{-}99}$. Labels above gel lanes indicate whether a reaction was prepared with FAM-ODN$_{60\text{-}99}$ alone or in combination with one of the 3rd generation blockers, and whether that sample was irradiated with 365 nm light. Arrow and bracket to the right of the gel indicates bands which correspond to single-stranded FAM-ODN$_{60\text{-}99}$ and crosslinked FAM-ODN$_{60\text{-}99}$ respectively. Gel was visualized by FAM fluorescence. **Figure 4.26b:** Denaturing PAGE analyze of B3.1 irradiated with 365 nm light for up to 2 min to track photo-cleavage. Arrows to the right of the gel highlight which bands correspond to the full-length B3.1 30-mer, partially cleaved 20-mers, or fully cleaved 10-mers. Gel was visualized by SYBR-Gold staining.

gel suggest that each blocker is capable of crosslinking to its target sequence. We observed additional faint bands above the main product in the B3.1 and B3.2 crosslinking reactions which were absent in the B3.3 reaction. Because B3.1 and B3.2 have two PC spacers as opposed to one, we believe these bands likely correspond to incomplete photo-cleavage.

To confirm this theory, we performed a UV-irradiation time course of B3.1 by itself to observe its rate of photo-cleavage. A stock of B3.1 was irradiated for 2 min, with aliquots removed every 30 s. The time points were analyzed by denaturing PAGE, as shown in **Figure 26.b**. We observed substantial photo-cleavage after 30 s of irradiation, but 1 min or longer is required for complete cleavage. Some variation is present in the size of the 20-mer and 10-mer bands, which corresponds to the sequence heterogeneity of the individual fragments after cleavage.

4.5.5) 3rd Generation Blocker *in vitro* Photo-Regulation

After characterizing the photo-cleavage and photo-crosslinking properties of 3rd generation ^{CNV}K-blockers, we proceeded to assess their translation regulation capabilities with *in vitro* UV irradiation. Similar to the experiment described in Section 4.4.1, IRES-mCherry mRNA was mixed with a x10 molar excess of each 3rd generation blocker, which was then annealed, irradiated with 365 nm light for 60 s, and then 312 nm light for 1 min. An aliquot was removed after each step and transfected into HEK293 cells, which were analyzed by flow cytometry after 24 h. An additional sample was prepared with B1.2 to compare the 3rd generation activity to the 1st generation. mCherry fluorescence was normalized to a positive control as reported in **Figure 4.27**.

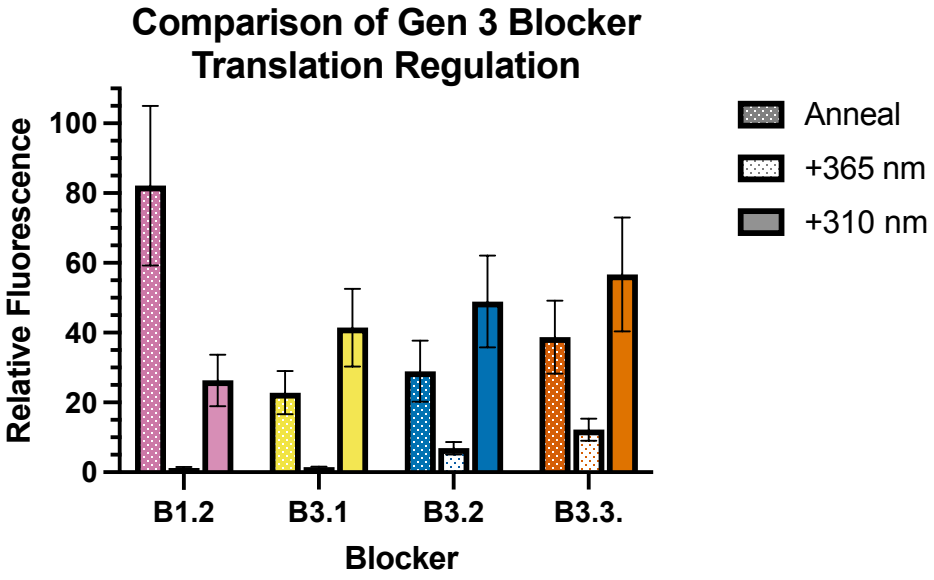


Figure 4.27: Relative mCherry fluorescence from 3rd generation ^{CNV}K-blocker photo-regulation. The background fluorescence of each cell population was removed by subtracting the average fluorescence of mock-transfected cells with propagation of error. The average fluorescence was then normalized to the positive control with propagation of error, which was set to 100%. Cell populations are grouped by which blocker was used for photo-regulation and subdivided by photo-crosslinking state (annealed, crosslinked (+365 nm), and photo-split (+310 nm)).

Excitingly, we found that the 3rd generation blockers were overall more effective at inhibiting the translation of mRNA than the 2nd generation blockers. B3.1 was able to repress 98.8% of translation compared to the positive control population, comparable to the 98.9% silencing achieved by B1.2 in this instance. Additionally, mCherry expression from B3.1-primed mRNA was more than 50% greater than B1.2 after 312 nm photo-splitting.

We observed that the translation of mRNA primed with 3rd generation blockers was significantly lower than the positive control even without crosslinking the blockers, similarly to the 2nd generation blockers. This result was expected, as the 3rd generation blockers were larger than most of the 2nd generation blockers prior to photo-cleavage. However, we also observed that the translation from photo-split blockers (312 nm) was greater than the annealed state for each of the 3rd generation blockers. This indicates that our strategy to

reduce the stability of the ^{CNV}K-blockers after photo-crosslinking was remediating the steric effect we had previously observed.

The ranges of mRNA expression observed from photo-regulation with B3.2 and B3.3 were greater than that of B3.1, but neither produced as robust an off-state as B3.1. We reasoned that we could improve their off-states by increasing the relative concentration of each blocker during mRNA priming. Because these blockers did not appear to produce a steric effect on translation as significantly as the 2nd generation blockers, larger excesses of blocker should not impact translation after photo-cleavage and photo-splitting. To test this, we repeated the experiment described above but applied a x10, x30, and x50 excess of B3.2 and B3.3 to mRNA. Primed or photo-split mRNA was transfected into cells, which were analyzed by flow cytometry after 24 h. As above, the expression from each transfection condition is

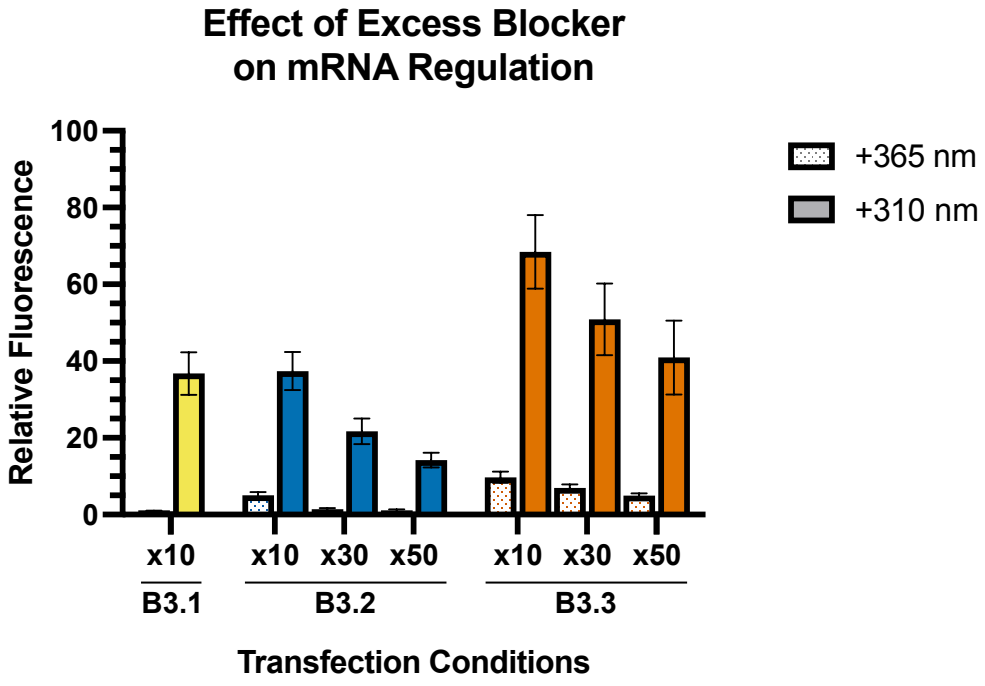


Figure 4.28: Photo-regulation of IRES-mCherry mRNA with excess amounts of 3rd generation ^{CNV}K-blockers. The background fluorescence of each cell population was removed by subtracting the average fluorescence of mock-transfected cells with propagation of error. The average fluorescence was then normalized to the positive control with propagation of error, which was set to 100%. Cell populations are grouped by which blocker was used for photo-regulation and subdivided by the molar excess used during crosslinking. Fluorescence values are reported for blocker/mRNA after crosslinking (+365 nm) and photo-splitting (+310 nm).

reported relative to an mRNA positive control (**Figure 4.28**). The application of excessive blockers was able to improve silencing rates. A 50x excess of B3.2 was able to produce an off-state of 0.5% expression, comparable to the rate achieved with a 10x excess of B3.1 (0.6% expression, in this case). The silencing rate of B3.3 was also increased, with 5.0% expression achieved for mRNA primed with a 50x excess of blocker.

Interestingly though, we observed that the relative expression of mRNA after 312 nm irradiation decreased with increasing excesses of B3.2 and B3.3. While a mild decrease might be expected, it was striking that application of x30 and x50 excesses of B3.2 yielded mRNA activation of 21.7 and 14.2% respectively. This was especially striking as B3.1, which has a larger ^{CNV}K-blocker than B3.2 after photo-cleavage, had an activation rate of 36.8% as well as an off-state of <1% expression. We expected that these two blockers would at least have a comparable rate of activation, which led us to conclude that the excess quantities of B3.2 were either crosslinking to off-target sites or inhibiting translation through steric effect. Therefore, while the silencing rates of this generation could be improved through application of excess blocker, it did not seem like a viable strategy to improve photo-regulation.

4.5.6) 2nd and 3rd Generation Blocker Photo-Regulation of mRNA in Cells

After assessing the activation rates of 2nd generation and 3rd generation blockers with *in vitro* light activation, we wanted to quantify the gains of these designs with an in-cell photo-regulation assay. We chose the best-performing blocker from each generation: B1.2, B2.3, which was not as effective as silencing expression as B1.2 but had a greater range of expression between its off and on-states, and B3.1.

IRES-mCherry mRNA was primed with blockers B1.2, B2.3, or B3.1, and was then transfected into HEK293 cells. 4 hours after transfection, the media was changed, and half the transfected cells were irradiated with 310 nm light for 45 s to activate translation. Cells were then incubated for an additional 24 h before being analyzed by flow cytometry. The average fluorescence of each population was normalized to a positive control, as reported in **Figure 4.29**.

Consistent with our previous in-cell translation activation assay **(Figure 4.22)**, we observed that B1.2 was able to produce a robust off-state but had low levels of activation (7.4% S.D. 1.7%). B2.3 was able to achieve a relatively higher level of mRNA activation but had a leakier off-state, consistent with the results we found from *in vitro* photo-regulation with this blocker. B3.1 had the best attributes of both designs, achieving a translational off-state of

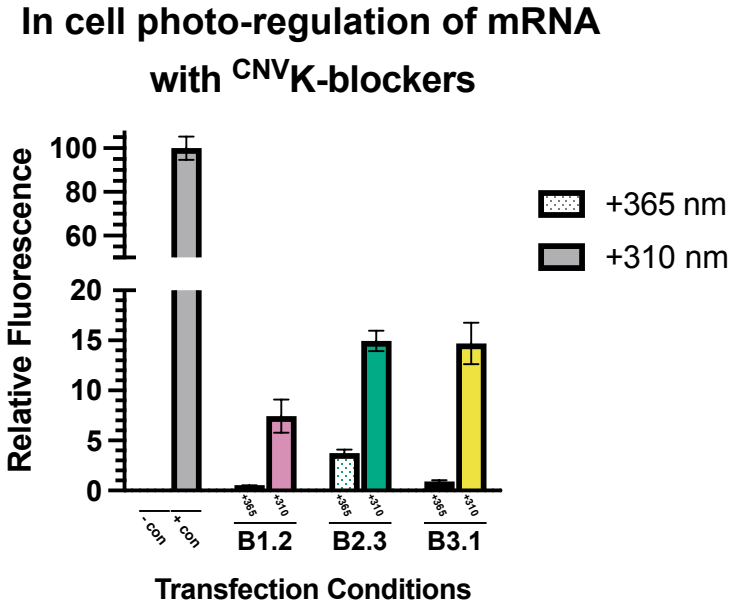


Figure 4.29: Relative mCherry fluorescence from HEK293 cells after *in vivo* mRNA activation. Background fluorescence was removed by subtracting the average fluorescence of mock-transfected cells from each population with propagation of error. The average fluorescence of each population was then normalized to the positive control, which was set to 100%. Populations are grouped by CNVK-blocker used for photo-regulation. +365 nm refers to the fluorescence from primed-mRNA, and +310 nm refers to photo-activated mRNA.

<1% of the expression of the positive control, and a 14.7% S.D. 2.1% activation rate. B3.1 was therefore able to improve upon the design of the 2nd generation blockers without sacrificing its activation rate while achieving mRNA silencing rates comparable to 1st generation blockers.

While these results confirmed our success with improving the activation rates of ^{CNV}K-blockers through design alone, the activation rates achieved with B3.1 were still relatively low compared to published mRNA photo-regulation methods. The design of B3.1 could likely be optimized to improve its mRNA activation rates; however, this would require additional photocleavable linkers to further reduce the size of anchor and blocker components while maintaining a similar T_M. The increased complexity would make the oligonucleotide relatively more difficult to synthesize and less accessible for other users. Because of this, we began to consider alternative strategies to improve photo-regulation rates.

4.6 Alternative Carbazoles Improve Photo-Regulation

4.6.1) Pyranocarbazole Nucleotides

The work of the previous section (Section 4.5) yielded ^{CNV}K-blocker designs with slightly better activation rates than our 1st generation design. However, our designs still suffered from the meager photo-splitting rate of ^{CNV}K crosslinks under physiological conditions. Because of this, we considered that an alternative crosslinking agent with a more favorable photo-splitting rate could be used in place of the 3-cyanovinylcarbazole. In particular, I was interested in the more recently described pyranocarbazole [48] and 4-methylpyranocarbazole [49] nucleotides, which were previously discussed in Section (1.4.1). Both carbazole nucleotides are reported to undergo crosslinking when irradiated with 365-450 nm and can be at least partially photo-split with 330 nm irradiation. Because crosslinking can be performed

with visible light, photo-crosslinking and photo-cleavage steps could be performed orthogonally in a 3rd generation style blocker. Additionally, the higher photo-splitting wavelength of pyranocarbazole nucleosides would be less toxic to cells compared to the 3-cyanovinylcarbazole. We also hypothesized that the greater distinction between photo-crosslinking and photo-splitting wavelengths in pyranocarbazole nucleosides would improve the ratio of crosslinking to splitting (forward/reverse) when irradiated with 310 nm light, which would improve the activation rates of the mRNA blockers.

Neither the pyranocarbazole nor 4-methylpyranocarbazole nucleotide has been extensively characterized *in vitro* or *in vivo*. Of the two, a more robust body of photo-splitting data has been reported for the pyranocarbazole. The 4-methylpyranocarbazole nucleoside is also incapable of crosslinking to cytosine residues, making it less versatile overall. For these reasons, we decided to investigate whether the pyranocarbazole nucleotide (^{PC}K) would be more effective to use in our translational blocker strategy. The ^{PC}K phosphoramidite monomer was kindly synthesized by Dr. Alice Kennett according to the method described in [48].

4.6.2) *In* vitro Characterization of ^{PC}K

To characterize ^{PC}K photo-crosslinking and photo-splitting, we synthesized a short ODN (^{PC}K-ODN)(Appendix 3) which was isosequential to the previously described ^{CNV}K-ODN. To confirm that the ^{PC}K monomer was functional, we first tested its crosslinking with comp-ODN^{FAM}. ^{PC}K-ODN and ^{CNV}K-ODN were annealed with comp-ODN^{FAM} and irradiated with either 30 s 365 nm or 300 s 420 nm light. The crosslinking reactions were then analyzed by denaturing PAGE alongside a non-irradiated negative control **(Figure 4.30)**. As expected,

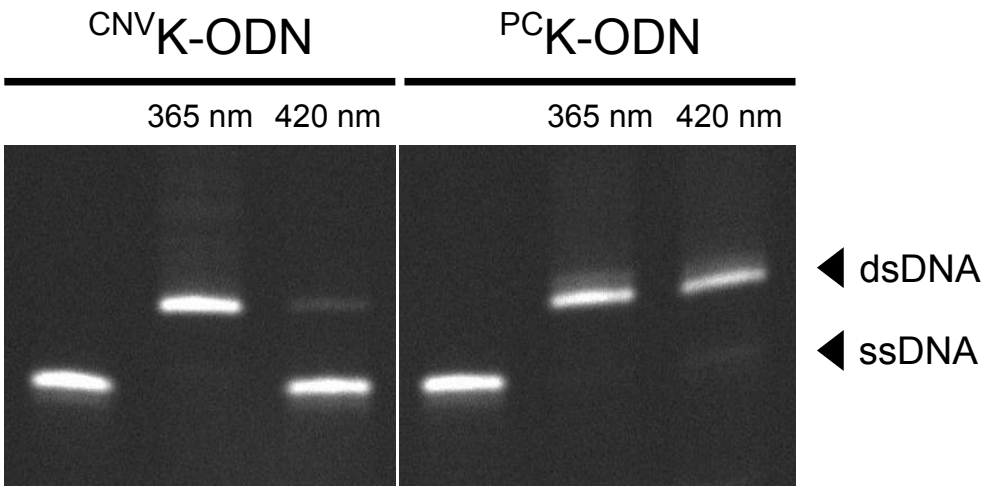


Figure 4.30: Denaturing PAGE analysis of ^{CNV}K and ^{PC}K crosslinking. ^{CNV}K-ODN and ^{PC}K-ODN were annealed with comp-ODN^{FAM} and irradiated with 30 s 365 nm light or 300 s 420 nm light (as listed above gel). Samples were run on gel and analyzed by FAM fluorescence. Arrows to the right of the gel indicate bands which correspond to non-crosslinked (ssDNA) and crosslinked (dsDNA) comp-ODN^{FAM}.

365 nm light was able to crosslink both carbazole-ODNs to comp-ODN^{FAM}. 420 nm irradiation was also able to catalyze the crosslinking of ^{PC}K-ODN, while only minimally affecting the ^{CNV}K-ODN.

We next proceeded to investigate the photo-crosslinking/photo-splitting equilibrium of the ^{PC}K-ODN when irradiated with 310 nm light in physiologically relevant conditions. Similar to the experiment described in Section 4.4.3, two pools of carbazole-ODN/comp-ODN^{FAM}, duplex were prepared, one of which was crosslinked and one was not. The pools were then irradiated at 37 °C with 310 nm light for up to 2 min to track the rate of photo-splitting (reverse) and crosslinking (forward) reactions of the duplexes respectively. Time points from each reaction were run on denaturing PAGE, visualized by FAM fluorescence, and were then quantified by FIJI. Three replicates were prepared for each reaction **(Figure 4.31)**.

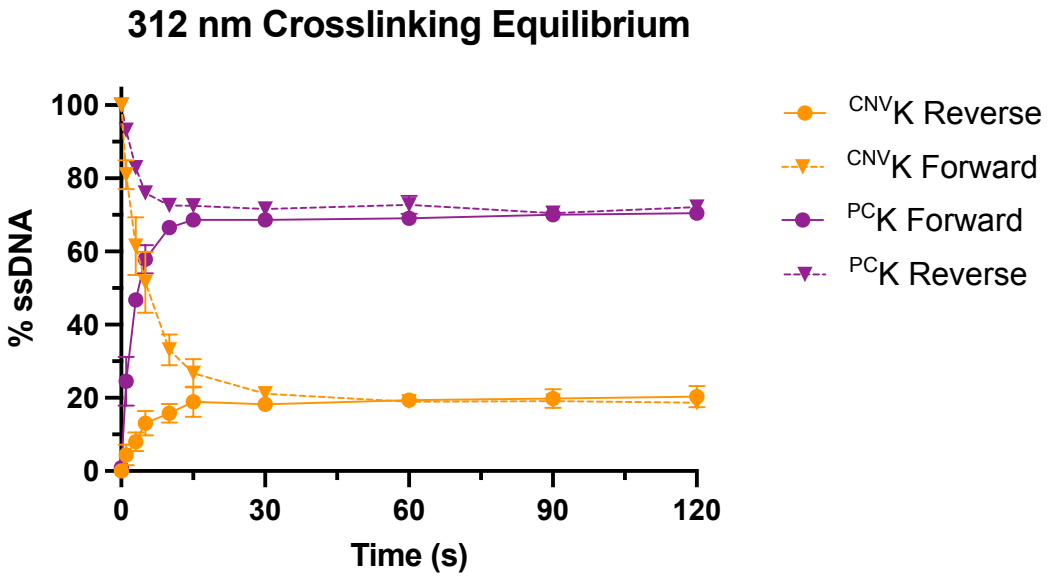


Figure 4.31: Carbazole crosslinking/photo-splitting rates when catalyzed by 310 nm light at 37°C. The reverse reaction (circles) tracks the photo-splitting of crosslinked duplex. The forward reaction (triangles) tracks the crosslinking of carbazole-ODN to comp-ODN-FAM. The % ssDNA was determined by running each time point on denaturing PAGE and subsequent analysis with Fiji. Data points reflect the average of n=3 replicates, error bars = SD.

