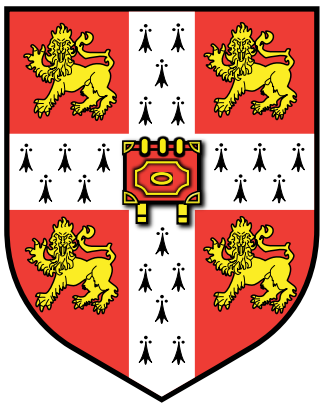
Nucleic Acid Scaffold-dependent Proximity-mediated Enzyme Response (NASPER) – A Proof of Concept Study



Nilesh Chatterjee

Christ's College
University of Cambridge

April 2022

This dissertation is submitted for the degree of Doctor of Philosophy

DECLARATION

This dissertation is submitted for the degree of Doctor of Philosophy. It is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. It is not substantially the same as any that I have submitted or is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the text. It does not exceed the prescribed word limit of 60,000 words.

Nilesh Chatterjee

April 2022

ABSTRACT

Nucleic **A**cid **S**caffold-dependent **P**roximity-mediated **E**nzyme **R**esponse (**NASPER**) – A Proof of Concept Study

Nilesh Chatterjee

Telomerase *hTERT* RNA is overexpressed in around 90% of all cancers but targeting it has been unsuccessful to date due to the inability of this approach to kill telomerase-expressing cells leading to the evolution of telomerase-independent cells. The approach proposed in the thesis (NASPER), aims to target cells overexpressing *hTERT* RNA and cause their apoptosis, preventing this evolution and debulking the tumour mass. NASPER involves bringing two fusion proteins, each of which comprises a custom-designed PUF (Pumilio and FBF) RNA-binding protein and a protease, onto the *hTERT* RNA into close proximity to activate the protease which should lead to cell death. The proteases tested in this study are the HIV protease (HIVPR) and the split-TEV protease, which require dimerisation/fragment complementation for catalytic activity. In Chapter 3, the two designed PUF proteins were first purified and tested *in vitro* using fluorescence polarisation experiments to assess RNA binding. The results indicate that the PUF proteins bind specifically to their cognate sequences both independently and in combination. In Chapter 4, expression in *E. coli* and purification of HIVPR was optimised, and its activity was confirmed in vitro. In cells, the individual PUF-HIVPR fusion proteins appeared to auto-activate without their dimerisation partner despite the use of mutations to reduce auto-activity, and it was concluded that a split system is required such that the protease can only undergo activation when both proteins are present. In Chapter 6, a split-GFP system (fusing two GFP fragments to each of the two PUF proteins) most notably show that, as intended, the fusion proteins bind better when cognate RNA is used versus scrambled RNA, and that engineering the fusion proteins such that the split-GFP are domains oriented towards each other results in improved re-constitution of GFP. Subsequently, a NASPER system designed with the split-TEV protease was tested, in which two TEV protease fragments are fused to each of the two PUF domains. The results were

consistent with the split-GFP findings, and NASPER was also able to target overexpressed *hTERT* mRNA in HeLa cells. No cleavage of a procaspase-3 construct could be detected in apoptosis assays in Chapter 8, and further work is required to elucidate the cause. Chapter 9 describes further experiments to understand the binding characteristics of designed PUF domains using crosslinked RNA-seq methods. Overall, the results presented in the thesis provide new insights into the behaviour of designed PUF proteins in the cell and lay the groundwork for a new therapeutic approach based on targeted protease-induced cell death, as discussed in Chapter 10.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Professor Laura Itzhaki for inviting me into her lab to carry out a completely new project that I had thought of – her generosity and support is what has helped me thrive during my time in Pharmacology. I was able to go out of my comfort zone and explore different areas of research because of how open-minded she was to new types of experiments and technologies. She was also incredibly supportive when I decided the next steps for me were to study Medicine, and for that I will be forever grateful.

I would also like to thank every member of the Itzhaki lab with whom I have had the pleasure of meeting, working, and relaxing with. I'd like to thank Pam for always being there to answer my 'quick questions' and essentially teaching me everything I know about how to do experiments and how to be competent in a lab, and for dealing with me when I made mistakes! I would also like to say thanks to Janet who has always been there to reassure me when I would get worried and overthink about my experiments, thesis, data, you name it! A special thanks to my Pharmacology buddies Jas and Michelle – I would never have survived this long without knowing you guys were going through it too, and without our unending banter on a depressing day.

Even though I was not able to see my family for two years during the PhD, I would like to thank all of them, especially my mother and father who have always been there for me no matter what. Even though sometimes I do not call and annoy you to no end, you have always managed to cheer me up on a sad day and completely supported my desire to move to Medicine, without which I may never have had the confidence to do it. Special mention to my little sister who decided to trap herself in the UK for three months just to hang out with me. She made life a thousand-fold more fun until she decided to leave.

Finally, I'd like to thank the person without whom I, without question, would have never made it this far – Alejandra Martin Segura. You held my hand on every step, went through all the stress, anxiety, and worries with me without ever wavering. There are not enough words to express how grateful I am for you being in my life. This thesis is as much yours as it is mine.

TABLE OF CONTENTS

COPYRIGHT

Reuse licenses or permissions have been obtained for the following figures:

Figure	Publisher	License Number
1.2 A	Elsevier	5290241305551
1.2 B	Elsevier	5290241429995
1.3	F1000 Research	Creative Commons Attribution License
7.2	Springer Nature	5333561081348

LIST OF ABBREVIATIONS

6-FAM	6-carboxyfluoroscein
A_{260nm}	Absorbance at 260 nm
A_{280nm}	Absorbance at 280 nm
ALT	Alternative Lengthening of Telomeres
BSA	Bovine Serum Albumin
CLIP	Crosslinked Immunoprecipitation
DEAE	Diethylaminoethyl
DN	Dominant Negative
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotides
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic Acid
EMSA	Electrophoretic Mobility Shift Assay
FL	Full length
FP	Fluorescence Polarisation
FPLC	Fast Protein Liquid Chromatography
FRET	Förster Resonance Energy Transfer
GFP	Green Fluorescent Protein
GST	Glutathione S Transferase
HA	Haemagglutinin tag
HIV	Human Immunodeficiency Virus
HIVPR	Human Immunodeficiency Virus Protease
Hislip	His-lipoyl
HRV3C	Human Rhinovirus 3C
IL	Interleukin
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β- d-1-thiogalactopyranoside
K_d	Dissociation Constant
kDa	Kilo Daltons

LB	Luria Bertani
MBS	Multiple Binding Sites
mtRNA	Mitochondrial Ribonucleic Acid
MWCO	Molecular Weight Cutoff
NASPER	Nucleic Acid Scaffold-dependent Proximity-mediated Enzyme Response
NGS	Next-generation Sequencing
Ni-NTA	Nickel Nitrilo Triacetic Acid
NMR	Nuclear Magnetic Resonance
NSCLC	Non-Small Cell Lung Cancer
OD_{600nm}	Optical Density at 600 nm
OOPS	Organic Orthogonal Phase Separation
PFS	Progression-free Survival
PB	Phosphate Buffer
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PROTAC	Proteolysis Targeting Chimera
PUF	Pumilio and FBF
RFU	Relative Fluorescence Units
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
rpm	Revolutions Per Minute
RTK	Receptor Tyrosine Kinase
RTPCR	Reverse Transcription Polymerase Chain Reaction
SBS	Single Binding Site
SEC	Size Exclusion Chromatography
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SOC	Super Optimal Broth with Catabolite Repression
STED	Stimulated Emission Depletion Microscopy
TAE	Tris Acetate EDTA
TALE	Transcription Activator-like Effectors
TEV	Tobacco Etch Virus

TEVPR	Tobacco Etch Virus Protease
TGS	Tris Glycine SDS
TIRF	Total Internal Reflection Microscopy
TNF	Tumour Necrosis Factor
Tris	tris(hydroxymethyl)aminomethane
UTR	Untranslatable Region
UV	Ultraviolet
UV-CLIP	Ultraviolet Crosslinking Immunoprecipitation
V	Volts
HIVPR	Human Immunodeficiency Virus Protease
YT	Yeast Tryptone

CHAPTER 1 – A GENERAL INTRODUCTION

1.1 EPIDEMIOLOGY OF CANCER

Cancer is one of the leading causes of death in wealthy countries causing around 1 in 6 deaths worldwide (WHO). GLOBOCAN (https://gco.iarc.fr/, updated in 2018) currently estimates a global 5-year cancer prevalence of 43,841,302 people, however, this suggestion of approximately 0.57% of the population living with cancer may be misleading[1]. First, the CONCORD-3 study[2] has shown that the age-standardised 5-year net survival rate for the most common cancers is increasing with time and is over 90% in some cases such as breast cancer, suggesting that a 5-year cancer prevalence statistic may underestimate the true prevalence of the disease. Second, contrary to its high prevalence, global age-standardised mortality rates and disease burdens are falling[3] (Figure 1.1), but the incidence of cancer is rising rapidly with 18.1 million new cases in 2018 (compared to previous years) and predicted to increase significantly by 2030[1]. Although age-standardised indices show falling death rates, this implies that by 2030 the absolute number of deaths due to cancer will rise significantly as well (trend in total deaths also seen in Figure 1.1A), suggesting that treatment strategies, although steadily improving, cannot balance the rapid population increases causing the increase in absolute cancer mortality. The slow reduction in standardized mortality rates and disease burden is a positive view on the current global state of cancer; however, this improvement does not change the fact that the number of people dying from the disease is rapidly rising and, thus, more effective therapies that can balance the rising absolute mortality are needed.

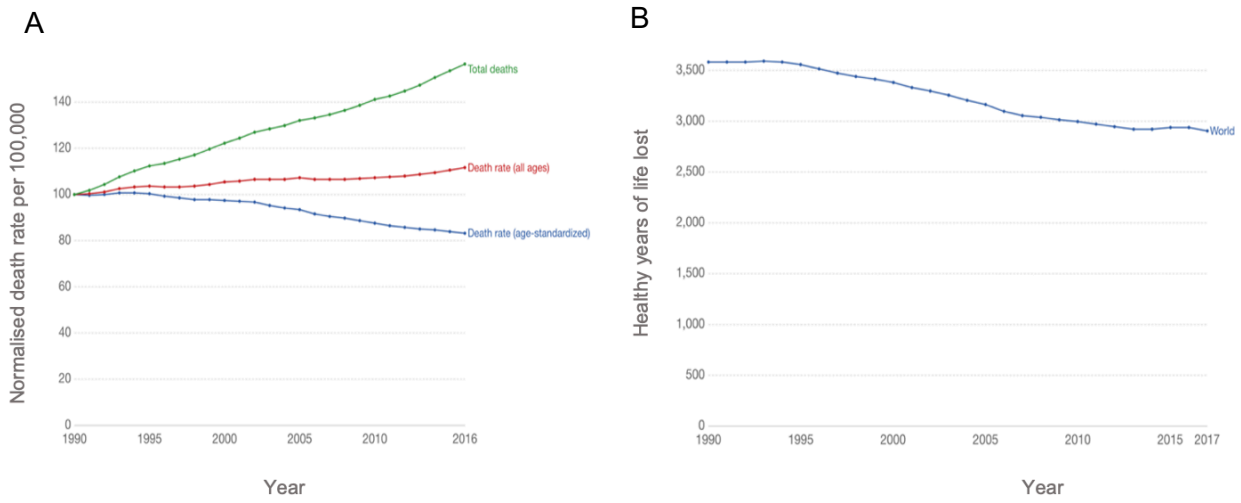


Figure 1.1: Taken from the Institute for Health Metrics and Evaluation. **A** – Total and age-standardised cancer death rates per 100,000 normalised to the year 1990. **B** – Global disease burden from cancer shown as disability-adjusted life years lost.

1.2 OVERVIEW OF BIOLOGY OF CANCER

To begin developing novel therapies for the treatment of cancer, the biology must first be understood. Cancer is a diverse set of diseases that can originate within nearly any organ and from nearly any cell type due to intrinsic or extrinsic factors[4]; however, all cancers are generally characterised by a set of hallmarks[5] as shown in Figure 1.2. Acquisition of some or all of these capabilities described leads to the pathogenesis of cancer, resulting in a cell that can freely proliferate past the Hayflick limit (a set number of cell divisions prior to inevitable senescence[6]) and that may be able infiltrate other organs due to vascular access and metastatic ability resulting in a spread of the cancer. In addition to these hallmarks, novel emerging hallmarks have been described (Figure 1.2B), which further enhances the ability of a cancer cell to survive and proliferate. The most common mechanisms by which somatic cells acquire these hallmarks are through the activation of oncogenes or the deactivation of tumour suppressor genes[5]. These phenomena can occur in cells due to mutation, chromosomal amplifications and deletions, and epigenetic mechanisms leading to aberrant expression of proteins resulting in the deregulation of pathways essential for cellular homeostasis such as the apoptotic pathways[5].

The number of cellular pathways that can be dysregulated in order to attain the phenotypes shown in Figure 1.2 are vast, thereby resulting in cancer being a hugely diverse disease. In addition, cancer cells are genotypically dynamic due to their genomic instability and mutational capacity (Figure 1.2B). This instability is the driving force in cancer cell evolution[7], resulting in inter- and intra-tumour heterogeneity. For example, there are various different subtypes of breast cancer tumours, such as luminal A and B, HER2+ and triple-negative[8], underlining inter-tumour heterogeneity. Even within a single tumour type there may be clonal genetic variations[8,9], as determined robustly by single-cell deep sequencing, resulting in intra-tumour heterogeneity and diversity.

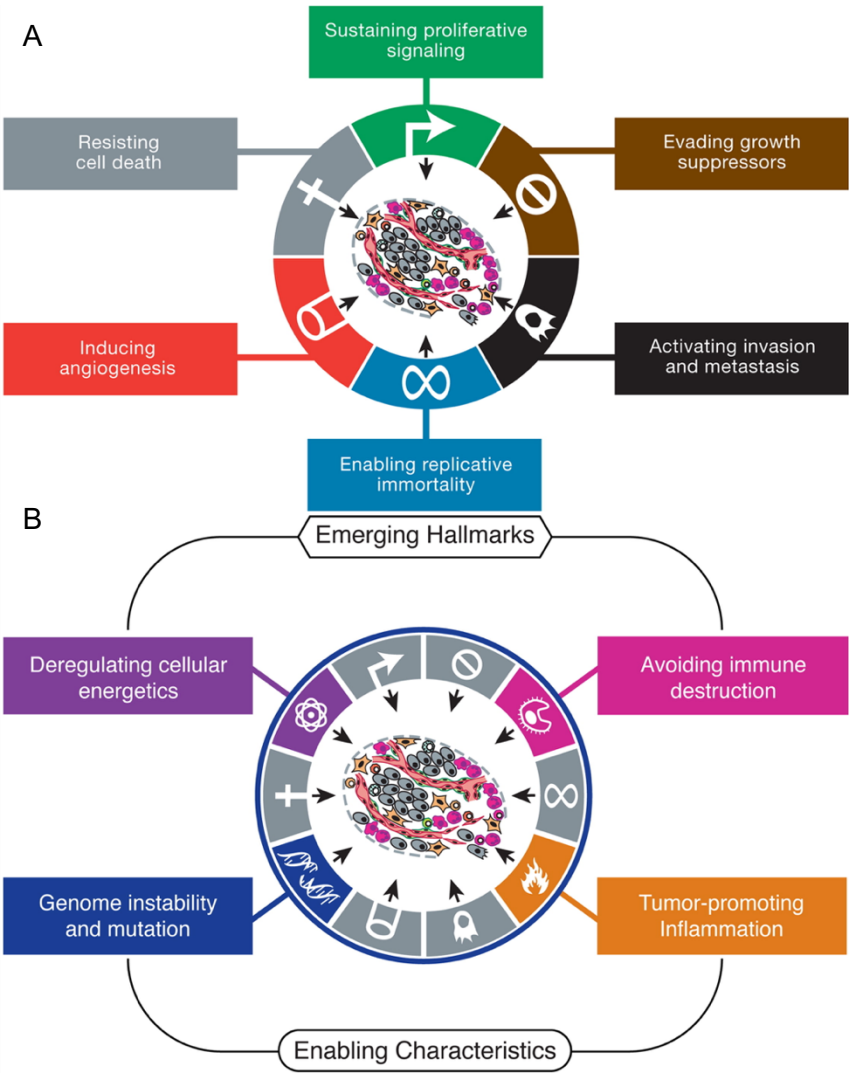


Figure 1.2: Adapted from Hanahan et al[5]. **A** – Original six hallmarks of cancer. **B** – Novel emerging hallmarks and enabling characteristics for cancer development and progression.

However, there is a spectrum of genomic instability in cancer, for example, the same cancer subtypes in different patients may have different degrees of genomic instability which correlates with disease progression and prognosis[10]. Despite this, tumour heterogeneity is omnipresent in cancer suggesting the presence of mechanisms alternative to genomic instability for this heterogeneity to develop, such as the effects of the tumour microenvironment[11] and drug treatment itself.

Various studies have shown that the presence of intra-tumour heterogeneity and clonal diversity in cancer results in poorer patient prognoses[12-15], sometimes due to the evolutionary drug resistance mechanism shown in Figure 1.3. In addition to this mechanism, susceptible clones can also evolve and develop resistance to treatments over time by various mechanisms, further increasing clonal diversity[16]. Hence, intra-tumour heterogeneity and clonal diversity pose a barrier to successful cancer therapy and there are different ways to overcome this barrier: 1) multidrug therapy to eradicate single drug-resistant clones, and 2) targeting a shared phenotype across cancer cells coupled to rapid killing of the positive cells.

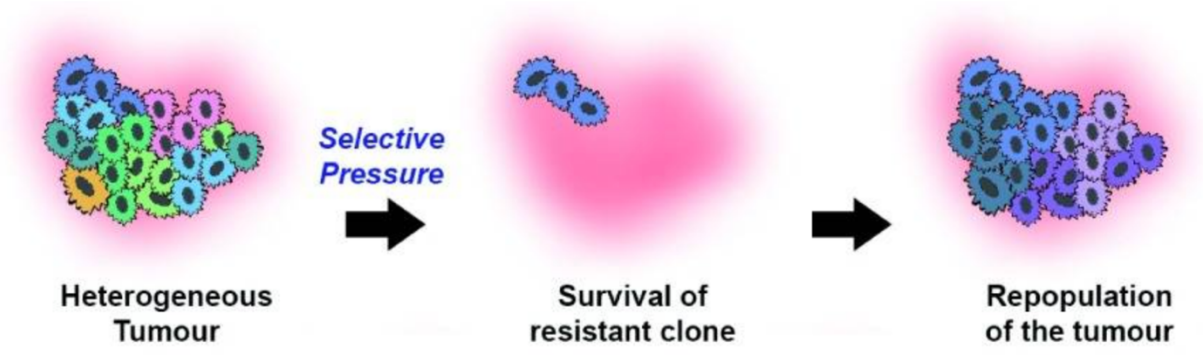


Figure 1.3: Adapted from Gay et al[13]. Tumour relapse due to survival of a drug resistant clone from a heterogenous tumour under selection pressure due to drug treatment.

One such phenotype that is necessary for cancer cell survival is 'enabling replicative immortality', which is restricted to cancer cells and pluripotent stem cells, and unseen in somatic cells. This phenotype can be achieved through different genotypic mechanisms such as aberrations in the p53 pathway[17], which occurs in around 50% of all cancers[18]. The protein p53 is a nuclear transcription factor often referred to as the "guardian of the genome" due to

its role in maintaining genomic integrity. Under normal conditions, it is maintained at very low levels in the cell due to ubiquitination and degradation by MDM2 and the proteosome respectively. When DNA damage occurs, p53 is phosphorylated and stabilised allowing translocation to the nucleus and subsequent transcription of its target genes which induce apoptosis[19]. This prevents potentially malignant cells from replicating, thereby acting as a tumour suppressor. Loss-of-function mutations in the p53 gene, or its co-factors, can therefore result in the ability of cells to divide regardless of DNA damage leading to genomic instability and replicative immortality. Although this results in immorality of 50% of cancers, there is a more universal change that results in immortality of cells, i.e., the overexpression of active telomerase[20] in 90% of cancers. Thus, targeting telomerase would allow a single therapy to be used for a huge range of different cancer types. Even in heterogenous tumours containing the 10% of telomerase-independent cancer cells, targeting and killing telomerase-positive cells could be used to debulk tumours and make them more amenable to other interventions, such as surgery.

1.3 TELOMERASE AND ITS FUNCTIONS

Telomerase is a ribonucleoprotein holoenzyme responsible for the maintenance and stabilisation of chromosome ends known as telomeres (Figure 1.4). The holoenzyme is composed of an RNA template-containing subunit, TERC, and a catalytic protein subunit, TERT, which is rate-limiting in complex formation[21]. Under normal circumstances, human chromosomes are capped by telomeric TTAGGG repeats because incomplete replication of the lagging strand results in erosion of chromosome ends – the end replication problem[22,23]. Telomere shortening leads to DNA damage signal induction resulting in a temporary arrest of the cell cycle. DNA damage repairs are initiated and if the damage due to telomere shortening is irreparable, the cell undergoes senescence. If not, the damage is repaired, and the cell cycle continues as normal. Thus, it is the cumulative effect of DNA damage at a single or multiple telomeres that leads to senescence. This 'age timer' phenomenon is also known as the Hayflick limit, i.e. a limit to the number of replicative cycles that a somatic cell can undergo. Telomerase has the ability to elongate these shortening telomeres by reverse transcription. Telomerase binds to telomeres using the TERC subunit and adds telomeric repeats sequentially using the TERT subunit[23]. Human telomeres are highly polymorphic, and the telomerase enzyme is not

known to prefer the telomeres of any particular chromosome. However, experiments in *S. cerevisiae* have shown that telomerase does not extend all telomeres in one cell cycle, but rather acts preferentially on chromosomes with shorter telomeres which are in an extendible state due to the absence of secondary structures (such as G-quadruplexes shown in Figure 1.5), promoting accessibility[24]. Thus, cells expressing active telomerase can evade replicative senescence.

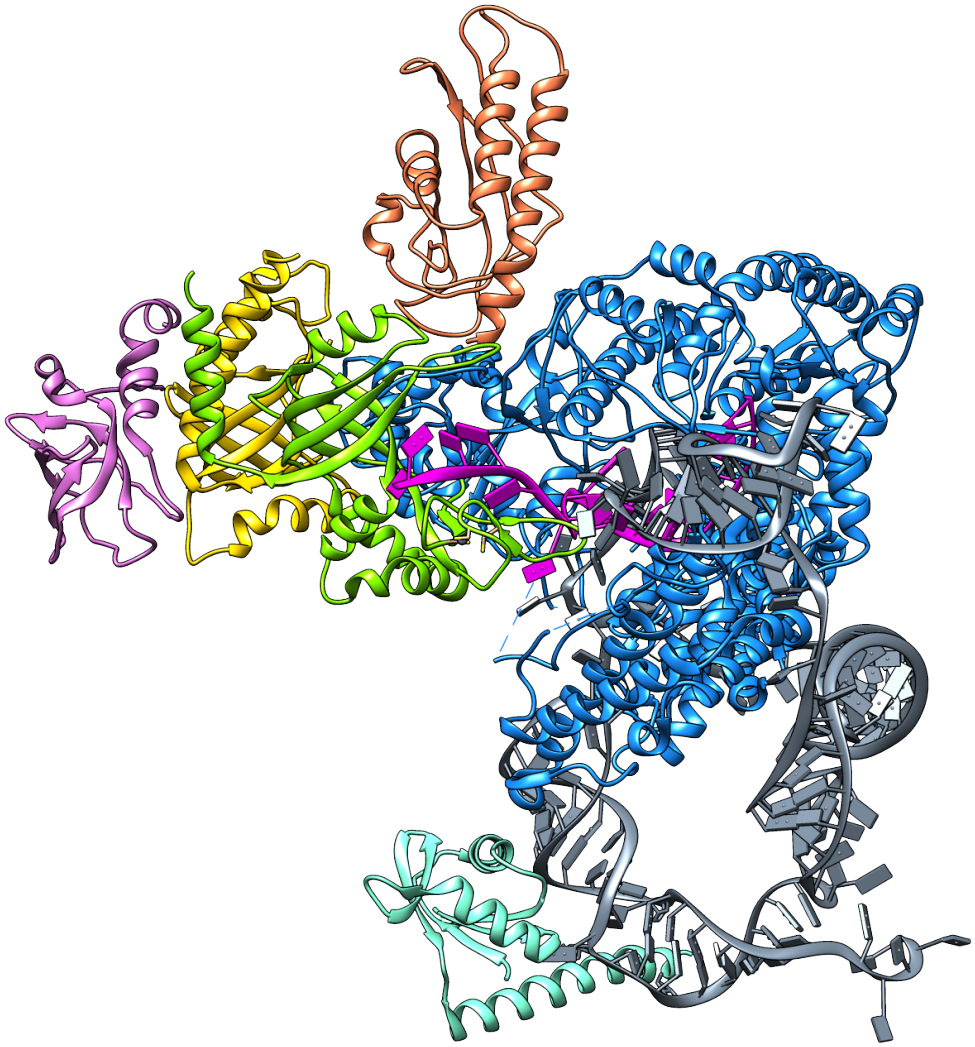


Figure 1.4: **A**: Schematic of the structure of telomerase holoenzyme (PDB: 6D6V). Blue – TERT subunit, grey – TERC subunit, magenta – telomere DNA, coral – telomerase associated protein 50, aquamarine – telomerase associated protein 65, chartreuse – telomerase associated protein 82, gold – TEB2, pink – TEB3.

Figure 1.5: Schematic of the structure of a G-quadruplex at a telomere (PDB: 143D). Grey – phosphate backbone, blue – sugars and nucleotides.

In 90% of cancers[20] this Hayflick limit is evaded through the overexpression of full-length *hTERT* mRNA. Subsequent increase in active telomerase concentration facilitates the maintenance of shortened telomeres and promotes the evasion of apoptosis[25]. There are various genetic and epigenetic mutational mechanisms of FL-*hTERT* upregulation in cancer cells including *hTERT* amplifications (3%), structural variants (3%), promoter mutations (31%) and promoter hypermethylation (53%)[25,26]. Conventionally, promoter hypermethylation leads to gene silencing, however *hTERT* is an exception to this perhaps because transcriptional repressors require unmethylated CpG islands in the promoter to bind[27]. Thus, hypermethylation prevents this binding and repression. The diversity of mutations within these classes makes targeting these individual mutations inefficient[25,26,28], and thus *hTERT*-driven treatment strategies focus on the consequence of these mutations, i.e., increased *hTERT* activity. In somatic cells, telomere shortening with every replicative cycle which eventually results in cellular senescence, and this was originally believed to be due to a lack of *hTERT* expression. However, this misunderstanding was due to a flawed methodology

in quantitating *hTERT* expression by RT-PCR. Initially, the primers used targeted the inner reverse transcriptase domain of the gene[29], but it has now been discovered that several variants of *hTERT* exist lacking parts of this domain[22,29-32] resulting in the false conclusion that somatic cells do not express *hTERT*.

To date, several alternative splice variants of *hTERT* have been discovered, including the Δα (deletion in exon 6), Δβ (deletion in exon 7 and 8), ΔαΔβ and Δ4-13 (deletion in exon 4 – 13) variants[29-32]. Of all the variants, however, none of them have produced a protein with telomerase activity, although, certain variants may be able to promote cell proliferation and inhibit apoptosis[29]. Two studies, by *Hrdličková et al*[31] and *Yi et al*[32], have revealed differential expression of the variants between normal, stem and cancer cells.

Hrdličková et al[31] showed that most normal cell types also express *hTERT*, but, they express large deletion variants, such as Δ4-13, that lack any telomerase activity. However, the spleen, thymus, brain, testes, and placenta appear to also express variants lacking the Δα and Δβ mutations. Furthermore, the thymus, testes, and placenta – tissues containing stem cell populations[33-35] – do contain full-length *hTERT*[31], as these stem cells require the ability to replicate infinitely. Their analysis of lung, heart, and ovarian tissues (lacking stem cell capabilities) however suggests that no full-length *hTERT* is present, only the Δ4-13 variant and other inactive variants, that ablate telomerase activity. This finding further suggests that tissues lacking any stem cell activity are devoid of full-length *hTERT* and thus active telomerase. These data have been extended to normal cell lines, such as the BJ fibroblasts and IMR90 cells, which do not contain any full-length *hTERT*[30-32], and cancer cell lines such as the HeLa, MCF-7 and MDA-MB-231 cells, which contain full-length *hTERT*[32] and display telomerase activity.

These data show that somatic cells indeed contain *hTERT* RNA variants and non-functioning telomerase protein. Hence, using telomerase as a target must rely upon targeting regions that are present in the full-length protein and not the variants found in somatic cells to minimize side-effects. On the other hand, the activity of telomerase itself may be targeted at a different stage, such as by stabilising the telomeres in the inaccessible G-quadruplex state, rather than by using a small molecule inhibitor to bind to the enzyme itself.

Interestingly, recent findings suggest that telomerase may also have other functions in the cell besides simply the elongation of telomeres. It has been implicated in transcriptional regulation of various signalling pathways and apoptosis regulation. Choi *et al.*[36] found that overexpression of *hTERT* led to transcriptional changes that mimicked those mediated by Wnt signalling[37,38] due to TERT interactions with SMARCA4. This results in cell proliferation suggesting a non-telomere-based proliferative role for telomerase in cells. Furthermore, telomerase can also activate the NF- ϰB pathway by interacting with the p65 subunit, resulting in increased transcription of NF- ϰB regulated genes such as IL-6 and TNF-α, both of which are implicated in cancer progression[39]. The influence of telomerase in these pathways also suggests a role for telomerase in epithelial-to-mesenchymal transition and cancer invasiveness as well[40]. Studies have also shown that telomerase can act as an inhibitor of the intrinsic pathway of apoptosis by enhancing the activity of the anti-apoptotic protein Bcl-2[41]. Recent advances have shown that telomerase contains a BH3-like motif – a peptide sequence found in Bcl-2 family proteins – which may allow it to interact with Bcl-2 and enhance its activity[42]. Telomerase has also been found to induce a conformational change in the pro-apoptotic protein Bax (potentially through interactions via the BH3-like domain since Bax belongs to the Bcl-2 family of proteins), preventing its activation and conferring cellular resistance to cytotoxic agents such as cisplatin[43]. Thus, besides 'enabling replicative immortality', telomerase is also implicated in other hallmarks of cancer such as 'resisting cell death', 'sustained proliferative signalling', 'activation of invasion and metastasis' and 'tumour-promoting inflammation'.

1.4 TELOMERASE-BASED THERAPIES

A number of cancer therapies targeting telomerase have been developed, including anti-sense TERC oligonucleotides such as Imetelstat[44,45], G-quadruplex stabilisers[44,45,46] such as CX-5461, and mutant hTERT expression[21,44,45].

Imetelstat, a 13-mer anti-sense TERC oligonucleotide that competitively inhibits telomerase enzymatic activity, has been evaluated in several Phase clinical 2 trials for non-small cell lung cancer (NSCLC)[47] and myeloid malignancies (NCT02426086). The Phase II clinical trial for NSCLC with imetelstat reported by *Chiappori et al*[47] did not show a significant improvement in progression free survival (PFS) in NSCLC, however, the trial design resulted in small patient

sub-groups in addition to the 2:1 randomisation which may have biased the end results. Haematological side-effects were significantly increased in those receiving imetelstat, likely due to effects on haematopoietic precursor cells[48]. In another study, *Wang et al.*[49] showed the opposite effect on haematopoietic stem cells, and hence the haematological effects may be mediated by the inhibition of extra-telomeric functions by imetelstat of telomerase in these haematopoietic precursor and stem cells[40]. The likely reason for the failure of imetelstat is that it does not cause cancer cells to undergo rapid cell death, thereby providing the necessary selection pressure allowing the cells to evolve and acquire alternate methods of lengthening telomeres (ALT)[50].

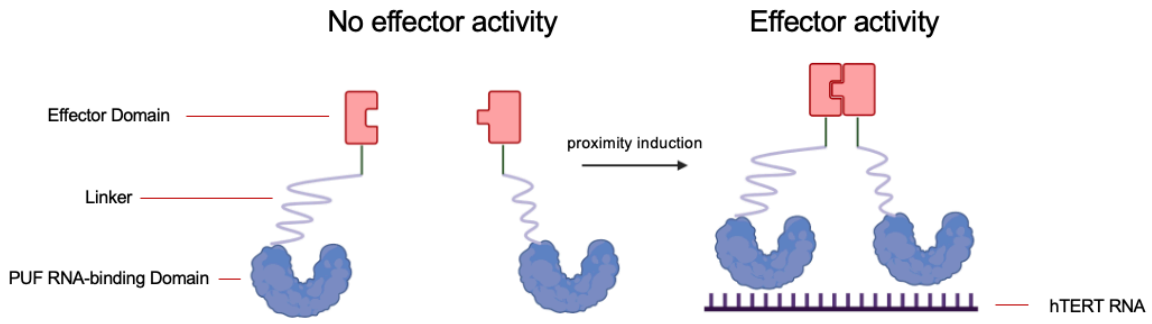
CX-5461 is a G-quadruplex-stabilising ligand that has been tested in a Phase 1 trial for BRCA1/2 deficient tumours[46]. G-quadruplexes (Figure 1.5) are four-stranded secondary structures of DNA formed due to the stacking of guanine residues in a planar arrangement found commonly at telomeres[44]. These structures prevent telomeric access to telomerase, thus, stabilising them results in the prevention of telomere maintenance by telomerase[44]. This results in enhanced DNA damage at G-quadruplexes resulting in apoptosis, albeit, only in DNA damage repair deficient cancers[46]. This area of research has been applied widely to different types of cancers using many different G-quadruplex ligands[51], however very few have proceeded on to clinical trials due to the challenges associated with these ligands such as specificity towards G-quadruplexes only while not binding to other nucleic acids[52].

The expression of a mutant, dominant negative *hTERT* (DN-*hTERT*) in cancer cell lines was also tested in order to reduce functional telomerase in cancer cells[45]. This method was shown to induce cell arrest in MCF-7 cells, however there are issues in translating this into pre-clinical models. The expression of DN-*hTERT* needs to be maintained in treated cells, which is difficult to achieve due to transgene promoter methylation *in vivo*[53]. Furthermore, imetelstat works using a very similar principle, i.e. interfering with the enzymatic activity of the telomerase protein and has been unsuccessful in clinical trials so far. Thus, it appears that these drugs or treatments may work in cells, their activity is difficult to translate into humans and this may be due to the inability of these treatments to result in apoptosis. The inhibition of telomerase in these cells leads to chromosomal instability, which should result in senescence. However, this p53-dependent mechanism may also be impaired in cancer cells, and thus senescence does

not occur[54]. Hence, there is a time lag from telomerase inhibition to telomere shortening to the point of senescence (or continued growth), during which cells can acquire more mutations potentially allowing them to continue proliferating through other mechanisms such as ALT (delay issue of *hTERT*-driven therapies). However, despite the hurdles faced by all three types of therapies, they were specific and had minimal effects on non-transformed somatic cells, suggesting that *hTERT*-targeting treatment strategies are cancer cell-specific[44-46].

1.5 THE NASPER SYSTEM

It is clear that although telomerase is a highly specific discriminator of cancer cells, its inhibition has not yet been fruitful in cancer therapies. Thus, I attempted to devise a strategy where cancer cells are specifically targeted using the presence of the *hTERT* mRNA (since overexpression of this mRNA is implicated in increasing telomerase activity in cancer) and this targeting is used to induce immediate cell death. I named this approach the Nucleic Acid Scaffold-dependent Proximity-mediated Enzyme Response system (NASPER system) and it is designed to overcome the intrinsic flaws in aforementioned *hTERT*-driven treatment strategies. The figure below illustrates the hypothesised mechanism of action of NASPER.



The primary protein used in this system is the PUF RNA-binding protein which will be discussed in Section 1.10. The rationale behind NASPER is that the two PUF proteins will only bind in the correct orientation in *hTERT*+ cells such that the fused effector domains undergo dimerisation and functional reconstitution due to proximity, thereby resulting in

activity. Three effector domains will be tested in this thesis – the HIV protease, split-GFP and split-TEV. The HIV protease and split-TEV effectors are intended to induce apoptosis through non-specific proteolysis and engineered procaspase-3 respectively (a landmark study using split-TEV to activate an engineered procaspase-3 and induce apoptosis by Gray *et al.*[55] provides additional confidence that NASPER can be used to dimerise split-TEV and induce apoptosis). On the other hand, the split-GFP effector was used to visualise whether the two PUF proteins can bind concurrently to a single stretch of RNA in cells and to select cells containing the target RNA using flow cytometry. Thus, the NASPER system is expected to circumvent the delay issue observed in current *hTERT*-driven therapeutic strategies by inducing apoptosis in *hTERT*+ cells before they can undergo any further divisions and acquire mutations that would allow *hTERT*-independent proliferation. The hypothesis is that the system will result in highly specific, sensitive, and rapid killing of *hTERT*+ cancer cell types. Since the NAPSER system kills cancer cells by targeting the overexpression of *hTERT*, I also believe that this system should be applicable to cancers overexpression other genes, and even other diseases which originate due to gene overexpression.

1.6 ALTERNATE POTENTIAL TARGETS OF THE NASPER SYSTEM

In a broader context, the NASPER system could potentially be applied to any condition where overexpression of a gene is the aetiology (since more PUF complexes would be formed in cells where the gene is being overexpressed compared to normal cells with baseline mRNA levels resulting in specific targeting of the overexpressing cells), and rapid cell death is the appropriate means to correct it. It may not be applicable to cases when the disease-causing RNA has other types of mutations, such as point mutations, as the PUF proteins may not be specific enough to target these discriminately. In cancer, gene overexpression is a very common phenomenon that increases the levels of oncogenes resulting in the cancerous phenotype. The most commonly overexpressed genes in some of the most common cancers that could potentially be targeted using the NASPER system are shown in Table 1.1.

Cancer Type	Overexpressed Gene
Breast	*FGFR1*[56], *Myc*[57], *AURKA*[58], *ERBB2*[59,60]
Colorectal	*EGFR*[61], *Myc*[62]
Endometrial	*ERBB2*[59]
Gastric	*Myc*[63], *ERBB2*[64]
Lung	*Myc*[65]
Melanoma	*CDK4*[66]
Ovarian	*ERBB2*[60]
Pancreatic	*SMURF1*[67]
Prostate	*AR*[68]

Table 1.1: Overexpressed genes in various cancer subtypes

ERBB2 is overexpressed in 15-30% of breast cancers, 25-30% of endometrial cancers, 25-30% of ovarian cancers and 10-30% of gastric cancers. *Myc* is another gene that is commonly overexpressed in various cancers such as breast, colorectal, gastric and lung cancers. Thus, targeting the overexpression of these genes may be of particular interest as these can be applied to a wider group of cancer sub-types (similar to the intention using *hTERT* as a target). Additionally, some of these cancers, such as lung and pancreatic cancers, have poor prognoses with current treatment regimens and may benefit from innovative approaches such as the NASPER system. On the other hand, the absolute numbers of people suffering from other cancers such as breast and prostate are extremely high resulting in high number of deaths despite generally good prognoses, and therefore using novel approaches to slightly improve the efficacy of treatment even for these cancers can have an enormous impact.

Although cell death was primarily a therapeutic strategy for cancer, other conditions such as autoimmune diseases, where an inappropriate activation of the immune system results in disease[69], could also benefit from the NASPER system. Hyperactive B-cells overexpressing *BAFF* can produce autoantibodies leading to autoimmune disease[70]; targeting these cells using NASPER could result in their apoptosis, and thus the prevention of autoantibody production.

However, treating other conditions associated with RNA overexpression, such as *SNCA* in Parkinson's disease[71] using an apoptosis-inducing method such as the NASPER system would be inappropriate because killing these cells would have a detrimental effect on physiological functioning as well. Thus, the NASPER system has several potential applications, primarily in cancer, however with the potential for application to other diseases where overexpression is the aetiology using the apoptosis-inducing effector domains or using alternate effector domains that should not induce cell death.

1.7 MULTISPECIFIC THERAPEUTICS

As NASPER goes beyond the traditional antagonist/agonist drug mechanism of blocking/activating a functional site to produce a response, it falls within the fourth wave of drug discovery as described by Deshaies[72]. It is essential to understand this fourth wave to see where the NASPER system fits in this novel area of drug discovery, and to see if other emerging technologies may be applicable to improve or optimise the NASPER design. This fourth generation is defined by small molecules or biologics that engage two or more partners in the cell giving them their name – multispecific drugs. NASPER broadly falls into this category of multispecific drugs, as it targets RNA at two positions and subsequently activates a third molecule to bring about the desired effect. Currently, there are three classes of multispecific drugs defined as follows: 1) sequentially binding obligate multispecific drugs that mediate localization (SOMs), 2) concurrently binding obligate multispecific drugs that mediate localization (COMLs), and 3) concurrently binding obligate multispecific drugs that function as matchmakers (COMMs).

SOMs bind to a 'dock' and a 'target' in two different cellular compartments to bring about their effects. A classic example of this mechanism is the GalNAc-siRNA drug conjugate[73]. GalNAc binds solely to the liver specific asialoglycoprotein receptor; GalNAc-siRNA binding to this receptor (the 'dock') results in endocytosis of the complex and release of the siRNA which goes on to silence the 'target' RNA. The first SOM to be approved for clinical use was the antibody-drug conjugate gemtuzumab-ozogamicin for acute myeloid leukaemia. Gentuzumab is the antibody which binds the 'dock' CD33 receptor on leukaemic blast cells subsequently releasing ozogamicin into the cell to cleave the 'target' nuclear DNA.

COMLs are variations of SOMs that must bind the 'dock' and the 'target' simultaneously for an effect to be produced. This drug mechanism is of great interest for the delivery of pro-inflammatory payloads such as IL-2 and TNFs[74]. The pro-inflammatory molecules are modified to reduce affinity for their cognate receptors and tethered to an antibody specific for a cancer cell surface antigen. When the antibody binds the receptor, the pro-inflammatory mediator can saturate receptors in proximity resulting in a localized effect. This is where the NASPER system fits best, as the PUF fusion proteins, the overexpressed RNA (the 'dock') and the 'target' (cellular proteins for the HIVPR domains and engineered procaspase-3 for the split-TEV domains) must all be present in the same compartment to activate by proximity. The key difference is that currently, COMLs have only been designed to work extracellularly, whereas the NASPER system must enter the cell which is a disadvantage, however NASPER is applicable to a broader range of cancers simultaneously depending on the RNA target (eg: *hTERT* is overexpressed in 90% of cancers) compared to targeting antigens which vary greatly between cancers[75], and should result in apoptosis of the cells.

COMMs are also known as matchmaker drugs, as they bring a target and effector into proximity such that the effector acts on the target to bring about an effect. Thus, an endogenous function is manipulated to create a therapeutic effect. This induction of proximity can be carried out by small molecules or biologics. Cyclosporin, the first small molecule COMM, induces an interaction between calcineurin and a proline isomerase which causes the recruitment of cyclophilin to the complex impeding the access of natural calcineurin substrates[76]. Proteolysis-targeting chimaeras (PROTACs) are hetero-bifunctional molecules that bring disease-causing proteins into proximity of an E3 ubiquitin ligase resulting in ubiquitination and proteosomal degradation of the protein[77]. Novel biologic versions of PROTACs being tested in the Itzhaki lab have advantages over traditional small molecule PROTACs. These hetero-bifunctional protein degraders are functionalized using 1) a degron sequence to bind an E3 ligase, and 2) a peptide sequence to bind a target protein[78]. Concurrent binding brings these molecules into proximity and exerts the same function as a PROTAC. However, the key difference is that peptide binding sequences are used rather than small molecules which opens up access to an array of targets that are 'undruggable' by small molecules due to the lack of a discrete binding pocket or groove. Thus, there is great excitement in the field of multispecific therapeutics, as it increases the potential of traditional

pharmacology exponentially by increasing the number of potential targets (such as RNA itself as in the NASPER system) and increasing the specificity of treatments (as in NASPER where multiple molecules coming together and activating is much less likely in healthy cells).

1.8 APOPTOSIS-INDUCING THERAPEUTICS

NASPER also falls into the category of drugs designed to induce apoptosis of cancer cells. This avenue of research has primarily been focused on activating the extrinsic pathway of apoptosis by targeting the TRAIL receptor as it does not result in apoptosis of normal cells[79]. However, targeting TRAIL is the first point of the apoptosis cascade and complex pathways affected by cancer result in inefficient induction of apoptosis in tumors[80]. Furthermore, although activation of the TRAIL receptor does not cause apoptosis in normal cells, systemic delivery of TRAIL ligands such as TNFs have been shown to be hepatotoxic[81] and the excitement in targeting this receptor has generally died down. Despite this, some efforts are still being made to improve the design of TRAIL ligands to reduce cytotoxicity and improve efficacy resulting in second- and third-generation therapies, although none have yet been approved for clinical use. As mentioned previously, there are also novel COMLs that are being tested to prevent systemic effects of TNFs by targeting the ligand to cancer cells using antibodies.

Inhibiting the activity of anti-apoptotic proteins such as Bcl-2 has also been of interest in inducing apoptosis in cancer cells[80]. Venetoclax, a Bcl-2 inhibitor, was the first approved drug for use in chronic lymphoblastic leukaemia[82] and has also been recently approved for use in acute myeloid leukaemia. There are disadvantages of this method such as toxicity and ineffective induction of apoptosis due to the overexpression of other anti-apoptotic proteins, and loss of pro-apoptotic proteins (such as BAX and BAK)[83], resulting in resistance. Targeting of inhibitor of apoptosis proteins (IAPs) has also been attempted by inducing their proteasomal degradation using E3 ligases. Small molecule inhibitor and siRNA therapies have also been tested, however none have progressed to the clinic due to a lack of efficacy, likely due to resistance of cancer cells to apoptosis[80].

The above challenges may seem demotivating. However, all of these methods have targeted upstream mediators in the apoptosis pathways, whereas one version of NASPER is designed

to activate procaspase-3, the final executioner caspase in the apoptotic pathway. Inducible caspase therapies to date have used caspase-6 and caspase-9 which have shown great promise in inducing apoptosis *in vivo*[81]. The most similar though is a novel SOM treatment being pioneered by Amgen which uses caspase-3 linked to a HER2 targeting antibody. This treatment has been shown to induce apoptosis in tumor cells and a xenograft model, although the data from clinical trials have not been published yet. Nevertheless, this shows that there is a huge potential in using procaspase-3 to induce apoptosis which provides a strong foundation to develop the NASPER system.

1.9 RNA-TARGETING THERAPEUTICS

To date, targeting cancer cells based on their RNA has been attempted with little success. Attempts have mainly been focused on using anti-sense oligonucleotides (ASOs), siRNA and miRNAs, all of which result in the degradation of the target mRNA thereby downregulating the gene in question[84]. siRNA and miRNA therapies for cancer are in their infancy; siRNAs have been used to target overexpressed genes such as *PLK1*, *KRAS(G12D)*, and *myc*[84,85]. miRNA mimics have been used to target key pathways in cancer cell proliferation; miR-16 and miR-34 both act as tumour suppressors by targeting oncogenic mRNAs involved in cell cycle regulation and cell proliferation[86,87]. Although early-stage trials have shown good tolerance in patients, late-stage data are yet to be released. Compared to siRNAs and miRNAs, several ASOs have been taken forward to clinical trials for the treatment of a wide range of cancers, however none of these have successfully made it to the clinic.

The main issues faced by these therapeutics are inefficient delivery and low activity. Drug delivery would be a barrier to efficacy of the NASPER system as well as the proteins required for the system would have to be provided as a gene or protein therapy, however there is a key difference compared to the current field of RNA targeting. NASPER does not directly affect the RNA target in any way. It will simply use the RNA as a scaffold to re-constitute its effector domains to bring about cell death. Regulating a single RNA at the transcript level may not be enough to produce a significant anti-cancer effect; however, the intended mechanism of NASPER should overcome this problem.

1.10 THE PUF PROTEINS

The Pumilio/Puf (hereafter referred to as PUF) proteins are a ubiquitous family of eukaryotic RNA-binding proteins that were initially discovered in *D. melanogaster* and *C. elegans*[88,89] (Figure 1.7). The PUF domain typically consists of eight tandem PUF repeats (and two pseudo repeats at the N- and C-terminus) comprising thirty-six amino acids of which three amino acids, known as the tripartite recognition motif, are responsible for binding to a particular RNA nucleotide[88-90] (Figure 1.8). Each PUF repeat can recognise one RNA nucleotide, and thus, most natural PUF domains have eight repeats that bind to an 8-mer RNA sequence. However, not all PUF domains contain the canonical eight repeat structure. In fact, the PUF family can be broadly divided into three subfamilies: 1) Classical PUF proteins as described here, 2) PUF-A/PUF-6 and 3) Nop-9[88,89]. There is less than 30% sequence homology between these families, however the classic PUF proteins are very well conserved with the main variations being in the tripartite recognition motifs of each PUF repeat.

The PUF-A/PUF-6 sub-family has no apparent sequence specificity and binds to both DNA and RNA using the phosphate backbone. Compared to the classical PUF proteins, they contain eleven repeats and are localized to the nucleus rather than the cytoplasm[91]. The Nop-9 sub-family also contains eleven repeats and is localized to the nucleus, however it specifically recognises RNA and can target both RNA sequences and structures such as stem-loops[92]. Both these sub-families also differ from the classic PUF proteins based on their shape. The classical PUF proteins are crescent shaped, whereas the other sub-families tend to be L-shaped which may influence their binding to nucleic acids[88-92].

Although the classical PUF proteins bind their RNA in a one-to-one manner with one repeat binding one RNA nucleotide, this is not always the case. Yeast PUF4p can accommodate additional bases into the target sequence by stacking or flipping away of bases due to the curvature of the protein. Additionally, PUF1 and PUF2 from novel fungal sub-family seem to only contain six repeats and have been shown in crystal structures to bind four nucleotide sequences and interestingly, two PUF1 fungal proteins were shown to bind to a single RNA containing two of the consensus binding motifs[93]. Thus, despite being evolutionarily conserved, PUF proteins and families may have quite divergent behaviours.

Classical PUF proteins are primarily involved in post-transcriptional regulation of mRNA stability and translation by various mechanisms involving binding to the 3' untranslatable region (UTR)[94]. PUF proteins have been shown to 1) de-adenylate target RNA resulting in repression, 2) interfere with co-factors binding to the mRNA cap resulting in reduced translation, and 3) influence mRNA stability through interactions with Argonaute proteins[94]. Classical PUF proteins are involved in stem cell regulation in nearly all organisms that contain them. Indeed, the mRNA targets of PUF proteins show over-representation of GO terms such as cell cycle, cell division and nuclear division[94]. In *D. melanogaster*, mutation of PUF proteins results in a loss of female germline stem cells[95]. Similarly, in *C. elegans*, silencing of PUF proteins results in a loss of germline stem cell maintenance and induces meiosis and spermatogenesis of these cells[96]. In mammals, PUF proteins control essential activities such as neurogenesis and expansion of haematopoetic stem cells[97,98]. In contrast, the non-classical PUF proteins PUF-6 and Nop-9 are both involved in ribosome biogenesis through their influence on rRNA processing[91,92].

Due to their modular nature, classical PUF proteins have been of great interest in producing designer RNA binding domains. Wang *et al.*[99] discovered the tripartite recognition motif code to design PUF repeats to target individual bases, however they were missing the code required to target cytosine. Two groups then developed a tripartite recognition motif for cytosine[100,101] opening up the field to custom engineer classical PUF proteins to bind any eight nucleotide RNA sequence of choice. Abil *et al.*[90] used this code to develop a high-throughput method to generate these PUF proteins using golden gate cloning. Furthermore, Zhao *et al.*[102] and Adamala *et al.*[103] expanded on the canonical eight repeat nature of classical protein by concatenating up to sixteen repeats in order to target sixteen nucleotide sequences on RNA. Although the engineered PUF proteins were found to have extremely high binding affinities[90,102,103], their non-cognate binding was not rigorously tested despite classical PUF proteins having been shown to bind non-cognate sequences through various mechanisms[104]. Indeed, *Zhao et al.*[102] have also suggested that longer PUF proteins may display promiscuity because not all modules may need to be bound to result in stable PUF-RNA complex formation. However, the degree of promiscuity of each PUF module has not been elucidated and thus there is no penalty-based system for predicting or scoring the non-specificity of

engineered PUF proteins. Additionally, no kinetic data on association or dissociation of custom PUF proteins has been collected.

PUF protein binding to RNA has been shown to interfere with the binding of certain proteins, such as eIF4E[105], however this is to repress translation of mRNAs by canonical PUF proteins and therefore may be protein specific. In fact, other studies have shown that PUF domains do not interfere with the binding of certain proteins such as pAbp[106] to mRNA. This further suggests that the influence of PUF proteins on the binding of other proteins to mRNA may be specific and related to the function of the PUF protein itself. Such studies on engineered PUF domains have not been carried out to date.

Canonical PUF protein binding has been shown to be negatively affected by RNA secondary structure formation *in vitro*, however this effect did not translate *in vivo* for most RNA secondary structures; those with exceptionally high stability ($\Delta\Delta G_{fold} > 8.6$ kcal/mol) showed slightly reduced PUF binding *in vivo*[107]. The difference between *in vitro* and *in vivo* characteristics may be due to the presence of different PUF co-factors *in vivo* or the ability of PUF domains to modulate RNA secondary structure *in vivo*, improving accessibility of hidden binding sites. Nevertheless, several studies have used these engineered PUF domains successfully in cells[90,102,108,109,110] to image individual RNAs and to repress translation, providing the initial foundation to use these proteins in the NASPER system.

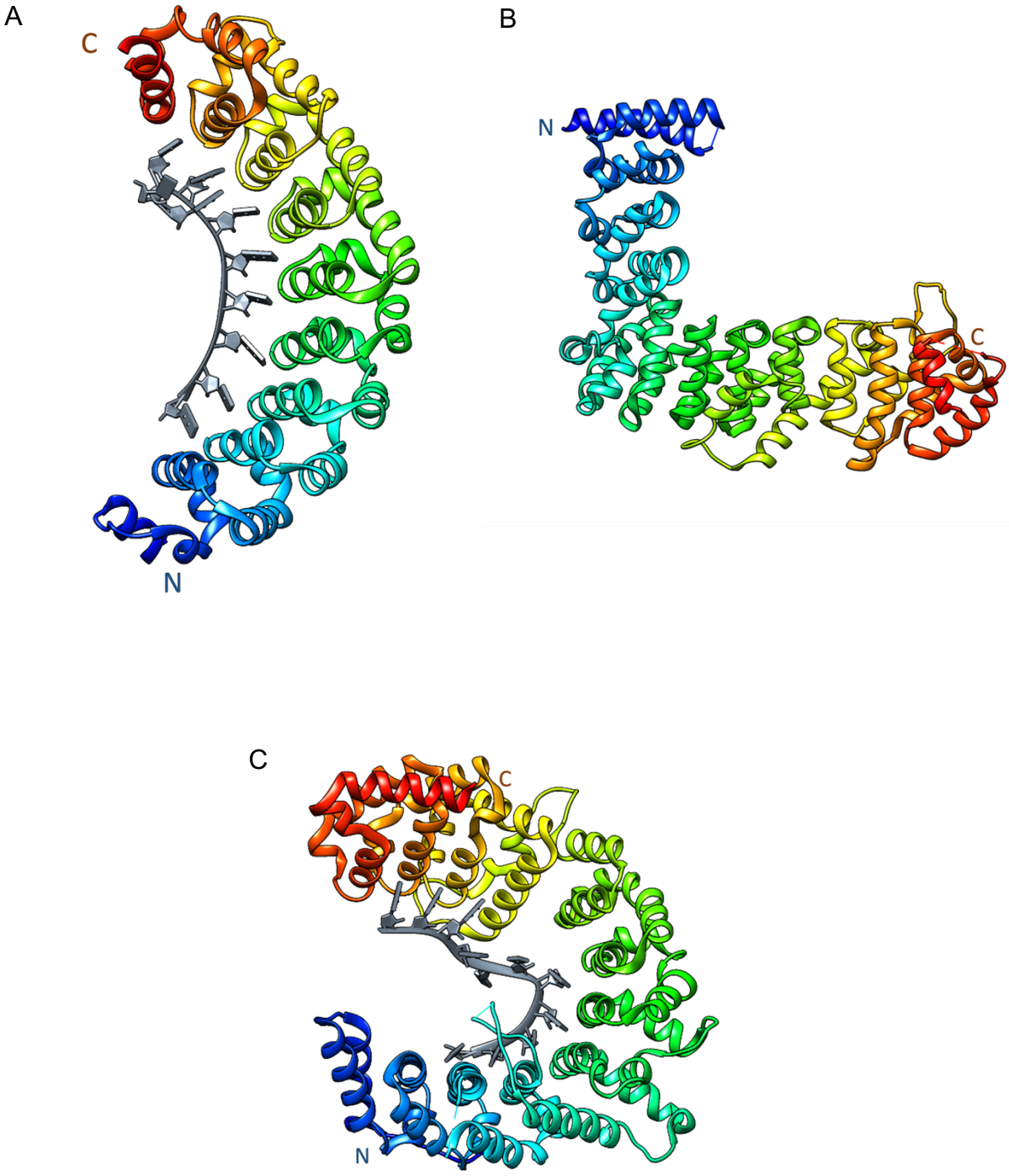


Figure 1.7: **A** – Schematic of the structure of a classical PUF protein bound to RNA (PDB: 1M8Y), **B** – Schematic structure of human PUF-A (PDB: 4WZR), **C** – Schematic structure of Nop9 bound to RNA (PDB: 5WTY). All structures are coloured from blue to red corresponding to the N-terminus to the C-terminus. Nucleic acids are coloured in grey.

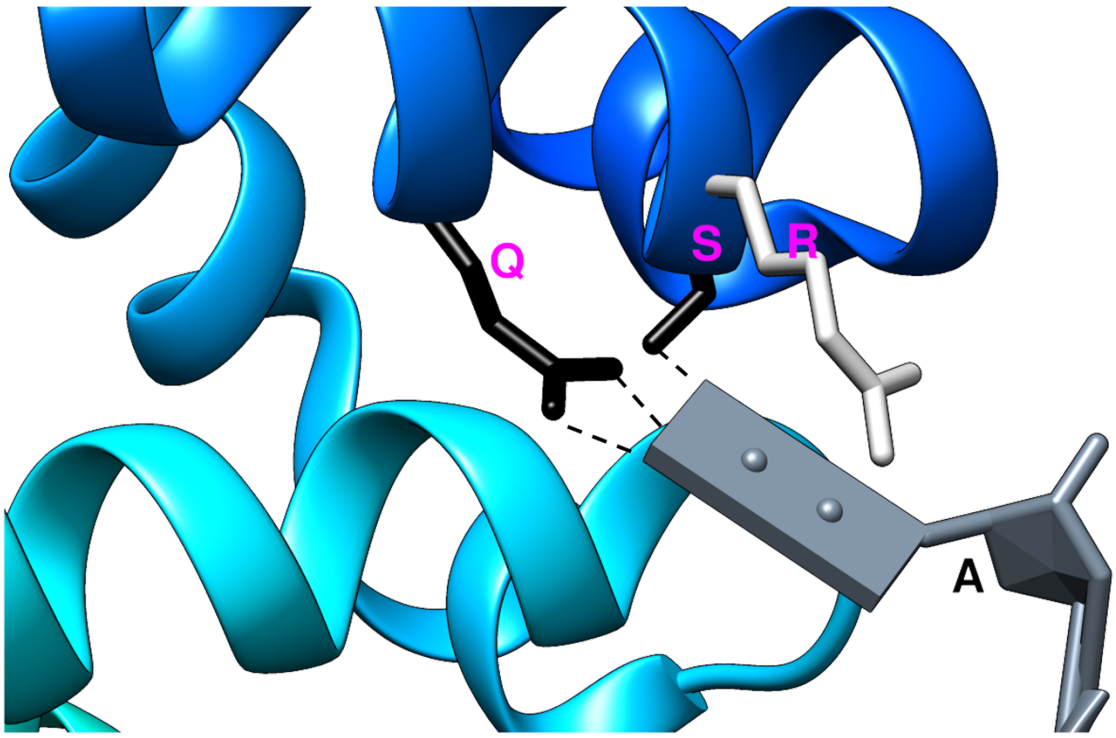


Figure 1.8: Close-up of the structure of the classical PUF protein bound to RNA highlighting the tripartite recognition motif interacting with the adenine nucleotide. The structure is coloured from blue to red corresponding to the N-terminus to the C-terminus. Nucleic acids are coloured in grey. Amino acid letters and RNA bases are shown in magenta and black letters respectively.

1.11 PROJECT AIMS

The aims of this thesis were to demonstrate that the NASPER system can function as hypothesized and ultimately to test whether it can target the cancer overexpressed *hTERT* mRNA. Several steps of optimization were required to get a working model of the NASPER system which was then tested in HeLa cells which are known to overexpress *hTERT*.

Chapter 3 describes efforts to design and test custom-engineered PUF domains *in vitro* using fluorescence polarisation assays. Although custom engineered PUF domains have been used and tested widely using various biophysical techniques[90,102,103], there have not been attempts to test whether two PUF domains can bind a single RNA molecule *in vitro*. My hypothesis was that if there is indeed dual binding occurring, the fluorescence polarisations measured would be significantly higher compared to single and non-binding controls. This would verify that these custom-engineered PUF domains are applicable for use in the NASPER system, as a pre-requisite is the ability to concurrently bind RNA.

Chapter 4 attempts to solubilize and purify the HIV protease – the first effector domain that was tested for use in the NASPER system. *In vitro* work was carried out to ensure that the HIV protease being used was active. To date, most work using the HIV protease *in vitro* has focused on denaturation and re-folding of the protease from inclusion bodies[111-113] and methods to solubly purify the protein could not be replicated in the lab. Thus, a novel His-lipoyl tag was used to consistently generate soluble HIV protease. The purified protease was then used to cleave a synthetic substrate using fluorescence intensity measurements to check its activity.

In Chapter 5, the fusion proteins required for the NASPER system were generated using the designed PUF domains and HIV protease as the effector domain. These fusion proteins were tested using cleavage assays in HEK293T cells by Western blotting. The hypothesis was that when co-expressed with the cognate RNA, the fusion proteins would cleave themselves and auto-activate. Several optimisations were also carried out using site-directed mutagenesis of the generated constructs and many control experiments were also carried out to troubleshoot the system.

Chapter 6 describes experiments to investigate the specificity of 8-, 9- and 10-mer PUF domains in a cellular context using a novel approach. Organic orthogonal phase separation (OOPS)[114,115] was used to selectively purify protein bound RNA from PUF protein expressing cells. The RNA was subsequently used in next-generation sequencing and compared using bioinformatics to control RNA that was also isolated using OOPS from cells lacking the PUF protein. The hypothesis is that the bioinformatic analysis will be able to determine PUF binding regions in the cell by comparing it to the control. Since the 8-mer PUF protein was first validated to bind its cognate sequence in Chapter 3, the OOPS-seq results would provide an initial indicator as to how the different PUF domains interact with cellular RNA. Subsequently, the HIV protease effector domains used in Chapter 5 were substituted for the split-GFP domains in order to assess whether the PUF domains were concurrently binding to RNA using confocal microscopy. If the PUF domains did bind, split-GFP reconstitution was expected to occur to a greater degree when using cognate RNA compared to the scrambled control. Several variables were tested by changing the distance between PUF binding sites, adding multiple sites, changing the orientations of the effector domains, and trying longer PUF

domains. Select constructs from the confocal microscopy were used in plate-reader assays and flow cytometry to establish whether cell populations could be differentiated based on the split-GFP reconstitution, as this would provide a useful method to distinguish live cell populations based on RNA content.

In Chapter 7, the split-GFP domains were replaced with the split-TEV domains, which were optimized using the same conditions as for the split-GFP domains. The split-TEV effector domain activity is cleavage, which was expected to produce higher signals due to amplification (one re-constituted split-TEV would cleave multiple substrates, whereas one re-constituted split-GFP would not amplify the fluorescence signal produced). Furthermore, split-TEV was essential to the original aim of the NASPER system, which was inducing cell death, since split-TEV domains have been used to cleave a custom TEV-cleavable procaspase-3 construct to induce cell death[55]. The hypothesis was identical to that for Chapter 6 in that the system would have higher cleavage when using cognate RNAs compared to the scrambled RNA control. This system was further optimized using novel mutations of the split-TEV domains which were intended to reduce non-specific activity and ultimately, the optimized lead candidates were used to design PUF domains to target the *hTERT* mRNA. These constructs were then tested in HeLa cells which overexpress this mRNA endogenously using Western blot.

Chapter 8 describes experiments to test the use of lead split-TEV NASPER candidate to activate the procaspase-3 construct and thereby induce apoptosis in cells containing the cognate RNA. Cell viability, caspase-3 activity and mCherry fluorescence were used to monitor the effects of the NASPER system in the HEK293T cells, and the changes in viability and mCherry fluorescence upon procaspase-3 addition was used to compare the different constructs. Troubleshooting of this system was carried out using Western blotting.

CHAPTER 2 – GENERAL METHODS AND MATERIALS

2.1 REAGENTS AND SOLUTIONS

All chemical reagents were purchased from either Sigma-Aldrich, Qiagen, Gibco, New England BioLabs or Thermofisher Scientific unless otherwise stated.

2.2 *E. coli* BACTERIAL STRAINS

Three *E. coli* strains were used throughout the thesis and are listed below:

1. DH5-α – K-12 strain derivative optimised for molecular cloning applications purchased from New England BioLabs.
2. C41 – BL21(DE3) derivative modified for increased transformation efficiency, recombinant protein expression and toxic protein resistance were from lab culture stock propagated in house.
3. Bronze cells – lab stock DH5-α cells propagated in house.

In house propagated strains were prepared from a glycerol stock. They were inoculated in 5 mL of LB medium and incubated at 37°C in a shaking incubator overnight. The overnight culture was diluted 1:100 in fresh LB medium and grown while shaking at 37°C until the OD_{600} reached 0.25 – 0.3. The cell suspension was transferred in to 50 mL tubes and cooled on ice for 10 minutes before centrifugation at 4000 x g for 5 minutes at 4°C. The resulting cell pellets were re-suspended thoroughly in 10 mL cold transformation buffer 1, combined together and incubated on ice for 5 minutes before another round of centrifugation. The cell pellet was subsequently re-suspended in transformation buffer 2 and incubated on ice for 15 minutes. The bacterial aliquots were dispensed in 50 µL aliquots into pre-chilled microfuge tubes and stoed at -80°C. Their competence was checked with heat shock transformation with the appropriate plasmids and antibiotics.

2.3 PLASMIDS AND PRIMERS

All primers and G-blocks were purchased from either Integrated DNA Technologies or Thermofisher Scientific. Primer sequences are listed in the Appendices. Two plasmid vectors were used in this thesis.

1. pRSET B – modified to to contain either an N-terminus 6xHis, GST or His-lipoyl (Hislip) tag purification tag, and a TEV protease cleavage site for bacterial protein expression
2. pcDNA 3.1(-) – modified to contain a Kozak sequence and initiator ATG downstream of the CMV promoter used for mammalian protein expression.

G-block protein sequences used in this thesis are provided below:

1. HIVPR (bacterial)

```
GSPQITLWKRPLVTIKIEGQLKEALLDTGADDTVIEDINLPGKWKPKMIGGIGGFIKVRQYDQIIIEI
AGKKAIGTVLVGPTPINIIGRNLLTQIGATLNF**
```

2. p6-HIVPR (mammalian)

```
QGTVSFNFPQITLWKRPLVTIKIEGQLKEALLDTGADDTVIEDINLPGKWKPKMIGGIGGFIKVRQYD
QIIIEIAGKKAIGTVLVGPTPINIIGRNLLTQIGATLNF**
```

3. Flexible Linker

```
EGKSSGSGSESKST
```

4. 3xFLAG-N-TEV

```
DYKDHDGDYKDHDIDYKDDDDKGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKH
LFRRNNGTLLVQSLHGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTT
NFQT**
```

5. 3xHA-C-TEV

```
YPYDVPDYAGYPYDVPDYAYPYDVPDYAKSMSSMVSDTSCTFPSSDGIFWKHWIQTKDGQCGSPLVST
RDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMV**
```

6. 3xFLAG-N-GFP

DYKDHDGDYKDHDIDYKDDDDKVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKF
ICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKF
EGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQ**

7. 3xHA-C-GFP

YPYDVPDYAGYPYDVPDYAYPYDVPDYAKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPD
NHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK**

8. TEV-cleavable mCherry (TEV site underlined)

DYKDHDGDYKDHDIDYKDDDDKGGSGSENLYFQSGSGGMVSKGEEDNMAIIKEFMRFKVHMEGSVNGH
EFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKW
ERVMNFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEI
KQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDEL
YKYPYDVPDYAGYPYDVPDYAYPYDVPDYANLRQLLPS

9. N-terminus-FLAG TEV-cleavable Procaspase-3 (TEV site underlined)

DYKDHDGDYKDHDIDYKDDDDKMENTENSVDSKSIKNLEPKIIHGSESMDSGISLDNSYKMDYPEMGL
CIIINNKNFHKSTGMTSRSGTDVDAANLRETFRNLKYEVRNKNDLTREEIVELMRDVSKEDHSKRSSF
VCVLLSHGEEGIIFGTNGPVDLKKITNFFRGDRCRSLTGKPKLFIIQACRGTELDCGIETENLYFQSG
VDDDMACHKIPVEADFLYAYSTAPGYYSWRNSKDGSWFMFIQSLCAMLKQYADKLEFMHILTRVNRKV
ATEFESFSFDATFHAKKQIPCIVSMLTKELYFYH**

10. PUF1$_8$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIQLKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIRKFFEFGSLEQKLALAERIRGHVLSLALQMYGCRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVEKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIRRILEHCLPDQTLPILEELHQ
HTEQLVQDQYGSYVIEHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFANNVVQKCVTHASRTERAVLID
EVCTMNDGPHSALYTMMKDQYASYVVRKMIDVAEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKY
YMKNGVDLG**

11. PUF2$_8$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIRLKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIEKFFEFGSLEQKLALAERIRGHVLSLALQMYGNRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVRKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIERILEHCLPDQTLPILEELHQ
HTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFASYVVRKCVTHASRTERAVLID
EVCTMNDGPHSALYTMMKDQYASYVVEKMIDVAEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKY
YMKNGVDLG**

12. PUF1$_9$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIQLKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIRKFFEFGSLEQKLALAERIRGHVLSLALQMYGCRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVEKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIRRILEHCLPDQTLPILEELHQ

ALYTMMKDQYACYVVQKMIDVAEPGQRKIVMHKIRPHTEQLVQDQYGSYVIEHVLEHGRPEDKSKIVA
EIRGNVLVLSQHKFANNVVQKCVTHASRTERAVLIDEVCTMNDGPHSALYTMMKDQYASYVVRKMIDV
AEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKYYMKNGVDLG**

13. $PUF2_{9}$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIRLKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIEKFFEFGSLEQKLALAERIRGHVLSLALQMYGNRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVRKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIERILEHCLPDQTLPILEELHQ
ALYTMMKDQYACYVVQKMIDVAEPGQRKIVMHKIRPHTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVA
EIRGNVLVLSQHKFASYVVRKCVTHASRTERAVLIDEVCTMNDGPHSALYTMMKDQYASYVVEKMIDV
AEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKYYMKNGVDLG**

14. $PUF1_{10}$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIQLKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIRKFFEFGSLEQKLALAERIRGHVLSLALQMYGCRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVEKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIRRILEHCLPDQTLPILEELHQ
NVLVLSQHKFANNVVQKCVTHASRTERAVLIDEVCTMNDGPHSALYTMMKDQYACYVVQKMIDVAEPG
QRKIVMHKIRPHTEQLVQDQYGSYVIEHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFANNVVQKCVTH
ASRTERAVLIDEVCTMNDGPHSALYTMMKDQYASYVVRKMIDVAEPGQRKIVMHKIRPHIATLRKYTY
GKHILAKLEKYYMKNGVDLG**

15. $PUF2_{10}$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIRLKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIEKFFEFGSLEQKLALAERIRGHVLSLALQMYGNRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVRKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIERILEHCLPDQTLPILEELHQ
NVLVLSQHKFANNVVQKCVTHASRTERAVLIDEVCTMNDGPHSALYTMMKDQYACYVVQKMIDVAEPG
QRKIVMHKIRPHTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFASYVVRKCVTH
ASRTERAVLIDEVCTMNDGPHSALYTMMKDQYASYVVEKMIDVAEPGQRKIVMHKIRPHIATLRKYTY
GKHILAKLEKYYMKNGVDLG**

16. Endogenous Cognate $PUF2_{8}$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIRLKLERATPAERQLVFNEILQAAYQLMVD
VFGCYVIQKFFEFGSLEQKLALAERIRGHVLSLALQMYGNRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVEKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIRRILEHCLPDQTLPILEELHQ
HTEQLVQDQYGSYVIEHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFASYVVRKCVTHASRTERAVLID
EVCTMNDGPHSALYTMMKDQYASYVVEKMIDVAEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKY
YMKNGVDLG**

17. Endogenous Cognate $PUF1_{8}$

GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIRLKLERATPAERQLVFNEILQAAYQLMVD
VFGCYVIQKFFEFGSLEQKLALAERIRGHVLSLALQMYGSRVIEKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGNHVVQKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIERILEHCLPDQTLPILEELHQ
HTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFACNVVQKCVTHASRTERAVLID
EVCTMNDGPHSALYTMMKDQYASYVVEKMIDVAEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKY
YMKNGVDLG**

18. Endogenous Scrambled $PUF2_8$

```
GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIELKLERATPAERQLVFNEILQAAYQLMVD
VFGSYVIRKFFEFGSLEQKLALAERIRGHVLSLALQMYGSRVIRKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGCYVVQKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIERILEHCLPDQTLPILEELHQ
HTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFASNVVEKCVTHASRTERAVLID
EVCTMNDGPHSALYTMMKDQYASYVVRKMIDVAEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKY
YMKNGVDLG**
```

19. Endogenous Scrambled $PUF1_8$

```
GRSRLLEDFRNNRYPNLQLREIAGHIMEFSQDQHGSRFIELKLERATPAERQLVFNEILQAAYQLMVD
VFGNYVIQKFFEFGSLEQKLALAERIRGHVLSLALQMYGCRVIQKALEFIPSDQQNEMVRELDGHVLK
CVKDQNGSHVVEKCIECVQPQSLQFIIDAFKGQVFALSTHPYGSRVIRRILEHCLPDQTLPILEELHQ
HTEQLVQDQYGCYVIQHVLEHGRPEDKSKIVAEIRGNVLVLSQHKFASNVVEKCVTHASRTERAVLID
EVCTMNDGPHSALYTMMKDQYANYVVQKMIDVAEPGQRKIVMHKIRPHIATLRKYTYGKHILAKLEKY
YMKNGVDLG**
```

2.4 BUFFERS FOR MOLECULAR BIOLOGY

Several homemade buffers and solutions were used across all Chapters for general molecular biology and are listed below:

1. 50X TAE Buffer (Stock) – 242 g Trizma Base, 57.1 mL glacial acetic acid and 18.6g EDTA were made up to a final volume of 1 L in Milli-Q water. Buffer was stored at room temperature and used at 1X for DNA or RNA agarose gel electrophoresis and running buffer.
2. Ampicillin (Amp) Stock – 50 mg/ml in water. Sterile filtered and used at 1:1000 dilution.
3. Kanamycin (Kan) Stock – 50 mg/ml in water. Sterile filtered and used at 1:1000 dilution.
4. Isopropylthiogalactoside (IPTG) – 1 M in water, sterile filtered
5. Dithiothreitol (DTT) – 1M in water, sterile filtered.
6. Transformation Buffer 1 – 30 mM KOAc pH 5.8, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 3 mM hexamine cobalt chloride, 15% (v/v) glycerol
7. Transformation Buffer 2 – 10 mM MOPS pH 6.5, 10 mM RbCl, 75 mM $CaCl_2$, 15% (v/v) glycerol
8. 2xYT Bacterial Culture Medium – 1.6% (w/v) tryptone, 1.0% (w/v) yeast extract, 0.5% (w/v) NaCl (Formedium)

9. Miniprep Lysis Buffer – 200 mM NaOH, 1% (w/v) SDS
10. Miniprep Neutralisation Buffer – 4.2 M guanidium-HCl, 0.9 M KOAc, pH 4.8 with AcOH
11. Miniprep Wash Buffer – 10 mM Tris-HCl pH 7.5, 80% (v/v) EtOH
12. Western Blot Wash Buffer – 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween-20
13. Western Blot Blocking Buffer - 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween-20, 5% (w/v) skim milk (Marvel)

2.5 CLONING

2.5.1 RESTRICTION ENZYMES

Restriction enzyme-based cloning was only used for bacterial expression vector pRSET B. DNA sequences were purchased as G-blocks with BamHI and HindIII sites added at the 5' and 3' ends, respectively, for restriction enzyme cloning into pRSET B vectors modified to contain either an N-terminus 6xHis, GST or His-lipoyl (Hislip) tag purification tag, and a TEV protease cleavage site. Enzymatic cloning was carried out as follows:

2 μg of pRSET B expression vectors were linearized using FastDigest BamHI and HindIII in FastDigest reaction buffer (Thermofisher Scientific) in a 20 μL reaction volume at 37°C for 1 hour. Subsequently, FastAP (Thermo Fisher Scientific) was added to the reaction and incubated at 37°C for 10 minutes. All enzymes were inactivated together by incubation at 85°C for 10 minutes. The digest was purified using agarose gel electrophoresis and gel extraction. 60 ng of G-block was digested in a 10 μL reaction, without the use of FastAP, in the same manner as the expression vectors, however, the digested fragment was instead purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instruction. Ligation of the digested PUF1 G-block into the digested expression vectors was carried out using a 3:1 ratio of insert:vector, calculated using NEBioCalculator (https://nebiocalculator.neb.com/#!/ligation), using Anza T4 DNA Ligase (Thermo Fisher Scientific) for 15 minutes at room temperature. The ligation mixture was used to transform DH5-α competent cells.