Consistent with the results presented earlier in **Figure 4.20**, we observed that ^{CNV}K-ODN duplexes approach equilibrium with approximately 20% of material photo-split after 30 s. The percent of photo-split duplex does not increase after this time point, underscoring the preferential ^{CNV}K crosslinking reaction over splitting in physiologically relevant conditions. Impressively, the ^{PC}K-ODN duplexes reached an equilibrium of 71% photo-splitting, which is approximately 3.5x higher than the ^{CNV}K-ODN photo-splitting. If this result is consistent in mRNA crosslinking, the activation rate of ^{PC}K-based translational blockers would be significantly higher compared to ^{CNV}K-blockers. The ^{PC}K-ODN duplexes also appeared to reach an equilibrium state approximately 10 s earlier than the ^{CNV}K-ODN, indicating that it might require less UV irradiation to drive the photo-splitting reaction to equilibrium. Either of these qualities would be beneficial for activating translation *in vivo*; taken together, it shows that ^{PC}K residues could massively improve mRNA photo-regulation over our older ^{CNV}K designs.

Before proceeding to design a ^{PC}K-blocker, we decided to investigate the extent to which ^{PC}K crosslinks could be photo-split at higher wavelengths. Fujimoto *et al.* [48] previously reported that 65% of crosslinked PCK duplexes could be photo-split with 10 min 330 nm irradiation, but did not provide more details of the reaction apart from buffer composition. To better understand photo-splitting dynamics within conditions relevant for cell work, we prepared crosslinked ^{CNV}K and ^{PC}K-ODN/comp-ODN^{FAM} duplex and attempted to photo-split them with 30 s of 310, 325, or 340 nm irradiation at 37 °C. Photo-splitting reactions were subsequently analyzed by denaturing PAGE and visualized by FAM fluorescence **(Figure 4.32)**. Unsurprisingly, a decrease in photo-splitting rate was observed as the wavelength of light increased. The difference in photo-splitting at 310 nm and 325 nm was modest however, which could be partly explained by the overlapping range of wavelengths produced by the two light sources. However, the reported intensity of light produced by our 325 nm LED light source is less than 1% of maximum intensity below 310 nm. The higher-spectrum wavelengths of the 325 nm LED along with its minimal reduction in ^{PC}K photo-splitting could therefore make it less toxic and more appropriate for cell-based assays.

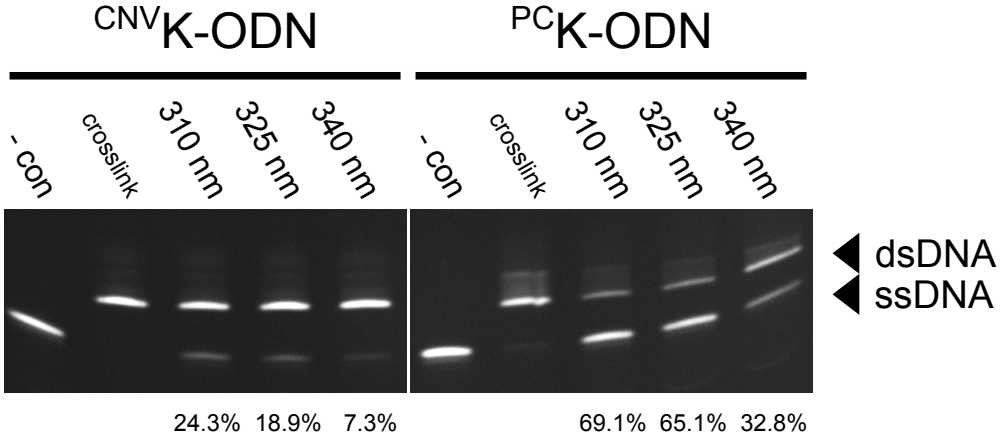


Figure 4.32: Denaturing PAGE analysis of ^{CNV}K and ^{PC}K photo-splitting at alternative wavelengths. ^{CNV}K-ODN and ^{PC}K-ODN were annealed with comp-ODN^{FAM} (- con) and irradiated with 30 s 365 nm light (crosslink). Aliquots were taken of each duplex and irradiated with 30 s of either 310, 325, or 340 nm light. Samples were run on gel and visualized by FAM. Arrows to the right of the gel indicate bands which correspond to non-crosslinked (ssDNA) and crosslinked (dsDNA) comp-ODN^{FAM}. The % of photo-split material is listed under relevant lanes, as calculated with FIJI.

4.6.3) ^{PC}K blocker *in vitro* Photo-Regulation

To compare the activity of ^{PC}K and ^{CNV}K photo-regulation, B1.2 was resynthesized with a ^{PC}K residue in place of a ^{CNV}K residue, yielding ^{PC}K-B1.2. Similar to the experiment previously illustrated in **Figure 4.16a**, IRES-mCherry mRNA was pooled with a 5x molar excess of ^{PC}K-B1.2. The blocker was then annealed with mRNA, crosslinked with 30 s 365 nm light at 4 °C, and photo-split for 30 s at 37 °C with one of 310, 325, or 340 nm irradiation. Aliquots were taken after each step and transfected into HEK293 cells in triplicate. The average mCherry fluorescence of each transfected cell population was normalized to a positive control, as reported in **Figure 4.33**.

Excitingly, we observed that ^{PC}K-B1.2 was able to inhibit translation to a similar degree as we had previously achieved with B1.2. Reversing the crosslink with 312 nm light was able

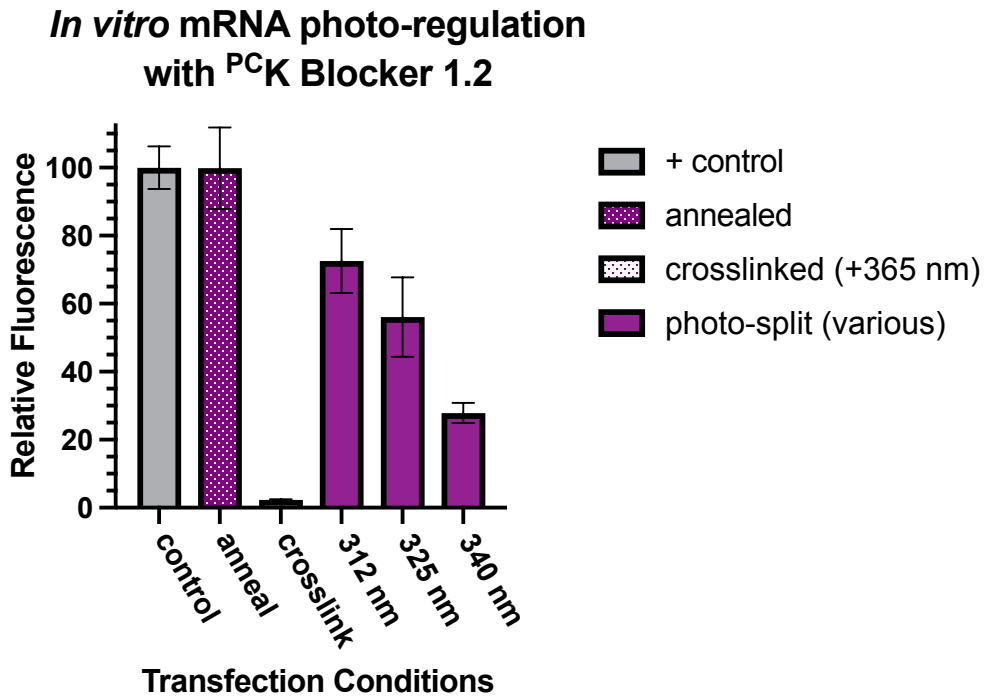


Figure 4.33: Relative mCherry fluorescence from ^{PC}K-B1.2 photo-regulation. The background fluorescence of each cell population was removed by subtracting the average fluorescence of mock-transfected cells with propagation of error. The average fluorescence was then normalized to the positive control with propagation of error, which was set to 100%. Error bars = SD.

to restore 72.5% SD 9.4% mRNA expression compared to the positive control. This represents a significant improvement in light-based activation compared to the original ^{CNV}K B1.2 and is consistent with our *in vitro* characterization of ^{PC}K activity. Here again, attempting to photo-split ^{PC}K crosslinks with higher wavelength light (325, 340 nm) resulted in a lower activation rate. However, both of these conditions still resulted in greater mRNA expression than we achieved with our previous blocker designs.

Designing a translational blocker with a ^{PC}K nucleoside resulted in a strategy with relatively high activation rates while maintaining the robust silencing we had previously achieved with our 1st generation blockers. Unfortunately, due to a series of delays, we were unable to further characterize the ^{PC}K blocker behavior in cells. However, the initial data presented in **Figure 4.33** is promising however and demonstrates that ^{PC}K-blockers can form the basis of a very effective mRNA regulation strategy.

4.7 Photo-Regulation of Exotic mRNA Species

At this project's inception, we envisaged that carbazole-blockers could produce a more robust translation off-state than conventional mRNA regulation strategies while maintaining a high dynamic activation range. Through continuous design improvement and then the incorporation of ^{PC}K residues, we were able to accomplish this goal. The carbazole-blocker strategy offers additional advantages however, such as providing a platform to tag mRNA with chemical modifications. This benefit was briefly illustrated in Section 4.3.3, in which the presence of primed mRNA in cells was tracked with FAM fluorescence. Another unique advantage of our blocker strategy is that it does not require pre-defined mRNA sequences, such as 5′ caps or structure motifs, in order to function. This potentially allows our strategy to be implemented in a wider variety of mRNA than existing photo-regulation strategies.

Concurrent with the blocker optimization work that took place in Sections 4.5 and 4.6, we also began to explore whether ^{CNV}K-blockers would be compatible with alternative mRNA species.

4.7.1) Photo-Regulation of Polycistronic mRNA

The first alternative species of mRNA we examined were polycistronic mRNA: mRNA with more than one start codon and ORF. Synthetic polycistronic mRNAs are typically produced by placing one ORF under the control of a 5′ cap and alternate ORFs under the control of IRES elements and allow multiple genes to be expressed from a single mRNA independently of each other. Polycistronic mRNA are an incredibly powerful for expressing multiple genes in cells while ensuring they are transfected in equal proportion. This is especially relevant for applications such as cell reprogramming, in which several transcription factors need to be over-expressed to alter cell lineage [294, 295]. Our goal was to demonstrate that carbazole-blockers could regulate the expression of single cistrons in a polycistronic mRNA regardless of their primary sequence order. Activating the translation of cistrons independently of each other would allow for highly precise spatial-temporal control of multiple genes without having to perform multiple mRNA transfections. This would additionally highlight the site-specific nature of ^{CNV}K-blocker translation inhibition.

To this end, we cloned two new mRNA templates into the pRSET vector (Appendix 5) for the expression of mCherry and GFP. The two templates differed only in the order of their ORFs and in which feature drove ORF translation. In mRNA prepared from pRSET-mCherry-IRES-EGFP **(Figure 4.34a)**, mCherry translation is promoted by a 5′ cap and EGFP translation by IRES element. For mRNA prepared from pRSET-EGFP-IRES-mCherry **(Figure 4.34b)**, EGFP translation is promoted by a 5′ cap and mCherry translation by IRES

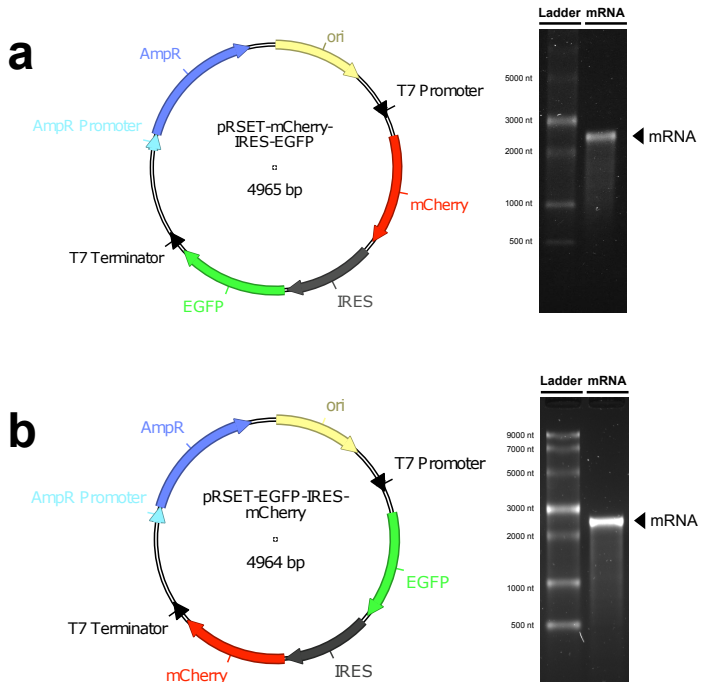


Figure 4.34a: Diagram of pRSET-mCherry-IRES-EGFP T7 IVT template and an agarose gel analysis of the mRNA produced through its transcription. Expected mRNA length = 2391 nt. **Figure 4.34b:** Diagram of pRSET-EGFP-IRES-mCherry T7 IVT template and an agarose gel analysis of the mRNA produced through its transcription. Expected mRNA length = 2390 nt.

element. The two species are abbreviated according to the order of their elements: mCherry-IRES-EGFP (MIG) and EGFP-IRES-mCherry (GIM) mRNA.

To test if ^{CNV}K-blockers can specifically inhibit target cistrons, GIM and MIG mRNA were individually pooled with a 5x molar excess of B1.2, annealed, and crosslinked with 30 s 365 nm irradiation at 4 °C. The annealed and crosslinked mRNA were transfected into HEK293 cells in triplicate and analyzed by flow cytometry after 24 h to observe mCherry and EGFP fluorescence. Representative flow plots were selected for MIG **(Figure 4.35a)** and GIM **(Figure 4.35b)** mRNA-transfected populations. For both mRNAs, we observed a robust EGFP and mCherry fluorescence signal after transfection. For both fluorescent proteins, 5′ cap-driven translation appeared to generate a relatively larger signal compared to IRES-driven translation. This made the expression of the first ORF more pronounced than the

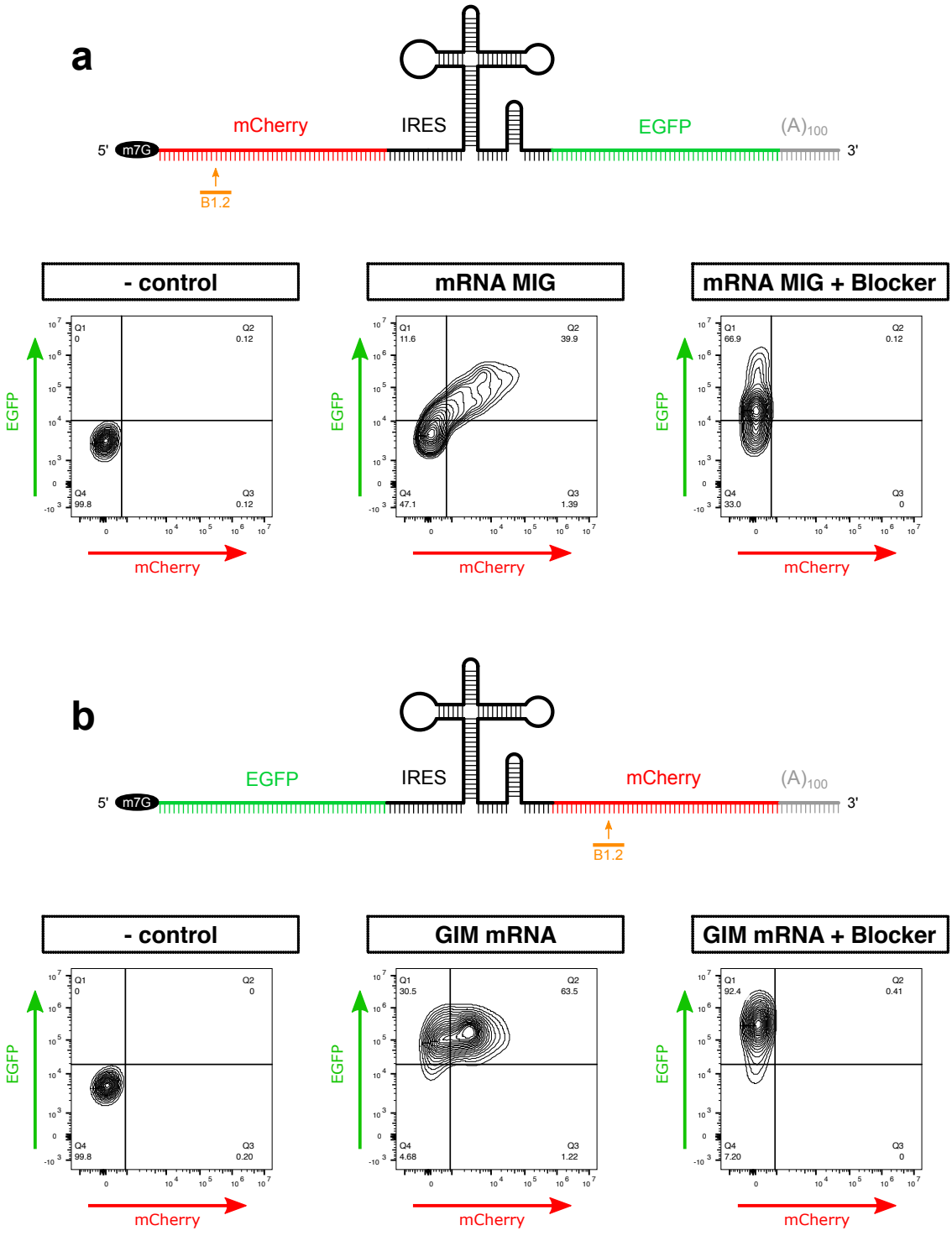


Figure 4.35: mRNA illustration and flow cytometry plots for **(a)** mCherry-IRES-EGFP mRNA and **(b)** EGFP-IRES-mCherry mRNA. In both mRNA illustrations, the order of cistrons is denoted along with the driver of translation (7-methylguanosine cap or IRES). B1.2 and arrow (orange) denote which cistron will be inactivated. Flow cytometry plots display EGFP and mCherry fluorescence for singlet HEK293 cells. For each mRNA, a non-transfected control population, mRNA transfected population, and blocker-primed mRNA transfected population is shown. Thresholds for mCherry and EGFP positive cells were set using the negative control. The percent of total cells in each quadrant is displayed.

second for each mRNA. Upon crosslinking with B1.2, we observed a near-total knockdown in mCherry expression for both mRNA species, regardless of the order of ORFs. Additionally, EGFP fluorescence was uninhibited in either case, indicating that translation inhibition was sequence and cistron-specific.

The ability to individually regulate the ORFs of a polycistronic mRNA could be a valuable research tool that we intend to further explore. We are also interested in applying ^{CNV}K-blockers to self-amplifying RNA (saRNA), which are similar to polycistronic mRNA in that they have distinct RNA polymerase gene and replicable cargo components. By selectively inhibiting the translation and replication of the saRNA cargo gene while allowing the translation of its RNA polymerase gene, ^{CNV}K-blockers could be used to produce burst kinetics upon light-activation.

4.7.2) Photo-Regulation of circRNA

After our experiments with polycistronic mRNA, we next wanted to demonstrate that carbazole-blockers can also effectively regulate circRNA translation. Because translation is driven from an IRES element in circRNA and not a 5′ cap, most common methods of mRNA light-regulation are incompatible with circRNAs. Our ^{CNV}K-blocker method does not require a cap or structured RNA regions, which could potentially make it a valuable tool to study and control the activity of circRNAs.

One of the initial reasons we had for including IRES elements in most of our mRNA sequences was so they could later be ligated in order to convert them into circRNA. To this end, we initially attempted to circularize IRES-mCherry mRNA through self-ligation catalyzed by T4 RNA Ligase 1. However, we were unable to obtain circularized ligation

products (data not shown), likely due to the inherent difficulty of catalyzing intramolecular ligations and the apparent sensitivity of T4 RNA Ligase 1 for certain 5′ substrates which were present in our RNA [296].

Because enzymatic preparation of circRNA was ineffective, we instead decided to implement the self-splicing circRNA system recently optimized by Chen *et al.* [159]. With this method, an RNA produced through *in vitro* transcription is co-transcriptionally circularized by upstream and downstream self-splicing introns without the need for additional enzymes. We acquired a plasmid from the Chang group (circRNA-SynIRES-R25-mNeonGreen [159]) which housed a T7 RNA polymerase template for an mRNA under the control of an engineered IRES, flanked by the two introns required for self-splicing. Using a Gibson assembly strategy, we prepared a new circRNA plasmid housing an IRES-mCherry sequence, as illustrated in **Figure 4.36a**. A set of primers was designed to PCR amplify the T7 transcription template within the plasmid (Appendix 4), and the resulting amplicon was used

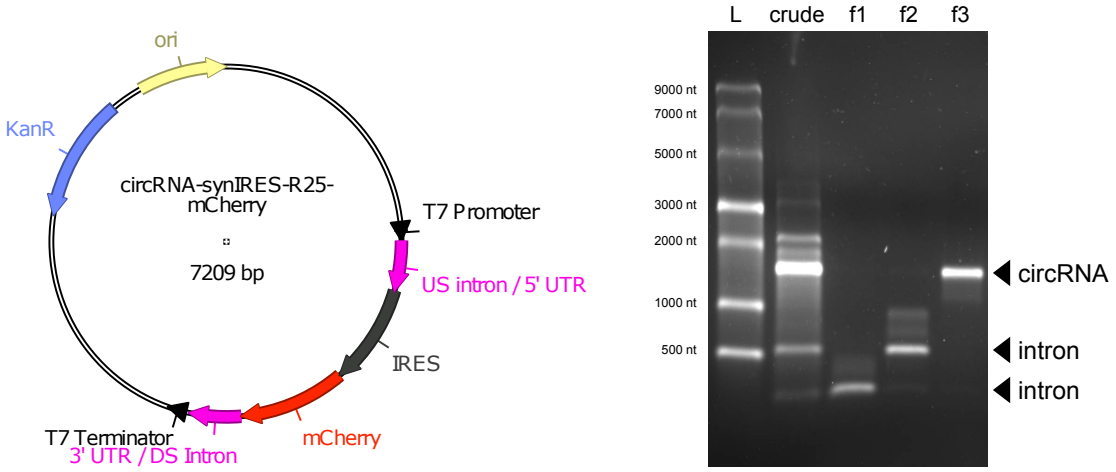


Figure 4.36a: Diagram of circRNA-synIRES-R25-mCherry plasmid used to template circRNA synthesis. **Figure 4.36b:** Agrose gel electrophoresis of circRNA. From left to right, lanes correspond to a ssRNA ladder (L), the crude circRNA transcription reaction (crude), and three fractions isolated with RP-HPLC purification of the crude transcription reaction (f1, f2, f3). Arrows to the right of the gel indicate which bands are presumed to be the circRNA product or the excised introns.

to template circRNA synthesis (Section 6.7.1). The crude transcription reaction was then purified by RP-HPLC (Section 6.3.1), from which we isolated three predominate fractions **(Figure 4.36b).** Based on their size and data supplied by Chen *et al.* [159], we believe that these fractions correspond to the post-splice introns and circRNA product. To confirm its identity, we transfected the presumed circRNA product into HEK293 cells alongside linear IRES-mCherry mRNA in triplicate and assessed their expression by flow cytometry after 24 h **(Figure 4.37)**. The populations transfected with circRNA were 5-fold more fluorescent than those transfected with linear mRNA. This result is consistent with the increase in translation generated from circRNA, which confirmed that the product we isolated was successfully spliced circRNA.

After optimizing our preparation and purification of circRNA, we next attempted to regulate its expression with ^{CNV}K-Blocker B1.2. Similar to earlier experiments, IRES-mCherry mRNA was pooled with a 10x molar excess of B1.2, annealed, and crosslinked. After

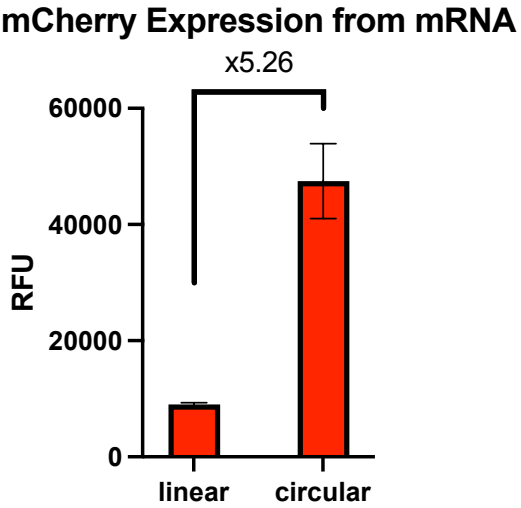


Figure 4.37: Relative fluorescence of HEK293 cells transfected with linear IRES-mCherry mRNA or mCherry circRNA. Values represent the average fluorescence of n=3 transfections, error bars = SD.

crosslinking, an aliquot was removed and subjected to photo-splitting conditions of 30 s 310 nm irradiation at 37 °C *in vitro*. The crosslinked (+365 nm) and photo-split (+310 nm) circRNA was transfected in triplicate to HEK293 cells alongside an untreated positive control circRNA. 24 h after transfection, cell fluorescence was measured by flow cytometry and normalized to the positive control, as shown in **Figure 4.38**. Excitingly, we observed that B1.2 photo-crosslinking was able to produce a robust off-state of >0.5% expression of the positive control. Reversing the crosslink with 310 nm irradiation partially restored expression to 10.8% SD 3.0% of the positive control, which was slightly lower than the activation rates we achieved with B1.2 and linear mRNA. Activation rates can likely be improved by applying the more photo-active ^{PC}K-B1.2, which unfortunately had not been

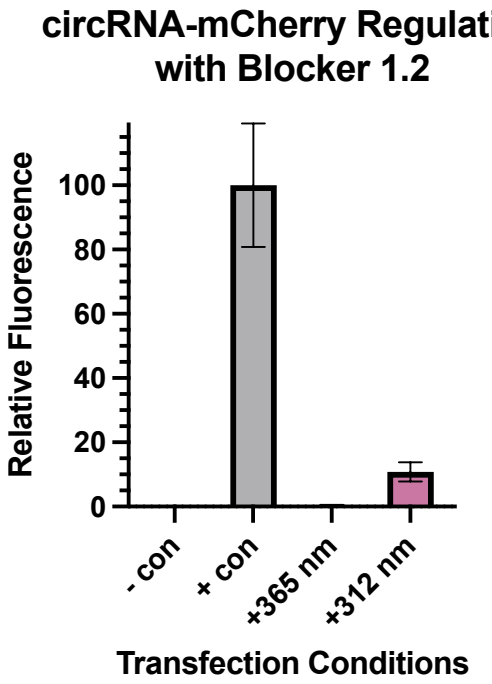


Figure 4.38: Relative mCherry fluorescence from HEK293 cells after transfection with mCherry circRNA. Cells were transfected with either a mock transfection (- con), circRNA (+ con), circRNA primed with B1.2 (+365 nm), or blocker-primed circRNA after photo-activation (+312 nm). Background fluorescence was removed by subtracting the average fluorescence of mock-transfected cells from each population with propagation of error. The average fluorescence of each population was then normalized to the positive control.

synthesized at the time of this work. Regardless, this does illustrate that ^{CNV}K-blockers can be applied to circRNA and represents the first instance of circRNA light-regulation.

4.8 Conclusion and Future Works

Photo-crosslinking is a viable strategy for regulating mRNA translation. We began this project by demonstrating that short 2′OMe ^{CNV}K-blockers are capable of inhibiting translation in a crosslink-dependent manner. The silencing achieved is robust and appears to occur regardless of the target within the mRNA ORF. During the initial characterization of mRNA activation by photo-splitting, we were able to confirm that our strategy fundamentally worked but struggled with low activation rates. An exploration of ^{CNV}K dynamics in physiologically relevant conditions revealed photo-splitting is less favorable than photo-crosslinking in a range of conditions, which had not previously been reported. After exploring strategies to improve the dynamics of ^{CNV}K-blocker photo-crosslinking and photo-splitting, we explored the novel ^{PC}K nucleotide and found that it had a significantly more favorable crosslink/splitting profile than ^{CNV}Ks. Implementing the ^{PC}K nucleotide in our original blocker designs improved its activation rate to above 70% while maintaining its robust silent state.

At certain points in this project, we also succeeded in demonstrating some of the unique advantages of our carbazole-blocker strategy over existing mRNA regulation techniques. Importantly, we were regularly able to silence the expression of mRNA to less than 1% relative to positive controls, which we believe to be a significant improvement compared to existing methods. In applications of mRNA where leaky expression cannot be tolerated, our method appears to be much more suitable than cap-based photo-regulation strategies. We were additionally able to demonstrate that ^{CNV}K-blockers are capable of regulating the

translation of polycistronic and circRNA, with which current photo-regulation strategies are incompatible. Future work on this project will apply ^{CNV}K-blockers to saRNA, another promising vector for RNA therapies and vaccines. Lastly, fluorescently labeled blockers were also used to track the presence of mRNA in cells, which highlights their ability to serve as a platform for tagging mRNA with chemical modifications. More work is needed to explore the full potential of this attribute.

In short, we have developed a novel method for the photo-regulation of mRNA based on crosslinking nucleosides. The next step for this project is to continue characterizing the activity of ^{PC}K-blockers to further validate them as a platform for photo-regulation. Additionally, the data presented in this thesis has only demonstrated photo-regulation on a population level. To highlight the versatility of our method, it will also be advantageous to demonstrate timed or targeted activation between sub-populations. It is unlikely that we will be able to improve the activation rate of ^{PC}K-blockers above 70% without altering the design of the blocker. However, this activation rate is comparable to several existing strategies and is an acceptable trade-off given the incredibly robust silent state our blocker strategy produces. Carbazole-blockers provide several new avenues to control and study diverse mRNA species which we are excited to share with the wider scientific community.

5

Conclusions

The varied physiological roles of oligonucleotides have increasingly been leveraged to produce novel therapeutics and research tools. To overcome bottlenecks in their application, new methods are required to effectively translate advances in nucleic acid chemistry to therapeutic RNAs, and to control their activity with high spatial and temporal precision. The primary goal of this thesis was to explore photo-crosslinking carbazole nucleosides as a platform to manipulate RNA.

In Chapter 2, ^{CNV}K crosslinking was explored as a tool to construct functional Cas9 sgRNAs which are too long to regularly produce by chemical synthesis. By crosslinking smaller RNA precursors to form sgRNAs opens the possibility of extensive chemical modification and library preparation with lower cost and effort. The work in this chapter was begun by characterizing the efficiency of ^{CNV}K crosslinking in dsRNA and minimal-length duplexes, which had not previously been reported. The information gleaned from these experiments helped guide the design of our first ^{CNV}K-sgRNA, which was constructed from a chemically

synthesized ^{CNV}K-crRNA and an IVT tracrRNA. After optimizing the sgRNA crosslinking reaction and demonstrating it could direct Cas9-mediated DNA cleavage *in vitro* and in cells, we proceeded to design a ^{CNV}K-tracrRNA which contained the invariable structural components of the Cas9 gRNA. The ^{CNV}K-tracrRNA effectively crosslinks to an EMX1-targetting crRNA, forming an easily purifiable ^{CNV}K-sgRNA. Side-by-side comparisons demonstrated that the ^{CNV}K-sgRNA outperformed non-crosslinked controls in both *in vitro* digest assays and gene knockouts in a human cell line. We then prepared an extensively modified ^{CNV}K-tracrRNA with 2'OMe residues and PS linkages, which had comparable crosslinking and cleavage activity compared to the non-modified ^{CNV}K-tracrRNA. Lastly, we demonstrated that a library of sgRNAs could easily be prepared by crosslinking various crRNAs to our common ^{CNV}K-tracrRNA scaffold. The resulting ^{CNV}K-sgRNAs had variable activity but were all functional. Further work is needed to accurately quantify the activity of ^{CNV}K-sgRNAs when transfected into cells with Cas9 mRNA, as opposed to Cas9-RNP delivery. Crosslinking also permits several approaches to construct alternative gRNAs, such as the recently described prime-editor gRNAs (pegRNAs). The results of this chapter establish ^{CNV}K crosslinking as an alternative method for the construction of functional, structural RNAs which could be used in place of or in conjunction with enzymatic or chemical ligation strategies.

The work in Chapter 3 aimed to expand upon the sgRNA construction method devised in Chapter 2. To this end, we explored three methods of ligating large dsDNA cargos onto Cas9 gRNAs. The two methods which appeared the most successful required some combination of oligonucleotide modification and enzymatic ligation. In one, gRNAs were click-ligated to DNA hairpins which could then be enzymatically ligated to DNA cargo, producing circular hgRNAs with no exposed termini in the DNA element. In the other, PCR amplification of

DNA cargo which was designed to leave a ssDNA extension allowed direct enzymatic ligation of sgRNA and Cargo. We then attempted to use these DNA/RNA hybrids to create dCas9 transcriptional state reporters (TSRs), which would produce a measurable signal if the underlying target gene were actively transcribed. We designed a TSR cargo based on a split GFP assay, but unfortunately found that the background expression of our cargo was too high to detect additional expression caused by target expression. The work of this chapter did highlight some creative methods of ligating large DNA and RNA molecules which might have application in other areas. While we did not work with ssDNA in this project, direct click-ligation of an alkyne-ssDNA prepared through chemical synthesis with an azide-^{CNV}K-sgRNA might be a more effective strategy moving forward. For future work, the utility of click and crosslink compatible sgRNAs might be better demonstrated with a less conceptually ambitious Cas9 application, such as HDR gene editing with sgRNA/donor template conjugates.

In Chapter 4, rather than constructing RNAs, we used the reversible crosslinking of ^{CNV}K to control the translation of mRNA in cells. Several strategies to control the activity of mRNA have been published, but most suffer from inflexibility and inadequate silencing of translation. We envisaged that a short ^{CNV}K-blocker oligonucleotide could be crosslinked to mRNA ORFs to completely impede the progression of ribosomes. Crosslinking would be performed *in vitro*, and translation then could be activated by photo-splitting after transfection in cells. We were able to demonstrate that short ^{CNV}K-blockers can produce an incredibly robust translational off-state in a crosslink-dependent manner. mRNA translation could subsequently be activated in cells through 310 nm induced ^{CNV}K photo-splitting. Our activation rates were initially low (~20%), which we discovered was due to unfavorable ^{CNV}K photo-splitting rates in physiologically relevant conditions. We attempted to improve the

translation activation rates of our blockers by destabilizing the local blocker/mRNA duplex around the ^{CNV}K residue. This culminated in the design of our 3rd generation ^{CNV}K blockers, which had photo-cleavable linkers to reduce the effective size and T_M of the blocker after crosslinking. The 3rd generation blockers maintained the robust silencing of the 1st generation blockers and improved in-cell activation rates. Despite our efforts, the overall activation rate remained low compared to existing methods of photo-regulation. Because of this, we explored the novel pyranocarbazole (^{PC}K) as an alternative crosslinking agent and found that it had much more favorable photo-splitting properties. With a ^{PC}K-blocker, we were able to achieve near-total silencing of mRNA expression, and a translation activation rate of over 70%. Carbazole-blockers were then applied to regulate the activity of polycistronic and circRNA, which cannot be accomplished by existing photo-regulation strategies. We additionally demonstrated that carbazole-blockers provide an easy platform for tagging mRNA with chemical modifications. Further work needs to be done to characterize the activity of ^{PC}K-blockers, demonstrate their compatibility with exotic mRNA species, and regulate mRNA translation in cells. However, our initial data indicates that ^{PC}K-blockers can form an effective mRNA regulation strategy with robust silent and active states. Overall, this chapter shows that carbazoles are a powerful yet overlooked tool for controlling mRNA activity and have a significant amount of potential to benefit research and therapeutic applications of mRNA.

In summary, the work presented in this thesis shows that carbazole nucleosides are a versatile tool for the manipulation of biofunctional RNAs and lays the foundation for future studies on Cas9 gRNAs and mRNA.

6

Methods and Materials

6.1 Oligonucleotide Synthesis

6.1.1) DNA Synthesis

DNA synthesis was performed on an Applied Biosystems 394 automated DNA/RNA synthesizer using a standard phosphoramidite cycle of detritylation, coupling, capping, and oxidation using TCA (3% in dichloromethane), tetrazole (0.45 M in acetonitrile), Cap A (10% acetic anhydride in 10% lutidine and 80% tetrahydrofuran (THF)) / Cap B (16% N-methylimidazole (NMI) in THF) and iodine (0.02 M in THF, pyridine, and water) respectively on a 0.2 or 1.0 μmol scale. Appropriate solid supports were packed into a twist column (Glen research) for synthesis. β-cyanoethyl protected phosphoramidites (dA-bz, dG-ib, dC-bz, dT, Sigma-Aldrich) were dissolved in anhydrous acetonitrile (0.1 M) immediately prior to use. The coupling time for dA, dC, dG, and dT monomers was 50 sec, and 10 min for modified monomers. Stepwise coupling efficiencies were determined by automated trityl cation conductivity monitoring and were >98% in all cases.

6.1.2) RNA Synthesis

RNA synthesis was performed on an Applied Biosystems 394 automated DNA/RNA synthesizer using a standard phosphoramidite cycle of detritylation, coupling, capping, and oxidation on a 1.0 μmole scale. Appropriate solid supports were packed into a twist column for synthesis.

For the TBDMS method, RNA phosphoramidites (A-tac, C-tac, G-tac, U, Sigma-Aldrich) were 2′-O-TBDMS protected. Monomers were dissolved in anhydrous acetonitrile (0.1 M) immediately prior to use. Coupling, capping, and oxidation reagents were benzylthiotetrazole (0.3 M in acetonitrile, Link Technologies), fast deprotection Cap A (5% t-butylphenoxyacetyl acetic anhydride in THF)/Cap B (16% N-methylimidazole (NMI) in THF) and iodine (0.1 M in THF, pyridine, and water) respectively. The coupling time for all monomers during RNA synthesis was 10 min. Stepwise coupling efficiencies were determined by automated trityl cation conductivity monitoring and in all cases were >97%.

For the TC method, RNA phosphoramidites (A-bz, C-ac, G-ib, U, Sigma-Aldrich) were 2′-O-TC protected. Monomers were dissolved in anhydrous toluene/acetonitrile (1:1 v/v) (0.1 M) immediately prior to use. Coupling, capping, and oxidation reagents were identical to those used in DNA synthesis except a solution of ethylthiotetrazole (0.25 M in acetonitrile, Link Technologies) was used as a coupling reagent. The coupling time for all monomers during RNA synthesis was 3 min. Stepwise coupling efficiencies were determined by automated trityl cation conductivity monitoring and in all cases were >97%.

6.1.3) Modified Monomers and Solid Supports

Monomer	Supplier	Cat. No.
PC Linker Phosphoramidite (PC)	Glen Research	10-4920
Spacer Phosphoramidite 18 (HEG)	Glen Research	10-1918
Spacer C12 CE Phosphoramidite (C12)	Glen Research	10-1928
3-Cyanovinylcarbazole Phosphoramidite (^{CNV}K)	Glen Research	10-4960
5-Octadiynyl-dU CE-Phosphoramidite	Link Technologies	BA 0308
DMT-dPC-OCEP (^{PC}K) **(Figure 6.1)**	Synthesized by Alice Kennett	N/A
2′-OMe-A-CE Phosphoramidite	Glen Research	10-3100
2′-OMe-Ac-C-CE Phosphoramidite	Glen Research	10-3115
2′-OMe-ibu-G-CE Phosphoramidite	Glen Research	10-3120
2′-OMe-U-CE Phosphoramidite	Glen Research	10-3130
Chemical Phosphorylation Reagent (PO_4)	Link Technologies	2101- F100x10
5′-Hexynyl Phosphoramidite	Glen Research	10-1908
5′-Amino-Modifier C12 (Amine-C12)	Glen Research	10-1912
5′-Fluorescein-CE Phosphoramidite (FAM)	Link Technologies	2134-F050

Support	Supplier	Cat. No.
3′-Amino-Modifier C7 CPG 1000 (C7-Amine)	Glen Research	20-2958

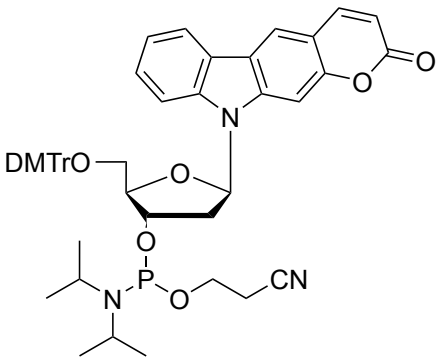


Figure 6.1: Chemical structure DMT-dPC-OCEP (^{PC}K phosphoramidite monomer)

6.2 Oligonucleotide Deprotection

6.2.1) DNA Deprotection

DNA was cleaved from the solid support and deprotected by exposure to a concentrated solution of aqueous ammonia in a sealed vial for 5 h at 55 °C. After drying *in vacuo*, oligonucleotides were dissolved in water and subject to further purification.