Colonies were picked the next day, and colony PCR was used to check for the presence of insert in selected colonies. Colony PCR was carried out in 10 µL reactions using DreamTaq Polymerase (Thermo Fisher Scientific) at a final concentration of 0.5 U/µL, DreamTaq Green PCR Master Mix, appropriate forward and reverse primers at a final concentration of 10 µM and water. The PCR protocol was as follows:

1. 95°C for 30 seconds
2. 30 cycles of
    a. 95°C for 30 seconds
    b. 55°C for 30 seconds
    c. 72°C for 60 seconds
3. 72°C for 5 minutes
4. Hold indefinitely at 10°C

Colonies positive for insert were grown in LB broth with ampicillin, at a final concentration of 100 µg/ml, at 37°C overnight at 200 rpm shaking. Plasmid DNA was extracted the next day using the QIAprep Spin Miniprep Kit (Qiagen) as per the manufacturer's instructions and DNA was eluted in 50 µL of plasmid elution buffer. Concentration was determined using absorbance at 260 nm on a NanoDrop2000 (Thermo Fisher Scientific). Sanger sequencing was used to verify the sequence of this extracted plasmid DNA (Eurofins).

2.5.2 SITE-DIRECTED MUTAGENESIS

Site-directed mutagenesis was carried out using the QuikChange Site-Directed Mutagenesis protocol. In brief, primers were designed using the QuikChange Primer Design Software (Agilent) and can be found in the Appendices. Subsequently, PCR reactions were set up according to the protocol, using 50 ng of template plasmid per reaction. PCR was run using the following settings:

1. 95°C for 30 seconds
2. 30 cycles of
    a. 95°C for 30 seconds

b. 55°C for 60 seconds

c. 68°C for 600 seconds

3. 68°C for 7 minutes
4. Hold indefinitely at 10°C

Reactions were incubated with 2 µL FastDigest DpnI (ThermoFisher) for 3 hours at 37 °C and subsequently transformed into lab made bronze *E. coli* cells as mentioned previously. Transformed cells were plated on ampicillin-containing LB agar plates and incubated overnight at 37°C.

Colonies were picked the next day for miniprep, and the resulting DNA were sent for Sanger sequencing and sequences were verified prior to use.

2.5.3 NEBUILDER

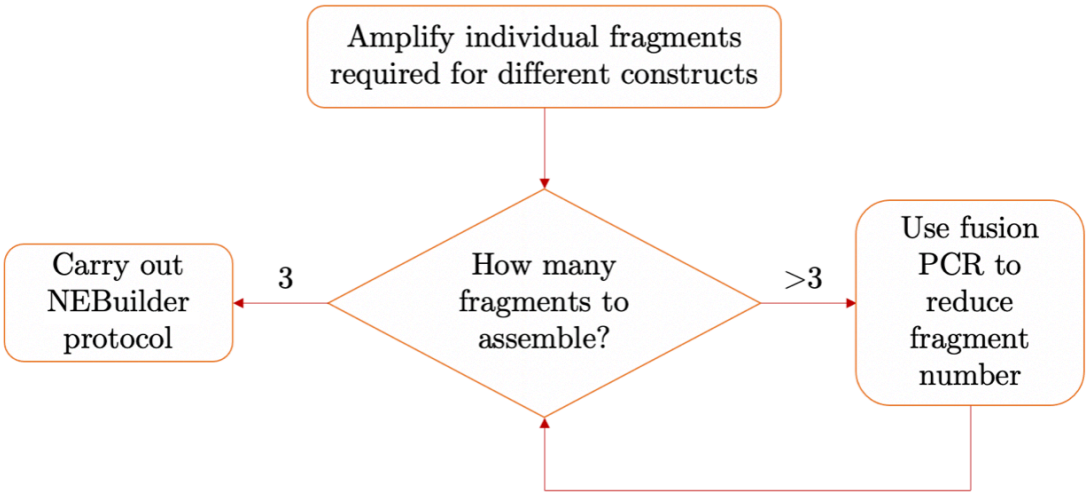


Figure 2.1: Flow chart for NEBuilder assembly used in this thesis.

The NEBuilder HiFi technology was used to assemble the various cloned G-blocks into the required fusion transgenes. Each assembly included the required G-block fragments and insertion vector (pcDNA3.1 containing a Kozak consensus sequence), thereby resulting in 2, 3, 4, 5 and 6 fragment assemblies. The flowchart above shows the different techniques used for different assemblies.

2.5.3.1 PRIMER DESIGN AND FRAGMENT AMPLIFICATION

Primers used to amplify fragments for subsequent NEBuilder assembly must contain overlaps corresponding to adjacent fragments in order to anneal and fuse different fragments. Such primers were designed initially using the NEBuilder Assembly Tool, and then manually to tweak overlap length and sequences. Overlap lengths ranged from 20 – 35 nucleotides.

PCR for fragment amplification was carried out using Q5 polymerase (New England BioLabs) as per the manufacturer's instructions using 1 µL of template in a 25 µL reaction volume. The following cycling conditions were used:

1. 99°C for 30 seconds
2. 30 cycles of
    a. 99°C for 30 seconds
    b. Appropriate T_m for primer pair for 30 seconds
    c. 72°C for 90 seconds
3. 72°C for 5 minutes
4. Hold indefinitely at 10°C

5 µL of 6X DNA loading dye was added to each reaction and the mixture was run on a 1) 1% TAE-agarose gel for fragments larger than 200 bp or 2) 3% TAE-agarose gel for smaller fragments as described. The band corresponding to the fragment was excised and gel purified as mentioned.

2 µg of the vector, Fragment A, was linearised using FastDigest BamHI and HindIII in FastDigest reaction buffer (Thermofisher Scientific) in a 20 µL reaction volume at 37°C for 1 hour. This fragment was then purified the same way as fragments amplified by PCR.

2.5.3.2 FUSION PCR

Fusion PCR was used to combine single fragments when the total number of fragments to assemble exceeded 3. This PCR was carried out in two stages.

STAGE 1

A 25 µL Q5 polymerase PCR reaction was set up according to the manufacturer's instructions, without the primers or template. The fragments requiring fusion were added to the reaction mixture in a 1:1 molar ratio. This mixture was thermocycled using the following 2-step protocol:

1. 99°C for 30 seconds
2. 15 cycles of
    a. 99°C for 30 seconds
    b. 72°C for 120 seconds
3. 72°C for 5 minutes
4. Hold indefinitely at 10°C

STAGE 2

The mixture was removed from the thermocycler after Stage 1, and the appropriate forward and reverse primers, corresponding to the 5' and 3' fragments respectively, were added to the reaction mixture. This mixture was then thermocycled using the following 2-step protocol:

1. 99°C for 30 seconds
2. 30 cycles of
    a. 99°C for 30 seconds
    b. 72°C for 120 seconds
3. 72°C for 5 minutes
4. Hold indefinitely at 10°C

Once both stages were complete, the mixtures were treated, and DNA was gel purified.

2.5.3.3 NEBUILDER PROTOCOL

The NEBuilder assembly reactions were set up as follows:

1. 6 µL 2X NEBuilder HiFi DNA Assembly Master Mix (New England BioLabs)
2. 1 ng Fragment A (vector)
3. Other required fragments in a 1:7 molar ratio
4. Nuclease-free water up to 12 µL

Each reaction mixture was incubated at 50°C for 30 minutes, after which 5 µL of the mixture was used to transform DH5-alpha competent *E. coli* cells (New England BioLabs) by heat-shock as mentioned. Transformed cells were then plated on ampicillin or kanamycin LB-agar plates and incubated overnight at 37°C. Colony PCR was used to verify the presence of the transgene in the vector as described previously and positive colonies were grown, DNA miniprepped and DNA sequenced to ensure correct transgene insertion.

2.6 AGAROSE GEL ELECTROPHORESIS AND EXTRACTION

DNA or RNA were separated using 0.8 – 3% (w/v) agarose (Appleton Woods) gels in 1X TAE buffer with 1x SYBR Safe reagent. Gels were run in 1X TAE buffer at 110 V for 25 minutes and were subsequently imaged by blue light illumination. Bands corresponding to the DNA of required length were excised and extracted from the agarose gel under blue light illumination. DNA in the gel was purified using the QIAquick Gel Extraction Kit (Qiagen) as per the manufacturer's instructions.

2.7 TRANSFORMATION OF BACTERIAL CELLS

Transformation mixture was added to cells at a maximum of 5 µL and incubated on ice for 20 minutes. Cells were then heated at 42°C for 30 seconds and returned to ice for 2 minutes. 200 µL of SOC Outgrowth Medium was added to the cells which were then incubated with shaking

at 37°C for 1 hour. Cells were then plated on ampicillin LB-agar plates and incubated overnight at 37°C.

2.8 DNA MINIPREPS

Bacterial colonies grown on agar plates were picked and incubated overnight at 37°C with shaking in 2xYT or LB medium. Cells were harvested by centrifugation at 4000 rpm for 5 minutes. The resulting cell pellets were re-suspended in P1 miniprep buffer with RNAse A (Qiagen). Pellets were vortexed till homogenous. Subsequently, miniprep lysis buffer was added and the tube was agitated till the solution became clear. Then, miniprep neutralization buffer was added, mixed and the precipitate pelleted by centrifugation at 13,000 rpm for 5 minutes. The supernatant was passed through a DNA spin column (NBS Biologicals) by centrifugation at 7000 rpm for 1 minute. The column was washed twice in miniprep was buffer by centrifugation at 13,000 rpm for 1 minute. Trace amounts of wash buffer were removed by an additional round of centrifugation at 13,000 rpm for 1 minute. The columns were placed in fresh 1.5 mL microfuge tubes and DNA was eluted in 50 µL endotoxin-free elution buffer (Qiagen)

2.9 SMALL-SCALE PROTEIN EXPRESSION

C41 (DE3) competent *E. coli* cells (Lucigen) were transformed with the required plasmid DNA using heat-shock as previously described and plated on ampicillin LB-agar plates at 37°C overnight. A single colony was picked and grown in 15 ml of 2xYT broth, with ampicillin at a final concentration of 100 µg/ml, at 37°C with shaking at 200 rpm till the $OD_{600nm} = 1$. Once this was achieved, cells were induced using 1 mM IPTG and grown for a further 3 hours at 37°C with shaking at 200 rpm. Cells were then harvested by centrifugation at 5,000 x g for 10 min at 4°C and the supernatant discarded.

Harvested cells were lysed by re-suspending with 1 ml BugBuster Mastermix Protein Extraction Reagent (Novagen) and incubated for 20 min at room temperature with shaking. A sample was removed at this stage for analysis (total cell protein) The solution was pelleted at 13,000 x g for 10 minutes at room temperature. An aliquot of the supernatant was removed

at this stage for analysis (soluble protein). Supernatant was discarded and the pellet was re-suspended in 10% Bugbuster and centrifuged at 13,000 x g for 10 minutes at room temperature. This step was repeated once more. Supernatant was discarded and the pellet was re-suspended in 100% Bugbuster. An aliquot of this was used for analysis (insoluble protein).

2.10 EXPRESSION AND PURIFICATION OF TEVPR

C41 (DE3) competent *E. coli* cells (Lucigen) were transformed with His-TEVPR plasmid DNA using heat-shock as previously described and plated on ampicillin LB-agar plates at 37°C overnight. All colonies on the plate were scraped into 700 ml of 2xYT broth, with ampicillin at a final concentration of 100 µg/ml, at 37°C with shaking at 200 rpm till the $OD_{600nm} = 0.6$. Once this was achieved, cells were induced using 1 mM IPTG and grown overnight at 20°C with shaking at 200 rpm. Cells were then pelleted at 35,000 x g for 10 minutes and the supernatant discarded.

Pellets were re-suspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole), at 2 ml/g of cell paste, with 1 mM DTT and SigmaFast protease inhibitor cocktail tablet, EDTA-free (SigmaAldrich), at a 1X final concentration, freshly added. Cells were homogenised using the EmulsiFlex homogeniser under 3 rounds of pressure at 10,000 psi. Homogenised cells were centrifuged at 35000 x g for 35 minutes and the supernatant containing the soluble fraction was collected.

Soluble fraction was applied in upflow to a 5 ml HisTrap Excel column (GE Healthcare) equilibrated with IMAC wash buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole and 1 mM DTT freshly added) at a flow rate of 2 ml/min using an AKTA Pure FPLC system (GE Healthcare). The column was washed in 10 column volumes of IMAC wash buffer and then eluted in 2 ml fractions using IMAC elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole and 1 mM DTT freshly added). 10 µL aliquots of soluble, flow-through, wash (only first and last) and elution fractions (chromatogram selected) were added to 2 µL of 6x SDS loading dye and boiled for 10 minutes. These samples were run on a 15% SDS-PAGE gel in 1x TGS running buffer at 80V for 20 minutes and 180V for 50 minutes. Gels were stained with Coomassie Brilliant Blue stain and imaged under white light. Fractions

containing protein were pooled and dialysed into TEV cleavage buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl and 1 mM DTT freshly added) and flash frozen in 25% glycerol for long term storage at -80°C.

2.11 MAMMALIAN CELL CULTURE, SEEDING, AND TRANSFECTION

HEK293T or HeLa cells were cultured in 12 mL of cell growth medium consisting of Dulbecco's Modified Eagles Medium (DMEM) GlutaMax supplemented with 10% fetal bovine serum (FBS) and 1X penicillin/streptomycin in T75 flasks. Cells were incubated at 37°C, 5% CO_2 and upon reaching confluency they were subcultured at the required ratio. For subculturing, cells were washed once with 1X PBS and detached from the surface using trypsin-EDTA. The trypsin was subsequently inactivated in growth medium and subcultured.

For seeding, cells were washed once in 1X PBS and detached from the surface with trypsin. Trypsin was inactivated with growth medium, and the cells were collected and counted using the Scepter Cell Counter (Millipore) with a 60 µM sensor. Cells were diluted to the required density and seeded in the required vessel. Several seeding vessels were used and are shown in Table 1.

Seeding Vessel	Cell Density (cells/well)	Growth Media Volume (mL)
12-well plate	300,000	1
24-well plate	50,000	0.5
96-well plate	10,000	0.1
iBidi 8-well slide	90,000	0.3
10-cm dish	3,000,000	10

Table 2.1: Seeding vessels, densities and volumes used throughout this thesis.

For transfection, Lipofectamine 2000 was used according to the manufacturer's protocol. In brief, the requires mass of DNA was added to the appropriate volume of reduced serum Opti-MEM medium, and separately, the required volume of lipofectamine 2000, at a 1:3 DNA:lipofectamine, was added to the appropriate volume of Opti-MEM medium. These were incubated for 5 minutes at room temperature and subsequently mixed at a 1:1 ratio and

incubated for a further 15 minutes at room temperature. The required volume of the mixture was then dispensed dropwise to the cells. Various DNA masses were used depending on the nature of the transfection and these are listed below:

Seeding Vessel	Number of Plasmids	DNA Mass per well (ng)
12-well plate	1	500
12-well plate	3	1000
24-well plate	1	250
24-well plate	3	500
24-well plate	4	685
96-well plate	4	115
iBidi 8-well slide	3	300
10-cm dish	1	20,000

Table 2.2: Seeding vessels, number of plasmids and DNA mass used for transfections.

Transfected cells were incubated for either 24 or 48 hours as required and subsequently analysed as needed.

2.12 SDS-PAGE AND WESTERN BLOT

Cells were harvested on ice for 15 minutes using RIPA buffer supplemented with EDTA-free cOmplete protease inhibitor pill (SigmaAldrich), 15 µL BaseMuncher endonuclease (abcam) and when required, darunavir (Sigma) at a final concentration of 100 µM. Harvested cells were then further lysed using SDS-PAGE loading buffer and boiled at 99°C for 20 minutes.

Once boiled, the samples were allowed to cool to room temperature and were loaded on a 12% SDS-PAGE gel along with the Spectra Multicolor Broad Range Protein Ladder (ThermoFisher). Samples were run for 10 minutes at 80V and then at 45 minutes for 180V in 1X TGS running buffer.

For solely SDS-PAGE gels, the gels were stained with Coomassie brilliant blue G-250 stain and imaged in white light using the LiCor Odyssey Fc System (Licor).

For Western blots, gels were then transferred onto nitrocellulose membrane, which were activated in methanol for 1 minute and then washed in 1X transfer buffer, for 35 mins at 15V using a Pierce Power Blot Cassette (ThermoFisher).

Blots were blocked in Western blot blocking buffer and subsequently probed with the appropriate primary antibody at 1:1000 dilution at room temperature for 2 hours with rotation. Blots were then washed 3x for 5 mins each in and subsequently incubated with the appropriate HRP-conjugated secondary antibody at a 1:1000 dilution for 1 hour at room temperature with rotation. Blots were washed 3x once again and developed using either Amersham ECL or Amersham ECL Select (Cytiva Life Sciences) for 1 minute at room temperature and imaged with a 2-minute exposure time for chemiluminescent detection, or 30 seconds for detection of ladders at 700 nm using the LiCor Odyssey Fc System (Licor). Bands were quantified using the ImageJ. Blots were also probed for housekeeping genes as a loading control. The same probing method as mentioned previously was carried out using primary antibodies against hsp60 or GAPDH.

2.13 FLUORESCENCE MICROSCOPY

Cells were analysed when needed for mCherry expression using the FLoid Cell Imaging Station (ThermoFisher). Images were obtained at 20x objective magnification using the excitation wavelength of 586 nm and an emission wavelength of 646 nm.

CHAPTER 3 – CELL-FREE STUDY OF PUF PROTEINS

3.1 INTRODUCTION

As mentioned previously, the PUF proteins are modular RNA-binding proteins wherein each module binds a single base of RNA in a highly specific manner. However, there are different methods of designing these proteins. Abil *et al.*[90] use the native PUF architecture where there are eight PUF repeats which specifically bind to an eight-nucleotide sequence in a one-to-one manner. The amino acids at the 12th and 16th positions of each repeat determines specificity of the repeat for a particular RNA nucleotide. This yields a total of four potential modules per position and thus thirty-two modules in total (for eight repeats).

Adamala *et al.*[103] on the other hand have used the architecture of native PUF proteins to develop four Pumilio-based assembly ('Pumby') modules. Instead of having distinct modules with unique sequences at each of the 8 positions, they have created a one-module-per-base system where the four resulting modules can be concatenated to target the RNA required.

Both these methods appear equally valid for generating sequence specific PUF proteins; however, the method by Abil *et al.*[90] has been used more widely[90.102,108,109,110] and has been shown to work under many different conditions *in vitro* and *in cellulo*. Hence, this method was selected for further here. Initially, two PUF proteins were designed to target two adjacent sites on the *hTERT* intron RNA as would be required in the final NASPER system. Thus, the aim of this Chapter was to assess whether engineered PUF proteins do indeed function as expected from the literature.

Several methods exist for detecting protein-RNA interactions *in vitro*, but the two most commonly used with the PUF proteins are the electrophoretic mobility shift assay (EMSA) and the fluorescence polarisation assay (FP). Typically, EMSA assays require radio-isotopic labelling of the probe RNA such that the protein-bound RNA is shifted to a higher molecular weight relative to unbound RNA on a polyacrylamide gel[116]. Due to the challenges associated with the use of radio-isotopes, I opted for an FP assay, which has been used previously to

assess PUF-RNA binding *in vitro*[90,103] and is routinely used in the Itzhaki lab to monitor protein-protein interactions. The principle of the FP assay is shown schematically in Figure 3.1.

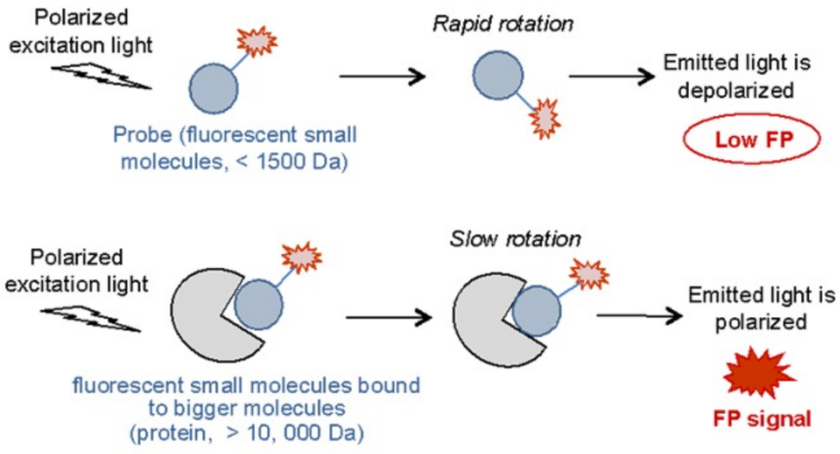


Figure 3.1: Mechanism of binding measurement using the FP assay.

When unbound, a fluorescently labelled RNA molecule will have a low FP in solution due to rapid tumbling; upon binding of the PUF protein, the FP signal increases in a concentration-dependent manner allowing the analysis of binding to be carried out[117]. This assay would not only allow me to test the binding of each individual PUF, but it would also allow me to monitor the binding of a single RNA molecule containing two cognate sites to their two respective PUF proteins. When two PUF proteins are bound simultaneously to the labelled RNA, the FP signal will be significantly higher than when only one PUF protein is bound. Thus, I could get an initial idea on the feasibility of the NASPER system design by checking whether two PUF proteins can bind simultaneously to a single stretch of RNA.

Indeed, Qiu *et al.*[93] showed that the concurrent binding of two PUF proteins to a tandem repeat of cognate sites is possible using X-ray crystallography (Figure 3.2). Figure 3.2 shows a PUF protein that can recognise the UAAU motif, and two molecules of this PUF protein can bind onto a single RNA containing two UAAU motifs separated by a four-nucleotide spacer. However, this study uses native PUF1 from *S. pombe* which recognises four base pair sequences rather than the canonical PUF proteins that recognise eight base pair sequences[118]. Additionally, in this case two molecules of the same PUF protein are binding the single RNA, whereas in my NASPER system, one molecule of each of two different PUF proteins would

need to bind to a single RNA. However, despite these differences, the study by Qiu *et al.*[93] provides confidence that the designed PUF proteins should be able to bind simultaneously to the RNA.

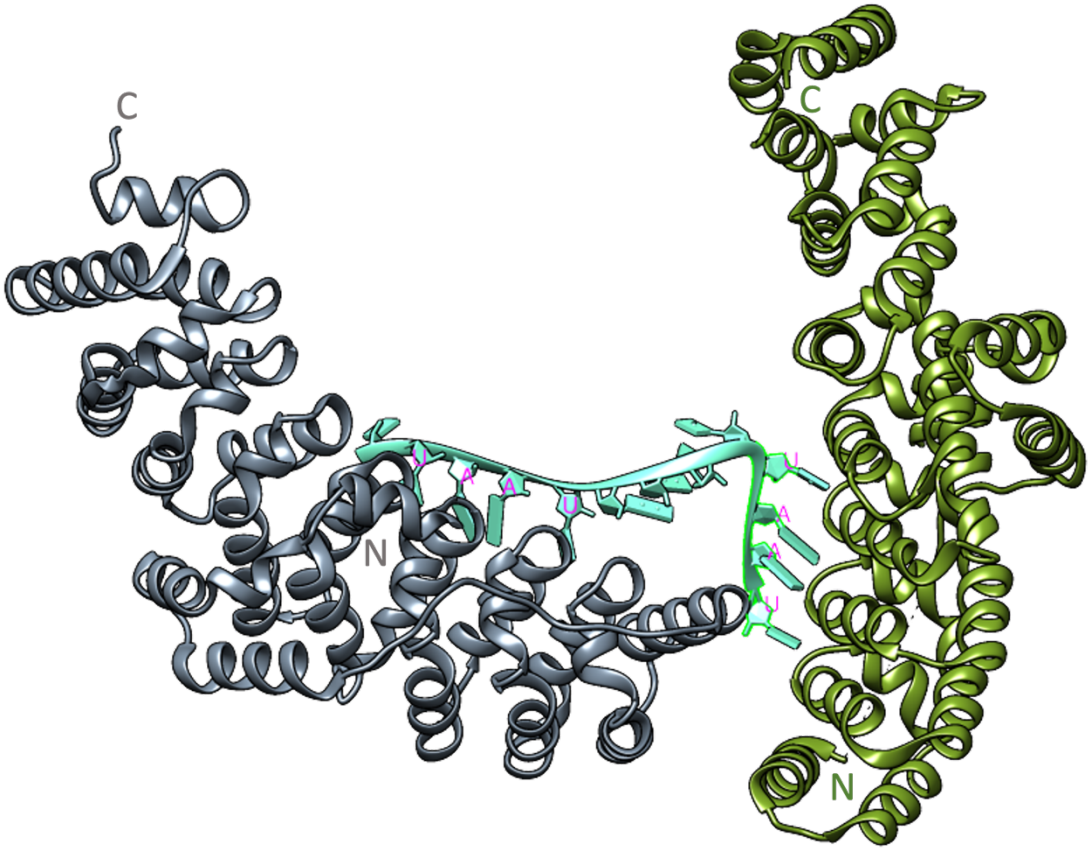


Figure 3.2: Crystal structure of two molecules of *S. pombe* PUF1 bound to one RNA molecule (PDB: 6NY5). RNA molecule is shown in cyan with interacting nucleotides highlighted in magenta.

3.2 METHODS AND MATERIALS

3.2.1 LARGE-SCALE EXPRESSION AND PURIFICATION

C41 (DE3) competent *E. coli* cells were transformed with GST-PUF1 plasmid DNA using heat-shock as previously described and plated on ampicillin LB-agar plates at 37°C overnight. All colonies on the plate were scraped into 700 ml of 2xYT broth, with ampicillin at a final concentration of 100 µg/ml, at 37°C with shaking at 200 rpm till the $OD_{600nm} = 1$. Once this was achieved, cells were induced using 1 mM IPTG and grown for a further 3 hours at 37°C with shaking at 200 rpm. Cells were then pelleted at 5,000 x g for 10 minutes and the supernatant discarded.

Cells were re-suspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl), at 2 ml/g of cell paste, with 1 mM DTT and SigmaFast protease inhibitor cocktail tablet, EDTA-free (SigmaAldrich), at a 1X final concentration, freshly added. Cells were homogenised using the EmulsiFlex homogeniser (Avestin) at 10,000 psi for 5 rounds. Homogenised cells were centrifuged at 35,000 x g for 35 minutes and the supernatant containing the soluble fraction was collected.

3.2.1.1 OPTIMISATION OF PURIFICATION

1. Batch Purification on GST resin: 4 ml of Amintra Glutathione Affinity Resin (Expedeon) was equilibrated in 46 ml of GST wash buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 1 mM DTT freshly added). Resin was centrifuged at 1000 x g for 5 min and the supernatant discarded. The soluble fraction was applied to the resin and left to rotate at 4°C for 1 h. The mixture was centrifuged as before and the supernatant was collected as the 'unbound' fraction. The resin was then washed using the GST wash buffer and centrifuged as before repeatedly till the A_{280nm} of the supernatant was less than 0.1. 2 ml of the wash buffer was then added to the resin to which TEV protease was added at a final concentration of 2 µM. The resin was then left to rotate at 4°C overnight. Cleaved protein was then eluted in 2 ml fractions using the GST wash buffer until the A_{280nm} of the fraction was less than 0.1. 10 µL aliquots of the unbound, wash (first and last only) and elution fractions were added to 2 µL of 6x SDS loading

dye and boiled for 10 minutes. These samples were run on a 15% SDS-PAGE gel in 1x TGS running buffer at 80V for 20 minutes and 180V for 50 minutes. Gels were stained with Coomassie Brilliant Blue stain and imaged using white light. Fractions containing protein were pooled, concentrated using a 10 kDa MWCO concentrator (VWR) and flash frozen in 10% glycerol for long term storage at -80°C.

2. Purification on GST resin with nuclease treatment: As the PUF proteins contained a significant amount of RNA/DNA contamination the following modification was made: A modified lysis buffer including 1 mM EDTA and 5mg DNase I (Sigma Aldrich) was used to aid lysis of the cells used in this method. To the soluble fraction obtained, RNase Cocktail Enzyme Mix (Thermo Fisher Scientific) was added at a 200x dilution and left to rotate at room temperature for 2 hours or at 4°C overnight. Both were then applied to the equilibrated GST resin and purified as described in 1.

3.2.2 OPTIMISATION OF FLUORESCENCE POLARISATION ASSAY

1. **Using FP Buffer[90,102]:** RNA oligonucleotides with 6-FAM attached to the 5' end (sequences provided in Appendix A, Integrated DNA Technologies) were diluted to the required concentration (3x final concentration, i.e 3 nM) in FP buffer (20 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 50 mM KCl, 0.01% Tween-20 and 0.1 mg/ml BSA freshly added). 20 µL of the 3x RNA was added to 20 wells of an OptiPlate 384-well Black Microplate (Perkin Elmer). To the first well, protein was also added at the required concentration (as determined by a Bradford assay) and the final reaction volume was made up to 60 µL using FP buffer (in the first well). A 2/3rd dilution series of PUF1 across 21 wells was made. Control wells included only RNA and only protein. All experiments were conducted in triplicate. The plate was sealed and incubated at room temperature for 30 minutes and measured using a Clariostar Microplate Reader (BMG). Gain and focal height were automatically adjusted prior to each run. An excitation wavelength of 495 nm and emission wavelength of 520 nm were used. Data were fitted on GraphPad Prism 5 (GraphPad) using a one-site binding equation provided by GraphPad: $\boldsymbol{Y = \frac{B_{max} \times X}{K_d + X}}$ (where Y is

fluorescence polarisation, X is concentration of free ligand, B_{max} is maximum binding, and K_d is the dissociation constant.

2. **Using PBS**: Instead of using FP buffer, PBS either at pH 6.0 or 7.4 was used containing 0.5 mM EDTA, 0.01% Tween-20 and 0.1 mg/ml BSA was freshly added.

3. **Using PB:** Instead of using FP buffer, PB (20 mM sodium phosphate pH 6.0, 1 mM EDTA, 0.01% Tween-20, 1 mM DTT and 0.1 mg/ml BSA both freshly added) was used with either 0.8 mM or 900 mM NaCl (HIVPR assay buffer).

3.2.3 RNA TARGETS USED

RNA targets were designed according to initial plans to target a region in the *hTERT* intron RNA. These RNA targets are shown below, and alignment comparisons are made between cognate and scrambled RNA (to display the shuffling of the PUF1 cognate RNA site to produce the corresponding scrambled sequence) and between L0_YY and L0_YN (to display the shuffling of the PUF2 cognate RNA site to produce the corresponding scrambled sequence)

Key:

Bold – PUF1 cognate sequence
Italics – PUF2 cognate sequence
Underlined – scrambled sequences
CAPITALS – END SPACERS

1. Cognate RNA: **AAUcugcgaca**--------CUU
2. Scrambled RNA: AAUcaaucgag--------CUU

3. L0_YY: **AAUcugcgaca***gcugcugc*CUU
4. L0_YN: **AAUcugcgaca**cuccgggu CUU

5. L0_NY: AAUcaaucgag*gcugcugc*CUU

6. L0_NN: AAUcaaucgagcuccgggu CUU

3.3 RESULTS

3.3.1 GST TAG PARTIALLY SOLUBILISES PUF1 CONSTRUCT

In order to express PUF1 in the soluble fraction in *E. coli*, three different purification tags – 6xHis (hereafter referred to as His), GST and Hislip – were tested. These constructs were expressed and purified on a small scale in order to ascertain which tag produced the most soluble protein. The results of this expression test are shown in Figure 3.3.

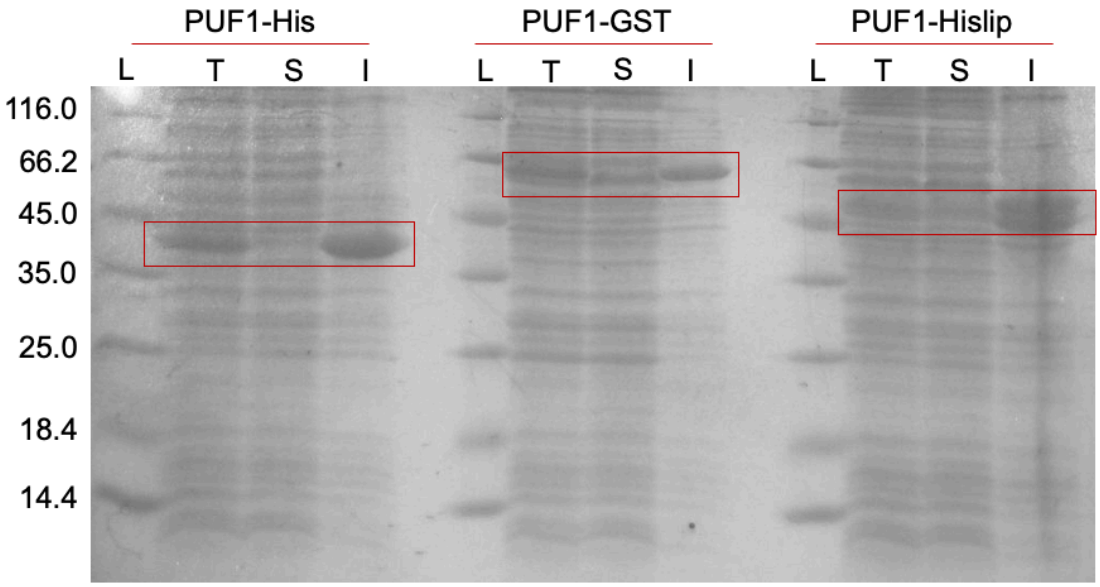


Figure 3.3: 15% SDS-PAGE gel, showing expression of the three PUF1 constructs in the total, soluble and insoluble fractions after cell lysis. The bands corresponding to the construct are boxed. The unstained protein ladder (Thermo Fisher Scientific) is in kDa. Abbreviations: L – Ladder, T – total protein fraction, S – soluble protein fraction, I – insoluble protein fraction.

Neither His nor Hislip tags solubilised the PUF1 construct as effectively as the GST tag, as there was no clear overexpressed band in the soluble fractions when these tags were used, whereas the GST tag did produce approximately 50% of PUF1 in the soluble fraction. All constructs were at approximately the calculated molecular weights (molecular weight of PUF1 in addition to that of the purification tag). Therefore, PUF1-GST was chosen for large-scale purification.

3.3.2 NUCLEASE-FREE PURIFICATION OF PUF1-GST

Large-scale expression and purification of PUF1-GST was carried out using GST affinity resin without prior nuclease treatment, and this resulted in the production of pure, cleaved PUF1 protein (Figure 3.4).

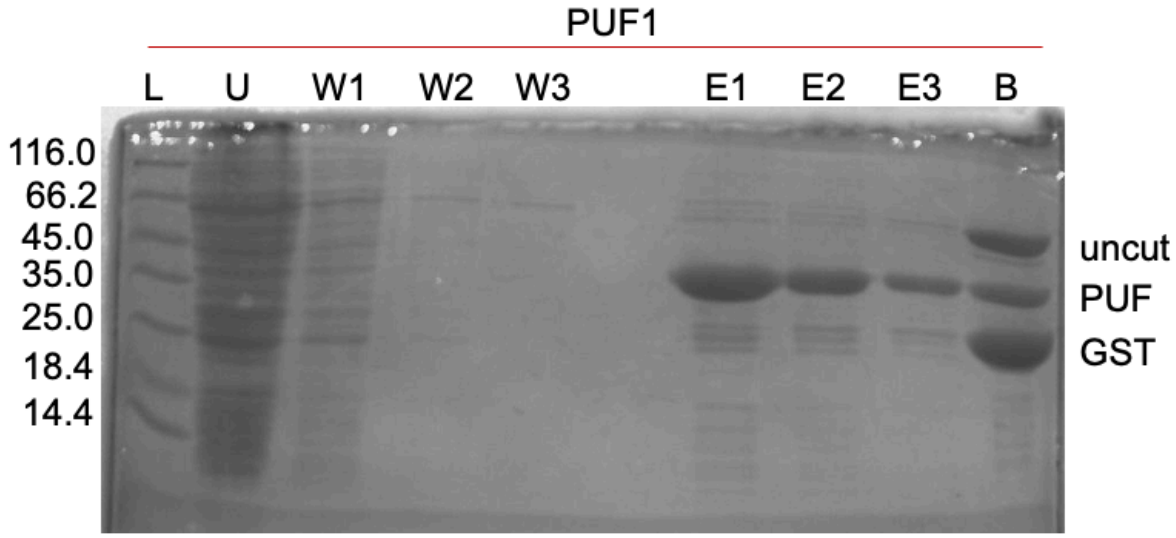


Figure 3.4: 15% SDS-PAGE gel, showing the different fractions obtained from large-scale purification and overnight TEV protease cleavage of PUF1-GST. The unstained protein ladder (ThermoFisher Scientific) is in kDa. Abbreviations: L – Ladder, U – unbound fraction, W – wash fractions, E – elution fractions, B – GST resin.

The GST affinity resin bound PUF1-GST well, as indicated by the lack of a band in the unbound and wash fractions. Washing was conducted well as seen by the lack of any protein in washes 2 and 3. Overnight TEV protease cleavage was successful and produced $> 95\%$ pure PUF1 protein (labelled as 'PUF' at approximately 35 kDa) in all three elution fractions. However, there was still uncleaved PUF1-GST bound to the resin even after TEV cleavage, suggesting the TEV protease used was not very active, thus, a different prep of TEV protease was used in subsequent purifications. When the protein was quantified by NanoDrop, the A_{260nm}/A_{280nm} was greater than 2.5 for all three elution fractions, and a peak at 280 nm was not visible due to skewing by the large peak at 260 nm suggesting that the protein was contaminated by nucleic acids.

3.3.3 NUCLEASE-TREATED PURIFICATION OF PUF1-GST

In order to reduce the A_{260nm}, the next preparation of PUF1-GST was treated with nucleases, and PUF1-GST was purified in the same manner. Figure 3.5 shows the effect of the nuclease treatment on the purification of PUF1-GST. At room temperature for 2 hours (Figure 3.5A), no PUF1 was present in the elution fractions, yet the resin had PUF1-GST bound. At 4°C overnight (Figure 3.5B), there was no PUF1 present in the elution fractions, however, there was cleaved PUF1 still bound to the resin in addition to uncleaved PUF1-GST. Due to the lack of protein in the elution fraction, no quantitation using NanoDrop was carried out. To exclude inactivity of TEV as a factor, a different batch of highly active TEVPR was used, which did not change the results obtained.

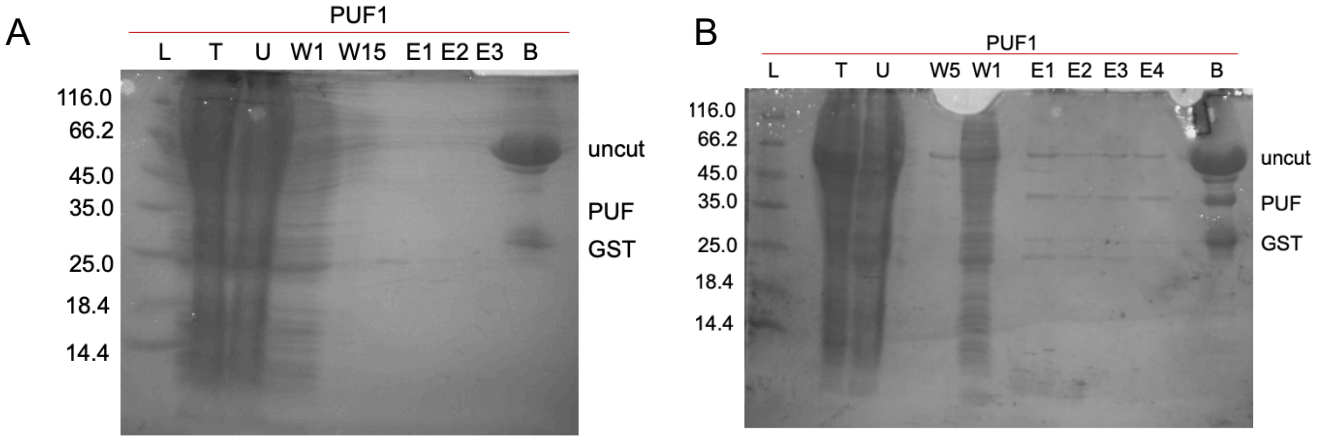


Figure 3.5: 15% SDS-PAGE gel, showing the different fractions obtained from large-scale purification and overnight TEV protease cleavage of PUF1-GST after **A** – nuclease treatment for 2 hours at room temperature, and **B** – nuclease treatment overnight at 4°C. The unstained protein ladder shown (ThermoFisher Scientific) is in kDa. Abbreviations: L – Ladder, U – unbound fraction, W – wash fractions, E – elution fractions, B – GST resin.

3.3.4 PUF1 BINDING RNA TARGETS

PUF1 was diluted in various buffers in order to test its binding to cognate RNA in a more physiological buffer and in HIVPR assay buffer. It was hypothesized that the preparation of PUF1 with a high A_{260nm} would still be partially active due to competition of labelled RNA with bacterial RNA from the purification. Figure 3.6 shows that the labelled RNA did succeed in competing out the contaminant RNA.

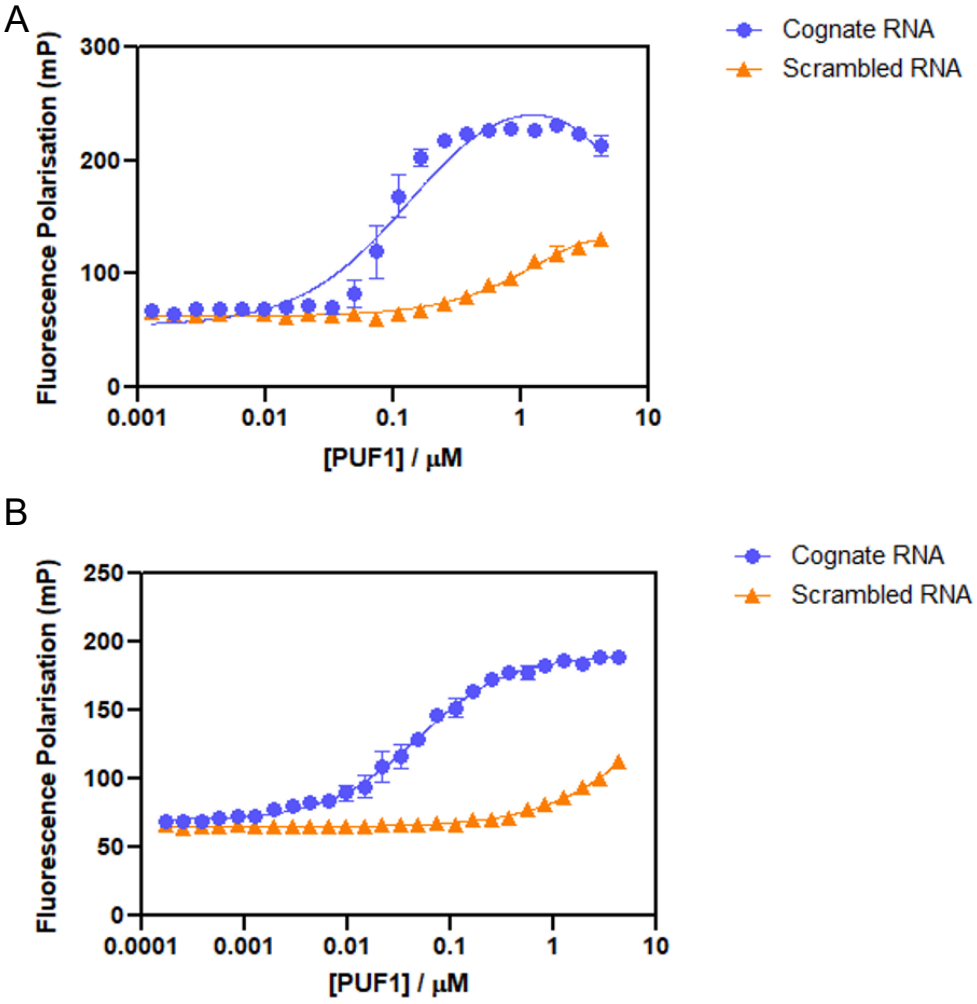


Figure 3.6: Fluorescence polarisation assay of PUF1 binding to its cognate, labelled RNA in two buffers. 5-FAM labelled RNA (1 nM) and purified PUF1 protein at varying concentrations were incubated at room temperature for 30 mins and then read. Curves were then fit to a one-site total binding equation in GraphPad Prism. **A** – using FP buffer, **B** – using PBS at pH 7.4. Error bars represent SEM from three biological replicates.

Thus, the PUF1 protein binds to its cognate RNA (K_d of approximately 100 nM) with significantly higher affinity than to a scrambled negative control (estimated from the log plot; data did not reach saturation and therefore could not be fitted) as seen by the sigmoidal increase in fluorescence polarisation with incremental increases in protein concentration. The baseline is non 0 due to the background fluorescence polarisation of the RNA, equivalent to the only RNA and buffer control. The protein and buffer control showed no change fluorescence polarisation.

The protein did not bind its cognate RNA in either the low-salt or high-salt HIVPR assay buffer (data not shown), thus, for subsequent experiments, PBS at pH 7.4 was used so that the results are more representative of physiological conditions, and as this buffer also gave the best titration curves.

The same assay was then repeated with a new set of RNAs, which will be used when testing both PUF1 and PUF2 together as they contain both, one, or none of the cognate binding sites for those proteins. There is no RNA spacer between the cognate binding sites of PUF1 and PUF2 in these RNAs, and thus they are called the L0 RNAs hereafter. The PUF1 binding site is 5' to the PUF2 binding site and thus, the nomenclature used to distinguish between the RNAs are:

1. **L0_YY** = PUF1 and PUF2 cognate sites
2. **L0_YN** = PUF1 cognate site and scrambled PUF2 site
3. **L0_NY** = scrambled PUF1 site and cognate PUF2 site
4. **L0_NN** = scrambled PUF1 and PUF2 sites

Cognate and scrambled are used to designate whether the RNA contains the cognate or scrambled site for that specific PUF protein. Figure 3.7 shows that the PUF1 protein binds as expected to the cognate RNAs (K_d of approximately 1 μM) and does not bind to the scrambled L0 RNAs. The fluorescence polarisation increases with increasing PUF1 concentration sigmoidally only for the RNAs containing the cognate site compared to those containing the scrambled sites for PUF1.

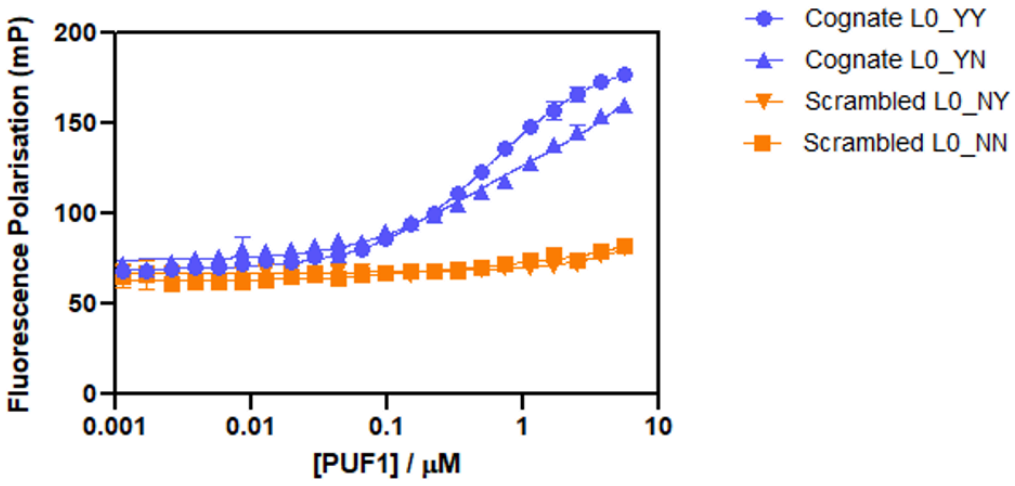


Figure 3.7: Fluorescence polarisation assay of PUF1 binding to L0 RNAs in PBS pH 7.4. 5-FAM labelled RNA (1 nM) and purified PUF1 protein at varying concentrations were incubated at room temperature for 30 mins and then read. Curves were then fit to a one-site total binding equation in GraphPad Prism. Error bars represent SEM from three biological replicates.

3.3.5 PUF2 BINDING RNA TARGETS

PUF2 was expressed and purified as a GST fusion in the same manner as PUF1 (with > 95% purity and an A_{260nm}/A_{280nm} of ~ 2.5). Fluorescence polarisation assays were then conducted in PBS pH 7.4 using the L0 RNAs to determine the binding capacity of PUF2 as shown in Figure 3.8.

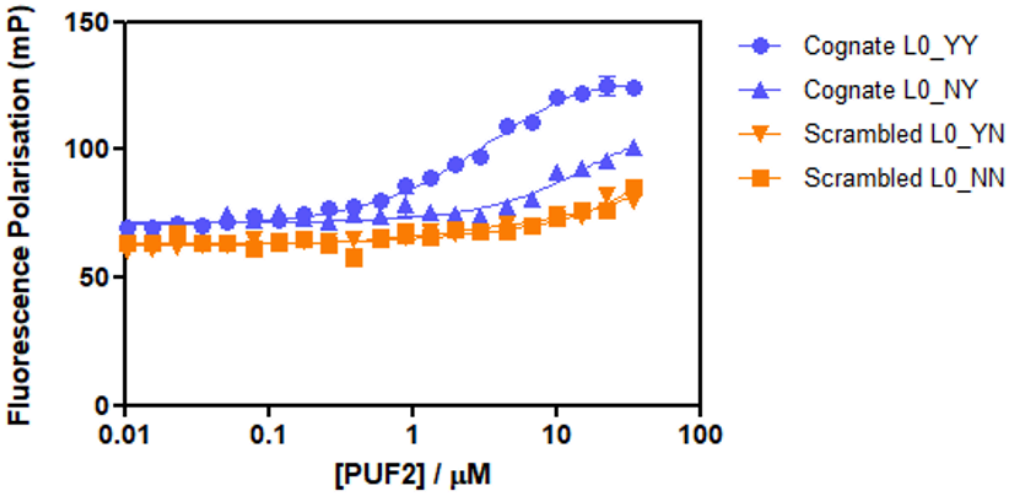


Figure 3.8: Fluorescence polarisation assay of PUF2 binding to L0 RNAs in PBS pH 7.4. 5-FAM labelled RNA (1 nM) and purified PUF2 protein at varying concentrations were incubated at room temperature for 30 mins and then read. Curves were then fit to a one-site total binding equation in GraphPad Prism. Error bars represent SEM from three biological replicates.

Fluorescence polarisation increases sigmoidally as expected for the L0_YY RNA, but this does not occur for the L0_NY RNA. As expected, both negative controls L0_YN and L0_NN did not show any binding. The fluorescence polarisation at saturation for the L0_YY RNA was approximately 120 mP units is significantly lower than what was observed using PUF1 (~ 175 mP units); this phenomenon is explained further in the discussion. In addition, the lack of similarity between fluorescence polarisation of L0_YY and L0_NY when binding PUF2 when compared to PUF1 may suggest binding of PUF2 to a non-cognate site, or, more likely, instability and degradation of L0_NY RNA. The K_d for PUF2 binding to its cognate RNA is approximately 5 µM.

3.3.6 PUF1 AND PUF2 SIMULTANEOUS BINDING TO RNA TARGETS

As PUF1 and PUF2 were shown to bind their cognate sites individually, simultaneous binding assays by fluorescence polarisation were carried out using a higher concentration of PUF1 and PUF2. Both proteins were mixed in an equimolar ratio prior to use in the assay.

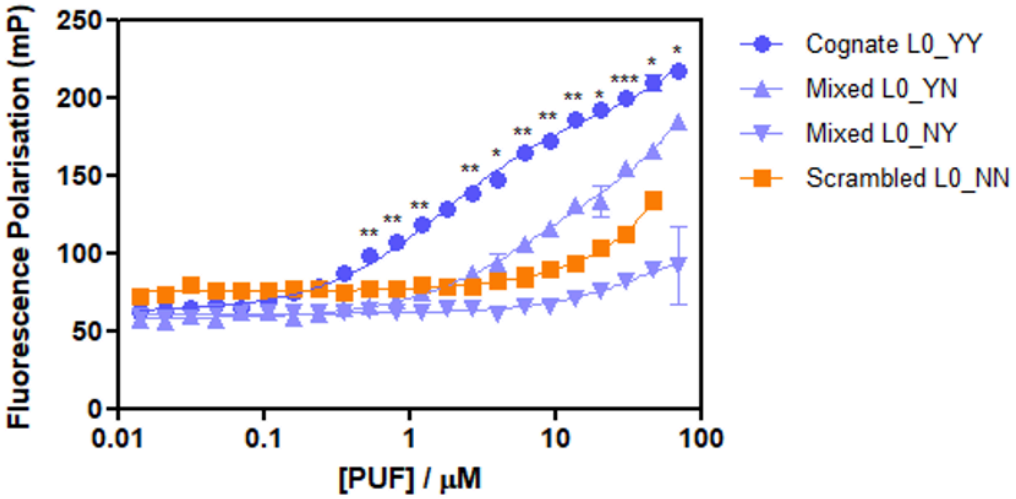


Figure 3.9: Fluorescence polarisation assay of PUF1 and PUF2 simultaneously binding to L0 RNAs in PBS pH 7.4. 5-FAM labelled RNA (1 nM) and 1:1 mixture of purified PUF1 and PUF2 proteins at varying total concentrations were incubated at room temperature for 30 mins and then read. Curves were then fit to a one-site total binding equation in GraphPad Prism. Error bars represent SEM from three biological replicates. Significance levels; * - $p<0.05$, ** - $p<0.01$, *** - $p<0.001$.

The data were fit using a one-site binding equation as it could not be fit accurately to the two-site equation (for L0_YY) due to incomplete saturation of binding. Binding could not be saturated in this assay because when the PUF proteins are mixed, their concentrations halve meaning double the concentration, relative to single binding assays, would be required to saturate the RNA. The PUF proteins could not be concentrated to these levels due to precipitation and sticking to the concentrator membrane. Despite this, Figure 3.9 shows that when both PUF1 and PUF2 are exposed to the L0 RNAs, the fluorescence polarisation of the L0_YY RNA is significantly higher than the L0_YN RNA at nearly every concentration level. The estimated K_d of dual binding to the L0_YY RNA is 10 µM which is weaker than that for PUF1 and PUF2 individually. As expected, the L0_NN negative control RNA shows much

weaker binding compared to the two aforementioned RNAs. As seen in Figure 3.9, unexpectedly, the L0_NY RNA failed to register even slightly increased fluorescence polarisation even at the highest concentration (lower than the negative control). Thus, this assay has shown with great promise that PUF1 and PUF2 can simultaneously bind a single RNA with no spacing between their cognate sites.

3.4 DISCUSSION

3.4.1 PURIFICATION OF THE PUF1 PROTEIN

Solubilisation protein tags are a common method used to increase the yield of protein in the soluble fraction when expressed in *E. coli*. It involves fusing the PUF to a more soluble protein (tag) and subsequently cleaving the tag after purification in order to yield just the PUF[113,119]. The tags used in this study were the 6xHis, GST and Hislip tags. In order to assess the effect of the tags on solubility, small-scale expression tests were carried out. Figure 3.3 shows that only the GST tag was able to solubilise the PUF1 protein[120]. When purified on a large scale, the PUF1 protein had significant nucleic acid contamination, the majority of which was expected to be RNA due to the RNA binding properties of the protein. Thus, RNase treatment was attempted several times to reduce this. However, RNase treatment causes the GST affinity resin to become clumped (not shown) and unable to elute correctly as seen in Figures 3.5A and B. Inefficient elution could be due to the inability of TEVPR to access its cleavage site due to steric hindrance by the nucleases, instability of nuclease-free PUF, or simply by an unknown negative interaction between the nucleases and TEVPR. Additionally, the nuclease treatment itself may not be able to digest bound RNA due to a 'protection' effect from PUF binding which prevents RNase-mediated degradation[121].

Passing the sample through the Emulsiflex homogenizer multiple times, however, reduced the nucleic acid contamination due to shearing of bacterial nucleic acids. Any remaining nucleic acid contamination was likely to be RNA which could not be removed (even using high salt to disrupt the electrostatic interactions) due to very high affinity binding to the PUF1 protein. Thus, this PUF1 protein was used in subsequent fluorescence polarization assays with the expectation that the K_d observed would likely be an underestimation due to the presence of competitor RNAs from the bacteria.

3.4.2 FLUORESCENCE POLARISATION ASSAYS OF PUF-RNA INTERACTIONS

To assess the binding of PUF1 to RNA, fluorescently-tagged RNA molecules were used to carry out a fluorescence polarization assay. This assay is widely used to measure the binding of proteins to nucleic acids, including the PUF proteins[90,102]. The principle behind this assay is that the binding of a higher molecular weight molecule (PUF1 protein) to a lower molecular weight labelled molecule (RNA) reduces the tumbling of the labelled molecule in a solution thereby increasing the polarization of the light emitted by the fluorophore upon excitation[122].

Initially, PUF1 binding was assayed using labelled RNA molecules containing either one cognate binding site or a scrambled negative control. Binding was tested in a buffer (F_p buffer) used in a similar published assay[102] and in PBS pH 7.4 in order to test the protein in more physiological conditions. Binding was seen in both buffers to the cognate RNA but not to the scrambled RNA (Figure 3.6) and the difference between the K_d's obtained is likely due to differences in buffer compositions affecting binding. Some binding can be seen to the scrambled RNA as the PUF concentration increases, however, this is likely a non-specific interaction with the RNA due to concentration-dependent crowding[123]. RNA only controls matched the baseline values for all figures shown, and no fluorescence polarisation change was observed for protein only controls in any experiments. A high-salt (mM) buffer was also tested, as the HIVPR to which the PUF1 will be fused is only active in high salt concentrations[124]. No binding was observed, likely due to the inhibition of the formation of electrostatic binding interactions. Thus, PBS pH 7.4 was used for subsequent assays.

When testing dual binding of PUF proteins, a longer RNA was used containing a cognate site for each PUF protein; thus, the binding of PUF1 was tested on the aforementioned L0 RNAs. As with the previous set of RNAs, binding of PUF1 was only seen when the RNA molecule contained a PUF1 binding site (L0_YY and L0_YN) and not in the presence of scrambled sites (L0_NY and L0_NN). The difference in K_d compared to the single-site RNA tested initially is likely due to structural dynamics of a longer RNA affecting binding.

As seen in Figure 3.8, PUF2 was able to bind the L0_YY RNA, but not the L0_NY RNA. After thorough analysis of the RNA sequences, it was determined that PUF2 is very unlikely to be binding a non-cognate site on the L0_YY RNA due to the presence of several mismatches to the PUF2 target sequence and since it does not bind to the other scrambled L0_YN RNA. When Figure 3.9 is examined, we can see that the L0_NY RNA exhibits lower fluorescence polarisation than even the negative control RNA L0_NN. These observations together suggest that it is not the PUF2 protein binding non-specifically; the L0_NY RNA is likely to be unstable and very easily degraded, thereby preventing PUF2 binding and producing the observed results. Thus, binding of PUF2 to the undegraded L0_YY RNA, and not to L0_YN, is considered evidence of binding to the correct cognate site.

PUF2 binding to L0_YY showed a lower peak fluorescence polarisation than PUF1 likely due to the fact that PUF2 binds at the 3' end of the RNA, leaving the flexible 5' end able to undergo some movement resulting in a relatively lower measured fluorescence polarisation compared to PUF1 which binds at the 5' end and does not allow for this movement by flexibility. This decoupling of fluorescence polarisation from binding activity due to probe flexbility is known as the propeller effect[122].

When used together, PUF1 and PUF2 resulted in significantly higher fluorescence polarisation on the L0_YY RNA than the other single-binding controls ($p<0.01$ at almost every concentration), suggesting that they are able to bind simultaneously to the same RNA molecule even with no spacing between the cognate sites, which is consistent with the structural and biophysical data obtained by Qiu *et al.*[93] using the *S. pombe* PUF1 protein. Interestingly, the K_d for dual binding was greater than that for single binding of the PUF proteins individually suggesting there may be a negative co-operativity in binding due to steric clashes between the two PUF proteins, although this should be repeated to allow fitting to the two-site binding equation to extract more accurate binding affinities. Furthermore, the experiment should be repeated by adding PUF proteins to the RNA consecutively, rather than simultaneously, to further verify them binding concurrently to the RNA. The low binding affinity detected is likely an underestimation, as mentioned previously, but the degree of underestimation is unknown. Since the NASPER system requires the binding of two PUF fusion proteins to a single RNA, higher affinities would provide a higher likelihood that this dual binding occurs

in the cell and results in the reconstitution of effector domains, hence the low affinity observed *in vitro* may suggest that the PUF domains may not bind to their target RNA effectively enough in the cell to induce dimerisation and reconstitution of the effector domains. However, previously observed K_d values for PUF domains are in the low nanomolar region, suggesting the underestimation of the K_d is high.

Despite the low affinities observed *in vitro*, I was confident in using these PUF1 and 2 RNA-binding domains for the NASPER system, as they can bind their cognate sites specifically, bind simultaneously on to the same RNA molecule, and have been shown in the literature to have low nanomolar binding affinities, which would be beneficial to NASPER. Hence, I moved on to using these proteins in cell-based assays testing the NASPER system.

CHAPTER 4 – CELL-FREE STUDY OF HIV PROTEASE

4.1 INTRODUCTION

The HIV-1 protease (HIVPR) is a 99-amino acid aspartate protease containing of the conserved triplet (D25-T26-G27) active site found in most known aspartyl proteases. However, unlike cellular monomeric aspartate proteases, the HIVPR is only active in its dimeric state, as each monomer contributes one of the two required catalytic aspartate residues (D25)[111,125]. During HIV-1 infection, the protease cleaves the HIV Gag-Pol polyprotein to liberate individual proteins[111] and thus form mature virions. There are two structural interfaces in the protease that contribute to dimerization: 1) the termini interface comprising residues 1 – 4 from one monomer, and 96 – 99 from the other to form a stable beta sheet that contributes 75% of the energy of dimerization[111]; and 2) the active-site interface formed by the catalytic triplet providing the remaining required energy[125] (Figures 4.1 and 4.2).

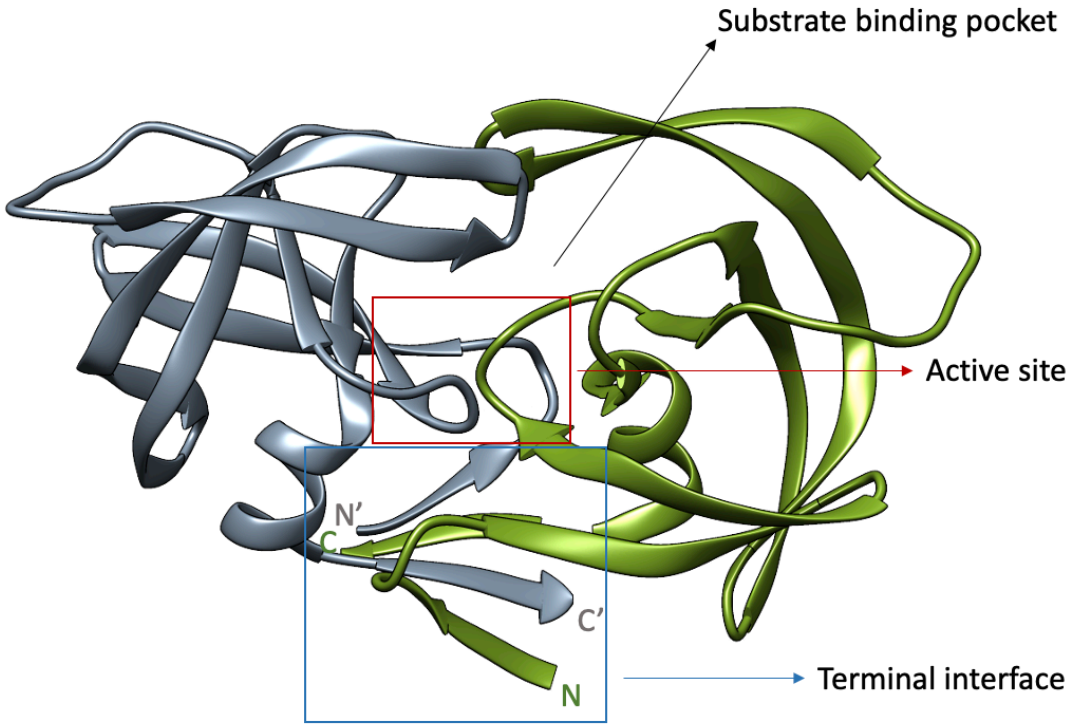


Figure 4.1: Schematic of the structure of the HIVPR (PDB: 1DMP). Active site, terminal interface and substrate binding pocket are highlighted.

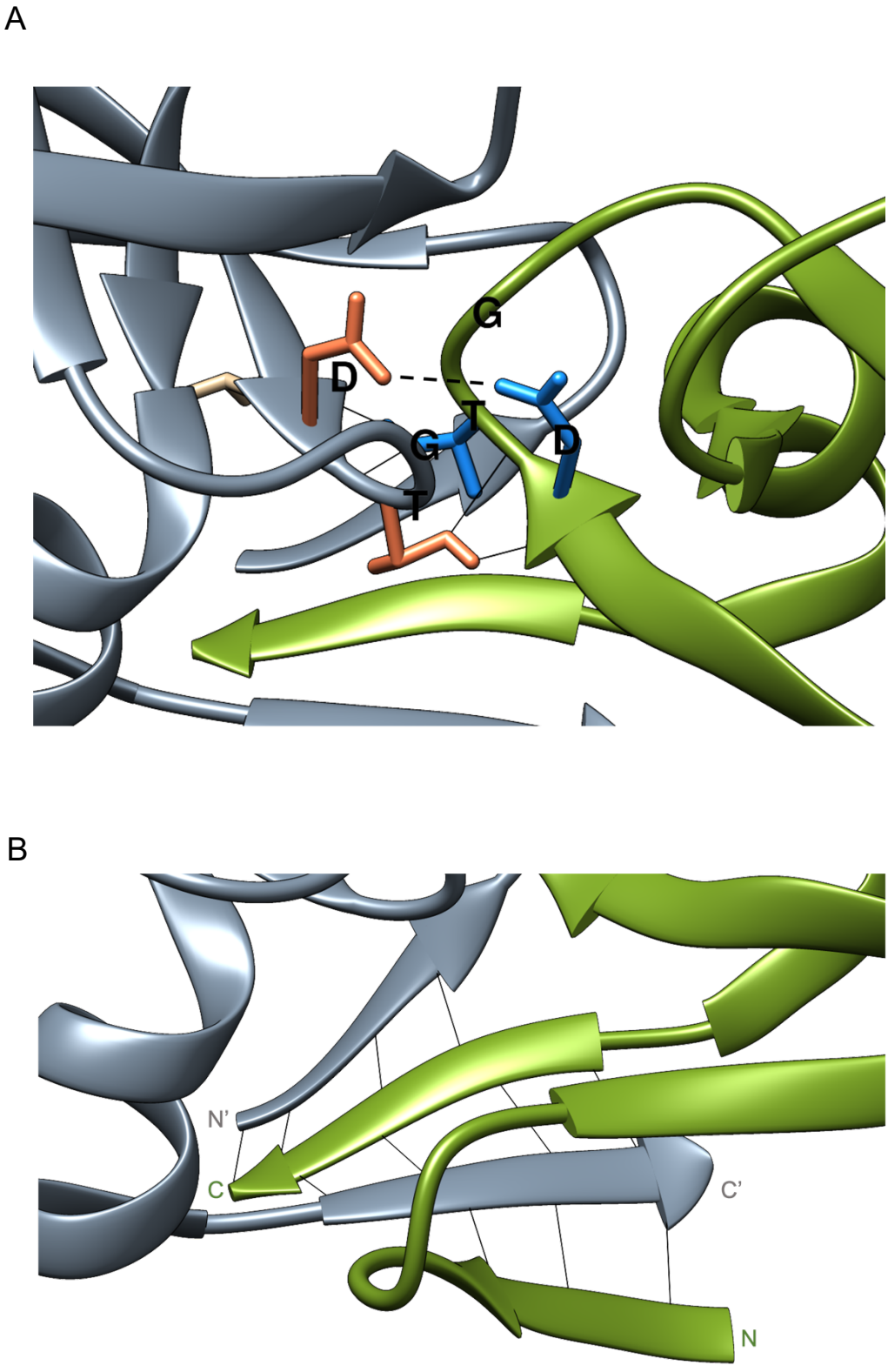


Figure 4.2: **A** – Schematic of the structure of the active site of HIVPR with catalytic DTG triad labelled. Solid black lines show hydrogen bonding within the active site. Dashed black line shows catalytic residue. **B** – termini interface with hydrogen bonding shown as solid black lines.

The protease substrate binds within the substrate binding pocket adjacent to the active site (Figure 4.1) in an extended conformation allowing interactions to occur with the appropriate amino acid side chains thereby conferring specificity to the enzyme. Hydrophobic peptides are the preferred substrate as the dimerisation interface is also hydrophobic in nature[126]. The mechanism of proteolysis by this enzyme is still unclear; however, several mechanisms have been put forward[127–129]. The main distinction between the mechanisms is the presence of a discrete peptide intermediate. Meek *et al.*[128] propose that a discrete peptide intermediate is produced, whereas Jasloski *et al.*[127] propose a concerted mechanism with no intermediate. However, despite the differences, all the mechanisms require a water molecule in the mechanism.

The protease has a broad substrate specificity, and there is currently no consensus on a conserved sequence required for catalysis. Indeed, the native Gag-pol polyprotein that is cleaved by the HIV protease during HIV infection has ten asymmetric and non-homologous cleavage sites[130]. Computational models are being used to predict potential cleavage sites by using the Gag-pol sites as a baseline. These models are good predictors of amino acids required within the substrate as they agree with experimental findings, but they have yet not yielded a consensus cleavage motif. This may be because there is in essence no particular conserved amino acid sequence that is recognised by the protease and, instead it recognises a substrate shape while other factors such as the availability of water molecules within the active site are important even when the substrate is bound in its pocket[130].

Current efforts to purify and test the HIV protease *in vitro* have relied on chemical denaturation and re-folding from inclusion bodies, as the protein tends to be insoluble when expressed in *E. coli*[111-113]. When used as a fusion protein, urea denaturation and re-folding may cause problems, as the protein domains may not fold correctly to their native conformation without the presence of the appropriate cellular chaperones and machinery[131,132]. Hence it is necessary to optimize the expression of the HIV protease such that it is present in the soluble form. The production of proteins in the soluble form primarily relies on the use of solubilization tags[119,133] such as thioredoxin and glutathione-S-transferase (GST). Since the HIV protease is hydrophobic, I used the solubilising tag – the His-lipoyl tag[134]. This tag is derived from *B. Stearothermophilus* and is prone to lipoylation is *E. coli*[135], which may be the

reason it can improve solubility (although this has not been verified experimentally). It has also been suggested that the lipoyl tag can act as an intramolecular chaperone thereby facilitating native folding of the protein[134]. Additionally, the conventional His and GST tags will also be tested.

Thus, the aims of this Chapter are to solubilize the HIV protease using different tags and subsequently test the purified protein for activity using a known substrate before testing the protein in cells. Initial plans were to test the NASPER concept *in vitro* before moving on to cell-based experiments and thus, optimising the purification of active HIVPR *in vitro* was essential. Although subsequently, further *in vitro* characterisation was not carried out, the experiments in this Chapter will demonstrate whether the HIVPR being used is active and can be used for future *in vitro* work using the HIVPR domains.

4.2 METHODS AND MATERIALS

4.2.1 LARGE-SCALE EXPRESSION AND PURIFICATION

C41 (DE3) competent *E. coli* cells (Lucigen) were transformed with Hislip-HIVPR plasmid DNA using heat-shock as previously described and plated on ampicillin LB-agar plates at 37°C overnight. All colonies on the plate were scraped into 700 ml of 2xYT broth, with ampicillin at a final concentration of 100 µg/ml, at 37°C with shaking at 200 rpm till the OD_{600nm} = 1. Once this was achieved, cells were induced using 1 mM IPTG and grown for a further 3 hours at 37°C with shaking at 200 rpm. Cells were then pelleted at 35,000 x g for 10 minutes and the supernatant discarded. Pellets were re-suspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole), at 2 ml/g of cell paste, with 1 mM DTT and SigmaFast protease inhibitor cocktail tablet, EDTA-free (SigmaAldrich), at a 1X final concentration, freshly added. Cells were homogenised using the EmulsiFlex homogeniser under 5 rounds of pressure at 10,000 psi. Homogenised cells were centrifuged at 35000 x g for 35 minutes and the supernatant containing the soluble fraction was collected.

4.2.1.1 OPTIMISATION OF PURIFICATION

1. **Immobilised Metal Affinity Chromatography (IMAC) only**: Soluble fraction was applied in upflow to a 5 ml HisTrap Excel column (GE Healthcare) equilibrated with IMAC wash buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole and 1 mM DTT freshly added) at a flow rate of 2 ml/min using an AKTA Pure FPLC system (GE Healthcare). The column was washed in 10 column volumes of IMAC wash buffer and then eluted in 2 ml fractions using IMAC elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole and 1 mM DTT freshly added). 10 µL aliquots of soluble, flow-through, wash (only first and last) and elution fractions (chromatogram selected) were added to 2 µL of 6x SDS loading dye and boiled for 10 minutes. These samples were run on a 15% SDS-PAGE gel in 1x TGS running buffer at 80V for 20 minutes and 180V for 50 minutes. Gels were stained with Coomassie Brilliant Blue stain and imaged under white light. Fractions containing protein were pooled and dialysed into TEV cleavage buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl and 1 mM

DTT freshly added) to which TEV protease was added at a final concentration of 2 µM. The sample was then left to cleave either at room temperature for 2 hours or at 4°C overnight after which it was flash frozen in 25% glycerol for long term storage at -80°C.

2. **Size-Exclusion Chromatography (SEC) with low-salt buffer**: After overnight TEV cleavage, the sample was applied to a HiLoad 16/600 Superdex 75 pg SEC column (Cytiva) equilibrated in HIVPR storage buffer (20 mM MES pH 6.0, 5 mM EDTA, 0.01% v/v Triton-X100, 10% v/v glycerol and 1 mM DTT freshly added) at a flow rate of 1 ml/min. Elution fractions (chromatogram selected) were treated with TEVPR and run on SDS-PAGE gels as mentioned earlier.

3. **IMAC and SEC with high-salt:** The sample was initially purified as per 'IMAC only'. However, after overnight TEV cleavage, the sample was applied to a HiLoad 16/600 Superdex 75 pg SEC column (GE Healthcare) equilibrated in high salt TEV cleavage buffer (50 mM Tris-HCl pH 8.0, 900 mM NaCl and 1 mM DTT freshly added) at a flow rate of 1 ml/min. Elution fractions (chromatogram selected) were treated and run on SDS-PAGE gels as mentioned earlier.

4.2.2 HIVPR ACTIVITY ASSAY

HIVPR (stored in TEV cleavage buffer with 25% glycerol) was dialysed into HIVPR assay buffer[113] (20 mM sodium phosphate pH 6.0, 900 mM NaCl and 1 mM DTT freshly added) using a 3 ml Slide-A-Lyzer Dialysis Cassette with a 3 kDa MWCO (Thermo Fisher) for 1 hour. The concentration was then estimated using the A_{280nm} on a Nanodrop Spectrophotometer (ThermoFisher) using the extinction coefficient of cleaved HIVPR, calculated on ProtParam (https://web.expasy.org/protparam/), of $12490 M^{-1}cm^{-1}$. HIVPR was 2-fold serially diluted in the same buffer. Anthranilyl HIVPR substrate (Bachem) was diluted to a working stock of 100 µM in DMSO (Fisher Scientific) and used at a final concentration of 10 µM.

5 µL of 100 µM HIVPR substrate was added to 45 µL of HIVPR at concentrations of 1 – 11 µM in an OptiPlate 384-well Black Microplate (Perkin Elmer) and incubated at 25°C for 10

minutes. The plate was then read on a Clariostar Microplate Reader (BMG) with an excitation wavelength of 280 nm and an emission wavelength of 430 nm. All experiments were conducted in triplicate. Data were analysed for statistical significance using a one-way analysis of variance followed by Tukey's honest significant difference test as the sample sizes of different groups were the same and data were normally distributed.

Details of the HIVPR substrate used are shown below:

Key:

1. **Bold – Anthranilic acid**
2. *Italics – p-nitro phenylalanine*
3. Underlined – HIVPR Cleavage Site

Substrate Amino Acid Sequence: **Abz**-TI-Nle|*Phe*-QR-NH_2

4.3 RESULTS

4.3.1 HISLIP TAG SOLUBILISES HIVPR

In order to express HIVPR in the soluble fraction using *E. coli* cells, three different purification tags – His, GST and Hislip – were tested. These constructs were expressed and purified on a small scale in order to ascertain which tag produced the most soluble protein. Figure 4.3 shows that neither the His tag nor the GST tag were able to solubilise the HIVPR construct, as there was no clear overexpressed band (darkened regions) in the soluble fractions when these tags were used. However, the Hislip tag did produce over 90% of HIVPR protein in the soluble fraction. All constructs were at approximately the expected molecular weights (molecular weight of HIVPR plus purification tag). HIVPR-Hislip migrated as two separate bands, an artefact also seen in the PUF1-Hislip construct. HIVPR-Hislip was subsequently taken further into large-scale purification.