6.2.2) TBDMS RNA Deprotection

RNA was cleaved from the solid support and the nucleobases deprotected using concentrated aqueous ammonia: ethanol (3:1 v/v) in a sealed vial for 2 h at 55 °C unless stated otherwise. After cooling to room temperature, the solution was filtered, and then ammonia was removed *in vacuo*. The ammonia-free solution was then freeze-dried, re-dissolved in a 1:1 mixture of dry DMSO (300 μL) and trimethylamine trihydrofluoride (300 μL), and heated for 2.5 h at 65 °C. After cooling down to room temperature, sodium acetate (3 M pH 5.2, 50 μL) and butanol (3 mL) were added, and the RNA was stored for 30 min at -80 °C. The RNA was then pelleted by centrifugation (8,500 rpm, 30 min, 4 °C), the supernatant discarded, and the pellet washed twice with 70% ethanol (750 μL). The pellet was then dried *in vacuo*, dissolved in water, and desalted as described in the Section 7.3.1 before further purification.

6.2.3) 2′-O-TC RNA Deprotection

The solid support was exposed to dry ethylenediamine: toluene (1:1 v/v) for 6 h at room temperature, washed with toluene (3 x 1 mL), and dried using argon. The cleaved RNA was eluted with water and purified further.

6.2.4) Selective β-cyanoethyl Removal

Oligonucleotides bearing primary amines were treated on-column with diethylamine (20% in anhydrous acetonitrile) for 20 min at room temperature. The resin was then washed with acetonitrile (3 x 1 mL) and dried with argon before cleaving from resin and further deprotection.

6.3 Oligonucleotide Purification

6.3.1) Desalting:

Crude desalting was performed on oligonucleotides using one of two commercially available desalting filters: NAP™-10 Columns (Cytivia cat. no. 17085402) were used to desalt oligonucleotides according to the manufacturer's instructions following click-ligation and post-synthetic labeling. Amicon®Ultra-4 0.5 mL centrifugal filters (Merck cat. no. UFC5x) were used to desalt and concentrate oligonucleotide samples following HPLC.

Prior to their application in cells, oligonucleotides were desalted with ethanol precipitation. 0.1 volumes of 3 M sodium acetate (pH 5.2) were added to 1 volume of oligonucleotide suspended in water. 3 volumes of 100% ethanol were added to this suspension, which was then thoroughly mixed and left to precipitate at -20 °C for at least 1 h. The precipitated oligonucleotide was pelleted with centrifugation (13,000 rpm, 10 min, 4 °C) and washed twice with 1 volume 70% ethanol before air drying and resuspension in ddH_20

6.3.2) Short Oligonucleotide RP-HPLC Purification

DNA and some short RNA oligonucleotides were purified using a Gilson HPLC system with ACE® C8 column (10 mm x 250 mm, pore size 100 Å, particle size 10 μm), a gradient of buffer A (0.1 M TEAB, pH 7.5) to buffer B (0.1 M TEAB, pH 7.5 containing 50% v/v

acetonitrile), and flow rate of 4 mL/min. For unmodified oligonucleotides, the gradient was 10–40% buffer B over 20 min. For other oligonucleotides, the gradient was suitably adjusted.

6.3.3) Long Oligonucleotide RP-HPLC Purification

Long oligonucleotides (>100 nt), chemically synthesized RNA, and ^{CNV}K-crosslink products were purified using an Agilent system with Kinetex® C18 column (10 mm x 250 mm, pore size 100 Å, particle size 5 μm), a gradient of buffer A (0.1 M TEAA, pH 7.5) to buffer B (0.1 M TEAA, pH 7.5 50% v/v acetonitrile), and flow rate of 5 mL/min. A gradient of 0%-50% buffer B over 28 min at 55 °C was used in all cases.

6.3.4) mRNA RP-HPLC Purification

IVT mRNA was purified using an Agilent system with DNAPac™ 4 μM column (3 mm x 100 mm, particle size 4 μM), a gradient of buffer A (0.1 M TEAA, pH 7.5) to buffer B (0.1 M TEAA, pH 7.5 50% v/v acetonitrile), and flow rate of 0.5 mL/min. A gradient of 38%-45.9% buffer B over 19 min at 60 °C was used in all cases.

6.3.5) Gel Purification

Gels were prepared and run as described below (Section 6.9). Large DNA fragments (>200 bp) were separated by agarose gel, excised, and purified using a Monarch® DNA Gel Extraction Kit (NEB #T1020) according to the manufacturer's instructions.

Some long oligonucleotides (>90 nt) and click-ligation products were separated by denaturing PAGE, excised, and snap-frozen with liquid nitrogen. The gel was then crushed in up to 5 mL of ddH_2O and incubated overnight at 37 °C, 700 RPM. After incubation, gel

was removed with filtration through cotton, and the supernatant was concentrated and desalted as described above.

6.4 Oligonucleotide Mass Spectrometry

All oligonucleotides were characterized by negative-mode electrospray using an UPLC-MS Waters XEVO G2-QTOF mass spectrometer and an Acquity UPLC system with a BEH C18 1.7 μm column (Waters). A gradient of methanol in TEA and hexafluoroisopropanol (HFIP) was used (buffer A, 8.6 mM TEA, 200 mM HFIP in 5% methanol/water (v/v); buffer B, 20% v/v buffer A in methanol). Buffer B was increased from 0–70% over 7.5 min or 15–30% over 12.5 min for normal oligonucleotides and 50–100% over 7.5 min for very hydrophobic oligonucleotides. The flow rate was set to 0.2 mL/min. Raw data were processed and deconvoluted using the deconvolution software MassLynx v4.1.

6.5 Chemical labeling and ligation of oligonucleotides

6.5.1) Post-Synthetic Oligonucleotide Labeling

Freeze-dried oligonucleotide (20 nmol) was dissolved in $NaHCO_3$ buffer (0.5 M, pH 8.5, 30 μL) and mixed with the NHS ester (200 nmol, 20 μL) dissolved in DMSO. The reaction was then left for 4 h at 25 °C with 750 rpm shaking. After dilution in water to halt the reaction, the labeled oligonucleotides were desalted using a NAP™-10 column prior to RP-HPLC purification.

NHS Ester	Supplier	Cat. No.
6-azidohexanoic acid NHS ester	Synthesized by Thanishta Mungur	n/a
FAM NHS Ester	Lumiprobe	65120

6.5.2) Non-Templated CuAAC

3′-alkyne RNA (750 pmol in 1.25 µL H_2O) and 5′-azido RNA (500 pmol in 1.25 µL H2O) were mixed with triethylammonium acetate (TEAA) buffer (2 M, pH 7, 1 µL), DMSO (5 µL) and fresh ascorbic acid (125 mM, 1 µL). While degassing the oligonucleotide solution with argon in a 0.5 mL Eppendorf tube, a 55% DMSO solution of CuSO4-THPTA complex (250 mM, 0.5 µL) was added, after which the reaction (final volume = 10 µL) was left for 3 h at room temperature. After incubation, the sample was diluted with water to halt the reaction and desalted using a NAP^{TM}-5 column (G.E. Healthcare cat. no. GE17-0853-01) prior to the purification of the ligation product.

6.6 Cloning and Plasmid Preparation

6.6.1) Plasmid Construction

All plasmids constructed during this project were prepared with Gibson assembly [297]. A description of the vector, insert sources, and PCR amplicons used for each assembly can be found in (Appendix 4). Prior to assembly, linearized vector and inserts would be purified using Monarch® PCR & DNA Cleanup Kit (5 µg) according to the manufacturer's protocol (NEB cat. no. T1030S). Assembly reactions consisted of 50 ng vector, x3 molar equivalent of each insert, 5 µL Gibson Assembly® Master Mix (NEB cat. no. E2611), and water (final volume = 10 µL). The reaction was incubated for 15 min at 50 °C before transferring onto ice until transformation.

6.6.2) Transformation

A 50 µL aliquot of NEB 5-alpha Competent *E. coli* cells (NEB cat. no. C2987H) was thawed on ice before the addition of 2 µL of a Gibson assembly reaction or 5 ng of a purified plasmid. After mixing plasmid and bacteria by gentle flicking, the aliquot was returned to ice

for 30 min prior to heat-shock transformation by incubation for 30 sec at 42 °C followed by 2 min on ice. Cells were then diluted in 950 µL SOC media (supplied with cells) and incubated for up to 1 h at 37 °C, 250 rpm shaking. 20 or 200 µL of the cell suspension was spread onto an LB agar plate with appropriate selection agent (ampicillin (100 µg/mL) or kanamycin (50 µg/mL)), which was then incubated for 16 h at 37 °C to allow colony growth.

6.6.3) Colony PCR

For plasmids assembled with pRSET vector, transformant colonies were picked and suspended in 10 µL sterile water. A PCR master mix was prepared with primers 9395 (5 µL, 10 µM), 9396 (5 µL, 10 µM), water (36 µL), and Q5® High-Fidelity 2X Master Mix (50 µL, NEB cat. no. M0492L). The PCR master mix was then split into 9 µL aliquots, to which 1 µL of bacterial suspension was added. The PCRs were then heated to 98°C for 3 min to lyse bacterial cells, followed by 35 cycles of denaturation (30 sec, 98°C), primer annealing (30 sec, 66°C), and primer extension (30 sec per 1 kb of amplicon, 72°C), followed by a final extension step (3 min, 72°C). Amplicons were then analyzed with 0.8% agarose gel to identify successfully assembled plasmids.

6.6.4) Plasmid Isolation

Liquid cultures of transformants were prepared by inoculating 5 mL LB broth containing 100 µg/mL ampicillin with either single colonies from LB agar selection plates or suspended colonies used to template colony PCR. Cultures were incubated for 16 h at 37°C, 250 rpm. After incubation, a stock of transformants was prepared by mixing a 500 µL aliquot of cell culture 1:1 with 50% glycerol, which was then stored at -80°C. The remaining cell culture was pelleted by centrifuging for 15 min, 8,500 rpm at RT. Plasmids were extracted from cell pellets using the QIAprep® Spin Miniprep Kit (Qiagen cat. no. 27104) according to the

manufacturer's protocol. The concentration of the purified plasmid DNA was determined using a NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific).

For larger-scale plasmid preparations, liquid bacteria cultures were prepared as described above with 250 mL LB, 100 μg/mL ampicillin inoculated from glycerol bacteria stocks. After incubation and pelleting cells, plasmid was extracted using QIAGEN Plasmid Maxi Kit (Qiagen cat. no. 12162) according to the manufacturer's protocol.

6.7 *In vitro* Transcribed RNAs and mRNA Processing

6.7.1) *In vitro* Transcription

Transcription reactions were performed with HiScribe™ T7 High Yield RNA Synthesis Kit (NEB cat. no. E2040S). Templates for short RNAs (<120 nt) were prepared by annealing a negative-sense DNA template (1 nmol) and a positive-sense T7 promoter sequence (1 nmol) in water (200 μL). 10 pmol of the dsDNA was used to template a 20 μL reaction prepared according to the manufacturer's instructions, which would be incubated for 16 h at 37 °C. To prepare templates for mRNA synthesis, PCR products were prepared from relevant plasmids with primers 9395 and 9396 and were purified using Monarch® PCR & DNA Cleanup Kit (5 μg) according to the manufacturer's protocol (NEB cat. no. T1030S). 0.5 μg of PCR product was used to template a 20 μL reaction prepared according to the manufacturer's instructions, which would be incubated for 4 h at 37 °C. Transcription reactions were halted by the addition of 1 μL DNAse 1 (provided with kit) and were then purified with a Monarch® RNA Cleanup Kit (NEB cat. no. T2030L or T2050L) according to the manufacturer's protocol.

6.7.2) RNA Poly(A)-Tail Addition

10 µg of mRNA prepared from IVT was mixed with ATP (2 µL, 10 mM), NEB *E. coli* poly(A) polymerase buffer (2 µL, x10, supplied with enzyme), *E. coli* poly(A) polymerase (1 µL, NEB cat. no. M0276), and water (to 20 µL). The reaction was incubated for 30 min at 37 °C before purifying with Monarch® RNA Cleanup Kit according to the manufacturer's protocol.

6.7.3) 7-Methylguanylate 5′ RNA Capping

10 µg of mRNA prepared from IVT or poly(A) tail addition reaction was mixed with GTP (1 µL, 10 mM), S-adenosylmethionine (1 µL, 2 mM), Vaccinia capping buffer (2 µL, x10, supplied with enzyme), Vaccinia capping enzyme (1 µL, NEB cat. no. M2080S), and water (to 20 µL). The reaction was incubated for 30 min at 37 °C before purifying with Monarch® RNA Cleanup Kit according to the manufacturer's protocol.

6.8 ^{CNV}K and ^{PC}K Photo-Crosslinking and Photo-Splitting

6.8.1) *In vitro* Photo-Crosslinking

A solution of up to 100 µL was prepared with 100 mM NaCl, 10 mM $MgCl_2$, 10 Tris-HCL, pH 7.2, and between a 1:1.5 to 1:10 molar of ^{CNV}K/^{PC}K-strand to non-carbazole strand. Reactions were annealed by incubating for 1 min at either 60 °C (short oligonucleotide) or 75 °C (mRNA) and then cooling at a rate of 2 °C/min until 4 °C was reached. The oligonucleotide solution was then transferred to a PCR tube cooler rack chilled to 4 °C and irradiated with a 1350 mW 365 nm LED light source (Thor Labs cat. no. M365LP1) operated at 1700 mA. Where specified, ^{PC}K oligonucleotides were crosslinked with a 750 mW 420 nm LED light source (Thor Labs cat. no. M420L3) operated at 1000 mA.

6.8.2) *In vitro* Photo-Splitting

Crosslinked oligonucleotides in up to 100 µL of any aqueous solution were heated to 37 °C for 3 min in a PCR thermocycler. Samples were then held at 37 °C and irradiated for up to 1 min with a 25 mW 310 nm LED light source (Thorlabs cat. no. M310L3) operated at 650 mA. Where specified, crosslinks with alternatively photo-split with a 25 mW 325 nm LED light source (Thorlabs cat. no. M325L5) or a 25 mW 340 nm LED light source (Thorlabs cat. no. M340L5).

6.9 Gel Electrophoresis

6.9.1) Agarose Gel Electrophoresis

Agarose gels were prepared by melting between 0.65-2.0% w/v of agarose (Lonza cat. no. 50002) in 1x TBE buffer (100 mM Tris-base, 100 mM boric acid, 2 mM EDTA, pH 8.3). While cooling, SYBR™ Gold Stain (0.5x concentration, Thermo Fisher Scientific cat. no. S11494) was mixed into the gel, which was then poured into a cast and allowed to solidify. Before being loaded onto the gel, DNA samples were mixed with purple gel loading dye (NEB cat. no. B7024S). RNA samples were mixed with 95% formamide, 18 mM EDTA, and 0.025% SDS, xylene cyanol, and bromophenol blue before heating for 3 min at 95 °C and then cooling to RT. Gels were run at 60-100 V for 45-90 min in 1x TBE buffer. Gels were visualized using a Syngene G:Box imager unless otherwise stated.

6.9.2) Native PAGE

Native polyacrylamide gels (1x TBE) were prepared according to the recipe below:

ddH2O	to 70 mL
40% acrylamide:bis-acrylamide (29:1)	15-25 mL
10x TBE (1 M Tris-base, 1 M boric acid, 20 mM EDTA, pH 8.3)	7 mL

Polymerization was initiated by the addition of 560 µL 10% ammonium persulfate (APS) and 56 µL TEMED (dimension of gel = W x D x H = 18 x 0.2 x 24.4 cm). After the gel

solidified, sample loading wells were washed with 1x TBE. Oligonucleotide samples were mixed 1:1 with 50% v/v glycerol. Samples were then loaded onto the gel and run for 2-4 h, 300 V, at 4 °C.

6.9.3) Denaturing PAGE

Denaturing polyacrylamide gels (1x TBE, 7 M urea) were prepared according to the recipe below:

Urea	29.4 g
ddH2O	to 70 mL
40% acrylamide:bis-acrylamide (29:1)	15-25 mL
10x TBE (1 M Tris-base, 1 M boric acid, 20 mM EDTA, pH 8.3)	7 mL

Polymerization was initiated by the addition of 560 µL 10% APS and 56 µL TEMED (dimension of gel = W x D x H = 18 x 0.2 x 24.4 cm). After the gel solidified, sample loading wells were washed with 1x TBE. Oligonucleotide samples were mixed 1:1 with a solution of 95% formamide, 18 mM EDTA, and 0.025% SDS, xylene cyanol, and bromophenol blue before heating for 3 min at 95 °C and then cooling to RT. Samples were then loaded onto the gel and run for 2-4 h, 300-400 V, at RT. Gels were analyzed by staining with SYBR™ Gold Stain (0.5x concentration) and then were visualized using a Syngene G:Box imager.

6.10 *In vitro* Cas9 Assays

6.10.1) *In vitro* Cas9 Digestion Assay

Cas9 (1 µL, 1 µM, NEB cat. no. M0386T) was mixed with an appropriate gRNA (1 µL, 1 µM), NEBuffer 3.1 (1 µL, NEB cat. no. B7203), and H_2O (7 µL) (final volume = 10 µL), which were then incubated for 10 min at 37 °C to form Cas9-RNP complex. An aliquot of this Cas9-RNP solution (1 µL, 0.1 µM) was then mixed with DNA substrate (1 µL, 0.1 µM), NEBuffer 3.1 (0.9 µL), and H_2O (7.1 µL) (final volume = 10 µL, 100 mM NaCl, 50 mM Tris-

HCl, 10 mM $MgCl_2$, 100 µg/ml BSA), which were incubated for 1 h at 37 °C. DNA digestion was halted with the addition of Proteinase K (1 µL, 0.8 U/µL, NEB cat. no. P8107S) and incubated for another 15 min at 37 °C.

6.10.2) *In vitro* Cas9 Binding Assay

A 1 µM solution of dCas9-RNP was prepared by mixing dCas9 (0.5 µL, 20 µM, NEB cat. no. M0652T) and appropriate gRNA (2 µL, 5 µM) with NEBuffer 3.1 (1 µL), and H_2O (6.5 µL) (final volume = 10 µL), which were then incubated for 37 °C to form dCas9-RNP complex. 1-2 µL of dCas9-RNP complex was then applied to 1 pmol of a FAM-labeled DNA substrate in 20 µL of 100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl2, and 100 µg/ml BSA. The binding reaction was incubated for 30 min at 37 °C before loading onto an 8.5% Native PAGE and run at 200 V for 2 h. Gels were imaged with an Amersham Typhoon imager using the Cy2 channel.

6.11 Mammalian Cell Culture

6.11.1) Cell Culture

Human embryonic kidney (HEK) cell line HEK293 was cultured in DMEM (4.5 g/L D-Glucose, - pyruvate, GlutaMAX™, phenol red, Thermo Fisher Scientific cat. no. 10566016) supplemented with 10% v/v FBS (Thermo Fisher Scientific cat. no. A4736301) in a humidified incubator at 37 °C with 5% CO_2.

6.11.2) Cas9-RNP Electroporation

Electroporation of Cas9-RNP was performed using a Neon™ Transfection System (Thermo Fisher Scientific) with reagents from a Neon™ Transfection System 10 µL Kit (Thermo Fisher Scientific cat. no. MPK1025), according to the manufacturer's protocol. HEK293 cells were grown to between 70-90% confluency in a T75 flask and were harvested by trypsinization. 1.5 x 10^5 cells per transfection condition, plus x1.2 excess for ease of pipetting, were transferred to a sterile tube and pelleted by centrifugation for 5 min, 300 g. Cells were then resuspended in an equal volume of DBPS and pelleted again to remove residual culture media. The washed cell pellet was resuspended in Buffer R to a concentration of 1.5 x 10^5 cells / 5 µL.

While the cells were pelleting, Cas9-RNP solution was prepared by mixing gRNA (0.75 µL, 10 µM) and TrueCut™ Cas9 Protein v2 (1 µL, 1.25 µg/µL, Thermo Fisher Scientific cat. no. A36499) in 3.25 µL Buffer R (final volume = 5 µL). The Cas9 mixture was incubated for at least 10 min at RT to allow RNP formation.

After incubation, 5 µL of cell suspension was mixed into the Cas9-RNP solution (final volume = 10). The suspension (7.5 pmol Cas9-RNP, 1.5 x 10^5 cells) was electroporated with 1150 V, 20 ms, 2 pulses and then immediately transferred into 0.5 mL DMEM + 10% v/v FBS. Cells were incubated for 48 h before characterizing editing efficiency.

6.11.3) dCas9 RNP Lipofection

Cas9 lipofection was performed with Lipofectamine™ CRISPRMAX™ transfection reagent (Thermo Fisher Scientific cat. no. CMAX00001). Cells were seeded to be 70% confluent at the time of transfection in a 96-well plate. On the day of transfection, cell culture media was replaced with 50 µL Opti-MEM. Between 0.125-0.5 pmol of dCas9 (NEB cat. no. M0652T)

was mixed with an equimolar amount of gRNA, Cas9 Plus™ Reagent (supplied with lipofection reagent), and Opti-MEM (to 5 µL). Separately, 0.3 µL of CRISPRMAX™ was diluted in 5 µL Opti-MEM. The two solutions were mixed and incubated at RT for 10 min before applying to cells. Lipofection was incubated with cells for 6 hours before the media was replaced with DMEM + serum.