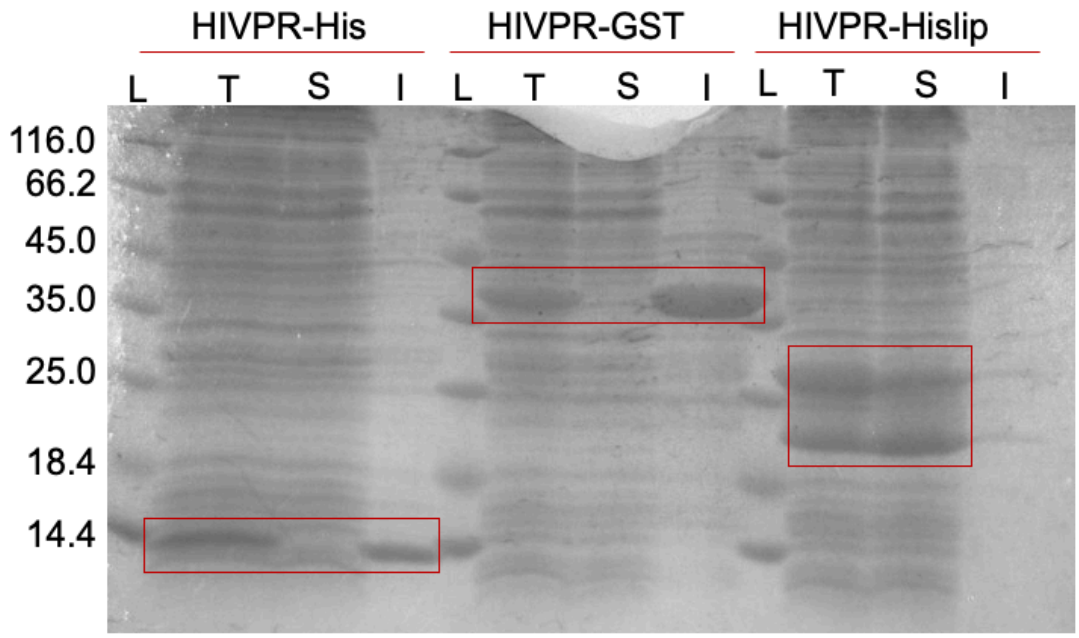


Figure 4.3: 15% SDS-PAGE gel showing expression of the three HIVPR constructs in the total, soluble and insoluble fractions after cell lysis. The bands corresponding to the construct are boxed. The unstained protein ladder (ThermoFisher Scientific) is in kDa. Abbreviations: L – Ladder, T – total protein fraction, S – soluble protein fraction, I – insoluble protein fraction.

4.3.2 IMAC PURIFICATION OF HIVPR-HISLIP

Large-scale expression and purification of HIVPR-Hislip was carried out using IMAC without prior nuclease treatment. HIVPR-Hislip purified well using IMAC (> 90% purity with an A_{260nm}/A_{280nm} ratio of ~0.6), as shown in Figure 4.4A, and the molecular weight of the major band in the figure corresponded to expected molecular weight of the HIVPR-Hislip construct. A minor band at a lower molecular weight can also be seen post-IMAC. Figure 4.4B shows the results of the TEV cleavage of HIVPR-Hislip at room temperature and 4°C. Incomplete cleavage is suggested by the presence of a band corresponding to uncleaved HIVPR-Hislip, which has the same intensity as that of the Hislip and cleaved HIVPR bands. However, cleavage was partially successful as shown by the presence of the free Hislip tag and HIVPR at their appropriate molecular weights. In order to separate the uncleaved fraction and Hislip tag from the cleaved HIVPR, the samples were subjected to low or high-salt SEC.

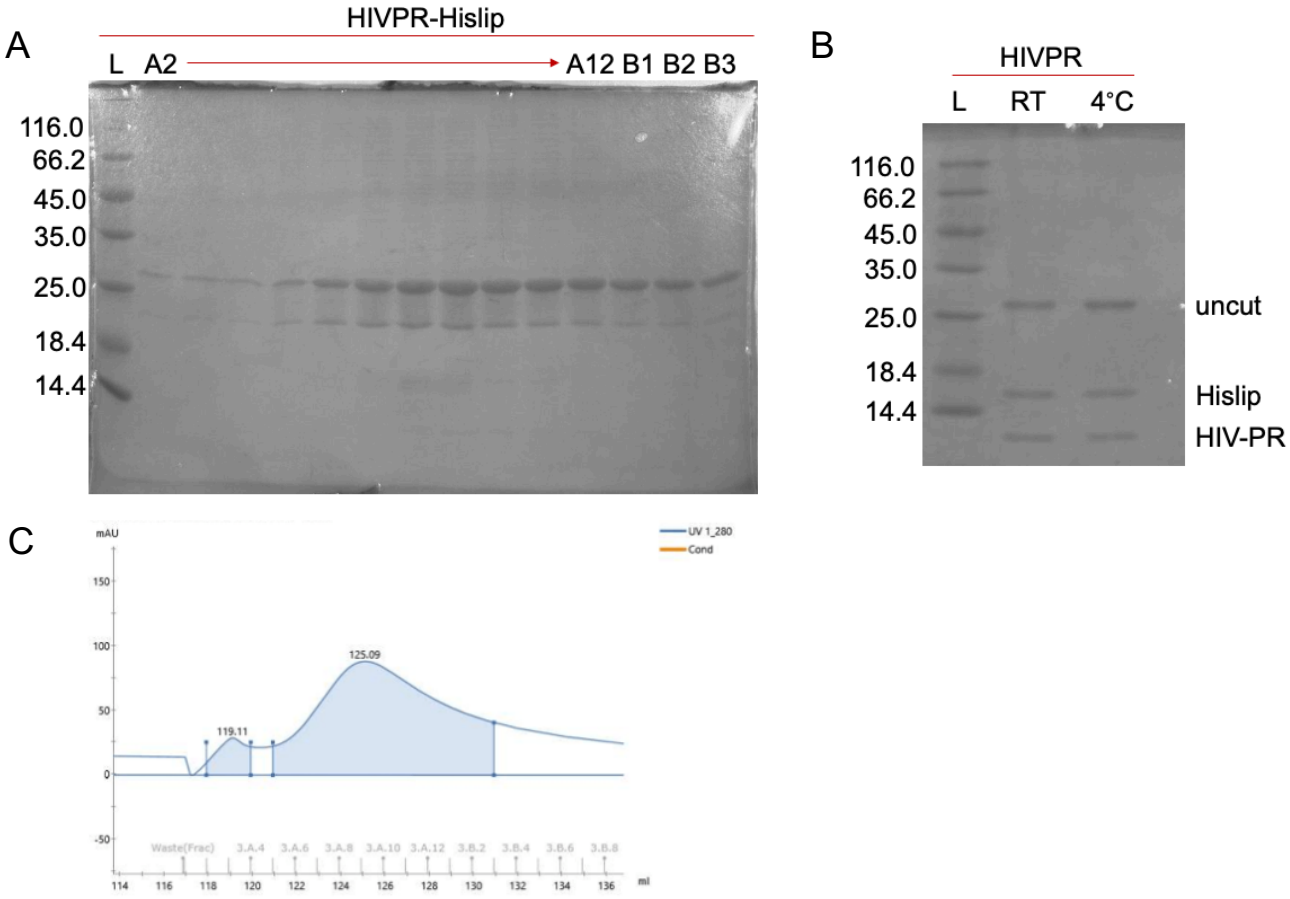


Figure 4.4: **A** – 15% SDS-PAGE gel showing elution fractions of the IMAC purified HIVPR-Hislip construct. **B** – 15% SDS-PAGE gel of HIVPR-Hislip after TEV protease cleavage either at room temperature or 4°C. **C** – UV elution profile of the HIVPR-Hislip protein from transformed bacterial cell lysate using IMAC showing the UV 280 nm profile. The unstained protein ladder (ThermoFisher Scientific) is in kDa. Abbreviations: L – Ladder, RT – room temperature.

4.3.3 SEC PURIFICATION OF HIVPR-HISLIP

In order to separate the cleaved HIVPR, the sample was subjected to SEC. However, the SEC only managed to separate the free Hislip tag during the run (Peak 2 in Figure 4.5). Upon inspection of the chromatogram (not shown), a few minor peaks were seen during cleaning using a high-salt solution at the end of the SEC run. These are referred to as the 'end peaks' in Figure 4.5 and contained the uncleaved and cleaved HIVPR (very faint bands).

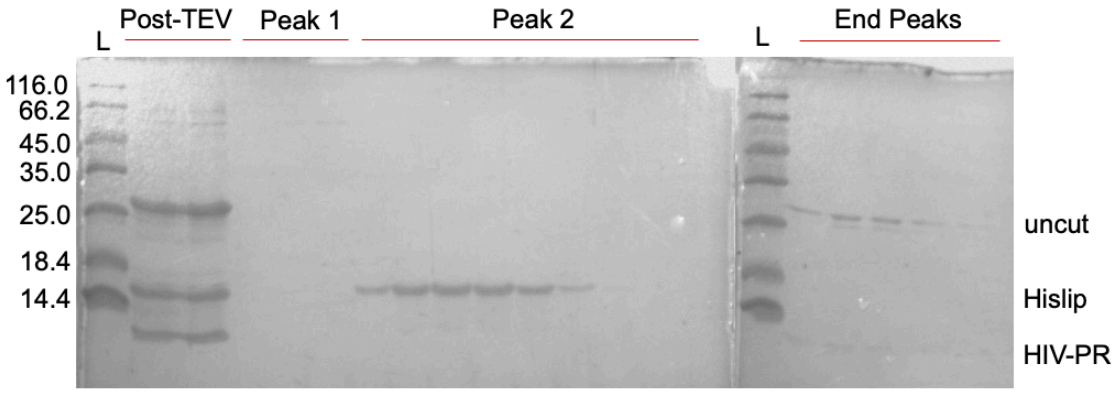


Figure 4.5:15% SDS-PAGE gel showing elution fractions of the SEC-purified cleaved HIVPR construct in a low-salt buffer. Peak 1, 2 and End Peaks refer to those seen on SEC chromatogram and lanes under these headings represent each individual elution fraction in that particular peak on the chromatogram. The ladder shown is an unstained protein ladder (ThermoFisher Scientific) and is in kDa. Abbreviations: L – Ladder, U – unbound fraction, W – wash fractions, E – elution fractions.

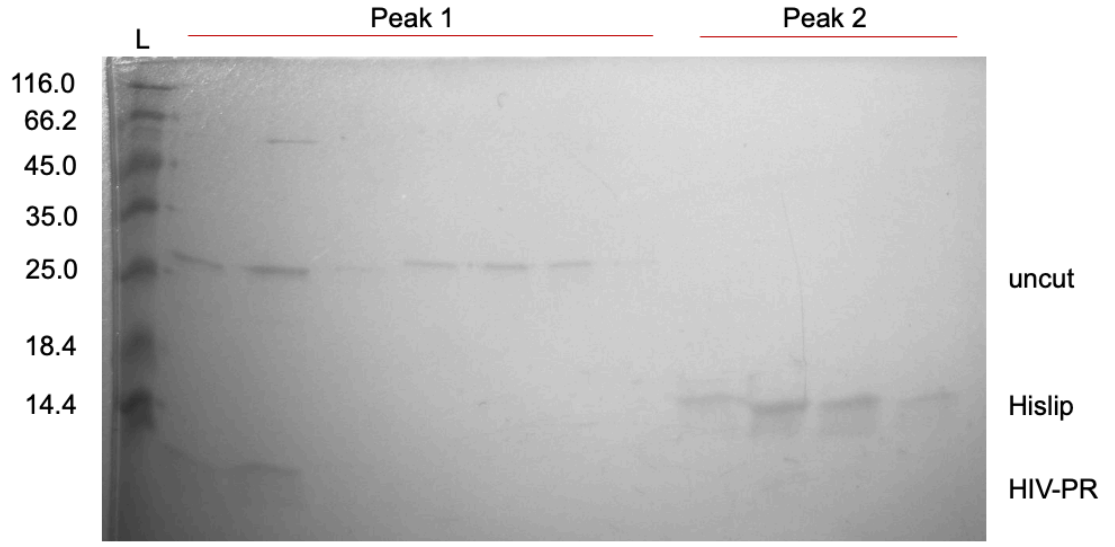


Figure 4.6: 15% SDS-PAGE gel showing elution fractions of peak 1 and peak 2 of the SEC-purified cleaved HIVPR construct after TEV cleavage using a high salt buffer. Lanes under these headings represent each individual elution fraction in that particular peak on the chromatogram

The SEC was repeated using a high-salt buffer to improve the elution of HIVPR, and the gel obtained from the elution fractions is shown in Figure 4.6. Two peaks were observed during this run and Figure 4.6 shows that the first peak contained both cleaved HIV protease and uncut HIV protease, whereas the second peak showed solely the Hislip tag (as in Figure 4.5). However, no "end peaks" were observed during this purification unlike the previous SEC using low-salt buffer. Chromatograms for the SEC purifications were not provided due to technical issues that caused their deletion from the system.

4.3.4 HIVPR ACTIVITY ASSAY

HIVPR purified using IMAC was tested for activity using an anthranilyl HIVPR substrate, the fluorescence of which increases upon cleavage. A control with TEV protease at a final concentration of 0.5 μM was also used in order to exclude the effects of any remaining TEV protease in the prep.

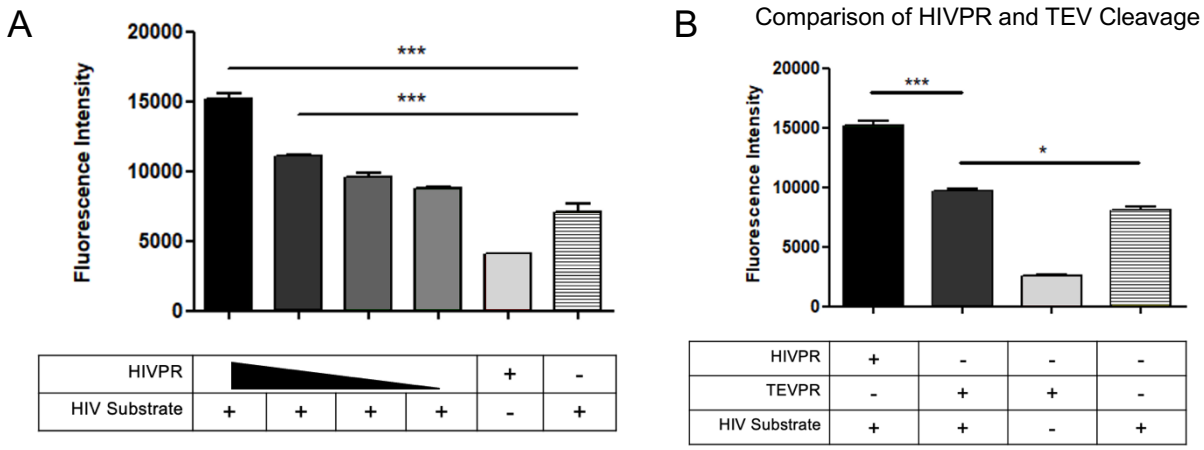


Figure 4.7: **A –** Fluorescence intensity upon cleavage of anthranilyl HIVPR substrate (10 μM) by HIVPR at various concentrations. Control experiments using HIVPR and anthranilyl HIVPR substrate alone in HIVPR assay buffer are also shown. Significance levels; * - $p<0.05$, ** - $p<0.01$, *** - $p<0.001$. n=3 **B –** Fluorescence intensity upon cleavage of anthranilyl HIVPR substrate (10 μM) by undiluted HIVPR and TEVPR (0.5 μM) . Control experiments using TEVPR and anthranilyl HIVPR substrate alone in TEV cleavage buffer are also shown. Significance levels; * - $p<0.05$, ** - $p<0.01$, *** - $p<0.001$. Error bars represent SEM from three independent biological experiments.

Figure 4.7A shows that anthranilyl HIVPR substrate alone in the HIVPR assay buffer produces an average background fluorescence intensity of around 6500 RFU. However, when combined with HIVPR, the average fluorescence intensity is significantly higher ($p<0.001$) and decreases with decreasing HIVPR concentration. In contrast, only the two highest used

concentrations of HIVPR exhibited this significant difference. To rule out any effects of TEVPR, a control experiment was carried out. Figure 4.7B shows that there is a very slight difference between the fluorescence intensities of HIVPR substrate alone and with TEVPR in TEV cleavage buffer (likely due to contaminants in the TEVPR preparation), whereas there is a very significant ($p<0.001$) difference between the fluorescence intensities exhibited by HIVPR and TEVPR when combined with the substrate (in their respective buffers).

4.4 DISCUSSION

4.4.1 HIVPR IS SOLUBILISED BY THE HISLIP TAG

The HIVPR dimer consists of a hydrophobic core, to which each monomer contributes 19 amino acids[126]. Thus, the monomeric form of HIVPR will have an exposed hydrophobic core and is consequently likely to be prone to aggregation and the formation of inclusion bodies when expressed in bacterial cells[112,113]. Although it is possible to purify the HIVPR from inclusion bodies in an active form[111-113], we decided to optimize the expression of the protein such that a soluble form of the protein can be acquired. Thus, we needed to find conditions under which HIVPR is expressed in the soluble fraction in order to properly test the proteolytic function of HIVPR. To assess the effect of the three tags on solubility, small-scale expression tests were carried out. Figure 13 showed that only the Hislip tag was able to produce HIVPR in the soluble fraction. Although Volontè *et al.*[113] showed that a GST tag was able to solubilize HIVPR, under the conditions used here it failed to do so. In further attempts to see whether GST can solubilize HIVPR to even a small degree, a large-scale expression and purification was carried out, but the elution fractions from the GST affinity resin showed that there was no protein present (data not shown). It is possible that the Hislip tag, which contains a soluble lipoyl domain and has been used to solubilise intrinsically disordered proteins[134], was effective in solubilizing HIVPR because the termini of HIVPR are relatively disordered[136]. Thus, the Hislip-tagged form of the protein was used in large scale expression and purification.

4.4.2 HISLIP-HIVPR CAN ONLY BE PURIFIED BY IMAC

IMAC produced pure Hislip-HIVPR (Figure 4.4A), and subsequent TEV cleavage liberated the HIVPR alone, but the cleavage was inefficient. In order to separate the cleaved and uncleaved proteins (shown in Figure 4.4B), size exclusion chromatography (SEC) was used. In HIV storage buffer, the elution peaks only showed the presence of the Hislip tag with no uncleaved or cleaved HIVPR seen (Figure 4.5). At the end of the run, a transient increase in salt concentration due to the original sample buffer resulted in the elution of more erratic ‘end peaks’. On a gel, these showed the presence of very small quantities of the uncleaved and cleaved HIVPR (Figure 4.5). From this, I thought that the lack of salts in the HIVPR MES

buffer may be affecting how the HIVPR interacts with the column material. It is possible that the HIVPR is retained in the column for a longer duration due to electrostatic interactions within the column which are then disrupted in the presence of higher salt concentrations[137]. Thus, another SEC run using high salt was performed. The high-salt conditions still resulted in elution of only very small quantities of uncleaved and cleaved HIVPR (Figure 4.6), although this elution occurred during the run, rather than during column cleaning after the run had completed, suggesting salt concentration was indeed the reason for the proteins eluting in the 'end peaks' during the first run. When the peaks were run on an SDS-PAGE gel, the uncleaved and cleaved HIVPR appeared to elute in the same chromatogram peak despite their different molecular weights. This may be due to the cleaved HIVPR dimerising and thereby eluting with the uncleaved HIVPR. Consequently, the HIVPR purified by IMAC, which was then cleaved by TEV was used in HIVPR proteolysis assays in order to test its function.

Thus, despite being able to solubilize the HIVPR using a Hislip tag, the purification still remains a challenge that requires further optimization to isolate only the HIVPR separated from its uncut form and the Hislip tag. New studies have shown that using tandem column setups may be of use in purifying the HIV protease. Sherry *et al.*[138] used a tandem HiTrap DEAE and CM column set up and subsequently re-ran the eluate on a HiTrap DEAE column. Although this was done for a denatured inclusion body containing HIV protease, the sequential ion exchange set up may be applicable to our soluble form as well.

4.4.3 PURIFIED HIVPR IS PROTEOLYTICALLY ACTIVE

A fluorogenic peptide substrate containing a known HIVPR cleavage site was used in order to determine the proteolytic activity of HIVPR. The fluorescence of the uncleaved substrate is very low due to the presence of an intramolecular quencher. The fluorescent dye was anthranilic acid, and the quencher was p-nitrophenylalanine[139], a known FRET pair. HIVPR was diluted serially, however, the absolute protein concentration was estimated to be 1 – 11 μM due to the presence of uncleaved HIVPR and Hislip tag in the sample. The assay showed that at the highest two concentrations, HIVPR is proteolytically active, as the fluorescence is significantly higher ($p<0.001$) than the background fluorescence of the substrate in the HIV assay buffer. This effect is likely entirely due to cleaved HIVPR and not uncleaved HIVPR for three reasons:

1. HIVPR dimerisation has shown to be inhibited *in vitro* by the presence of an N-terminal extension[125,136].
2. HIVPR dimerisation is essential for its proteolytic activity[111].

As there may have been some TEV protease remaining in the sample (although this was not seen on the gel in Figure 4.7B), a series of control experiments were conducted to exclude the effects of TEV protease contamination. TEV protease, at the same concentration that was used to cleave the Hislip-HIVPR, was incubated with the HIVPR substrate in TEV cleavage buffer in the same way as for HIVPR. This assay showed that the fluorescence from TEV protease is only slightly higher ($p<0.05$) than background fluorescence from the HIVPR substrate in TEV cleavage buffer. As TEV protease is a highly specific protease, it was not expected to cleave the HIVPR substrate (which did not contain a TEV cleavage site). Additionally, TEV protease is likely to have very little activity in the HIVPR assay buffer due to the high salt concentration and relatively low pH, both of which reduce TEV activity[140,141]. Therefore, the cause of increased fluorescence of the HIVPR substrate upon exposure to TEV protease is likely to be caused by proteolysis by contaminants in the TEV protease preparation (the TEV protease was expressed and purified in the lab).

These control data, together with the fact that substrate fluorescence is significantly higher when the HIVPR is used rather than the TEV protease ($p<0.001$), suggest that the activity seen is indeed due to the HIVPR. Thus, it was established that the HIVPR construct that I had made was active, and therefore could be used for the cell-based experiments.

Nevertheless, the experiments highlight areas for future research, such as optimizing the purification of soluble HIVPR and increasing the efficiency of cleavage of the Hislip-tagged HIVPR potentially by using a different protease and protease site such as the PreScission protease[141].

CHAPTER 5 – HIVPR IN THE NASPER SYSTEM

5.1 INTRODUCTION

Having confirmed that the purified HIVPR protein was active, I was confident in moving on to testing the HIV protease domain in cells, along with the first iteration of the NASPER system using the HIV protease domain fused to the RNA-binding PUF protein domain. A schematic diagram for the mechanism of action of this system is shown below in Figure 5.1.

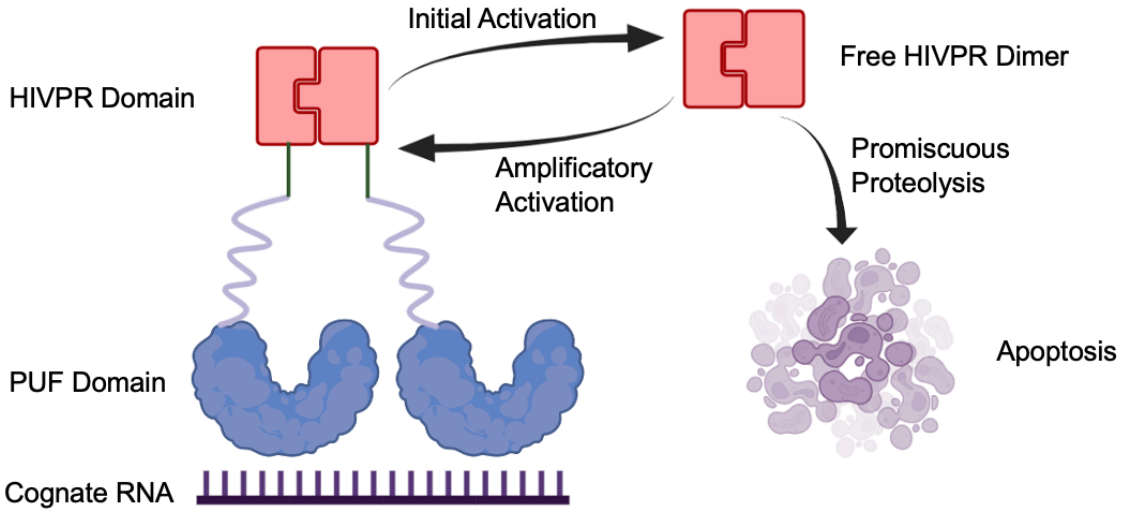


Figure 5.1: Proposed mechanism of HIVPR-based NASPER system.

However, auto-dimerization of the fusion protein (PUF-HIVPR) must be avoided in order to prevent proteolysis in cells lacking the cognate RNA. Thus, the protease will need to be modified accordingly such that auto-dimerisation does not occur but the active D25-T26-G27 triplet structure is conserved to retain proteolytic activity when the two monomers are forced into proximity by binding of the fused PUF domains to the cognate RNA scaffold. Thus, of the two dimerisation interfaces (active site and termini interfaces) available to disrupt, the termini interface will be modified in order to preserve the active site.

Louis *et al.*[125] and Hayashi *et al.*[111] have studied HIVPR variants designed to modulate dimerisation. Louis *et al.*[125] classified their variants based on the concentration at which dimer populations were observable by NMR and mass spectrometry (estimated by each of the four classes) with Classes C and D being primarily monomeric even at 0.5 mM concentrations. The primary variant presented in Classes C and D that preserves the active site triplet is $HIVPR_{1\text{-}95}$ (although this can result in a more disordered, accessible active site). Results obtained by Hayashi *et al.*[111] using mass spectrometry agreed with those obtained by Louis *et al.*[125] using NMR and sedimentation equilibrium. In addition, Louis *et al.*[125] also observed that N-terminus extensions in the protease can also inhibit dimerization and increase the K_d to approximately 0.6 mM, similar to that of the Class C and D variants. This difference was with a 4 amino-acid extension, and they have also shown that longer extensions are even more effective at preventing dimerisation.

Mutant Name	Class	Approximate Dimerisation K_d (nM)
Wild-type	N/A	<1
R8Q	A	500
D29N	B	10000
T26A	C	100000
N-extension	C	100000
ΔC	D	1000000

Table 5.1: HIVPR mutations, classes and approximate dimerisation affinities determined by Louis *et al.*[125].

Studies into the N-terminus extension phenomenon have shown that it is the lack of dimerisation initially which leads to inactivity. N-terminal fusions of HIVPR with GST, GFP and hsp70[142-146] (an anti-parallel dimer[147]) and the p6 region (found upstream of HIVPR in the HIV polyprotein containing the HIVPR autoprocessing cleavage site) can undergo dimerisation (due to the fused domain) resulting in an increased local concentration and subsequent autoactivation of the protease. Lindsten *et al.*[143] used similar GFP-p6-HIVPR fusion to *Huang et al.*[146] and showed that in such a fusion construct, the HIVPR can effectively autoprocess and kill mammalian cells through proteolysis. Blanco *et al.*[148] have also shown that the HIVPR alone is capable of causing cellular apoptosis, likely due to the cleavage of key cellular cytoskeletal proteins such as actin, vimentin and desmin[149], thereby suggesting its suitability for use as a driver of apoptosis. These studies suggest that the HIV protease domain can indeed be fused to domains which induce dimerisation successfully and subsequent dimerisation of these fusion proteins can then drive apoptosis through promiscuous proteolysis. This research provides confidence that the HIV protease domain is highly suited to the NASPER system conceived within this thesis.

Based on the literature, I chose the N-terminus extension modification, as it fits well with the arrangement of domains in the fusion protein. The PUF RNA-binding domain is placed on the N-terminus of the HIVPR domain, therefore acting as the N-terminus extension required to inhibit spontaneous dimerisation, and consequently activity. If the N-terminus extension alone was not effective at preventing spontaneous activity, the other mutations found in the literature could also be used in addition. Thus, several fusion constructs were designed to be used for the assays in the Chapter, as shown in Figure 5.2.

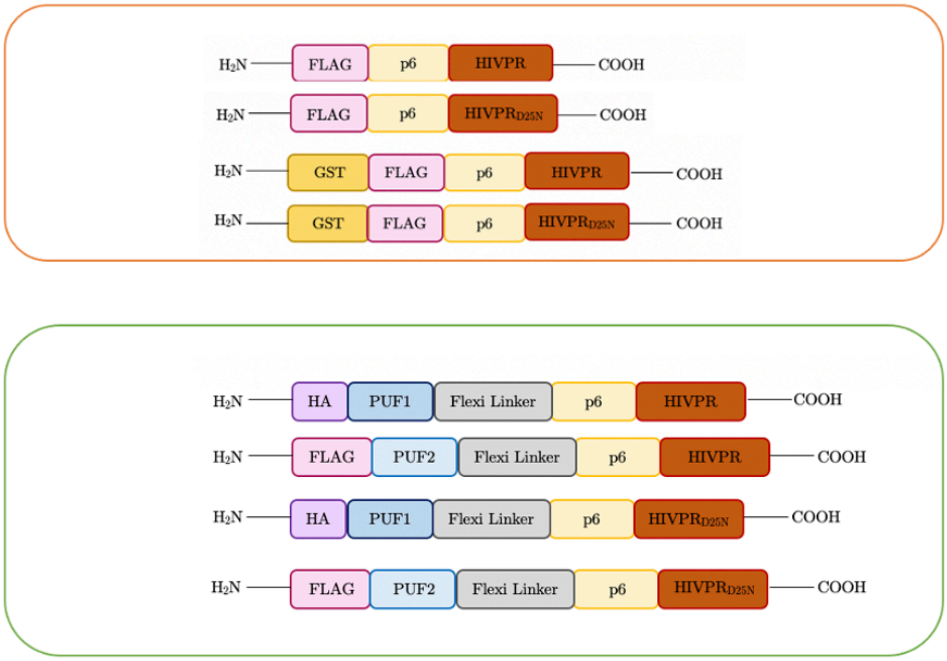


Figure 5.2: Fusion proteins designed for cell-based assays. Eight fusion proteins were generated to study the effects of HIVPR fusion proteins in cells. Different groups of fusion proteins are highlighted in different colours. HIVPR and GST-HIVPR constructs are shown in an orange box. PUF-HIVPR constructs are shown in a green box

Each protein was fused to either a 3xFLAG or 3xHA tag for immuno-detection. An N-terminal p6 domain was added to facilitate auto-activation. The p6 domain is a mini-precursor form of the wild-type N-terminal flanking sequence of the HIV protease in the polyprotein which contains a cleavage site to liberate the active wtHIVPR[142,144,145]. Studies have shown that this auto-activation is essential to HIVPR activity in cells, and thus, the eight amino acid p6 sequence was added to the fusion protein to allow for this to occur. Furthermore, as a positive control, an N-terminal GST sequence, known to dimerise, was added to allow HIVPR to dimerise and auto-process despite the N-terminal extension[142-146].

For the PUF-HIVPR fusion proteins (in the green box in Figure 5.2), a flexible linker was used to separate the PUF and HIVPR domains, in order to allow HIVPR domain dimerisation, while at the same time providing enough separation from the PUF domain to allow both domains to function independently. The sequence for the flexible linker was from published data[150,151] on the effects of artificial linkers on protein structure and has the sequence EGKSSGSGSESKST.

The aims in this Chapter are to test the HIVPR alone in cells first before moving on to testing this iteration of the NASPER system as a whole using triple transfections in HEK293 cells by comparing auto-activation between cognate and scrambled RNA conditions (cognate and scrambled sequences were fused to mCherry mRNA such that expression could be visually assessed prior to assays). The hypothesis is that firstly, the HIVPR domain will only be active when fused to GST, and secondly the PUF-HIVPR fusions will be active only when cognate RNA is transfected and not when scrambled RNA is used. Finally, these constructs will be transiently transfected, rather than stably transfected, to enable screening of a large number of constructs more efficiently, and also because as a therapeutic, these proteins would be overexpressed in the cells they target either through gene-based or protein-based delivery methods which is better modelled by transient transfections. Hence, transient transfections will be used in the experiments throughout this thesis. However, if lead constructs are found, future work can use stable cell lines to verify the results of transient transfections.

5.2 RESULTS

5.2.1 OPTIMISATION OF TRIPLE TRANSFECTION

Prior to beginning testing the fusion constructs in cells, the transfection method had to be optimized in order to minimize the influence of transfection efficiency on the observed results. As triple transfections had been planned, they were optimized based on mass of DNA transfected.

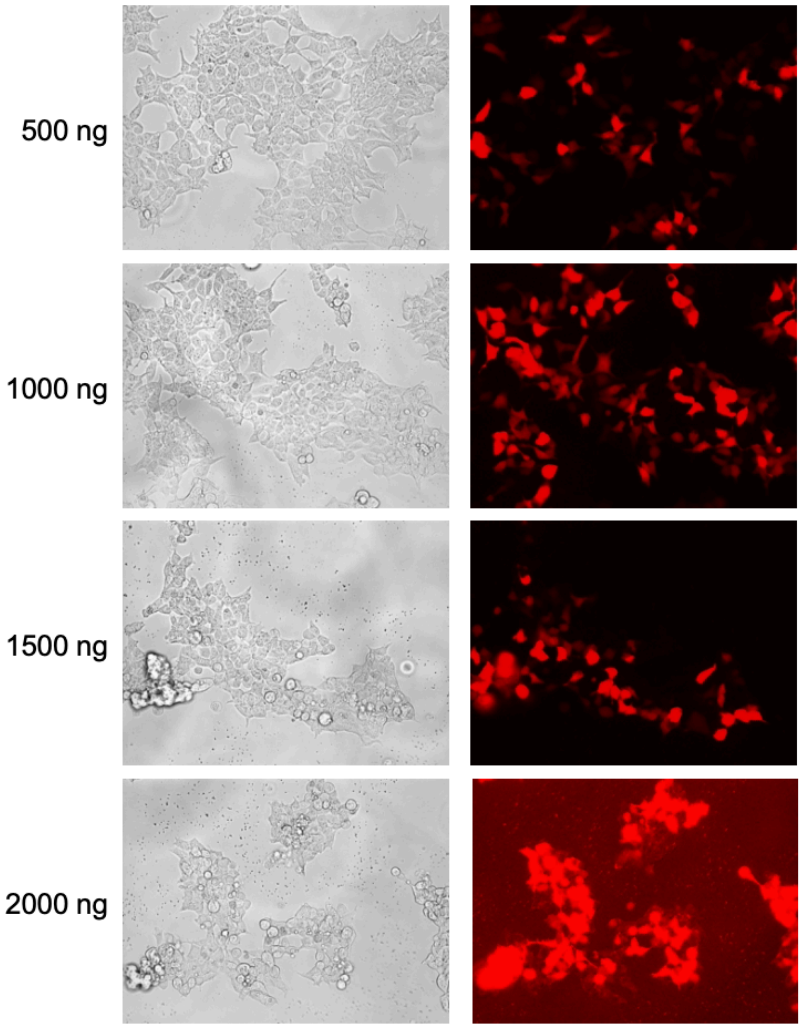


Figure 5.3: 300,000 HEK293T cells in a 12-well plate were triple transfected using various DNA masses using a 1:3 ratio of DNA to Lipofectamine 2000 over 24 hours.

Triple transfections were optimized using two different plasmids co-transfected with an mCherry expressing plasmid transfected at a 1:1:1 molar ratio with varying total DNA masses. Efficiency of triple transfection was assessed using mCherry fluorescence. Figure 5.3 shows these results and, although the red fluorescence appears to increase with increasing mass of DNA transfected, so does the blebbing observed in the brightfield images. Thus, a total transfected DNA mass of 1000 ng showed the ideal balance between maximized transfection and minimal cell death.

5.2.2 EXPRESSION OF HIVPR FUSION PROTEINS

The expression of a set of the fusion proteins was assessed using the optimized conditions.

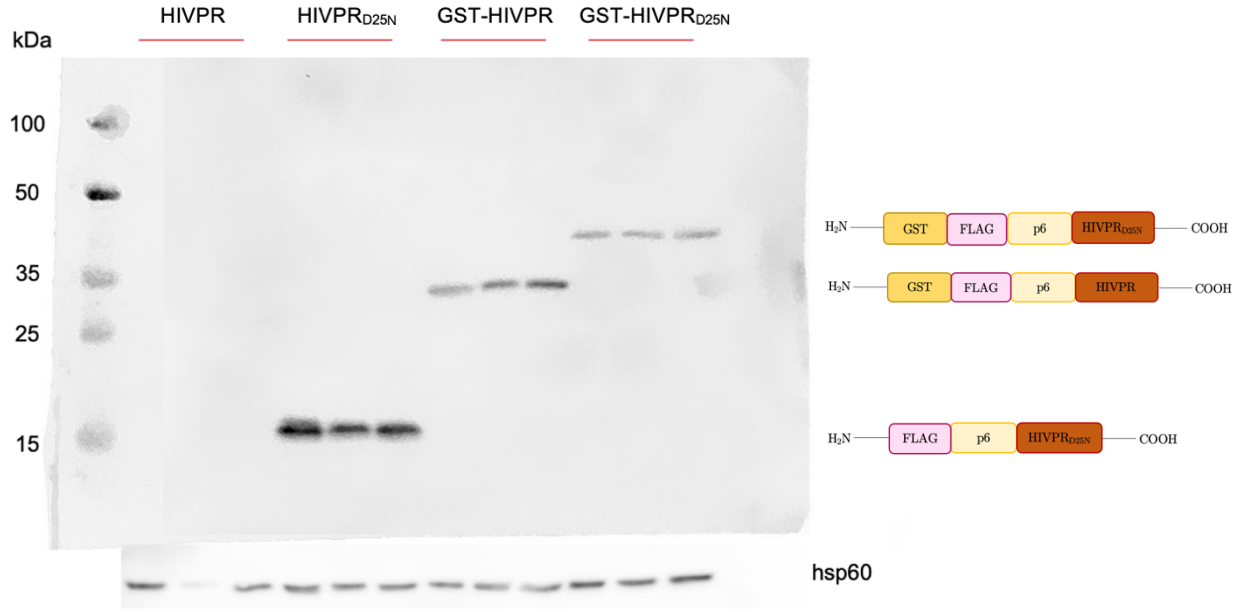


Figure 5.4: Western blot of HIVPR, HIVPR$_{D25N}$, GST-HIVPR and GST-HIVPR$_{D25N}$. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-FLAG mouse monoclonal primary antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Construct cartoons referring to the Western blot band are shown alongside the band. The three lanes under each title represent the three independent biological replicates.

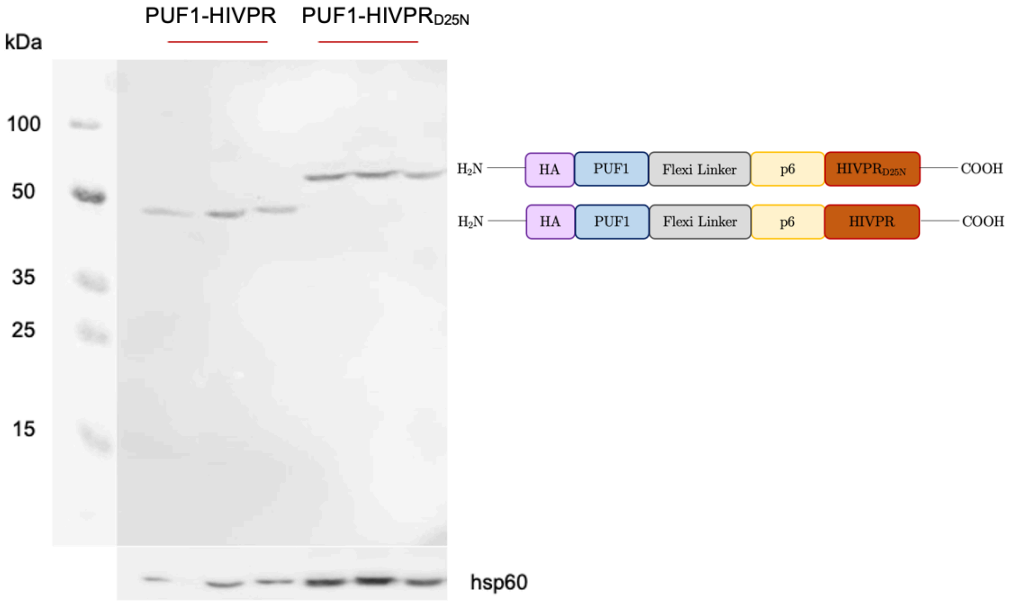


Figure 5.5: Western blot of PUF1-HIVPR and PUF1-HIVPR$_{D25N}$. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA rat monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Construct cartoons referring to the Western blot band are shown alongside the band. The three lanes under each title represent the three independent biological replicates.

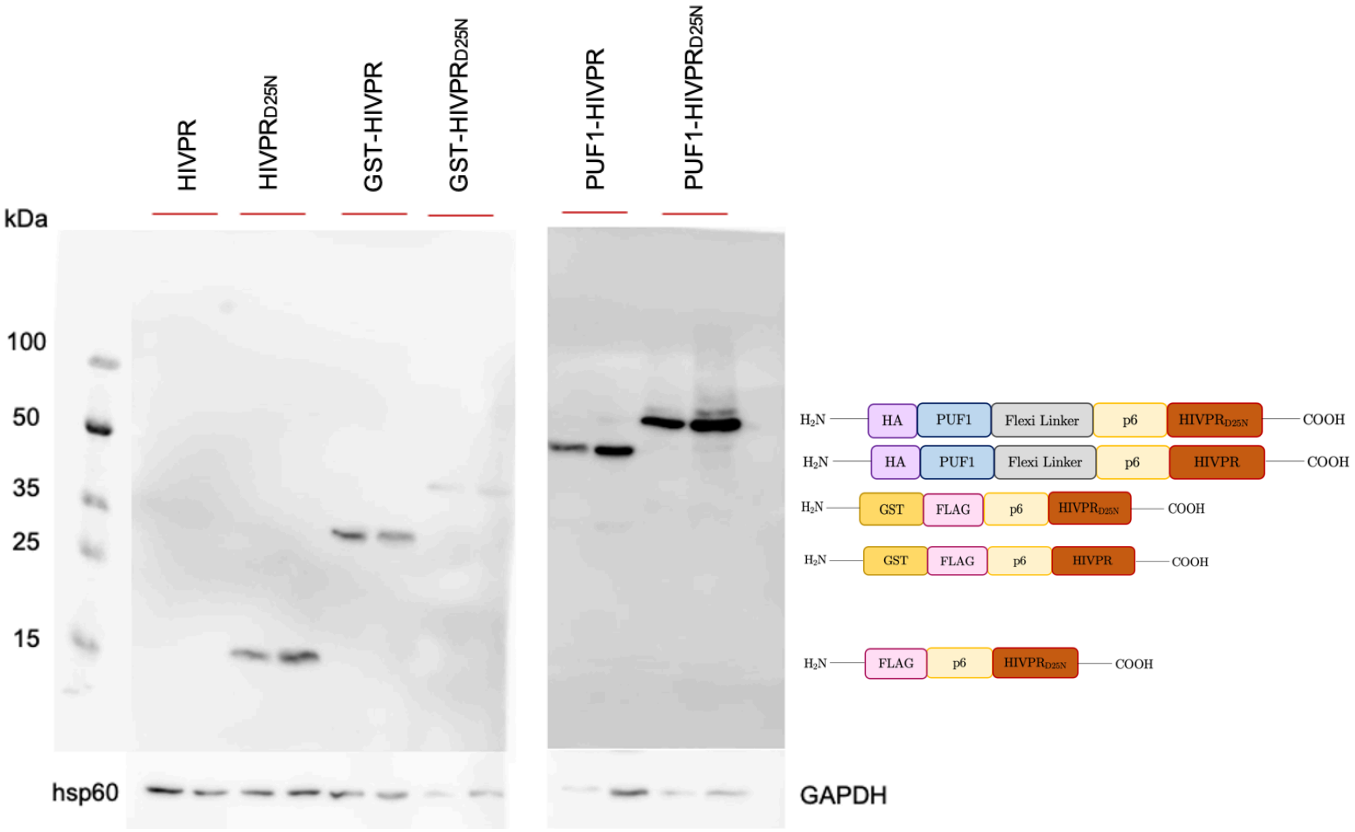


Figure 5.6: Western blot of HIVPR, HIVPR$_{D25N}$, GST-HIVPR, GST-HIVPR$_{D25N}$, PUF1-HIVPR and PUF1-HIVPR$_{D25N}$. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with the addition of darunavir in the lysis buffer. The blot was stained using either an anti-FLAG monoclonal or an anti-HA rat monoclonal primary antibody and subsequently with an anti-Mouse or anti-Rat HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 or GAPDH was used as a loading control. Construct cartoons referring to the Western blot band are shown alongside the band. The two lanes under each title represent the two independent biological replicates.

The HIVPR constructs generated for cell work were tested for expression in HEK293 cells using Western blot. Figure 5.4 shows the expression of HIVPR, HIVPR$_{D25N}$, GST-HIVPR and GST-HIVPR$_{D25N}$.

The 3x-FLAG tagged HIVPR protein an approximate molecular weight of 15 kDa. In Figure 5.4, no band corresponding to this protein is observed, however a single band is observed at the correct molecular weight for the D25N variant of the HIVPR domain, i.e., the proteolytically inactive variant used throughout this Chapter as the negative control. Thus, it is likely that the wild-type HIVPR domain auto-activates resulting in a protein fragment too small to be resolved by gel electrophoresis.

The GST-HIVPR protein has an approximate molecular weight of 41 kDa, however a band corresponding to around 30 kDa is observed. For the D25N negative control version, the correct band is seen at 41 kDa. Hsp60 loading control is mostly even across all samples.

Figure 5.5 shows the blots for the PUF1-HIVPR and PUF1-HIVPR$_{D25N}$ fusion proteins and a similar pattern as observed in Figure 5.4 is seen. The negative control shows a band at the correct molecular weight of around 55 kDa, whereas the wild-type construct 5 shows a band approximately 11 kDa lower. To ensure that the cleavage of proteins containing an active HIVPR domain was not due to trace amounts of active HIVPR cleaving the full-length proteins during cell lysis, the Western blots were repeated using a lysis buffer containing darunavir (an HIVPR inhibitor). The results of this are shown in Figure 5.6 and no change was observed of the molecular weights of the bands, and no true additional bands were seen. Faint bands and smears seen in Figure 5.6 are artefacts of Western blotting. This suggests that the effects of any HIVPR activity after cell lysis are negligible and this was not the cause for complete auto-cleavage of the fusion proteins observed.

5.2.3 EXPRESSION OF HIVPR VARIANT FUSION PROTEINS

Thus, the next step to reduce the spontaneous activity of the HIVPR domains was to try and increase the dimerisation K_d of the HIVPR domains such that spontaneous dimerisation and activation does not occur even at high protein concentrations within the cell. Five mutants were produced, each of which has a different effect on the dimerisation K_d of the HIVPR domains. The five mutant variants of the HIVPR domain used were – R8Q, T26A, D29N, R87K and ΔC. These constructs were expressed, and the results are shown below (Figures 5.7, 5.8 and 5.9). Expression of the HIVPR variants resulted in bands corresponding to the full-length construct for variants T26A, D29N, R87K and ΔC as compared to the negative control band and previously observed molecular weights. However, no bands were seen for the R8Q

variant as was for the wild-type HIVPR. Hsp60 expression was even throughout, even for the R8Q variant of HIVPR.

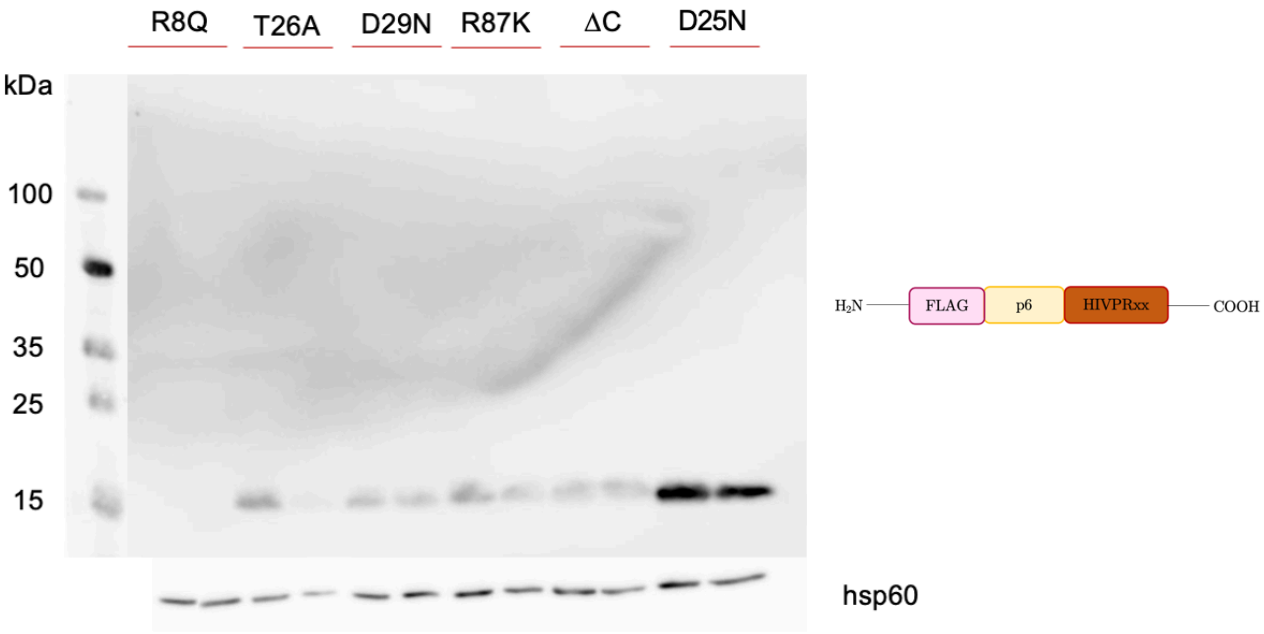


Figure 5.7: Western blot of HIVPR variants. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-FLAG monoclonal antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Original construct cartoon is shown alongside the blot. HIVPRxx refers to the fact that variants (corresponding to lane titles) of the HIVPR were used in this Western blot. The two lanes under each title represent the two independent biological replicates.

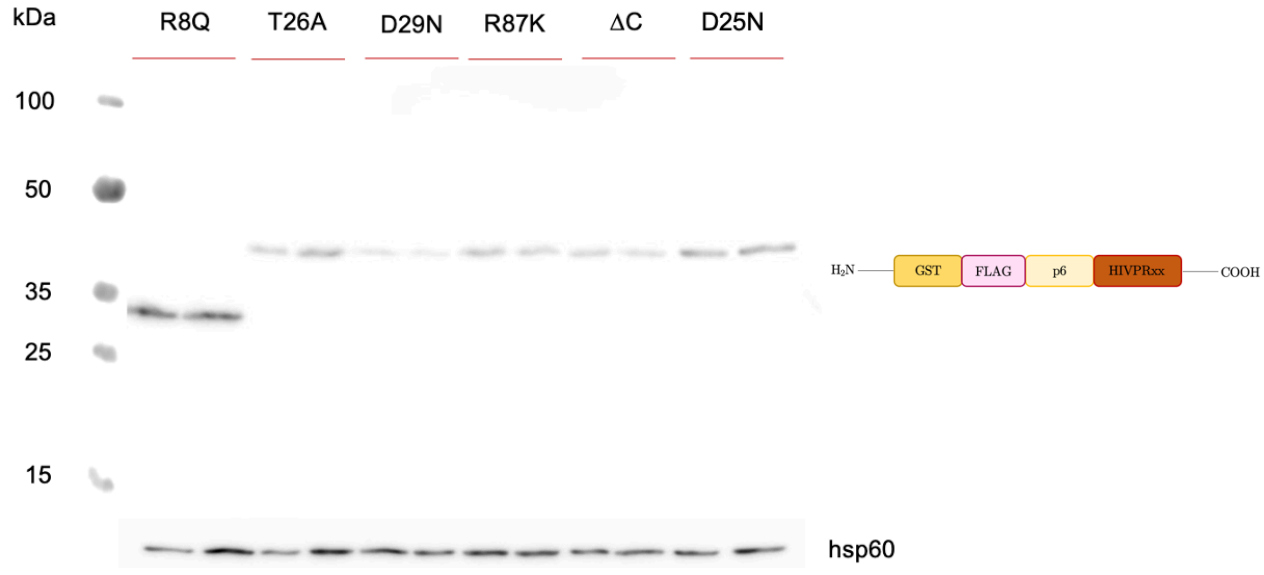


Figure 5.8: Western blot of GST-HIVPR variants. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-FLAG monoclonal primary antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Original construct cartoon is shown alongside the blot. HIVPRxx refers to the fact that variants (corresponding to lane titles) of the HIVPR were used in this Western blot. The two lanes under each title represent the two independent biological replicates.

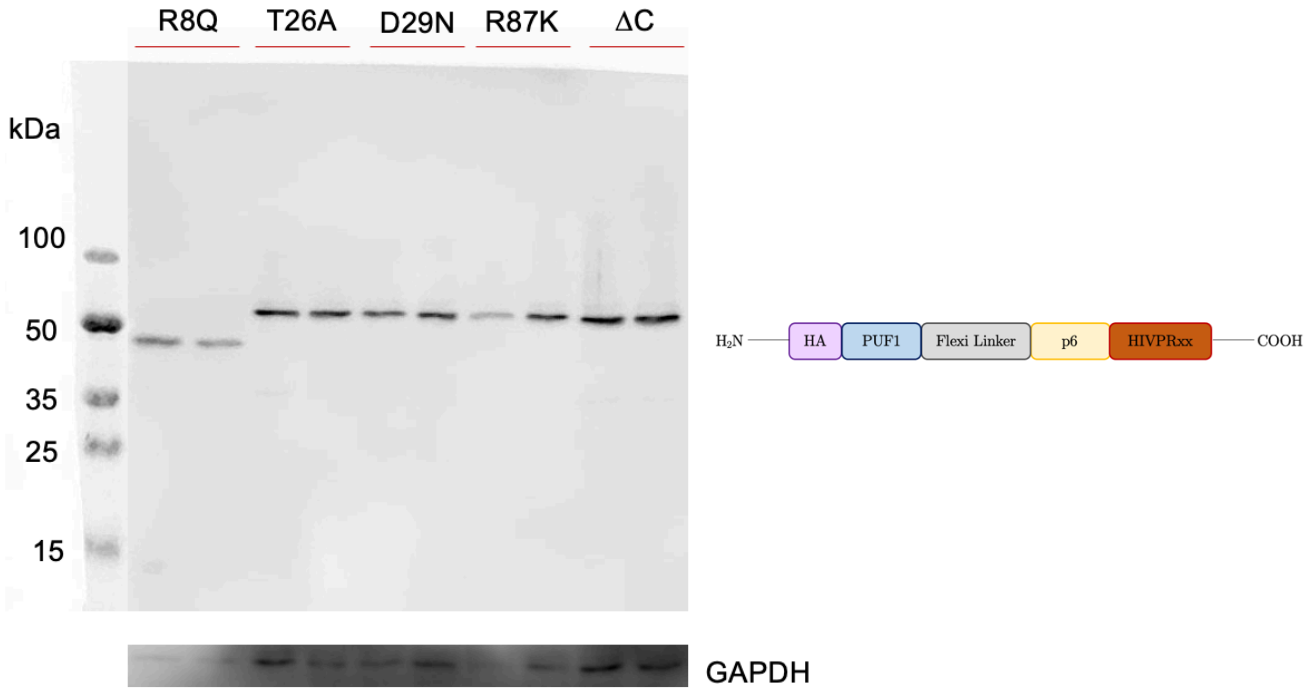


Figure 5.9: Western blot of PUF1-HIVPR variants. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-FLAG monoclonal primary antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Original construct cartoon is shown alongside the blot. HIVPRxx refers to the fact that variants (corresponding to lane titles) of the HIVPR were used in this Western blot. The two lanes under each title represent the two independent biological replicates.

This was initially promising as it suggested spontaneous dimerisation was eliminated for the T26A, D29N, R87K and ΔC variants. Thus, the GST-HIVPR and PUF1-HIVPR variants were also tested (Figures 5.8 and 5.9) to ensure that spontaneous dimerisation was eliminated in the fusion constructs as well, and whether the GST-fused variants could activate (as GST is a known dimerisation domain).

Figure 5.8 shows the GST-HIVPR variants. Compared to the negative control, the same variants – T26A, D29N, R87K and ΔC – show a band at the same molecular weight, whereas in this case, bands are observed for the R8Q variant at a molecular weight approximately 11 kDa lower than the negative control and at the same molecular weight seen for the wild-type GST-HIVPR (~ 30 kDa) in Figures 5.4 and 5.6.

Again, a similar such pattern is observed in Figure 5.9 for the PUF1-HIVPR variants. Although the negative control is not shown on the blot, the T26A, D29N, R87K and ΔC variants show a band that corresponds to its molecular weight, whereas the R8Q variant shows a band at approximately 11 kDa lower, at the same weight of the wild-type PUF1-HIVPR as seen previously. Although this suggests that the variant HIVPR domains cannot activate in proximity, it was later discovered that the GST domain is not an efficient inducer of dimerisation.

5.2.4 ACTIVITY OF PUF-HIVPR VARIANTS IN COMBINATION

Thus, the activity of PUF1-HIVPR variants were tested in combination with the matching PUF2-HIVPR variants using cognate and scrambled RNA to see whether any activity could be observed, potentially demonstrating concurrent binding. Cognate and scrambled RNA sequences were fused as DNA to the 3' end of mCherry DNA so that upon transcription, the PUF binding sites would be present on the mCherry mRNA (Figure 5.10).

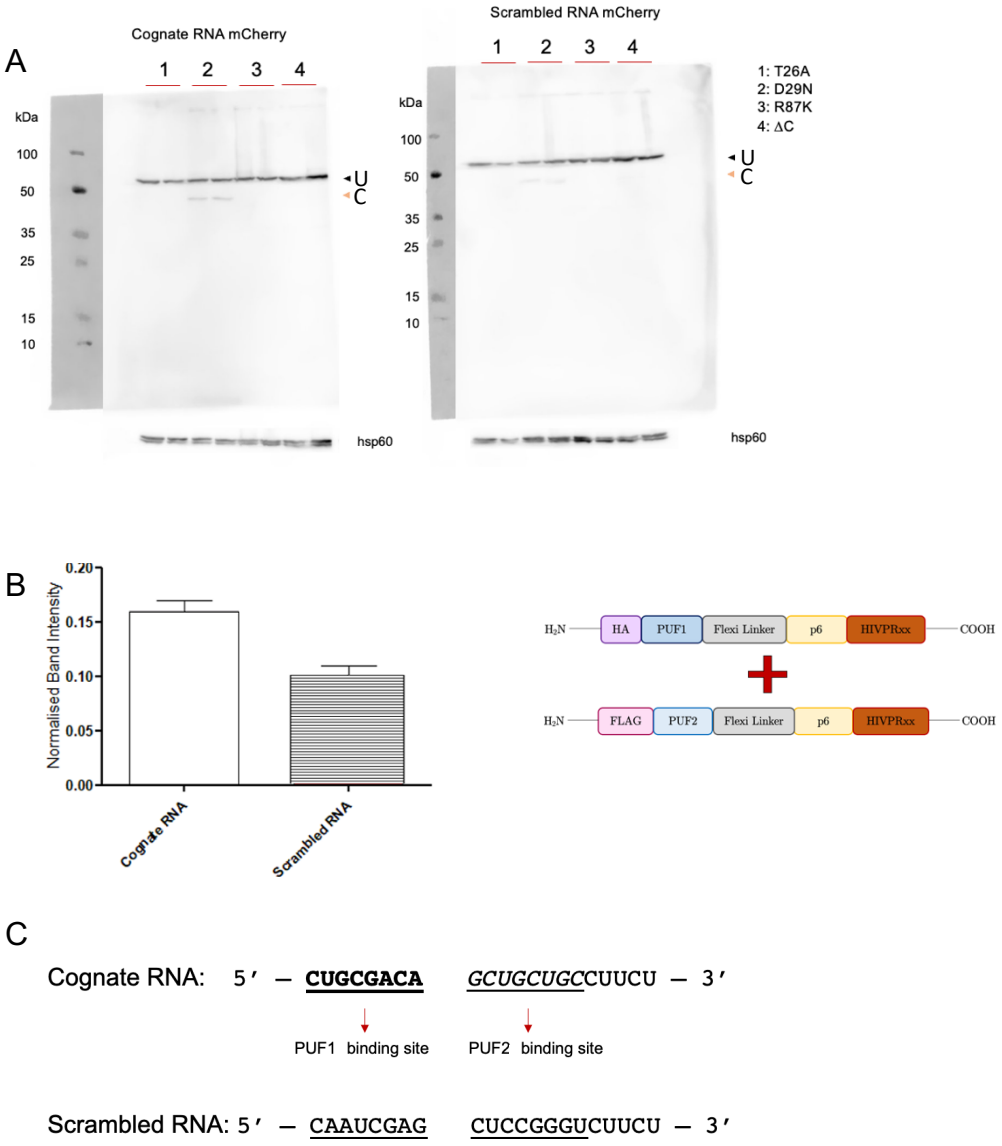


Figure 5.10: A – Western blot of PUF-HIVPR variants probed with an **anti-FLAG primary antibody**. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-FLAG monoclonal primary antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are indicated with a U and C respectively. The two lanes under each title represent the two independent biological replicates. **B** – Quantification of normalised proportion cleaved PUF-HIVPR for the D29N variant. Error bars represent SEM from two independent biological replicates. Original construct cartoon is shown alongside the graph. HIVPRxx refers to the fact that variants (corresponding to lane titles) of the HIVPR were used in this Western blot. **C** – Schematic of the PUF binding sites fused to the 3' end of the mCherry RNA. No spacer was present between the $PUF1_8$ and $PUF\ 2_8$ binding sites. **Bold** – $PUF1_8$ site, *italics* – $PUF\ 2_8$ site, underlined only – scrambled sites.

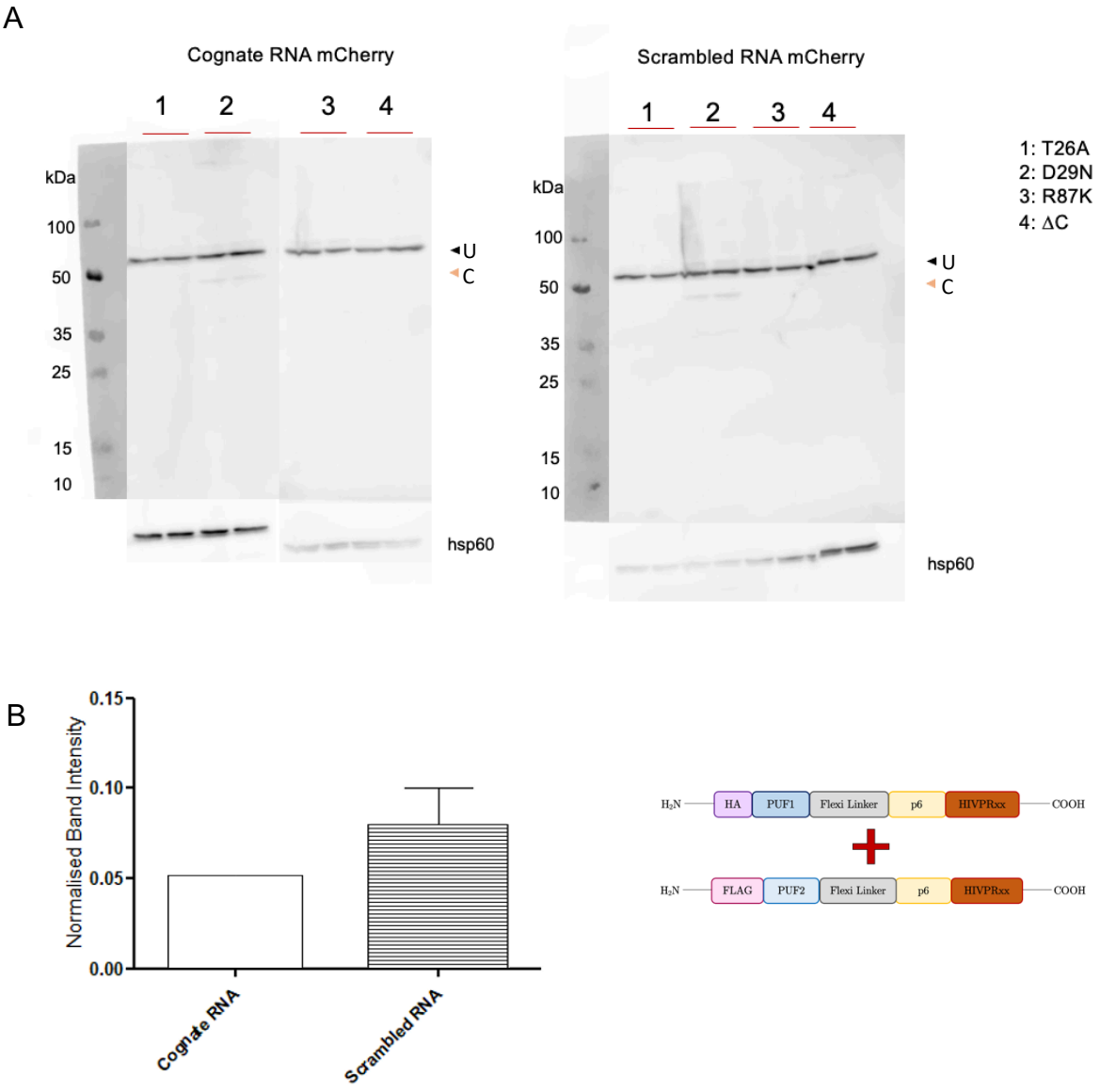


Figure 5.11: A – Western blot of PUF-HIVPR variants probed with an **anti-HA primary antibody**. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-FLAG monoclonal primary antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are indicated with a U and C respectively. The two lanes under each title represent the two independent biological replicates. **B** – Quantification of normalised proportion cleaved of PUF-HIVPR for the D29N variant. Error bars represent SEM from two independent biological replicates. Original construct cartoon is shown alongside the graph. HIVPRxx refers to the fact that variants (corresponding to lane titles) of the HIVPR were used in this Western blot.

PUF1-HIVPR and PUF2-HIVPR variant (T26A, D29N, R87K and ΔC) co-transfection Western blots were probed with either an anti-FLAG primary antibody to look for PUF2-HIVPR variants (Figure 5.10), or an anti-HA primary antibody to look for PUF1-HIVPR variants (Figure 5.11).

Figure 5.10A shows that upon co-transfection with either the cognate site mCherry or the scrambled site mCherry, only the D29N variant of PUF2-HIVPR showed a band around 11 kDa below the expected molecular weight of the negative control (at ~ 45 kDa compared to ~ 55 kDa). Densitometric quantification of the bands were used to calculate the ratio between the cleaved (lower) and uncleaved (higher) bands for the D29N variants (Figure 5.10B), which allowed for normalization, and it was found that when co-transfected with the cognate mCherry RNA, the cleavage of the PUF1-HIVPR$_{D29N}$ was not significantly higher than when co-transfected with the scrambled mCherry RNA. The same results are observed when using the anti-HA antibody to probe for PUF1-HIVPR (Figure 5.11). Again, the D29N variant is the only one which shows a band corresponding to the cleaved construct, with densitometry revealing that there is no significant difference between the ratio when co-transfected with a cognate or scrambled mCherry. The hsp60 band for these blots looks like a doublet due to the full-length PUF fusion constructs having a similar molecular weight to the hsp60 protein. Several control experiments using the PUF1-HIVPR$_{D29N}$ or PUF2-HIVPR$_{D29N}$ alone with cognate or scrambled RNA were carried out in order to establish the reason for a lack in significant difference observed in Figures 5.10 and 5.11.

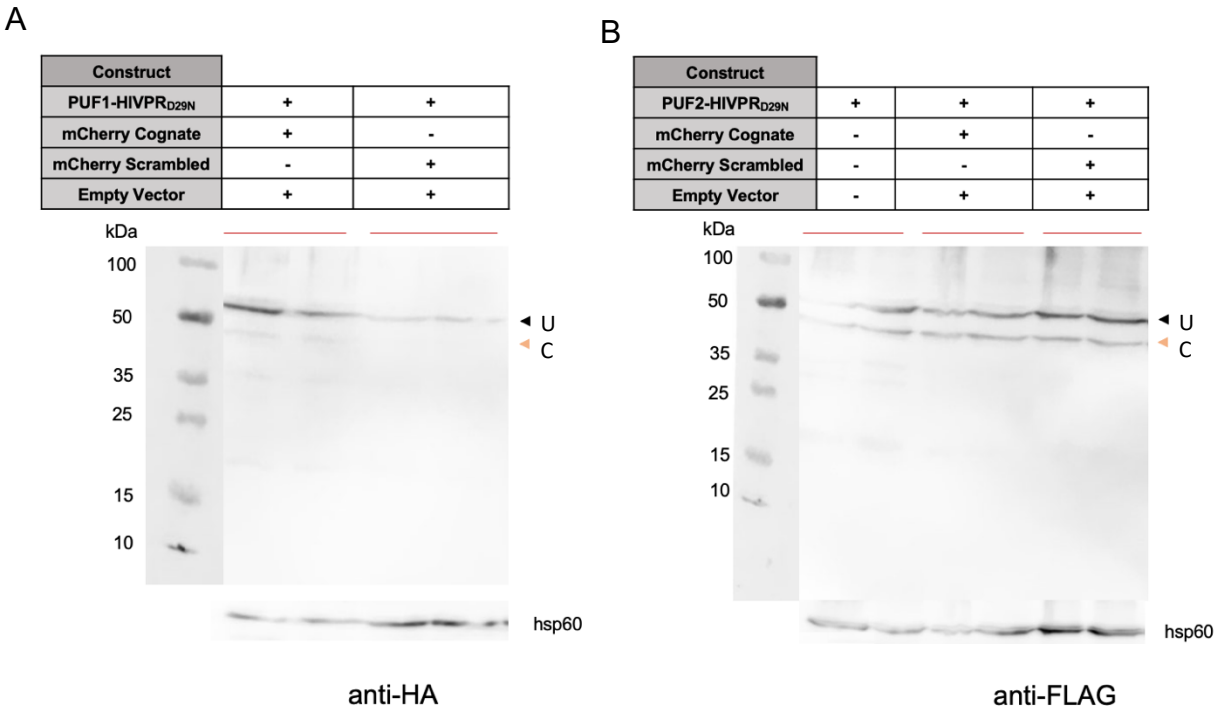


Figure 5.12: A – Western blot of PUF1-HIVPR$_{D29N}$ controls probed with an **anti-HA primary antibody**. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are indicated with a U and C respectively. The two lanes under each title represent the two independent biological replicates. **B** – Western blot of PUF2-HIVPR$_{D29N}$ controls probed with an **anti-FLAG primary antibody**. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The western blot was run according to the methods section with darunavir. The blot was stained using an anti-FLAG monoclonal primary antibody and subsequently with an anti-Mouse HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. The two lanes under each title represent the two independent biological replicates Uncleaved and cleaved bands are indicated with a U and C respectively. The two lanes under each title represent the two independent biological replicates.

Figure 5.12A shows the control experiments probed using the anti-HA antibody. When PUF1-HIVPR$_{D29N}$ alone was co-transfected with scrambled RNA no auto-activation product was seen, however, when this was carried out with the cognate RNA, a faint band at the molecular weight of the auto-activated PUF1-HIVPR$_{D29N}$ construct was seen. Figure 5.12B shows the same controls probed for with an anti-FLAG antibody. In contrast to the PUF1-HIVPR$_{D29N}$, the PUF2-HIVPR$_{D29N}$ showed auto-activation under all control conditions, as seen by the presence of an auto-activation band present in all lanes including when transfected alone.

5.3 DISCUSSION

5.3.1 LIPOFECTAMINE 2000 IN HEK293T CELLS SHOWS BEST TRANSFECTION

Plasmid transient transfection is the first step to good protein expression in mammalian cells. Single plasmid transfection is routinely carried out in our lab and requires minimal optimization. However, the system here requires three plasmids being simultaneously co-transfected into the cell. Three- and even four-plasmid transfection is routinely used in adenovirus or lentivirus production in HEK293 cells[152-154], but it needs to be optimized for the user in order to achieve maximal triple transfection efficiency with minimal effects of cell viability and physiology[155]. Different masses of DNA were tested for transfection efficiency using mCherry expression using 3 plasmids in a 1:1:1 ratio. As expected, as the total mass of DNA transfected was increased, the mCherry expression was increased, however this was also true for cytotoxicity assessed by blebbing of cells under a light microscope. This dose dependency was consistent with previous data from the lab and enabled me to select 1000 ng of total DNA mass for future triple transfection efficiencies, as it has the highest transfection efficiency with no cell death seen. This quantity can also be scaled up or down depending on the application.

5.3.2 HIV PROTEASE FUSION PROTEINS EXPRESS FULLY CLEAVED

Fusion proteins with an active HIVPR domain ran at a lower molecular weight than those with the inactive (D25N) HIVPR domain (Figures 5.4 and 5.5). This was unexpected, as previous literature has shown that any N-terminal extension on the HIVPR significantly reduces dimerisation of the HIVPR domains which in turn reduces activity[125]. My initial hypothesis was that the HIVPR is non-specifically activating in cells to a very small degree, and the free protease was then auto-activating the remainder of the fusion precursor upon cellular lysis for Western blotting. This was not the case, however, as addition of excess of the HIVPR inhibitor darunavir did not change the results seen in the blot (Figure 5.6).

Going back to the literature, I found that the experiments by Louis *et al.*[125] showing the effects of N-terminal extension on dimerisation K_d were carried out *in vitro*. Data from cellular studies shows that the HIV protease may indeed be active in cells even in the presence of an N-terminus extension[156]. Taken together, this suggests that although the N-terminus extension increases the K_d of the HIVPR domain, the cellular protein concentration is great enough to allow dimerisation to occur. Therefore, in order to further reduce dimerisation and prevent auto-activation even within a cellular environment, I tested the different mutations in the literature that were used to reduce dimerisation *in vitro* and *in cellulo*[111,125].

5.3.3 HIV PROTEASE VARIANTS REDUCE AUTO-ACTIVATION

In vitro studies by Louis *et al.*[125] and Hayashi *et al.*[111] have shown that R8Q, T26A, R87K or ΔC mutations can also significantly impair dimerisation of the HIV protease. Thus, these mutations were generated in combination with the already present N-terminal extension. A comparison between the effects of these mutations *in vitro* (modelled from literature data) and *in cellulo* (experimentally obtained in this project) is shown in Figure 5.13.

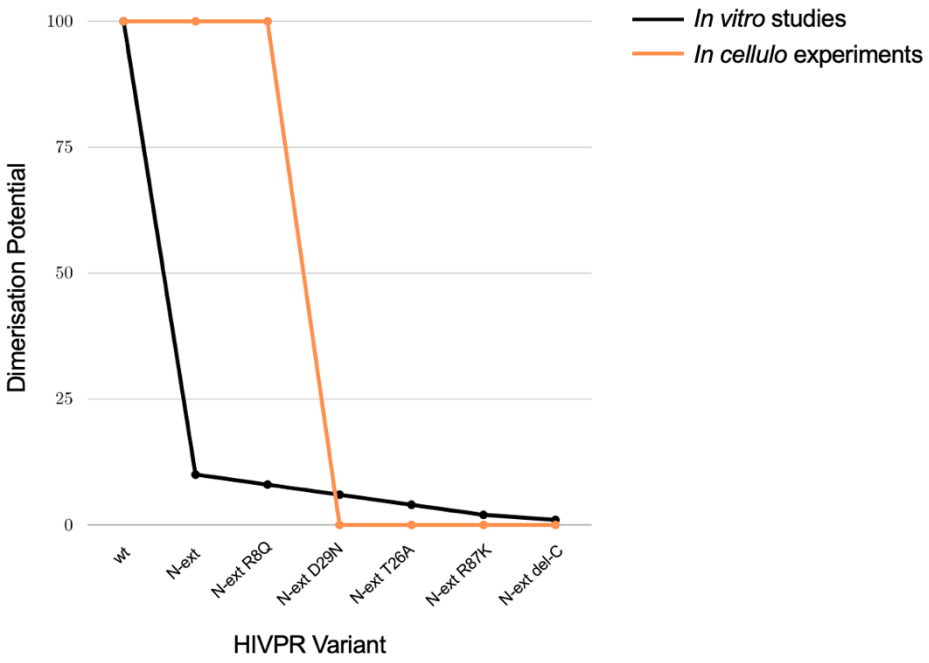


Figure 5.13: Comparison of HIVPR dimerisation potential in vitro and in cellulo. Dimerisation potential is estimated using the *in vitro* dimerisation affinities for the HIVPR variants determined experimentally by Louis *et al.*[125] (black line), and from the experimentally observed cleavage of the different variants *in cellulo* in this chapter (orange line).

The discrepancy between the *in vitro* and *in cellulo* data for the N-terminus extension initially was thought to be a concentration dependent effect. However, this may not entirely be the case as FRET-based dimerisation data of the HIV protease has shown that the D29N, T26A and R87K mutations alone can reduce the dimerisation of the protease to 0%[111]. The dimerisation K_d of the N-terminus extended HIV protease is significantly higher than that of the D29N, T26A and R87K mutations alone[125] which suggests that the N-terminus extended protease should also have 0% dimerisation, yet this is not what was observed in my experiments or in those of Tien *et al.*[156]. What this result suggests is that the dimerisation effects seen in cell-based experiments may not be predictable from *in vitro* data on dimerisation and may be specific to the D29N, T26A and R87K mutants.