6.11.4) mRNA Lipofection

1.0 x 10^4 cells/well were seeded into a 96 well plate which had been pretreated with 50 µL poly-D-lysine (Thermo Fisher Scientific cat. no. A3890401) for 5 min. 24 h after seeding, cell culture media was replaced with Opti-MEM™ reduced serum media (50 µL, Thermo Fisher Scientific cat. no. 31985062). For each transfection reaction, mRNA (1 pmol) was mixed with Lipofectamine 3000 (0.15 µL, Thermo Fisher Scientific cat. no. L3000015), P3000 reagent (0.33 µL, supplied with Lipofectamine 3000) and Opti-MEM™ (to 10 µL) according to the manufacturer's protocol. After incubating at RT for 10 min, the transfection mixture was applied to cells. Cells were incubated with transfection reagent for 24 h at 37 °C before analysis unless otherwise stated.

6.11.5) Plasmid Lipofection

1.0 x 10^4 cells/well were seeded into a 96-well plate which had been pretreated with 50 µL poly-D-lysine for 5 min. 48-72 h after seeding, cell culture media was replaced with 50 µL Opti-MEM™**.** For each transfection, plasmid DNA (50 ng) was mixed with Lipofectamine 3000, p3000 (0.15 µL), P3000 reagent (0.2 µL), and Opti-MEM™ (to 10 µL) according to the manufacturer's protocol. After incubating at RT for 10 min, the transfection mixture was applied to cells. Media was exchanged with full DMEM after 6 h to remove excess

transfection reagent. Cells were incubated for another 42 h at 37 °C before analysis unless otherwise stated.

6.11.6) UV Irradiation

To minimize UV spill-over between separated cell culture wells during irradiation, all UV cell experiments were performed in a black-wall 96 well plate (Merck cat. no. CLS9102) with cells cultured in non-adjacent wells. The cover of the plate was measured with a UV/vis spectrophotometer to confirm it was translucent at 310 nm. The cover was then coated in a metallic film, and a second blacked-out cover was added to the bottom of the plate, as illustrated below. Lastly, a box stand was constructed to both house the 312 nm lamp and reduce its aperture to one well. For any UV-irradiation experiment, a 25 mW 310 nm LED light source operated at 650 mA was used unless stated otherwise. Experimental procedure is illustrated in **Figure 6.2**.

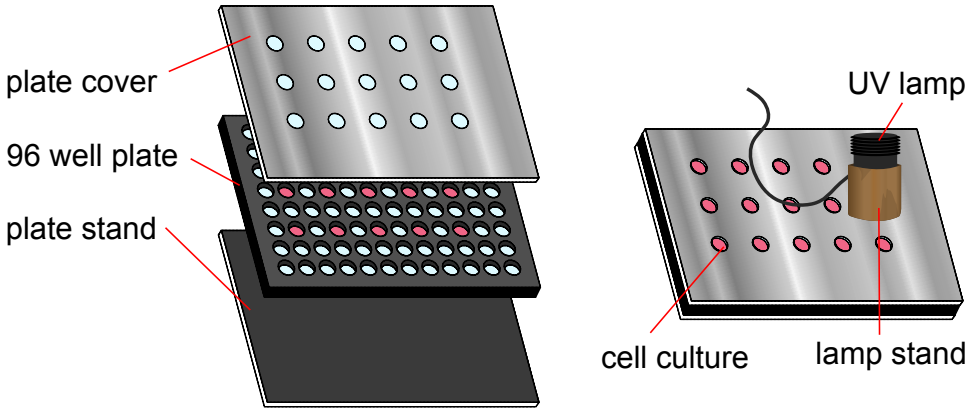


Figure 6.2: Illustration of the steps taken to reduce UV spill-over between wells during photo-irradiation experiments.

6.11.7) Flow Cytometry

Prior to flow cytometry, HEK293 cells were trypsinized for 5 min and then mixed with flow cytometry buffer (DPBS, 2% FBS) (2:1 volumes flow buffer: trypsin reagent) to halt digestion. Cells were gently pipetted to dislodge clusters. All flow cytometry was performed using a CytoFLEX LX (Beckman Coulter). HEK293 cells were gated from debris and other material using FSC-A/SSC-A plots. Single cells were gated with FSC-A/FSC-H plots. An example of the gating strategies used to identify singlet HEK293 cells is shown in **Figure 6.3**. EGFP and FAM fluorescence was measured with 488 nm laser and 525/40 BP filter. mCherry fluorescence was measured with 561 nm laser and 610/20 BP filter. Flow cytometry results were analyzed using FlowJo™ v10.8 Software (BD Life Sciences).

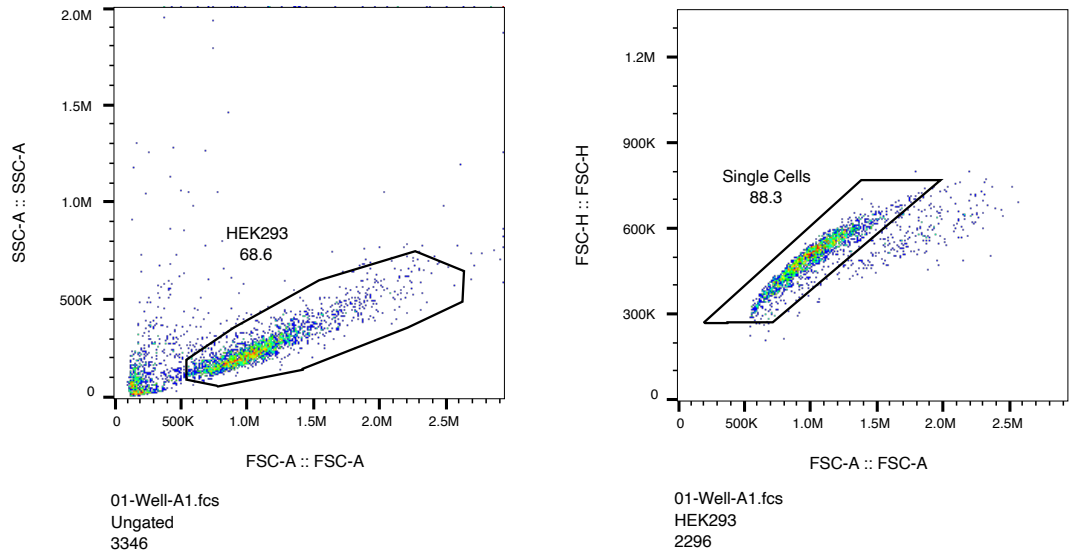


Figure 6.3: Gating strategy used to identify singlet HEK293 cells. FSC-A/SSC-A is used to identify cells from debris, and FSC-A/FSC-H is used to separate singlet cells from clumps. The population analyzed above was treated with a mock transfection (Lipofectamine 3000, no DNA/RNA).

6.11.8) Microscopy

All microscopy images were taken with an EVOS™ M7000 microscope (Thermo Fisher Scientific cat. no. AMF7000).

6.11.9) Cell Viability Assay

Cell viability was determined using the WST-8 reduction assay. 1.0 x 10^4 cells/well were seeded into a 96-well plate which had been pretreated with 50 µL poly-D-lysine for 5 min. 48 h after seeding, wells were irradiated with 312 nm light for various lengths of time. 2 h before a viability reading, cell media was replaced with 100 µL DPBS supplemented with 10% v/v WST-8 (Merck cat. no. 96992). After incubating for 2 h at 37°C, the WST-8 containing cell media was transferred to a clean 96-well plate. The A_{450} of cell media was measured using a CLARIOstar® (BMG Labtech) plate reader to determine cell viability.

6.11.10) Genomic DNA Extraction

Genomic DNA was extracted from cells by incubating in a crude lysis buffer. 1 L of buffer was prepared by dissolving Tris (1.211 g, 10 mM), KCl (3.728 g, 50 mM), $MgCl_2$ hexahydrate (0.304 g, 1.5 mM), Triton X-100 (4.5 mL, 0.45%), and Tween-20 (4.5 mL, 0.45%) in Milli-Q water (final volume 1 L), which was then adjusted to pH 8.5 with HCl (2 M). Before applying to cells, the lysis buffer was supplemented with Proteinase K (NEB, cat. no. P8107S) to a final concentration of 20 mU/µL.

Cells were harvested by incubating in 0.5 mL 0.25% Trypsin-EDTA solution (Thermo Fisher Scientific cat. no. 25200056) for 3 min at 37 °C, followed by the addition of 0.5 mL DMEM + 10% FBS and vigorous pipetting. Cell suspensions were transferred to a sterile 1.5 mL Eppendorf tube, pelleted by centrifugation (100 rcf, 5 min, 4 °C), and aspirated of media. Cell pellets were then resuspended in crude lysis (100 µL buffer / 1 x 10^5 cells) and incubated for 4-12 h, 50 °C, 700 rpm. After incubation, samples were held for 15 min, 95 °C, 700 rpm to heat-inactivate Proteinase K. The concentrations of the genome extracts were approximated using a NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific).

6.11.11) Sanger Sequencing

Sequencing was performed by Source Bioscience using either user-supplied (for gene edit sequencing)(Appendix 4) or stock primers (for plasmid sequencing).

6.12 Open-Source Programs

6.12.1) Chimera X

Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases [298].

6.12.2) TIDE

Analysis of Sanger sequencing reads for gene edits was performed with TIDE, developed by the Bas van Steensel lab at the Netherlands Cancer Institute [204].

6.12.3) FIJI

Gel and microscopy image analysis was performed with Fiji [299].

Appendix 1

Oligonucleotides Introduced in Chapter 2

Name: As referred to in this thesis.

Lab ID: Unique identifier for the oligonucleotide used within our lab.

Type: Refers to the species of oligonucleotide and backbone modifications, which could be DNA, RNA, 2′OMe, or a combination of 2′OMe and phosphorothioate modifications (MPS). 2′OMe bases are identified in **bold text**, and phosphorothioate linkages are underlined.

Source: Refers to how the oligonucleotide was synthesized and can be from one of AB 394 (chemically synthesized in our lab by Prof. Afaf El-Sagheer), IVT (produced through *in vitro* transcription), click ligation, or ^{CNV}K crosslinking. Alternatively, some oligonucleotides were purchased from ATD Bio and IDT.

Sequence: Composed of the universal single letter codes for oligonucleotides. Chemical modifications are listed as named in Section 6.1.3

Calcd and **Found**: Refer to the calculated and observed mass of each oligonucleotide.

Appendix Table 1: Oligonucleotides Introduced in Section 2.3

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
^{CNV}K-ODN	8840	DNA	ATD Bio	CCGGTGA-^{CNV}K-CAGCTCCTC	5213	5212
^{CNV}K-ORN	8841	RNA	ATD Bio	CCGGUGA-^{CNV}K-CAGCUCCUC	5427	5429

comp-ODN	8842	DNA	ATD Bio	GAGGAGCTGTTCACCGG	5251	5252
comp-ORN	8843	RNA	ATD Bio	GAGGAGCUGUUCACCGG	5481	5482
comp-ODN^{-4}	9730	DNA	AB 394	CTGTTGACCGG	3348	3350
comp-ODN^{-8}	9732	DNA	AB 394	GTTGACCGG	2755	2757
comp-ODN^{-9}	9733	DNA	AB 394	TTGACCGG	2426	2425
comp-ODN^{-10}	9734	DNA	AB 394	TGACCGG	2121	2120

Appendix Table 2: Oligonucleotides Introduced in Section 2.4

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
^{CNV}K-crRNA	8847	RNA	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAUGA-^{CNV}K-CAGCUC	13670	13670
crRNA	9652	RNA	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAUGAACAGCUC	13604	13603
^{CNV}K-2′OMe-crRNA	9588	RNA/ 2′OMe	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUA**UGA**-^{CNV}K-**CAGCUC**	13796	13799
2′OMe-crRNA	9589	RNA/ 2′OMe	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUA**UGAACAGCUC**	13742	13746

^{CNV}K-crRNA^{-10}	9890	RNA	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUA-^{CNV}K-G	11138	11140
T7 promoter reverse	9210	DNA	AB 394	TCTAATACGACTCACTATAG	6060	6061
tracrRNA IVT template	8844	DNA	ATD Bio	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATGAACAGCTCCCCTATAGTGAGTCGTATTAGA	5427	5429
tracrRNA	10603	RNA	IVT	GGGAGCUGUUCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU	5251	5252
sgRNA^{EMX1} IVT template	7515	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTCTTCTTCTGCTCGG ACTCCTATAGTGAGTCGTATTAGA	36791	36790
sgRNA^{EMX1}	10603	RNA	IVT	GGGAGCUGUUCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU	5251	5252
^{CNV}K-sgRNA	8847/ 10603	RNA	^{CNV}K Crosslink	See ^{CNV}K-crRNA (8847) and tracrRNA (10603)	38797	38795
^{CNV}K-2′OMe-sgRNA	9588/ 10603	RNA/ 2′OMe	^{CNV}K Crosslink	See ^{CNV}K-2′OMe -crRNA (9588) and tracrRNA (10603)	38923	38925

Appendix Table 3: Oligonucleotides Introduced in Section 2.5

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
crRNA^{EMX1}	9883	RNA	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAUGUUCAGCUC	13558	13559
crRNA$^{EMX1, -5}$	9960	RNA	AB 394	GAGUCCGAGCAGAAGAAGAAGUUUUAGAGCUAUGUUC	11967	11966

^{CNV}K-tracrRNA	9884	RNA	ATD Bio	C8-GAGCUGA-^{CNV}K-CAUAGCAAGUUAAAAUAAGGCUAGUCC GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-Propanol	24233	24235
^{CNV}K-tracrRNA^{9700}	9700	RNA	AB 394	CUCGAC-^{CNV}K-AGUUAGCAAGUUAAAAUAAGGCUAGUCCG UUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	23840	23841
^{CNV}K-sgRNA^{EMX1}	9883/ 9884	RNA	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-tracrRNA (9884)	33791	37793

Appendix Table 4: Oligonucleotides Introduced in Section 2.6

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
MPS-tracrRNA	10109	RNA/ MPS	AB 394	Amine-C12-**GAGCUGAACAUAGCAA**GUU**A**A**A**UA**AGGCUA**GU**CC**GU UA**U**CAA**CUUGAAAAAGUG**G**CACCGAGUCG**<u>GUGCUUUU</u>	25026	25025
^{CNV}K-MPS-tracrRNA	10110	RNA/ MPS	AB 394	Amine-C12-**GAGCUGA-^{CNV}K-CAUAGCAA**GUU**A**A**A**UA**AGGCUA**GU**CC** GUUA**U**CAA**CUUGAAAAAGUG**G**CACCGAGUCG**<u>GUGCUUUU</u>	25076	25079
^{CNV}K-MPS-sgRNA^{EMX1}	10110/ 9883	RNA/ MPS	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-MPS-tracrRNA (10110)	38634	38633

Appendix Table 5: Oligonucleotides Introduced in Section 2.7

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
crRNA^{CCR5}	10104	RNA	AB 394	GUGUUCAUCUUUGGUUUUGUGUUUUAGAGCUAUGUUCAGCUC	13297	13299
crRNA^{HBB}	10105	RNA	AB 394	CUUGCCCCACAGGGCAGUAAGUUUUAGAGCUAUGUUCAGCUC	13384	13385

crRNA^{IL2RG}	10106	RNA	AB 394	UGGUAAUGAUGGCUUCAACAGUUUUAGAGCUAUGUUCAGCUC	13411	13412
crRNA^{VEGFA}	10107	RNA	AB 394	GGUGAGUGAGUGUGUGCGUGGUUUUAGAGCUAUGUUCAGCUC	13555	13556
Non-^{CNV}K tracrRNA IVT template	10214	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATGTTCAGCTCCCTATAGTGAGTCGTATTAGA	29197	29200
sgRNA^{CCR5} IVT template	10065	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATTTCTAGCTCTAAAACACAAAACCAAAGATGA ACACCCCTATAGTGAGTCGTATTAGA	37474	37475
sgRNA^{HBB} IVT template	10063	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTACTGCCCTGTGGGG CAAGCCCTATAGTGAGTCGTATTAGA	37518	37519
sgRNA^{IL2RG} IVT template	10066	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATTTCTAGCTCTAAAACTGTTGAAGCCATCATT ACCACCCTATAGTGAGTCGTATTAGA	37445	37446
sgRNA^{VEGFA} IVT template	10064	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATTTCTAGCTCTAAAACCACGCACACACTCACT CACCCCCTATAGTGAGTCGTATTAGA	37305	37306
Non-^{CNV}K tracrRNA	10602	RNA	IVT	GGGAGCUGAACAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGGCACCGAGUCGGUGCUUUU	24828	24824
sgRNA^{CCR5}	10598	RNA	IVT	GGGGUGUUCAUCUUUGGUUUUGUGUUUUAGAGCUAGAAAUAGCAA GUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA GUCGGUGCUUUU	33436	33439
sgRNA^{HBB}	10597	RNA	IVT	GGGCUUGCCCCACAGGGCAGUAAGUUUUAGAGCUAGAAAUAGCAA GUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA GUCGGUGCUUUU	33523	33526
sgRNA^{IL2RG}	10598	RNA	IVT	GGGUGGUAAUGAUGGCUUCAACAGUUUUAGAGCUAGAAAUAGCAA GUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA GUCGGUGCUUUU	33550	33548
sgRNA^{VEGFA}	10599	RNA	IVT	GGGGGUGAGUGAGUGUGUGCGUGGUUUUAGAGCUAGAAAUAGCAA GUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA GUCGGUGCUUUU	33694	33697

^{CNV}K-sgRNA^{CCR5}	10104/ 9884	RNA	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-tracrRNA (9884)	37530	37531
^{CNV}K-sgRNA^{HBB}	10105/ 9884	RNA	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-tracrRNA (9884)	37617	37619
^{CNV}K-sgRNA^{IL2RG}	10106/ 9884	RNA	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-tracrRNA (9884)	37644	37646
^{CNV}K-sgRNA^{VEGFA}	10107/ 9884	RNA	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-tracrRNA (9884)	37788	37790
^{CNV}K-MPS-sgRNA^{CCR5}	10104/ 10110	RNA/ MPS	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-MPS-tracrRNA (10110)	38373	38371
^{CNV}K-MPS-sgRNA^{HBB}	10105/ 10110	RNA/ MPS	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-MPS-tracrRNA (10110)	38460	38459
^{CNV}K-MPS-sgRNA^{IL2RG}	10106/ 10110	RNA/ MPS	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-MPS-tracrRNA (10110)	38487	38385
^{CNV}K-MPS-sgRNA^{VEGFA}	10107/ 10110	RNA/ MPS	^{CNV}K Crosslink	See crRNA^{EMX1} (9883) and ^{CNV}K-MPS-tracrRNA (10110)	38631	38630

Appendix 2

Oligonucleotides Introduced in Chapter 3

Appendix Table 6: Oligonucleotides Introduced in Chapter 3

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
Azide-crRNA^{GAPDH}	9236	RNA	AB 394	AUUAUCAGGUCCAGGCUACAGUUUUAGAGCUAUGAUCAGCUC-C7-Amide-C6-Azide	13742	13743
hp1	9505	DNA	AB 394	PO_4-GATGCAGGTCGAGATGTTTTTTTTTTTTTTTTTTTTCATCTCGACCTG	15096	15098
hp2	9506	DNA	AB 394	PO_4-GTCCCAGGTCGAGATGTTTTTTTTTT-Octadiynyl dU-TTTTTTTTTTCATCTCGACCTG	15123	15125
GFP11$_F$	9356	DNA	AB 394	Hexynyl-GAAGGTCTCCGGACC	4738	4739
GFP11$_R$	9357	DNA	AB 394	TCCCATCGAGATCCAGAC	5429	5429
crRNA^{GAPDH}-hp1	9236/ 9506	RNA/ DNA	Click Ligation	See crRNA^{GAPDH} (9236) and hp1 (9505)	28839	28864 (+Mg^{+2})
crRNA^{GAPDH}-GFP11$_F$	9236/ 9356	RNA/ DNA	Click Ligation	See crRNA^{GAPDH} (9236) and GFP11$_F$ (9356)	18477	18481
GAPDH Target	9352	DNA	AB 394	ACCTCGCTGATGAGCATAGATTATCAGGTCCAGGCTACAAGGCGTCGACGGAGCGC	17300	17299

GAPDH Non-target	9353	DNA	AB 394	FAM-GCGCTCCGTCGACGCCTTGTAGCCTGGACCTGATAATCTATGCTCATCAGCGAGGT	17720	17723
sgRNA^{GAPDH} IVT template	9370	DNA	AB 394	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTGTAGCCTGGACCTGATAATCCCTATAGTGAGTCGTATTAGA	37502	37502
sgRNA^{GAPDH}	9370	RNA	IVT	GGGAUUAUCAGGUCCAGGCUACAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU	33509	33506
sgRNA/PCR Splint	9384	DNA	AB 394	CGCCTAAAGCTGGAGCGCCGAAAAGCACCGACTCGGTGCC	12286	12286
GFP11vZ F Primer	9540	DNA	AB 394	CGGCGCTCCAGCTTTAGGCG-HEG-CCAGTACATGACCTTATGGG	12633	12633
GFP11vZ R Primer	9504	DNA	AB 394	TTGTAGCGTCTCTCATCTCAGAAGCCATAGAGCCC	10667	10668