Combining the N-terminus extension with the inhibitory mutations from the FRET-based study (D29N, T26A and R87K) indeed resulted in data consistent with what was observed by Hayashi *et al.*[111], which further suggests that the effects on dimerisation observed *in cellulo*

are due to an effect of the particular mutations used, possibly on folding, post-translational modification[157], or synthesis in mammalian cells[158], rather than their effects on K_d. Another possibility is that the D29N, T26A and R87K mutations inhibit the proteolytic activity of the HIV protease, which is why they in particular do not show any activity. In particular, residues 26 and 29 are in proximity to the active site of the protease, and their mutation may affect the catalytic activity.

To exclude this possibility, GST controls were carried out, as GST is a commonly used positive control for dimerisation[159,160]. Thus, if the HIVPR domains are indeed active, but not dimerising, fusing them to GST should result in proteolytic activity in the variants. However, this did not seem to be the case, as the D29N, T26A, R87K and ΔC variants were inactive even when fused to GST (Figure 5.8). Initially, I thought that the activity of the protease was also affected in the variants tested, however after obtaining further data from the split-GFP system used later on, it seemed that GST was not an efficient inducer of dimerisation. Hence, I attempted to test the PUF-HIVPR variants in combination with their cognate and scrambled RNAs to see whether this resulted in any proteolytic activity being detected.

5.3.4 PUF-HIVPR VARIANTS IN COMBINATION

The activity of PUF1-HIVPR$_{D29N}$ and PUF2-HIVPR$_{D29N}$ was found to not be significantly different using cognate RNA compared to the scrambled RNA control (Figures 5.10 and 5.11). In order to assess why this was the case, several controls were carried out. PUF1-HIVPR$_{D29N}$ appeared to activate alone in the presence of its cognate sequence, yet not when transfected with solely the scrambled sequence (Figure 5.12), which suggests that the activity is not simply non-specific binding to any overexpressed RNA in the cell. This result could suggest that the PUF1 RNA-binding domain may be interacting partially with its cognate site – for example, half the site may be bound by one PUF, whereas the other half could be bound by another resulting in enough proximity to allow the HIVPR domains to activate. However, a similar hypothesis was tested by Shinoda *et al.*[108] and showed that a 16-mer PUF RNA binding domain does not exhibit this partial binding effect. To explore the cause of this result, the dynamics of this interaction could potentially be elucidated using NMR to characterize the protein-RNA[161,162] and protein-protein interactions[163] involved.

The PUF2-HIVPR$_{D29N}$ protein, on the other hand, was active under all control conditions tested, including when transfected alone (Figure 5.12). Since we have seen that the D29N variant is not spontaneously active by itself, these results are suggestive of non-specific binding of the PUF2 domain to cellular RNA. Non-specific binding of PUF domains has been observed previously[104] and two potential mechanisms – base 'ejection' and PUF repeat promiscuity – have been discovered through X-ray crystallography. These non-specific interactions of the PUF2$_8$ will be explored further using next generation crosslinked RNA sequencing analysis[164] in Chapter 6.

Thus, I found that there is significant background in this system due to the use of the HIV protease domain. This background skews the signal-to-noise ratio which in turn prevents any true effects of the system from being detected. The primary source of background was found to be due to an individual PUF-HIVPR fusion protein being able to produce proteolytic activity alone, either due to spontaneous interactions of the HIVPR domain, or non-specific proximity enhancement by the PUF domain. In addition, it was found that the free HIVPR$_{D29N}$ does not completely cleave all of the precursor PUF2-HIVPR$_{D29N}$, likely as it cannot dimerise in the free form[111] resulting in incomplete activation and lack of an amplificatory effect. This finding suggests that the cleaved protease may not have the ability to cause apoptosis via protein degradation, which was the envisaged mechanism of action. In order to eliminate these issues, I decided to instead test out split-protein systems[165-167]. In this way, an individual fusion protein cannot activate alone, which should reduce the background and allow any true effects to be detectable. In addition, this system would go on to activate a specific pro-apoptotic zymogen to result in certain cell death unlike in the current version of the system where cell death would have been due to promiscuous protein degradation.

CHAPTER 6 – *IN CELLULO* STUDY OF PUF DOMAINS

6.1 INTRODUCTION

The *in vitro* data from Chapter 3 showed that the PUF proteins can bind RNA individually and concurrently, and next it was essential to ensure that both of these can occur in the cell and thereby drive the dimerisation and activation of their fused domains. Furthermore, in Chapter 5, it was found that the PUF2$_8$-HIVPR$_{D29N}$ being active when transfected by itself. The reasons are due the specificity and the number of endogenous binding sites (which also depends on their specificity) of the proteins. Thus, experiments were carried out to elucidate the binding characteristics of PUF proteins of different lengths within the cell using next-generation sequencing and elucidate how specificity changes with PUF length. This method tests the binding characteristics of PUF domains individually, however their concurrent binding must also be verified *in cellulo*. As mentioned in Chapter 5, this can be carried out using split-protein systems to monitor simultaneous binding which would have less background and non-specific activity than the HIV protease-based system that was tested. Two groups have previously used PUF domains fused to split-GFP and split-mCitrine to localize individual molecules of RNA within the cell[168–170], and thus we selected the split-GFP system to fuse to our PUF domains to assess concurrent binding and activation of the fusion proteins and to optimize the system itself.

6.1.1 METHODS FOR DETECTION OF RNA BINDING SITES

Traditionally, crosslinking immunoprecipitation and next-generation sequencing (CLIP-seq) based methodologies are used to detect RNA binding sites of specific proteins. These techniques generally involve UV-based crosslinking of cellular proteins to RNA, along with subsequent purification of protein-RNA complexes by immunoprecipitation, radiolabelling and SDS-PAGE gel extraction. The protein is degraded using a protease and the subsequent RNA is used to generate cDNA libraries for next-generation sequencing[164]. Several different variations of CLIP have been developed to overcome limitations of the conventional method such as the loss of cDNAs due to loss of the 5' adaptor due to truncation. To this end, techniques such as iCLIP and eCLIP have been developed. iCLIP involves inserting the Illumina barcodes by circularization of resulting cDNAs which only requires addition of the combined adaptor at the 3' end of the RNA, thus preventing loss of truncated cDNAs[171,172]. eCLIP on the other hand involves ligation of adaptors in the same way as iCLIP, initially to the 3' end of the RNA, however no circularization is involved and the 5' adaptor is blunt-end ligated to the resulting cDNA fragment[173].

Due to the immunoprecipitation, only specific protein bound RNAs are recovered, and thus subsequent next-generation sequencing (NGS) and bioinformatic analysis can be used to determine the sequences of these RNAs and thus form a motif that the protein of interest binds to[164,171-173]. Although this method is very specific, due to the use of RNA radio-labelling, X-ray imaging of SDS-PAGE gels and incompatibility with available NGS library preparation kits, an alternate method had to be found to use in this thesis.

6.1.2 ORGANIC ORTHOGONAL PHASE SEPARATION (OOPS)

The Lilley Lab (Dept. of Biochemistry, Cambridge) have developed a novel method of purifying protein-bound RNA called organic orthogonal phase separation (OOPS)[114,115]. Traditional phase separation using acidic guanidine phenol chloroform results in migration of free RNA to the upper aqueous phase and proteins to the lower organic phase. UV crosslinking of RNA and protein results in RNA-protein adducts that share the physicochemical properties of both molecules and are therefore found in the interface of the organic and aqueous phases[114].

Purification of this interface and subsequent protease digestion yields the protein-bound RNA which can then be used in downstream sequencing analysis. Additionally, this method has been applied by the group to detect protein binding sites on RNA by looking for read depletion compared to a non-crosslinked control since crosslinking induces protein adducts on the RNA which cause premature termination of reverse transcription during library preparation resulting in lower read coverage at protein binding sites[174]. However, the issue was that this method would isolate all protein-bound RNA, and therefore it would be impossible to find PUF binding sites compared to a non-crosslinked control. To overcome this challenge, I used a crosslinked control lacking the PUF protein. As such, the PUF library would contain read depletion at PUF binding sites, along with binding sites for other proteins, whereas the control library would only contain read depletion at binding sites for other proteins. Thus, bioinformatically looking for depletions in the PUF library compared to this library would theoretically yield the PUF protein binding sites. Thus, OOPS-seq was selected as an alternative to CLIP-based methods. Furthermore, since I have already validated a binding site for the $PUF2_8$ protein *in vitro*, OOPS-seq could also be validated using the $PUF2_8$ protein.

6.1.3 NEXT-GENERATION SEQUENCING (NGS)

Next-generation sequencing (NGS) is an umbrella term referring to several different forms of high-throughput DNA sequencing. It allows millions of DNA molecules to be sequenced in parallel[175] and therefore has applications when huge numbers of molecules need to be sequenced, which would be inefficient by simple Sanger sequencing. For example, for the experiments in this thesis, millions of different RNA molecules bound to proteins will be purified by OOPS and sequencing each individual resulting cDNA by Sanger sequencing would be impossible and separating the different species of RNA in the pool would also be impossible. Thus, NGS is employed, as it can sequence a large proportion of the generated cDNAs at once and separate the different species as individual 'reads'. The NGS technology used will be Illumina 'Sequence by Synthesis'.

During library preparation, adaptor DNA oligonucleotides that allow annealing of these fragments onto different regions of the sequencing flow cell, and primer binding sites that allow amplification to occur are added to the generated cDNA fragments. Bridge amplification occurs

using the primer sequences and this results in the formation of a cluster of the same cDNA. This process happens in parallel for the millions of cDNA fragments annealed on the flow cell (Figure 6.1). Once this amplification is completed, sequencing takes place by synthesis of the complementary strand (hence the name Sequencing by Synthesis) using fluorescently labelled nucleotides. Thus, nucleotides of the cDNA are detected using the fluorescence generated per cluster for each round of sequencing synthesis yielding the full cDNA sequence for the millions of cDNA molecules annealed on the flow cell. In this way, NGS allows the OOPSed RNAs to be sequenced simultaneously

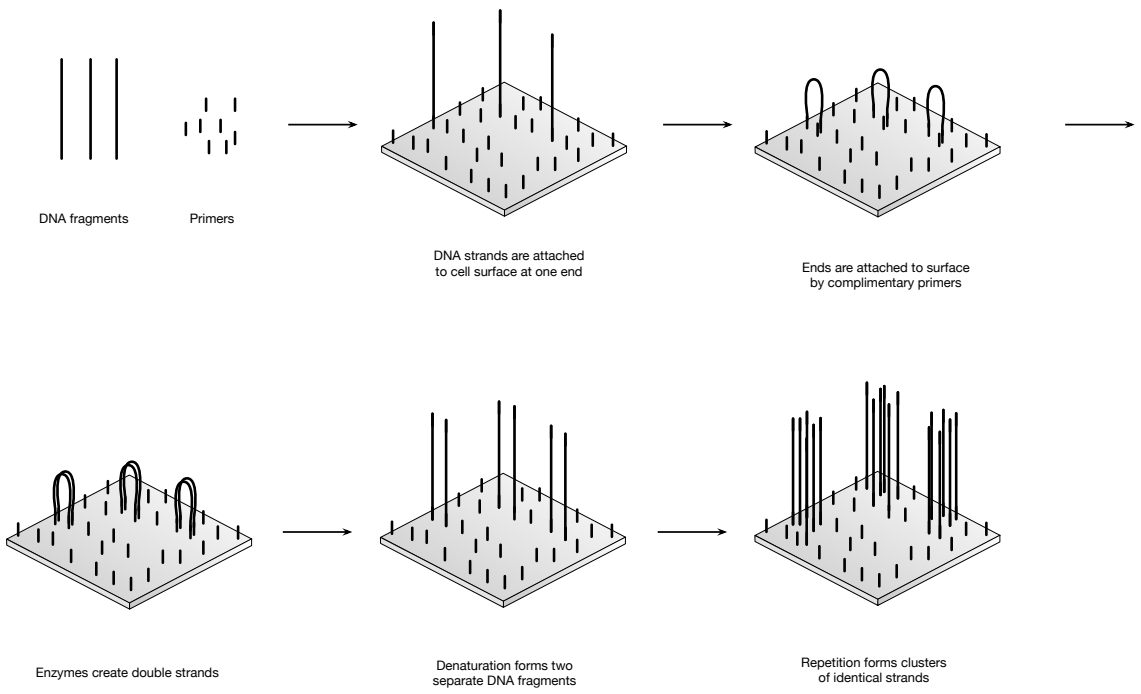


Figure 6.1: Schematic of Illumina flow cell cluster generation for NGS.

6.1.4 GREEN FLUORESCENT PROTEIN

The green fluorescent protein (GFP) is an exceptionally versatile protein used regularly in molecular biology, medicine, and cell biology. Discovered by Shimomura *et al.*[176] in 1962, GFP was found to be useful in several fields such as detecting gene expression[177] and protein localization[178]. It is a cylindrical-shaped protein made of eleven β-strands in a pleated sheet arrangement with five α-helices forming caps on either ends of the cylinder (Figure 6.2). In addition, the immature GFP chromophore with the amino acid sequence Ser-Tyr-Gly is present within the barrel itself and forms an internal α-helix[179]. It is this β-barrel structure that is responsible for the high stability of the GFP protein[180], maturation of the chromophore, and subsequent protection of the chromophore from quenching by water molecules[179].

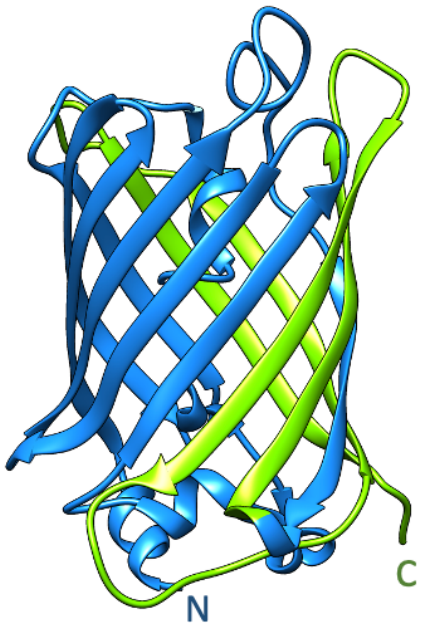


Figure 6.2: Schematic of the structure of GFP (PDB: 4KW4). The N-GFP structure is shown in blue, whereas the C-GFP structure is shown in green. The split point is between amino acids 157 and 158.

The GFP chromophore initially is non-fluorescent and requires complete folding of the GFP protein, producing the appropriate environment within the barrel for auto-cyclization of the chromophore amino acids and subsequent oxidation for chromophore maturation and fluorescence[179]. In agreement with this time-consuming maturation process, GFP fluorescence in cells is only observed approximately 1.5 – 4 hours after protein synthesis.

Using the knowledge that the β-barrel structure is essential for GFP fluorescence, splitting this barrel into two parts was attempted in order to create a split-GFP that could be used to monitor protein-protein interactions by reconstitution of this non-fluorescent split-GFP into a mature, fluorescent GFP molecule. Ghosh *et al.*[181] successfully split the canonical GFP protein between residues 157 and 158 resulting in two fragments that indeed form the β-barrel structure like the full-length GFP protein and can re-constitute when fused to leucine zipper domains that induce dimerisation. The reconstitution of the β-barrel structure then allows chromophore maturation to proceed resulting in fluorescence restoration. This bimolecular fluorescence complementation methodology has been used widely to study protein-protein interactions, as it can detect even weak and transient interactions[182]. Nevertheless, the system does have limitations, the most important of which is the spontaneous non-specific reassembly of the split-GFP halves, and efforts to overcome this problem have led to the use of modified split-GFPs and even trimolecular split-GFPs where three fragments are required for fluorescence recovery[183].

6.1.5 PUF DOMAINS FUSED TO SPLIT-GFP

These limitations, however, were not observed when groups fused the PUF RNA-binding domain to split-GFP in the bimolecular complementation method to localize individual RNAs within the cell. This method was initially used by Ozawa *et al.*[168] to image mitochondrial RNA (mtRNA) in the cell. Their pioneering work showed that the split-GFP halves lacking PUF domains do not show any non-specific restoration of fluorescence, and when transfected into cells expressing the cognate mtRNA there is fluorescence restoration, whereas when the mtRNA is absent there is no restoration. Although this does provide support for the use of PUF fused to split-GFP as a tool to locate RNAs within the cell, their study did not include a control that expressed scrambled mtRNA to monitor non-specific binding of the PUF fusion

proteins (which cannot be determined from their control using no cellular mtRNA). Their next work looked at imaging cytosolic RNA, the β-actin mRNA within the cell[169]. Again, the data from this study did not include the scrambled control, but it did attempt to co-localise re-constituted GFP with *in situ* TAMRA-labelled antisense oligonucleotide probes, which would be indicative of specificity. Interestingly, the images showed background GFP fluorescence that did not co-localise with the mRNA suggesting that there may some non-specific binding of the PUF domains in cells.

Other groups such as the Boyden[103] and Oparka[170] groups have also used PUF domains fused to split-fluorescent proteins to monitor concurrent PUF binding to a single RNA. However, all these studies used different methods of generating the PUF domains that were fused to the split-GFP halves. Ozawa *et al.*[168] and Tilsner *et al.*[170] found regions within their target RNA that had a high RNA sequence homology to the canonical target of the human PUF protein. By doing so, they had to modify very few repeats of the canonical PUF protein and did so using mutagenesis based on results that have now been expanded upon by Abil *et al.*[102] to include a wider PUF repeat recognition code. In contrast, Adamala *et al.*[103] used their own Pumby modules (a type of PUF repeat that can be concatenated and mutated as required to target any RNA sequence) fused to split-GFP.

The PUF domains used in my thesis were designed using the method suggested by Abil *et al.*[102], which have not yet been applied to the split-GFP bimolecular fluorescence complementation system of detecting protein-protein proximity. In addition, optimisation of the PUF-split-GFP binding and reconstitution has also not been carried by any of the aforementioned studies. Thus, it was essential to ensure that the PUF domains used in this thesis can indeed bind concurrently to RNA within the cell as well as *in vitro*, and the split-GFP system appeared to be ideal to explore and optimise this binding.

6.1.6 ASSAY DESIGN, AIMS AND OBJECTIVES

Thus, the first aim of this Chapter was to validate this novel OOPS-seq method as a potential alternative to CLIP-seq techniques and thereby detect binding sites of the 8-, 9- and 10-mer PUF proteins in the cell. The assay design simply involved the transfection of the 8-, 9- or 10-mer PUFs into HEK293T cells which were then subjected to OOPS-seq. A negative control sample was also used for OOPS-seq where an empty vector was used to transfect the cells rather than a PUF protein. The subsequent bioinformatic analysis used to compare the PUF and control libraries were simply a published workflow for the analysis of traditional CLIP data on the Galaxy web server.

The second aim of this Chapter was to ensure that the PUF domains being used in this thesis can also bind concurrently to the same RNA in the cell, as well as to optimise different variables that may be influencing PUF fusion protein binding and reconstitution and assess the specificity of this concurrent binding. To answer these questions, confocal microscopy, fluorescence plate reader assays and flow cytometry were used. Cells must be transfected with the two PUF-split-GFP fusions together with their appropriate cognate or scrambled RNAs. The RNA targets in question were fused to the 3'-end of the mCherry mRNA and co-transfected as plasmids into the cell with the PUF fusions such that expression of the RNA targets could be checked by simple fluorescence microscopy before moving on to plate reader assays and flow cytometry to sort samples based on the presence of RNA in a higher-throughput fashion. A schematic diagram of the assay design is shown in Figure 6.3.

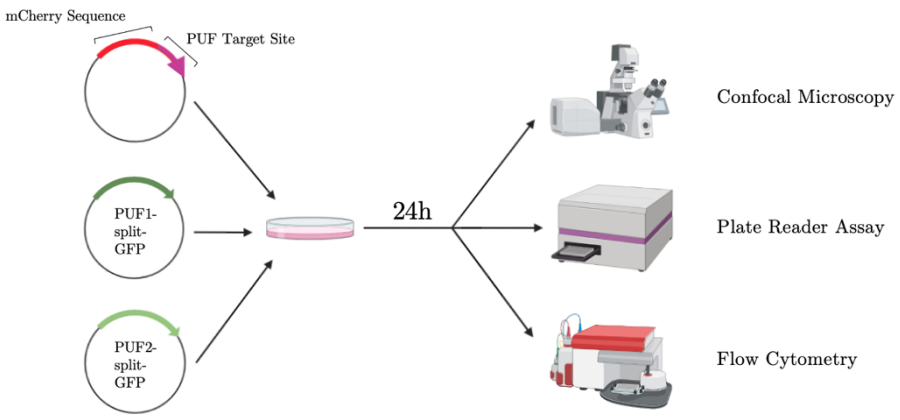


Figure 6.3: Schematic diagram of assay design.

6.2 METHODS AND MATERIALS

6.2.1 ORGANIC ORTHOGONAL PHASE SEPARATION (OOPS)

Cells were seeded in 10-cm dishes (Thermofisher) at a density of 2,000,000 cells per dish and transfected with the appropriate plasmids as mentioned in previous sections. A total DNA mass of 10 micrograms was used per plate. Transfected plates were incubated at 37°C for 24 hours and OOPS was carried out once this period lapsed.

Media was aspirated and cells were washed gently in sterile PBS (Gibco) three times after which all liquids were aspirated from the plate. Cells were UV crosslinked at 254 nm using 400 mJ/cm^2 without the lid. 1 ml of TRIzol reagent (Thermofisher) was added to each plate and cells were scraped and placed in an RNase-free 1.5 ml Eppendorf tube. Each tube was completely homogenized by vortexing vigourously at max speed until no clumps remained. 200 µL (1:5 vol/vol, chloroform:TRIzol) of chloroform was added to each tube and homogenised by vortexing at maximum speed. Tubes were then centrifuged at 12,000 x g for 15 minutes at 4°C. The first organic phase was retained for precipitation to analyse total cellular protein. The aqueous phase was discarded such that only the interface remained. This constitutes one phase separation cycle. This was carried out two more times to remove non-specific protein and RNA in the interface by adding 1 ml of TRIzol to the interfaces and repeating.

The interfaces were cleaned by adding 900 µL of methanol to each sample. Vortexing was used to mix the samples and interfaces were pelleted by centrifugation at 14,000 x g for 10 minutes at 4°C. The supernatant was discarded, and the process was repeated. From this point on, the interfaces were treated either to collect RNA or protein.

PROTEIN COLLECTION FOR WESTERN BLOT

Interfaces were solubilised by adding 100 µL of DEPC-treated water (Thermofisher) and 16 µL of RNase A/T1 mix (Thermofisher) were added to each sample. These were incubated overnight with rotation at 4°C. The following day, SDS-PAGE loading dye was added to each sample, and they were run on a 12% SDS-PAGE gel and blotted as previously mentioned using

an anti-FLAG primary antibody and anti-Mouse secondary antibody conjugated to HRP. Blots were imaged using ECL (Cytiva).

RNA COLLECTION FOR NGS LIBRARY PREPARATION

30 µL Proteinase K (Thermofisher) was added to 300 µL Proteinase K Buffer (10mM Tris-HCl pH 8.0, 10 mM EDTA) per sample and incubated at 50°C for 15 minutes with shaking at 400 r.p.m. before adding 300 µL to each interface. The interfaces were incubated with the Proteinase K mix at 50°C for 2 hours with 400 r.p.m. shaking. 300 µL of ultrapure phenol:chloroform:isoamyl alcohol (25:24:1) was added to each sample (1:1 vol/vol) and vortexed at maximum speed to homogenise. The samples were centrifuged at 12,000 x g for 15 minutes at 4°C. The aqueous phase was then transferred to a separate tube. 50 µL of sodium acetate and 600 µL of isopropanol were added to each aqueous phase (in that order) and incubated on ice for 10 minutes to precipitate the RNA. Centrifugation was carried out at maximum speed for 10 minutes at 4°C to pellet the RNA and supernatant was discarded. 900 µL of 100% ethanol was added to each pellet and centrifuged at maximum speed for 5 minutes at 4°C. The supernatant was discareded and the process was repeated using 70% (vol/vol) ethanol freshly prepared using DEPC-treated water. The pellets were left to air dry for 5 minutes after which they were re-suspended on ice in 100 µL DEPC-treated water for 5 minutes. RNA was quantified using NanoDrop and 1 µg per sample was run on a 1% agarose gel stained with SYBR safe (Thermofisher) for 30 minutes at 90V. Imaging was carried out using a Li-Cor Odyssey (Licor).

6.2.2 METHANOL PROTEIN EXTRACTION

Methanol extraction was used to precipitate proteins from the organic phase of the first round of OOPS phase separation. 150 µL of organic phase per sample was precipitated by adding 1350 µL of 100% methanol. Protein was pelleted at maximum speed for 10 minutes at 4°C and the supernatant was discarded. 1 ml of freshly prepared 80% ethanol (vol/vol) was added to each pellet and centrifuged at maximum speed for 10 minutes at 4°C. The supernatant was discarded, and the pellets were air dried for 2 minutes. The pellets were re-suspended in buffer (50 mM Tris-HCl, 225 mM KCl, 0.1% NP-40, 1% Tween-20 and 1% Triton-X100) with

vigorous vortexing and pipetting up and down. They were left to further solubilise in the buffer overnight at 4°C on a rotator. The samples were then treated for Western blotting in the same way as protein that was collected from the interface.

6.2.3 NGS LIBRARY PREPARATION AND SEQUENCING

rRNA depletion was carried out using the NEBNext rRNA Depletion Kit v2 (Human/Mouse/Rat) (New England BioLabs) and subsequent library preparation was done using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England BioLabs).

rRNA DEPLETION

100 ng of RNA (extracted using OOPS) in 11 µL was added to 2 µL of rRNA depletion solution and 2 µL of probe hybridization buffer on ice and mixed well by pipetting. The reaction was then incubated in a pre-heated thermal cycler using the following settings:

1. 95°C for 2 minutes
2. Ramp down to 22°C at 0.1°C/s
3. 22°C for 5 minutes

The reaction was immediately placed on ice and 2 µL RNase H reaction buffer, 2 µL Thermostable RNase H and 1 µL nuclease-free water was added to it and mixed well. This reaction was incubated in a thermal cycler for 30 minutes at 50°C with the lid set to 55°C.

The reaction was then immediately placed on ice and 5 µL DNase I reaction buffer, 2.5 µL DNase I (RNase-free) and 22.5 µL nuclease-free water were added to it and mixed well. The reaction was then incubated at 37°C for 30 minutes with the lid off.

rRNA depleted RNA was purified using NEBNext RNA Sample Purification Beads (New England BioLabs). 90 µL (1.8x) beads were added to the depleted RNA sample, mixed, and incubated on ice for 15 minutes. The beads were pelleted using a magnetic rack and supernatant was discarded. 200 µL freshly prepared 80% ethanol was added to the tube while

in the magnetic rack, incubated for 30 seconds at room temperature and then discarded without disturbing the beads. This step was repeated for a total of 2 washes. All the ethanol was removed using a pipette and the beads were air-dried for 5 minutes. 7 µL of nuclease-free water was added to the beads and incubated at room temperature for 2 minutes. The beads were then pelleted in the rack and 5 µL of the rRNA depleted, purified RNA was transferred to a separate nuclease-free tube.

LIBRARY SYNTHESIS

5 µL of the rRNA depleted RNA was placed on ice. 1 µL of random primers was added to the sample and mixed well. The reaction was ncubated in a pre-heated thermocycler for 5 minutes at 65°C with the lid set to 105°C. The temperature was ramped down at maximum speed to 4°C and the sample was immediately placed on ice.

8 µL of nuclease-free water, 4 µL First Strand Synthesis buffer and 2 µL First Strand Synthesis Enzyme Mix were added to the sample and mixed well. The reaction was incubated in a thermocycler with the heated lid set to 105°C with the following cycle settings:

1. 10 minutes at 25°C
2. 15 minutes at 42°C
3. 15 minutes at 70°C
4. Hold at 4°C

The reaction was immediately placed on ice. 8 µL Second Strand Synthesis buffer, 4 µL of Second Strand Synthesis Enzyme Mix and 48 µL nuclease-free water were added to the reaction and mixed thoroughly. The reaction was incubated at 16°C for 1 hour with the heated lid off.

The cDNA was purified using SPRISelect Beads (Beckman Coulter). 144 µL (1.8x) beads were added to the reaction, mixed well, and incubated at room temperature for 5 minutes. The beads were pelleted using a magnetic rack and the supernatant was discarded. Two 80% ethanol washes were carried out and beads were dried for 5 minutes (as was done for the rRNA depletion). 53 µL of 0.1x TE buffer was added to the beads, mixed, and incubated for 2 minutes

at room temperature. Beads were pelleted using the rack and 50 µL of the purified cDNA mixture was transferred to a new nuclease-free PCR tube.

The mix was placed on ice and 7 µL of NEBNext Ultra II End Prep reaction buffer, and 3 µL of NEBNext Ultra II End Prep Enzyme mix were added to it. The reaction was mixed well and incubated in a thermocycler with the lid set to 105°C with the following cycling settings:

1. 30 minutes at 20°C
2. 30 minutes at 65°C
3. Hold at 4°C

NEBNext Adaptor was diluted 25-fold in Adaptor Dilution buffer and 2.5 µL of this was added to the end prepped cDNA on ice. 1 µL of ligation enhancer and 30 µL of Ligation Master mix were also added to the reaction and mixed well by pipetting. This mix was incubated for 15 minutes at 20°C with the heated lid off. 3 µL of USER enzyme was added to the mix and it was further incubated for 15 minutes at 37°C with the heated lid set to 105°C.

The ligation reaction was purified using SPRISelect beads. 87 µL beads were added to the sample, mixed well, and incubated for 10 minutes at room temperature. The beads were pelleted using a magnetic rack and the supernatant was discarded. They were washed with 80% ethanol twice and dried as mentioned previously. 17 µL 0.1x TE buffer was then added, mixed, and incubated for 2 minutes at room temperature. Beads were pelleted using the rack and 15 µL of supernatant was transferred to a new PCR tube.

25 µL of Ultra II Q5 Master mix, 5 µL of universal primer and 5 µL of the appropriate index primer from the NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) (New England BioLabs) was added to the adaptor ligated cDNA and mixed. PCR amplification was carried out using the following settings:

1. 98°C for 30 seconds
2. 98°C for 10 seconds, and 65°C for 75 seconds (13 cycles)
3. 65°C for 5 minutes
4. Hold at 4°C

PCR amplified DNA was purified using SPRISelect beads. 45 µL of beads were added to the reaction, mixed, and incubated for 5 minutes at room temperature. The supernatant was discarded and beads were washed and dried with 80% ethanol as previously mentioned. 23 µL 0.1x TE was added to the beads, mixed, and incubated for 2 minutes at room temperature. 20 µL of the purified library was transferred to a fresh nuclease-free PCR tube and stored at -20°C till required.

Libraries were quality controlled using the Agilent TapeStation System (Agilent) and the DNA High Sensitivity Chip according to the manufacturer's instructions. Multiplex libraries were pooled in an equimolar fashion to a final concentration of 4 nM in 20 µL of nuclease free water. Pooled libraries were then sequenced using the NextSeq 500 Sequencing System (Illumina) for 75 cycles in single-end mode.

6.2.4 BIOINFORMATIC ANALYSIS PIPELINES

All bioinformatic analyses were conducted using the worldwide or EU Galaxy servers and the general pipeline used was also from a Galaxy tutorial. Prior to analysis, FastQC was carried out using default settings on Galaxy. FastQC is a quality control program that checks sequencing files for metrics such as base calling quality, sequence length distribution and GC content to ensure sequences are of appropriate quality for downstream analysis. Subsequently, reads were mapped to the hg38 genome using RNA STAR using default settings except without soft clipping of reads. Peak calling was then carried out using PEAKachu on the EU Galaxy server with default settings except:

1. Maximum insert size = 200
2. Pairwise replicates = False
3. Adaptive Mode
4. Mad Multiplier = 0.0

For depletion analysis, controls were provided as "experimental libraries" and experiments were provided as "control libraries" and vice-versa for enrichment analysis.

Subsequently, peak regions were extended at either end by 20 bp using bedtools SlopBed and chromosome sizes of hg38, and the subsequent sequences were extracted using bedtools GetFastaBed using hg38. Finally, MEME analysis was carried out using MEME-ChIP with default settings, except the following:-

1. Limit of sequences to pass to MEME: none
2. Subsampling = random, seed = 123
3. Search given strand only = true
4. Expected motif site distribution = any number of repetitions
5. Minimum motif width = 5
6. Maximum motif width = 20

6.2.5 CONFOCAL MICROSCOPY

Confocal microscopy was carried out using the Zeiss Axiovert 100M Confocal Microscope (Zeiss). Images were taken using the 40x objective lens in Type F Immersion Liquid (Leica Microsystems) using appropriate detection filters for GFP. The argon laser power was set to 50% and the 488 nm laser attenuation was set to 15.9%. A pinhole width of 1000 µM was used and a gain of 1000 volts was used to amplify the signal. The offset was set to -0.5%. Images were taken in a 1024 X 1024 format with a pixel time of 12.8 µs and a field line average of 4.

6.2.6 PLATE-READER ASSAY

Cell seeding and transfection was scaled down to black 96-well flat clear-bottom plates (Greiner). On the day of analysis, media was aspirated, and cells were washed in sterile PBS. The plate was immediately analysed using a Clariostar Plate Reader (BMG). The excitation and emission wavelengths along with bandwidths used were 561-15 nm and 610-20 nm for mCherry respectively, and 483-14 nm and 530-30 nm for split-GFP respectively. Focal height was adjusted automatically for the wells using the mCherry emission as a reference for both the mCherry and split-GFP emission channels. Gain was adjusted automatically for the mCherry emission channel. For the split-GFP channel, various gains were tested, and a gain of 2850 was used for final experiments.

6.2.7 FLOW CYTOMETRY

Cell seeding and transfection were carried out in 24-well plates and incubated for 24 hours (as previously described). On the day of analysis, cells were washed in PBS and detached from wells using 50 µL 1x Trypsin-EDTA and trypsin was neutralised by the addition of 100 µL fresh culture medium. Cell suspensions were transferred to 96-well plates and then pelleted at 500 x g for 5 minutes at 4°C. The supernatant was discarded, and cell pellets were re-suspended in FACS buffer (PBS + 2% FBS). This washing process was repeated, and final cell pellets were re-suspended in 200 µL FACS buffer.

The Attune™ NxT Acoustic Focusing Flow Cytometer was the instrument used for these experiments. Voltages for forward- and side-scatter, and the appropriate lasers were set for mCherry and GFP, and compensation was carried out using single colour controls. Data acquisition was then carried out using the 96-well plate autosampler at a flow rate of 100 µL/min.

6.3 RESULTS

6.3.1 OOPS CAN PURIFY RNA-BOUND PUF PROTEINS

OOPS coupled to Western blotting was used to assess whether the protocol could be used with the PUF proteins. Three lengths of PUFs were generated, i.e., 8-mer, 9-mer and 10-mer PUFs which bind 8-, 9- and 10- nucleotide target sites respectively. These will be referred to as $PUF2_{8,}$ $PUF2_{9}$ and $PUF2_{10}$ throughout the thesis. To check whether RNA-bound PUFs could be purified, the organic phase and interfaces of crosslinked and non-crosslinked conditions were probed for PUF presence.

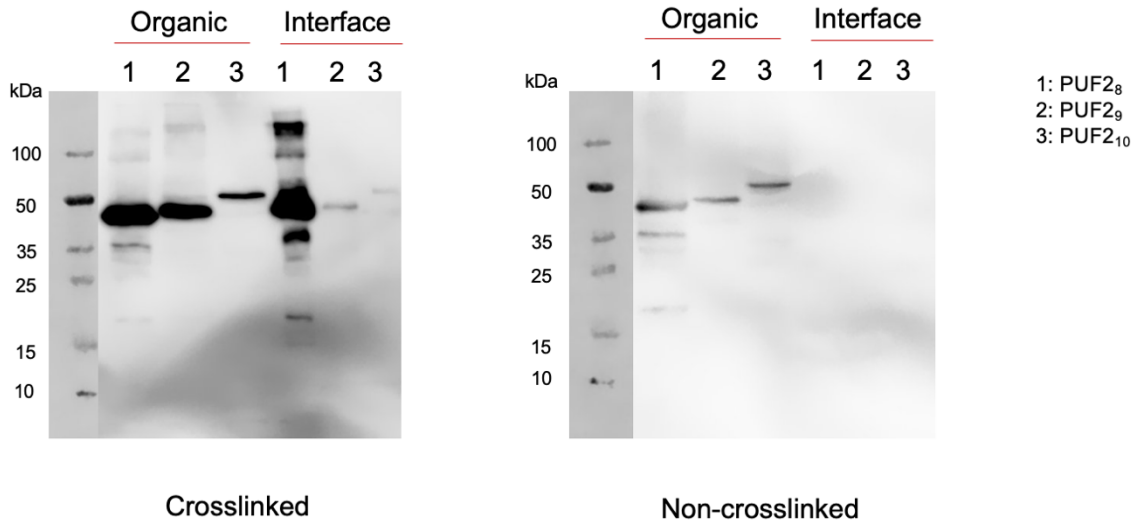


Figure 6.4: Western blot of OOPSed cellular lysates transfected with either $PUF2_{8}$, $PUF2_{9}$ or $PUF2_{10}$. OOPS and Western blot were carried out according to the methods section using an anti-FLAG primary antibody and anti-Mouse secondary antibody. The blots are representative of three independent biological replicates.

The organic phase extraction in both the crosslinked condition and non-crosslinked control shows the presence of all three PUF proteins tested at the expected molecular weights. The crosslinked experiment shows higher molecular weight bands in the organic phase for the $PUF2_{8}$ and $PUF2_{9}$ proteins which are not seen in the non-crosslinked control.

The interface in the crosslinked experiment shows the presence of bands representing all three PUF proteins. The band densities of $PUF2_9$ and $PUF2_{10}$ in the interface are lower than within the organic phase. An interesting observation is that $PUF2_8$ shows purification of the higher and lower molecular weights bands in the interface as compared to the organic phase corresponding to RNA-bound PUF protein which was not completed digested by the RNase treatment. No bands are seen for either of the three proteins in the non-crosslinked control interface suggesting only RNA-bound PUF is being isolated upon crosslinking and that subsequent analyses are valid and that the RNA from crosslinked PUF proteins can be isolated.

.

6.3.2 LIBRARY PREPARATION AND QUALITY CONTROL

Figure 6.5 shows the gel electrophoresis images of RNA from the extractions using OOPS. The All the samples show heavy smearing in the 500 – 1000 bp region. Some bands are visible within the smears of the $P2_{10}$ and negative control samples, however, these bands are less clear in the $PUF2_8$ and $PUF2_9$ samples (Figure 6.5). Although this suggests RNA degradation which may affect library preparation[184], the kit being used is known to handle degraded RNA well. Thus, NGS library preparation was carried out on these RNA samples, and the libraries generated were subject to Tapestation analysis to verify the quality of the libraries.

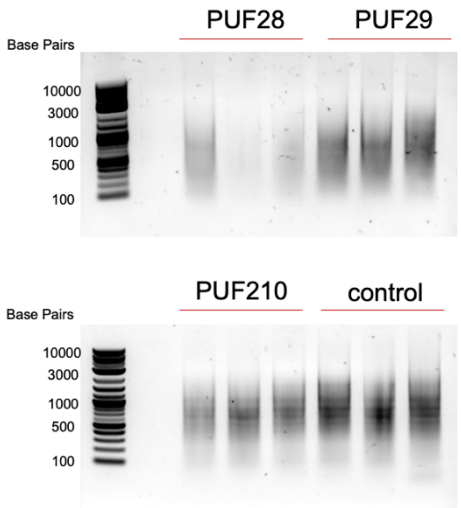


Figure 6.5: RNA gel electrophoresis of samples obtained from OOPS. Samples were run on a 1% TAE gel stained with SYBR safe at 75V till the dye reached three-quarters down the gel. The three lanes under each heading represents the three biological replicates for each sample.

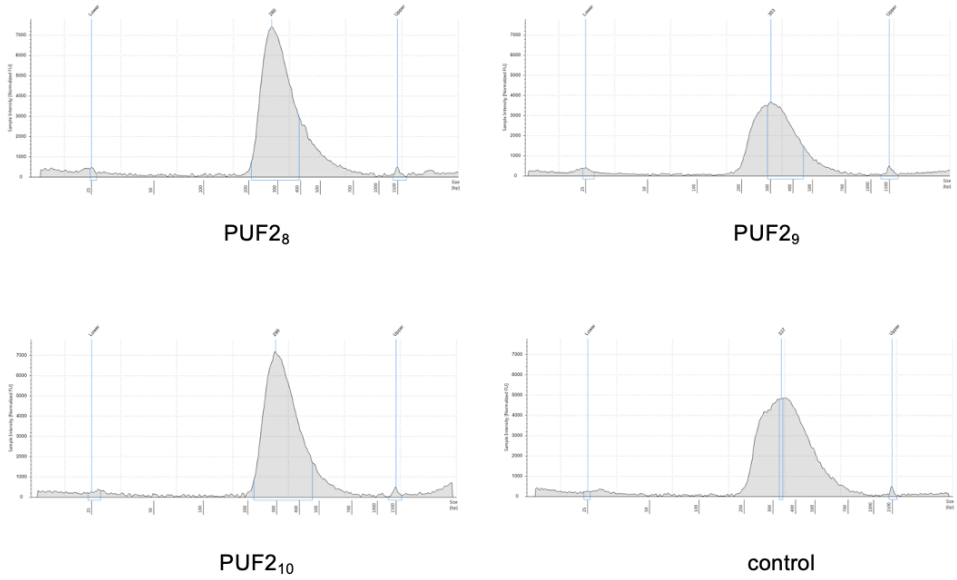
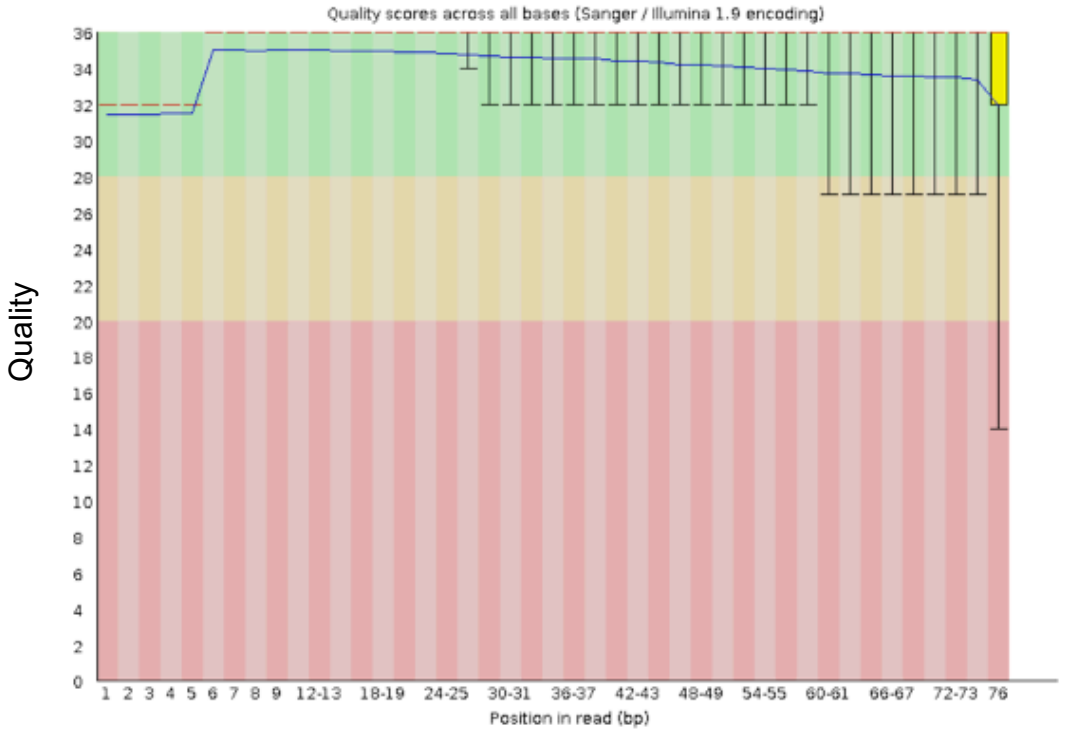


Figure 6.6: Tapestation fluorescence intensity traces using high sensitivity DNA chip of NGS library samples. Upper and lower ladders are marked on the trace along with the average size of the library. Traces are representative of three independent biological replicates.

The Tapestation quality control for the library samples was used to assess four main factors: 1) remaining index primers that would appear as peaks of less than 85 nucleotides, 2) presence of adaptor-dimers which would be found as a peak at ~ 127 bp, 3) PCR overamplification seen as a peak at a higher molecular weight than the library size ~ 1000 bp and 4) a broad library size distribution. Figure 6.6 shows the Tapestation traces for the libraries generated and none of these traces show any of the aforementioned quality issues suggesting the libraries were ready for NGS. The quality of the sequencing reaction was then assessed using FastQC to determine the average quality score per base in each read.

High quality scores indicate good base calling in the sequencing reaction, which corresponds to the likelihood that a called base is indeed that base. Figures 6.7 and 6.8 show that the sequence quality at all positions in the read are within the "green" region which denotes very good quality base calls which suggests that the libraries have very high read qualities. Thus, subsequent mapping, peak calling, and motif analysis can be carried out robustly without being affected by poor quality reads. No other quality issues were detected using FastQC using the other metrics tested.



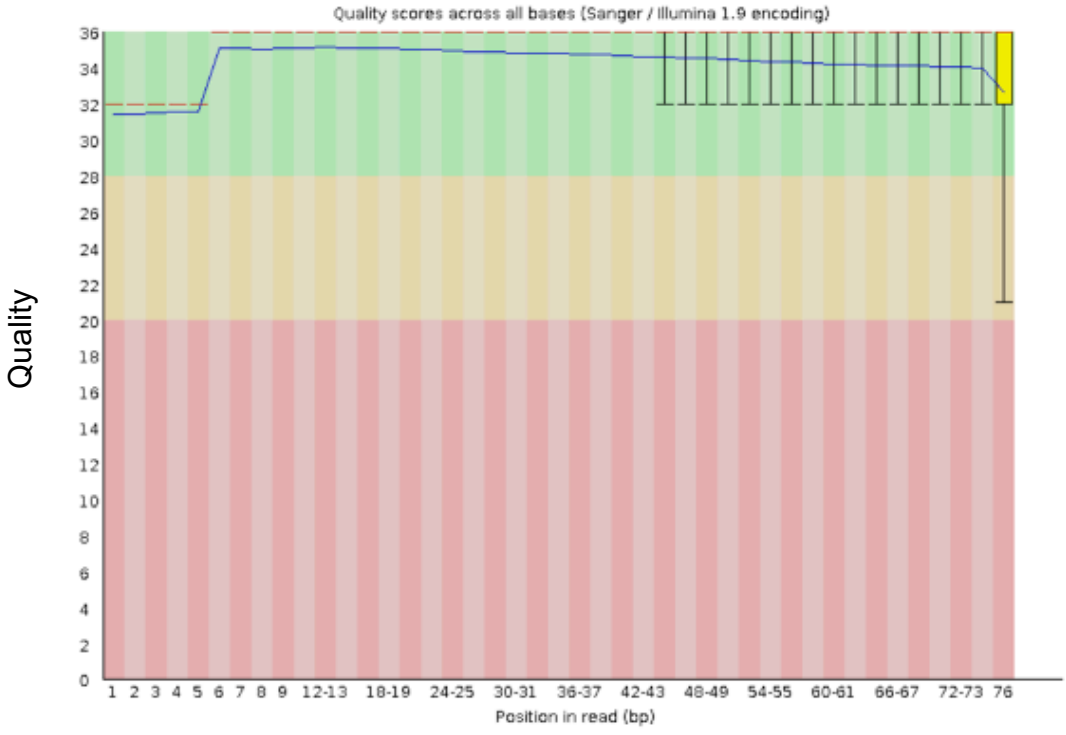
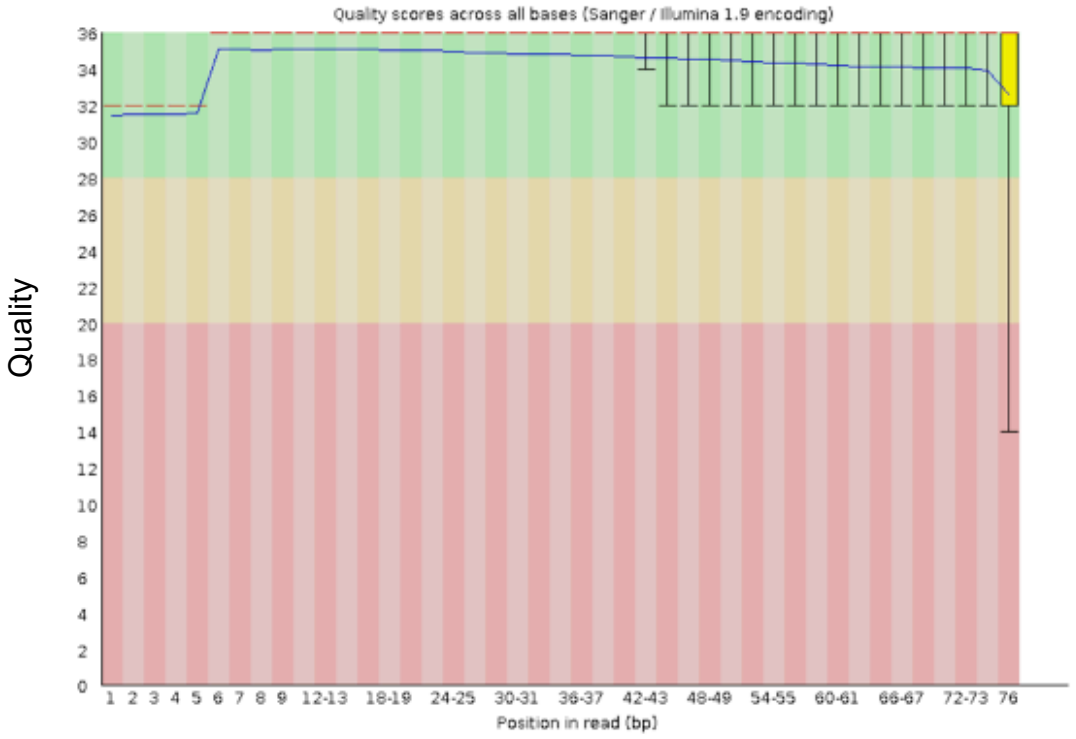


Figure 6.7: FastQC analysis of base calling quality score for each read in the sequencing reaction for PUF2$_8$ and PUF2$_9$. Analyses are representative of three independent biological replicates.



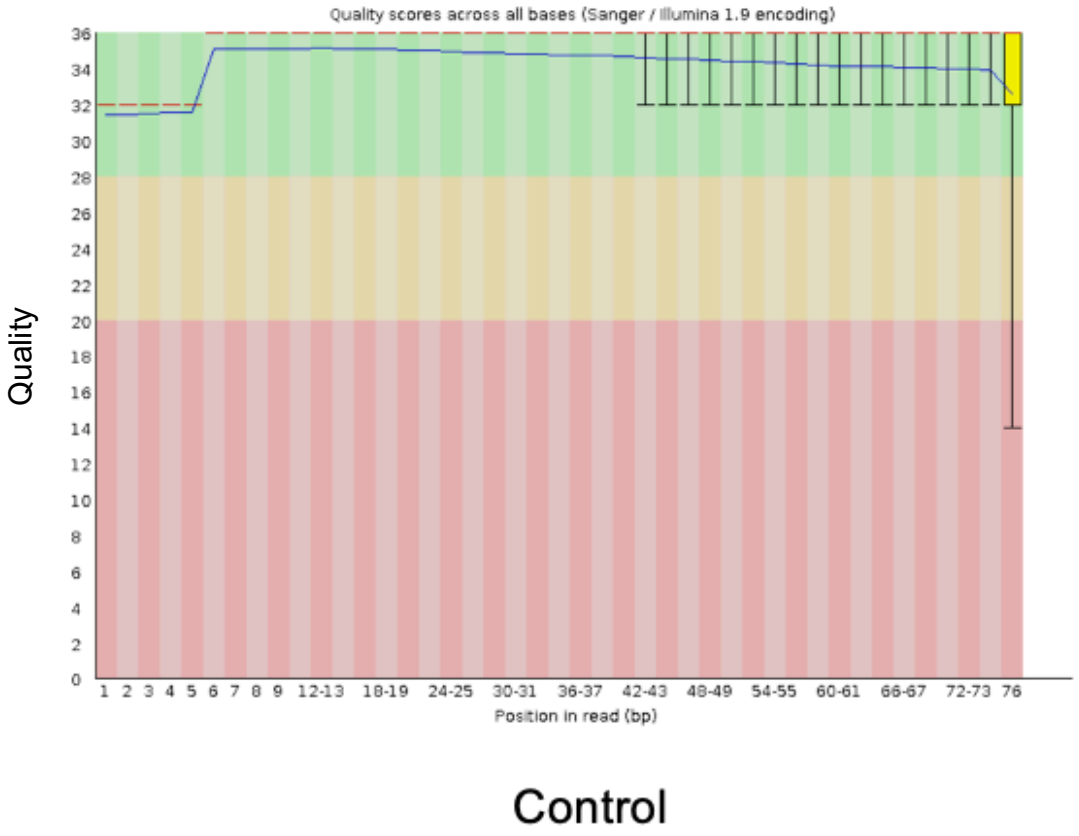


Figure 6.8: FastQC analysis of base calling quality score for each read in the sequencing reaction for $PUF2_{10}$ and the control. Analyses are representative of three independent biological replicates.

6.3.3 READ DEPLETION ANALYSIS

Once the sequencing quality was verified, read mapping was carried out to the human genome build 38 (hg38) using the RNA STAR aligner which accounts for RNA splicing. Subsequently, peak calling was carried out to detect peaks in the control samples compared to the PUF samples as this would essentially denote regions in the hg38 genome that are depleted in the PUF samples. These significantly depleted regions can then be subject to motif analysis in order to find the actual sequences of the depleted regions and identify PUF binding sites. Indeed, peak calling using PEAKachu detected several peaks that were enriched in the control samples compared to the $PUF2_8$, $PUF2_9$ and $PUF2_{10}$ samples.

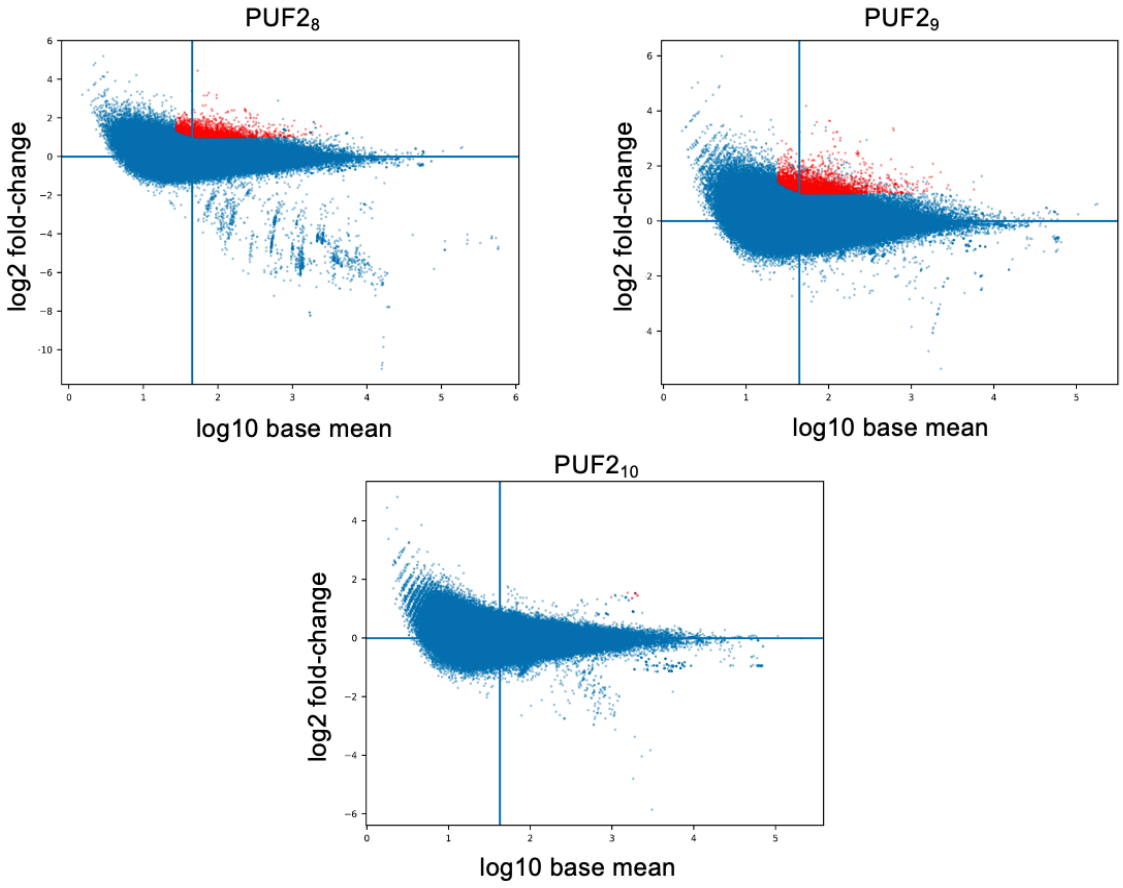


Figure 6.9: MA plots of PEAKachu peak depletion analysis. Red points represent significantly enriched peaks with an adjusted p-value < 0.05 as determined by DESEQ2 using the Benjamini-Hochberg correction.

The MA plots in Figure 6.9 plot the average (between control and experimental libraries) normalized read count of a particular region against the fold change in normalized reads of that region (between control and experimental libraries). This plot is used to visualize significantly enriched peaks in the experimental library compared to the control as shown in red that are present in all three biological replicates. In this version of the analysis, the significant peaks in red represent depleted peaks in the PUF libraries. The different PUF proteins have similar fold changes (between 1.5 and 4.5) in their depleted regions compared to the control and, interestingly, there are several peaks in the MA plots for all three proteins that show negative fold changes suggesting they are enriched in the PUF libraries. Surprisingly, most of the significant peaks are clustered within the region of peaks that are not significantly enriched, and have low log-2 fold-changes, suggesting they may be artefactual. The significantly depleted peaks are then used for motif analysis by MEME (multiple EM for motif elicitation) to determine the sequences of the significantly depleted regions.

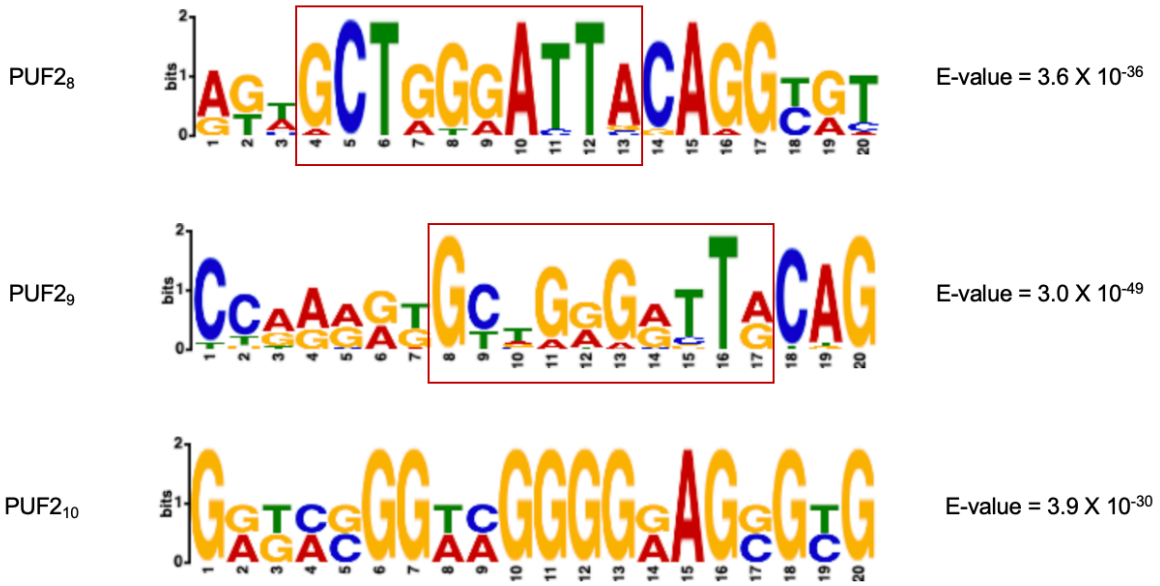


Figure 6.10: Motifs obtained from MEME analysis of depletion peaks. Red box represents repeated motif partially matching designed target sites for corresponding PUFs. Significance of the discovered motif is represented by the E-value shown in the Figure.

None of the exact target sites for the PUF proteins were obtained in the MEME analysis, however partial sites were seen. PUF2$_8$ and PUF2$_9$ had a common motif detected – GCTGGGATTA (Figure 6.10) – which partially matches their cognate sites: GCTGCTGC (PUF2$_8$) and GCTAGCTGC (PUF2$_9$). The bit depths of individual nucleotides at the different

positions of these sites were lower in $PUF2_9$ than $PUF2_8$. Bit depths refer to the height of the nucleotides at any given position of the motif and represent the probability of a particular nucleotide being present at that position in the motif. In contrast, no target sites partially matching that of $PUF2_{10}$ were detected in the analysis, but GC-rich regions were enriched. This suggests that the analysis using the depletion methodology has several limitations, and that the motifs detected may be artefacts. Thus, a complementary experiment was carried out to detect regions that were enriched in the PUF protein libraries compared to the control libraries to find RNAs that were specifically enriched due to PUF binding.

6.3.4 READ ENRICHMENT ANALYSIS

Similar MA plots were generated as for the read depletion analysis. Peak calling in this case was used to isolate peaks enriched in the PUF samples corresponding to RNAs that have been enriched due to PUF binding.

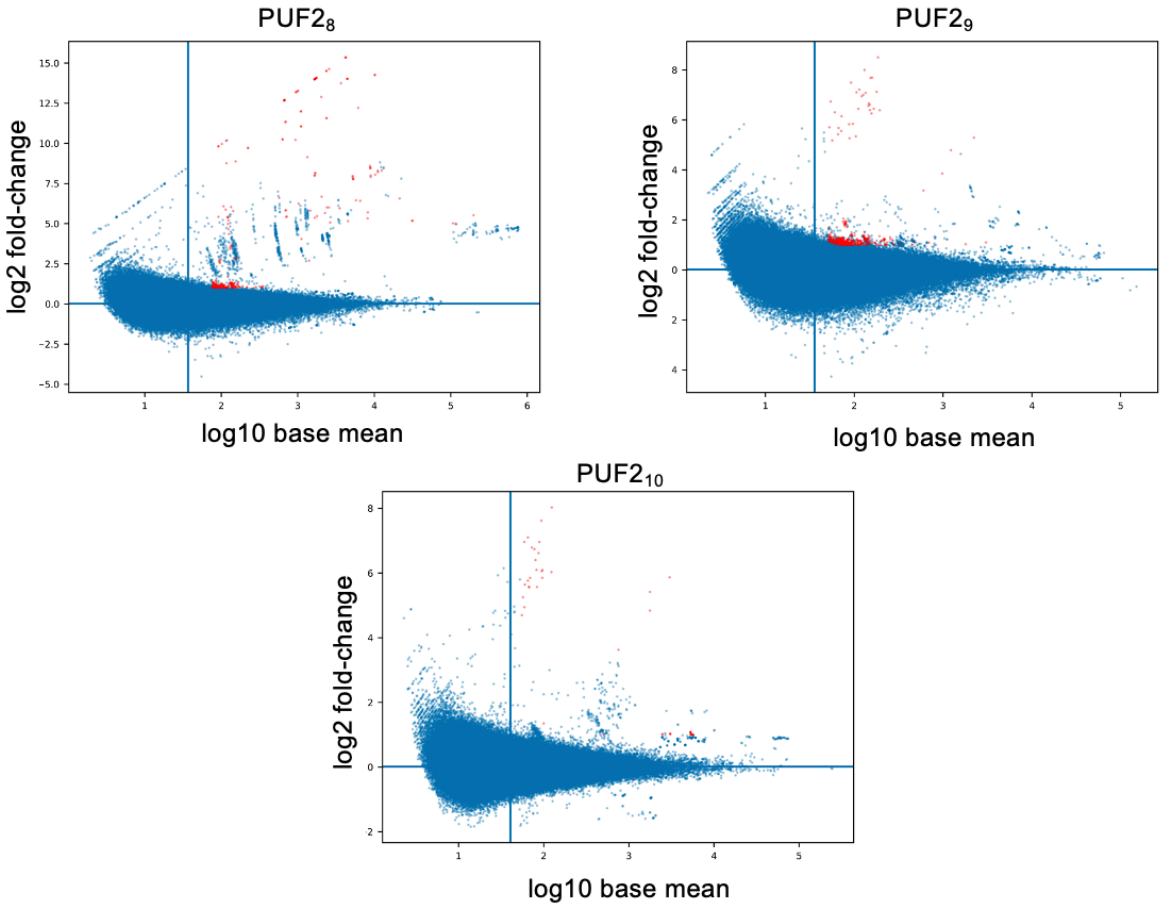


Figure 6.11: MA plots of PEAKachu peak enrichment analysis. Red points represent significantly enriched peaks with an adjusted p-value < 0.05 as determined by DESEQ2 using the Benjamini-Hochberg correction.

All PUF samples have significantly enriched peaks compared to the control libraries (Figure 6.11). PUF2$_8$ has peaks with higher fold changes and mean read coverages than the PUF2$_9$ and PUF2$_{10}$ samples which tend to have lower fold changes and read coverages. Compared to the read depletion analysis, these plots are more promising. The significantly enriched peaks here are not clustered near the non-significant peaks and have much higher log-2 fold-changes suggesting that they represent true peaks. These significantly enriched peaks were then used for motif analysis.

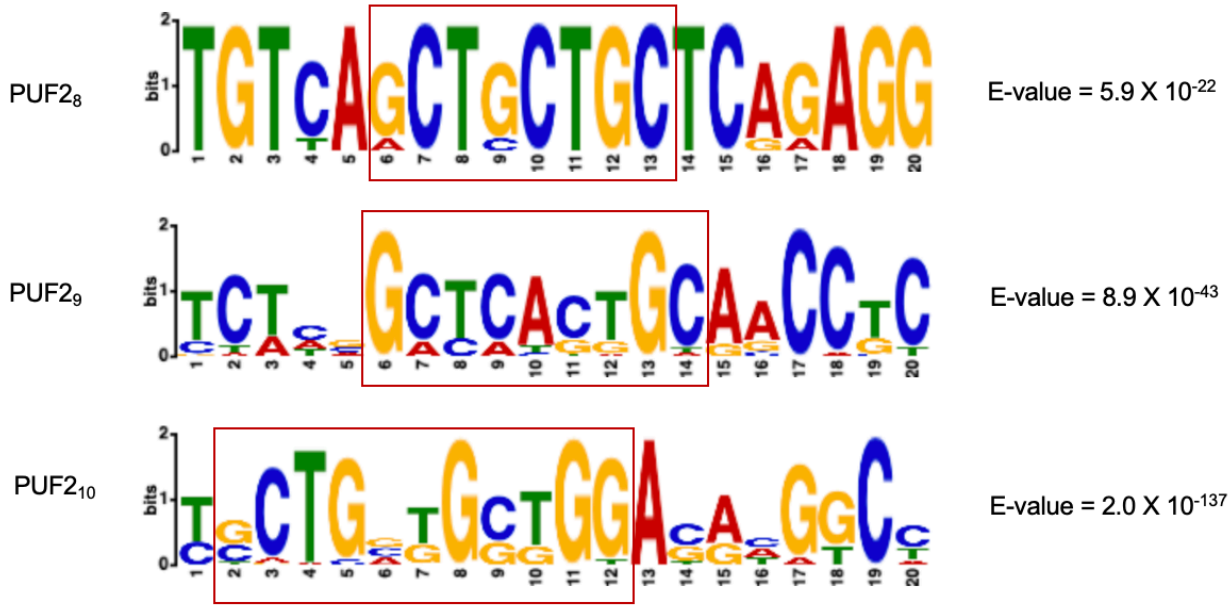


Figure 6.12: Motifs obtained from MEME analysis of enrichment peaks. Red box represents repeated motif partially matching designed target sites for corresponding PUFs. Significance of the discovered motif is represented by the E-value shown in the Figure.

The enrichment motif analysis found the exact target site for the PUF2$_8$ protein: GCTGCTGC as highlighted in Figure 6.12. A partial motif containing two mismatches is detected for the PUF2$_9$ protein: GCTCACTGC compared to the designed target site: GCTAGCTGC. However, the C in position 4 of the target site (position 9 of the discovered motif) can be an A with a lower frequency suggesting a partial mismatch at this position. Compared to PUF2$_8$, nearly every position has a slight mismatch in PUF2$_9$. For the PUF2$_{10}$ target site detected, the nucleotide at position 6 of the discovered motif has a very low bit depth, suggesting it is a non-essential nucleotide in the target site and a minority of RNAs being bound by PUF2$_{10}$ contain an additional nucleotide at this position. Elimination of this position results in the target site: GCTGTGCTGG being detected which contains only two mismatches compared to the designed site: GCTATGCTGC. However, the PUF2$_{10}$ site also has a lower bit-depth for guanine at position 1 of the target site (position 2 in the motif) compared to the PUF2$_8$ and PUF2$_9$ proteins. Additionally, the final CT doublet decreases in bit-depth from PUF2$_8$, to PUF2$_9$ and then to PUF2$_{10}$. Again, nearly every position in the PUF2$_{10}$ motif displays some promiscuity compared to PUF2$_8$. Thus, PUF2$_8$ has the fewest deviations from its cognate site with only 2 partial mismatches, whereas PUF2$_9$ has 1 complete mismatch along with 6 partial

mismatches, and $PUF2_{10}$ has 2 complete mismatches and 6 partial mismatches. Additional enriched bases outside the target motif are likely artefactual due to the bioinformatics pipeline.

6.3.5 EXPRESSION OF SPLIT-GFP FUSION CONSTRUCTS

OOPS-seq allowed us to elucidate that longer PUF domains may display more promiscuity, however it cannot provide any information about the extent to which this promiscuity affects binding, how many mismatches can be tolerated, how this affects the number of natural sites for the PUFs within the transcriptome, whether this can be translated to proteins with fused PUF domains, and how this affects concurrent binding of PUF domains. Furthermore, for $PUF2_9$ and $PUF2_{10}$, the exact designed target site was not detected. To examine how these results translate to concurrent binding of PUF fusion proteins, experiments were carried out using PUF domains fused to split-GFP. Prior to analysis by confocal microscopy, the expression of the various split-GFP constructs was assessed using Western blot. Figures 6.13A and B show the results of expression. Again, as samples were run on various gels, the blots shown are a collage of different blots.

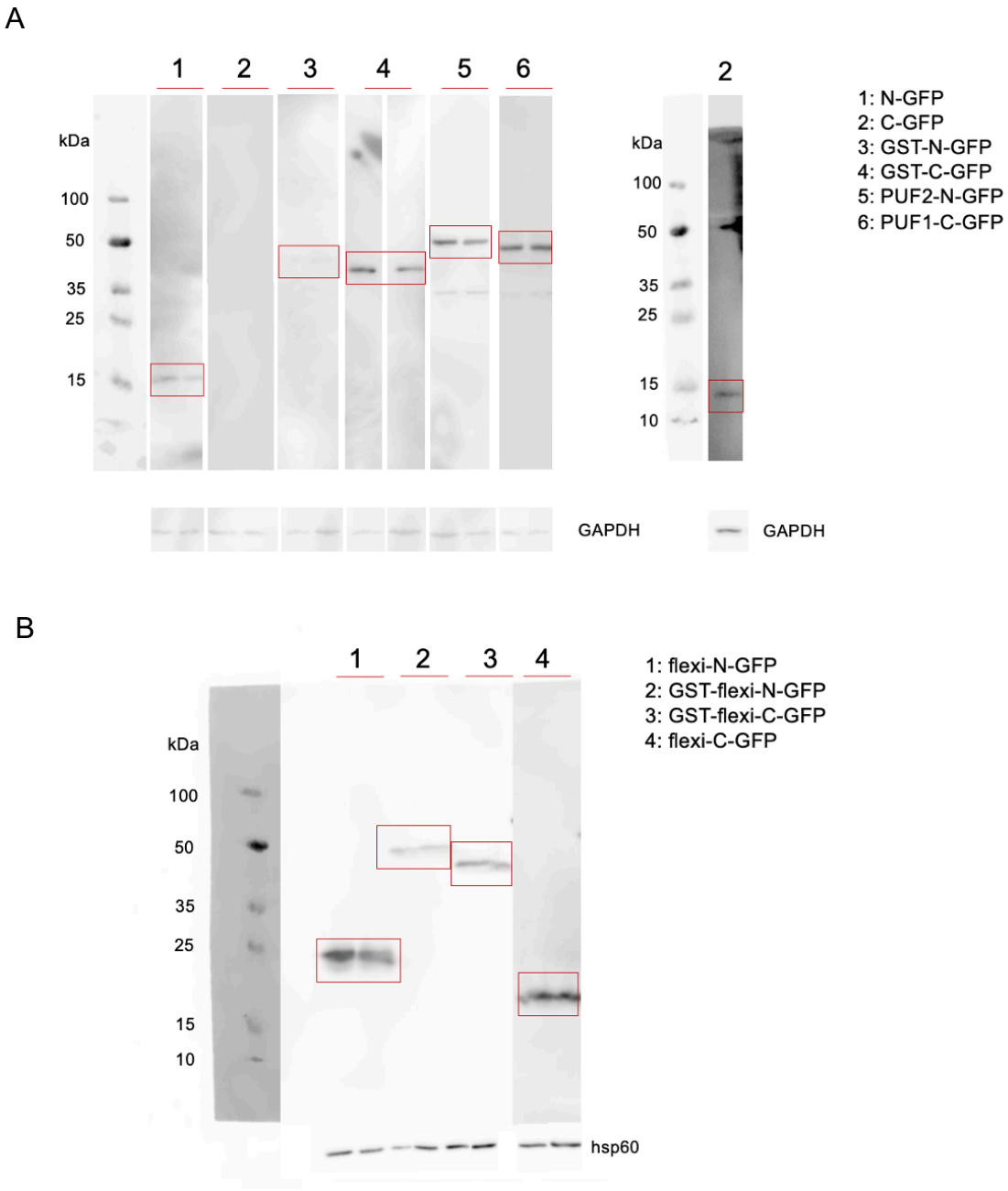


Figure 6.13A and B: Western blot of split-GFP constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG mouse monoclonal or anti-HA rat monoclonal primary antibody and subsequently with an anti-Mouse or anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. GAPDH or hsp60 was used as a loading control. Red boxes are used to highlight bands of interest. The two lanes under each title represent the two independent biological replicates.

A single band per sample per replicate is seen in all samples corresponding to the approximate molecular weight of the construct (Table 6.1). The bands for C-GFP were not detected at first attempt as they most likely ran off the gel but were seen after running the samples for a shorter time. The flexi in the construct names refer to a flexible linker (same as in Chapter 5) that was inserted into the construct to separate the different domains within the constructs.

Construct Name	Approximate Molecular Weight (kDa)
N-GFP	16
C-GFP	14
GST-N-GFP	42
GST-C-GFP	40
PUF2-N-GFP	50
PUF1-C-GFP	48
Flexi-N-GFP	24
Flexi-C-GFP	20
GST-flexi-N-GFP	50
GST-flexi-C-GFP	48

Table 6.1: Approximate molecular weights of split-GFP constructs.

6.3.6 RECONSTITUTION OF THE SPLIT-GFP SYSTEM

Once the expression of the constructs at the correct molecular weight was verified, confocal microscopy was used to detect reconstitution of split-GFP upon triple transfection into cells along with cognate or scrambled site containing mCherry to assess whether the PUF domains can bind RNA concurrently. The images are shown below.