Appendix Table 7: Sequence of GFP11 cargo v1

GAA*GGTCTC*$_1$C**GGAC**CCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGC
GTGTACGGTGGGAGGTCTATATAAGCAGAGCT**TACT**G*GAGACG*$_2$ATACCC$_2$*CGTCTC*G**TCTG**GCCACCATGGTTAGAGATCATA
TGGTGCTGCACGAGTACGTGAACGCCGCCGGCATCACCTAG**GCTT**AG*GTCTTC*$_3$TTCCGA$_3$*GAAGAC*TC**CGCT**CCATGGCCCAA
CTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACTGCATT
CTAGTTGTGGTTTGTCCAAACTCA TCAATGTATCTTATCATGTCTGGATCTC**GATG**G$_1$*GAGACC* GAC

Sequence	**Feature**
CCAAAAT...	CMVd1 minimal promoter
GCCACC...	Kozak consensus sequence
AGAGAT...	GFP11 coding sequence
CCATGG...	SV40 terminator and polyadenylation signal
GGTCTC$_1$	BsaI restriction site (5′ to 3′)
GAGACG$_2$	BsmBI restriction site (5′ to 3′)
GTCTTC$_3$	BbsI restriction site (5′ to 3′)
NNNN	4 nt overhangs produced by neighboring IIS restriction digest

Appendix 3

Oligonucleotides Introduced in Chapter 4

Appendix Table 8: Oligonucleotides Introduced in Section 4.3

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
Blocker 1.1	10601	DNA	AB 394	GTTCACGGA-^{CNV}K-CCCTC	4972	4973
Blocker 1.2	10600	2′OMe	AB 394	**GUUCACGGA**-^{CNV}K-**CCCUC**	5130	5135
Blocker 1.1F	9381	DNA	AB 394	FAM-GTTCACGGA-^{CNV}K-CCCTC	5508	5510
Blocker 1.2F	9380	2′OMe	AB 394	FAM-**GUUCACGGA**-^{CNV}K-**CCCUC**	5666	5667
FAM-ODN$_{60\text{-}99}$	10300	DNA	AB 394	FAM-GGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAG	13030	13025
B1.3	10291	2′OMe	AB 394	**GGGGA**-^{CNV}K-**GUCGGCGGG**	5226	5230
B1.4	10292	2′OMe	AB 394	**CUGCUUGA**-^{CNV}K-**CUCGCCCUUC**	6228	6230
B1.5F	9592	2′OMe	AB 394	FAM-**UGUGGCCA**-^{CNV}K-**AUUAUC**	5534	5539

Appendix Table 9: Oligonucleotides Introduced in Section 4.4

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
comp-ODN^{FAM}	9886	DNA	AB 394	FAM-GAGGAGCTGTTCACCGG	5788	5789

Appendix Table 10: Oligonucleotides Introduced in Section 4.5

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
B2.1	10111	2′OMe	AB 394	Amine-C12-**GGAGTCCTGGGTCAC**-C12-**CCA**-^{CNV}K-**GC**	7603	7603
B2.2	10112	2′OMe	AB 394	Amine-C12-**CCTGCAGGGAGGAGT**-C12-**GTCA**-^{CNV}K-**GGT**	8386	8386
B2.3	10113	2′OMe	AB 394	Amine-C12-**GTCCTGCAGGGAGGA**-C12-**GGGTCA**-^{CNV}K-**GGTC**	9424	9422
B3.1	10169	2′OMe	AB 394	**UCGAACUCGU**-PC-**CGUUCA**-^{CNV}K-**GGA**-PC-**CUCCAUGUGC**	10686	10687
B3.2	10170	2′OMe	AB 394	**UGGCCGU**-PC-**CGGA**-^{CNV}K-**CCC**-PC-**AUGUGCAC**	8399	8401
B3.3	10171	2′OMe	AB 394	**CCGUUCACGGAG**-PC-**UCCA**-^{CNV}K-**GUGC**	7377	7377

Appendix Table 11: Oligonucleotides Introduced in Section 4.6

Name	Lab ID	Type	Source	Sequence (5′→ 3′)	Calcd (Da)	Found (Da)
^{PC}K-ODN	10352	2′OMe	AB 394	CCGGTGA-^{PC}K-CAGCTCCTC	5230	5230
^{PC}K-B1.2	10557	2′OMe	AB 394	**GUUCACGGA**-^{PC}K-**CCCUC**	4990	4992

Appendix 4

PCR Primer Pairs

Appendix Table 12: PCR Primers Introduced in Chapter 2

Product	Length	Primer	Sequence (5′→ 3′)
EMX1	622 bp	9119	ACCACCCTTCTCTCTGGC
		9120	CAGGCAGGCTCTCCG
EMX1 TIDE	687 bp	9954	TGCCCGTGTCATTAAGAG
		9955	CCTATGTAGCCTCAGTC
CCR5 TIDE	731 bp	10081	CAGTTTGCATTCATGGAGGG
		10082	CCCGAGTAGCAGATGACC
HBB TIDE	753 bp	10083	CAAGGACAGGTACGGCT
		10084	TGCAATCATTCGTCTGTTTCC
IL2RG TIDE	743 bp	10085	GTTTACCACCTTACAGCAGC
		10086	GAGTCAGTGGGCATAGTGG
VEGFA TIDE	747 bp	10087	AAACTCTGTCCAGAGACACG
		10088	AGACTCCACAGTGCATACG

Appendix Table 13: PCR Primers Introduced in Chapter 3

Product	Length	Primer	Sequence (5′→ 3′)
GFP11 v1	388 bp	9356	Hexynyl-GAAGGTCTCCGGACC
		9357	TCCCATCGAGATCCAGAC
GFP11 vZ	1005 bp	9503	ACACAACGTCTCAGGACCCAGTACATGACCTTATGGG
		9504	TTGTAGCGTCTCTCATCTCAGAAGCCATAGAGCCC
GFP11 vZ w/ extension	998 bp + 20 nt	9540	CGGCGCTCCAGCTTTAGGCG-HEG-CCAGTACATGACCTTATGGG
		9504	TTGTAGCGTCTCTCATCTCAGAAGCCATAGAGCCC

Appendix Table 14: PCR Primers Introduced in Chapter 4

Product	Length	Primer	Sequence (5′→ 3′)
pRSET Insert	variable	9395	TTCTTTCCTGCGTTATCCCC
		9396	GCCTCTTCGCTATTACGCC
circRNA Insert	variable	10423	ATCATTTTGGCAAAGCGG
		10424	CAAAAAACCCCTCAAGACC

Appendix 5

Plasmids

Appendix Table 15: Acquired Plasmids

Plasmid	Addgene #	Source
pRSET-mCherry	Gift from Kalina Hristova (108857)	[300]
pCas9-GFP	Gift from Kiran Musunuru (44719)	[301]
pCAG-EOMES-IRES-EGFP	Gift from the Robertson Lab (N/A)	N/A
pCMV-PE2	Gift from David Liu (132775)	[93]
pCXLE-EGFP	Gift from Shinya Yamanaka (27082)	[302]
ZipGFP_TEV	Gift from Xiaokun Shu (81243)	[303]
circRNA-synIRES-R25-mNeonGreen	Gift from Howard Chang (188115)	[159]

Appendix Table 16: Plasmids Cloned in Chapter 2

pRSET-Cas9-GFP (Backbone = pRSET-mCherry, BamHI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
Cas9-T2A-GFP	pCas9-GFP	9725	9726	63	5036

Appendix Tables 17: Plasmids Cloned in Chapter 3

pCXLE-mCherry-GFP1-10 (Backbone = pCXLE-EGFP, EcoRI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
mCherry	pRSET-mCherry	9124	9125	62	761
GFP1-10	IDT G Block	9122	9123	62	715

pCMV-GFP11-v1 (Backbone = pCMV-PE2, SnaBI+EcoRI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
GFP11-v1	IDT G Block	9388	0434	68	457

pCMV-GFP11-v2 (Backbone = pCMV-GFP11-v1, SnaBI+EcoRI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
GFP11-v1	IDT G Block	9388	0434	68	457

Appendix Tables 18: Plasmids Cloned in Chapter 4

pRSET-IRES-mCherry (Backbone = pRSET-mCherry, BamHI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
IRES	pCAG-EOMES-IRES-EGFP	9350	9351	64	646

pRSET-IRES-EGFP (Backbone = pRSET-mCherry, BamHI+EcoRI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
IRES-EGFP	pRSET-IRES-EGFP-mCherry	9350	10301	67	1383

pRSET-IRES-mCherry-EGFP (Backbone = pRSET-IRES-mCherry, BsrGI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
EGFP	pRSET-mCherry-IRES-EGFP	10302	10301	69	764

pRSET-IRES-EGFP-mCherry (Backbone = pRSET-mCherry, BamHI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
IRES	pCAG-EOMES-IRES-EGFP	9350	9594	67	646
EGFP	pCAG-EOMES-IRES-EGFP	9595	9596	67	752

pRSET-mCherry-IRES-EGFP (Backbone = pRSET-mCherry, EcoRI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
IRES-EGFP	pRSET-IRES-EGFP	9653	9654	69	1386

pRSET-EGFP-IRES-mCherry (Backbone = pRSET-IRES-mCherry, NotI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
EGFP	pRSET-mCherry-IRES-EGFP	10304	10305	67	740

pRSET-IRES-mCherry-T2A-EGFP (Backbone = pRSET-IRES-mCherry, BsrGI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
T2A-EGFP	pCas9-GFP	10306	10307	67	805

circRNA-mCherry (Backbone = circRNA-synIRES-R25-mNeonGreen, BamHI+MfeI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
mCherry	pRSET-mCherry	10377	10378	64	761

circRNA-EGFP (Backbone = circRNA-synIRES-R25-mNeonGreen, BamHI+MfeI digest)

Insert	Insert Source	F Primer	R Primer	T_M (° C)	Length (bp)
EGFP	pRSET-IRES-EGFP	10377	10378	64	770

Appendix Table 18: Primers used in Plasmid Cloning Schemes

Lab ID	Source	Sequence (5 3)
9122	AB 394	GGCGGAGGCGGCAGCGGCGGTGGCGGCAGCATGTCCAAAGGAGAAGAACTG
9123	AB 394	TAGTCCCCGAAGCTTGAATTCACGACTAGAAGCTTGAATTCC
9124	AB 394	CATCATTTTGGCAAAGAATTCACCATGGTGAGCAAGGGC
9125	AB 394	GCTGCCGCCACCGCCGCTGCCGCCTCCGCCCTTGTACAGCTCGTCCATGC
9350	AB 394	TGTACGACGATGACGATAAGCTGGAGGCCAAGGTGTGAT
9351	AB 394	CTCACCATGGTGGCGGGATCGGTTGTGGCCATATTATCATCG
9388	AB 394	CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGAAGGTCTCCGGACC
9434	AB 394	TTCCTCTTCTTCTTGGGCTCGAATTTCCCATCGAGATCCAGAC
9495	AB 394	CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACGTGATGCGGTTTTGGCAG
9496	AB 394	ACGGAATTCCGTCTCCAGTAAGCTCTGCTTATATAGACCTCCC
9594	AB 394	ACCATGGTGGCGAATTCTTTGCCAAGGTTGTGGCCATATTATCATCGTG
9595	AB 394	ACACGATGATAATATGGCCACAACCTTGGCAAAGAATTCGCCACC
9596	AB 394	CCTTGCTCACCATGGTGGCGGGATCGCTGCCGCCTCCGCCCTTGTACAGCTCGTCCATGCC
9653	AB 394	GACGAGCTGTACAAGTAAGAATTCGCTGGAGGCCAAGGTGTGATG
9654	AB 394	TTGTTAGCAGCCGGATCAAGCTTCGTTACTTGTACAGCTCGTCCATGCC
9725	AB 394	GGTTGTGGCCATATTATCATCGTGTGAATTCGCGGCCGCCAC
9726	AB 394	TAGCAGCCGGATCAAGCTTCGAATTGTGATGCTATTGCTTTATTTGTAACC
10301	IDT	CCGGATCAAGCTTCGAATTTTACTTGTACAGCTCGTCCATGC
10302	IDT	ACCGGCGGCATGGACGAGCTGTACAAGGGCGGAGGCGGCAGCATGGTGAGCAAGGGCGA
10303	IDT	AAGCTTCGAATTCTTACTTGTACCTCAGCTTCCTTTCGGGCT
10304	IDT	TGGAGGCCAAGGTGTGATGCTTGGCAAAGAATTCGCCACC
10305	IDT	TCAAGCTTATCGAGCGGCCGTTACTTGTACAGCTCGTCCATGC
10377	AB 394	TATATAACATATACTGTGATCATGGTGAGCAAGGGCGAGGAG
10378	AB 394	CCACCGAGGCTCCAGCCAATTGTTACTATTTACTTGTACAGCTCGTCCATG

References

1. Beaucage, S.L. and M.H. Caruthers, *Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis.* Tetrahedron Lett, 1981. **22**: p. 1859–1862.
2. Matteucci, M.D. and M.H. Caruthers, *Synthesis of deoxyoligonucleotides on a polymer support.* J Am Chem Soc, 1981. **103, 3185–3191**(11): p. 3185-3191.
3. Gait, M.J. and R.C. Sheppard, *Rapid synthesis of oligodeoxyribonucleotides: a new solid-phase method.* Nucleic Acids Res, 1977. **4**(4): p. 1135-1158.
4. Caruthers, M.H., *A brief review of DNA and RNA chemical synthesis.* Biochem Soc Trans, 2011. **39**(2): p. 575-580.
5. Saiki, R.K., et al., *Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia.* Science, 1985. **230**(4732): p. 1350-1354.
6. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* J Mol Biol, 1975. **94**(3): p. 441-448.
7. Hughes, R.A. and A.D. Ellington, *Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology.* Cold Spring Harb Perspect Biol, 2017. **9**(1).
8. Seeman, N.C. and H.F. Sleiman, *DNA nanotechnology.* Nat Rev Mater, 2017. **3**(1): p. 17068.
9. Kulkarni, J.A., et al., *The current landscape of nucleic acid therapeutics.* Nat Nanotechnol, 2021. **16**(6): p. 630-643.
10. Epple, S., A.H. El-Sagheer, and T. Brown, *Artificial nucleic acid backbones and their applications in therapeutics, synthetic biology and biotechnology.* Emerg Top Life Sci, 2021. **5**(5): p. 691-697.
11. Elskens, J. and A. Madder, *Crosslinker-modified nucleic acid probes for improved target identification and biomarker detection.* RSC Chem Biol, 2021. **2**(2): p. 410-422.
12. Peckler, S., et al., *Structure of a Psoralen-Thymine Monoadduct Formed in Photoreaction with DNA.* J Mol Biol, 1982. **162**: p. 157-172.
13. Hearst, J.E., *Photochemistry of the Psoralens.* Chem Res Toxicol, 1989. **2**(2): p. 69-75.
14. Song, P.S. and K.J. Tapley, *Photochemistry and Photobiology of Psoralens.* Photochem Photobiol, 1979. **29**(6): p. 1177-1197.
15. Tavakoli, A. and J.H. Min, *Photochemical modifications for DNA/RNA oligonucleotides.* RSC Adv, 2022. **12**(11): p. 6484-6507.
16. Rastogi, R.P., et al., *Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair.* J Nucleic Acids, 2010. **2010**: p. 592980.
17. Yoshimura, Y. and K. Fujimoto, *Ultrafast reversible photo-cross-linking reaction: toward in situ DNA manipulation.* Org Lett, 2008. **10**(15): p. 3227-3230.

18. Yoshimura, Y., et al., *A new approach for reversible RNA photocrosslinking reaction: application to sequence-specific RNA selection.* Chembiochem, 2009. **10**(9): p. 1473-6.
19. Sakamoto, T., Y. Tanaka, and K. Fujimoto, *DNA photo-cross-linking using 3-cyanovinylcarbazole modified oligonucleotide with threoninol linker.* Org Lett, 2015. **17**(4): p. 936-9.
20. Sakamoto, T., M. Ooe, and K. Fujimoto, *Critical Effect of Base Pairing of Target Pyrimidine on the Interstrand Photo-Cross-Linking of DNA via 3-Cyanovinylcarbazole Nucleoside.* Bioconjug Chem, 2015. **26**(8): p. 1475-8.
21. Fujimoto, K., H. Yang-Chun, and S. Nakamura, *Strong Inhibitory Effects of Antisense Probes on Gene Expression through Ultrafast RNA Photocrosslinking.* Chem Asian J, 2019. **14**(11): p. 1912-1916.
22. Fujimo, K., K. Konishi-Hiratsuka, and T. Sakamoto, *Quick, Selective and Reversible Photocrosslinking Reaction between 5-Methylcytosine and 3-Cyanovinylcarbazole in DNA Double Strand.* Int J Mol Sci, 2013. **14**(3): p. 5765-74.
23. Fujimoto, K., et al., *Details of the ultrafast DNA photo-cross-linking reaction of 3-cyanovinylcarbazole nucleoside: cis-trans isomeric effect and the application for SNP-based genotyping.* J Am Chem Soc, 2013. **135**(43): p. 16161-7.
24. Fujimoto, K., et al., *Quick and reversible photocrosslinking reaction of 3-cyanovinylcarbazole nucleoside in a DNA triplex.* Org Biomol Chem, 2013. **11**(31): p. 5065-8.
25. Zhang, D., et al., *Crystal structure of a DNA/Ba2+ G-quadruplex containing a water-mediated C-tetrad.* Nucleic Acids Res, 2014. **42**(21): p. 13422-9.
26. Fujimoto, K., et al., *Effect of linker length on photo-cross-linking position mediated by click chemistry via [2 + 2] photocycloaddition.* Photochem Photobiol Sci, 2020. **19**(6): p. 776-782.
27. Yoshimura, Y., H. Okada, and K. Fujimoto, *Photoreversible DNA end capping for the formation of hairpin structures.* Org Biomol Chem, 2010. **8**(7): p. 1523-6.
28. Tagawa, M., et al., *Stabilization of DNA nanostructures by photo-cross-linking.* Soft Matter, 2011. **7**(22).
29. Nakamura, S. and K. Fujimoto, *Creation of DNA array structure equipped with heat resistance by ultrafast photocrosslinking.* J Chem Technol Biotechnol, 2014. **89**(7): p. 1086-1090.
30. Zhou, F., et al., *Mutations in artificial self-replicating tiles: A step toward Darwinian evolution.* Proc Natl Acad Sci U S A, 2021. **118**(50).
31. Chao, D., et al., *Covalent stabilization of DNA nanostructures on cell membranes for efficient surface receptor-mediated labeling and function regulations.* Sci China Chem, 2022. **65**(11): p. 2327-2334.
32. Harimech, P.K., et al., *Reversible Ligation of Programmed DNA-Gold Nanoparticle Assemblies.* J Am Chem Soc, 2015. **137**(29): p. 9242-5.
33. De Fazio, A.F., et al., *Light-Induced Reversible DNA Ligation of Gold Nanoparticle Superlattices.* ACS Nano, 2019. **13**(5): p. 5771-5777.

34. Fujimoto, K., M. Ichikawa, and S. Nakamura, *Photoinduced aggregation of liposome modified with DNA containing ultrafast DNA photo-cross-linker.* J Chem Technol Biotechnol, 2021. **97**(1): p. 295-298.
35. Nakamura, S., et al., *Phototriggered Sequence-specific DNA Transportation into Liposomes Using Ultrafast DNA Photocrosslinking.* Chem Lett, 2017. **46**(12): p. 1839-1841.
36. Vieregg, J.R., et al., *Selective nucleic acid capture with shielded covalent probes.* J Am Chem Soc, 2013. **135**(26): p. 9691-9.
37. Nakamura, S., et al., *Photochemical Acceleration of DNA Strand Displacement by Using Ultrafast DNA Photo-crosslinking.* Chembiochem, 2017. **18**(20): p. 1984-1989.
38. Nakamura, S., H. Kawabata, and K. Fujimoto, *Double duplex invasion of DNA induced by ultrafast photo-cross-linking using 3-cyanovinylcarbazole for antigene methods.* Chem Commun (Camb), 2017. **53**(54): p. 7616-7619.
39. Fujimoto, K., et al., *The Inhibition Effect of Photo-Cross-Linking between Probes in Photo-Induced Double Duplex Invasion DNA.* Chembiochem, 2021. **22**(24): p. 3402-3405.
40. Fujimoto, K., et al., *RNA fluorescence in situ hybridization using 3-cyanovinylcarbazole modified oligodeoxyribonucleotides as photo-cross-linkable probes.* Bioorg Med Chem Lett, 2016. **26**(21): p. 5312-5314.
41. Nakamura, S., C. Kano, and K. Fujimoto, *Wash-free RNA FISH Using a Photoresponsive Beacon Probe via Photochemical Crosslinking.* Chem Lett, 2017. **46**(12): p. 1711-1713.
42. Liu, N., et al., *Super-resolution labelling with Action-PAINT.* Nat Chem, 2019. **11**(11): p. 1001-1008.
43. Kishi, J.Y., et al., *Light-Seq: light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing.* Nat Methods, 2022. **19**(11): p. 1393-1402.
44. Shigeno, A., et al., *Quick regulation of mRNA functions by a few seconds of photoirradiation.* Org Biomol Chem, 2012. **10**(38): p. 7820-5.
45. Sakamoto, T., et al., *Photo-regulation of constitutive gene expression in living cells by using ultrafast photo-cross-linking oligonucleotides.* Biomater Sci, 2014. **2**(9): p. 1154-1157.
46. Watanabe, Y. and K. Fujimoto, *Complete Photochemical Regulation of 8-17 DNAzyme Activity by Using Reversible DNA Photo-crosslinking.* Chembiochem, 2020. **21**(22): p. 3244-3248.
47. Masubuchi, T., et al., *Construction of integrated gene logic-chip.* Nat Nanotechnol, 2018. **13**(10): p. 933-940.
48. Fujimoto, K., et al., *DNA Photo-cross-linking Using Pyranocarbazole and Visible Light.* Org Lett, 2018. **20**(10): p. 2802-2805.
49. Mihara, J.I. and K. Fujimoto, *Photocrosslinking of DNA using 4-methylpyranocarbazole nucleoside with thymine base selectivity.* Org Biomol Chem, 2021. **19**(45): p. 9860-9866.

50. Fujimoto, K., et al., *DNA photo-cross-linking using a pyranocarbazole-modified oligodeoxynucleotide with a d-threoninol linker.* RSC Adv, 2019. **9**(53): p. 30693-30697.
51. Zhao, G., et al., *Switchable DNA-Encoded Chemical Library: Interconversion between Double- and Single-Stranded DNA Formats.* Chembiochem, 2022. **23**(14): p. e202200025.
52. Mochizuki, Y., et al., *A versatile puromycin-linker using cnvK for high-throughput in vitro selection by cDNA display.* J Biotechnol, 2015. **212**: p. 174-80.
53. Adli, M., *The CRISPR tool kit for genome editing and beyond.* Nat Commun, 2018. **9**(1): p. 1911.
54. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-821.
55. Gasiunas, G., et al., *Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* Proc Natl Acad Sci USA, 2012. **109**(39): p. E2579-E2586.
56. Cong, L., et al., *Multiplex Genome Engineering Using CRISPR/Cas Systems.* Science, 2013. **339**(6121): p. 819-823.
57. Mali, P., et al., *RNA-Guided Human Genome Engineering via Cas9.* Science, 2013. **339**(6121): p. 823-826.
58. Rouet, P., F. Smih, and M. Jasin, *Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.* Mol Cell Biol, 1994. **14**(12): p. 8096-8106.
59. Miller, J.C., et al., *An improved zinc-finger nuclease architecture for highly specific genome editing.* Nat Biotechnol, 2007. **25**(7): p. 778-785.
60. Christian, M., et al., *Targeting DNA double-strand breaks with TAL effector nucleases.* Genetics, 2010. **186**(2): p. 757-761.
61. Lander, Eric S., *The Heroes of CRISPR.* Cell, 2016. **164**(1): p. 18-28.
62. Hille, F. and E. Charpentier, *CRISPR-Cas: biology, mechanisms and relevance.* Philos Trans R Soc Lond B Biol Sci, 2016. **371**(1707).
63. Hillary, V.E. and S.A. Ceasar, *A Review on the Mechanism and Applications of CRISPR/Cas9/Cas12/Cas13/Cas14 Proteins Utilized for Genome Engineering.* Mol Biotechnol, 2023. **65**(3): p. 311-325.
64. Jinek, M., et al., *Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation.* Science, 2014. **343**(6176): p. 1247997.
65. Jiang, F., et al., *A Cas9 guide RNA complex preorganized for target DNA recognition.* Science, 2015. **348**(6242): p. 1477-1481.
66. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-67.
67. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity.* Nature, 2010. **463**(7280): p. 568-571.
68. Bikard, D., et al., *CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during In Vivo Bacterial Infection.* Cell Host Microbe, 2012. **12**(2): p. 177-186.

69. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-573.
70. Jiang, F., et al., *Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage.* Science, 2016. **351**(6275): p. 867-71.
71. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes.* Proc Natl Acad Sci USA, 2014. **111**(27): p. 9798-9803.
72. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-49.
73. O'Geen, H., et al., *A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture.* Nucleic Acids Res, 2015. **43**(6): p. 3389-404.
74. Kuscu, C., et al., *Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease.* Nat Biotechnol, 2014. **32**(7): p. 677-683.
75. Wu, X., et al., *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells.* Nat Biotechnol, 2014. **32**(7): p. 670-676.
76. Sternberg, S.H., et al., *Conformational control of DNA target cleavage by CRISPR–Cas9.* Nature, 2015. **527**(7576): p. 110-113.
77. Josephs, E.A., et al., *Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage.* Nucleic Acids Res, 2015. **43**(18): p. 8924-41.
78. Mekler, V., et al., *Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3'-terminal segment of guide RNA.* Nucleic Acids Res, 2016. **44**(6): p. 2837-2845.
79. Wright, A.V., et al., *Rational design of a split-Cas9 enzyme complex.* Proc Natl Acad Sci USA, 2015. **112**(10): p. 2984-2989.
80. Jinek, M., et al., *RNA-programmed genome editing in human cells.* eLife, 2013. **2**: p. e00471.
81. Fu, Y., et al., *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nat Biotechnol, 2013. **31**(9): p. 822-826.
82. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nat Biotechnol, 2013. **31**(9): p. 827-832.
83. Fu, Y., et al., *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nat Biotechnol, 2014. **32**(3): p. 279-284.
84. Graf, R., et al., *sgRNA Sequence Motifs Blocking Efficient CRISPR/Cas9-Mediated Gene Editing.* Cell Rep, 2019. **26**(5): p. 1098-1103.e3.
85. Thyme, S.B., et al., *Internal guide RNA interactions interfere with Cas9-mediated cleavage.* Nat Commun, 2016. **7**(1): p. 11750.
86. Riesenberg, S., et al., *Improved gRNA secondary structures allow editing of target sites resistant to CRISPR-Cas9 cleavage.* Nat Commun, 2022. **13**(1): p. 489.
87. Jasin, M. and R. Rothstein, *Repair of strand breaks by homologous recombination.* Cold Spring Harb Perspect Biol, 2013. **5**(11): p. a012740.
88. Komor, A.C., et al., *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage.* Nature, 2016. **533**(7603): p. 420-424.

89. Gaudelli, N.M., et al., *Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage.* Nature, 2017. **551**(7681): p. 464-471.
90. Chen, S.P. and H.H. Wang, *An engineered Cas-transposon system for programmable and site-directed DNA transpositions.* CRISPR J, 2019. **2**(6): p. 376-394.
91. Chaikind, B., et al., *A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells.* Nucleic Acids Res, 2016. **44**(20): p. 9758-9770.
92. Wang, C., et al., *dCas9-based gene editing for cleavage-free genomic knock-in of long sequences.* Nat Cell Biol, 2022. **24**(2): p. 268-278.
93. Anzalone, A.V., et al., *Search-and-replace genome editing without double-strand breaks or donor DNA.* Nature, 2019. **576**(7785): p. 149-157.
94. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Mol Cell, 2010. **40**(2): p. 179-204.
95. Chapman, J.R., M.R. Taylor, and S.J. Boulton, *Playing the end game: DNA double-strand break repair pathway choice.* Mol Cell, 2012. **47**(4): p. 497-510.
96. Chiruvella, K.K., Z. Liang, and T.E. Wilson, *Repair of double-strand breaks by end joining.* Cold Spring Harb Perspect Biol, 2013. **5**(5): p. a012757.
97. Brinkman, E.K., et al., *Kinetics and fidelity of the repair of Cas9-induced double-strand DNA breaks.* Mol Cell, 2018. **70**(5): p. 801-813. e6.
98. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system.* Science, 2014. **343**(6166): p. 80-84.
99. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells.* Science, 2014. **343**(6166): p. 84-87.
100. Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation.* Nat Biotechnol, 2014. **32**(12): p. 1262-1267.
101. Wang, W., et al., *Gene editing of the wheat homologs of TONNEAU 1-recruiting motif encoding gene affects grain shape and weight in wheat.* Plant J, 2019. **100**(2): p. 251-264.
102. Xu, Z.-S., et al., *Changing carrot color: insertions in DcMYB7 alter the regulation of anthocyanin biosynthesis and modification.* Plant Physiol, 2019. **181**(1): p. 195-207.
103. Uluisik, S., et al., *Genetic improvement of tomato by targeted control of fruit softening.* Nat Biotechnol, 2016. **34**(9): p. 950-952.
104. Wang, H., et al., *One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.* Cell, 2013. **153**(4): p. 910-918.
105. Shams, F., et al., *Advance trends in targeting homology-directed repair for accurate gene editing: An inclusive review of small molecules and modified CRISPR-Cas9 systems.* Bioimpacts, 2022. **12**(4): p. 371-391.
106. Rouet, P., F. Smih, and M. Jasin, *Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.* Proc Natl Acad Sci U S A, 1994. **91**(13): p. 6064-8.
107. Song, F. and K. Stieger, *Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks.* Mol Ther Nucleic Acids, 2017. **7**: p. 53-60.

108. Anzalone, A.V., L.W. Koblan, and D.R. Liu, *Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors.* Nat Biotechnol, 2020. **38**(7): p. 824-844.
109. Li, T., et al., *CRISPR/Cas9 therapeutics: progress and prospects.* Signal Transduct Target Ther, 2023. **8**(1): p. 36.
110. Larson, R.C., et al., *CAR T cell killing requires the IFNγR pathway in solid but not liquid tumours.* Nature, 2022. **604**(7906): p. 563-570.
111. Heyer, W.-D., K.T. Ehmsen, and J. Liu, *Regulation of homologous recombination in eukaryotes.* Annu Rev Genet, 2010. **44**: p. 113-139.
112. Moynahan, M.E. and M. Jasin, *Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis.* Nat Rev Mol Cell Biol, 2010. **11**(3): p. 196-207.
113. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.* elife, 2014. **3**: p. e04766.
114. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annu Rev Biochem, 2010. **79**: p. 181-211.
115. Jayavaradhan, R., et al., *CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites.* Nat Commun, 2019. **10**(1): p. 2866.
116. Charpentier, M., et al., *CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair.* Nat Commun, 2018. **9**(1): p. 1133.
117. Rees, H.A., W.-H. Yeh, and D.R. Liu, *Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks.* Nat Commun, 2019. **10**(1): p. 2212.
118. Richardson, C.D., et al., *Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA.* Nat Biotechnol, 2016. **34**(3): p. 339-344.
119. Chen, P.J. and D.R. Liu, *Prime editing for precise and highly versatile genome manipulation.* Nat Rev Genet, 2023. **24**(3): p. 161-177.
120. Gao, R., et al., *Genomic and transcriptomic analyses of prime editing guide RNA–independent off-target effects by prime editors.* CRISPR J, 2022. **5**(2): p. 276-293.
121. Schene, I.F., et al., *Prime editing for functional repair in patient-derived disease models.* Nat Commun, 2020. **11**(1): p. 5352.
122. Geurts, M.H., et al., *Evaluating CRISPR-based prime editing for cancer modeling and CFTR repair in organoids.* Life Sci Alliance, 2021. **4**(10).
123. Park, S.-J., et al., *Targeted mutagenesis in mouse cells and embryos using an enhanced prime editor.* Genome Biol, 2021. **22**(1): p. 1-11.
124. Liu, Y., et al., *Efficient generation of mouse models with the prime editing system.* Cell Discov, 2020. **6**(1): p. 27.
125. Jin, S., et al., *Genome-wide specificity of prime editors in plants.* Nat Biotechnol, 2021. **39**(10): p. 1292-1299.
126. Landrum, M.J., et al., *ClinVar: public archive of interpretations of clinically relevant variants.* Nucleic Acids Res, 2016. **44**(D1): p. D862-D868.

127. Chen, P.J., et al., *Enhanced prime editing systems by manipulating cellular determinants of editing outcomes.* Cell, 2021. **184**(22): p. 5635-5652. e29.
128. Zong, Y., et al., *An engineered prime editor with enhanced editing efficiency in plants.* Nat Biotechnol, 2022. **40**(9): p. 1394-1402.
129. Xu, W., et al., *A design optimized prime editor with expanded scope and capability in plants.* Nat Plants, 2022. **8**(1): p. 45-52.
130. Nelson, J.W., et al., *Engineered pegRNAs improve prime editing efficiency.* Nat Biotechnol, 2022. **40**(3): p. 402-410.
131. Doman, J.L., et al., *Designing and executing prime editing experiments in mammalian cells.* Nat Protoc, 2022. **17**(11): p. 2431-2468.
132. Boczkowski, D., et al., *Dendritic cells pulsed with RNA are potent antigen-presenting cells in vitro and in vivo.* J Exp Med, 1996. **184**(2): p. 465-472.
133. Polack, F.P., et al., *Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine.* N Engl J Med, 2020. **383**(27): p. 2603-2615.
134. Baden, L.R., et al., *Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine.* N Engl J Med, 2021. **384**(5): p. 403-416.
135. Qin, S., et al., *mRNA-based therapeutics: powerful and versatile tools to combat diseases.* Signal Transduct Target Ther, 2022. **7**(1): p. 166.
136. Kis, Z., et al., *Rapid development and deployment of high-volume vaccines for pandemic response.* J Adv Manuf Process, 2020. **2**(3): p. e10060.
137. Chaudhary, N., D. Weissman, and K.A. Whitehead, *mRNA vaccines for infectious diseases: principles, delivery and clinical translation.* Nat Rev Drug Discov, 2021. **20**(11): p. 817-838.
138. Vishweshwaraiah, Y.L. and N.V. Dokholyan, *mRNA vaccines for cancer immunotherapy.* Front Immunol, 2022. **13**: p. 1029069.
139. Conry, R.M., et al., *Characterization of a messenger RNA polynucleotide vaccine vector.* Cancer Res, 1995. **55**(7): p. 1397-1400.
140. Pardi, N., et al., *mRNA vaccines - a new era in vaccinology.* Nat Rev Drug Discov, 2018. **17**(4): p. 261-279.
141. Rohner, E., et al., *Unlocking the promise of mRNA therapeutics.* Nat Biotechnol, 2022. **40**(11): p. 1586-1600.
142. Gillmore, J.D., et al., *CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis.* N Engl J Med, 2021. **385**(6): p. 493-502.
143. Collén, A., et al., *VEGFA mRNA for regenerative treatment of heart failure.* Nat Rev Drug Discov, 2022. **21**(1): p. 79-80.
144. Finn, J.D., et al., *A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing.* Cell Rep, 2018. **22**(9): p. 2227-2235.
145. Sürün, D., et al., *Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors.* Genes, 2020. **11**(5): p. 511.
146. Sultana, N., et al., *Optimization of 5' untranslated region of modified mRNA for use in cardiac or hepatic ischemic injury.* Mol Ther Methods Clin Dev, 2020. **17**: p. 622-633.

147. Gebre, M.S., et al., *Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine.* Nature, 2022. **601**(7893): p. 410-414.
148. Sample, P.J., et al., *Human 5′ UTR design and variant effect prediction from a massively parallel translation assay.* Nat Biotechnol, 2019. **37**(7): p. 803-809.
149. Grier, A.E., et al., *pEVL: a linear plasmid for generating mRNA IVT templates with extended encoded poly (A) sequences.* Mol Ther Nucleic, 2016. **5**: p. e306.
150. Henderson, J.M., et al., *Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription.* Curr Protoc, 2021. **1**(2): p. e39.
151. Andries, O., et al., *N(1)-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice.* J Control Release, 2015. **217**: p. 337-44.
152. Anderson, B.R., et al., *Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation.* Nucleic Acids Res, 2010. **38**(17): p. 5884-5892.
153. Kim, D.Y., et al., *Enhancement of protein expression by alphavirus replicons by designing self-replicating subgenomic RNAs.* Proc Natl Acad Sci USA, 2014. **111**(29): p. 10708-10713.
154. McKay, P.F., et al., *Self-amplifying RNA SARS-CoV-2 lipid nanoparticle vaccine candidate induces high neutralizing antibody titers in mice.* Nat Commun, 2020. **11**(1): p. 3523.
155. Fuller, D.H. and P. Berglund, *Amplifying RNA vaccine development.* N Engl J Med, 2020. **382**(25): p. 2469-2471.
156. Chen, Y.G., et al., *Sensing Self and Foreign Circular RNAs by Intron Identity.* Mol Cell, 2017. **67**(2): p. 228-238.e5.
157. Santer, L., C. Bär, and T. Thum, *Circular RNAs: a novel class of functional RNA molecules with a therapeutic perspective.* Mol Ther, 2019. **27**(8): p. 1350-1363.
158. Wesselhoeft, R.A., et al., *RNA circularization diminishes immunogenicity and can extend translation duration in vivo.* Mol Cell, 2019. **74**(3): p. 508-520. e4.
159. Chen, R., et al., *Engineering circular RNA for enhanced protein production.* Nat Biotechnol, 2022.
160. Ochoa, S. and V.T. Milam, *Modified Nucleic Acids: Expanding the Capabilities of Functional Oligonucleotides.* Molecules, 2020. **25**(20).
161. Hendel, A., et al., *Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells.* Nat Biotechnol, 2015. **33**(9): p. 985-989.
162. Yin, H., et al., *Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing.* Nat Biotechnol, 2017. **35**(12): p. 1179-1187.
163. Mir, A., et al., *Heavily and fully modified RNAs guide efficient SpyCas9-mediated genome editing.* Nat Commun, 2018. **9**(1): p. 2641.
164. Basila, M., M.L. Kelley, and A.V.B. Smith, *Minimal 2′-O-methyl phosphorothioate linkage modification pattern of synthetic guide RNAs for increased stability and efficient CRISPR-Cas9 gene editing avoiding cellular toxicity.* PLoS One, 2017. **12**(11): p. e0188593.

165. Finn, J.D., et al., *A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing.* Cell Rep, 2018. **22**(9): p. 2227-2235.
166. Rahdar, M., et al., *Synthetic CRISPR RNA-Cas9-guided genome editing in human cells.* Proc Natl Acad Sci U S A, 2015. **112**(51): p. E7110-7.
167. Ryan, D.E., et al., *Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs.* Nucleic Acids Res, 2018. **46**(2): p. 792-803.
168. Cromwell, C.R., et al., *Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity.* Nat Commun, 2018. **9**(1): p. 1448.
169. Hao Yin, C.-Q.S., Sneha Suresh, Suet-Yan Kwan, Qiongqiong Wu, Stephen Walsh, Junmei Ding, Roman L Bogorad, Lihua Julie Zhu, Scot A Wolfe, Victor Koteliansky, Wen Xue, Robert Langer & Daniel G Anderson, *Partial DNA-guided Cas9 enables genome editing with reduced off-target activity.* Nat Chem Biol, 2018. **14**: p. 311-316.
170. Rueda, F.O., et al., *Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease.* Nat Commun, 2017. **8**(1): p. 1610.
171. Kartje, Z.J., et al., *Chimeric Guides Probe and Enhance Cas9 Biochemical Activity.* Biochemistry, 2018. **57**(21): p. 3027-3031.
172. O'Reilly, D., et al., *Extensive CRISPR RNA modification reveals chemical compatibility and structure-activity relationships for Cas9 biochemical activity.* Nucleic Acids Res, 2019. **47**(2): p. 546-558.
173. Behlke, M.A., *Chemical modification of siRNAs for in vivo use.* Oligonucleotides, 2008. **18**(4): p. 305-19.
174. Allen, D., M. Rosenberg, and A. Hendel, *Using Synthetically Engineered Guide RNAs to Enhance CRISPR Genome Editing Systems in Mammalian Cells.* Front Genome Ed, 2021. **2**.
175. Rozners, E., *Chemical Modifications of CRISPR RNAs to Improve Gene-Editing Activity and Specificity.* J Am Chem Soc, 2022. **144**(28): p. 12584-12594.
176. Schubert, M.S., et al., *Chemical Modification of CRISPR gRNAs Eliminate type I Interferon Responses in Human Peripheral Blood Mononuclear Cells.* J Cytokine Biol, 2018. **3**(1).
177. Li, B., et al., *Engineering CRISPR-Cpf1 crRNAs and mRNAs to maximize genome editing efficiency.* Nat Biomed Eng, 2017. **1**(5).
178. McMahon, M.A., et al., *Chemically Modified Cpf1-CRISPR RNAs Mediate Efficient Genome Editing in Mammalian Cells.* Mol Ther, 2018. **26**(5): p. 1228-1240.
179. Jiang, T., et al., *Chemical modifications of adenine base editor mRNA and guide RNA expand its application scope.* Nat Commun, 2020. **11**(1): p. 1979.
180. Li, H., et al., *Highly efficient generation of isogenic pluripotent stem cell models using prime editing.* eLife, 2022. **11**: p. e79208.
181. Zhuo, C., et al., *Spatiotemporal control of CRISPR/Cas9 gene editing.* Signal Transduct Target Ther, 2021. **6**(1): p. 238.
182. Moroz-Omori, E.V., et al., *Photoswitchable gRNAs for Spatiotemporally Controlled CRISPR-Cas-Based Genomic Regulation.* ACS Cent Sci, 2020. **6**(5): p. 695-703.