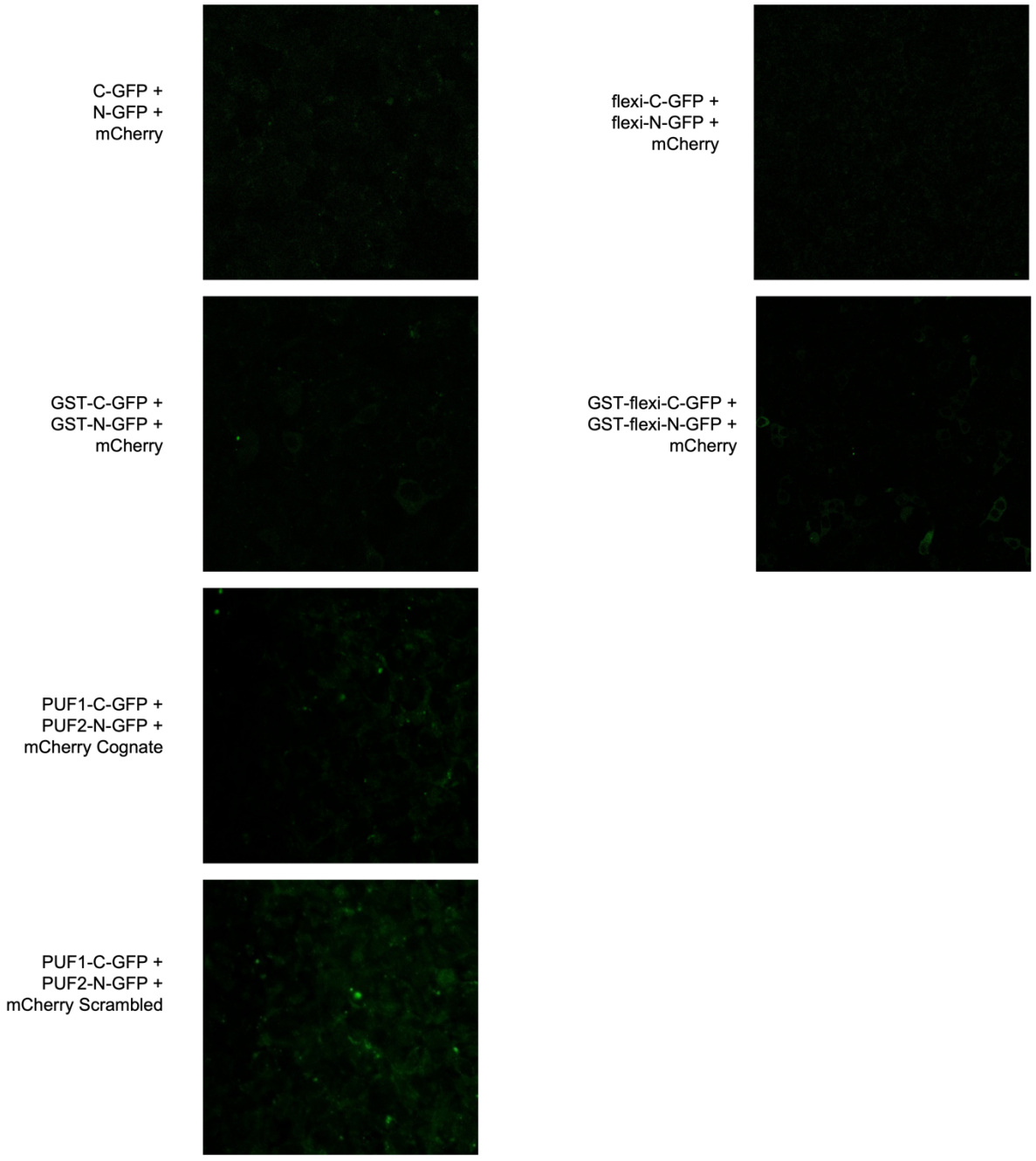


Figure 6.14: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with different combinations of plasmids listed in an iBidi 8-well slide, incubated for 24h, and subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells. All images shown are the field of view of the microscopy and are representative of two biological replicates.

Figure 6.14 shows the images taken by confocal microscopy with a filter to illuminate and detect GFP fluorescence. mCherry fluorescence was visualised prior to confocal microscopy and was even across all samples. The panels for the C-GFP + N-GFP, and GST-split-GFP combinations show no visible green fluorescence. In contrast, all combinations containing both the PUF1-C-GFP and PUF2-N-GFP, regardless of which mCherry (cognate or scrambled) was co-transfected, show significantly higher green fluorescence compared to the other combinations. It was noted that the PUF-split-GFP constructs used contained a flexible linker separating the PUF and split-GFP domains which may have improved GFP re-constitution, and thus the microscopy was repeated for C-GFP + N-GFP, and GST-split-GFP constructs after adding a flexible linker to them (Figure 6.14). The flexi-C-GFP + flexi-N-GFP combination shows no fluorescence re-constitution, but the GST flexi combination does show a low level of GFP fluorescence.

However, it appears that the scrambled RNA shows more GFP fluorescence than the cognate RNA (Figure 6.14). In order to find the best conditions to maximise GFP fluorescence in the presence of cognate RNA compared to scrambled RNA, the system was optimized as described in the next section.

6.3.7 OPTIMISATION OF THE SPLIT-GFP SYSTEM

Various different split-GFP constructs and target RNAs were screened to determine the optimal ones. The variables tested were PUF cognate site length, PUF cognate site spacing and split-GFP domain orientations. In addition, multiple binding site (MBS) variations of the cognate sites were produced to assess the effect of the presence of more available binding sites on activity. Figure 6.15 shows a schematic of the different optimisations tested in this Chapter. Outside-oriented domains refer to when the appropriate split-domain is on the C-terminus of the PUF1 (on the outer-side compared to the PUF2 fusion), whereas inside-oriented domains refer to when the split-domain is on the N-terminus of PUF1 (on the inner side compared to the PUF2 fusion). A summary of the constructs used to achieve this is listed in Table 6.2, and a comparison of the RNA target sites in the different mCherry constructs is shown in Table 6.3. The mCherry cognate constructs provided in Table 6.2 were generated for 8-, 9- and 10-mer PUF cognate sites (Table 6.3) and also contained a TEV cleavage site for use in the next

Chapter. The MBS target sites are not shown in Table 6.3 are the font size required to display them is too small, however they are simply two SBS RNAs separated by a 5'-CUUCU-3' spacer. For example: the 8-mer mCherry SC0 is 5'-**CUG—CGACA**----------*GCU—GCUGC*-3'and the corresponding 8-mer mCherry MC0 is 5'-**CUG—CGACA**----------*GCU—GCUGC* CUUCU **CUG—CGACA**----------*GCU—GCUGC*-3'.

Scrambled construct 1 contains scrambled 8-nt PUF sites separated by a 10-nt spacer, whereas scrambled construct 2 contains scrambled 16-nt PUF sites separated by a 5-nt spacer. Initial test experiments showed that both these constructs resulted in identical results for 8-, 9- and 10-mer PUF domains (Appendix E) indicating spacer and target site length do not affect non-specific binding to scrambled RNA. Thus, scrambled construct 1 was used as the control for all further experiments. Subsequently, 8-mer PUFs will be referred to as $PUF1_8$ or $PUF2_8$, 9-mer PUFs as $PUF1_9$ or $PUF2_9$, and 10-mers as $PUF1_{10}$ and $PUF2_{10}$.

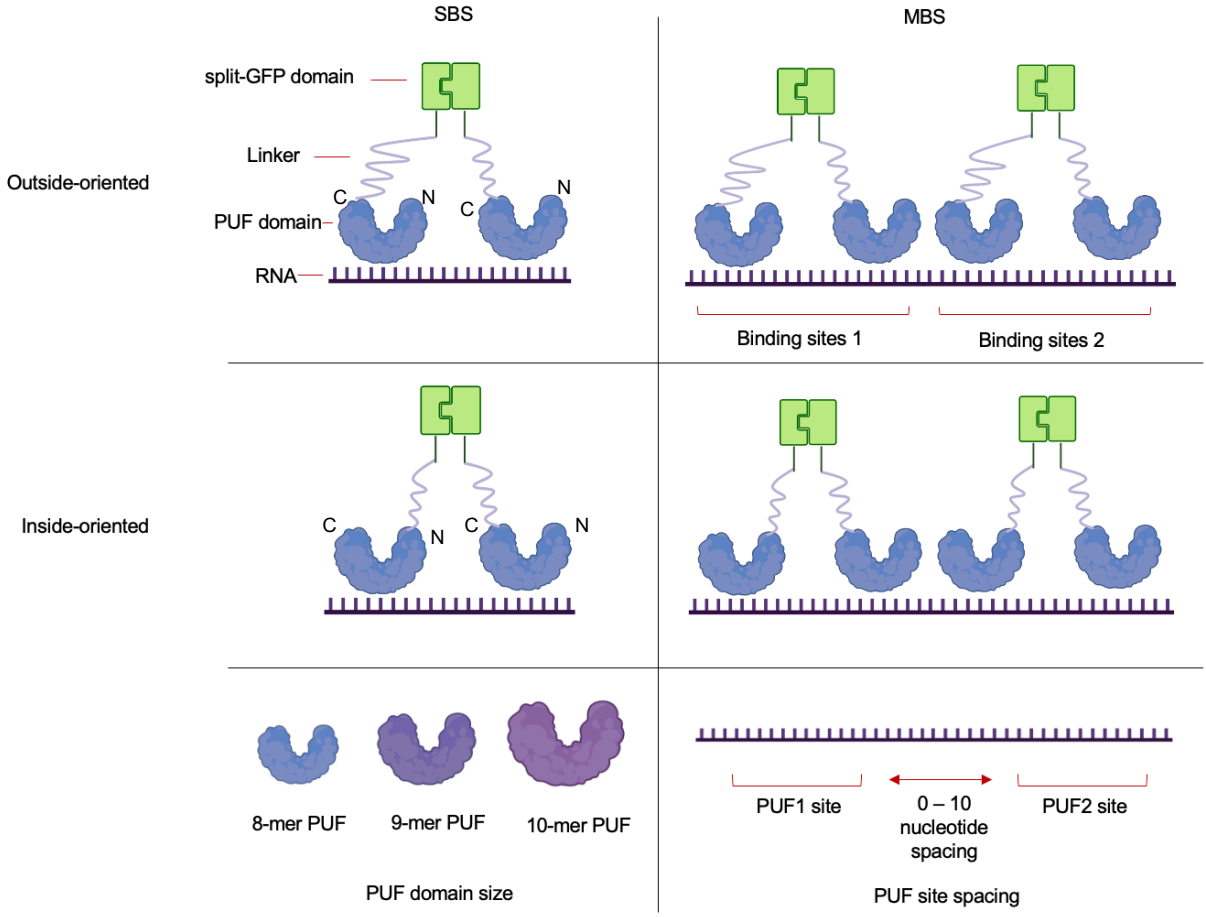


Figure 6.15: Schematic diagram of different parameters tested using the PUF-split-GFP system.

Construct Name	Protein Tag	Notes
$PUF2_8$-N-GFP	3xFLAG	-
$PUF2_9$-N-GFP	3xFLAG	-
$PUF2_{10}$-N-GFP	3xFLAG	-
$PUF1_8$-C-GFP	3xHA	Outside-oriented
$PUF1_9$-C-GFP	3xHA	Outside-oriented
$PUF1_{10}$-C-GFP	3xHA	Outside-oriented
C-GFP-$PUF1_8$	3xHA	Inside-oriented
C-GFP-$PUF1_9$	3xHA	Inside-oriented
C-GFP-$PUF1_{10}$	3xHA	Inside-oriented
mCherry SC0	3xFLAG and 3xHA	Cognate, SBS, 0-nt spacer
mCherry SC2	3xFLAG and 3xHA	Cognate, SBS, 2-nt spacer
mCherry SC4	3xFLAG and 3xHA	Cognate, SBS, 4-nt spacer
mCherry SC6	3xFLAG and 3xHA	Cognate, SBS, 6-nt spacer
mCherry SC8	3xFLAG and 3xHA	Cognate, SBS, 8-nt spacer
mCherry SC10	3xFLAG and 3xHA	Cognate, SBS, 10-nt spacer
mCherry MC0	3xFLAG and 3xHA	Cognate, MBS, 0-nt spacer
mCherry MC2	3xFLAG and 3xHA	Cognate, MBS, 2-nt spacer
mCherry MC4	3xFLAG and 3xHA	Cognate, MBS, 4-nt spacer
mCherry MC6	3xFLAG and 3xHA	Cognate, MBS, 6-nt spacer
mCherry MC8	3xFLAG and 3xHA	Cognate, MBS, 8-nt spacer
mCherry MC10	3xFLAG and 3xHA	Cognate, MBS, 10-nt spacer
mCherry Scrambled 1	3xFLAG and 3xHA	*
mCherry Scrambled 2	3xFLAG and 3xHA	*

Table 6.2: Summary of constructs generated to optimise split-GFP system. Scrambled constructs labelled with a * are discussed in Section 6.3.3. C0 – C10 in the construct names refer to the number of nucleotides between the PUF binding sites.

Construct Name	Target Site		
8-mer mCherry SC0	**CUG—CGACA**	----------	*GCU--GCUGC*
8-mer mCherry SC2	**CUG—CGACA**	AA--------	*GCU--GCUGC*
8-mer mCherry SC4	**CUG—CGACA**	AAAA------	*GCU--GCUGC*
8-mer mCherry SC6	**CUG—CGACA**	AAAAAA----	*GCU--GCUGC*
8-mer mCherry SC8	**CUG—CGACA**	AAAAAAAA--	*GCU--GCUGC*
8-mer mCherry SC10	**CUG—CGACA**	AAAAAAAAAA	*GCU--GCUGC*
9-mer mCherry SC0	**CUGA-CGACA**	----------	*GCUA-GCUGC*
9-mer mCherry SC2	**CUGA-CGACA**	AA--------	*GCUA-GCUGC*
9-mer mCherry SC4	**CUGA-CGACA**	AAAA------	*GCUA-GCUGC*
9-mer mCherry SC6	**CUGA-CGACA**	AAAAAA----	*GCUA-GCUGC*
9-mer mCherry SC8	**CUGA-CGACA**	AAAAAAAA--	*GCUA-GCUGC*
9-mer mCherry SC10	**CUGA-CGACA**	AAAAAAAAAA	*GCUA-GCUGC*
10-mer mCherry SC0	**CUGAUCGACA**	----------	*GCUA-GCUGC*
10-mer mCherry SC2	**CUGAUCGACA**	AA--------	*GCUAUGCUGC*
10-mer mCherry SC4	**CUGAUCGACA**	AAAA------	*GCUAUGCUGC*
10-mer mCherry SC6	**CUGAUCGACA**	AAAAAA----	*GCUAUGCUGC*
10-mer mCherry SC8	**CUGAUCGACA**	AAAAAAAA--	*GCUAUGCUGC*
10-mer mCherry SC10	**CUGAUCGACA**	AAAAAAAAAA	*GCUAUGCUGC*
mCherry Scrambled 1	CAATCGAG	AAAAAAAAAA	CTCCGGGT
mCherry Scrambled 2	CAATCGAGCTCCGGGTCTTCTCAATCGAGCTCCGGGT		

Table 6.3: Summary of PUF target sequences fused to the 3'-end of mCherry mRNA. Scrambled constructs labelled with a * are discussed in Section 6.3.3. C0 – C10 in the construct names refer to the number of nucleotides between the PUF binding sites. MBS RNAs are not shown in this table as the font size requires to display them is too small. **Bold and underlined** = PUF1 binding site, *italics and underlined* = PUF2 binding site, underlined-only = scrambled site, unmodified text = spacer between the PUF sites.

6.3.7.1 8-MER SITES

Outside-oriented 8-mer PUF proteins were tested for activity with various spacer lengths present between the sites for the PUF domains. This was expected to improve the ability of the PUF fusions to bind by reducing any steric hindrance present. These assays were conducted using mCherry constructs containing either an SBS or MBS of both PUF proteins.

First, a linear series of spacing between 8-mer PUF sites on the SBS RNA was tested on the outside-oriented split-GFP domains together with MBS RNAs (Figure 6.16). The panels show that GFP fluorescence presents as minute puncta rather than bright, diffuse green as expected when an intact GFP protein is used, although diffuse GFP fluorescence is observable in certain conditions such as C4 SBS and C10 MBS (Figure 6.16). For the SBS RNAs, there are very few puncta seen for all spacer lengths, but the C8 SBS RNA shows the brightest and was thus selected for further study using flow cytometry. There are green puncta of similar brightness visible in the scrambled control; differences between the cognate and scrambled puncta cannot be distinguished in these images. This could be due to low overall split-GFP fluorescence intensity and thus, MBS versions of the RNAs were used in an attempt to increase the number of re-constituted split-GFP domains and therefore increase the split-GFP fluorescence.

Compared to their SBS counterparts and scrambled control, the MBS RNAs in general show higher GFP fluorescence. The C4 condition was selected for flow cytometry. No differences in fluorescence could be observed as the spacer length was changed for either the SBS or MBS RNAs. Although the GFP re-constitution was improved due to the MBS RNAs, for the NASPER system, a high level of re-constitution is required using SBS RNAs. Thus, to improve the orientation of the split-GFP domains, the fusion proteins were re-engineered to have the split-GFP domains facing towards each other (inside-oriented, Figure 6.15). These inside-oriented split-GFP 8-mer PUFs were also tested using the MBS versions of the RNAs in order to detect the GFP re-constitution more easily.

Figure 6.17 shows the same experiment as Figure 6.16, except that inside-oriented split-GFP domains are used instead of outside-oriented. The fluorescence observed for this experiment looks significantly different to Figure 6.16. All the panels, SBS and MBS RNAs, show a mixture

of minute puncta and diffuse green fluorescence. The brightest green fluorescence for the SBS RNAs was observed with the C10 spacer length, which looks brighter than the scrambled control and was selected for flow cytometry. The C0, C2, C6 and C10 MBS RNAs show brighter green fluorescence compared to their SBS counterparts (and the scrambled control), and of these, the C0 was the brightest and was selected for flow cytometry. Although the GFP fluorescence is greater in these conditions, no differences in fluorescence could be observed when the spacer length was changed for SBS or MBS RNAs.

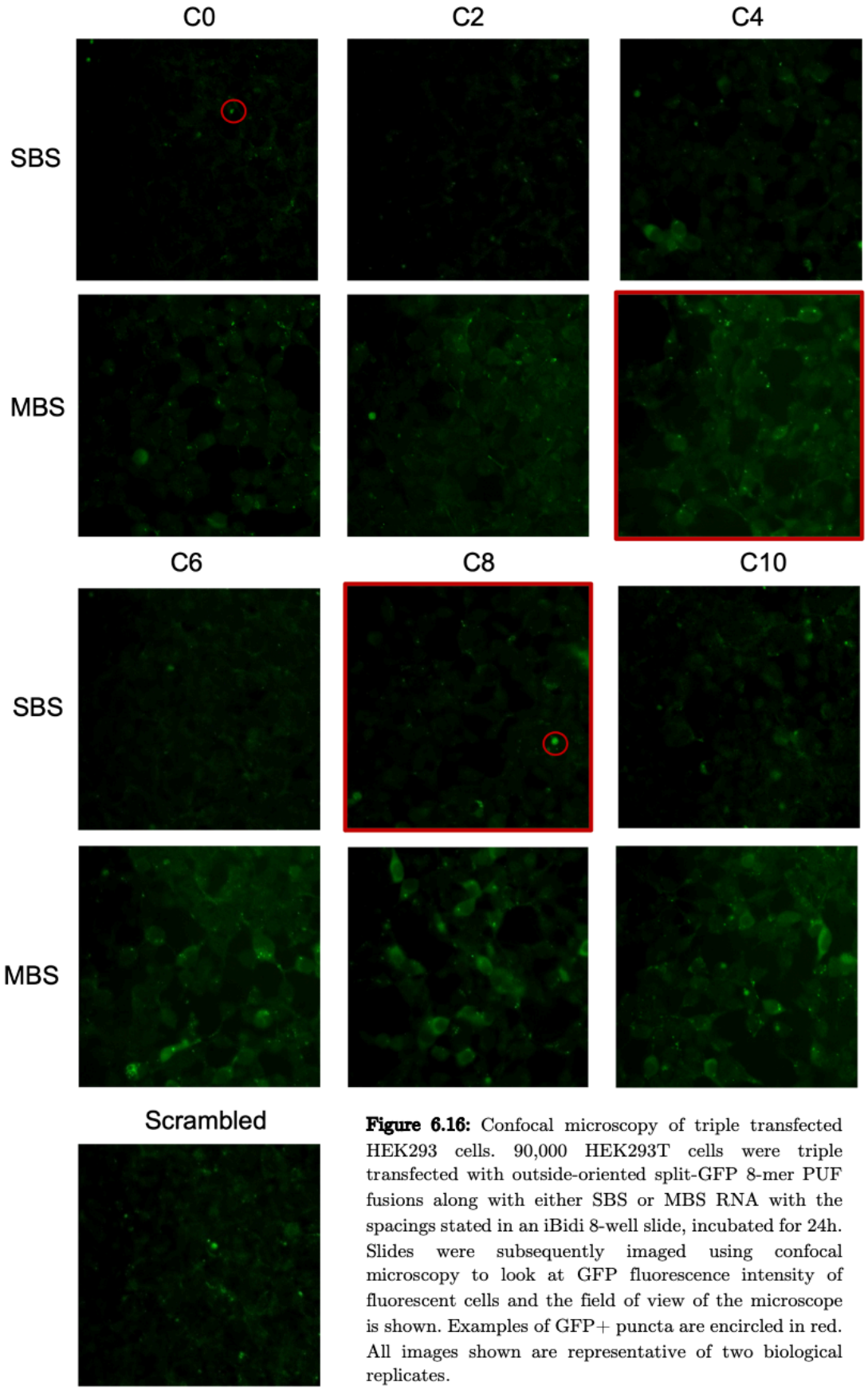


Figure 6.16: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with outside-oriented split-GFP 8-mer PUF fusions along with either SBS or MBS RNA with the spacings stated in an iBidi 8-well slide, incubated for 24h. Slides were subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells and the field of view of the microscope is shown. Examples of GFP+ puncta are encircled in red. All images shown are representative of two biological replicates.

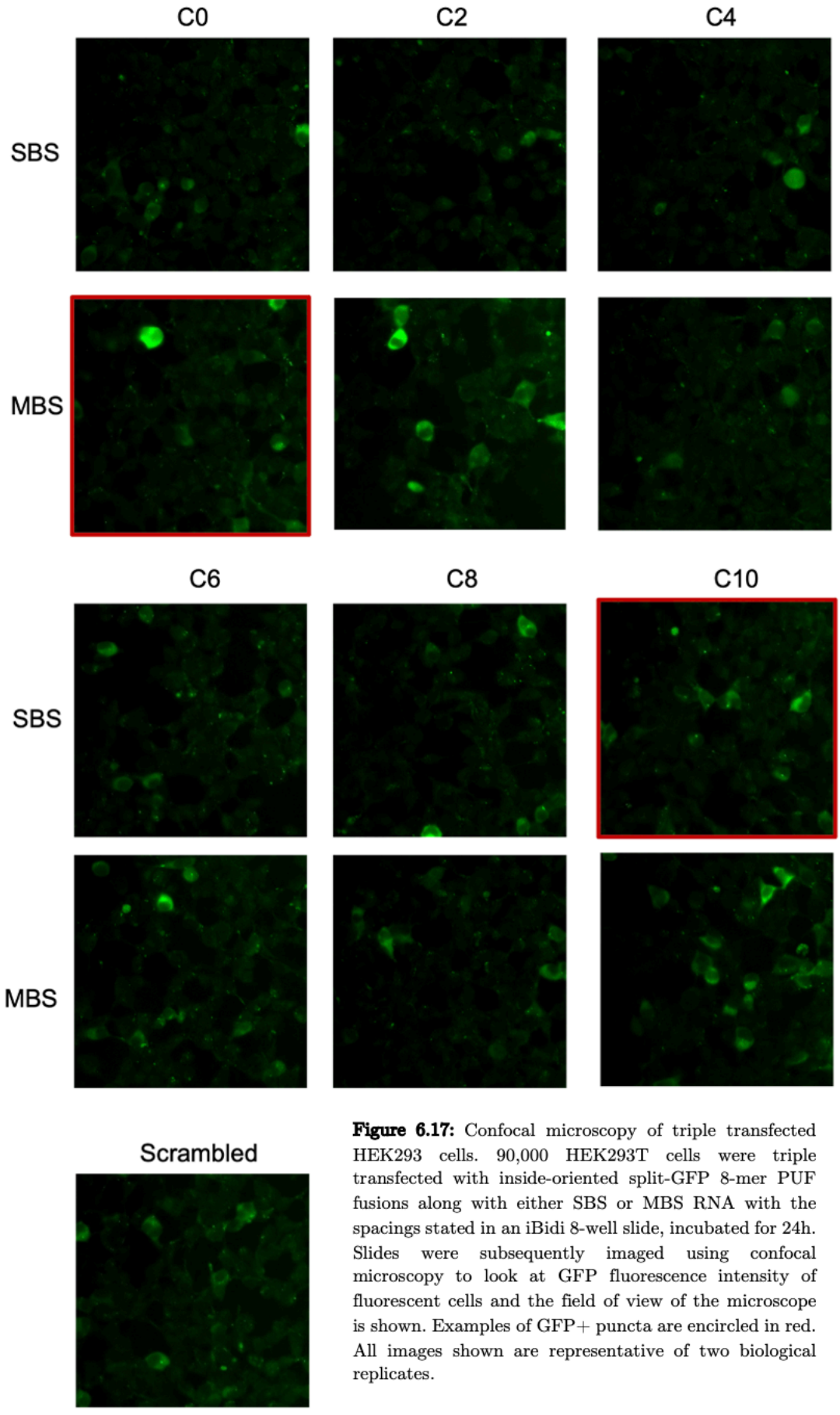


Figure 6.17: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with inside-oriented split-GFP 8-mer PUF fusions along with either SBS or MBS RNA with the spacings stated in an iBidi 8-well slide, incubated for 24h. Slides were subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells and the field of view of the microscope is shown. Examples of GFP+ puncta are encircled in red. All images shown are representative of two biological replicates.

6.3.7.2 9-MER SITES

In order to improve the PUF-split-GFP system, 9- and 10-mer PUF proteins fusions designed and produced with the presumption being that 9- and 10-nucleotide sites would occur less frequently within the cellular transcriptome as compared to 8-nucleotide sites and thereby would reduce the background fluorescence observed when using 8-mer fusions with the scrambled RNAs. Although OOPS-seq showed that longer PUF fusions may be more promiscuous, this does not necessarily mean that they bind more RNAs in the cell because OOPS-seq did not yield any information on binding affinities. Thus, it was important to assess whether they did in fact reduce split-GFP re-constitution when using scrambled RNA. The same variations of these proteins were tested as for the 8-mer PUF proteins.

Figure 6.18 shows the use of 9-mer PUF domains together with outside-oriented split-GFP domains on SBS and MBS RNAs. The panels only show minute green puncta representative of GFP reconstitution with no clear trend observable as spacer length increases. In the SBS condition, the C2 and C6 spacer lengths showed the brightest puncta. The C2 was chosen as the puncta were slightly larger, potentially facilitating detection in flow cytometry. MBS RNAs were tested here as well to increase the split-GFP re-constitution, as observed for the 8-mer fusions, to be more easily detected by the confocal microscope. The C0 and C4 MBS RNAs showed brighter green fluorescence compared to their SBS counterparts, and the C4 was selected for further study. The brightest conditions selected also look brighter than the scrambled RNA, suggesting that they may be distinguishable by flow cytometry. Subsequently, to improve re-constitution using SBS RNAs, and to compare to the data for the 8-mer split-GFP fusions, inside-oriented split-GFP 9-mer PUF fusions were also tested. Figure 6.19 shows the same experiment as Figure 6.18, except using inside-oriented split-GFP domains. For the SBS RNAs, no trend can be seen as spacer length increases. The brightest and largest puncta were seen when the C4 RNA was used, and thus this was selected for further study. The puncta in this panel also had higher fluorescence intensity than the scrambled control. The C0 and C2 MBS RNAs showed brighter and larger puncta than their SBS versions. Of these two, the C0 MBS looked better and was selected for flow cytometry. Interestingly, both scrambled controls (outside- and inside-oriented split-GFP) with the 9-mer PUF domains showed fewer and less intense puncta compared to the scrambled controls with 8-mer PUF domains.

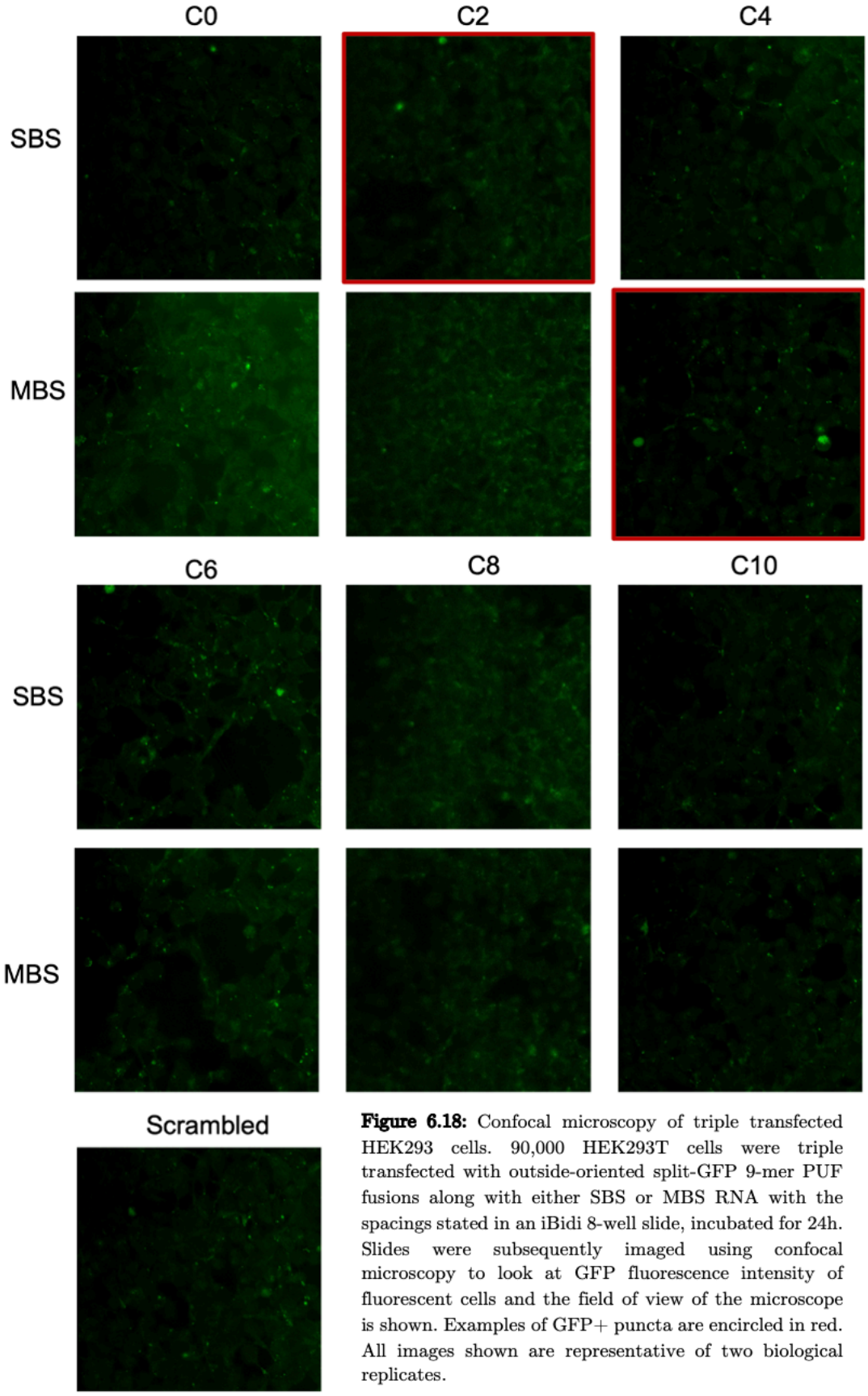


Figure 6.18: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with outside-oriented split-GFP 9-mer PUF fusions along with either SBS or MBS RNA with the spacings stated in an iBidi 8-well slide, incubated for 24h. Slides were subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells and the field of view of the microscope is shown. Examples of GFP+ puncta are encircled in red. All images shown are representative of two biological replicates.

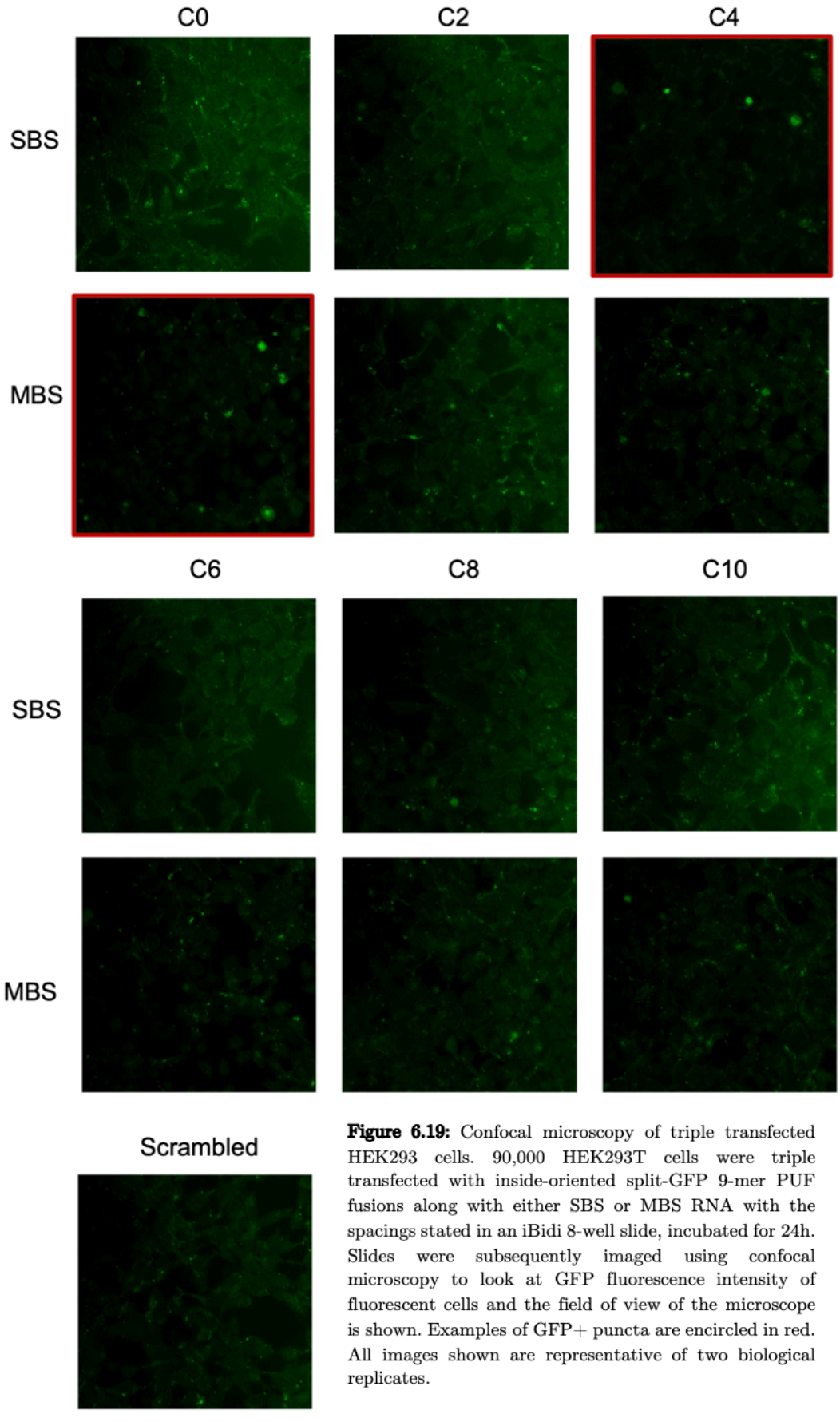


Figure 6.19: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with inside-oriented split-GFP 9-mer PUF fusions along with either SBS or MBS RNA with the spacings stated in an iBidi 8-well slide, incubated for 24h. Slides were subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells and the field of view of the microscope is shown. Examples of GFP+ puncta are encircled in red. All images shown are representative of two biological replicates.

6.3.7.3 10-MER SITES

Although the 9-mer PUF fusions did not yield split-GFP re-constitution that was easily detectable by the confocal microscope which may suggest higher specificity, it could also have been because the 9-mer PUF domains were unable to bind the RNA targets well. Thus, 10-mer PUF fusions were designed and tested to see whether they had higher specificity and more PUF binding and re-constitution than the 9-mer fusions.

Figure 6.20 shows the use of 10-mer PUF domains together with outside-oriented split-GFP domains on SBS and MBS RNAs. The panels only show green puncta representative of GFP reconstitution with no clear trend observable as spacer length increases for the SBS or MBS RNAs. In the SBS condition, the C4 and C8 spacer lengths showed the brightest puncta, and of these, the C8 was chosen as the puncta were observed to be larger and brighter. The C0 and C2 MBS RNAs showed brighter green fluorescence compared to their SBS counterparts and the C0 was selected for further study. These constructs also look brighter than the scrambled RNA, suggesting that they may be distinguishable by flow cytometry.

Figure 6.21 shows the same experiment as Figure 6.20, except using inside-oriented split-GFP domains. For the SBS RNAs, no trend can be seen as spacer length increases. The brightest and largest puncta were seen when the C0 RNA was used and thus this was selected for further study. The puncta in this panel also had higher fluorescence intensity than the scrambled control. The C0 MBS RNA showed brighter and larger puncta than its SBS versions. The C0 MBS RNA was also the brightest of all the MBS RNAs and was selected for flow cytometry.

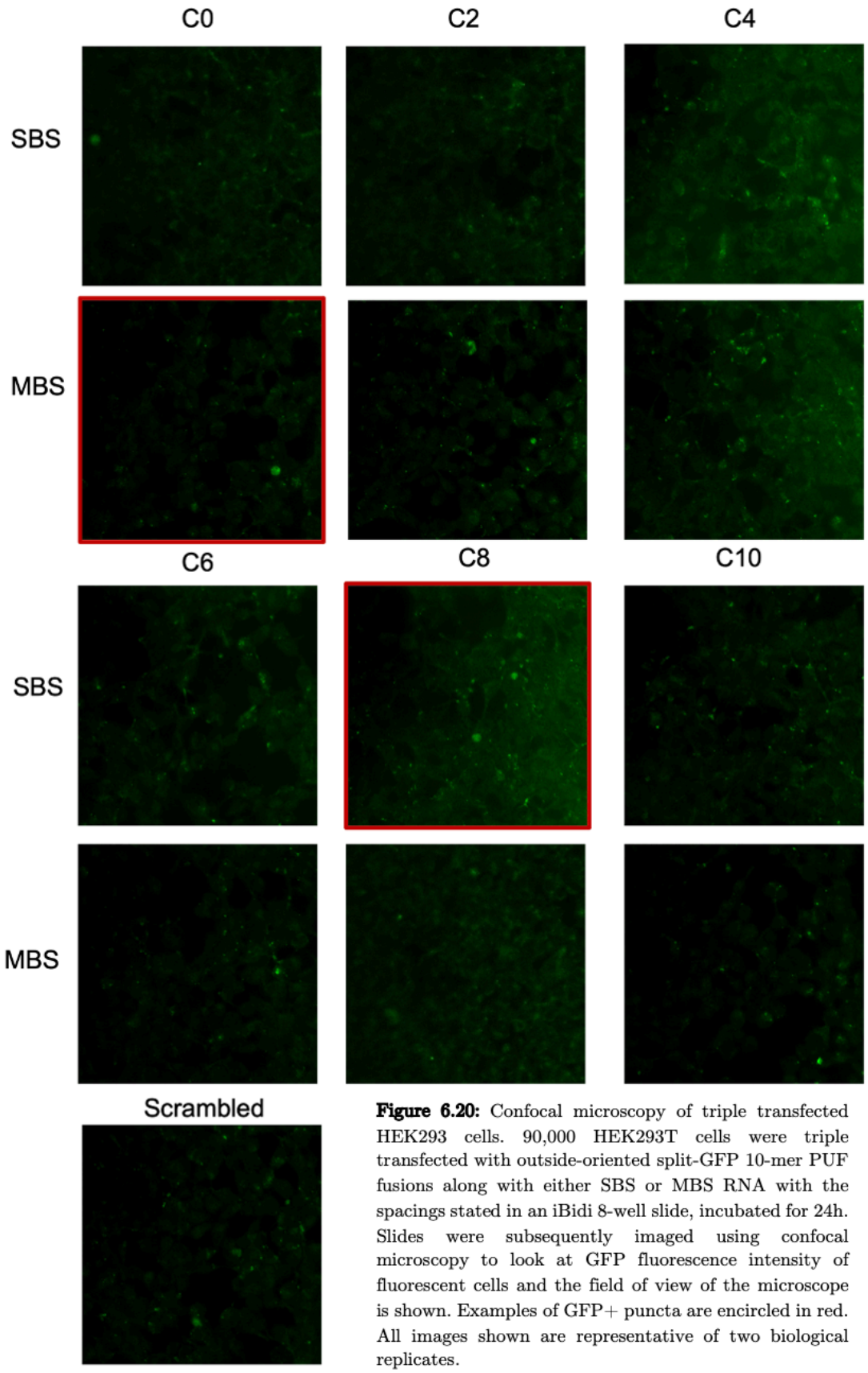


Figure 6.20: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with outside-oriented split-GFP 10-mer PUF fusions along with either SBS or MBS RNA with the spacings stated in an iBidi 8-well slide, incubated for 24h. Slides were subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells and the field of view of the microscope is shown. Examples of GFP+ puncta are encircled in red. All images shown are representative of two biological replicates.

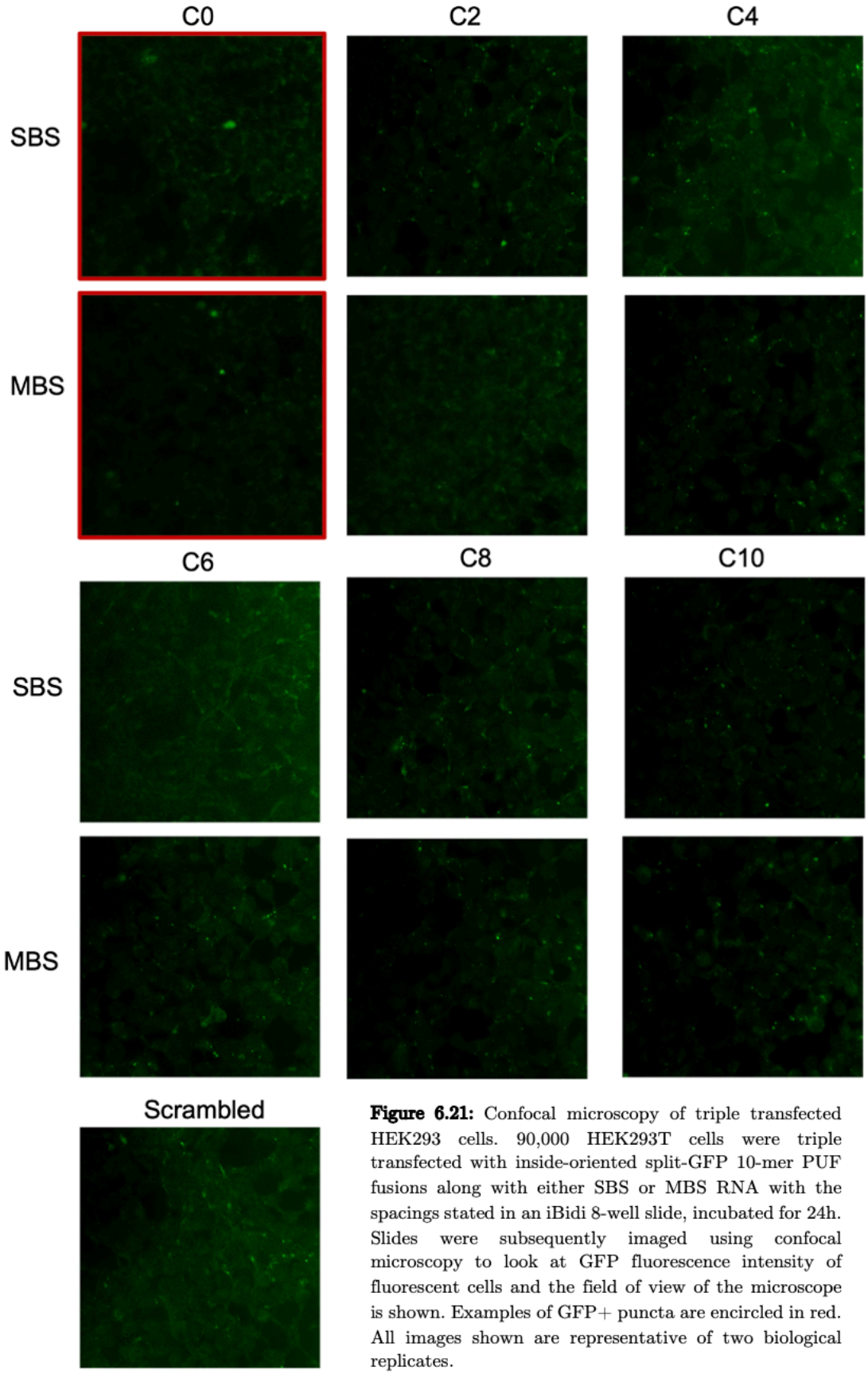


Figure 6.21: Confocal microscopy of triple transfected HEK293 cells. 90,000 HEK293T cells were triple transfected with inside-oriented split-GFP 10-mer PUF fusions along with either SBS or MBS RNA with the spacings stated in an iBidi 8-well slide, incubated for 24h. Slides were subsequently imaged using confocal microscopy to look at GFP fluorescence intensity of fluorescent cells and the field of view of the microscope is shown. Examples of GFP+ puncta are encircled in red. All images shown are representative of two biological replicates.

6.3.8 FLUORESCENCE INTENSITY OF SELECTED CONSTRUCTS

As there were difficulties with quantification of the fluorescence intensities using confocal microscopy, a plate reader assay was used to try and quantify the split-GFP re-constitution of the brightest sets of constructs selected from confocal microscopy to quantitatively validate the differences observed (Figure 6.22).

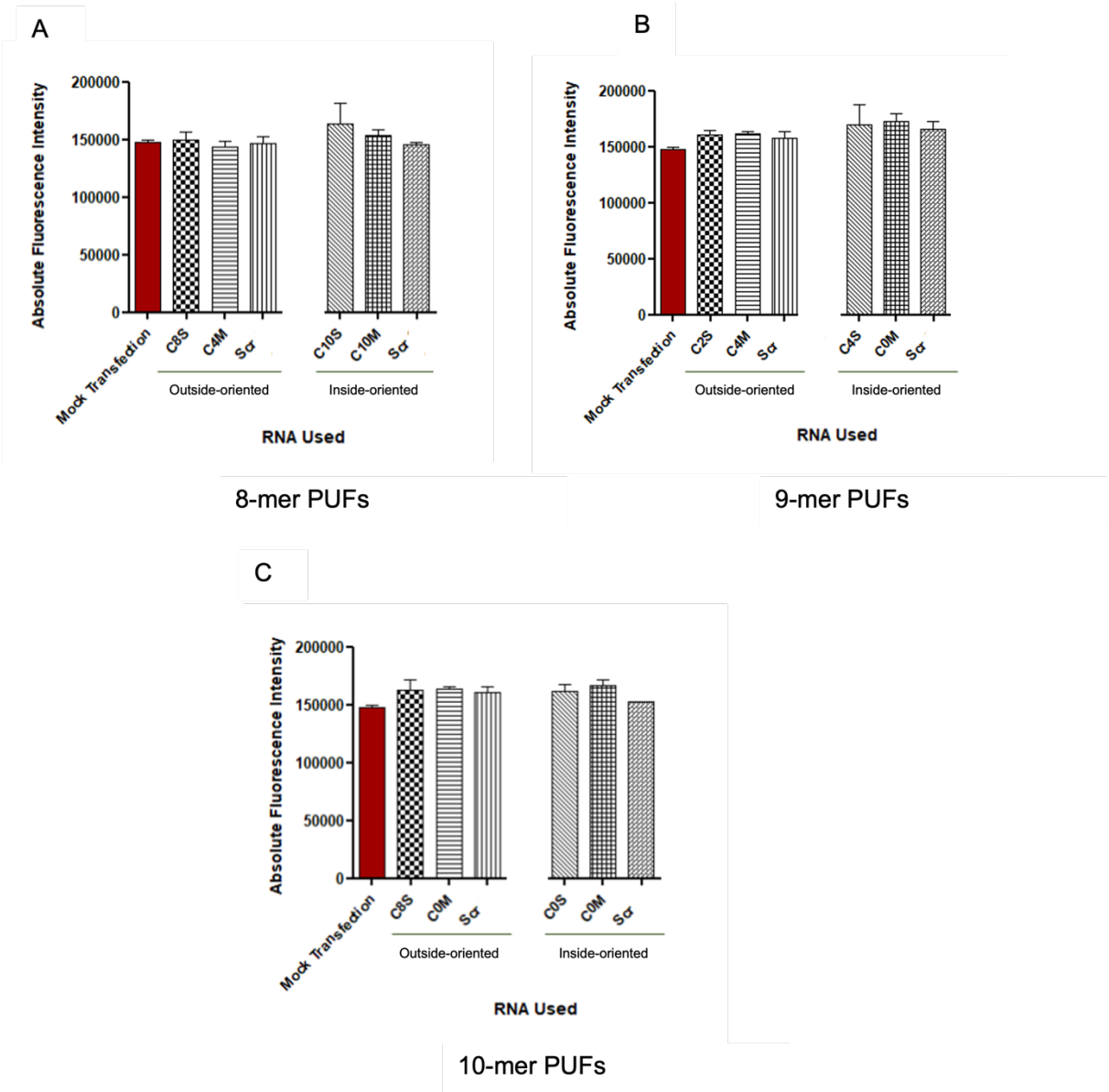


Figure 6.22: Absolute fluorescence intensities of selected 8-, 9-, and 10-mer PUF constructs. 10,000 HEK293T cells were transfected with the brightest constructs determined in confocal microscopy for 24h and then analysed using a plate reader to measure total GFP fluorescence intensity per well of cells. Error bars represent the SEM from three biological replicates.

The absolute split-GFP fluorescence intensities for all RNA conditions were nearly the same for the three PUF lengths tested. A one-way ANOVA coupled with Dunnett's post-hoc test to compare these intensities to the mock transfection (untransfected cells) showed no significant differences between the control and any of the constructs tested. Thus, flow cytometry was attempted instead because it uses a more powerful light source and more sensitive fluorescence detectors which would enable the detection of minute fluorescence intensities that are invisible to the plate reader.

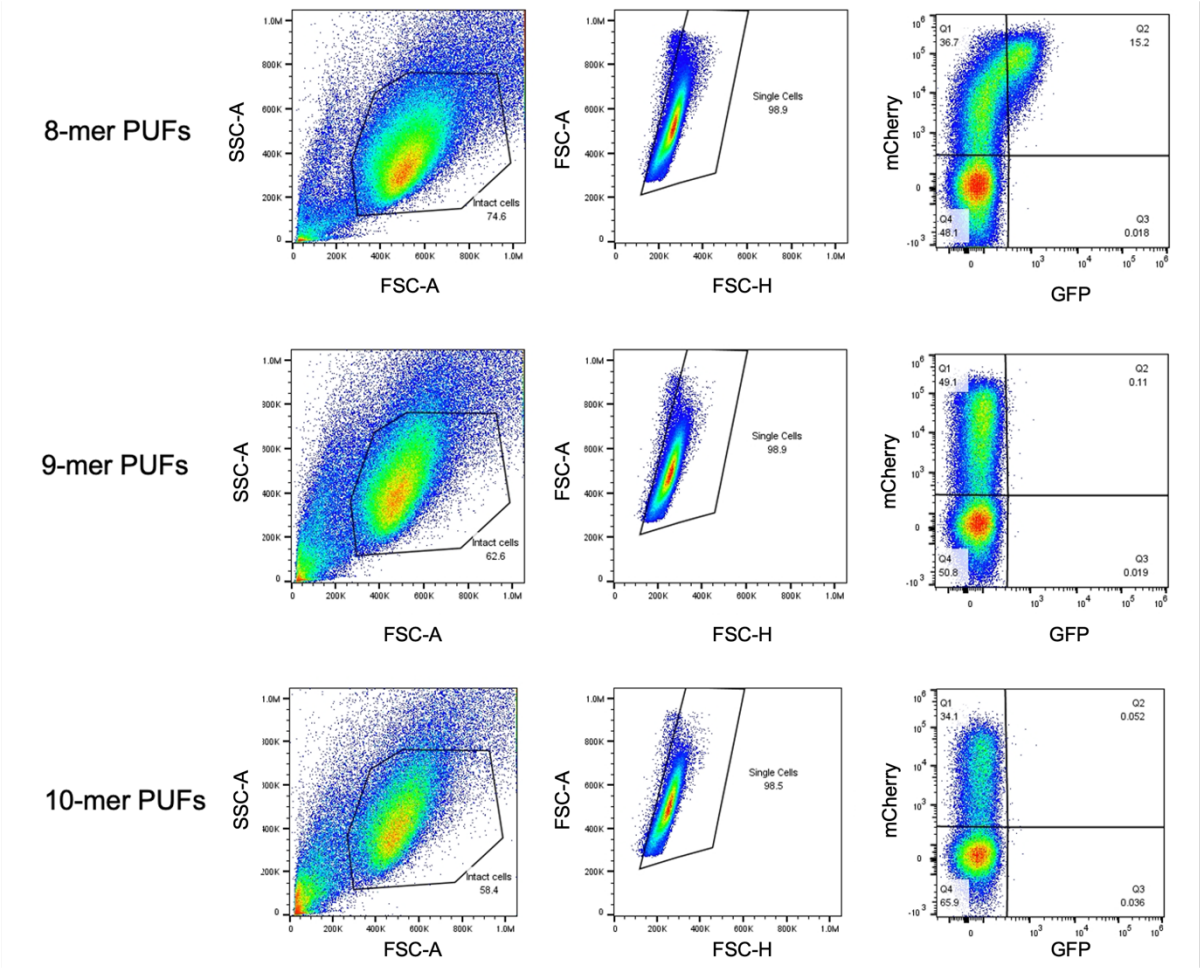


Figure 6.23: Flow cytometry of selected split-GFP constructs. 50,000 HEK293T cells were transfected with the brightest constructs determined in confocal microscopy for 24h and then analysed using flow cytometry to measure total GFP and mCherry fluorescence intensity per cell. Representative dot plots from three independent biological replicates for 8, 9 and 10-mer PUF proteins showing intact cell selection, single cell selection and GFP fluorescence intensity.

Figure 6.23 shows the representative dot plots obtained from the 8-mer, 9-mer and 10-mer PUF split-GFP constructs tested. All 8-mer PUFs tested had cells present in the second quadrant which represents mCherry and GFP positive cells, whereas the 9- and 10-mer PUF

conditions did not show any cells in this quadrant. In order to compare between the fluorescence intensities of the different 8-mer PUF-split-GFP conditions, overlays of the GFP intensity histograms were produced (Figure 6.24). However, the median fluorescence intensity (MFI) values may not be accurate because there is a shift in GFP fluorescence occurring in the GFP- quadrant 1 of the dot plots (Figure 6.23) and thus a significant amount of signal is being ignored in the analysis due to systematic issues with flow cytometer voltage setting.

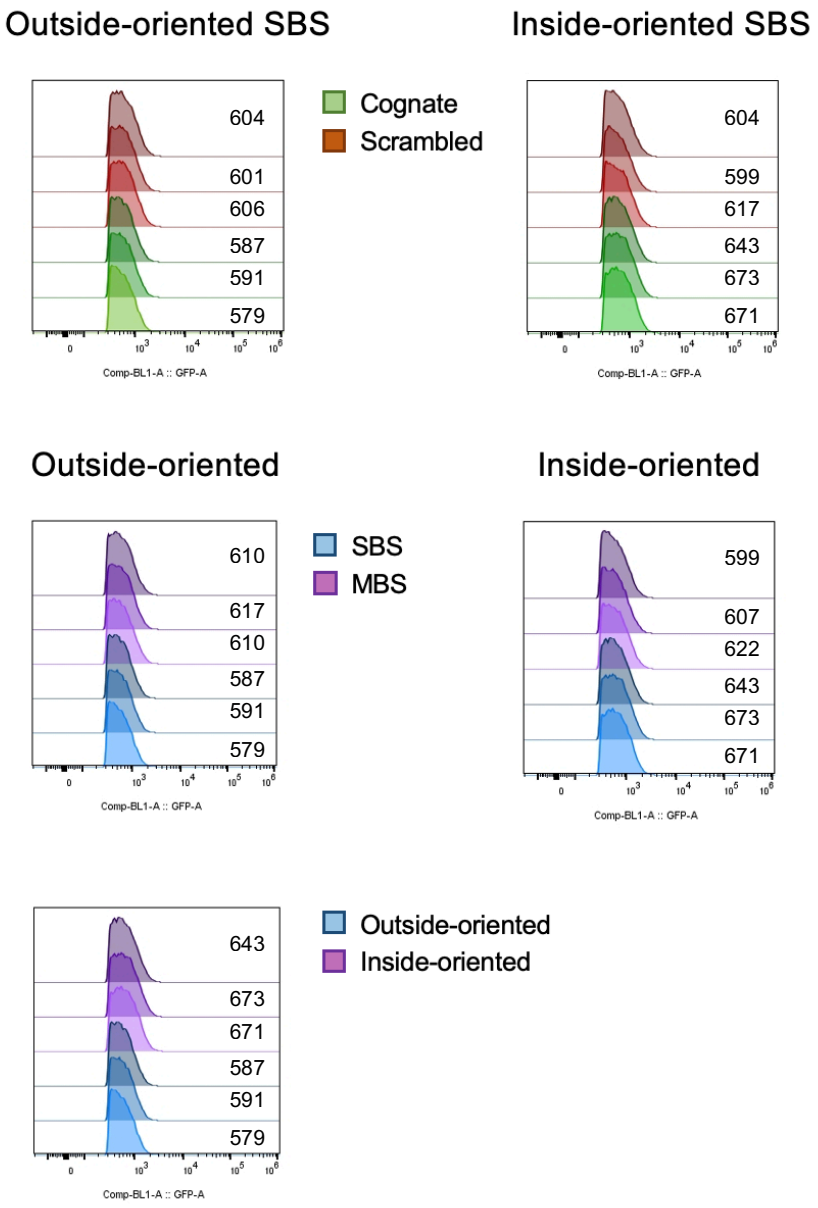


Figure 6.24: split-GFP intensity histogram overlays for 8-mer PUFs. Overlays showing the fluorescence intensities of selected constructs of 8-mer PUF-split-GFP measured as median fluorescence intensity by flow cytometry in FlowJo. Three independent biological replicates are shown as three histograms for each condition in the key – eg: the three green histograms for outside-oriented SBS represent the three biological replicates testing the outside-oriented 8-mer PUF fusions using cognate SBS RNA.

The histogram overlays (Figure 6.24) showed no shift in GFP intensity when cognate and scrambled SBS RNA were used for both the outside- and inside-oriented split-GFP domains. Although the MFIs for the inside-oriented split-GFP fusions was greater when using the cognate RNA compared to the scrambled RNA, due to systematic errors these values may be artefactual. Similarly, comparison of the MFIs between SBS and MBS constructs, and outside- and inside-oriented constructs, is unreliable.

6.4 DISCUSSION

The first aim of this Chapter was to employ OOPS, a novel high-throughput method of protein-bound RNA purification, to enrich PUF-bound RNA for subsequent next-generation sequencing and bioinformatic comparison to a control sample for PUF binding site analysis as an alternative to CLIP-seq. Furthermore, this Chapter describes the experiments carried out in attempts to understand the nature of concurrent binding of two PUF proteins using the split-GFP system using confocal microscopy. Studies using the split-GFP system with PUF proteins have been carried out, but they all use super-resolution techniques such as total internal reflection fluorescence (TIRF) microscopy[168–170]. These studies have shown that the signal obtained from PUF-RNA-induced split-GFP reconstitution is rather low, and there may be background non-specific binding of the probes[169,170]. Thus, my use of confocal microscopy was not the ideal method, but it was the most sensitive instrument available, and demonstrating GFP reconstitution using more readily available confocal microscopes would also broaden the use of this PUF-split-GFP system in the RNA imaging field.

6.4.1 NGS REVEALS PROMISCUITY OF PUF BINDING

Prior to any downstream bioinformatic analyses, sequencing quality was checked using FastQC, and read qualities were extremely high allowing for robust downstream analysis. Initially, I aimed to use the protein-binding fingerprint method suggested in Queiroz *et al.*[115] to detect genomic regions in the PUF libraries that were depleted compared to the control libraries. Such depletion occurs because UV crosslinking-induced protein adducts on RNA block reverse transcriptase during library preparation resulting in protein-bound regions being depleted in the final library[174]. The MA plots and motif analysis for this experiment suggests that the results are likely an artefact; a common sequence was detected for the $PUF2_8$ and $PUF2_9$ suggesting a potential systematic bias in this analysis (Figures 6.9 and 6.10). Upon further consideration, two sources of systematic bias in this mode of analysis were apparent – 1) read-through rate and 2) RNA enrichment due to PUF binding.

Firstly, read-through is a phenomenon that occurs when the reverse transcriptase does not terminate due to the presence of protein adducts on the RNA[174]. Thus, if read-through

frequency is high, PUF binding and crosslinking to RNA would not produce the expected depletions. This may indeed be the case here since the fold changes of significant in the depletion analysis MA plots are very low, almost identical to those for non-significantly depleted peaks. A novel technique developed by Su *et al.*[185] known as LACE-seq, does not use proteinase K digestion of the protein crosslinked to RNA thus resulting in a much larger adduct remaining on the RNA. This has a more profound effect on reverse transcription termination by minimizing read-through rate. Thus, using this technique in the future for depletion analysis may be significantly more sensitive than the approach adopted here.

Secondly, PUF-bound RNAs would be enriched in the OOPS interface compared to the control libraries. This enrichment leads to a higher concentration of PUF-bound RNAs in the PUF libraries compared to the controls. Due to this, the read depletion may not be enough to compensate for the RNA enrichment and thus, depleted regions may not truly be detected (Figure 6.25)

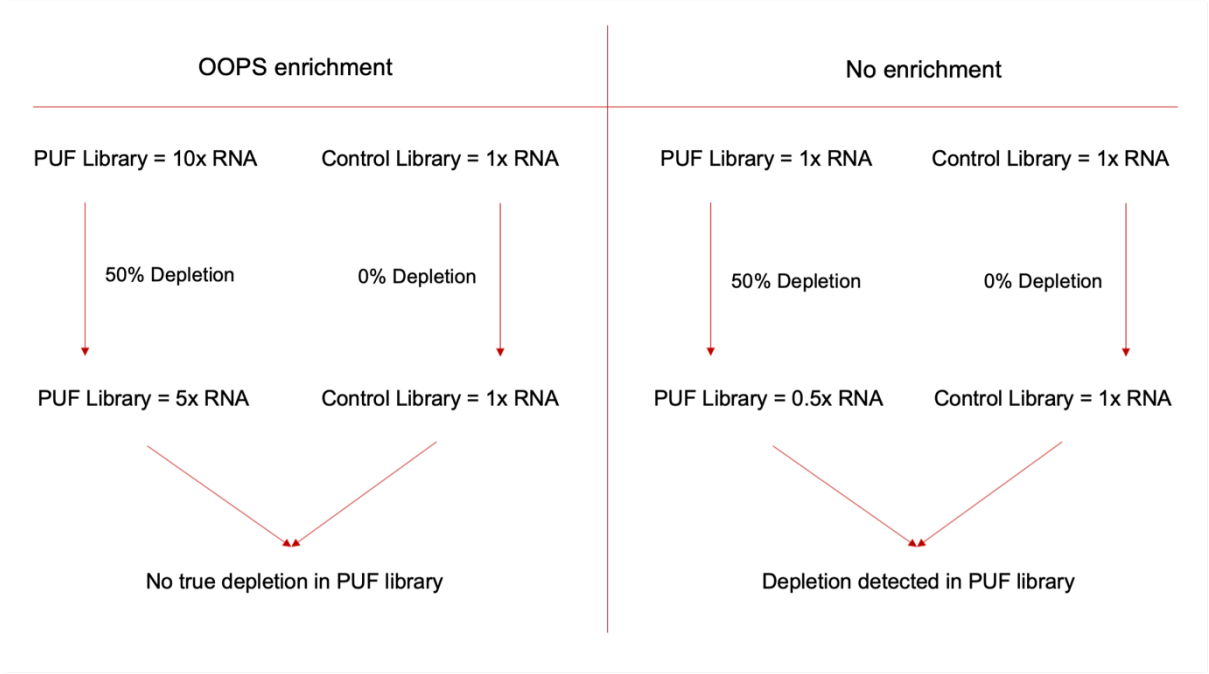


Figure 6.25: Schematic for effect of enrichment on read depletion.

The effect of enrichment on read depletion is affected by the degree of enrichment and degree of depletion, which may be why Queiroz *et al.*[115] detected depletions in some genes at protein-binding regions. However, this work was only applied to highly expressed RNAs (as this may

mitigate the effects of enrichment) and not developed further and validated for all known protein binding sites and on low expression RNAs. In my experiments, looking at only highly expressed RNAs for depletions would greatly bias the downstream motif analysis as it may exclude significant binding regions that are present on low expression RNAs. Hence, a better bioinformatic analysis method was required.

Since the PUF proteins seemed to be causing an enrichment of their RNA binding partners in the OOPS interface, I instead decided to apply the peak calling protocol to search for peaks that were enriched in the PUF libraries compared to the control libraries. Motif analysis yielded nearly the exact designed target sequence of the $PUF2_8$ protein (which was also validated by *in vitro* fluorescence polarisation in Chapter 3). Nucleotides in positions 1 and 4, however were amenable to slight promiscuity with some RNAs containing adenine and cytosine in those position respectively. Such promiscuity has been observed previously[104] in PUF proteins, however the $PUF2_8$ protein seems to have a generally high specificity since this promiscuity is minimal as seen by the low bit depth of A and C at positions 1 and 4, and since nucleotides at other positions match the designed sequence exactly.

Surprisingly, in contrast, $PUF2_9$ displays lower specificity for the designed target and $PUF2_{10}$ is even less specific for the designed target compared to $PUF2_9$. Zhao *et al.*[102] have suggested that increasing the number of PUF repeats may allow binding in imperfect modes if the bound repeats (promiscuous and non-promiscuous) provide enough energy to stabilize the PUF protein on the non-target RNA. Additionally, as PUF repeats can be promiscuous, the non-target RNA binding of longer $PUF2_9$ and $PUF2_{10}$ begins to make sense as the potential for such promiscuous binding would increase due to the increased number of binding repeats. In the future, competition FP assays between target and non-target cognate sequences for $PUF2_9$ and $PUF2_{10}$ detected here should be carried out to quantitatively assess the preference of the $PUF2_8$, $PUF2_9$ and $PUF2_{10}$ proteins for these sites. The presence of complete mismatches in motifs compared to the designed target sequence of $PUF2_9$ and $PUF2_{10}$ begs the question of whether they would even bind to their designed target sequences. This is a question of how much of the required binding energy for the PUF protein is provided by those particular repeats. Furthermore, these positions may simply appear as complete mismatches in these experiments because the particular designed target RNA was not available to bind in the cell,

and thus was not enriched by OOPS and work in Chapter 7 shows that the 10-mer PUFs can indeed bind their designed target cognate sequences suggesting this may be the case. *in vitro* binding studies as mentioned earlier would also help further answer this question and verify this hypothesis. Furthermore, the MEME motif analysis in these experiments does not account for contextual binding of the PUF domains, i.e., how binding to a given nucleotide is affected by its neighbouring nucleotides. Additional bioinformatics should be carried out to assess whether a contextual motif search yields different results.

Thus, we have found that as PUF target length increases, binding promiscuity significantly increases, and the proteins are able to bind many more potential cognate sequences rather than simply the cognate sequence they were designed to target. This huge increase in potential endogenous non-target cognate sites may overcome the theoretical reduction in the natural frequency of sites as PUF length increases (which was originally the hypothesis to explain the observations seen in previous Chapters). Thus, since $PUF2_9$ and $PUF2_{10}$ have so many more potential cognate sequences than $PUF2_8$, they may undergo sequestration in the cell on these non-target cognate sequences[90,104] which would reduce the available pool of these PUF proteins to bind transfected RNA. This phenomenon has also been suggested by Abil *et al.*[90] and is consistent with the decrease in split-GFP intensity when using 9- and 10-mer PUFs. This hypothesis should be explored further in future studies using competition FP assays and high-throughput *in vitro* binding assays, to test the sheer number of potential binding RNAs detected by NGS, such as HiTS-Kin[186,187].

HiTS-Kin is a method that tests an RNA-binding protein against all possible RNA sequences in a single reaction. For example, the 8-mer PUF protein would be tested for binding against all potential 8-mer RNA sequences (generated through *in vitro* transcription or synthesis) at different protein concentrations. Once the binding reaction is equilibrated, the bound RNA, at the given protein concentrations, is separated and used for NGS library preparation. Sequencing is subsequently carried out and the data can be used to calculate kinetic rate constants and binding affinities for all the randomised 8-mer sequences tested[188]. This would allow a study of the affinity landscape of the PUF proteins for all potential 8-, 9- and 10-mer RNAs, along with an analysis of pairwise interactions of multiple deviations from the designed target. In this way, it would delineate the NGS data into affinity data and provide a clearer

picture of exactly how many RNA species can indeed be bound by the 8-, 9- and 10-mer PUFs. For example, an RNA containing all the partial mismatches suggested by NGS may not bind the PUF due to a lack of sufficient binding energy, but some numbers of partial mismatches are clearly tolerated. Thus, elucidating the affinity map for these proteins would be crucial in contextualizing the NGS data and verifying the sequestration hypothesis in relation to the reduction in frequency of natural binding sites in the transcriptome hypothesis.

The data here also do not explain why the $PUF2_8$-$HIVPR_{D29N}$ fusion was active by itself since $PUF2_8$ was not found to be significantly promiscuous. Thus, it likely the case that multiple GCTGCTGC binding targets for $PUF2_8$ are located in proximity naturally in the cell (confirmed by GGRNA – a tool to search for short RNA motifs in the transcriptome) allowing $PUF2_8$-$HIVPR_{D29N}$ to activate. Although this method has been able to detect the binding site of $PUF2_8$ that was validated *in vitro*, the robustness should further be verified using gold standard CLIP-based technologies to check whether the same non-target cognate RNAs and promiscuity are detected for the three PUF proteins (using both methods), and that any additional artefactual motifs detected for each PUF protein due to the bioinformatics pipeline, or lack of an RNase digestion step in the protocol, are indeed artefacts. Lastly, these experiments should also be carried out for the other PUF proteins used in this thesis including those fused to effector domains to ensure these results can be generalised. Since binding of individual PUF domains has characterised by OOPS-seq, the next steps were to elucidate how this affects concurrent binding of the PUF domains using split-GFP.

6.4.2 SPLIT-GFP FUSIONS ARE ACTIVE IN THE CELL

PUF-fused split-GFPs showed reconstitution, albeit there was no difference between the cognate and scrambled RNAs. Reconstitution in the scrambled condition was unexpected from literature claiming the PUF proteins are extremely specific. However, these studies were either carried out *in vitro*[90,93,102] or used a different method to generate the designer PUF domains[103]. Studies by Yamada *et al.*[169] and Tilsner *et al.*[170] confirms the observation that in a PUF-split-GFP system, the presence of non-cognate RNA can indeed result in GFP reconstitution, suggesting that there is a degree of non-specific binding by the PUF domains.

Even with a flexible linker, no fluorescence was observed in the wild-type constructs (Figure 6.14), indicating that any re-constituted GFP fluorescence is not due to random encounters between the split halves in the cell, but rather is due to PUF-RNA-induced dimerization. This result is consistent with other studies that have shown that split-GFP halves do not re-constitute any activity unless they are fused to the PUF domains[169,170]. In contrast, some fluorescence was observed for the GST-fused split-GFP constructs, however, the intensity was low compared to the intensity seen for the PUF-fused GFPs suggesting that although the flexible linker is important in maintaining GFP domain activity, GST may not be a good positive control for dimerization as observed in Chapter 5 using the HIVPR-based system as well.

6.4.3 UNDERSTANDING THE VARIABLES INFLUENCING PUF-INDUCED SPLIT-GFP RECONSTITUTION

6.4.3.1 8-MER PUF DOMAINS

Initially, creating spacers of different lengths between the two PUF binding sites was tested, as Qiu *et al.*[93] reported that spacing can increase the binding affinity of two PUF domains in tandem. No increases in the diffuse green fluorescence were observed when the spacer length was increased, and the fluorescence obtained for all spacer lengths was indistinguishable from that of the scrambled control (Figure 6.16). To assess whether this was caused by low overall split-GFP fluorescence, multiple binding site (MBS) versions of the cognate RNAs were generated to increase the number of re-constituted split-GFP domains and thereby the GFP fluorescence (Figure 6.26).

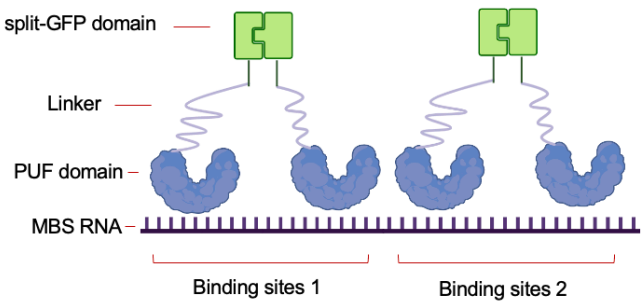


Figure 6.26: Schematic of expected PUF-split-GFP binding on MBS RNA.

Indeed, a slight increase in diffuse green GFP fluorescence is observed when using the MBS RNA compared to the SBS RNA (Figure 6.16) suggesting that the PUF fusions are not fully saturated using the SBS RNA. Although using MBS RNAs increases the fluorescence observed, which may be clinically relevant for cancer-overexpressed RNAs containing repetitive elements, such as *HSATII* in colorectal and pancreatic tumours[189,190], it is less important for single element overexpression such as in the *hTERT* RNA.

Using inside-oriented PUF-split-GFP fusions resulted in much higher GFP fluorescence for all SBS RNA spacer lengths compared to the outside-oriented fusions (Figure 6.17). In addition, increased GFP fluorescence when using the MBS RNA was also observed relative to SBS RNA for the inside-oriented fusion proteins combined with some of the RNA spacers (C0, C2, C6 and C10). The fact that this increase was not observed for all RNA spacer lengths suggests that in order to increase the binding interaction between PUFs and RNA, it is not necessarily sufficient to simply increase the number of sites available (explored further in Chapter 7). Thus, based on these data it seemed that orienting the split-GFP domains towards each other improved reconstitution likely due to more favourable orientations and proximity when facing inwards. Although there was some variability in split-GFP reconstitution between the cells for each set of constructs tested, these data provide a good foundation for further characterisation using flow cytometry and the split-TEV system (to quantitatively assess concurrent binding and reconstitution for much larger cell numbers) to reinforce these results since more reconstitution was observed in several cases using the cognate RNA compared to the scrambled RNA suggesting that concurrent binding of PUF fusions is occurring.

6.4.3.2 9- AND 10-MER PUF DOMAINS

For the 9- and 10-mer PUF domains, the striking reduction in the GFP fluorescence (Figures 6.18 – 6.21) suggests that these PUF domains concurrently target fewer RNAs compared to the 8-mer PUFs (as expected potentially due to the reduced natural frequency of these sites in the cell).

However, the reduction in fluorescence intensity may also be due to increased sequestering of individual 9- and 10-mer PUF fusions on endogenous non-target cognate RNA resulting in less

non-specific binding to exogenous RNA and therefore reduced GFP intensity (effects on specific binding cannot be determined from this data since GFP intensities were too low). The sequestering is due to increased promiscuity of longer PUF proteins as suggested in Zhao *et al.*[102] and supported by earlier OOPS-seq data in this Chapter of their increased non-target cognate binding. Although it may appear that an increase in endogenous non-target cognate binding of single PUF fusions would increase their ability to concurrently bind endogenous non-target cognate sites (and therefore increase GFP intensity), there will be a balance between PUF fusions being sequestered individually and concurrently. Additionally, there would also be a balance between the increase in concurrent endogenous non-target cognate binding for PUFs and the change in pool size of available fusion proteins to bind exogenous RNA. Although the increased endogenous concurrent binding could increase the GFP fluorescence, the consequent decrease in pool size of available fusion proteins to bind exogenous RNA would result in less binding of the fusion proteins to the overexpressed RNA which may have a greater effect, resulting in the overall reduction in observed GFP intensity.

However, this is simply a hypothesis, and the differences could also simply be due to the fact that the 9- and 10-mer fusions have weaker affinities for RNA. Thus, further studies are required to dissect the observed effects such as *in vitro* binding studies using the 9- and 10-mer PUF domains to elucidate their binding affinities in unfused and fused forms. Differences in specific binding in cognate conditions cannot be ascertained due to low fluorescence intensities. Although some differences were noted using the 9- and 10-mers, these differences were much smaller than for the 8-mers due to the significantly lower overall fluorescence intensity.

6.4.4 CHALLENGES OF USING CONFOCAL MICROSCOPY

As mentioned previously, studies of the PUF-split-GFP system have used super-resolution microscopy[168–170]. In order to detect the low split-GFP fluorescence using confocal microscopy (including for the 8-mer PUF domains), non-standard settings on the confocal microscope had to be used such as a high laser power, gain and pinhole width. Despite this, any potential differences between constructs for the 9- and 10-mer PUF fusions were below the sensitivity of the confocal. When attempting to quantify the fluorescence intensity to detect these smaller

differences for 9- and 10-mer fusions, highly variable results were obtained which showed no differences between conditions. The variability is likely due to artefacts generated due to the unusual microscope settings used, such as autofluorescence (seen clearly in 9- and 10-mer PUF fusion images). Additionally, using these settings also increased the intrinsic issues with confocal microscope image quantification[191] (such as fluctuations in laser power and diffraction artefacts), which led to an inability to accurately quantify any changes. Despite these challenges, the simultaneous binding of PUF-split-GFP fusions for the 8-mer PUF domains (showing these fusions can indeed bind RNA simultaneously in the cell), the improvement in reconstitution observed from inside-oriented split-GFP domains in the 8-mer PUF fusions relative to outside-oriented, and the reduction in diffuse GFP fluorescence when using 9- and 10-mer PUF fusions are all promising results.