183. Zhou, W., et al., *Spatiotemporal Control of CRISPR/Cas9 Function in Cells and Zebrafish using Light-Activated Guide RNA.* Angew Chem Int Ed Engl, 2020. **59**(23): p. 8998-9003.
184. Carlson-Stevermer, J., et al., *CRISPRoff enables spatio-temporal control of CRISPR editing.* Nat Commun, 2020. **11**(1): p. 5041.
185. Wang, S.R., et al., *Conditional control of RNA-guided nucleic acid cleavage and gene editing.* Nat Commun, 2020. **11**(1): p. 91.
186. Krysler, A.R., et al., *Guide RNAs containing universal bases enable Cas9/Cas12a recognition of polymorphic sequences.* Nat Commun, 2022. **13**(1): p. 1617.
187. Lee, K., et al., *Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair.* Nat Biomed Eng, 2017. **1**: p. 889-901.
188. Lee, K., et al., *Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering.* eLife, 2017. **6**: p. e25312.
189. Qiao, H., et al., *The Advance of CRISPR-Cas9-Based and NIR/CRISPR-Cas9-Based Imaging System.* Front Chem, 2021. **9**: p. 786354.
190. Wu, X., et al., *Progress and Challenges for Live-cell Imaging of Genomic Loci Using CRISPR-based Platforms.* GPB, 2019. **17**(2): p. 119-128.
191. Deng, W., et al., *CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells.* Proc Natl Acad Sci U S A, 2015. **112**(38): p. 11870-5.
192. Haifeng Wang, M.N., Timothy R. Abbott, Dehua Zhao, Kaiwen Luo,Cordelia Yu, Cindy M. Nguyen, Albert Lo, Timothy P. Daley, Marie La Russa,Yanxia Liu, Lei S. Qi, *CRISPR-mediated live imaging ofgenome editing and transcription.* Science, 2019. **365**: p. 1301-1305.
193. Kim, S., et al., *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.* Genome Res, 2014. **24**(6): p. 1012-9.
194. Chen, Q., Y. Zhang, and H. Yin, *Recent advances in chemical modifications of guide RNA, mRNA and donor template for CRISPR-mediated genome editing.* Adv Drug Deliv Rev, 2021. **168**: p. 246-258.
195. Cho, S.W., et al., *Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.* Nat Biotechnol, 2013. **31**(3): p. 230-2.
196. Liang, X., et al., *Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection.* J Biotechnol, 2015. **208**: p. 44-53.
197. Gholamalipour, Y., A. Karunanayake Mudiyanselage, and C.T. Martin, *3′ end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character-RNA-Seq analyses.* Nucleic Acids Res, 2018. **46**(18): p. 9253-9263.
198. Pleiss, J.A., M.L. Derrick, and O.C. Uhlenbeck, *T7 RNA polymerase produces 5′ end heterogeneity during in vitro transcription from certain templates.* RNA, 1998. **4**(10): p. 1313–1317.
199. Kim, S., et al., *CRISPR RNAs trigger innate immune responses in human cells.* Genome Res, 2018. **28**(3): p. 367-73.
200. Wienert, B., et al., *In vitro-transcribed guide RNAs trigger an innate immune response via the RIG-I pathway.* PLoS Biol, 2018. **16**(7): p. e2005840.
201. Hertler, J., et al., *Synthesis of point-modified mRNA.* Nucleic Acids Res, 2022. **50**(20): p. e115.

202. Gamper, H., et al., *Synthesis of Long RNA with a Site-Specific Modification by Enzymatic Splint Ligation.* bioRxiv, 2022: p. 2022.09.17.508400.
203. He, K., et al., *Conjugation and Evaluation of Triazole-Linked Single Guide RNA for CRISPR-Cas9 Gene Editing.* Chembiochem, 2016. **17**(19): p. 1809-1812.
204. Brinkman, E.K., et al., *Easy quantitative assessment of genome editing by sequence trace decomposition.* Nucleic Acids Res, 2014. **42**(22): p. e168.
205. Chen, Z., et al., *Tetrazine-Ligated CRISPR sgRNAs for Efficient Genome Editing.* ACS Chem Biol, 2022. **17**(5): p. 1045-1050.
206. Taemaitree, L., et al., *An artificial triazole backbone linkage provides a split-and-click strategy to bioactive chemically modified CRISPR sgRNA.* Nat Commun, 2019. **10**(1): p. 1610.
207. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-588.
208. Shechner, D.M., et al., *Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display.* Nat Methods, 2015. **12**(7): p. 664-70.
209. Carlson-Stevermer, J., et al., *Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing.* Nat Commun, 2017. **8**(1): p. 1711.
210. Jiang, M., Y. Ye, and J. Li, *Core Hairpin Structure of SpCas9 sgRNA Functions in a Sequence- and Spatial Conformation-Dependent Manner.* SLAS Technol, 2021. **26**(1): p. 92-102.
211. Gruber, A.R., et al., *The Vienna RNA websuite.* Nucleic Acids Res, 2008. **36**: p. W70-4.
212. Carrington, B., K. Bishop, and R. Sood, *A Comprehensive Review of Indel Detection Methods for Identification of Zebrafish Knockout Mutants Generated by Genome-Editing Nucleases.* Genes, 2022. **13**(5).
213. Hanna, R.E. and J.G. Doench, *Design and analysis of CRISPR–Cas experiments.* Nat Biotechnol, 2020. **38**(7): p. 813-823.
214. Liu, Z., et al., *Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector.* Sci Rep, 2017. **7**(1): p. 2193.
215. Cradick, T.J., et al., *CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity.* Nucleic Acids Res, 2013. **41**(20): p. 9584-92.
216. Dever, D.P., et al., *CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells.* Nature, 2016. **539**(7629): p. 384-389.
217. Chen, P.J. and D.R. Liu, *Prime editing for precise and highly versatile genome manipulation.* Nat Rev Genet, 2022.
218. Xu, R., et al., *Identification of herbicide resistance OsACC1 mutations via in planta prime-editing-library screening in rice.* Nat Plants, 2021. **7**(7): p. 888-892.
219. Jiao, Y., et al., *Random-PE: an efficient integration of random sequences into mammalian genome by prime editing.* Mol Biomed, 2021. **2**(1): p. 36.
220. Thomas, M.C. and C.M. Chiang, *The general transcription machinery and general cofactors.* Crit Rev Biochem Mol Biol, 2006. **41**(3): p. 105-78.
221. Haberle, V. and A. Stark, *Eukaryotic core promoters and the functional basis of transcription initiation.* Nat Rev Mol Cell Biol, 2018. **19**(10): p. 621-637.

222. Andersson, R. and A. Sandelin, *Determinants of enhancer and promoter activities of regulatory elements.* Nat Rev Genet, 2020. **21**(2): p. 71-87.
223. Gasperini, M., J.M. Tome, and J. Shendure, *Towards a comprehensive catalogue of validated and target-linked human enhancers.* Nat Rev Genet, 2020. **21**(5): p. 292-310.
224. Kadauke, S. and G.A. Blobel, *Chromatin loops in gene regulation.* Biochim Biophys Acta, 2009. **1789**(1): p. 17-25.
225. Hansen, A.S., et al., *CTCF and cohesin regulate chromatin loop stability with distinct dynamics.* eLife, 2017. **6**.
226. Merkenschlager, M. and E.P. Nora, *CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation.* Annu Rev Genomics Hum Genet, 2016. **17**(1): p. 17-43.
227. van Ruiten, M.S. and B.D. Rowland, *On the choreography of genome folding: A grand pas de deux of cohesin and CTCF.* Curr Opin Cell Biol, 2021. **70**: p. 84-90.
228. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
229. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre.* Nature, 2012. **485**(7398): p. 381-385.
230. Tena, J.J. and J.M. Santos-Pereira, *Topologically Associating Domains and Regulatory Landscapes in Development, Evolution and Disease.* Front Cell Dev Biol, 2021. **9**.
231. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
232. Maeder, M.L., et al., *CRISPR RNA-guided activation of endogenous human genes.* Nat Methods, 2013. **10**(10): p. 977-9.
233. Liu, X.S., et al., *Editing DNA Methylation in the Mammalian Genome.* Cell, 2016. **167**(1): p. 233-247 e17.
234. Pulecio, J., et al., *CRISPR/Cas9-Based Engineering of the Epigenome.* Cell Stem Cell, 2017. **21**(4): p. 431-447.
235. Li, J., et al., *Programmable human histone phosphorylation and gene activation using a CRISPR/Cas9-based chromatin kinase.* Nat Commun, 2021. **12**(1): p. 896.
236. Jovic, D., et al., *Single-cell RNA sequencing technologies and applications: A brief overview.* Clin Transl Med, 2022. **12**(3): p. e694.
237. Hwang, B., J.H. Lee, and D. Bang, *Single-cell RNA sequencing technologies and bioinformatics pipelines.* Exp Mol Med, 2018. **50**(8): p. 1-14.
238. Ma, H., et al., *CRISPR-Cas9 nuclear dynamics and target recognition in living cells.* J Cell Biol, 2016. **214**(5): p. 529-37.
239. Gebauer, F. and M.W. Hentze, *Molecular mechanisms of translational control.* Nat Rev Mol Cell Biol, 2004. **5**(10): p. 827-35.
240. Leppek, K. and G. Stoecklin, *An optimized streptavidin-binding RNA aptamer for purification of ribonucleoprotein complexes identifies novel ARE-binding proteins.* Nucleic Acids Res, 2014. **42**(2): p. e13.

241. Ming Ma , F.Z., Xiongbing Hu , and X.-Z.W. Bolun Wang, Jia-Fu Ji, Jianzhong Jeff Xi, *Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system.* Cell Res, 2017. **27**: p. 578–581.
242. Natasa Savic, F.C.R., Helen Lindsay, Christian Berk, Katja Bargsten, Yizhou Li, Dario Neri, Mark D Robinson, Constance Ciaudo, *Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair.* eLife, 2018.
243. Keppler, A., et al., *A general method for the covalent labeling of fusion proteins with small molecules in vivo.* Nat Biotechnol, 2003. **21**(1): p. 86-9.
244. Lovendahl, K.N., A.N. Hayward, and W.R. Gordon, *Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags.* J Am Chem Soc, 2017. **139**(20): p. 7030-7035.
245. Eric J. Aird, K.N.L., Amber St. Martin, Reuben S. Harris, Wendy R. Gordon, *Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template.* Commun Biol, 2018. **1**.
246. Danyang Wang, W.D., Jian Wu, Jinke Wang, *Improving transcriptional activity of human cytomegalovirus major immediate-early promoter by mutating NF-kB binding sites.* Protein Expr Purif, 2017. **142**: p. 16-24.
247. Xu, X., et al., *Gene activation by a CRISPR-assisted trans enhancer.* Elife, 2019. **8**.
248. Lesueur, L.L., L.M. Mir, and F.M. Andre, *Overcoming the Specific Toxicity of Large Plasmids Electrotransfer in Primary Cells In Vitro.* Mol Ther Nucleic Acids, 2016. **5**(3): p. e291.
249. Cabantous, S., T.C. Terwilliger, and G.S. Waldo, *Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein.* Nat Biotechnol, 2005. **23**(1): p. 102-7.
250. Romei, M.G. and S.G. Boxer, *Split Green Fluorescent Proteins: Scope, Limitations, and Outlook.* Annu Rev Biophys, 2019. **48**: p. 19-44.
251. Okita, K., et al., *A more efficient method to generate integration-free human iPS cells.* Nature Methods, 2011. **8**(5): p. 409-412.
252. Zhang, C., et al., *Advances in mRNA Vaccines for Infectious Diseases.* Front Immunol, 2019. **10**: p. 594.
253. Kallunki, T., et al., *How to Choose the Right Inducible Gene Expression System for Mammalian Studies?* Cells, 2019. **8**(8).
254. Hartmann, D., et al., *Controlling gene expression with light: a multidisciplinary endeavour.* Biochem Soc Trans, 2020. **48**(4): p. 1645-1659.
255. Yun, S.H. and S.J.J. Kwok, *Light in diagnosis, therapy and surgery.* Nat Biomed Eng, 2017. **1**.
256. Darrah, K.E. and A. Deiters, *Translational control of gene function through optically regulated nucleic acids.* Chem Soc Rev, 2021. **50**(23): p. 13253-13267.
257. Shah, S., S. Rangarajan, and S.H. Friedman, *Light-activated RNA interference.* Angew Chem Int Ed Engl, 2005. **44**(9): p. 1328-32.
258. XinJing Tang, S.M., Eric S. Weinberg, and Ivan J. Dmochowski, *Regulating Gene Expression in Zebrafish Embryos Using Light-Activated, Negatively Charged Peptide Nucleic Acids.* J Am Chem Soc, 2007. **129**: p. 11000-11001.

259. Shestopalov, I.A., S. Sinha, and J.K. Chen, *Light-controlled gene silencing in zebrafish embryos.* Nat Chem Biol, 2007. **3**(10): p. 650-1.
260. Govan, J.M., et al., *Optochemical control of RNA interference in mammalian cells.* Nucleic Acids Res, 2013. **41**(22): p. 10518-28.
261. Bardhan, A., A. Deiters, and C.A. Ettensohn, *Conditional gene knockdowns in sea urchins using caged morpholinos.* Dev Biol, 2021. **475**: p. 21-29.
262. Deiters, A., et al., *Photocaged Morpholino Oligomers for the Light-Regulation of Gene Function in Zebrafish and Xenopus Embryos.* J Am Chem Soc, 2010. **132**(44): p. 15644-15650.
263. Yamazoe, S., et al., *Sequential gene silencing using wavelength-selective caged morpholino oligonucleotides.* Angew Chem Int Ed Engl, 2014. **53**(38): p. 10114-8.
264. Tang, X., et al., *Photomodulating RNA cleavage using photolabile circular antisense oligodeoxynucleotides.* Nucleic Acids Res, 2010. **38**(11): p. 3848-55.
265. Nguyen, Q.N., et al., *Light controllable siRNAs regulate gene suppression and phenotypes in cells.* Biochim Biophys Acta, 2006. **1758**(3): p. 394-403.
266. Wang, Q., et al., *Photoregulation of Gene Expression with Ligand-Modified Caged siRNAs through Host/Guest Interaction.* Chembiochem, 2021. **22**(11): p. 1901-1907.
267. Wang, Y., et al., *Manipulation of gene expression in zebrafish using caged circular morpholino oligomers.* Nucleic Acids Res, 2012. **40**(21): p. 11155-62.
268. Hammill, M.L., C. Isaacs-Trépanier, and J.-P. Desaulniers, *siRNAzos: A New Class of Azobenzene-Containing siRNAs that Can Photochemically Regulate Gene Expression.* ChemistrySelect, 2017. **2**(30): p. 9810-9814.
269. Hammill, M.L., G. Islam, and J.P. Desaulniers, *Reversible control of RNA interference by siRNAzos.* Org Biomol Chem, 2019. **18**(1): p. 41-46.
270. Hammill, M.L., et al., *Synthesis, Derivatization and Photochemical Control of an ortho-Functionalized Tetrafluorinated Azobenzene-Modified siRNA.* Chembiochem, 2022. **23**(20): p. e202200386.
271. Berdnikova, D.V., *Photoswitches for controllable RNA binding: a future approach in the RNA-targeting therapy.* Chem Commun, 2021. **57**(83): p. 10819-10826.
272. Rotstan, K.A., et al., *Regulation of mRNA translation by a photoriboswitch.* Elife, 2020. **9**.
273. Reichert, D., H.D. Mootz, and A. Rentmeister, *Light-control of cap methylation and mRNA translation via genetic code expansion of Ecm1.* Chem Sci, 2021. **12**(12): p. 4383-4388.
274. Ando, H., et al., *Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos.* Nat Genet, 2001. **28**(4): p. 317-25.
275. Alexander, S.C., et al., *Site-Specific Covalent Labeling of RNA by Enzymatic Transglycosylation.* J Am Chem Soc, 2015. **137**(40): p. 12756-9.
276. Zhang, D., et al., *Light-Activated Control of Translation by Enzymatic Covalent mRNA Labeling.* Angew Chem Int Ed Engl, 2018. **57**(11): p. 2822-2826.
277. Zhang, D., et al., *Multiplexed Photoactivation of mRNA with Single-Cell Resolution.* ACS Chem Biol, 2020. **15**(7): p. 1773-1779.

278. Ovcharenko, A., F.P. Weissenboeck, and A. Rentmeister, *Tag-Free Internal RNA Labeling and Photocaging Based on mRNA Methyltransferases.* Angew Chem Int Ed Engl, 2021. **60**(8): p. 4098-4103.
279. Jackson, R.J., C.U.T. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 113-127.
280. Ogasawara, S. and M. Maeda, *Photoresponsive 5'-cap for the reversible photoregulation of gene expression.* Bioorg Med Chem Lett, 2011. **21**(18): p. 5457-9.
281. Ogasawara, S., *Control of cellular function by reversible photoregulation of translation.* Chembiochem, 2014. **15**(18): p. 2652-5.
282. Ogasawara, S., *Duration Control of Protein Expression in Vivo by Light-Mediated Reversible Activation of Translation.* ACS Chem Biol, 2017. **12**(2): p. 351-356.
283. Klocker, N., et al., *Photocaged 5' cap analogues for optical control of mRNA translation in cells.* Nat Chem, 2022. **14**(8): p. 905-913.
284. Bollu, A., et al., *Light-Activated Translation of Different mRNAs in Cells via Wavelength-Dependent Photouncaging.* Angew Chem Int Ed Engl, 2022.
285. Hagihara, S., et al., *Production of truncated protein by the crosslink formation of mRNA with 2'-OMe oligoribonucleotide containing 2-amino-6-vinylpurine.* Bioorg Med Chem Lett, 2012. **22**(12): p. 3870-2.
286. Sugihara, Y., et al., *Cross-Linking Antisense Oligodeoxyribonucleotides with a Photoresponsive alpha-Chloroaldehyde Moiety for RNA Point Mutations.* J Org Chem, 2016. **81**(3): p. 981-6.
287. Yamada, K., et al., *Synthesis and Properties of 2'-OMe-RNAs Modified with Cross-Linkable 7-Deazaguanosine Derivatives.* J Org Chem, 2018. **83**(16): p. 8851-8862.
288. Higuchi, M., et al., *Synthesis of antisense oligonucleotides containing 2'-O-psoralenylmethoxyalkyl adenosine for photodynamic regulation of point mutations in RNA.* Bioorg Med Chem, 2009. **17**(2): p. 475-83.
289. Kikuta, K., Y. Taniguchi, and S. Sasaki, *Study of the Inducible Cross-Linking Reaction to mRNA and the Effect on the Translation.* Chem Pharm Bull, 2019. **67**(8): p. 877-883.
290. Bochkov, Y.A. and A.C. Palmenberg, *Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location.* Biotechniques, 2006. **41**(3): p. 283-4, 286, 288 passim.
291. Kelkar, D.A., et al., *Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate.* J Biol Chem, 2012. **287**(4): p. 2568-78.
292. De Gregorio, E., T. Preiss, and M.W. Hentze, *Translational activation of uncapped mRNAs by the central part of human eIF4G is 5' end-dependent.* RNA, 1998. **4**(7): p. 828-836.
293. Nowakowska, M., et al., *Cap analogs containing 6-thioguanosine--reagents for the synthesis of mRNAs selectively photo-crosslinkable with cap-binding biomolecules.* Org Biomol Chem, 2014. **12**(27): p. 4841-7.

294. Mahdavi, B., et al., *Construction and functional evaluation of a polycistronic modified mRNA for human iPS cell generation.* Cytotherapy, 2019. **21**(5, Supplement): p. S23.
295. Yoshioka, N., et al., *Efficient generation of human iPSCs by a synthetic self-replicative RNA.* Cell stem cell, 2013. **13**(2): p. 246-254.
296. Muller, S. and B. Appel, *In vitro circularization of RNA.* RNA Biol, 2017. **14**(8): p. 1018-1027.
297. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-345.
298. Goddard, T.D., et al., *UCSF ChimeraX: Meeting modern challenges in visualization and analysis.* Protein Sci, 2018. **27**(1): p. 14-25.
299. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-682.
300. Sarabipour, S., C. King, and K. Hristova, *Uninduced high-yield bacterial expression of fluorescent proteins.* Anal Biochem, 2014. **449**: p. 155-7.
301. Ding, Q., et al., *Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs.* Cell Stem Cell, 2013. **12**(4): p. 393-4.
302. Okita, K., et al., *A more efficient method to generate integration-free human iPS cells.* Nat Methods, 2011. **8**(5): p. 409-12.
303. To, T.L., et al., *Rational Design of a GFP-Based Fluorogenic Caspase Reporter for Imaging Apoptosis In Vivo.* Cell Chem Biol, 2016. **23**(7): p. 875-882.