6.4.5 QUANTITATIVE ANALYSIS OF FLUORESCENCE INTENSITIES

Both plate reader assays and flow cytometry were carried out to quantify any fluorescence differences between the brightest conditions observed using confocal microscopy. No GFP fluorescence was seen for any of the conditions tested using the plate reader assay (Figure 6.22), likely because the light source in the plate reader is a xenon flash lamp coupled to a monochromator, which produces a single wavelength output into the sample. This source has a significantly lower excitation power than the lasers used in the confocal microscopy setup.

In flow cytometry, the 8-mer PUF domains indeed show a population of cells containing both mCherry and GFP, however, the 9- and 10-mer PUF domains do not have this population which is consistent with the data obtained from the confocal microscopy (Figure 6.23). As with the microscopy, this is likely to be due to increased sequestering with the 9- and 10-mers resulting in less GFP reconstitution falling below the sensitivity of the flow cytometer. Interestingly, the dot plots for the 8-, 9- and 10-mer PUF domains showed no GFP reconstitution in mCherry-negative cells (Figure 6.23), suggesting that any reconstitution from naturally occurring concurrent binding sites is negligible. This result also indicates that the GFP fluorescence for the 8-mer scrambled RNA must be due to non-specific binding potentiated by overexpressed RNA rather than binding to natural concurrent PUF sites, and

therefore that the reduction in fluorescence when using 9- and 10-mer PUF domains is due to sequestering, rather than due to a reduction in the natural frequency of binding sites in the cell. This result further contextualises the promiscuity observed in the OOPS-seq experiments as well and suggests that in fact, the promiscuity of longer PUF domains leads to the aforementioned sequestration effect rather than reduced binding to natural binding sites in the transcriptome. However, the hypothesis that the 9- and 10-mers simply bind more poorly cannot be excluded from these data.

Although this Chapter showed that PUF-split-GFP fusions seem to be concurrently binding RNA in the cell, that the inside-oriented split-GFP domains result in an improvement in reconstitution of split-GFP relative to outside-oriented domains, and that the 9- and 10-mer PUF domains have less reconstitution than the 8-mer domains, it was not possible to quantify smaller differences due to both a lack of sensitivity of the instrumentation and setup used, and the overall low level of GFP reconstitution. The latter is likely because the split-GFP system is not amplificatory, i.e a single re-constituted GFP produces a single signal output. Thus, I moved to a system that would allow a single re-constituted split molecule to amplify its signal output.

CHAPTER 7 – SPLIT-TEV IN THE NASPER SYSTEM

7.1 INTRODUCTION

As the split-GFP system was not able to resolve the differences between the different optimization conditions tested, there was need to use an amplificatory system to create higher signals and thereby provide insight into the effects of different variables on the NASPER system. A split-protease system satisfies this criteria – one re-constituted protease would cleave multiple substrates resulting in an amplified signal, and additionally, it could be used in future assays to activate a pro-apoptotic zymogen to induce apoptosis as is required for the NASPER system. Hence a literature search into split protease systems was carried out to select the ideal protease. Two potential split proteases were identified: 1) split-TEV[55,166,192-194] and 2) split-HRV3C[165]. The two proteases have similar mechanisms of actions, as they are both cysteine protease, and they also share stringent substrate-recognition specificities[195,196]. Although Wang *et al.*[165] showed that split-HRV3C outperforms the split-TEV protease, their study was carried out in *E. coli*, whereas all studies of the split-TEV protease have been carried out in mammalian cells. Furthermore, the split-TEV protease has also been used successfully to activate a zymogen procaspase-3[55] resulting in apoptosis which is the end goal of the NASPER system. Lastly, other studies optimizing the split-protease itself using site-directed mutagenesis have been carried out[197] making it easier to troubleshoot and adapt the system.

7.1.1 THE TEV PROTEASE

The TEV protease, derived from the Tobacco Etch Virus, is a 27 kDa cysteine protease with stringent substrate specificity[198] (Figure 7.1)

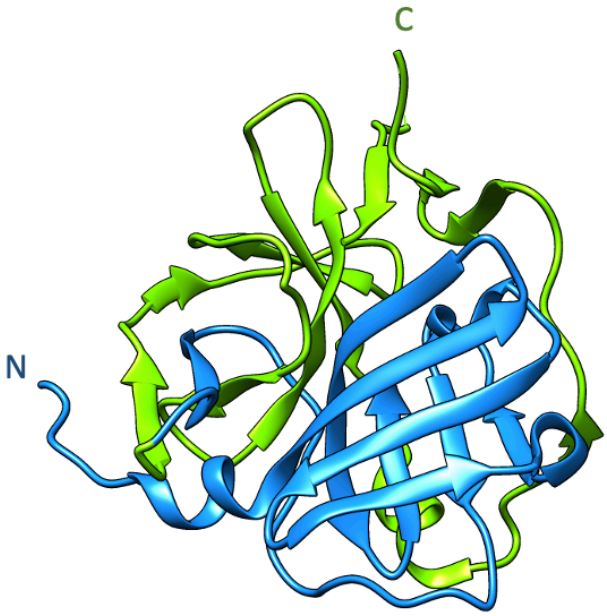


Figure 7.1: Schematic of the crystal structure of TEV Protease (PDB: 1Q31). The N-TEV structure is shown in blue, whereas the C-TEV structure is shown in green. The split point is between amino acids 118 and 119.

The protease recognises a consensus peptide sequence E-X-X-Y-X-Q-G/S (where X is any residue) and cleaves after the glutamine residue resulting in the glycine or serine residue becoming the new N-terminus of the protein[196,198]. The amino acids occupying the X positions are important in determining the rate of cleavage, and the most common sequence used is ENLYFQG[196]. This high specificity is why the TEV protease is commonly used *in vitro* as a tool for removing affinity tags after protein purification[199]. However, the wild-type TEV protease is relatively unstable, as it auto-cleaves to form a truncated protease with greatly diminished catalytic activity[200-202]. Mutagenesis resulted in the creation of the S219V variant, which has a significantly higher stability and catalytic activity and was shown to be active in *E. coli* cells. The Itzhaki lab regularly using this variant of TEV for *in vitro* protein production, and for these reasons this is the variant that will be split for use in the NASPER system in this Chapter.

7.1.2 SPLITTING THE TEV PROTEASE

Several studies have already been conducted using the split-TEV system. *Wehr et al.*[166] carried out the first study in determining an optimal split sites within the full length protein. Systematic screening identified candidate split sites, which were then verified using GBR1acc and GBR2cc domains. GBR1 and GBR2 are two $GABA_B$ receptors that contain coiled-coil domains which interact to form heterdimers[203]. Ultimately, the optimal split site was found to be after amino acid 118 resulting in N-TEV and C-TEV fragments that are able to re-constitute around 30% - 40% activity of the full TEV protease. The group went on to further validate the use of this split-TEV system in various cell-based models using both soluble and transmembrane-tethered forms of the domains. After validation, the split-TEV domains were used to monitor the biological interaction between the ErbB2 and ErbB4 receptor tyrosine kinases (RTKs) upon Nrg-1 ligand addition. The ligand-induced dimerisation of the split-TEV-fused RTKs at the cell membrane released a GAL4-VP16 protein which could then translocate to the nucleus and drive the expression of luciferase whose luminescence was the assay output[166]. Differences between Nrg-1 isoforms were successfully detected using this assay. Figure 7.2 shows the principle of the split-TEV system used by *Wehr et al.*[166] and how it would be used in the NASPER system; X and Y would represent the PUF1 and PUF2 RNA binding domains.

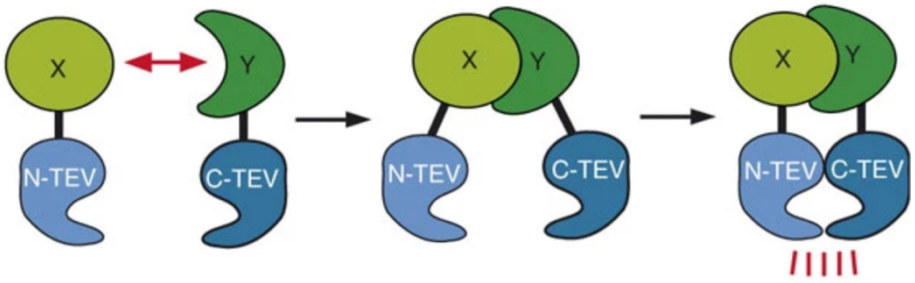


Figure 7.2: Schematic of the principle of the split-TEV assay used by their study to determine the optimal split site for the TEV protease. Adapted from Wehr et al.[166].

Wehr *et al.*[194,204] then moved on to apply this system to more transient biological protein-protein interactions such as the phosphorylation-dependent interaction between the Bad protein and adaptor protein 14-3-3ε, and Nrg-1 stimulated ErbB4 interactions with several other adaptor proteins. The Estévez group[193] used this system to characterise the interaction

between the $G_{\alpha i}$ subunit and its partner G-protein coupled receptor. Upon interaction, the transcription factor GV activates the expression of sGluc producing a measurable light signal. Most importantly this system has been used by Gray *et al.*[55] to activate a procaspase-3 zymogen and induce apoptosis in cells. They also optimised the system further and found that slightly truncating the C-TEV domain resulted in less background activity without reducing specific activity (this is the C-TEV domain that was used in my experiments). Using the procaspase-3 protein, they were able to effectively induce apoptosis over the short-term in mammalian cells as measured by DEVDase activity and cell staining for apoptotic markers. Our experiments will be for a longer duration than those carried out by this group, as they require transfection and expression of the constructs rather than simply small molecule treatment. However, it is a good foundation as it matches our NASPER approach closely.

7.1.3 ASSAY DESIGN, AIMS AND OBJECTIVES

Thus, the aim of this Chapter was to develop a more sensitive system than the split-GFP system and thereby optimize constructs and identify the lead split-TEV candidates to take forward in apoptosis assays. To achieve these aims, the mCherry construct that was used to deliver the cognate and scrambled RNAs was modified to contain a TEV cleavage site. Cleavage of this mCherry protein would then be detected and quantified using Western blotting. In this way, cells containing the substrate for TEV must also contain the RNA for the PUF fusions to bind and therefore, any changes in cleavage detected between split-TEV-only and PUF-split-TEV in cognate and scrambled RNA can be attributed to differences in PUF-RNA binding and split-TEV reconstitution (i.e, there is no background interference from a population of cells containing the mCherry substrate but lacking the RNA). Additionally, these changes can be normalized to total substrate present, which would allow an even more robust analysis independent of any variations in transfection efficiency.

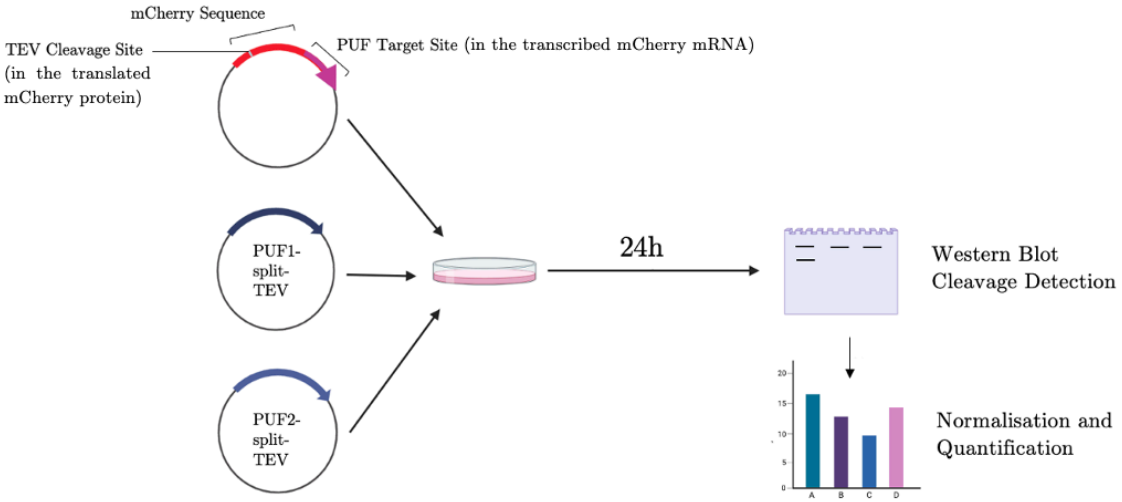


Figure 7.3: Schematic of assay design to detect activity of the split-TEV system

7.2 RESULTS

7.2.1 EXPRESSION OF SPLIT-TEV FUSION PROTEINS

The expression of the various split-TEV constructs, generated for initial testing of split-TEV activity, in HEK293 cells was assessed by Western blot prior to starting the TEV cleavage assay (Figure 7.4).

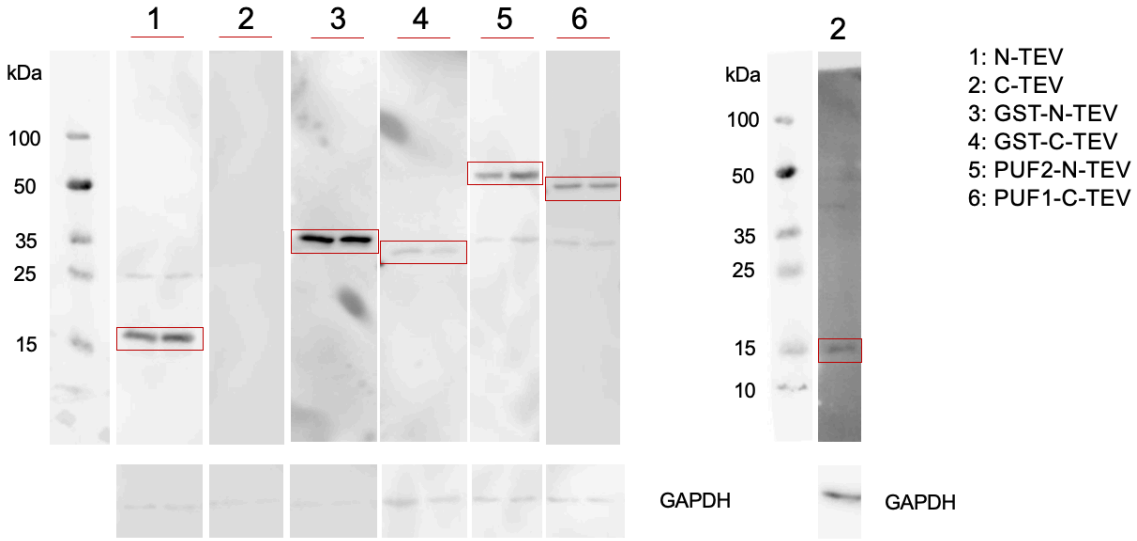


Figure 7.4: Western blot of split-TEV constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG monoclonal or anti-HA monoclonal primary antibody, and subsequently with an anti-Mouse or anti-Rat HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Red boxes are used to highlight bands of interest. GAPDH was used as a loading control. The two lanes under each title represent the two independent biological replicates.

Bands can be seen corresponding to the appropriate molecular weights of the constructs (Table 7.1). C-TEV was not detected initially, but upon re-running the sample with a shorter gel running time and without cutting off the end of the gel, it was seen at approximately the appropriate molecular weight.

Construct Name	Approximate Molecular Weight (kDa)
N-TEV	17
C-TEV	15
GST-N-TEV	36
GST-C-TEV	34
PUF2-N-TEV	51
PUF1-C-TEV	50

Table 7.1: Approximate molecular weight of split-TEV constructs.

mCherry constructs containing a 3xFLAG tag, TEV cleavage site, a 3xHA site and either the cognate or scrambled RNA sequences for the PUF1 and PUF2 proteins, were also generated and expressed.

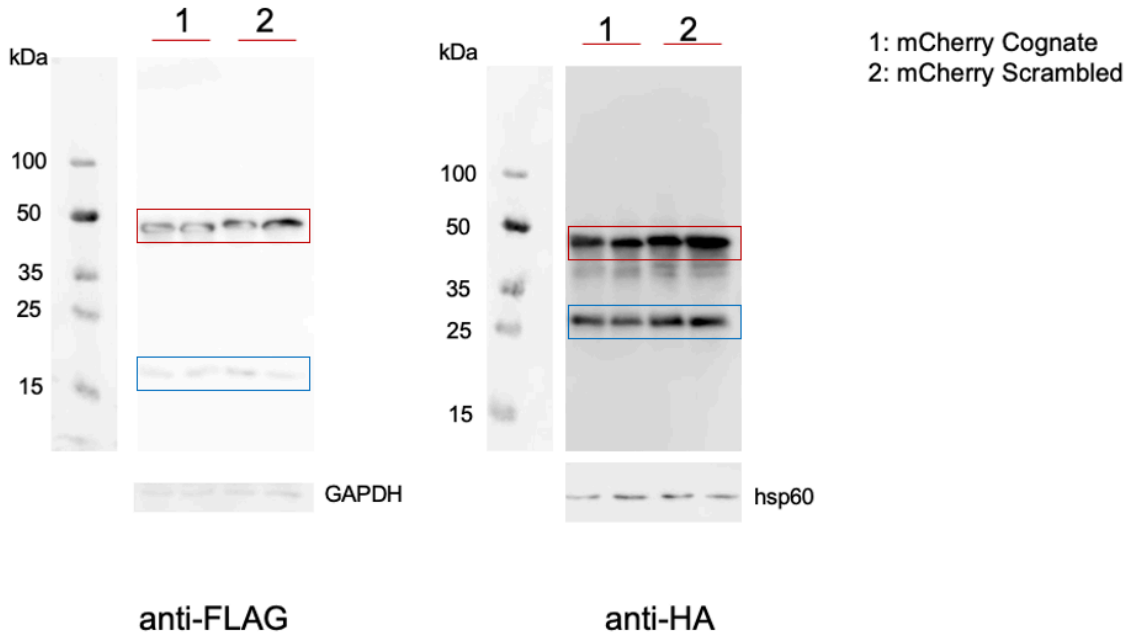


Figure 7.5: Western blot of TEV-cleavable mCherry. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG monoclonal or anti-HA monoclonal primary antibody, and subsequently with an anti-Mouse or anti-Rat HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. GAPDH or hsp60 was used as a loading control. Red boxes are used to highlight bands of interest (full-length mCherry protein), whereas blue boxes highlight degraded mCherry protein. The two lanes under each title represent the two independent biological replicates.

Figure 7.5 shows the expression of these constructs probed using either an anti-FLAG or anti-HA antibody to ensure detection by both. Both the cognate and scrambled mCherry showed the same pattern of cleavage. On the anti-FLAG blot, two bands were observed in each sample – one at approximately 48 kDa and one at approximately 18 kDa. On the anti-HA blot two bands were observed again in each sample – one at around 48 kDa and the other at around 27 kDa. It should be noted that the sum of molecular weights of the lower bands from the anti-FLAG and anti-HA blots add up to approximately the molecular weight of the higher band on both blots. The smearing observed below the 48 kDa band on the anti-HA gel is likely an artefact of the Western blot. Based on the molecular weights of the fragments observed on the Western blot, it was determined that cleavage was occurring somewhere within the mCherry protein itself, away from the TEV cleavage site. This phenomenon has been observed previously and is thought to be mCherry degradation within the cell[205]. Thus, because the TEV site was not thought to be affected, TEV cleavage could still be detected using this substrate. Hence, I moved on to testing it using split-TEV proteins lacking PUF domains, split-TEV domains fused to GST, and split-TEV domains fused to 8-mer PUF domains.

7.2.2 ACTIVITY OF SPLIT-TEV SYSTEM

Activity of the split-TEV system was then measured by the cleavage of TEV-cleavable mCherry in Western blots.

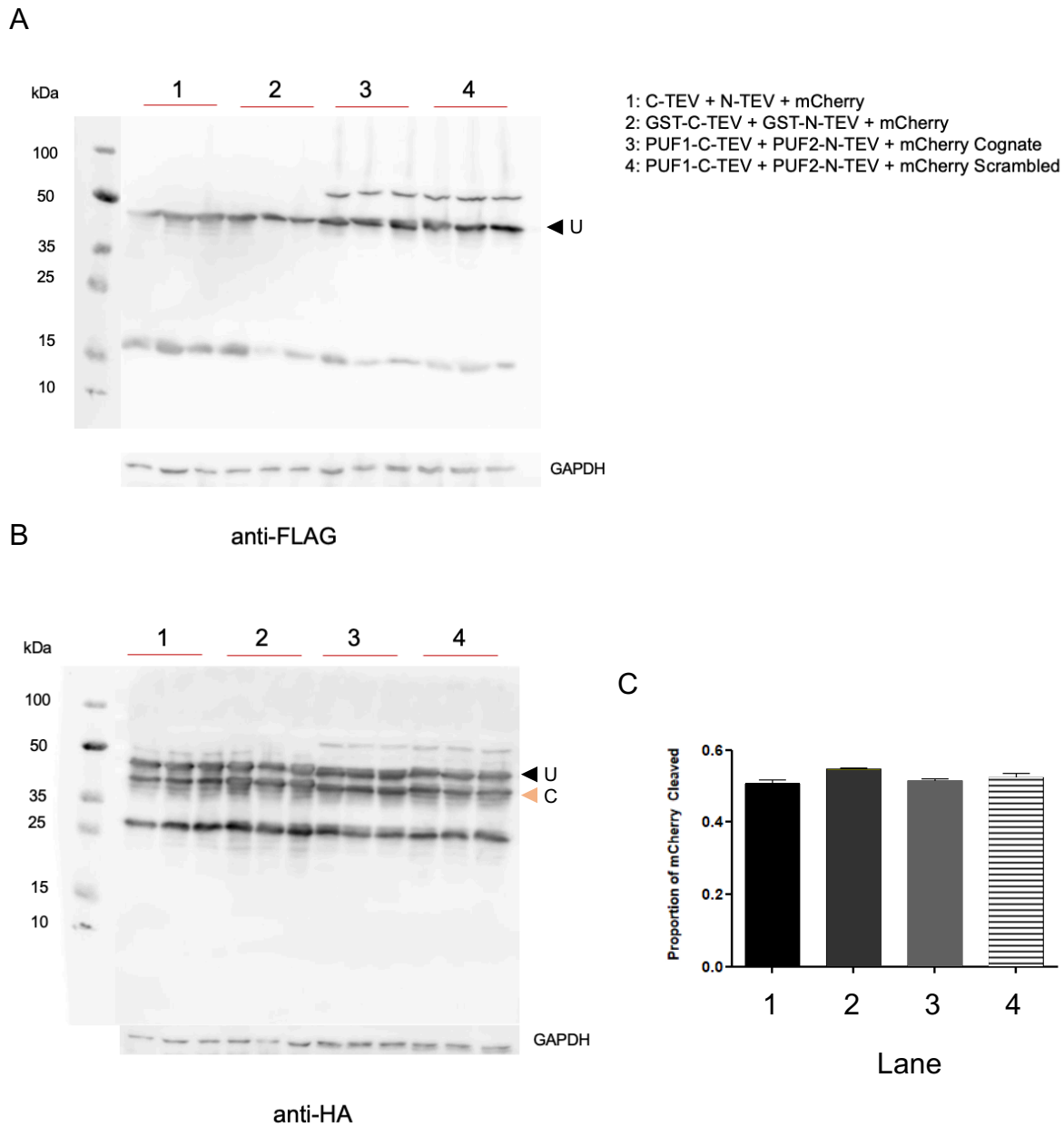


Figure 7.6: Western blot of triple transfected split-TEV constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG monoclonal **(A)** or anti-HA monoclonal primary antibody **(B)**, and subsequently with an anti-Mouse or anti-Rat HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. GAPDH was used as a loading control. The three lanes under each title represent the three independent biological replicates. Uncleaved and cleaved bands are marked as U and C respectively. **(C)**: The normalised proportion of mCherry cleaved as determined from the anti-HA blot is shown. Error bars represent the SEM of three independent biological experiments.

Figure 7.6 shows the results obtained after co-transfecting HEK293T cells with various different combinations of split-TEV constructs. Panel A shows the Western blot after being probed with an anti-FLAG primary antibody. Lanes 1 and 2 show the typical cleavage pattern for mCherry when probed with anti-FLAG as seen in Figure 7.5, whereas lanes 3 and 4 contain an additional band at a higher molecular weight corresponding to the PUF2-N-TEV constructs. There is no visible difference in band intensity of the full-length mCherry between any of the four combinations.

Panel B shows the anti-HA blot for the same samples as Panel A. The pattern seen across all combinations is typical for the anti-HA mCherry with bands at 48 kDa and 27 kDa. However, an additional band at approximately 42 kDa is seen across all samples, and a band at around 50 kDa is seen in samples 3 and 4 corresponding to the PUF1-C-TEV.

Panel C shows the quantification of cleavage as a proportion of total mCherry (a ratio of the 42 kDa, and 42 kDa + 48 kDa band (corresponding to cleaved and uncleaved, respectively) bands) from the anti-HA blot as this compensates for the loading and transfection efficiencies. No significant differences were observed between the different combinations, which suggests that any activity due to concurrent binding of the PUF domains and subsequent re-constitution of split-TEV as per the NASPER mechanism is masked by the non-specific re-constitution of the split-TEV domains. Thus, the system needed to be optimized such that NASPER activity could be increased and detected above this background re-constitution of split-TEV, and so that the activity is higher using cognate RNA versus scrambled RNA.

7.2.3 OPTIMISATION OF THE SPLIT-TEV SYSTEM

Various optimisations of the split-TEV system were carried out similarly to those described in Chapter 6 for the split-GFP (Figure 6.15) system to maximise activity compared to the scrambled control. First, the expression of these new constructs was tested. The blots also show the various constructs produced for the split-GFP system in the previous Chapter. These were run on the same blots and are hence shown together as they could not be separated well. The constructs generated for this Chapter are shown in Table 7.2, along with the PUF target sites fused to the 3'-end of the mCherry mRNA (Table 7.3), and an optimization schematic in

Figure 7.7. Similar to the split-GFP system, scrambled RNA 1 was used as the control for all assays as no differences in activity were observed compared to scrambled RNA 2 indicating RNA and spacer length do not influence non-specific binding (Appendix E). As seen in Figures 7.8 and 7.9, all new constructs expressed at the correct molecular weights as expected.

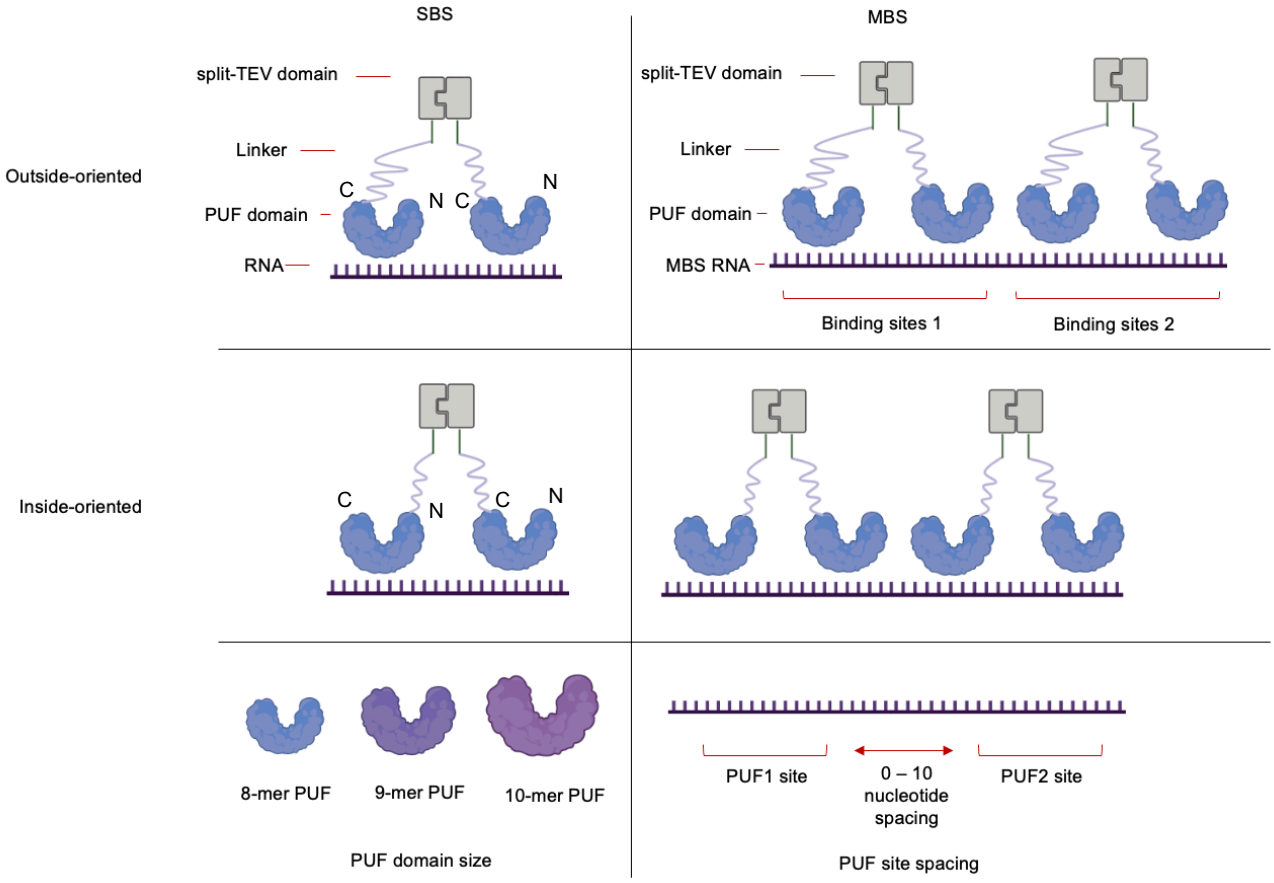


Figure 7.7: Schematic diagram of different parameters tested using the PUF-split-TEV system.

Construct Name	Protein Tag	Notes
$PUF2_8$-N-TEV	3xFLAG	-
$PUF2_9$-N-TEV	3xFLAG	-
$PUF2_{10}$-N-TEV	3xFLAG	-
$PUF1_8$-C-TEV	3xHA	Outside-oriented
$PUF1_9$-C-TEV	3xHA	Outside-oriented
$PUF1_{10}$-C-TEV	3xHA	Outside-oriented
C-TEV-$PUF1_8$	3xHA	Inside-oriented
C-TEV-$PUF1_9$	3xHA	Inside-oriented
C-TEV-$PUF1_{10}$	3xHA	Inside-oriented
mCherry SC0	3xFLAG and 3xHA	Cognate, SBS, 0-nt spacer
mCherry SC2	3xFLAG and 3xHA	Cognate, SBS, 2-nt spacer
mCherry SC4	3xFLAG and 3xHA	Cognate, SBS, 4-nt spacer
mCherry SC6	3xFLAG and 3xHA	Cognate, SBS, 6-nt spacer
mCherry SC8	3xFLAG and 3xHA	Cognate, SBS, 8-nt spacer
mCherry SC10	3xFLAG and 3xHA	Cognate, SBS, 10-nt spacer
mCherry MC0	3xFLAG and 3xHA	Cognate, MBS, 0-nt spacer
mCherry MC2	3xFLAG and 3xHA	Cognate, MBS, 2-nt spacer
mCherry MC4	3xFLAG and 3xHA	Cognate, MBS, 4-nt spacer
mCherry MC6	3xFLAG and 3xHA	Cognate, MBS, 6-nt spacer
mCherry MC8	3xFLAG and 3xHA	Cognate, MBS, 8-nt spacer
mCherry MC10	3xFLAG and 3xHA	Cognate, MBS, 10-nt spacer
mCherry Scrambled 1	3xFLAG and 3xHA	*
mCherry Scrambled 2	3xFLAG and 3xHA	*

Table 7.2: Summary of constructs generated to optimise split-GFP system. Scrambled constructs labelled with a * are discussed in Section 6.3.3. C0 – C10 in the construct names refer to the number of nucleotides between the PUF binding sites. All mCherry constructs contain a TEV cleavage site at the N-terminus and were generated as required for the 8-, 9- and 10-mer PUF domains.

Construct Name	Target Site		
8-mer mCherry SC0	**CUG—CGACA**	----------	*GCU--GCUGC*
8-mer mCherry SC2	**CUG—CGACA**	AA--------	*GCU--GCUGC*
8-mer mCherry SC4	**CUG—CGACA**	AAAA------	*GCU--GCUGC*
8-mer mCherry SC6	**CUG—CGACA**	AAAAAA----	*GCU--GCUGC*
8-mer mCherry SC8	**CUG—CGACA**	AAAAAAAA--	*GCU--GCUGC*
8-mer mCherry SC10	**CUG—CGACA**	AAAAAAAAAA	*GCU--GCUGC*
9-mer mCherry SC0	**CUGA-CGACA**	----------	*GCUA-GCUGC*
9-mer mCherry SC2	**CUGA-CGACA**	AA--------	*GCUA-GCUGC*
9-mer mCherry SC4	**CUGA-CGACA**	AAAA------	*GCUA-GCUGC*
9-mer mCherry SC6	**CUGA-CGACA**	AAAAAA----	*GCUA-GCUGC*
9-mer mCherry SC8	**CUGA-CGACA**	AAAAAAAA--	*GCUA-GCUGC*
9-mer mCherry SC10	**CUGA-CGACA**	AAAAAAAAAA	*GCUA-GCUGC*
10-mer mCherry SC0	**CUGAUCGACA**	----------	*GCUA-GCUGC*
10-mer mCherry SC2	**CUGAUCGACA**	AA--------	*GCUAUGCUGC*
10-mer mCherry SC4	**CUGAUCGACA**	AAAA------	*GCUAUGCUGC*
10-mer mCherry SC6	**CUGAUCGACA**	AAAAAA----	*GCUAUGCUGC*
10-mer mCherry SC8	**CUGAUCGACA**	AAAAAAAA--	*GCUAUGCUGC*
10-mer mCherry SC10	**CUGAUCGACA**	AAAAAAAAAA	*GCUAUGCUGC*
mCherry Scrambled 1	CAATCGAG	AAAAAAAAAA	CTCCGGGT
mCherry Scrambled 2	CAATCGAGCTCCGGGTCTTCTCAATCGAGCTCCGGGT		

Table 7.3: Summary of PUF target sequences fused to the 3'-end of mCherry mRNA. Scrambled constructs labelled with a * are discussed in Section 6.3.3. C0 – C10 in the construct names refer to the number of nucleotides between the PUF binding sites. MBS RNAs are not shown in this table as the font size requires to display them is too small. **Bold and underlined** = PUF1 binding site, *italics and underlined* = PUF2 binding site, underlined-only = scrambled site, unmodified text = spacer between the PUF sites.

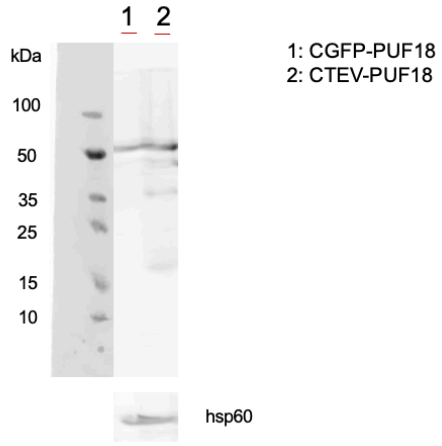


Figure 7.8: Western blot of inside-oriented PUF18 split-TEV and split-GFP constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody, and subsequently with an anti-Rat HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. hsp60 was used as a loading control

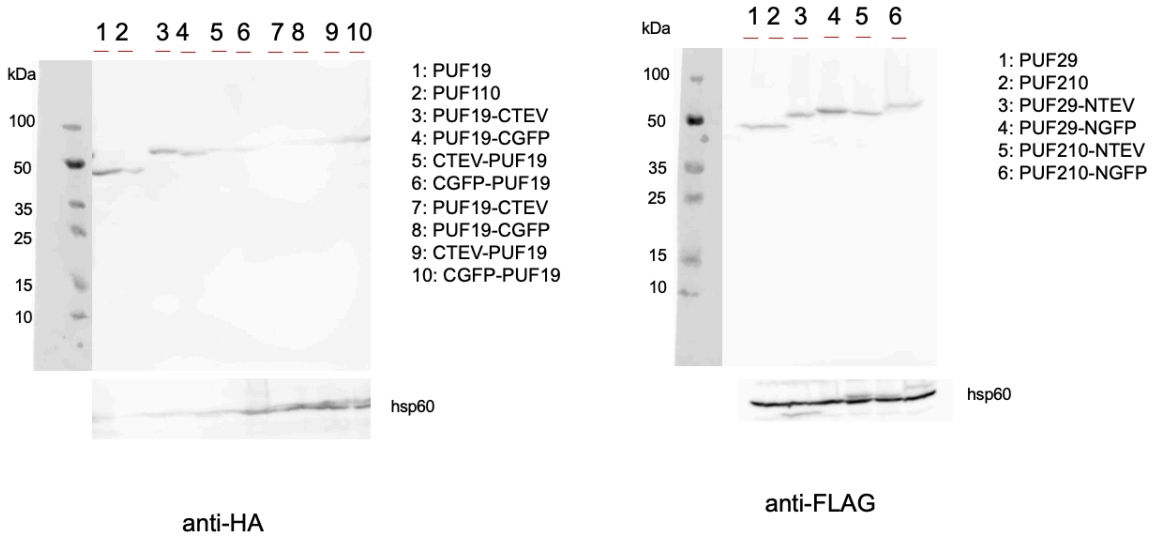


Figure 7.9: Western blots of PUF split-TEV, split-GFP constructs and non-fusion constructs using 9-mer or 10-mer PUF domains. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG or an anti-HA monoclonal primary antibody, and subsequently with an anti-Mouse or anti-Rat HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. hsp60 was used as a loading control

7.2.3.1 8-MER SITES

OUTSIDE-ORIENTED SPLIT-TEV DOMAINS

Once it was verified that all the constructs produced to optimise the split-TEV system expressed correctly, TEV activity assays could be carried out. Outside-oriented 8-mer PUF proteins were tested for activity with various spacer lengths in the RNA between the binding sites for the two PUFs as this has been shown previously to improve PUF concurrent binding and was therefore expected to increase split-TEV re-constitution. These assays were conducted using mCherry constructs containing a single binding site (SBS) of both PUF proteins. For all results obtained in this section, the data are provided as Western blot and its normalized quantification. Although differences between different spacers cannot be visualized qualitatively on the blot (Figure 7.10), the cleavage pattern seen is as expected from Figure 7.6. The quantitation shown in Figure 7.10 is also consistent with Figure 7.6 in that the C0 spacer does not have significantly increased cleavage over the scrambled control. In fact, the data suggest that none of the different spacers result in an increased cleavage proportion compared to the scrambled control. Thus, I attempted to change the orientation of the split-TEV domains to facilitate reconstitution.

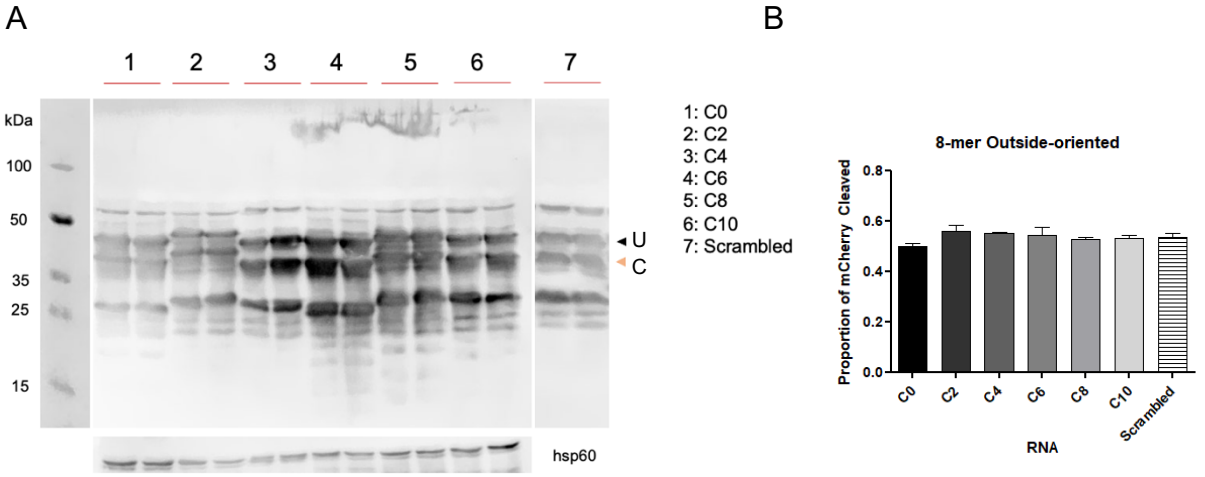


Figure 7.10: (A) Western blot of triple transfected outside-oriented split-TEV constructs with 8-mer PUF domains, and SBS RNAs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM of two independent biological experiments. Significance levels; * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are in comparison to scrambled control. Significance levels; # – $p < 0.05$, ## – $p < 0.01$, ### – $p < 0.001$ are in comparison between groups.

INSIDE-ORIENTED SPLIT-TEV DOMAINS

Inside-oriented 8-mer PUF proteins were tested as we hypothesised that due to the increased proximity of the split-TEV halves, activity may be improved. Comparisons were made between the activity of inside- and outside-oriented split-TEV domains (Figure 7.10).

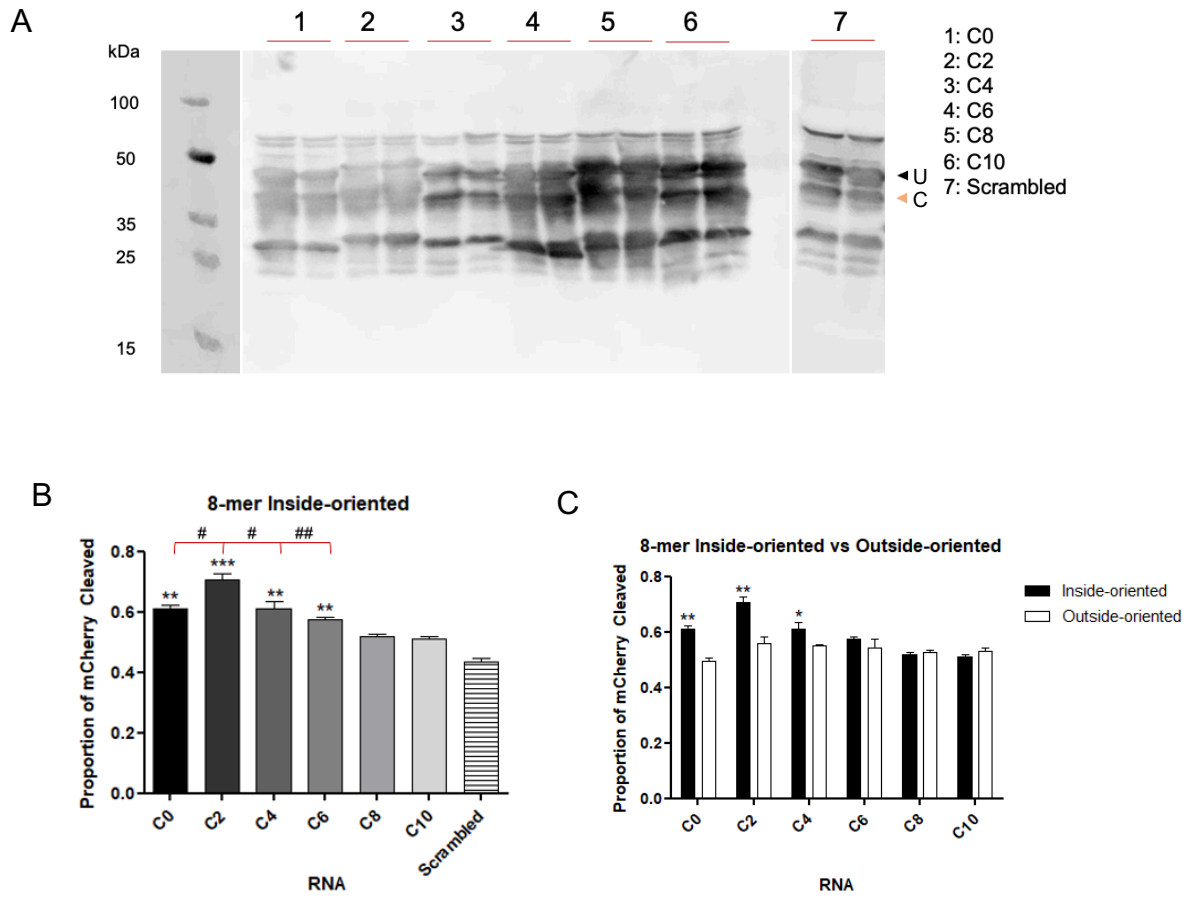


Figure 7.11: (A) Western blot of triple transfected inside-oriented split-TEV constructs with 8-mer PUF domains, and SBS RNAs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM of two independent biological experiments. **(C)**: The normalised proportion of mCherry cleaved for outside- versus inside-oriented split-TEV domains. Error bars represent the SEM of two independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are in comparison to control. Significance levels: # – $p < 0.05$, ## – $p < 0.01$, ### – $p < 0.001$ are in comparison between groups.

The inside-oriented split-TEV domains did indeed show enhanced cleavage activity (Figure 7.11). Cleavage using the C0 RNA is significantly increased ($p<0.01$) compared to the scrambled RNA control. The activity peaks for C2 RNA ($p<0.001$) and then shows a steady decrease in activity when the spacer length is further increased, and the Tukey's post-hoc test shows that these differences are significant. In particular, the activity when the C2 spacer is used is significantly higher than C0 ($p<0.05$), C4 ($p<0.05$) and C6 ($p<0.01$). Moreover, for the C0, C2 and C4 spacers, the inside-oriented split-TEV showed significantly higher activity ($p<0.01$, $p<0.001$, $p<0.01$ respectively) than the outside-oriented counterpart. Thus, due to the improved re-constitution of inside-oriented split-TEV domains, trends in PUF binding due to a change in spacer length also became apparent and correlate well with published data suggesting slightly increasing the spacer length improved PUF binding. The activity for the C2 spacer was also significantly higher than the activity using split-TEV domains alone and this set of constructs was selected as the lead candidate for NASPER. Although this was promising, further experiments were carried out using MBS RNAs to contextualise observations from the split-GFP system, and to assess whether the activity of the system could be increased even further.

In this regard, outside- and inside-oriented 8-mer PUF proteins with various spacer lengths were tested with an MBS of each PUF protein. Comparisons were made between the activity of SBS and MBS sites.

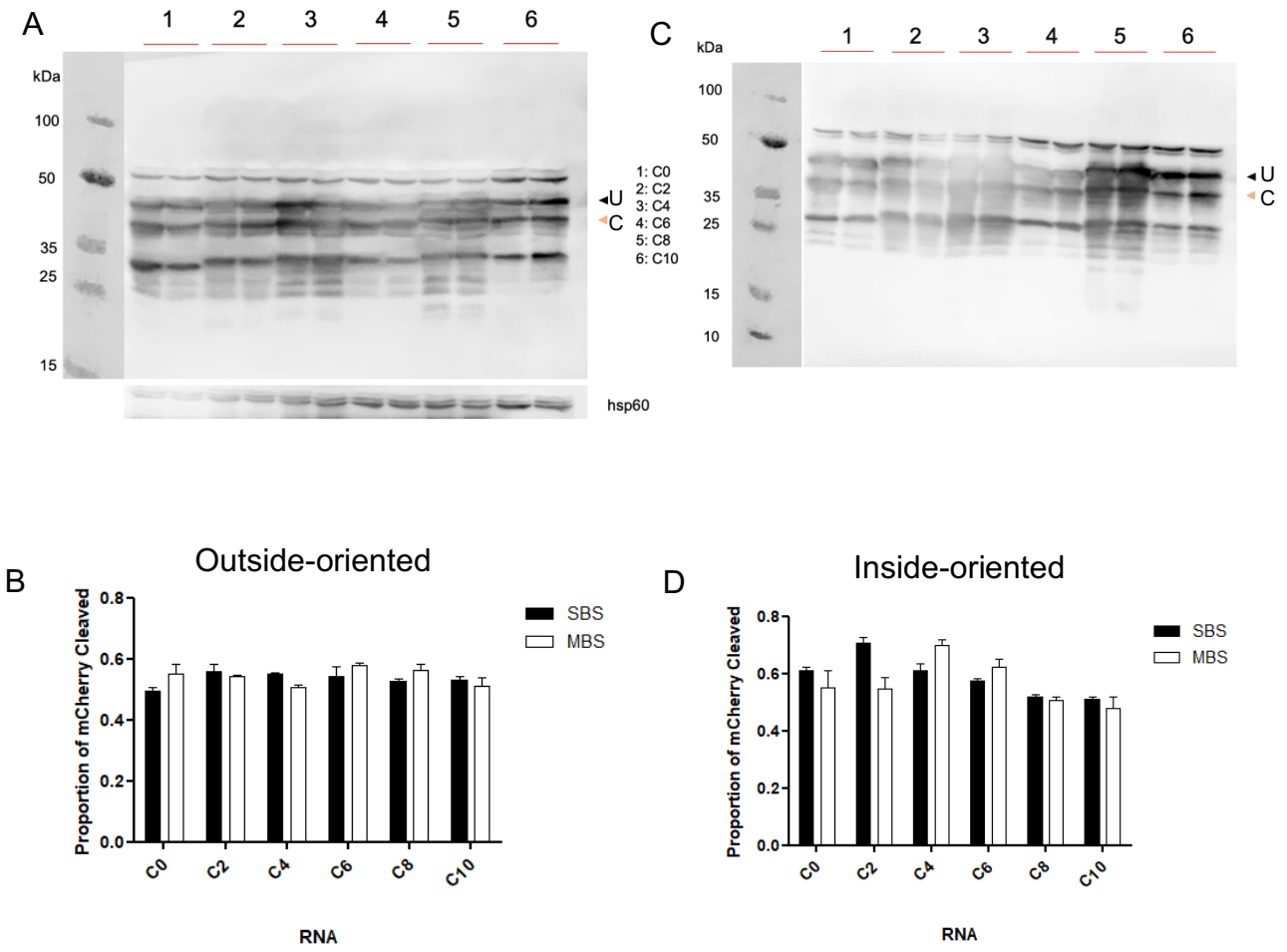


Figure 7.12: (A, C) Western Blot of Triple Transfected Outside- and inside-oriented split-TEV constructs with 8-mer PUF domains using MBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody, and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B, D)**: The normalised proportion of mCherry cleaved for the SBS versus MBS blots. Error bars represent the SEM from two independent biological replicates. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are comparisons within groups.

Figure 7.12 shows that in general, the MBS alternatives do not have greater activity than the SBS RNAs. The inside-oriented split-TEV with the MBS RNAs show that their activity peaks with the C4 RNA and decreases as spacer length is increased beyond this, which is similar to the trend observed using the SBS RNAs.

Spacing	Fold-change
C0	1.40
C2	1.90
C4	1.40
C6	1.32

Table 7.3: Fold-change in activity compared to scrambled control for 8-mer PUFs with inside-oriented split-TEV domains.

7.2.3.2 9-MER SITES

Although the data for the 8-mer PUF domains was promising, there was still background activity observed when using the scrambled RNA. From the split-GFP system, it seemed that using 9- and 10-mer PUF domains reduced this background activity, however, it was to such an extent that the microscopy was unable to detect differences between the cognate and scrambled RNA. Thus, 9- and 10-mer PUF domains were tested next using the split-TEV system to see whether similar reductions in background could be achieved while also detecting activity when using cognate RNA. The 9-mer PUFs tested yielded unexpected results compared to the 8-mer PUFs. The western blots in Figure 7.13 and show a similar cleavage pattern to the 8-mer PUF proteins, but the cleaved and uncleaved bands are less clear and there is laddering between the cleaved and uncleaved bands, and below the cleaved bands. Quantifying these blots showed that there were no significant differences between the cognate and scrambled sequences for any spacer length for outside- or inside-oriented split-TEV domains (Figure 7.13 and 7.14). There were also no significant differences in activity between the outside- and inside-oriented split-TEV domains.

OUTSIDE-ORIENTED SPLIT-TEV DOMAINS

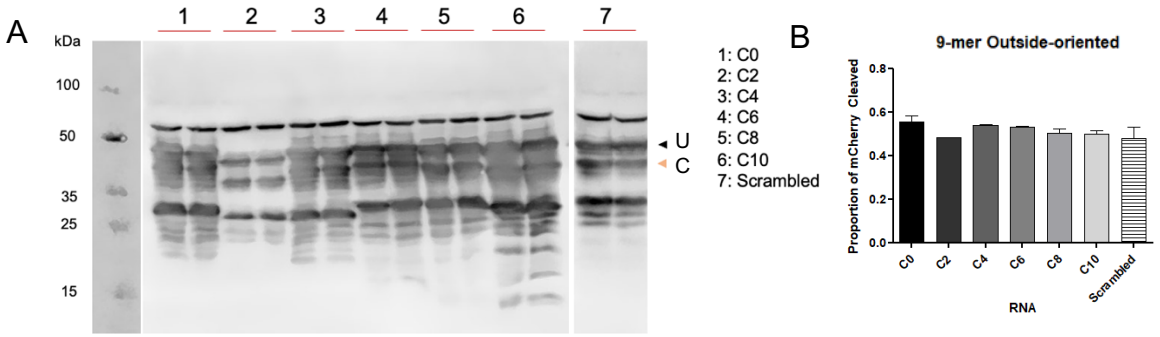


Figure 7.13: (A) Western blot of triple transfected PUF 9-mer outside-oriented split-TEV constructs using SBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM from two independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are in comparison to control. Significance levels: # – $p < 0.05$, ## – $p < 0.01$, ### – $p < 0.001$ are in comparison

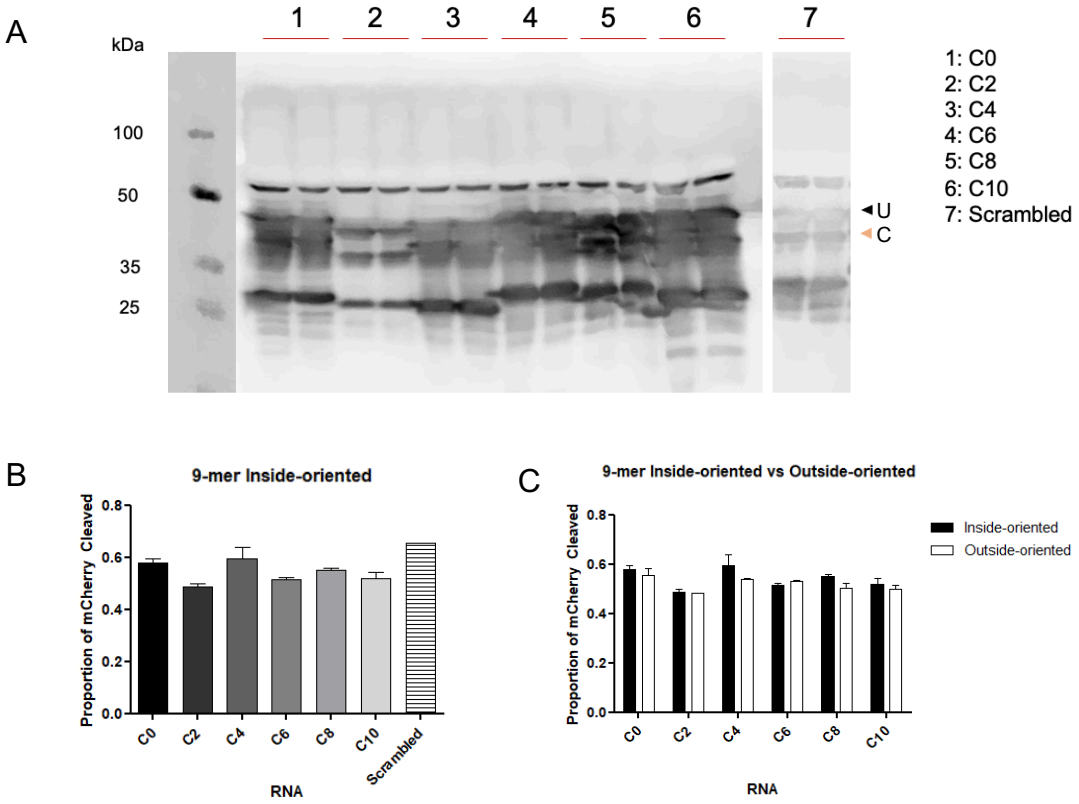


Figure 7.14: (A) Western blot of triple transfected PUF 9-mer inside-oriented split-TEV constructs using SBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM from two independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are in comparison to control. Significance levels: # – $p < 0.05$, ## – $p < 0.01$, ### – $p < 0.001$ are in comparison between groups. **(C)**: The normalised proportion of mCherry cleaved for the outside- versus inside-oriented split-TEV domains.

7.2.4 MBS RNA COGNATE SITES

Since no differences in activity was observed using SBS RNAs, I tried using MBS RNAs to see whether the number of re-constitutions can be increased and whether this can therefore increase activity enough to be detected over the activity from the scrambled RNA. The blots for the 9-mer MBS RNAs (Figure 7.15) are different from the corresponding 8-mer blots. There is significant laddering where the uncleaved and cleaved bands are expected to be to such an extent that the uncleaved and cleaved bands cannot be distinguished well from these unknown bands. Interestingly, the control hsp60 band was clearly resolved in both MBS blots with no unknown bands above or below. This pattern is also seen in the SBS blots, as mentioned, but to a much lesser degree. As such, the MBS blots could not be quantified reliably, and it is likely that the quantifications for the SBS blots are also unreliable due to extensive laddering. Potential explanations for this result are detailed in the Discussion section.

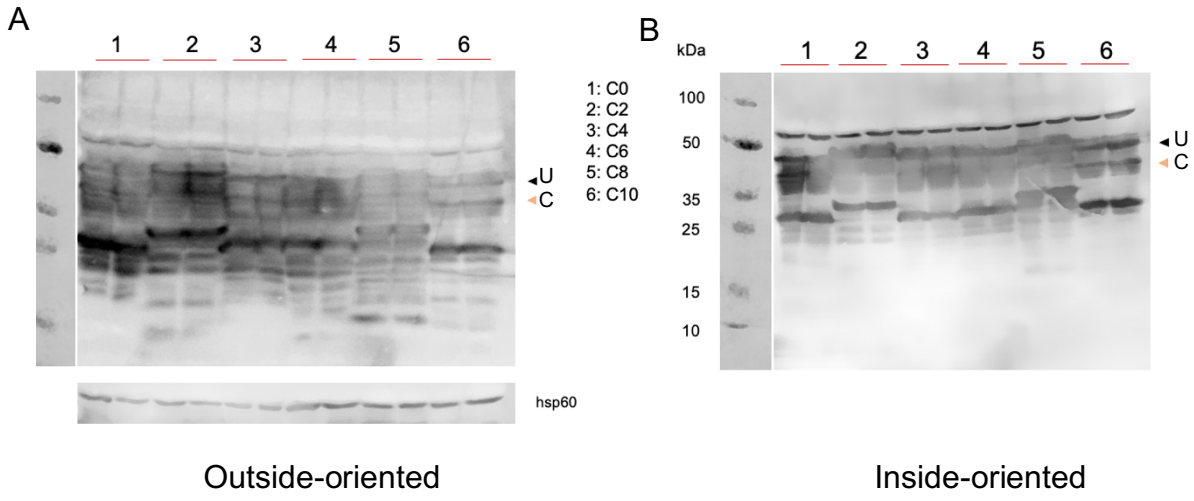


Figure 7.15: Western blot of triple transfected PUF 9-mer outside- **(A)** and inside-oriented **(B)** split-TEV constructs using MBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates.

7.2.3.3 10-MER SITES

Since the data and blots obtained using the 9-mer PUF domains were unreliable, I attempted to use 10-mer PUF domains instead as these had the same effects on re-constitution using scrambled RNA as the 9-mer PUF domains in the split-GFP system, and their blots may be more reliably quantifiable using the split-TEV system than the blots for the 9-mer PUF fusions.

OUTSIDE-ORIENTED SPLIT-TEV DOMAINS

Again, although no visual differences can be observed on the blots, the cleavage patterns seen are informative. All the blots in this section show slight smearing, between the uncleaved and cleaved bands, and below the cleaved band. This shading is more than seen in the 8-mer site blots, but less than in the 9-mer blots allowing for accurate quantification of the uncleaved and cleaved bands to be carried out. The quantification shown in Figure 7.16 (outside-oriented split-TEV) shows no significant differences between any spacer length and the scrambled control, however, a similar trend to the 8-mer inside-oriented split-TEV domains can be seen wherein the activity peaks at the C2 spacer and decreases consistently as spacer length increases. No decrease in scrambled activity was observed compared to the 8-mer PUFs.

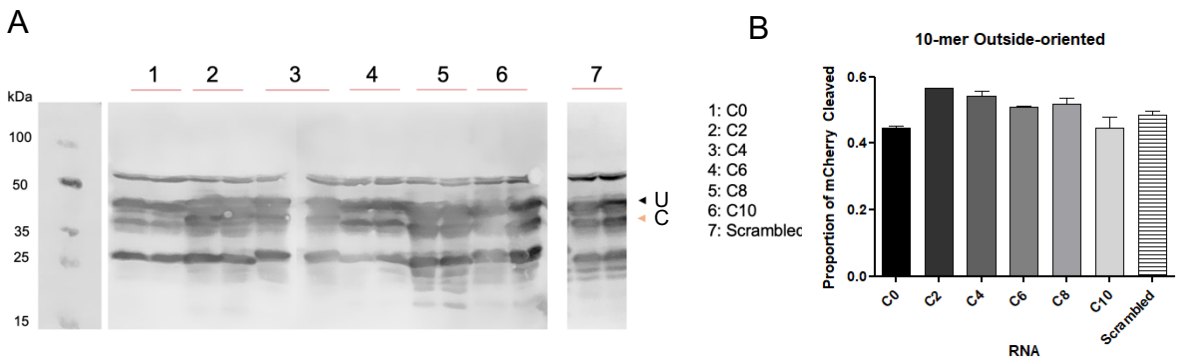


Figure 7.16: (A) Western blot of triple transfected PUF 10-mer outside-oriented split-TEV constructs using SBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM from two independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are in comparison to control. Significance levels: # – $p < 0.05$, ## – $p < 0.01$, ### – $p < 0.001$ are in comparison between groups.

INSIDE-ORIENTED SPLIT-TEV DOMAINS

Since no differences compared to scrambled RNA were detected using outside-oriented split-TEV domains, I tried using the more robust inside-oriented split-TEV domains to see whether there was indeed any concurrent PUF binding and re-constitution occurring using the 10-mer PUF domains. Inside-oriented split-TEV domains show a different trend to outside-oriented split-TEV (Figure 7.16 and 7.17C) in that activity increases slightly as the RNA spacer length is increased. Using a 10-nucleotide spacer showed significantly higher activity compared to the scrambled RNA control, however this was not taken forward as the activity was not higher than when split-TEV domains alone are used. There were no significant differences in activity between the outside and inside orientations except for when the C10 RNA spacer is used ($p<0.001$). No decrease in scrambled activity was observed compared to the 8-mer PUFs.

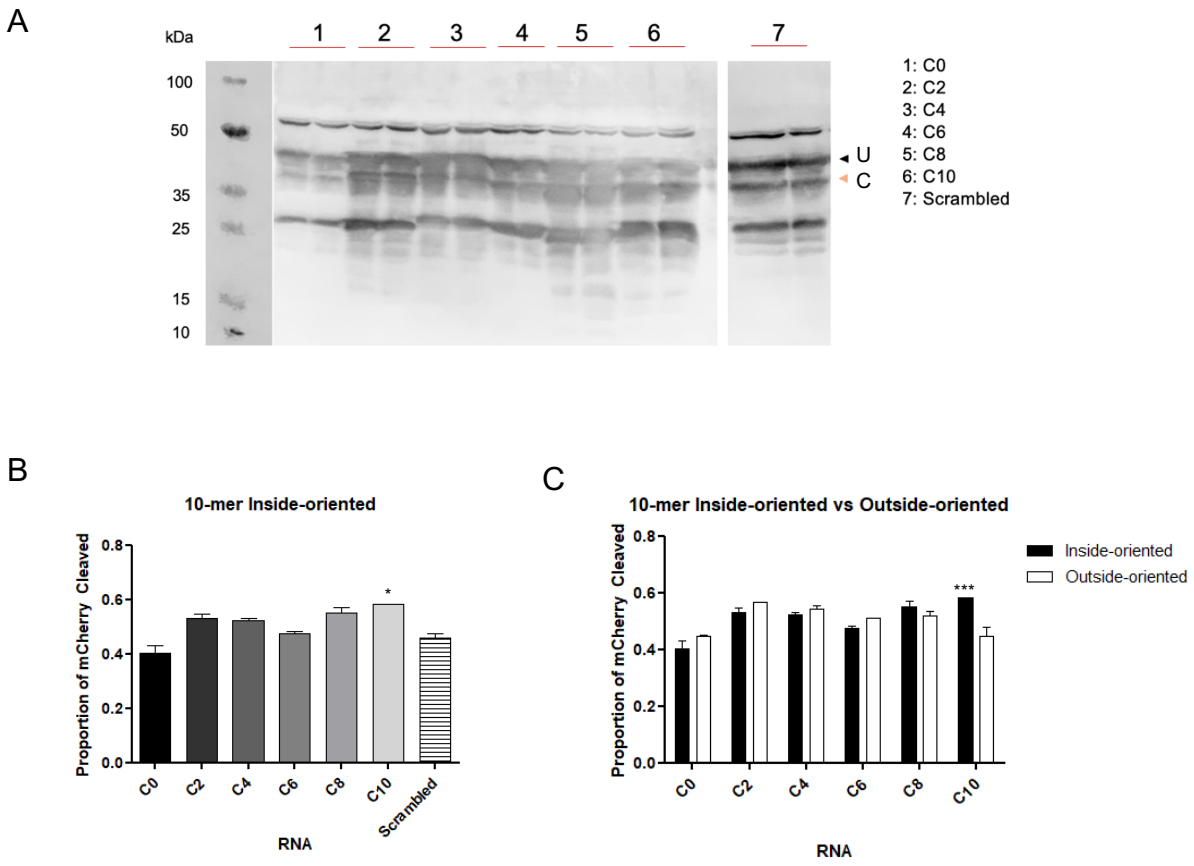


Figure 7.17: (A) Western blot of triple transfected PUF 10-mer inside-oriented split-TEV constructs using SBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM from two independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are in comparison to control. Significance levels: # – $p < 0.05$, ## – $p < 0.01$, ### – $p < 0.001$ are in comparison between groups. **(C)**: The normalised proportion of mCherry cleaved for the outside- versus inside-oriented split-TEV domains.

MBS RNA COGNATE SITES

Even though the 10-mer PUFs with inside-oriented split-TEV domains did show increased activity compared to scrambled RNA, this was lower than what was detected using the 8-mer PUF domains. MBS RNAs were used to assess whether this was because the PUF fusions were not being saturated and to increase the activity further. Figure 7.18 show that in general, the MBS alternatives do not have any different activity to the SBS for either the outside- or inside-oriented split-TEV domains. The exceptions to this are C2 spacer which when oriented either outwards or inwards has significantly higher activity in the SBS form compared to the MBS form.

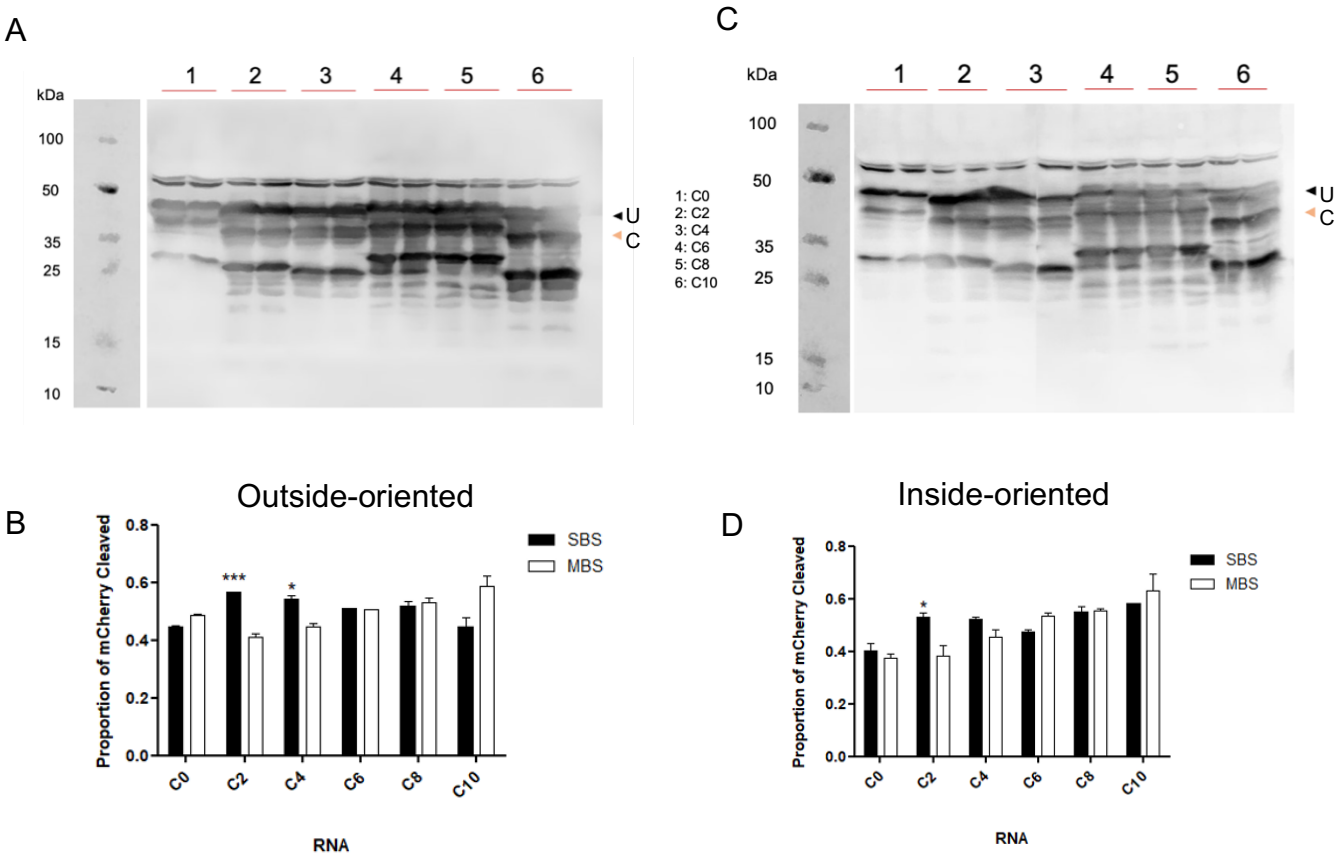


Figure 7.18: (A, C) Western blot of triple transfected PUF 10-mer outside- and inside-oriented split-TEV constructs using MBS mCherry constructs. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The two lanes under each title represent the two independent biological replicates. **(B, D)**: The normalised proportion of mCherry cleaved from the anti-HA blot is shown. Error bars represent the SEM from two independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$ are comparisons within groups.

7.2.4 MUTATIONAL OPTIMISATION OF THE SPLIT-TEV DOMAINS

Using 9- and 10-mers did not reduce activity when using scrambled RNA likely due to the spontaneous re-constitution of the split-TEV halves. Thus, mutation of the split-TEV domains was carried out in order to further optimize and reduce the background activity of the free split-TEV domains. The mutation pair (i.e, one mutation in each of N- and C-TEV fragments) with lowest activity without dimerisation induction was chosen, from a study by Dolberg *et al.*[197], to be H75S for N-TEV and I163V for C-TEV (amino acid numbers correspond to the appropriate amino acids from full length TEV protease). Western blots were then carried out to see if these mutations did in fact reduce the background activity.

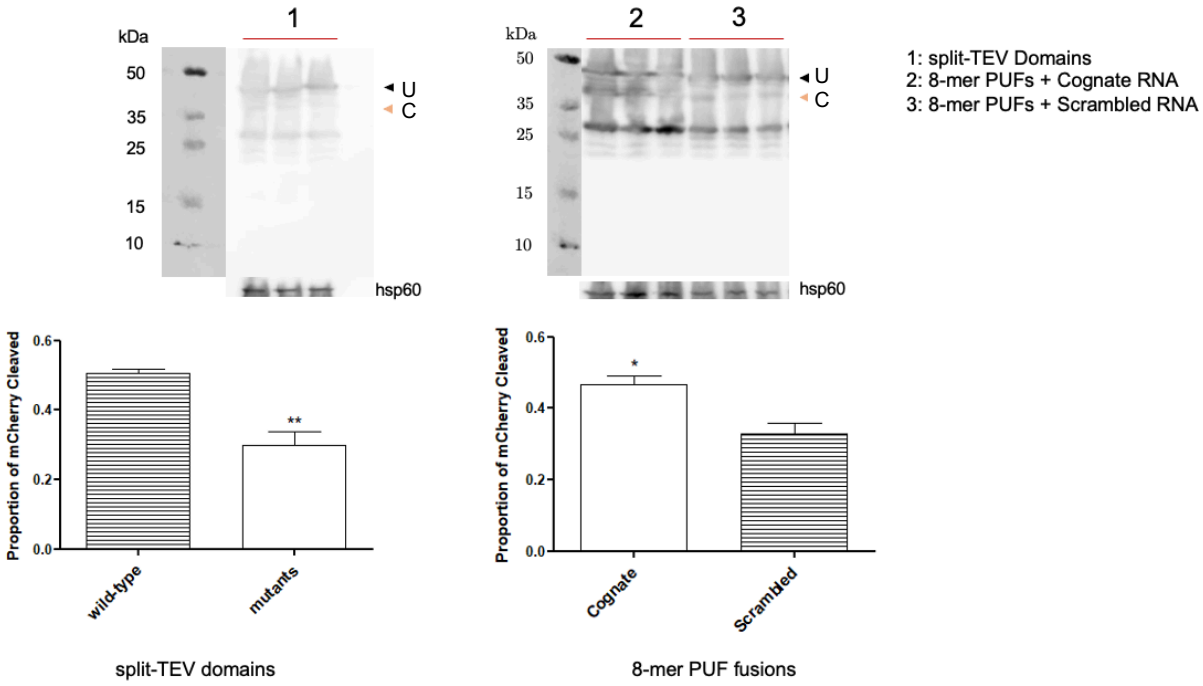


Figure 7.19: Western blots of triple transfected mutant split-TEV and 8-mer PUF fusions with inside-oriented split-TEV domains along with the SBS mCherry construct containing a two-nucleotide spacer between the PUF sites. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The three lanes under each title represent the three independent biological replicates. Densitometry compared wild-type and mutant split-TEV domains for split-TEV domains lacking PUF domains. For mutant PUF-split-TEV fusions, cognate and scrambled RNA conditions were compared. Error bars represent the SEM from three independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$.

The Western blot of the mutant split-TEV domains lacking PUF domains (Figure 7.19) shows significantly less cleavage compared to the blot in Figure 7.6 when wild-type split-TEV domains were used which is also confirmed by quantification ($p<0.01$). To assess the effects of the mutation on the lead construct from the previous optimization, the Western blots using cognate and scrambled RNA were repeated using the 8-mer PUF inside-oriented mutant split-TEV fusions with 2-nucleotide spacing. The intensity of the cleaved bands on the blot are lower when compared to blots of the PUF fusions with wild-type split-TEV domains, and this difference is reflected in the densitometry. Compared to wild-type data, the activities are lower. The 8-mer lead has a significantly higher ratio ($p<0.05$) using cognate RNA compared to the scrambled RNA control.

To test whether these effects are also present when targeting endogenous RNA, Western blots were carried out in HeLa cells using 8-mer PUF domains which target *hTERT* RNA and control PUF domains which do not. First, the expression of these new constructs was tested, and they express as expected in Figure 7.20.

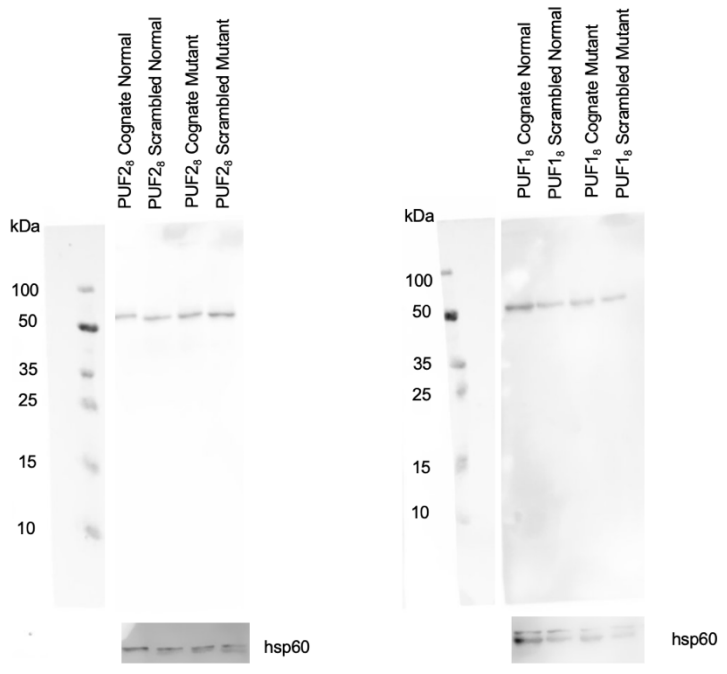


Figure 7.20: Western blots of inside-oriented split-TEV 8-mer PUF constructs designed to target endogenous *hTERT* mRNA. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control.

Once expression was confirmed, Western blots of the NASPER system were carried out in the HeLa cells (Figure 7.21 and 7.22) which contain the *hTERT* mRNA that is endogenously overexpressed (as HeLa cells are a telomerase-dependent cervical cancer cell line) to test whether the lead constructs can cleave an mCherry substrate more when cognate PUF domains are used compared to scrambled PUF domains in the context of endogenous RNA. The RNA sequences on the *hTERT* mRNA targeted by these PUF domains are shown below:

1. Cognate $PUF2_8$ Target: 5'-GCGCGUAC-3'
2. Scrambled $PUF2_8$ Target: 5'-CGUGACCG-3'
3. Cognate $PUF1_8$ Target: 5'-GAUGUGAC-3'
4. Scrambled $PUF1_8$ Target: 5'-UGACGAUG-3'

As expected from previous data, the wild-type TEV domains result in greater cleavage of the mCherry substrate as seen by the higher cleavage ratios (Figure 7.21B, 7.21C, 7.22B) compared to the mutants, and the greater cleaved band density (Figure 7.21A). When the wild-type split-TEV domains are used, no differences are seen between cognate and scrambled conditions however, when the split-TEV domains are mutated, the cognate condition shows a significantly higher cleavage ratio compared to the scrambled condition, due to reduction of non-specific activity in scrambled condition to free split-TEV domain levels (Figure 7.21B), suggesting they can indeed target endogenous RNA as well as overexpressed exogenous RNA.

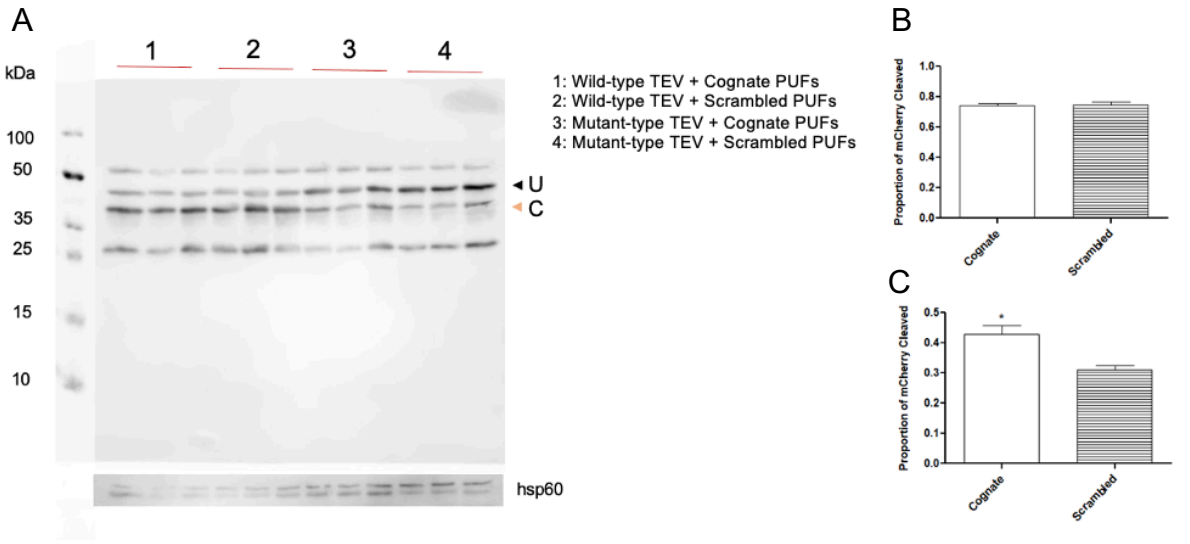


Figure 7.21: **(A)** Western blot and **(B)** densitometry of NASPER endogenous targeting in HeLa cells. 300,000 HeLa cells in a 12-well plate were transfected with the constructs shown, along with the TEV-cleavable mCherry construct, and harvested after 24h for Western blot. The blot was stained using an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The three lanes under each title represent the three independent biological replicates. Densitometry of normal TEV conditions **(B)** and mutant TEV conditions **(C)** was carried out in ImageJ. Error bars represent the SEM from three independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$.

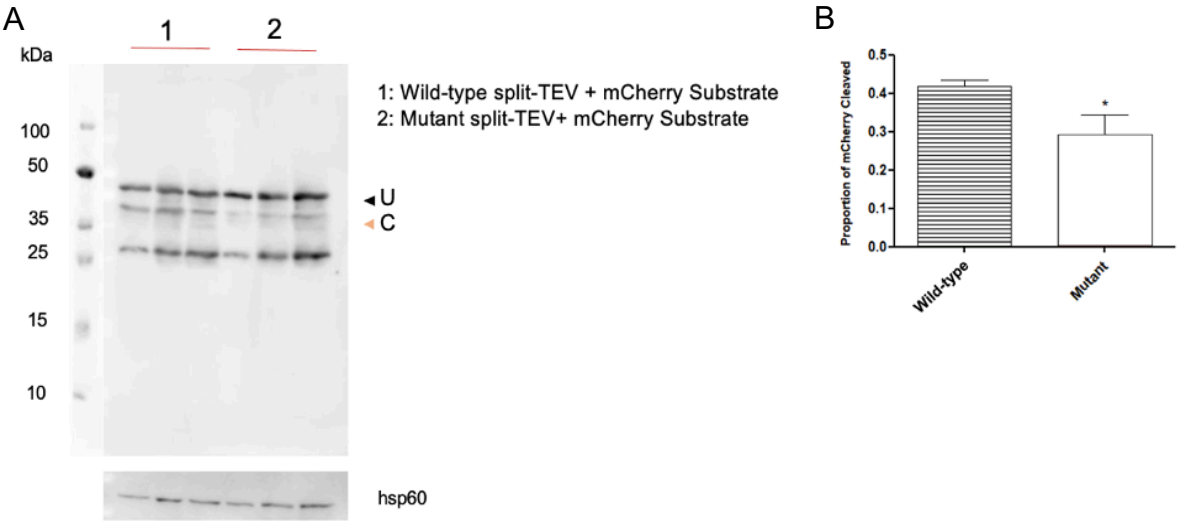


Figure 7.22: **(A)** Western blot and **(B)** densitometry of split-TEV domains lacking PUF domains in HeLa cells. 300,000 HeLa cells in a 12-well plate were transfected with the constructs shown, along with the TEV-cleavable mCherry construct, and harvested after 24h for Western blot. The blot was stained using either an anti-HA monoclonal primary antibody and subsequently with an anti-Rat HRP-conjugated secondary antibody. The blot was developed using ECL and imaged with an exposure time of 2 minutes. Hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The three lanes under each title represent the three independent biological replicates. Densitometry of normal compared to mutant TEV conditions **(B)** was carried out in ImageJ. Error bars represent the SEM from three independent biological experiments. Significance levels: * – $p < 0.05$, ** – $p < 0.01$, *** – $p < 0.001$.

7.3 DISCUSSION

In this Chapter, a novel PUF-split-TEV system was used in order to test the NASPER concept. Such split-TEV systems have been used in detecting protein-protein interactions by several groups[55,166,192-194] and have even been used to activate caspases to cause apoptosis in cells[55]. Hence, combining this with the PUF proteins seemed to be ideal to provide proof-of-concept for the NASPER system. In addition, it could be used to activate caspases in future experiments if this initial study showed that they functioned as expected. Furthermore, since the system using split-TEV domains is amplificatory, it would shed some light on the differences between the optimization conditions tested that were not visible using the split-GFP system due to its low signal.

7.3.1 SPLIT-TEV FUSIONS ARE ACTIVE IN THE CELL

On the anti-HA blot (Figure 7.6), a clear cleaved band is seen below the highest molecular weight band, which was not present on the blot with mCherry substrate alone. Thus, it was determined that the split-TEV proteins could indeed cleave this mCherry substrate. However, no differences were seen between the groups (Figure 7.6C). Even the GST-split-TEV experiment (meant to be the positive control since GST is known to dimerise[206,207]) did not show any higher activity than the wild-type. This is consistent with the split-GFP data where GST-fusion did not result in any higher split-GFP reconstitution compared to the wild-type further highlighting it may not be an efficient dimerisation inducer.

Since it was clear that the split-TEV could cleave the substrate, it was likely that cleavage when using PUF-split-TEV proteins was not enough for a difference to be detected. Hence, as for the split-GFP system, I next optimized various design features namely spacing between the PUF sites[93], orientation of the split-TEV domains, the use of multiple PUF binding sites, and using longer PUF domains, in order to improve cleavage efficiency.

7.3.2 UNDERSTANDING THE VARIABLES INFLUENCING SPLIT-TEV ACTIVITY

8-MER PUF DOMAINS

Initially, the outside-oriented split-TEV domains (8-mer PUFs) were used to optimize the spacing between the $PUF2_8$ and $PUF1_8$ sites on the RNA by changing the spacing linearly from 0 to 10 nucleotides. Quantification of the western blot revealed no significant differences between the different spacings or between cognate RNA and the scrambled RNA control (Figure 7.10). Using MBS RNA did not result in any significant changes to the SBS counterparts (Figure 7.12) suggesting that the PUF proteins are saturated with the highly expressed SBS RNA, and an increase in the number of potential binding sites does not result in an increase in the signal observed for any spacer lengths. This is different from what was observed in the split-GFP system, and the reasons will be discussed later in this section. However, changing the orientation of the of the split-TEV domains resulted in significantly higher activity consistent with data from the split-GFP system. The linear spacer series revealed a trend in the activity (Figure 7.11); the increase in activity upon increasing spacing to 2 nucleotides is consistent with data from Qiu *et al.*[93] which suggests that this change decreases the K_d of the PUF proteins by 2-fold. Spacing the PUF fusions on the RNA likely also improved the orientation of the inside split-TEV domains which would contribute to the increased cleavage observed. As the spacing increases further however, the activity decreases likely because the PUF fusions are being separated and their split-TEV domains are no longer in enough proximity to cleave the substrate.

The trends seen using SBS RNA were also observed when using the MBS RNA (Figure 7.12), however, the peak activity was seen for the 4-nucleotide spacer rather than the 2-nucleotide spacer, after which a decrease was observed. This result was surprising, as I expected the same effects as seen in the SBS, except with significantly higher cleavage ratios when using the MBS RNAs as seen in the split-GFP system. Additionally, for the 2-nucleotide spacer it even looks as though the SBS RNA had higher activity than the MBS counter-part. Together, this result suggests that PUF binding is in fact saturated with the SBS RNAs themselves and rather than simply providing additional binding sites, the MBS RNAs may also be influencing the binding

of the PUF fusions and the activity of their split-TEV domains by other means such as steric effects on binding and alternate modes of interaction of the split-TEV domains on the RNA (Figure 7.23).

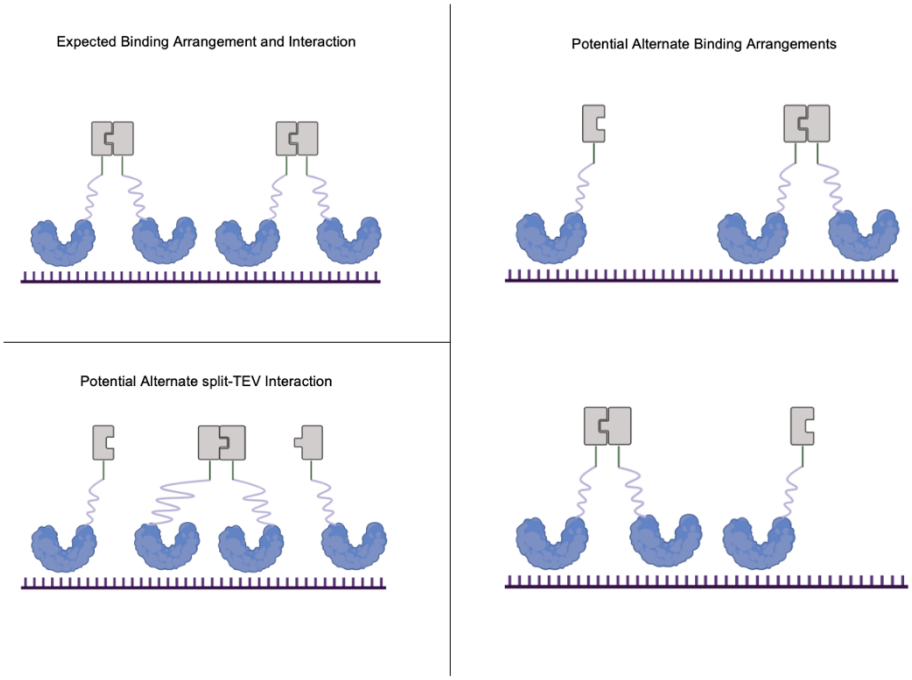


Figure 7.23: Schematic diagram of different binding orientations and interactions of PUF fusions and their split-TEV domains

These orientations and interactions of the PUF fusions and their split-TEV domains are likely the cause of the differences in cleavage seen between the SBS and MBS RNAs, despite there being no simple increases in cleavage. The observed increases in GFP fluorescence when using the MBS RNAs but not in split-TEV cleavage suggests that the PUF fusions can indeed bind in pairs onto MBS RNA (resulting in increased GFP fluorescence detected from a single RNA that is better resolved by the microscope). However, the total number of re-constituted sites are the same (resulting in the lack of increased signal using split-TEV as it is a bulk measurement of cleavage). To verify that these differences, NMR or X-ray crystallography should be carried out. Interactions between the proteins and RNA can be detected using NOESY[208,209] and chemical shift perturbation upon titration of RNA into the protein, whereas crystallography will allow protein-RNA interactions as well as interactions of the split-TEV domains to be visualised. In particular, crystallography may not be as limited as NMR by the repetitive nature of the RNA sequence and the fact that it is potentially the same proteins interacting in multiple ways on a single RNA molecule.

Thus, I found through these experiments that inside-oriented split-TEV PUF fusions were the way forward and discovered a lead construct, a 2-nucleotide spacer with inside-orientation on SBS RNA, with the highest fold change compared to the scrambled RNA (and split-TEV domains alone), to carry forward to apoptosis assays. However, before this, I decided to also test 9- and 10-mer PUF domains in the same conditions to try to improve the specificity of this PUF-split-TEV system while maintaining the fold changes observed using the 8-mer PUF fusions.

9-MER PUF DOMAINS

Using 9-mer PUF domains within the fusion proteins showed contrasting results to using the 8-mer domains. Neither the outside- nor inside-oriented split-TEV fusions displayed any significantly increased cleavage compared to a scrambled control (Figures 7.13 and 7.14). This was initially suspected to be because the 9-mer PUF domains are less effective in binding their sequences, due to misfolding or stability issues, thus even though they are expected to be more specific, they would bind their cognate sequences more poorly producing this result. However, after seeing the unusual results for the MBS RNAs, the cause of these results became clearer. When using the MBS RNAs, the blots show a number of fine bands below where the cleaved band (Figure 7.15) is expected to be. In most cases, the cleaved band blends into the shading greatly which likely affected the quantification. This is seen using SBS RNAs as well (Figures 7.13 and 7.14) and as a result, I believe the quantification for the 9-mer SBS RNA conditions are not reliable.

The presence of this laddering suggests that the mCherry substrate is being cleaved at more points than just the consensus TEV site. Another interesting factor is that these bands are much less obvious when using the 8-mer PUF domains making the resolution of the cleaved band much greater and amenable to accurate densitometry. Two properties of this system are likely to be the cause of this promiscuous cleavage. Firstly, the TEV protease normally has an extremely stringent specificity to its recognition site (Figure 7.24) primarily due to hydrogen bonding with specific residues within the protease. The substrate binds in a solvent-facing pocket of the protease[198,210], the interaction is stabilized by hydrogen bonding, and then it is

cleaved through nucleophilic attack of the carbonyl group via the thiolate-imidazolium ion pair within the active site as for other members of the cysteine protease family[211]. However, no crystallography studies have been conducted on the interaction of split-TEV domains. Splitting the protease and spacing the split domains apart could alter this binding groove compared to the intact TEV, potentially broadening substrate specificity and therefore allowing promiscuous cleavage of substrates. Only one other group used Western blots to identify split-TEV activity, but they used rapamycin-inducible dimerization which may not produce the correct orientations and kinetics for non-specific cleavage[55] compared to using the 9-mer PUF domains as I did here. The kinetics of binding of the 9-mer PUF fusions may also increase the half-life compared to the other fusions.

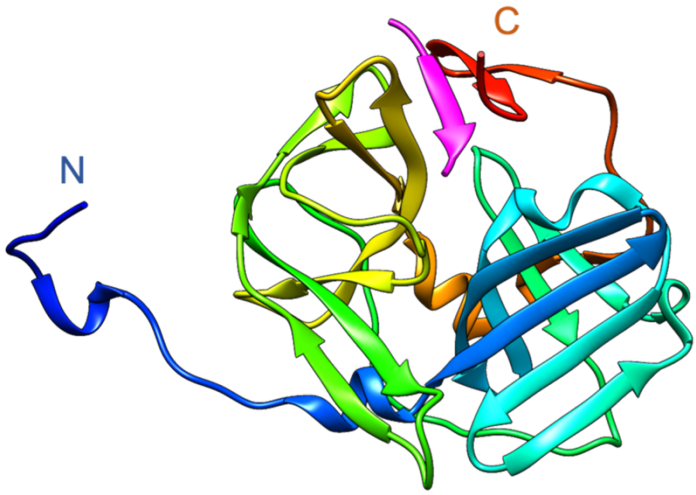


Figure 7.24: Schematic structures of the TEV protease bound to a substrate peptide. Structure is coloured from blue to red corresponding to the N-terminus to the C-terminus. Peptide substrate bound within the active site is shown in magenta.

Together, these properties would allow the 9-mer PUF fusions to bind to their RNA for a longer duration and initiate promiscuous split-TEV cleavage. Furthermore, this explains why the effect is enhanced when using MBS RNAs due to the presence of potentially two active split-TEV dimers which could cleave a single mCherry substrate in tandem. However, these hypotheses need to be tested further, opening avenues for further potential research into the effects of splitting the TEV protease and characterizing both single and dual PUF binding

kinetics using PUF domains of different lengths. The effects of splitting the TEV protease could be investigated by X-ray crystallography to understand their interactions, and how it compares to the full protease. Additionally, a high-throughput method developed by Salisbury *et al.*[212] using peptide microarrays wherein fluorescence is produced upon peptide cleavage, could be used to assess the change in specificity of the TEV protease upon splitting. Characterizing the on- and off-rates of the PUF domains may be more challenging due to the tendency of these proteins to be purified with bound RNA (see Chapter 3), but denaturation and re-folding of the proteins may be able to liberate, and therefore separate, the bound RNA facilitating these assays.

Due to these results however, the 9-mer PUF fusions were not suitable for the apoptosis assays. Since the 9-mer fusions did not work as expected, 10-mer PUF domains were tested.

10-MER PUF DOMAINS

Inside-oriented split-TEV domains fused to 10-mer PUF domains had significantly higher split-TEV activity when a 10-nucleotide spacer was used (Figure 7.17) further suggesting that inside-oriented split-TEV domains are more stable than outside-oriented split-TEV domains. The 10-mer PUF fusions with inside-oriented split-TEV domains showed a different trend (compared to the corresponding 8-mer fusions) that as spacer length was increased, split-TEV activity increased. This is likely because they are bulkier than the 8-mer PUF domains leading to more steric hindrance when binding concurrently. Hence, increasing the spacing between the two fusions improves their ability to bind and properly orient the split-TEV domains resulting in increased activity.

The MBS RNA for the inside-oriented split-TEV domains replicates this pattern of increasing activity with increasing spacer length, but there is no increase in activity relative to the SBS RNAs (Figure 7.18). At a spacing of 2 nucleotides, the SBS RNA seems to produce higher activity than the MBS counterpart, further suggesting that the MBS RNAs are not simply providers of additional binding sites and may be changing the binding modes of the PUF fusions as suggested in Figure 7.23. The equilibrium between these different modes likely produces the unexpected results using MBS RNAs observed in these experiments.

However, the aim of using the 10-mer PUF domains was to reduce concurrent binding to natural cognate sites and thus reduce split-TEV activity for the scrambled RNA compared to 8-mer PUF domains. No such decrease in activity was observed when using 10-mer PUF domains (unlike in the split-GFP system), suggesting any changes in activity due to reduction in binding to natural cognate sites, increases in sequestering[102], or poorer binding ability (Chapters 6 and 9) are masked by the background reconstitution of the split-TEV domains themselves which would be present even if the fusions are sequestered.

Surprisingly, using the 10-mer PUF fusions resulted in lower fold-changes in activity compared to when the 8-mer PUF fusions were used. This suggests that the 10-mer PUF fusions are forming fewer cognate split-TEV reconstitutions providing further evidence that there are potential sequestering effects (as supported by Chapter 6), as these would also reduce the number of these cognate reconstitutions since the 10-mers would have a smaller available pool of proteins to bind exogenous target cognate RNA. However, this evidence could also apply to the hypothesis that the 10-mer PUF domains bind less efficiently to their RNA targets and thus, biophysical studies are required to elucidate which of these hypotheses leads to the observed effects. These results do not support the hypothesis that there is a reduction in concurrent binding to natural cognate sites because if this were true, split-TEV activity would reduce equivalently when using cognate and scrambled RNAs, maintaining the fold changes observed using the 8-mer PUF domains.

The 10-mer PUF fusions were not used further as none of the tested constructs showed a significantly higher activity compared the scrambled RNA control and the split-TEV only control. In parallel, in an attempt to reduce the background activity of the split-TEV domains alone, site-directed mutagenesis of the split-TEV domains was attempted as explored by Dolberg *et al.*[197].

7.3.3 MUTATION OF SPLIT-TEV REDUCES ACTIVITY

Dolberg *et al.*[197] attempted to reduce the spontaneous reconstitution of the split-TEV halves through site-directed mutagenesis. They used the identical TEV split site as used in this thesis (amino acid 118/119), and previous studies of split-TEV[55,166,192-194], in a computational model

called SPORT which enabled them to carry out a mutational scan of residues buried within the full-length TEV protease to look for perturbations in the interfacial energy across the split-TEV interface and the total stability of the mutated protein. They verified their mutations experimentally using both a tethered and soluble split-TEV system in a luciferase-based reporter assay with and without inducing dimerisation. I chose the optimal mutation pair, having the lowest reporter activity when uninduced but significantly higher activity upon induction, and applied it to the NASPER system in this thesis. Activity of the NASPER lead when co-transfected with scrambled RNA was significantly decreased, however, so was the activity with cognate RNA (Figure 7.19) suggesting that the mutations, despite reducing non-specific activity may also decrease the specific activity of the split-TEV domains either through a reduction in catalytic efficiency[197] or ability to dimerise in proximity. Despite this, using the cognate RNA resulted in significantly higher activity compared to scrambled RNA.

To explore whether these effects were still present when targeting endogenous *hTERT* RNA, further experiments were carried out. The region on *hTERT* RNA targeted was the alpha region which is present on full length *hTERT* RNA that is required for telomerase activity[213], whereas it is not present in normal tissue and non-cancer cell lines[31,32], such as BJ fibroblasts, which contain the Δ4-13 variant. HeLa cells have been shown to be telomerase-dependent and naturally contain the full-length *hTERT* RNA which allows them to avoid cellular senescence and maintain a cancerous phenotype[31,32,203,213,214]. They also contain other *hTERT* variants lacking the alpha and beta regions of *hTERT* (absent in normal cells), however variants containing the alpha region are the most prominent[32] making the alpha region a better target. These experiments showed that the NASPER lead can target endogenous RNA successfully using mutant split-TEV. The effect was not observed using normal split-TEV domains likely because the activity using cognate proteins was not sufficiently high to be observed over the non-specific activity. Once this background was reduced by mutation to the level of mutant split-TEV domains lacking PUF domains, the differences were revealed (Figure 7.21). This extends the initial proof of concept of the NASPER system by showing that endogenous RNA can also be specifically targeted by the system. The lead NASPER construct chosen here will be tested further in downstream apoptosis assays, initially using exogenous RNA to show that cells can be killed in the presence of this RNA, and subsequently attempts will be made to target endogenous RNA in apoptosis assays also.

CHAPTER 8 – CAN NASPER INDUCE APOPTOSIS?

8.1 INTRODUCTION

In the previous Chapter, a lead candidate for the NASPER system was identified based on the cleavage of an mCherry substrate. Thus, experiments using this lead were carried out in order to determine whether it could also activate a TEV-cleavable procaspase-3 zymogen to induce apoptosis when co-expressed with cognate RNA.

8.1.1 APOPTOTIC PATHWAYS

Apoptosis is a highly regulated process that ultimately results in the controlled death of a cell. There are two primary pathways that can result in the apoptosis cascade (Figure 8.1) – the extrinsic and intrinsic pathways[215].

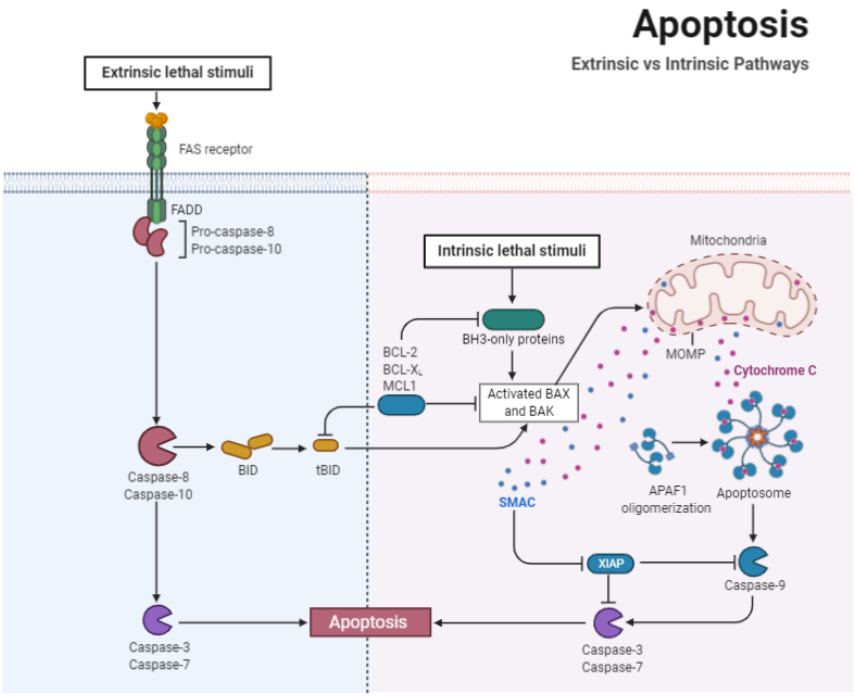


Figure 8.1: Schematic diagram of extrinsic and intrinsic apoptotic pathways in the cell.

The extrinsic pathway is initiated by the binding of a death ligand to a death receptor which in turn results in the activation of procaspase-8 and procaspase-10. These upstream caspases then go on to activate the effectors procaspase-3 and procaspase-7 which results in mass proteolysis in the cell creating a point of no return in apoptosis and inevitable cell death[215-217]. In contrast, the intrinsic pathway is activated by lethal stimuli such as hypoxia, which also results in the activation of the aforementioned effector caspases. However, this process occurs due to mitochondrial release of cytochrome c, which activates the apoptosome that goes on to activate caspase-9 and this subsequently activates caspases 3 and 7[215,218,219].

8.1.2 CASPASE-3

Caspase-3 is the primary effector caspase active in the demolition phase of apoptosis. It is expressed as a zymogen dimer that undergoes two cleavage events that result in its activation[220].

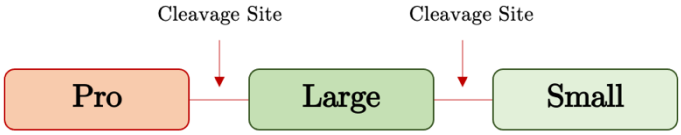


Figure 8.2: Schematic diagram of procaspase-3 domain structure.

Activation then leads to the cleavage of key cellular proteins such as poly(ADP-ribose) polymerase, cell cycle proteins and DNases[220]. Loss of these proteins and cleavage of DNA result in cell condensation, blebbing and ultimately death, and hence Gray *et al.*[55] used caspase-3 as an inducer of apoptosis in their split-TEV system. They modified the cleavage site between the large and small subunits of procaspase-3 to be TEV-cleavable, thereby creating a zymogen that can only be activated in the presence TEV protease activity. They used a rapamycin-inducible system to show that dimerisation of split-TEV halves results in subsequent activation of this new procaspase-3 which subsequently results in cell death. Hence, I decided to use this substrate as the apoptosis effector in the NASPER system as well. There are two key differences between the two systems – 1) Gray *et al.*[55] used small molecule induction of dimerisation, which is a much more rapid process compared to the NASPER

system which requires transfection and expression of mRNA and proteins to occur; and 2) they used procaspase-3 stable cell lines compared to overexpression by transient transfection. Overexpression of procaspase-3 itself can induce some apoptosis[221], however the NASPER system should increase apoptosis levels above this background. Additionally, the measurement time-course will have to be longer than that used by Gray *et al.*[55].

8.1.3 ASSAY DESIGN, AIMS AND OBJECTIVES

Thus, the aim of this Chapter was to test the NASPER candidate in apoptosis assays and evaluate whether it can induce apoptosis in the presence of cognate RNA. The ApoTox-Glo kit from Promega was used to measure cell viability and caspase-3 activity simultaneously. It uses cleavage of a fluorogenic cell permeable peptide to measure cell viability, and cleavage of a luminogenic caspase-3 substrate and luciferase to measure capsase-3 activity. Additionally, since the RNA sequences for the PUF proteins to bind are fused to the 3' end of the mCherry mRNA (as in previous Chapters), mCherry fluorescence intensity can be used as a measure of viability for the cell population containing the RNA of interest. Indeed, the use of fluorescent proteins as a measure of cell viability has been verified by Strebel *et al.*[222] and is now widely applied. This method works since apoptosis causes mass proteolysis, which includes fluorescent proteins such as mCherry, and thus results in a decrease in fluorescence. Hence, the schematic shown in Figure 8.3 was used to achieve these aims.

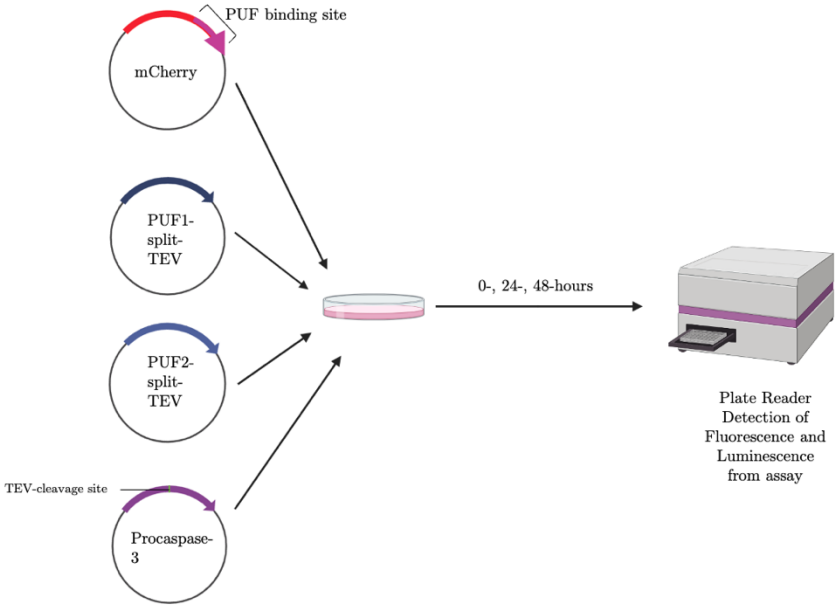


Figure 8.3: Schematic of assay design.

8.2 METHODS

APOTOX-GLO TRIPLEX ASSAY

HEK293T cell seeding was carried out as per the methods section, scaled down to 10,000 cells per well in a 96-well plate. Transfection of cells was carried out using an equimolar ratio of the required plasmids (and empty vector was used for the appropriate controls) in a total of 115 ng DNA per well in the 96-well plate. ApoTox-Glo Triplex Assay (Promega) was carried out at 0-, 24- and 48-hour time-points post-transfection according to the manufacturer's instructions. In brief, 10 µL of GF-AFC substrate was added to 2 ml of assay buffer, and the contents of the Caspase3/7 Glo buffer bottle was added to the Caspase3/7 Glo substrate to form the active reagent. At the appropriate time-points, 20 µL of prepared assay buffer was added per well, the plate was shaken for 30 seconds using an orbital shaker and incubated at 37°C for 1 hour. Fluorescence was then measured using a Clariostar plate reader (BMG Labtech) with an excitation wavelength of 398.2-56 nm and emission wavelength of 496-32 nm. The dichroic mirror used was 453.1 nm and gain was set to 600.

100 µL of Caspase3/7 reagent was then added per well, shaken for 30 seconds and incubated at 37°C for 1 hour. Luminescence was then measured using a Clariostar plate reader (BMG Labtech) without an emission filter using a gain of 2500. mCherry fluorescence intensity was measured prior to substrate addition using an excitation wavelength of 561-15 nm and emission wavelength of 610-20 nm. The dichroic mirror used was 584.2 nm and gain was set to 2614. Background emission was also measured using the same settings prior to substrate addition and was subtracted from measurements to correct readings.

8.3 RESULTS

8.3.1 PROCASPASE-3 EXPRESSION

To check whether the procaspase-3 construct is expressed in HEK293T cells, Western blot was used to probe for the FLAG-tag in the construct. Figure 8.4 shows that the procaspase-3 construct expressed at the expected molecular weight. Thus, subsequent cell viability, and caspase, and mCherry fluorescence assays could be carried out using this construct and the lead NASPER system candidate.

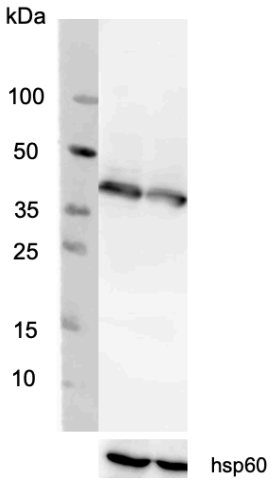


Figure 8.4: Western blot of procaspase-3 construct containing an N-terminus 3xFLAG-tag. 300,000 HEK293T cells in a 12-well plate were transfected with the constructs shown and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG primary antibody, and subsequently with an anti-Mouse HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. hsp60 was used as a loading control. The two lanes shown are the two biological replicates.

8.3.2 CELL DEATH ASSAYS

Both wild-type and mutant (H75S for N-TEV and I163V for C-TEV) split-TEV versions of the lead candidate were tested in cell death assays, together with controls transfected using procaspase-3-only, once the expression of procaspase-3 was verified. For controls without procaspase-3, an empty vector was transfected instead so that the same number of plasmids was used, maintaining comparable transfection efficiencies.

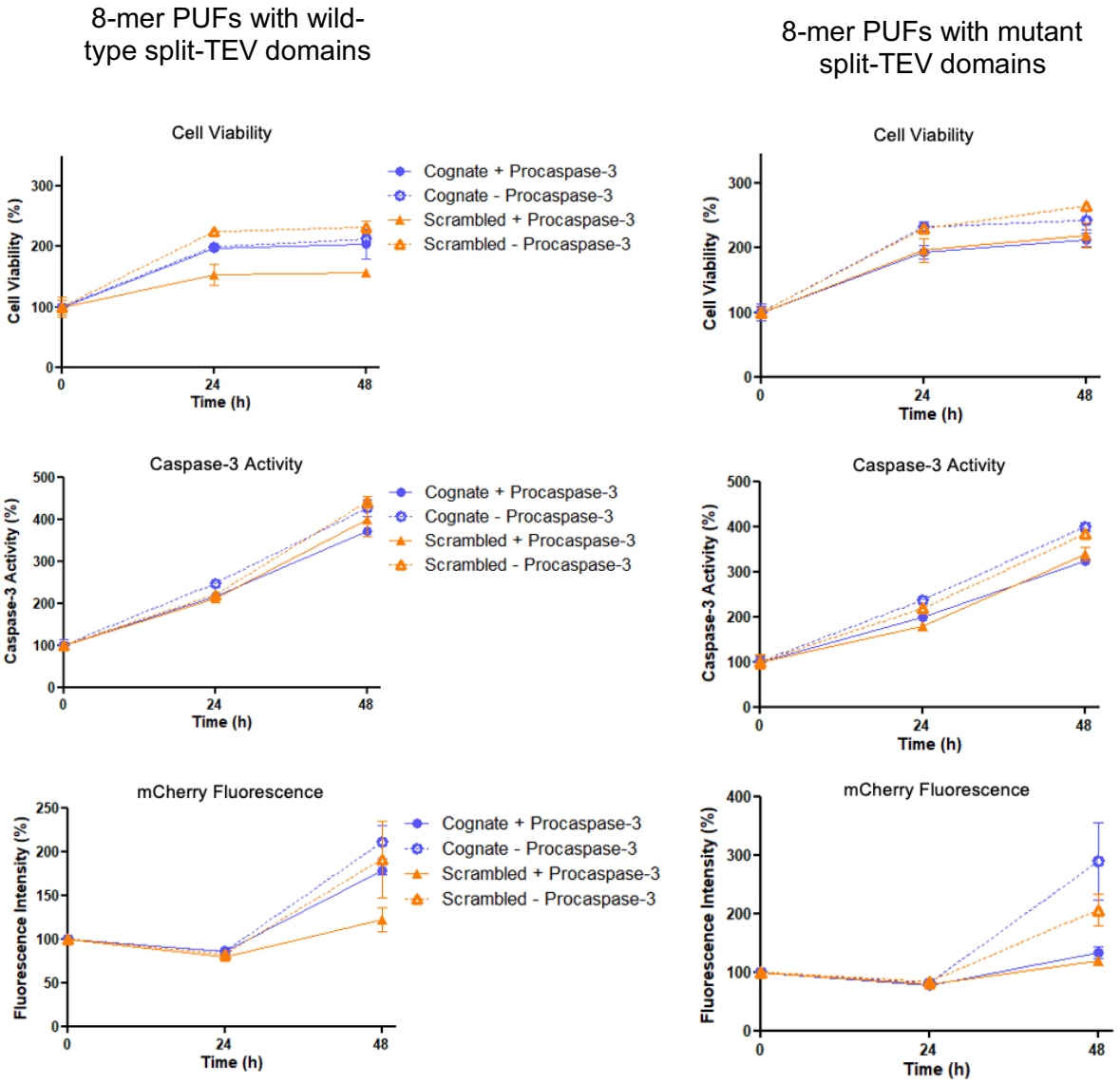


Figure 8.5: Cell death assays of NASPER lead candidate with either wild-type or mutant split-TEV domains. 10,000 HEK293T cells were transfected in a 96-well plate with 8-mer PUF fusions, mCherry RNA containing the PUF target sites separated by a 2-nucleotide spacer and with either the TEV-cleavable procaspase-3 construct or an empty vector. Cell viability, caspase-3 activity and mCherry fluorescence were measured using the ApoGlo Triplex Assay at different time-points over 48 hours and normalised to t=0. Error bars represent the SEM from three independent biological replicates.

For 8-mer fusions with both wild-type and mutant split-TEV domains, the cell viability, caspase-3 activity and mCherry fluorescence increases over the 48 hours. Differences can be observed between the procaspase-3+ and procaspase-3– groups at the 48-hour time-point – cell viability, caspase-3 activity and mCherry fluorescence were significantly lower when procaspase-3 was added, but these differences were observed for both the cognate and scrambled RNAs (Figure 8.5). For the wild-type split-TEV fusions, it appears that the cell viability and mCherry fluorescence with procaspase-3 is lower when using scrambled RNA compared to cognate RNA, however this difference is not seen using the mutant split-TEV 8-mer fusions.

This result is the opposite of what is expected – the cell viability and mCherry fluorescence was expected to be lower for the cognate RNA due to NASPER-induced activation of procaspase-3 and subsequent apoptosis. Additionally, there are differences in cell viability, caspase-3 activity and mCherry fluorescence between the cognate and scrambled RNA even when empty vector was used instead of procaspase-3 suggesting this needs to be considered. Thus, fold changes upon procaspase-3 addition would provide a more robust analysis (Figure 8.7). Interestingly, the differences between procaspase-3 + and procaspase-3 – groups appear to be greater for mCherry fluorescence than for cell viability for both the wild-type and mutant split-TEV fusions. Controls using only procaspase-3 were tested to assess whether the effects of procaspase-3 on these variables are due to the NASPER system or simply due to procaspase-3 overexpression.

Using procaspase-3-only as a control also resulted in a small reduction in cell viability, caspase-3 activity and mCherry fluorescence for procaspase-3 + groups at the 48-hour point (Figure 8.6).

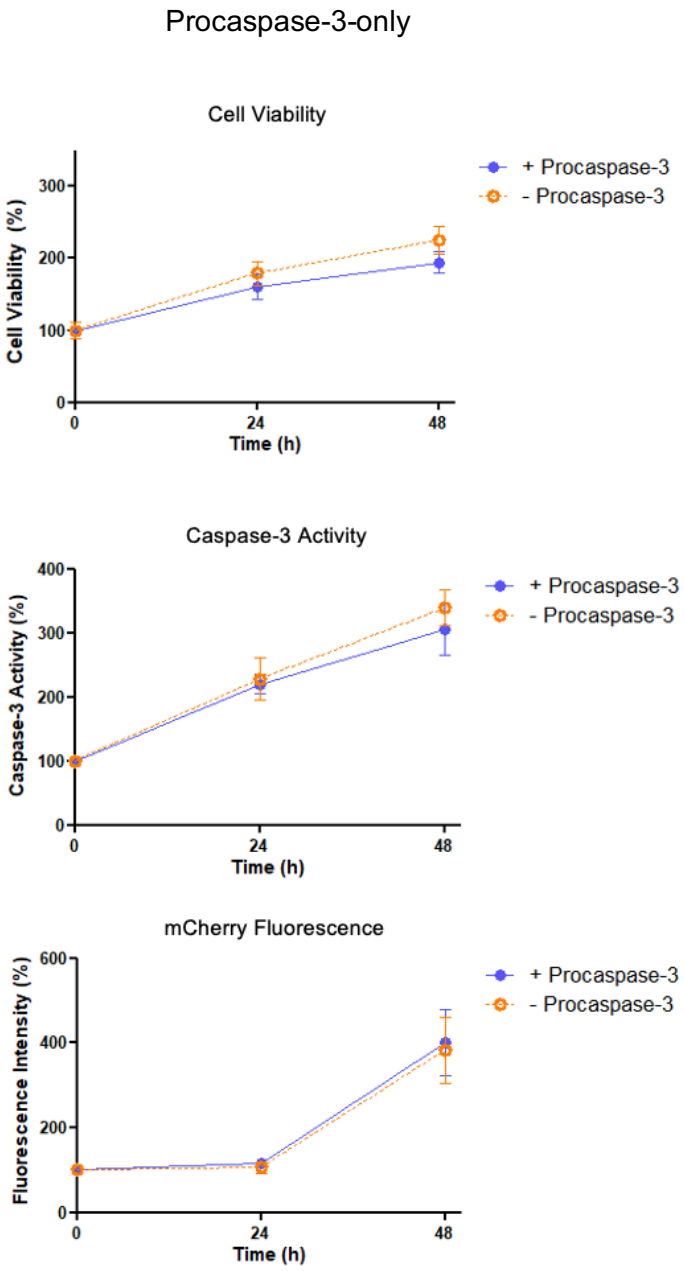


Figure 8.6: Cell death assays of procaspase-3-only controls. 10,000 HEK293T cells were transfected in a 96-well plate with mCherry RNA and either the TEV-cleavable procaspase-3 construct or an empty vector. Cell viability, caspase-3 activity, and mCherry fluorescence were measured using the ApoGlo Triplex Assay at different time-points over 48 hours and normalised to t=0. Error bars represent the SEM from three independent biological replicates.

To better visualise the differences between cell viability and mCherry fluorescence in cognate and scrambled conditions, and to compare this to the procaspase-3-only control, fold changes upon procaspase-3 addition were calculated (Figure 8.7).

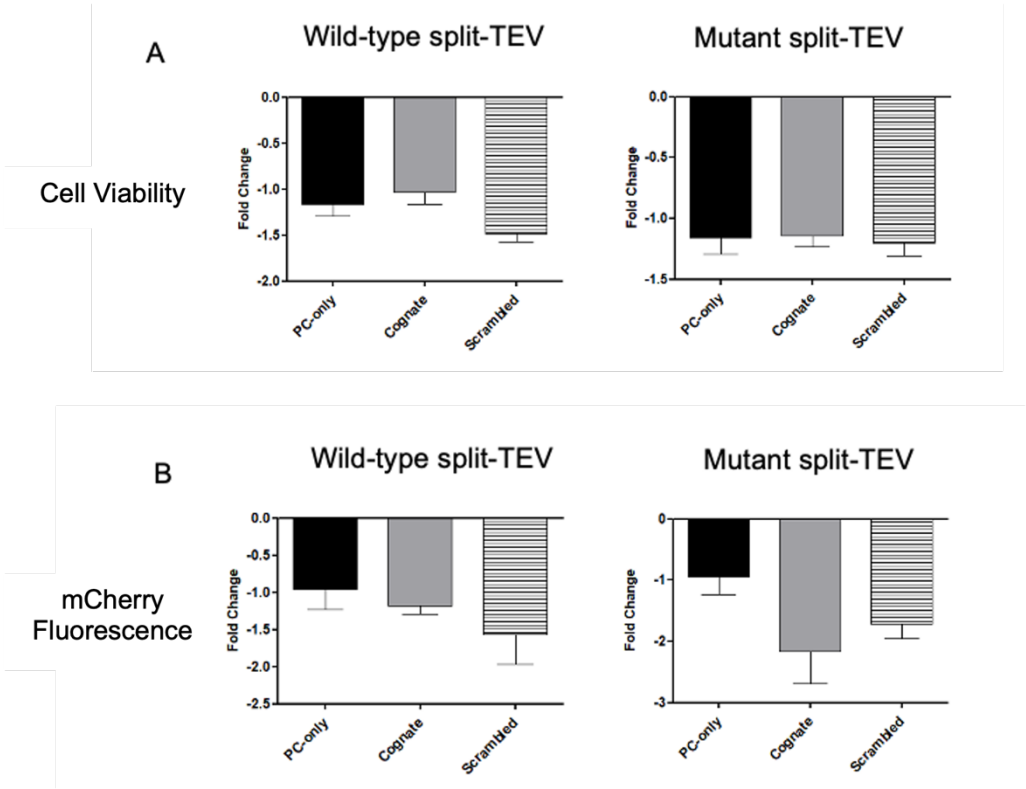


Figure 8.7: Fold changes upon procaspase-3 addition for cell viability **(A)** and mCherry fluorescence **(B)** at 48h. Error bars represent the SEM from three independent biological replicates.

Using 8-mer fusions with either wild-type or mutant split-TEV does not result in any significant changes in cell viability between cognate RNA, scrambled RNA and procaspase-3-only. This same pattern was seen when using mCherry fluorescence as a measure of cell death. Although the fold decreases in mCherry fluorescence using NASPER appear to be greater than that of the procaspase-3-only control, the difference is not significant. To further investigate the cause, Western blotting was carried out to check whether the procaspase-3 substrate is being cleaved by the split-TEV in the NASPER system.

Western blots show only the presence of a band at the molecular weight of the full-length procaspase-3 construct with no bands corresponding to the cleaved construct using the NASPER lead with either normal or mutant split-TEV domains (Figure 8.8). To verify whether this was due to the procaspase-3 construct being uncleavable, *in vitro* cleavage assays were carried out using HEK cell lysate containing procaspase-3. Addition of full-length TEV protease resulted in the formation of a band corresponding to the cleaved procaspase-3 construct in either 10 mM Tris-HCl pH 7.5 (Buffer A) or 20 mM HEPES pH 7.5 (with 5 mM KCl, 1.5 mM $MgCl_2$ and 1 mM DTT, Buffer B) buffers.

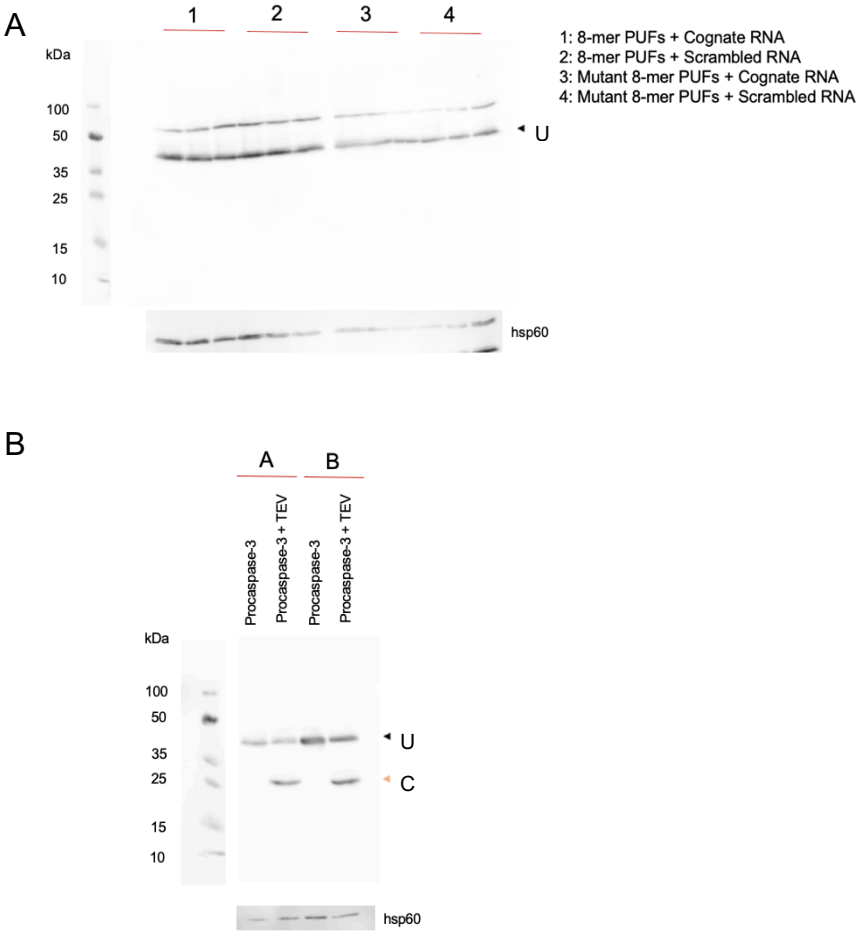


Figure 8.8: Western blot of procaspase-3 cleavage using 8-mer NASPER **(A)**. 300,000 HEK293T cells in a 12-well plate were transfected with the 8-mer PUF fusions, mCherry RNA containing the PUF target sites separated by a 2-nucleotide spacer and the TEV-cleavable N-terminus 3xFLAG-tag procaspase-3 and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG primary antibody, and subsequently with an anti-Mouse HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The three lanes shown are the three biological replicates. **(B)** shows an in vitro procaspase-3 cleavage assay using full-length TEV protease where lanes A and B refer to different buffers tested. Uncleaved and cleaved bands are marked as U and C respectively.

Lastly, to establish whether the lack of cleavage observed in the blots was due to cleavage of the prodomain resulting in loss of the FLAG-tag, the experiments were repeated using a procaspase-3 construct containing the FLAG-tag on the C-terminus (Figure 8.9). No bands corresponding to the cleaved procaspase-3 were visible on the blot (Figure 8.9) suggesting the NASPER system is not cleaving the procaspase-3 construct.

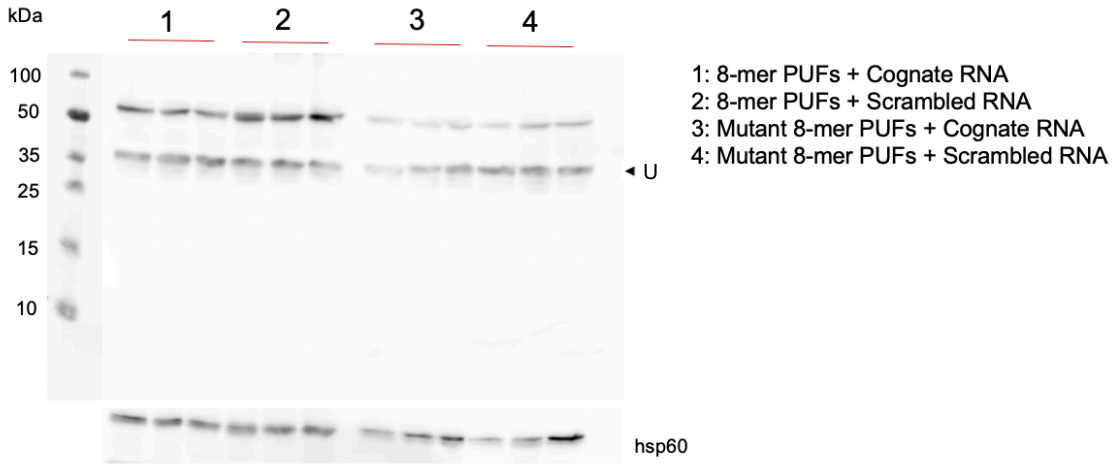


Figure 8.9: Western blot of C-terminal 3xFLAG-tag procaspase-3 cleavage using 8-mer NASPER **(A)**. 300,000 HEK293T cells in a 12-well plate were transfected with the 8-mer PUF fusions, mCherry RNA containing the PUF target sites separated by a 2-nucleotide spacer and the TEV-cleavable C-terminus 3xFLAG-tag procaspase-3 and harvested after 24h for Western blot. The blot was stained using either an anti-FLAG primary antibody, and subsequently with an anti-Mouse HRP-conjugated secondary antibody respectively. The blot was developed using ECL and imaged with an exposure time of 2 minutes. hsp60 was used as a loading control. Uncleaved and cleaved bands are marked as U and C respectively. The three lanes shown are the three biological replicates.

8.4 DISCUSSION

The aim of this Chapter was to test the lead candidate for the NASPER system identified in the previous Chapter using apoptosis assays to see whether they could indeed cause greater apoptosis in the cognate condition compared to the scrambled condition by cleaving a TEV-cleavable procaspase-3 zymogen as developed by Gray *et al.*[55]. The lead candidate was tested using wild-type and mutated split-TEV domains.

The data showed that all three variables measured – cell viability, caspase-3 activity and mCherry fluorescence – increase over the time-course tested. This is likely due to normal growth of the cells themselves resulting in greater viability and total caspase-3 activity. mCherry fluorescence increases over time due to increased expression of the mCherry protein in the cell. However, when procaspase-3 replaced the empty vector, the data for all these variables was shifted down suggesting although there is still growth of cells resulting in the increase over the time-course, the cells have lower viabilities, caspase-3 activity and mCherry fluorescence at any given time-point, indicative of cell death. Although caspase-3 activity was expected to increase due to increased cleavage of procaspase-3, a paradoxical decrease was observed. This phenomenon was also observed in Yoon *et al.*[223] and is because as cells with activated caspases die, the caspases get degraded as well and you get a decrease in the total caspase activity (similar to the reason why mCherry is lost as a cell dies). The effect was also seen by Gray *et al.*[55] – their experimental design allowed them to detect an increase in caspase-3 activity due to split-TEV cleavage, however this increase was incredibly transient and only present when the cell viabilities were unaffected. As the viability of their cells decreased due to the subsequent apoptosis, the caspase-3 activity also decreased.

Although Figures 8.5 and 8.6 only show that there decrease in the measured variables upon procaspase-3 addition, they cannot be used to identify differences between cognate and scrambled conditions because there are differences in the dependent variables even without procaspase-3 addition. Thus, it was necessary to calculate fold changes upon procaspase-3 addition and use these as a comparator since this would normalise for those differences. Theoretically, addition of procaspase-3 should show a greater fold decrease in viability and mCherry fluorescence in cognate conditions compared to scrambled if NASPER is indeed

increasing cleavage of procaspase-3 and thus causing apoptosis. Data for caspase-3 activity were not analysed further since it was only used as a method to detect procaspase-3 cleavage which did not work. Additionally, cell viability and mCherry fluorescence are better measures of apoptosis even though caspase-3 activity decreased upon procaspse-3 addition since this decrease was linked to the cell viability itself.

No differences in the fold changes for cell viability or mCherry fluorescence were detected when using the lead candidates compared to the procaspase-3-only control. Although fold changes would account for some of the variability in transfection efficiency (as seen by the error bars) this is still a limitation to the assays used in this Chapter. To troubleshoot the lack of effects due to NASPER, Western blotting was used to check whether the procaspase-3 substrate is being cleaved by the system. Figure 8.8 showed that in fact, it does not seem to be cleaved by the lead candidate tested, but it is cleavable by full length TEV protease *in vitro*. This suggests that procaspase-3 is not cleaved by the NASPER system, but this is contradictory to the cleavage observed by the split-TEV system used by Gray *et al.*[55]. Alternatively, the procaspase-3 may be cleaved but this cleavage is not detected either because the cleaved form of procaspase-3 is rapidly ubiquitinated and degraded[55,224,225] or because prodomain cleavage[220] results in loss of the FLAG-tag impairing Western blot detection. However, using a procaspase-3 construct containing a C-terminus FLAG-tag which would not be lost upon prodomain cleavage did not result in any visible cleavage either. This is unlikely to be due to ubiquitination and degradation of active caspase-3, because Gray *et al.*[55] observed the cleavage on Western blot using their system despite this. Alternatively, cleaved procaspase-3 may not be detected by Western blot because cells in which procaspase-3 is being activated would undergo apoptosis and subsequent degradation of the caspase-3. Although this phenomenon is observed in the cell death assays, it is very unlikely to be the reason that caspase-3 is not being detected by Western blot because there would still be a population of cells in the early phase of apoptosis when caspase-3 is not yet degraded (which is why Gray *et al*[55] were able to detect the cleavage using Western blot). Furthermore, if this was indeed the case, greater cell death would have been detected using the lead NASPER candidate compared to the procaspase-3-only control in the cell death assays (which it was not). However, to test this other explanation, the Western blots could be repeated using a caspase-3 inhibitor such as Z-DEVD-FMK to see whether caspase-3 can then be detected. The system used by Gray *et al.*[55]

should also be obtained to ensure the Western blots and ApoGlo Triplex assays were working as expected and shown by them.

Hence, the most likely explanation is that the orientations achieved by the PUF-based dimerisation are not appropriate for cleavage of the procaspase-3 substrate clearly indicating a substrate-dependency on split-TEV cleavage which has been observed previously[197,226]. There are two potential avenues for future work here – 1) re-optimising the NASPER system using the procaspase-3 construct to check whether any constructs used in Chapter 7 are able to cause cleavage and 2) using a new substrate that is more representative of the TEV cleavage site on the mCherry substrate such as a peptide-drug conjugate pro-drug[227] where activation occurs due to removal of the peptide domain by TEV releasing a cytotoxic drug such as staurosporine.

CHAPTER 9 – CONCLUSIONS AND FUTURE WORK

Cancer constitutes a diverse range of diseases with an even greater number of different potential treatments. Several of these treatments are for specific cancers, such as Enzalutamide and Abiraterone for prostate cancer[228,229], but there is currently no single treatment applicable to different types of cancers. In an attempt to find a common feature between cancers that could provide a broadly applicable target, I identified telomerase, which is upregulated in around 90% of cancers[20]. Therapies targeting this enzyme have been unsuccessful, as they do not kill cancer cells, allowing them to divide and survive with shorter telomeres promoting alternative methods of senescence prevention[230]. Hence, rather than targeting telomerase, a potentially better approach is to use the overexpression of *hTERT* mRNA as a marker for cancer and exploit this to kill the cell. I named this method Nucleic Acid Scaffold Proximity-mediated Enzyme Response (NASPER). It involves co-localising two fusion proteins (each comprising an RNA-binding PUF domain and an effector domain) on the *hTERT* mRNA to induce proximity and thereby activate the effector domains to kill the cell. This thesis describes experiments aimed to demonstrate proof of concept of the NASPER system, from which the following conclusions can be drawn, and areas of future research envisaged:

In Chapter 3 *I found that the PUF proteins can simultaneously bind to a single RNA in vitro* using fluorescence polarisation, which showed that polarisation obtained when the cognate sites for both PUF proteins are present is significantly higher than when only one cognate site is present, suggesting a larger complex is assembled on the RNA indicative of concurrent PUF binding. In Chapter 4, *the purification of one of the selected effector domains– the HIV protease – was optimised in order to verify that it was active before proceeding to cell-based assays.* This protease is a notoriously difficult molecule to produce recombinantly, and current purification methods all require the use of inclusion body denaturation followed by refolding[111-113,231]. There are some studies suggesting methods for soluble expression[113,232], but these could not be replicated in our hands. The His-lipoyl tag[134] was found to yield soluble expression of the HIVPR, but subsequent purification was challenging. Although the tagged HIVPR could be purified, cleavage of the tag caused the protein to stick to the SEC column. Furthermore, as the tagged protein has a very similar molecular weight to that of the dimeric cleaved HIVPR, SEC would not be able to separate them. Despite these challenges, a fluorescence-

based protease activity assay showed that the solubly purified HIV protease is indeed active. Further work could focus on improving the efficiency of tag cleavage and optimizing the purification of the cleaved HIV protease by finding a buffer for SEC where the cleaved HIV protease does not dimerise. In Chapter 5 it was shown that *the HIV protease cannot be used as an effector domain in the NASPER system*, as it has activity even when only a single PUF-HIVPR fusion protein is expressed in the cell, producing background activity that makes it impossible to detect any true NASPER activity. Additional mutations to those (including combinations) used in this Chapter could be used to try to reduce this background so that the HIVPR can work as an effector domain in the NASPER system. These could target the termini dimerisation interfaces, since targeting the active site dimerisation interface proved unsuccessful. The mutations could include truncations of the last four amino acids of the C-terminus, rather than a complete deletion, which may have a milder effect on the activity when the PUF fusions come into proximity.

In Chapter 6, *a novel assay using organic orthogonal phase separation linked to next-generation sequencing (NGS) was used to probe the specificity of the PUF domains and the dependence on PUF length.* 9- and 10-mer PUF domains were found to bind target sequences that were different from the sequences they were designed to target. These effects are expected to be due to the ability of PUF modules to be promiscuous[104]. Longer PUFs will have a greater tendency to bind these non-target RNAs, since the reduction in binding energy due to promiscuity could be offset by the increase in binding energy due to the number of PUF repeats present thus providing enough energy to bind despite there being mismatches between the alignment of the repeats and nucleotides. This led to the hypothesis that an increase in the number of potential endogenous sites that 9- and 10-mer PUF domains can bind could result in the sequestering of the PUF proteins. In other words, the equilibrium of PUF domain binding would be shifted away from the target RNA towards endogenous non-target RNAs (Figure 10.1).

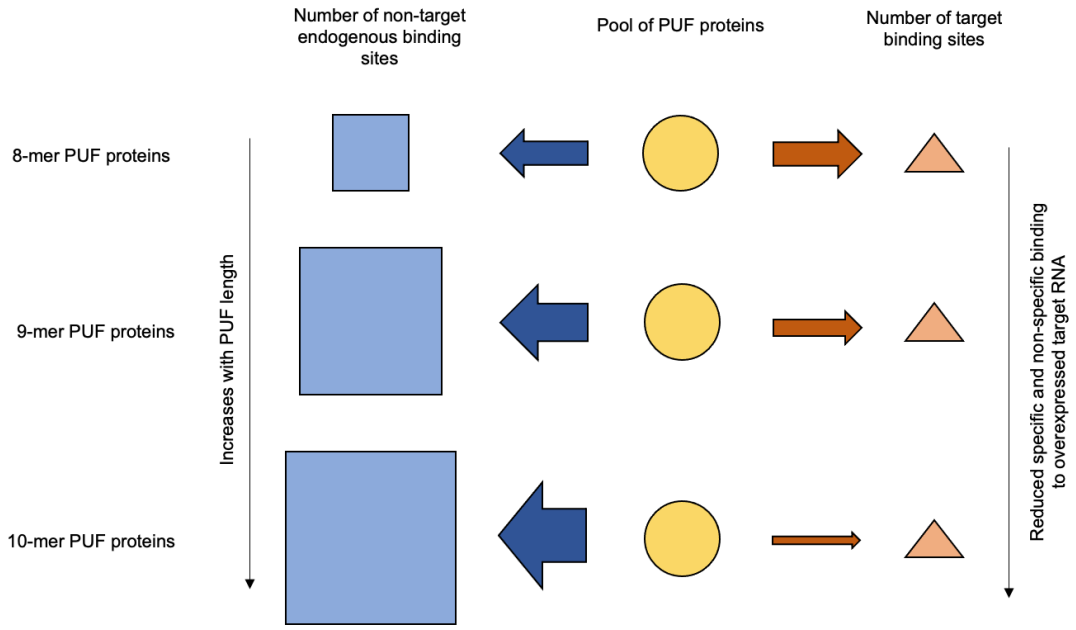


Figure 10.1: Schematic illustration of the sequestering hypothesis. As PUF length increases, the number of endogenous binding sites increases resulting in increased sequestering. This means there are fewer PUF proteins available to bind target RNA resulting in a decreased signal output. In other words, the binding equilibrium shifts to endogenous non-target sites rather than target sites.

This sequestering could explain the decrease in GFP fluorescence, when using 9- and 10-mer PUF domains (observed in later experiments using split-GFP), as binding to transfected RNA was greatly reduced compared to the 8-mer PUF domains. Increased binding to endogenous RNA did not result in increased GFP fluorescence likely because the proteins are sequestered away individually rather than in proximity. An alternative hypothesis could be that the 9- and 10-mer PUF domains simply do not bind their target RNA as effectively as the 8-mer PUF domains. To elucidate the true cause of these differences, high-throughput biophysical characterization should be carried out using a method such as HiTS-Kin[186,187]. Not only does this technique allow the measurement of binding affinity to the target RNA, HiTS-Kin allows an affinity landscape for RNA binding proteins to be established by testing binding against all potential RNA sequences for a given RNA binding protein in parallel. Thus, it would also

enable the contextualization of the NGS data to see which RNA sequences detected can actually be bound by the PUF domain and thus how many non-target RNAs they can effectively bind. Although this novel method was validated by *in vitro* fluorescence polarisation using $PUF2_8$, it should be further validated against gold standard CLIP methodologies as well to ensure that the detected non-target binding of the PUFs does indeed occur. Such data would demonstrate that the method is a suitable alternative to the very labour-intensive CLIP experiments. Lastly, the data for the single 8-, 9- and 10-mer are being extrapolated to all 8-, 9- and 10-mers used in this thesis (including those fused to effector domains). These experiments should therefore be repeated using all of the proteins used in this thesis to ensure that this generalization is valid and that fusion to an effector domain does not alter the binding properties of the PUF domains. Subsequently, *a split-GFP system was used to show that the concurrent binding of PUF proteins to RNA observed in vitro (Chapter 3) also occurs in cellulo.* Several variables influencing PUF binding and split-GFP reconstitution were optimized in this Chapter:

1. Spacing between the two PUF binding sites (0 – 10 nucleotide spacing)
2. Orientation of the split-GFP domains (outside- or inside-oriented)
3. Single (SBS) or multiple (MBS) binding sites for the PUFs on a single RNA
4. PUF domains of different lengths with different RNA target lengths (8-, 9- and 10-mers)

When using 8-mer PUF domains, higher GFP reconstitution was observed using multiple cognate binding sites compared to the scrambled RNA control. The data observed provided some initial insights into the complex network of variables governing concurrent PUF binding and effector domain reconstitution. Using inside-oriented split-GFP domains yielded significantly higher GFP intensities for the 8-mer PUF domains (compared to outside-oriented split-GFP domains) suggesting an improvement in orientation or stability of the re-constituted halves. Using SBS RNAs with inside-oriented split-GFP domains and 8-mer PUF domains showed greater GFP reconstitution for some spacings than the scrambled RNA control further indicating that the PUF domains can concurrently bind a single RNA molecule. However, using MBS RNAs did not always result in an increased GFP intensity compared to the SBS RNAs, suggesting that increasing the number of binding sites on a single RNA may not directly result in an increased GFP fluorescence. Additionally, using 9- and 10-mer PUF domains

resulted in a drastic decrease in GFP fluorescence compared to the 8-mer PUF domains to the extent that no differences between the constructs tested could be observed. Thus, an important experiment to carry out is using super-resolution microscopy techniques such as STED or TIRF[233]. Indeed, TIRF was the microscopy technique of choice for other labs using PUF domains fused to split-GFP to localize RNA[168-170] suggesting the signals from this system are intrinsically low. Although 8-mer PUF domains are compatible with confocal microscopy for split-GFP detection to an extent, 9- and 10-mer domains are not. Furthermore, it is unclear whether the best constructs in the split-GFP system are the best constructs in the split-TEV system and thus the apoptosis assays.

In Chapter 7, *using the amplificatory split-TEV system instead of split-GFP allowed for the detection of activity in the presence of cognate RNA that was higher than that in the presence of scrambled RNA for the 8- and 10-mer PUF domains. The 8-mer PUF fusion with inside-oriented split-TEV domains was used as a lead candidate for subsequent apoptosis assays.* This study demonstrates the first use of PUF domains fused to split-TEV halves. Previous efforts have focused on using directly dimerising domains such as the GCN4 coiled-coil domain[166], G-protein subunits[193] or the rapamycin-inducible FKBP-FRB system[55,234]. These dimerisation domains directly interact with each other, whereas the NASPER system uses PUF proteins that do not directly interact with each other, but rather come together via binding to an RNA molecule. This creates a lot more variables that can potentially influence the system (and need to be optimized), as mentioned for the split-GFP system. Using the split-TEV system yielded further information regarding the variables governing the NASPER system. Here as well, inside-oriented effector domains were better at producing cleavage activity, compared to outside-oriented domains, using the 8-mer and 10-mer PUF domains. Certain constructs (8-mer with 2 nucleotide spacing) produced greater activity than the scrambled RNA control, and the split-TEV domain only constructs, providing proof of concept that the NASPER system can target cells containing exogenous, overexpressed RNA. Using MBS RNAs did not result in the increased activity, as observed using split-GFP suggesting the total number of re-constituted effector domains are the same using SBS or MBS RNA even though multiple binding may be occurring. The effect of changing the spacing between the PUF binding sites also showed different trends for the different PUF proteins: activity for the 8-mer PUF proteins peaked at a short spacing and decreased as spacing was increased further,

whereas activity for the 10-mer fusions increased as RNA spacing increased. This difference is likely caused by steric hindrance and/or different orientations due to the differences in the overall structure of the different length PUF domains. While testing this system, a study using computational methods to reduce the background activity of the split-TEV halves was published, which led to testing of mutant split-TEV domains. The mutations indeed reduced the background activity as expected, which was thought to be due to reduced spontaneous reconstitution of the effector domains[197] and a reduction in catalytic activity since activity using cognate RNA was also reduced. Thus, there is great scope for the use of the NASPER system, and even further optimization using additional mutations tested in the Dolberg *et al.*[197] computational optimization study (such as N-TEV$_{75S}$ with C-TEV$_{wt}$, and N-TEV$_{wt}$ with C-TEV$_{163V}$. Both combinations show higher activity than the mutations tested, with a sufficiently low background signal). Further experiments also showed that the mutant split-TEV domains can be used with 8-mer PUF domains in order to specifically target endogenous *hTERT* RNA with significantly higher cleavage observed using cognate PUF domains compared to scrambled PUF domains, extending the initial proof of concept that NASPER can target exogenous RNA. Since this proof has been established, the proteins should be further tested in primary cancer cells rather than cultured cancer cells. Furthermore, these experiments should also be conducted using stem cells to assess the potential side-effects from treatment with the NASPER system as adult stem cells and germ cells also contain some full-length *hTERT* RNA[235]. However, one area significantly lacking in current literature is any structural or *in vitro* enzymatic data using the split-TEV domains. Consequently, there is no clear understanding of the folding of the split-TEV halves and whether the re-constituted enzyme has a similar structure to the intact TEV protease. Such data are essential in contextualizing the results of cell-based assays; non-specific cleavage of the TEV substrate was found when using 9-mer PUF domains suggesting a slightly different structure of the dimerized halves that may, in certain situations, allow promiscuous cleavage. Hence, future work should be directed towards improving the understanding of the split-TEV halves from a structural and biophysical perspective.

In Chapter 8, *the lead candidate from the optimization experiments was then tested in apoptosis assays, but no induction of apoptosis was observed relative to the procaspase-3 only controls.* Death of cells containing only the RNA of interest was monitored using mCherry

fluorescence as a measure of viability[222] (since assays of total cell viability did not show any differences relative to the procaspase-3 only controls likely due to the presence of significant background from RNA-negative cells). However, even with this assay neither wild-type nor mutant split-TEV domains showed increased cell death compared to the control. Troubleshooting using Western blot subsequently showed that the NASPER system is unable to cleave the procaspase-3 substrate, which is surprising since Gray *et al.*[55] showed that cleavage can be achieved using an FKBP-FRB split-TEV system. This work highlights a gap in the literature: - there is a clear substrate-dependent effect on the activity of the split-TEV system, which has also been observed previously[197,226], however the cause of this needs to be elucidated through structural studies of the split-TEV domains. Furthermore, the difference between the published split-TEV procaspase-3 cleavage and the lack of cleavage using NASPER must be due to the use of PUF domains causing sub-optimal dimerisation of split-TEV. Thus, structural work should follow to establish how this may occur. Alternatively, different effector domains could be used in the NASPER system, such as a receptor. Split-tyrosine phosphatases and kinases have been used by Camacho-Soto *et al.*[236] to tune molecular circuitry, and since phosphorylation is intricately linked to apoptosis, this system may be applicable to NASPER. In the extrinsic pathway, dephosphorylation of Y232 and Y291 of Fas by tyrosine phosphatase 1 results in a switch from the anti-apoptotic to pro-apoptotic state[237]. Thus, there is potential for these split-receptors to be used with PUF domains to induce apoptosis specifically in cancer cells.

Thus, based on the findings from this thesis, some final key conclusions can be drawn:

1. The split-GFP system did provide some insights into PUF concurrent binding to single RNA molecules, however these findings are not directly translatable to the split-TEV system, i.e, the brightest constructs from the split-GFP system are not the ones that show the most activity in the split-TEV system. This is likely because the split-GFP and split-TEV domains interact in different manners.

2. The split-TEV system showed that cells containing the RNA of interest can be specifically targeted using both transfected and endogenously overexpressed RNA, however these assays cannot be used to infer activity in apoptosis assays which is why

the lead construct selected from the split-TEV activity assays was not successful in apoptosis assays.

3. No apoptosis was detected using the split-TEV NASPER system, as it was unable to cleave the procaspase-3 substrate. This indicates a substrate dependency which has previously been noted. A follow-up study to the initial split-TEV procaspase-3 study by Gray *et al.*[55] also noted this substrate dependency as their split-TEV FKBP-FRB system was unable to efficiently cleave a different caspase-activated DNase (CAD) substrate to the extent that the group instead used a doxycycline-inducible intact TEV protease to achieve the cleavage required[226]. However, the original study[55] showed that procaspase-3 can be cleaved by split-TEV, and this proof of concept implies that the system can in principle work but is simply not working with the NASPER system.

Since the experiments in this thesis have shown that NASPER can target cells with the RNA of interest, and the Gray *et al.*[55] study showed that procaspase-3 can be cleaved by split-TEV, future work should focus on optimizing the NASPER system to induce apoptosis. There are different ways that this effect can be achieved:

1. Optimising the NASPER system to cleave the procaspase-3 construct. The experiments carried out in Chapter 7 to optimize the NASPER system using the mCherry substrate could be repeated to find the optimal PUF site spacing and split-TEV orientation to cleave the procaspase-3 construct. Since the TEV cleavage site in procaspase-3 is within the protein structure, spacings greater than the 2 nucleotides used in this thesis can be tested to better accommodate the bulkier procaspase-3 construct. Furthermore, Gray *et al.*[55] also found that swapping the N- and C-TEV fragments between the FKBP and FRB dimerisation domains had a significant effect on the activity of their system. This could be another variable that could be optimized in the NASPER system; rather than using PUF2-N-TEV and C-TEV-PUF1, the alternate combination, C-TEV-PUF2 and PUF1-N-TEV could be tested.

2. Different substrates, with a more accessible TEV cleavage site could be tested to induce apoptosis. Peptide-drug conjugates could serve as an alternate substrate[227]. A cytotoxic drug such as doxorubicin[238] or staurosporine can be inactivated by

conjugation to a peptide containing a TEV-cleavable linker (Figure 10.2). This site would be more representative of the flexible, exposed nature of the TEV site in the mCherry substrate and may translate better to the apoptosis assays compared to the procaspase-3 substrate used. The pro-drug would thus only be activated in cells where NASPER is active, i.e, cancer cells.

3. A split-caspase system can be used with the PUF domains instead of trying to activate a procaspase-3 through the use of a re-constituted protease. Chelur *et al.*[239] have developed a split-caspase-3 system which is able to kill cells upon re-constitution of the caspase-3 due to the induction of dimerisation. They found that there is no apoptotic activity without fusing the split halves to dimerisation inducing leucine zippers highlighting the specificity of this system. Using these domains fused to NASPER may be an exciting way forward to specifically kill cancer cells without needing to optimise protease cleavage and substrate type.

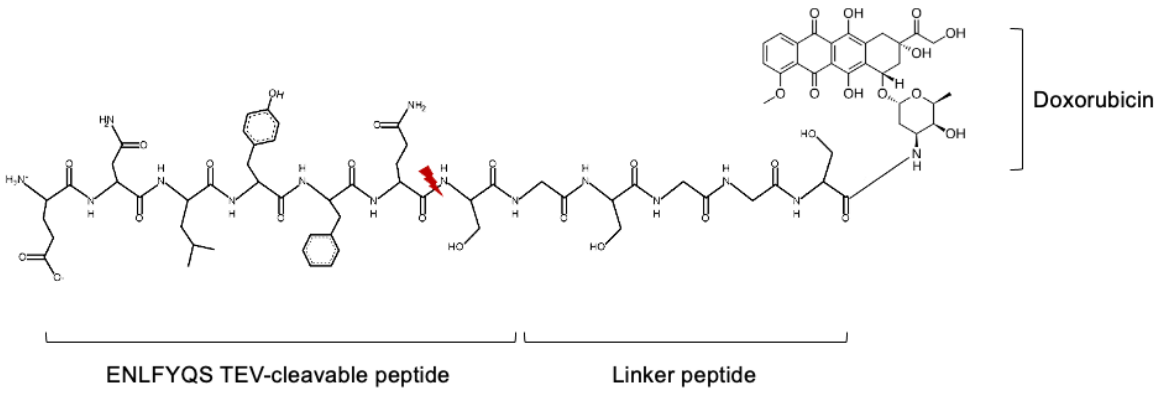


Figure 10.2: TEV-cleavable doxorubicin peptide-drug conjugate. TEV cleavage site is shown as a red lightning symbol

The NASPER system designed in this thesis represents a step forward in the fundamentals of rational drug design. While it fits in with current trends of designing multispecific therapeutics which can harness cellular machinery for therapeutic gain, it goes beyond this by expanding the types of cellular machinery that can be hijacked to include RNA as well. Most modern cancer drug development focuses on monoclonal antibodies to target epitopes on particular cancers either through intravenous delivery, or delivery in the form of CAR-T cells[240]. This

fits with a moving focus towards personalised clinical medicine in hopes of reducing side-effects and increasing efficacy, however the NASPER method deviates from this as it was intrinsically designed to be a broad-spectrum therapy. In this way it could overcome some of the limitations of these novel, personalised therapies such as cost – a single CAR-T therapy per patient can cost upwards of $500,000[240]. If this system is optimised to be used as a cancer therapeutic in the future, it could represent another shift towards the rational design of less personalised therapies with the high efficacy and specificity of current personalised therapies.

However, there are still limitations in the use of this system in patients. As is, it would require the delivery of at least two sets of genes, or proteins, into cancer cells. Delivery of the proteins themselves would pose a challenge due to their molecular weight, and the need to delivery them to the same cell[241]. Delivery of the genes would also pose similar challenges due to multiple genes, immunogenicity, long-term transgene expression issues, and efficient cell entry[242]. However, using multi-cistronic plasmid vectors, or adeno-associated viral vectors harbouring multiple genes in a single vector, to deliver these genes may overcome these barriers.

Despite being currently limited as a therapeutic option, the NASPER system could be applied for research purposes in the future as well. For example: specifically tuning protein activation based on the transcriptome allowing for much finer control of protein activation, activating phosphorylation related pathways in response to subtle transcriptomic changes[236], using RNases tethered to the PUF domains to provide a novel method of RNA silencing or tethering the PUF domains to modifying proteins such as m^6A demethylase for targeted RNA post-transcriptional modification[243].

Ultimately, although the experiments in this thesis provide evidence to show that the NASPER system can target cells containing an RNA of interest when tethered to a protease, further work is required to provide proof of concept that this can cause apoptosis of the targeted cells, and even if this is possible, further work will be needed to optimise delivery for clinical use of this system.

APPENDIX A (CHAPTER 3)

Appendix A contains the amino acid sequences for constructs used in experiments carried out in Chapter 3.

A.1 PROTEIN SEQUENCES

A.1.1 PUF MODULE SEQUENCES

Repeat	Recognition	AA sequence
1′		MGRSRLLEDFRNNRYPNLQLREIAG
1	A	HIMEFSQDQHGSRFIQLKLERATPAERQLVFNEILQ
1	G	HIMEFSQDQHGSRFIELKLERATPAERQLVFNEILQ
1	U	HIMEFSQDQHGNRFIQLKLERATPAERQLVFNEILQ
1	C	HIMEFSQDQHGSRFIRLKLERATPAERQLVFNEILQ
2	A	AAYQLMVDVFGCYVIQKFFEFGSLEQKLALAERIRG
2	G	AAYQLMVDVFGSYVIEKFFEFGSLEQKLALAERIRG
2	U	AAYQLMVDVFGNYVIQKFFEFGSLEQKLALAERIRG
2	C	AAYQLMVDVFGSYVIRKFFEFGSLEQKLALAERIRG
3	A	HVLSLALQMYGCRVIQKALEFIPSDQQNEMVRELDG
3	G	HVLSLALQMYGSRVIEKALEFIPSDQQNEMVRELDG
3	U	HVLSLALQMYGNRVIQKALEFIPSDQQNEMVRELDG
3	C	HVLSLALQMYGSRVIRKALEFIPSDQQNEMVRELDG
4	A	HVLKCVKDQNGCHVVQKCIECVQPQSLQFIIDAFKG
4	G	HVLKCVKDQNGSHVVEKCIECVQPQSLQFIIDAFKG
4	U	HVLKCVKDQNGNHVVQKCIECVQPQSLQFIIDAFKG
4	C	HVLKCVKDQNGSHVVRKCIECVQPQSLQFIIDAFKG
4	A	HVLKCVKDQNGCYVVQKCIECVQPQSLQFIIDAFKG
4	G	HVLKCVKDQNGSYVVEKCIECVQPQSLQFIIDAFKG
4	U	HVLKCVKDQNGNYVVQKCIECVQPQSLQFIIDAFKG
4	C	HVLKCVKDQNGSYVVRKCIECVQPQSLQFIIDAFKG
5	A	QVFALSTHPYGCRVIQRILEHCLPDQTLPILEELHQ
5	G	QVFALSTHPYGSRVIERILEHCLPDQTLPILEELHQ
5	U	QVFALSTHPYGNRVIQRILEHCLPDQTLPILEELHQ
5	C	QVFALSTHPYGSRVIRRILEHCLPDQTLPILEELHQ
6	A	HTEQLVQDQYGCYVIQHVLEHGRPEDKSKIVAEIRG
6	G	HTEQLVQDQYGSYVIEHVLEHGRPEDKSKIVAEIRG
6	U	HTEQLVQDQYGNYVIQHVLEHGRPEDKSKIVAEIRG
6	C	HTEQLVQDQYGSYVIRHVLEHGRPEDKSKIVAEIRG
7	A	NVLVLSQHKFACNVVQKCVTHASRTERAVLIDEVCTMNDGPHS
7	G	NVLVLSQHKFASNVVEKCVTHASRTERAVLIDEVCTMNDGPHS
7	U	NVLVLSQHKFANNVVQKCVTHASRTERAVLIDEVCTMNDGPHS
7	C	NVLVLSQHKFASYVVRKCVTHASRTERAVLIDEVCTMNDGPHS
8	A	ALYTMMKDQYACYVVQKMIDVAEPGQRKIVMHKIRP
8	G	ALYTMMKDQYASYVVEKMIDVAEPGQRKIVMHKIRP

8	U	ALYTMMKDQYANYVVQKMIDVAEPGQRKIVMHKIRP
8	C	ALYTMMKDQYASYVVRKMIDVAEPGQRKIVMHKIRP
8′		HIATLRKYTYGKHILAKLEKYYMKNGVDLG

APPENDIX B (CHAPTER 5)

Appendix B contains the primer sequences used for NEBuilder cloning of the constructs in Chapter 5.

B.1 NEBUILDER DATA

B.1.1 PRIMERS AND TEMPLATES THAT THEY CAN AMPLIFY

Primer Names	5'-3' Primer Sequences (*OVERLAP*-**ANNEAL**)	Plasmid Template
3xFLAG_1fwd	*CACCACACTGGACTAGTATG***GACTACAAAGACCATG**	3xFLAG
3xFLAG_1rev	*GGACACGGTGCCCTG***CTTGTCATCGTCATCCTTG**	3xFLAG
p6HIVPR_1fwd	*GATGACGATGACAAG***CAGGGCACCGTGTCCTTC**	p6-HIVPR/p6-HIVPR$_{D25N}$
p6HIVPR_1rev	*CTGATCAGCGGTTTAAACTT***TCATCAAAAGTTCAGTGTGGCG**	p6-HIVPR/p6-HIVPR$_{D25N}$
GST fwd	*CACCACACTGGACTAGTATG***TCCCCTATACTAGGTTATTG**	GST
GST rev	*GTCATGGTCTTTGTAGTC***CGATTTTGGAGGATGGTC**	GST
3xFLAG_2fwd	*ACCATCCTCCAAAATCG***GACTACAAAGACCATG**	3xFLAG
3xFLAG_2rev	*AAGGACACGGTGCCCTG***CTTGTCATCGTCATCCTTG**	3xFLAG
3xHA_fwd	*CACCACACTGGACTAGTATG***TACCCGTACGACGTCCCTG**	3xHA
3xHA_rev	*CAGCAGCCGAGATCTGCC***GGCGTAATCCGGCACATC**	3xHA
PUF1_fwd	*ATGTGCCGGATTACGCC***GGCAGATCTCGGCTGCTG**	PUF1
PUF1/2_rev	*CTGATCAGCGGTTTAAACTT***TCATCAGCCCAGGTCCAC**	PUF1/PUF2
3xFLAG_3rev	*AGCACGCGAGATCTGCC***CTTGTCATCGTCATCCTTG**	3xFLAG
PUF2_fwd	*AAGGATGACGATGACAAG***GGCAGATCTCGGCTGCTG**	PUF2
PUF1/2_flex_rev	*GCCGCTAGACTTTCCCTC***GCCCAGGTCCACGCCGTTCTTC**	PUF1/PUF2
PUF1/2_rig_rev	*GCTTCTTTAGCGGCGGCTTC***GCCCAGGTCCACGCCGTTCTTC**	PUF1/PUF2
flexi_ PUF1/2_fwd	*AGAAGTACTACATGAAGAACGGCGTGGACCTGGGC***GAGGGAAAGTCTAGCGGCTC**	Flexible Linker
Flexi_rev	*GTGATCTGAGGGAAGTTGAAGGACACGGTGCCCTG***TGTGCTCTTGCTCTCGCTGC**	Flexible Linker
Rigid_flexi_fwd	*CTAAAGAAGCTGCTGCCAAAGAGGCCGCTGCCAAG***GAGGGAAAGTCTAGCGGCTC**	Flexible Linker

Rigid_fwd	*AGAAGTACTACATGAAGAACGGCGTGGACCTGGGC***GAAGCCGCCGCTAAAGAAGC**	Rigid Linker
Flexi_rigid_rev	*TTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTC***CTTGGCAGCGGCCTCTTTGGC**	Rigid Linker
HIVPR_rigid_rev	*GTGATCTGAGGGAAGTTGAAGGACACGGTGCCCTG***CTTGGCAGCGGCCTCTTTGGC**	Rigid Linker
Flexi_HIVPR_fwd	*AGCGAGAGCAAGAGCACA***CAGGGCACCGTGTCCTTCAAC**	p6-HIVPR/p6-HIVPR$_{D25N}$
Rigid_HIVPR_fwd	*CCAAAGAGGCCGCTGCCAAG***CAGGGCACCGTGTCCTTCAAC**	p6-HIVPR/p6-HIVPR$_{D25N}$

B.1.2 FRAGMENTS AND PRIMER PAIRS TO AMPLIFY THEM

Fragment Names	Primer Pairs	Annealing T_m (°C)	Template
B	3xFLAG_1fwd and 3xFLAG_1rev	53	3xFLAG
C	p6HIVPR_1fwd and p6HIVPR_1rev	67	p6HIVPR
C_1	p6HIVPR_1fwd and p6HIVPR_1rev	67	$p6HIVPR_{D25N}$
D	GST fwd and GST rev	58	GST
E	3xFLAG2_fwd and 3xFLAG2_rev	53	3xFLAG
F	3xHA_fwd and 3xHA_rev	68	3xHA
G	PUF1_fwd and PUF1/2_rev	68	PUF1
H	3xFLAG_1fwd and 3xFLAG_3rev	53	3xFLAG
I	PUF2_fwd and PUF1/2_rev	68	PUF2
J	PUF1_fwd and PUF1/2_flex_rev	68	PUF1
J_1	PUF2_fwd and PUF1/2_flex_rev	68	PUF2
K	flexi_HIVPR_fwd and p6HIVPPR_1rev	67	p6HIVPR
K_1	flexi_HIVPR_fwd and p6HIVPPR_1rev	67	$p6HIVPR_{D25N}$
L	flexi_PUF1/2_fwd and flexi_rev	69	Flexible Linker
M	PUF1_fwd and PUF1/2_rigid_rev	68	PUF1
M_1	PUF2_fwd and PUF1/2_rigid_rev	68	PUF2
N	rigid_HIVPR_fwd and p6HIVPPR_1rev	67	p6HIVPR
N_1	rigid_HIVPR_fwd and p6HIVPPR_1rev	67	$p6HIVPR_{D25N}$
O	rigid_fwd and HIVPR_rigid_rev	71	Rigid Linker
P	rigid_fwd and flexi_rigid_rev	71	Rigid Linker
Q	rigid_flexi_fwd and flexi_rev	69	Flexible Linker

B.1.3 FRAGMENT COMPOSITION OF CONTRUCTS

Construct Number	Fragments Required
1	A + B + C
2	A + B + C_1
3	A + D + E + C
4	A + D + E + C_1
5	A + F + J + L + K
6	A + H + J_1 + L + K
7	A + F + J + L + K_1
8	A + H + J_1 + L + K_1

B.2 MUTAGENESIS

B.2.1 HIVPR MUTAGENESIS PRIMERS

Mutant Name	Primer Pair
R8Q	5'-gtgaccaggggctgcttccacaggg-3' 5'-ccctgtggaagcagcccctggtcac-3
T26A	5'-cggcgcctgcatccagcagggcctct-3' 5'-agaggccctgctggatgcaggcgccg-3'
D29N	5'-cgatcactgtatcgttggcgcctgtatccag-3' 5'-ctggatacaggcgccaacgatacagtgatcg-3'
R87K	5'-gatctgggtcagcagattcttgccgatgatgttgataggt-3' 5'-acctatcaacatcatcggcaagaatctgctgacccagatc-3'
ΔC	5'-ggtttaaactttcatcaggcgccgatctgggtc-3' 5'-gacccagatcggcgcctgatgaaagtttaaacc-3'

APPENDIX C (CHAPTER 6, 7 and 8)

Appendix C contains the primer sequences used in NEBuilder cloning carried out in Chapters 6, 7 and 8.

C.1 PRIMERS FOR NEBUILDER FRAGMENT GENERATION

C.1.1 FIRST SET OF CONSTRUCTS

PRIMER NAME	PRIMER SEQUENCE (5'-3')	PRIMER TEMPLATE	FRAGMENT PRODUCED
GST_rev_NGFP	ccttggacacCTTGTCATCGTCATCCTTG	GST-FLAG constructs	GST_NGFP
GST_rev_CGFP	tgccgttcttCTTGTCATCGTCATCCTTG	GST-FLAG constructs	GST_CGFP
GST_rev_NTEV	ggctctcgccCTTGTCATCGTCATCCTTG	GST-FLAG constructs	GST_NTEV
GST_rev_CTEV	tcatgctcttCTTGTCATCGTCATCCTTG	GST-FLAG constructs	GST_CTEV
PUF_rev_NGFP	ccttggacacTGTGCTCTTGCTCTCGCTG	PUF1/2 constructs	PUF2_NGFP
PUF_rev_CGFP	tgccgttcttTGTGCTCTTGCTCTCGCTG	PUF1/2 constructs	PUF1_CGFP
PUF_rev_NTEV	ggctctcgccTGTGCTCTTGCTCTCGCTG	PUF1/2 constructs	PUF2_NTEV
PUF_rev_CTEV	tcatgctcttTGTGCTCTTGCTCTCGCTG	PUF1/2 constructs	PUF1_CTEV
NTEV_GST_fwd	cgatgacaagGGCGAGAGCCTGTTCAAG	N-TEV constructs	NTEV_GST
NTEV_PUF_fwd	caagagcacaGGCGAGAGCCTGTTCAAG	N-TEV constructs	NTEV_PUF2
CTEV_GST_fwd	cgatgacaagAAGAGCATGAGCAGCATG	C-TEV constructs	CTEV_GST
CTEV_PUF_fwd	caagagcacaAAGAGCATGAGCAGCATG	C-TEV constructs	CTEV_PUF1
NGFP_GST_fwd	cgatgacaagGTGTCCAAGGGCGAAGAAC	N-GFP constructs	NGFP_GST
NGFP_PUF_fwd	caagagcacaGTGTCCAAGGGCGAAGAAC	N-GFP constructs	NGFP_PUF2
CGFP_GST_fwd	cgatgacaagAAGAACGGCATCAAAGTG	C-GFP constructs	CGFP_GST
CGFP_PUF_fwd	caagagcacaAAGAACGGCATCAAAGTG	C-GFP constructs	CGFP_PUF1

NEBuilder_fwd	CACCACACTGGACTAGTATG	All constructs	**N/A**
GFP/TEV_rev	CTGATCAGCGGTTTAAACTT	All constructs	**N/A**

* NEBuilder_fwd was used alongside the 'rev' primers
** GFP/TEV_rev was used alongside the 'fwd' primer

C.1.2 FLEXIBLE SPLIT-GFP SYSTEM CONSTRUCTS

PRIMER NAME	PRIMER SEQUENCE	PRIMER TEMPLATE	FRAGMENT PRODUCED
Flag_flexi_rev	5'-TGTGCTCTTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTCCTTGTCATCGTCATCCTTG-3'	GST-N-GFP	FLAG_flexi GST_FLAG_flexi
Flexi_NGFP_fwd	5'-GAGGGAAAGTCTAGCGGCTCTGGCAGCGAGAGCAAGAGCACAGTGTCCAAGGGCGAAGAAC-3'	GST-N-GFP	Flexi_NGFP
HA_flexi_rev	5'-TGTGCTCTTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTCGGCATAGTCGGGGACGTC-3'	HA-C-GFP	HA_flexi
Flexi_CGFP_fwd	5'-GAGGGAAAGTCTAGCGGCTCTGGCAGCGAGAGCAAGAGCACAAAGAACGGCATCAAAGTG-3'	HA-C-GFP	Flexi-CGFP

* NEBuilder_fwd was used alongside the 'rev' primers
** GFP/TEV_rev was used alongside the 'fwd' primers

C.1.3 PUF1 CONSTRUCTS WITH N-TERMINUS SPLIT DOMAIN

PRIMER NAME	PRIMER SEQUENCE	PRIMER TEMPLATE	FRAGMENT PRODUCED
Flexi_PUF1_fwd	5'-GAGGGAAAGTCTAGCGGCTCTGGCAGCGAGAGCAAGAGCACAGGCAGATCTCGGCTG-3'	PUF1 constructs	Flexi_PUF1
CG_flexi_rev	5'-TGTGCTCTTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTCCTTGTACAGCTCATCCATGC-3'	Flexi-C-GFP constructs	Flexi_CG_flexi
CT_flexi_rev	5'-TGTGCTCTTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTCCACCATGAACACCTTGTG-3'	CT constructs	CT_flexi

* NEBuilder_fwd was used alongside the 'rev' primers
** GFP/TEV_rev was used alongside the 'fwd' primer

C.1.4 FRAGMENT COMBINATIONS FOR SPLIT SYSTEMS CONSTRUCTS

CONSTRUCT NAME	FRAGMENT 1	FRAGMENT 2
GST-N-TEV	GST_NTEV	NTEV_GST
GST-C-TEV	GST_CTEV	CTEV_GST
GST-N-GFP	GST_NGFP	NGFP_GST
GST-C-GFP	GST_CGFP	CGFP_GST
FLEXI-N-GFP	FLAG_flexi	Flexi_NGFP
FLEXI-C-GFP	HA_flexi	Flexi_CGFP
GST-FLEXI-N-GFP	GST_FLAG_flexi	Flexi_NGFP
GST-FLEXI-C-GFP	GST_FLAG_flexi	Flexi_CGFP

PUF2-NTEV	PUF2_NTEV	NTEV_PUF2
PUF2-NGFP	PUF2_NGFP	NGFP_PUF2
PUF1-CTEV	PUF1_CTEV	CTEV_PUF1
PUF1-CGFP	PUF1_CGFP	CGFP_PUF1
CTEV-PUF1	Flexi_PUF1	CT_flexi
CGFP-PUF1	Flexi_PUF1	Flexi_CG_flexi

C.2 PRIMERS FOR mCherry CONSTRUCTS

C.2.1 TEV-CLEAVABLE mCherry

mCherry_T_0_SBS_Cog8 was used as the template for all reactions. Site-directed mutagenesis was used to generate the following constructs.

Mutant Name	Primer Pair
mCherry_T_2_SBS_Cog8	5'-gaaggcagcagctttgtcgcagattggcataatctgg-3' 5'-ccagattatgccaatctgcgacaaagctgctgccttc-3'
mCherry_T_4_SBS_Cog8	5'-ccagattatgccaatctgcgacaaaaagctgctgccttc-3' 5'-gaaggcagcagcttttttgtcgcagattggcataatctgg-3'
mCherry_T_6_SBS_Cog8	5'-ccagattatgccaatctgcgacaaaaaaagctgctgccttc-3' 5'-gaaggcagcagctttttttgtcgcagattggcataatctgg-3'
mCherry_T_8_SBS_Cog8	5'-ccagattatgccaatctgcgacaaaaaaaaagctgctgccttc-3' 5'-gaaggcagcagctttttttttgtcgcagattggcataatctgg-3
mCherry_T_10_SBS_Cog8	5'-gaaggcagcagctttttttttttgtcgcagattggcataatctgg-3' 5'-ccagattatgccaatctgcgacaaaaaaaaaaagctgctgccttc-3'
mCherry_T_10_SBS_Scr8	5'-catcaagaagacccggagtttttttttttctcgattgattggcataa-3' 5'-ttatgccaatcaatcgagaaaaaaaaaactccgggtcttcttgatg-3'

Mutant Name	Primer Pair
mCherry_T_0_SBS_Cog9	5'- CACCACACTGGACTAGTATG -3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTGTCGTCAGATTGGCATAATCTG GCACGTC -3'
mCherry_T_2_SBS_Cog9	5’-aaggcagctagctttgtcgtcagattggcataatct-3’ 5’-agattatgccaatctgacgacaaagctagctgccttt-3’
mCherry_T_4_SBS_Cog9	5’-agattatgccaatctgacgacaaaaagctagctgccttt-3’ 5’-aaggcagctagctttttgtcgtcagattggcataatct-3’
mCherry_T_6_SBS_Cog9	5’-agattatgccaatctgacgacaaaaaaagctagctgccttt-3’ 5’-aaggcagctagctttttttgtcgtcagattggcataatct-3’

mCherry_T_8_SBS_Cog9	5’-ccagattatgccaatctgacgacaaaaaaaaaagctagctgccttc-3’ 5’-gaaggcagctagctttttttttgtcgtcagattggcataatctgg-3’
mCherry_T_10_SBS_Cog9	5’-ccagattatgccaatctgacgacaaaaaaaaaaagctagctgccttc-3’ 5’-gaaggcagctagctttttttttttgtcgtcagattggcataatctgg-3’

Mutant Name	Primer Pair
mCherry_T_0_SBS_Cog10	5'- CACCACACTGGACTAGTATG -3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTGTCGATCAGATTGGCATAAT CTGGCACGTC -3'
mCherry_T_2_SBS_Cog10	5'-aggcagcatagctttgtcgatcagattggcataatc-3' 5'-gattatgccaatctgatcgacaaagctatgctgcct-3'
mCherry_T_4_SBS_Cog10	5'-gattatgccaatctgatcgacaaaaagctatgctgcct-3' 5'-aggcagcatagcttttgtcgatcagattggcataatc-3'
mCherry_T_6_SBS_Cog10	5'-aggcagcatagctttttttgtcgatcagattggcataatc-3' 5'-gattatgccaatctgatcgacaaaaaaagctatgctgcct-3'
mCherry_T_8_SBS_Cog10	5'-cagattatgccaatctgatcgacaaaaaaaaaagctatgctgcctt-3' 5'-aaggcagcatagctttttttttgtcgatcagattggcataatctg-3'
mCherry_T_10_SBS_Cog10	5'-cagattatgccaatctgatcgacaaaaaaaaaaaagctatgctgcctt-3' 5'-aaggcagcatagctttttttttttgtcgatcagattggcataatctg-3'

NEBuilder was used to generate the following constructs by insertion of the PCR product into a pcDNA3.1(-) vector.

Clone Name	Reverse Primer
mCherry_T_0_MBS_Cog8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTGTCGCAGAGAAGGCAGCAGCTGTCGCAG-3’
mCherry_T_2_MBS_Cog8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTTTGTCGCAGAGAAGGCAGCAGCTTTGTCGCAG-3’
mCherry_T_4_MBS_Cog8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTTTTTGTCGCAGAGAAGGCAGCAGCTTTTTGTCGCAG-3'
mCherry_T_6_MBS_Cog8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTTTTTTTGTCGCAGAGAAGGCAGCAGCTTTTTTTGTCGCAG-3’
mCherry_T_8_MBS_Cog8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTTTTTTTTTGTCGCAGAGAAGGCAGCAGCTTTTTTTTTTGTCGCAG-3’
mCherry_T_10_MBS_Cog8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTTTTTTTTTTTTGTCGCAGAGAAGGCAGCAGCTTTTTTTTTTTTGTCGCAG-3’
mCherry_T_10_MBS_Scr8	5’-CTGATCAGCGGTTTAAACTTTCATCAAGAAGACCCGGAGTTTTTTTTTTTCTCGATTGAGAAGACCCGGAGTTTTTTTTTTCTCGATTG-3’

Clone Name	Reverse Primer
mCherry_T_0_MBS_Cog9	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTGTCGTCAGAGAAGGCAGCTAGCTGTCGTCAG -3’
mCherry_T_2_MBS_Cog9	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTTTGTCGTCAGAGAAGGCAGCTAGCTTTGTCGTCAG -3’
mCherry_T_4_MBS_Cog9	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTTTTTGTCGTCAGAGAAGGCAGCTAGCTTTTTGTCGTCAG -3’
mCherry_T_6_MBS_Cog9	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTTTTTTTGTCGTCAGAGAAGGCAGCTAGCTTTTTTTGTCGTCAG -3’
mCherry_T_8_MBS_Cog9	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTTTTTTTTTGTCGTCAGAGAAGGCAGCTAGCTTTTTTTTTGTCGTCAG -3’
mCherry_T_10_MBS_Cog9	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTTTTTTTTTTTTGTCGTCAGAGAAGGCAGCTAGCTTTTTTTTTTTTGTCGTCAG -3’

Clone Name	Reverse Primer
mCherry_T_0_MBS_Cog10	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTGTCGATCAGAGAAGGCAGCATAGCTGTCGATCAG -3’
mCherry_T_2_MBS_Cog10	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTTTGTCGATCAGAGAAGGCAGCATAGCTTTGTCGATCAG -3’
mCherry_T_4_MBS_Cog10	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTTTTTGTCGATCAGAGAAGGCAGCATAGCTTTTTGTCGATCAG -3’
mCherry_T_6_MBS_Cog10	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTTTTTTTGTCGATCAGAGAAGGCAGCATAGCTTTTTTTGTCGATCAG -3’
mCherry_T_8_MBS_Cog10	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTTTTTTTTTGTCGATCAGAGAAGGCAGCATAGCTTTTTTTTTGTCGATCAG -3’
mCherry_T_10_MBS_Cog10	5’- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTTTTTTTTTTTGTCGATCAGAGAAGGCAGCATAGCTTTTTTTTTTTGTCGATCAG -3’

***** NEBuilder_fwd was used as the forward primer in all cases

****** The template used was the corresponding SBS clone

C.2.2 mCherry without TEV Cleavage Site

NEBuilder was used to generate the following constructs by insertion of the PCR product into a pcDNA3.1(-) vector.

Mutant Name	Primer Pair
mCherry_N_0_SBS _Cog9	5'- CACCACACTGGACTAGTATGGTGAGCAAGGGCGAGG-3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCTAGCTGTCGTCAGATTCTTGTACAGCTCGT CCATGCCG -3'
mCherry_N_0_SBS _Cog10	5'- CACCACACTGGACTAGTATGGTGAGCAAGGGCGAGG-3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTGTCGATCAGATTCTTGTACAGCTC GTCCATGCCG -3'
mCherry_N_2_SBS _Cog8	5'-CACCACACTGGACTAGTATGGGCTCTGGCGGCATGGTG-3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCAGCTTTGTCGCAGATTCTTGTACAGCTCGT CCATTCCGCC-3'
mCherry_N_10_SBS _Cog10	5'-CACCACACTGGACTAGTATGGGCTCTGGCGGCATGGTG-3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGGCAGCATAGCTTTTTTTTTTTGTCGATCAGATTCT TGTACAGCTCGTCCATTCCGCC-3'
mCherry_N_Scr_1	5'-CACCACACTGGACTAGTATGGGCTCTGGCGGCATGGTG-3' 5'- CTGATCAGCGGTTTAAACTTTCATCAAGAAGACCCGGAGTTTTTTTTTTCTCGATTGATTCTTGTA CAGCTCGTCCATTCCGCC-3'

APPENDIX D

Appendix D contains all the primer sequences used for NEBuilder cloning of the constructs for endogenous targeting in Chapter 7.

D.1 NEBUILDER PRIMERS

Fragment Name	Primer Pair
P28 Fragments	5'- CACCACACTGGACTAGTATG - 3' 5'- TGTGCTCTTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTCGCCGAGGTCCACGCCG -3'
P18 Fragments	5'- GAGGGAAAGTCTAGCGGCTCTGGCAGCGAGAGCAAGAGCACAGGCAGAAGCAGACTGCTGG -3' 5'- CTGATCAGCGGTTTAAACTT -3'
NT Mutant Fragment	5'- GAGGGAAAGTCTAGCGGCTCTGGCAGCGAGAGCAAGAGCACAGGCGAGAGCCTGTTCAAG -3' 5'- CTGATCAGCGGTTTAAACTT - 3'
CT Mutant Fragment	5'- CACCACACTGGACTAGTATG -3' 5'- TGTGCTCTTGCTCTCGCTGCCAGAGCCGCTAGACTTTCCCTCCACCATGAACACCTTGTGG -3'

- Appropriate templates containing the sequences to be amplified were used

CONSTRUCT NAME	FRAGMENT 1	FRAGMENT 2
P28C-NTEVm	P28C	NTEVm
P28S-NTEVm	P28S	NTEVm
CTEVm-P18C	P18C	CTEVm
CTEVm-P18S	P18S	CTEVm

APPENDIX E

Appendix E contains data showing that scrambled RNA 1 and scrambled RNA 2 were nearly identical in both split-GFP reconstitution, and TEV cleavage, assays.

E.1 CONFOCAL MICROSCOPY OF DIFFERENT SCRAMBLED RNA (OUTSIDE-ORIENTED)

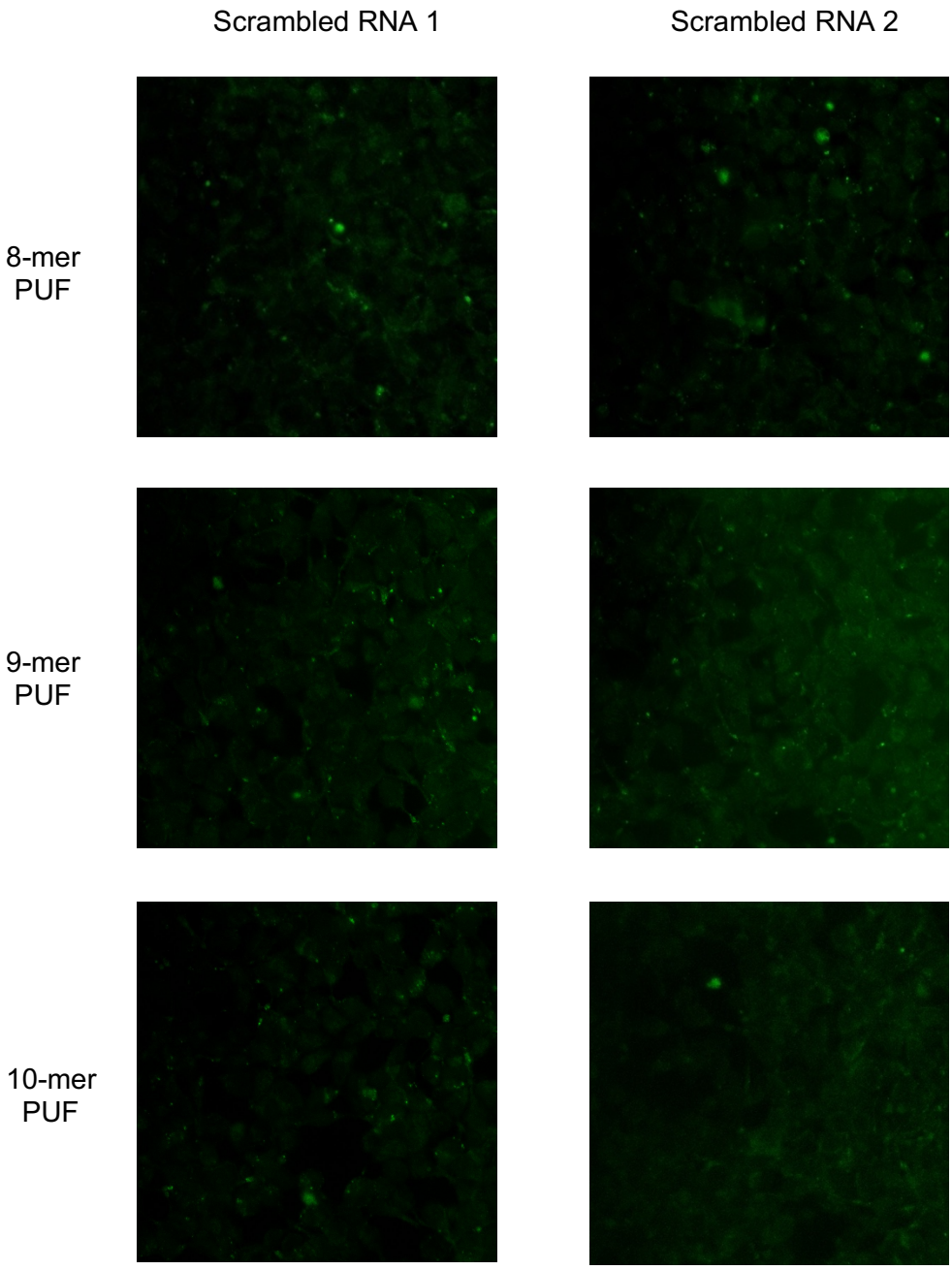


Figure E.1: Confocal microscopy of outside-oriented PUF-split-GFP fusions transfected with either scrambled RNA 1 or 2. No differences were observed between the two RNAs.

E.2 CONFOCAL MICROSCOPY OF DIFFERENT SCRAMBLED CONSTRUCTS (INSIDE-ORIENTED)

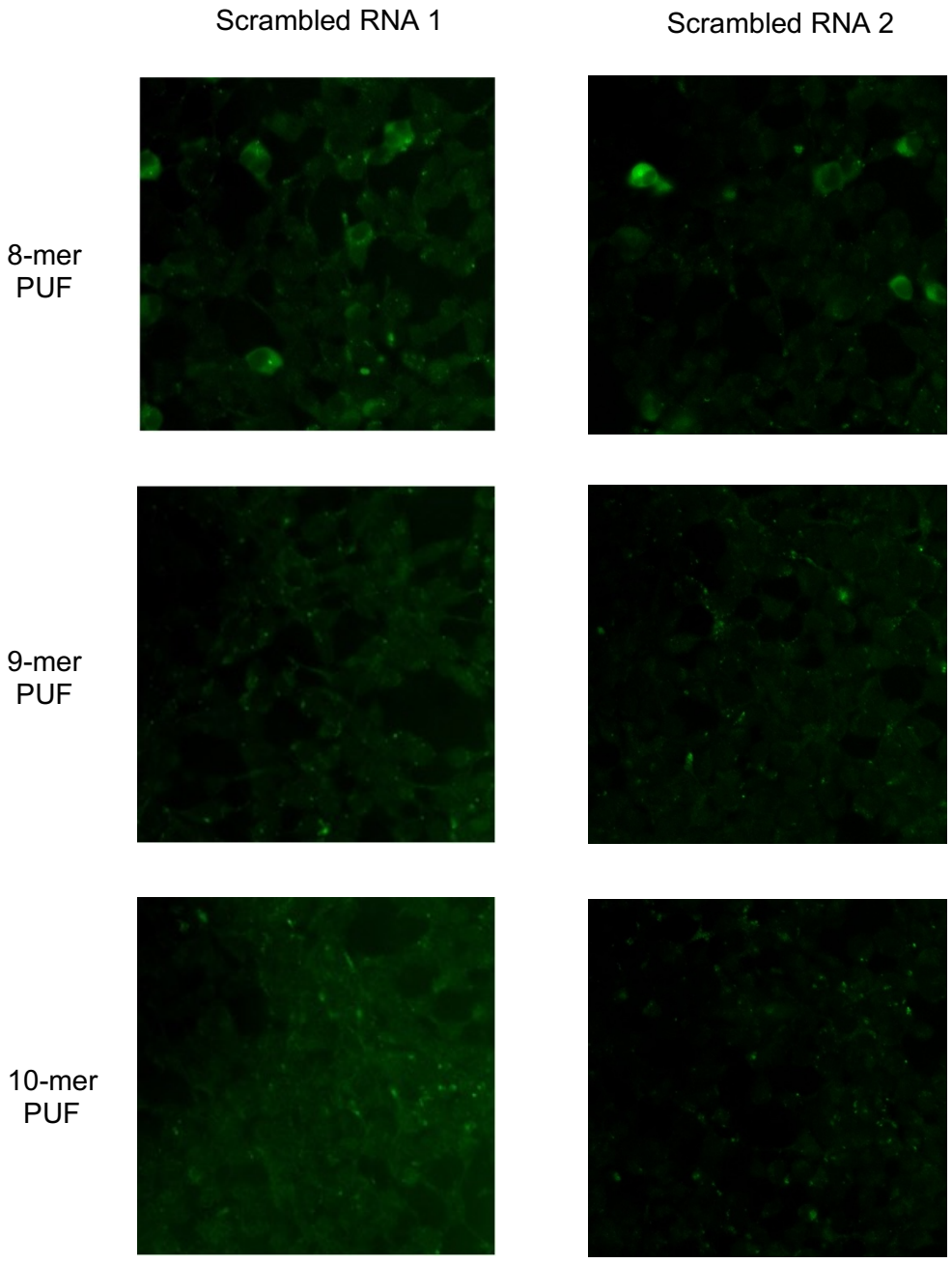


Figure E.2: Confocal microscopy of inside-oriented PUF-split-GFP fusions transfected with either scrambled RNA 1 or 2. Minimal differences were observed between the two RNAs.

E.3 DENSITOMETRY OF DIFFERENT SCRAMBLED CONSTRUCTS (OUTSIDE-ORIENTED)

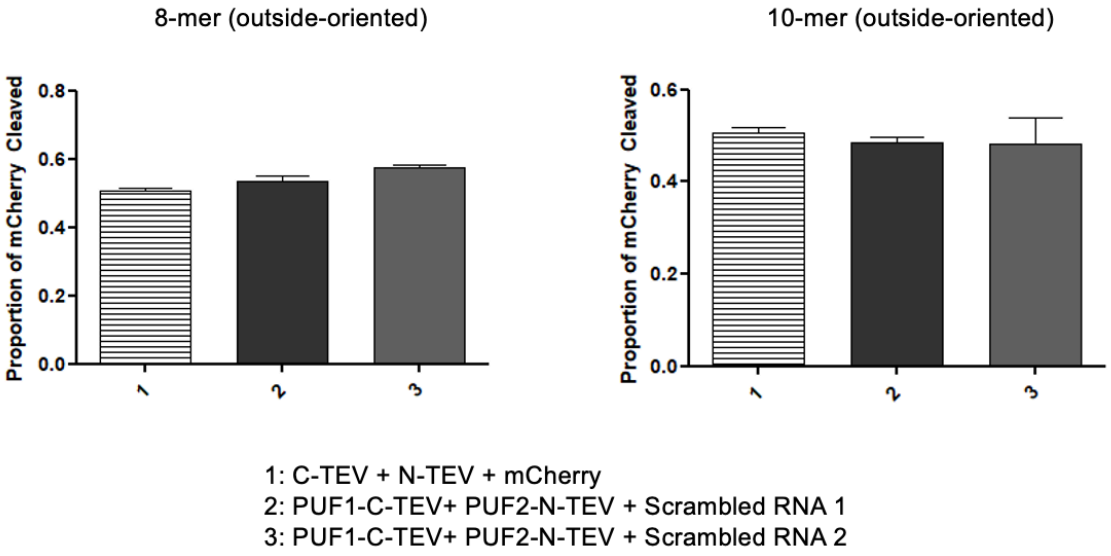


Figure E.3: Western blot densitometry of outside-oriented PUF-split-TEV fusions transfected with either scrambled RNA 1 or 2 and the split-TEV domains lacking the PUF domain. No differences were observed between the two RNAs and the PUF negative control for either the 8- or 10-mer PUF domains tested.

E.4 DENSITOMETRY OF DIFFERENT SCRAMBLED CONSTRUCTS (INSIDE-ORIENTED)

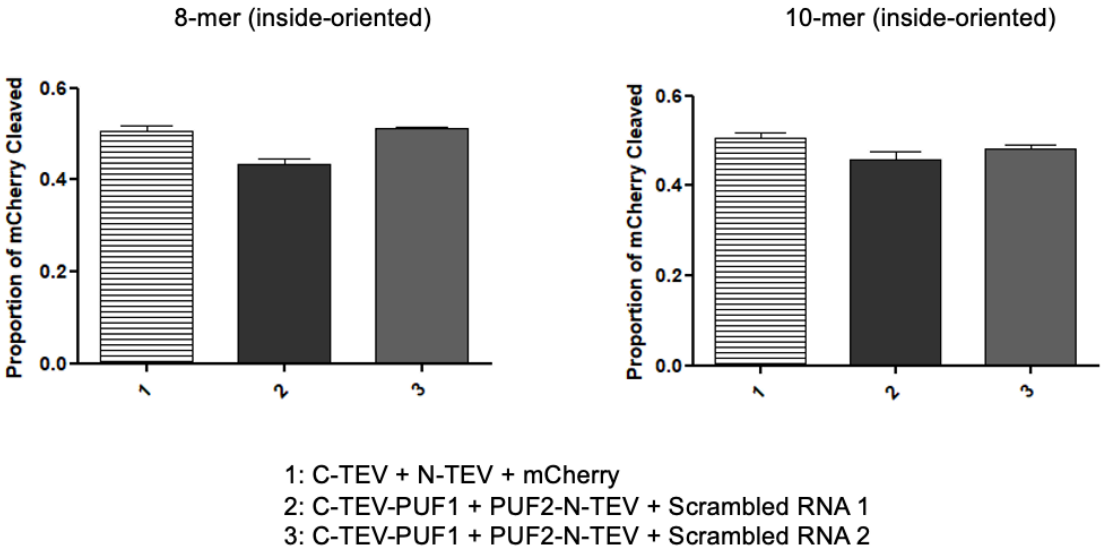


Figure E.4: Western blot densitometry of inside-oriented PUF-split-TEV fusions transfected with either scrambled RNA 1 or 2 and the split-TEV domains lacking the PUF domain. Minimal differences were observed between the two RNAs and the PUF negative control for either the 8- or 10-mer PUF domains tested. Differences did not affect statistical tests carried out in Chapter 7 for the respective constructs.

REFERENCES:

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018;68(6): 394–424. Available from: doi:10.3322/caac.21492
2. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *The Lancet*. 2018;391(10125): 1023–1075. Available from: doi:10.1016/S0140-6736(17)33326-3
3. Naghavi M, Abajobir AA, Abbafati C, Abbas KM, Abd-Allah F, Abera SF, et al. Global, regional, and national age-sex specifc mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*. 2017;390(10100): 1151–1210. Available from: doi:10.1016/S0140-6736(17)32152-9
4. Zhu L, Finkelstein D, Gao C, Shi L, Wang Y, López-Terrada D, et al. Multi-organ Mapping of Cancer Risk. *Cell*. 2016;166(5): 1132-1146.e7. Available from: doi:10.1016/j.cell.2016.07.045
5. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011. p. 646–674. Available from: doi:10.1016/j.cell.2011.02.013
6. Shay JW and Wright WE. Hayflick, His Limit and Cellular Ageing. *Nature Reviews Molecular Cell Biology*. 2000;1(1): 72 – 76. Available from: doi:10.1038/35036093
7. Heng HHQ, Stevens JB, Bremer SW, Liu G, Abdallah BY, Ye CJ. Evolutionary Mechanisms and Diversity in Cancer. *Advances in Cancer Research*. 2011. p. 217–253. Available from: doi:10.1016/B978-0-12-387688-1.00008-9
8. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World Journal of Clinical Oncology. 2014*. p. 412–424. Available from: doi:10.5306/wjco.v5.i3.412
9. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014;512(7513): 155–160. Available from: doi:10.1038/nature13600

10. Andor N, Maley CC and Ji HP. Genomic instability in cancer: Teetering on the limit of tolerance. *Cancer Research*. 2017;77(9): 2179 – 2185. Available from: doi: 10.1158/0008-5472.CAN-16-1553
11. Sonugur FG and Akbulut H. The Role of Tumor Microenvironment in Genomic Instability of Malignant Tumors. *Frontiers Genetics*. 2019;10(1063): 1 – 10. Available from: doi: 10.3389/fgene.2019.01063
12. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5): 979–993. Available from: doi:10.1016/j.cell.2012.04.024
13. Gay L, Baker AM, Graham TA. *Tumour Cell Heterogeneity*. 2016. Available from: doi:10.12688/f1000research.7210.1
14. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. 2017;168(4): 613–628. Available from: doi:10.1016/j.cell.2017.01.018
15. Oh BY, Shin HT, Yun JW, Kim KT, Kim J, Bae JS, et al. Intratumor heterogeneity inferred from targeted deep sequencing as a prognostic indicator. *Scientific Reports*. 2019;9(1). Available from: doi:10.1038/s41598-019-41098-0
16. Housman G, Byler S, Heerboth S, Lapinska K, Longacre M, Snyder N, et al. Drug resistance in cancer: An overview. *Cancers*. 2014. p. 1769–1792. Available from: doi:10.3390/cancers6031769
17. Yaswen P, MacKenzie KL, Keith WN, Hentosh P, Rodier F, Zhu J, Firestone GL, Matheu A, Carnero A, Bilsland A *et al*. Therapeutic targeting of replicative immortality. *Seminars in Cancer Biology*. 2015;35(Suppl): S104 – S128. Available from: doi: 10.1016/j.semcancer.2015.03.007
18. Perri F, Pisconti S and Scarpati GDV. P53 mutations and cancer: a tight linkage. *Annals of Translational Medicine*. 2016;4(24): 522. Available from: doi: 10.21037/atm.2016.12.40
19. Ozaki T and Nakagawara A. Role of p53 in Cell Death and Human Cancers. *Cancers*. 2011;3(1): 994 – 1013. Available from: doi: 10.3390/cancers3010994
20. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, et al. Specific association of human telomerase activity with immortal cells and cancer. *Science*. 1994;266(1): 2011–2015. Available from: doi:10.1126/science.7605428.

21. Hahn WC, Stewart SA, Brooks MW, York SG, Eaton E, Kurachi A, et al. Inhibition of telomerase limits the growth of human cancer cells. *Nature Medicine*. 1999;5(10): 1164–1170. Available from: doi:10.1038/13495
22. Wong MS, Wright WE and Shay JW. Alternative Splicing of Telomerase: A New Paradigm. *Trends in Genetics*. 2015;30(10): 430 – 438. Available from: doi:10.1016/j.tig.2014.07.006
23. Rubtsova MP, Vasilkova DP, Malyavko AN, Naraikina YV, Zvereva MI and Dontsova OA. Telomere Lengthening and Other Functions of Telomerase. *Acta Naturae*. 2012;4(2): 44 – 61. Available from: PMID:22872811
24. Teixeira MT, Arneric M, Sperisen P and Lingner J. Telomere Length Homeostasis Is Achieved via a Switch between Telomerase- Extendible and -Nonextendible States. *Cell*. 2004;117(3): 323 – 335. Available from: doi: 10.1016/S0092-8674(04)00334-4
25. Okamoto K, Seimiya H. Revisiting Telomere Shortening in Cancer. *Cells*. 2019;8(2): 107. Available from: doi:10.3390/cells8020107
26. Leão R, Apolónio JD, Lee D, Figueiredo A, Tabori U, Castelo-Branco P. Mechanisms of human telomerase reverse transcriptase (hTERT) regulation: Clinical impacts in cancer. *Journal of Biomedical Science*. 2018. Available from: doi:10.1186/s12929-018-0422-8
27. Salgado C, Roelse C, Nell R, Gruis N, van Doorn R and van der Velden P. Interplay between TERT promoter mutations and methylation culminates in chromatin accessibility and TERT expression. *PLoS One*. 2020;15(4): e0231418. Available from: doi: 10.1371/journal.pone.0231418
28. Vinagre J, Almeida A, Pópulo H, Batista R, Lyra J, Pinto V, et al. Frequency of TERT promoter mutations in human cancers. *Nature Communications*. 2013;4. Available from: doi:10.1038/ncomms3185
29. Liu X, Wang Y, Chang G, Wang F, Wang F and Geng X. Alternative Splicing of hTERT Pre-mRNA: A Potential Strategy for the Regulation of Telomerase Activity. *International Journal of Molecular Sciences*. 2017;18(3): 567. Available from: doi:10.3390/ijms18030567
30. Keith WN and Hoare SF. Detection of Telomerase hTERT Gene Expression and Its Splice Variants by RT-PCR. *Molecular Diagnosis of Cancer*. 2004;97(1): 297 – 309. Available from: doi:10.1385/1-59259-760-2:297
31. Hrdličková R, Nehyba J and Bose Jr, HR. Alternatively Spliced Telomerase Reverse Transcriptase Variants Lacking Telomerase Activity Stimulate Cell Proliferation.

Molecular and Cellular Biology. 2012;32(21): 4283 – 4296. Available from: doi:10.1128/MCB.00550-12

32. Yi X, Shay JW and Wright WE. Quantitation of Telomerase Components and hTERT mRNA Splicing Patterns in Immortal Human Cells. *Nucleic Acids Research*. 2001;29(23): 4818 – 4825. Available from: doi:10.1093/nar/29.23.4818
33. Hamazaki Y, Sekai M and Minato N. Medullary thymic epithelial stem cells: role in thymic epithelial cell maintenance and thymic involution. *Immunology Reviews*. 2016;271(1):38 – 55. Available from: doi: 10.1111/imr.12412
34. Chia WK, Cheah FC, Aziz NHA, Kampan NC, Shuib S, Khong TY, Tan GC and Wong YP. A Review of Placenta and Umbilical Cord-Derived Stem Cells and the Immunomodulatory Basis of Their Therapeutic Potential in Bronchopulmonary Dysplasia. *Frontiers in Pediatrics*. 2021;9(615508): 1 – 10. Available from: doi: 10.3389/fped.2021.615508
35. Bhang DH, Kim BJ, Kim BG, Schadler K, Baek KH, Kim YH, Hsiao W, Ding BS, Rafii S, Weiss MJ *et al*. Testicular endothelial cells are a critical population in the germline stem cell niche. *Nature Communications*. 2018;9(1): 4379. Available from: doi: 10.1038/s41467-018-06881-z
36. Choi J, Southworth LK, Sarin KY, Venteicher AS, Ma W, Chang W, Cheung P, Jun S, Artandi MK, Shah N, Kim SK and Artandi SE. TERT Promotes Epithelial Proliferation through Transcriptional Control of a Myc- and Wnt-Related Developmental Program. *PLoS Genetics*. 2008;4(1): e10. Available from: doi: 10.1371/journal.pgen.0040010
37. Martínez P and Blasco MA. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. *Nature Reviews Cancer*. 2011;11(1): 161 – 176. Available from: doi: 10.1038/nrc3025
38. Park J, Venteicher AS, Hong JY, Choi K, Jun S, Shkreli M, Chang W, Meng Z, Cheung P, Ji H *et al*. Telomerase modulates Wnt signalling by association with target gene chromatin. *Nature*. 2009;460(1): 66 – 72. Available from: doi: 10.1038/nature08137
39. Ghosh A, Saginc G, Leow SC, Khattar E, Shin EN, Yan TD, Wong M, Zhang Z, Li G, Sung WK *et al*. Telomerase directly regulates NF-κB-dependent transcription. *Nature Cell Biology*. 2012;14(1): 1270 – 1281. Available from: doi: 10.1038/ncb2621

40. Saretzki G. Extra-telomeric Functions of Human Telomerase: Cancer, Mitochondria and Oxidative Stress. *Current Pharmaceutical Design*. 2014;20(41): 6386 – 6403. Available from: doi: 10.2174/1381612820666140630095606

41. Bufalo DD, Rizzo A, Trisciuoglio D, Cardinali G, Torrisi MR, Zangemeister-Wittke U, Zupi G and Biroccio A. Involvement of hTERT in apoptosis induced by interference with Bcl-2 expression and function. *Cell Death and Differentiation*. 2005;12(1): 1429 – 1438. Available from: doi: 10.1038/sj.cdd.4401670

42. Jin Y, You L, Kim HJ and Lee HW. Telomerase Reverse Transcriptase Contains a BH3-Like Motif and Interacts with BCL-2 Family Members. *Molecular Cells*. 2018;41(7): 684 – 694. Available from: doi: 10.14348/molcells.2018.0206

43. Massard C, Zermati Y, Pauleau AL, Larochette N, Métivier D, Sabatier L, Kroemer G and Soria JC. hTERT: a novel endogenous inhibitor of the mitochondrial cell death pathway. *Oncogene*. 2006;25(1): 4505 – 4514. Available from: doi: 10.1038/sj.onc.1209487

44. Ivancich M, Schrank Z, Wojdyla L, Leviskas B, Kuckovic A, Sanjali A, et al. *Treating cancer by targeting telomeres and telomerase*. [Online] Antioxidants. MDPI AG; 2017. Available from: doi:10.3390/antiox6010015

45. Bentham Science Publisher BSP. Telomerase Inhibition in Cancer Therapeutics: Molecular-Based Approaches. *Current Medicinal Chemistry*. 2006;13(24): 2875–2888. Available from: doi:10.2174/092986706778521887

46. Xu H, Di Antonio M, McKinney S, Mathew V, Ho B, O'Neil NJ, et al. CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nature Communications*. 2017;8. Available from: doi:10.1038/ncomms14432

47. Chiappori AA, Kolevska T, Spigel DR, Hager S, Rarick M, Gadgeel S, et al. A randomized phase II study of the telomerase inhibitor imetelstat as maintenance therapy for advanced non-small-cell lung cancer. *Annals of Oncology*. 2015;26(2): 354–362. Available from: doi:10.1093/annonc/mdu550

48. Mosoyan G, Kraus T, Ye F, Eng K, Crispino JD, Hoffman R, et al. Imetelstat, a telomerase inhibitor, differentially affects normal and malignant megakaryopoiesis. *Leukemia*. 2017;31(11): 2458–2467. Available from: doi:10.1038/leu.2017.78

49. Wang X, Hu CS, Petersen B, Qiu J, Ye F, Houldsworth J, et al. Imetelstat, a telomerase inhibitor, is capable of depleting myelofibrosis stem and progenitor cells. *Blood Advances*. 2018;2(18): 2378–2388. Available from: doi:10.1182/bloodadvances.2018022012

50. Hu Y, Shi G, Zhang L, Li F, Jiang Y, Jiang S, et al. Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. *Scientific Reports*. 2016;6. Available from: doi:10.1038/srep32280
51. Kosiol N, Juranek S, Brossart P, Heine A and Paeschke K. G-quadruplexes: a promising target for cancer therapy. *Molecular Cancer*. 2021;20(40): 1 – 10. Available from: doi: 10.1186/s12943-021-01328-4
52. Savva L and Georgiades SN. Recent Developments in Small-Molecule Ligands of Medicinal Relevance for Harnessing the Anticancer Potential of G-Quadruplexes. *Molecules*. 2021;26(841): 1 – 10. Available from: doi: 10.3390/molecules26040841
53. Chanda D, Hensel JA, Higgs JT, Grover R, Kaza N, Ponnazhagan S. Effects of cellular methylation on transgene expression and site-specific integration of adeno-associated virus. *Genes*. ; 2017;8(9). Available from: doi:10.3390/genes8090232
54. Muraki K, Nyhan K, Han L and Murnane JP. Mechanisms of telomere loss and their consequences for chromosome instability. *Frontiers in Oncology*. 2012;2(135): 1 – 10. Available from: doi: 10.3389/fonc.2012.00135
55. Gray DC, Mahrus S and Wells JA. Activation of Specific Apoptotic Caspases with an Engineered Small Molecule-Activated Protease. *Cell*. 2010;142(4): 637 – 646. Available from: doi:10.1016/j.cell.2010.07.014
56. Cheng CL, Thike AA, Tan SYJ, Chua PJ, Bay BH and Tan PH. Expression of FGFR1 is an independent prognostic factor in triple-negative breast cancer. *Breast Cancer Research and Treatment*. 2015;151(1):99 – 111. Available from: doi: 10.1007/s10549-015-3371-x
57. Xu J, Chen Y and Olopade OI. MYC and Breast Cancer. *Genes & Cancer*. 2010;1(6): 629 – 640. Available from: doi: 10.1177/1947601910378691
58. Weier HG and Mao JH. Meta-analysis of Aurora Kinase A (AURKA) Expression Data Reveals a Significant Correlation Between Increased AURKA Expression and Distant Metastases in Human ER-positive Breast Cancers. *Journal of Data Mining, Genomics and Proteomics*. 2013;4(1): 127. Available from: doi: 10.4172/2153-0602.1000127
59. Buza N. HER2 Testing in Endometrial Serous Carcinoma: Time for Standardized Pathology Practice to Meet the Clinical Demand. *Archives of Pathology & Laboratory Medicine*. 2021;145(6): 687 – 691. Available from: doi: 10.5858/arpa.2020-0207-RA

60. Smith V, Hobbs S, Court W, Eccles S, Workman P and Kelland LR. ErbB2 overexpression in an ovarian cancer cell line confers sensitivity to the HSP90 inhibitor geldanamycin. *Anticancer Research*. 2002;22(4): 1993 – 1999. Available from: PMID: 12174876

61. Spano JP, Lagorce C, Atlan D, Milano G, Domont J, Benamouzig R, Attar A, Benichou J, Martin A, Morere JF *et al*. Impact of EGFR expression on colorectal cancer patient prognosis and survival. *Annals in Oncology*. 2005;16(1): 102 – 108. Available from: doi: 10.1093/annonc/mdi006

62. Rochlitz CF, Herrmann R and de Kant E. Overexpression and Amplification of c-myc during Progression of Human Colorectal Cancer. *Oncology*. 1996;53(1): 448 – 454. Available from: doi: 10.1159/000227619

63. Zhang L, Hou Y, Ashktorab, Gao L, Xu Y, Wu K, Zhai J and Zhang L. The impact of C-MYC gene expression on gastric cancer cell. *Molecular and Cellular Biochemistry*. 2010;344(0): 125 – 135. Available from: doi: 10.1007/s11010-010-0536-0

64. Kanayama K, Imai H, Usugi E, Shiraishi T, Hirokawa YS and Watanabe M. Association of HER2 gene amplification and tumor progression in early gastric cancer. *Virchows Archive*. 2018;473(1): 559 – 565. Available from: doi: 10.1007/s00428-018-2433-y

65. Dragoj M, Bankovic J, Podolski-Renic A, Buric SS, Pesic M, Tanic M and Stankovic T. Association of overexpressed *MYC* gene with altered *PHACTR3* and *E2F4* genes contributes to non-small cell lung carcinoma pathogenesis. *Journal of Medical Biochemistry*. 2019;37(1): 188 – 195. Available from: doi: 10.2478/jomb-2018-0022

66. Wang YL, Uhara H, Yamazaki Y, Nikaido T and Saida T. Immunohistochemical detection of CDK4 and 9 16INK4 proteins in cutaneous malignant melanoma. *British Journal of Dermatology*. 1996;134(2): 269 – 275. Available from: doi: 10.1111/j.1365-2133.1996.tb07612.x

67. Kwei KA, Shain AH, Bair R, Montgomery K, Karikari CA, van de Rijn M, Hidalgo M, Maitra A, Bashyam MD and Pollack JR. SMURF1 Amplification Promotes Invasiveness in Pancreatic Cancer. *PLoS ONE*. 2011;6(8): e23924. Available from: doi: 10.1371/journal.pone.0023924

68. Urbanucci A, Sahu B, Seppälä J, Larjo A, Latonen LM, Waltering KK, Tammela TLJ, Vessella RL, Lädesmäki H, Jänne OA and Visakorpi T. Overexpression of androgen receptor enhances the binding of the receptor to the chromatin in prostate cancer. *Oncogene*. 2012;31(17): 2153 – 2163. Available from: doi: 10.1038/onc.2011.401

69. Kaminskyy VO and Zhivotovsky B. Cell Death-based Treatment of Various Diseases: A Fifty Year Journey. *Cell Death and Disease*. 2018;9(110): 1 – 4. Available from: doi:10.1038/s41419-017-0168-3

70. Pers JO, Daridon C, Devauchelle V, Jousse S, Saraux A, Jamin C and Youinou P. BAFF overexpression is associated with autoantibody production in autoimmune diseases. *Annals of the New York Academy of Sciences*. 2005;1050(1): 34 – 39. Available from: doi:10.1196/annals.1313.004

71. Stefanis L. α-Synuclein in Parkinson's Disease. *Cold Spring Harbour Perspectives in Medicine*. 2012;2(2): a009399. Available from: doi: 10.1101/cshperspect.a009399

72. Deshaies R. Multispecific drugs herald a new era of biopharmaceutical innovation. *Nature*. 2020;580(1): 329 – 338. Available from: doi:10.1038/s41586-020-2168-1

73. Nair JK, Willoughby JLS, Chan A, Charisse K, Alam MR, Wang Q, Hoekstra M, Kandasamy P, Kel'in AV, Milstein S, Taneja N *et al*. Multivalent N-Acetylgalactosamine-Conjugated siRNA Localizes in Hepatocytes and Elicits Robust RNAi-Mediated Gene Silencing. *Journal of the American Chemistry Society*. 2014;136(49): 16958 – 16961. Available from: doi:10.1021/ja505986a

74. Neri D. Antibody-Cytokine Fusions: Versatile Products for the Modulation of Anticancer Immunity. *Cancer Immunology Research*. 2019;7(3): 348 – 354. Available from: doi: 10.1158/2326-6066.CIR-18-0622

75. Vigneron N. Human Tumor Antigens and Cancer Immunotherapy. *Biomed Research International*. 2015;2015(948501): 1 – 10. Available from: doi:10.1155/2015/948501

76. Liu J, Farmer Jr JD, Lane WS, Friedman J, Weissman I and Schreiber SL. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell*. 1991;66(4): 807 – 815. Available from: doi: 10.1016/0092-8674(91)90124-h

77. Paiva SL and Crews C. Targeted protein degradation: elements of PROTAC design. *Current Opinion in Chemical Biology*. 2019;50(1): 111 – 119. Available from: doi: 10.1016/j.cbpa.2019.02.022

78. Diamante A, Chaturbedy PK, Rowling PJE, Kumita JR, Eapen RS, McLaughlin SH, de la Roche M, Perez-Riba A and Itzhaki L. Engineering mono- and multi-valent inhibitors on a modular scaffold. *Chemical Science*. 2020;12(3): 880 – 895. Available from: doi: 10.1039/d0sc03175e

79. Gonzalvez F and Ashkenazi A. New insights into apoptosis signaling by Apo2L/TRAIL. *Oncogene*. 2010;29(34): 4752 – 4765. Available from: doi: 10.1038/onc.2010.221

80. Lim B, Greer Y, Lipkowitz S and Takebe N. Novel Apoptosis-Inducing Agents for the Treatment of Cancer, a New Arsenal in the Toolbox. *Cancers (Basel)*. 2019;11(8): 1087. Available from: doi: 10.3390/cancers11081087

81. Fischer U and Schulze-Osthoff K. Apoptosis-based therapies and drug targets. *Cell Death and Differentiation*. 2005;12(1): 942 – 961. Available from: doi: 10.1038/sj.cdd.4401556

82. Roberts AW, Davids MS, Pagel JM, Kahl BS, Puvvada SD, Gerecitano JF, Kipps TJ, Anderson MA, Brown JR, Gressick L *et al*. Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *New England Journal of Medicine*. 2016;374(4): 311 – 322. Available from: doi: 10.1056/NEJMoa1513257

83. Pfeffer CM and Singh ATK. Apoptosis: A Target for Anticancer Therapy. *International Journal of Molecular Sciences.* 2018;19(2): 448. Available from: doi: 10.3390/ijms19020448

84. Barata P, Sood AK and Hong DS. RNA-targeted therapeutics in cancer clinical trials: Current status and future directions. *Cancer Treatment Reviews*. 2016;50(1): 35 – 47. Available from: doi: 10.1016/j.ctrv.2016.08.004

85. Tolcher AW, Papadopoulos KP, Patnaik A, Rasco DW, Martinez D, Wood DL, Fielman B, Sharma M, Janisch LA, Brown BD *et al*. Safety and activity of DCR-MYC, a first-in-class Dicer-substrate small interfering RNA (DsiRNA) targeting MYC, in a phase I study in patients with advanced solid tumors. *Journal of Clinical Oncology*. 2015;33(15): 11006. Available from: doi: 10.1200/jco.2015.33.15_suppl.11006

86. Zhang L, Liao Y and Tang L. MicroRNA-34 family: a potential tumor suppressor and therapeutic candidate in cancer. *Journal of Experimental and Clinical Cancer Research*. 2019;53(1): 1 – 10. Available from: doi: 10.1186/s13046-019-1059-5

87. Yan X, Liang H, Deng T, Zhi K, Zhang S, Wang N, Jiang X, Wang X, Liu R, Zen K *et al*. The identification of novel targets of miR-16 and characterization of their biological functions in cancer cells. *Molecular Cancer*. 2013;12(92): 1 – 10. Available from: doi: 10.1186/1476-4598-12-92

88. Nishanth MJ and Simon Bindu. Functions, mechanisms and regulation of Pumilio/Puf family RNA binding proteins: a comprehensive review. *Molecular Biology Reports*. 2020;47(1): 785 – 807. Available from: doi: 10.1007/s11033-019-05142-6

89. Lu G, Hall TMT. Alternate Modes of Cognate RNA Recognition by Human PUMILIO Proteins. *Structure*. 2011;19(3): 361 – 367. Available from: doi: 10.1016/j.str.2010.12.019
90. Abil Z, Denard CA, Zhao H. Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *Journal of Biological Engineering*. 2014;8(1): 7. Available from: doi:10.1186/1754-1611-8-7
91. Qiu C, McCann KL, Wine RN and Hall TMT. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. *PNAS*. 2014;111(52): 18554 – 18559. Available from: doi: 10.1073/pnas.1407634112
92. Zhang J, McCann KL, Qiu C, Gonzalez LE, Baserga SJ and Hall TMT. Nop9 is a PUF-like protein that prevents premature cleavage to correctly process pre-18S rRNA. *Nature Communications*. 2016;7(13085): 1 – 10. Available from: doi: 10.1038/ncomms13085
93. Qiu C, Dutcher RC, Porter DF, Arava Y, Wickens M, Hall TMT. Distinct RNA-binding modules in a single PUF protein cooperate to determine RNA specificity. *Nucleic Acids Research*. 2019; Available from: doi:10.1093/nar/gkz583
94. Wang X and Voronina E. Diverse Roles of PUF Proteins in Germline Stem and Progenitor Cell Development in C. elegans. *Frontiers in Cell and Developmental Biology*. 2020;8(29): 1 – 10. Available from: doi: 10.3389/fcell.2020.00029
95. Forbes A and Lehmann R. Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. *Development*. 1998;125(4): 679 – 690. Available from: doi: 10.1242/dev.125.4.679
96. Haupt KA, Law KT, Enright AL, Kanzler CR, Shin H, Wickens M and Kimble J. A PUF Hub Drives Self-Renewal in Caenorhabditis elegans Germline Stem Cells. *Genetics*. 2020;214(1): 147 – 161. Available from: doi: 10.1534/genetics.119.302772
97. Zhang M, Chen D, Xia J, Han W, Cui X, Neuenkirchen N, Hermes G, Sestan N and Lin H. Post-transcriptional regulation of mouse neurogenesis by Pumilio proteins. *Genes and Development*. 2017;31(13): 1354 – 1369. Available from: doi: 10.1101/gad.298752.117
98. Naudin C, Hattabi A, Michelet F, Miri-Nezhad A, Benyoucef A, Pflumio F, Guilonneau F, Fichelson S, Vigon I, Dusanter-Fourt I and Lauret E. PUMILIO/FOXP1 signaling drives expansion of hematopoietic stem/progenitor and leukemia cells. *Blood*. 2017;129(18): 2493 – 2506. Available from: doi: 10.1182/blood-2016-10-747436

99. Wang X, McLachlan J, Zamore PD and Hall TMT. Modular Recognition of RNA by a Human Pumilio-Homology Domain. *Cell*. 2002;110(4): 501 – 512. Available from: doi: 10.1016/S0092-8674(02)00873-5

100. Filipovska A, Razif MFM, Nygård KKA and Rackham O. A universal code for RNA recognition by PUF proteins. *Nature Chemical Biology*. 2011;7(1): 425 – 427. Available from: doi: 10.1038/nchembio.577

101. Dong S, Wang Y, Cassidy-Amstutz C, Lu G, Bigler R, Jezyk MR, Li C, Hall TMT and Wang Z. Specific and Modular Binding Code for Cytosine Recognition in Pumilio/FBF (PUF) RNA-binding Domains. *Journal of Biological Chemistry*. 2011;286(30): 26732 – 26742. Available from: doi: 10.1074/jbc.M111.244889

102. Zhao YY, Mao MW, Zhang WJ, Wang J, Li HT, Yang Y, et al. Expanding RNA binding specificity and affinity of engineered PUF domains. *Nucleic Acids Research*. 2018;46(9): 4771–4782. Available from: doi:10.1093/nar/gky134

103. Adamala KP, Martin-Alarcon DA, Boyden ES. Programmable RNA-binding Protein Composed of Repeats of a Single Modular Unit. *PNAS*. 2016;113(19): E2579 – E2588. Available from: doi:10.1073/pnas.1519368113

104. Gupta YK, Nair DT, Wharton RP and Aggarwal AK. Structures of Human Pumilio with Noncognate RNAs Reveal Molecular Mechanisms for Binding Promiscuity. *Structure*. 2008;16(4): 549 – 557. Available from: doi:10.1016/j.str.2008.01.006

105. Dedow LK and Bailey Serres J. Searching for a Match: Structure, Function and Application of Sequence-Specific RNA-Binding Proteins. *Plant and Cell Physiology*. 2019;60(9): 1927 – 1938. Available from: doi: 10.1093/pcp/pcz072

106. Weidmann CA, Raynard NA, Blewett NH, Van Etten J and Goldstrohm AC. The RNA binding domain of Pumilio antagonizes poly-adenosine binding protein and accelerates deadenylation. *RNA*. 2014;20(8): 1298 – 1319. Available from: doi: 10.1261/rna.046029.114

107. Jarmoskaite I, Denny SK, Vaidyanathan PP, Becker WR, Andreasson JOL, Layton CJ, Kappel K, Shivashankar V, Sreenivasan R, Das R *et al*. A Quantitative and Predictive Model for RNA Binding by Human Pumilio Proteins. *Molecular Cell*. 2019;74(5): 966 – 981. Available from: doi: 10.1016/j.molcel.2019.04.012

108. K. Shinoda, S. Tsuji, S. Futaki and M. Imanishi. Nested PUF Proteins: Extending Target RNA Elements for Gene Regulation. *ChemBioChem*. 2017;19(2): 171 – 176. Available from: doi:10.1002/cbic.201700458

109. Sugimoto M, Suda A, Futaki S and Imanishi M. Effective RNA Regulation by Combination of Multiple Programmable RNA-Binding Proteins. *Applied Sciences*. 2020;10(19): 6803. Available from: doi:10.3390/app10196803

110. Abil Z, Gumy LF, Zhao H and Hoogenraad CC. Inducible Control of mRNA Transport Using Reprogrammable RNA-Binding Proteins. *ACS Synthetic Biology*. 2017;6(6): 950 – 956. Available from: doi:10.1021/acssynbio.7b00025

111. Hayashi H, Takamune N, Nirasawa T, Aoki M, Morishita Y, Das D, et al. Dimerization of HIV-1 protease occurs through two steps relating to the mechanism of protease dimerization inhibition by darunavir. *Proceedings of the National Academy of Sciences of the United States of America*. 2014;111(33): 12234–12239. Available from: doi:10.1073/pnas.1400027111

112. Nguyen HLT, Nguyen TT, Vu QT, Le HT, Pham Y, Trinh PL, Bui TP and Phan TN. An Efficient Procedure for the Expression and Purification of HIV-1 Protease from Inclusion Bodies. *Protein Expression and Purification*. 2015;116(1): 59 – 65. Available from: doi:10.1016/j.pep.2015.07.011

113. Volontè F, Piubelli L and Pollegioni L. Optimising HIV-1 Protease Production in Escherichia Coli as a Fusion Protein. *Microbial Cell Factories*. 2011;10(53): 1 – 10. Available from: doi:10.1186/1475-2859-10-53

114. Villanueva E, Smith T, Queiroz RML, Monti M, Pizzinga M, Elzek M, Dezi V, Harvey RF, Ramakrishna M, Willis AE and Lilley KS. Efficient recovery of the RNA-bound proteome and protein-bound transcriptome using phase separation (OOPS). *Nature Protocols*. 2020;15(1): 2568 – 2588. Available from: doi: 10.1038/s41596-020-0344-2

115. Queiroz RML, Smith T, Villanueva E, Marti-Solano M, Monti M, Pizzinga M, Mirea DM, Ramakrishna M, Harvey RF, Dezi V, Thomas GH, Willis AE and Lilley KS. Comprehensive identification of RNA–protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology*. 2019;37(1): 169 – 178. Available from: doi: 10.1038/s41587-018-0001-2

116. Pagano JM, Clingman CC and Ryder SP. Quantitative approaches to monitor protein–nucleic acid interactions using fluorescent probes. *RNA*. 2011;17(1): 14 – 20. Available from: doi:10.1261/rna.2428111

117. Arkin MR, Glicksman MA, Fu H, et al. Inhibition of Protein-Protein Interactions: Non-Cellular Assay Formats. In: Markossian S, Grossman A, Brimacombe K, et al., (eds). *Assay Guidance Manual.* Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004.

118. Wang M, Ogé L, Perez-Garcia MD, Hamama L, Sakr S. The PUF protein family: Overview on PUF RNA targets, biological functions, and post transcriptional regulation. *International Journal of Molecular Sciences*. 2018. Available from: doi:10.3390/ijms19020410

119. Costa S, Almeida A, Castro A and Domingues L. Fusion Tags for Protein Solubility, Purification and Immunogenicity in Escherichia Coli: The Novel FH8 System. *Frontiers in Microbiology*. 2014:5(63): 1 – 15. Available from: doi:10.3389/fmicb.2014.00063

120. Harper S and Speicher DW. Purification of Proteins Fused to Glutathione S Transferase. *Methods in Molecular Biology*. 2011;681(1): 259 – 280. Available from: doi:10.1007/978-1-60761-913-0_14

121. Carey MF, Peterson CL and Smale ST. The RNase protection assay. In: Carey MF, Peterson CL and Smale ST. *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*. New York (USA): CSHL Press; 2009.

122. Moerke NJ. Fluorescence Polarisation (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Current Protocols in Chemical Biology*. 2009;1(1): 1 – 15. Available from: doi:10.1002/9780470559277.ch090102

123. Köhn B, Schwarz P, Wittung-Stafshede P and Kovermann M. Impact of crowded environments on binding between protein and single-stranded DNA. *Scientific Reports*. 2021;11(17682): 1 – 10. Available from: doi:10.1038/s41598-021-97219-1

124. Szeltner Z and Polgár L. Conformational Stability and Catalytic Activity of HIV-1 Protease are Both Enhanced at High Salt Concentration. *Journal of Biological Chemistry*. 1996;271(1): 5458 – 5463. Available from: doi:10.1074/jbc.271.10.5458

125. Louis JM, Ishima R, Torchia DA, Weber IT. HIV-1 Protease: Structure, Dynamics, and Inhibition. *Advances in Pharmacology*. 2007. p. 261–298. Available from: doi:10.1016/S1054-3589(07)55008-8

126. Mittal S, Cai Y, Nalam MN, Bolon DNA and Schiffer CA. Hydrophobic Core Flexibility Modulates Enzyme Activity in HIV-1 Protease. *Journal of the American Chemical Society*, 2012;134(9): 4163 – 4168. Available from: doi:10.1021/ja2095766

127. Jaskolski M, Tomasselli AG, Sawyer TK, Staples DG, Heinrikson RL, Schneider J, Kent SB and Wlodawer A. Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry*. 1991;30(6): 1600 – 1609. Available from: doi: 10.1021/bi00220a023

128. Meek TD, Rodriguez EJ and Angeles TS. Use of steady state kinetic methods to elucidate the kinetic and chemical mechanisms of retroviral proteases. *Methods in Enzymology*. 1994;241(1): 127 – 156. Available from: doi:10.1016/0076-6879(94)41063-1

129. Pearl LH. The catalytic mechanism of aspartic proteinases. *FEBS Letter*. 1987;214(1): 8 – 12. Available from: doi:10.1016/0014-5793(87)80003-0

130. Prabu-Jeyabalan M, Nalivaika E and Schiffer CA. Substrate shape determines specificity of recognition for HIV-1 protease. *Structure*. 2002;10(3): 369 – 381. Available from: doi:10.1016/S0969-2126(02)00720-7

131. Rudolph R and Lilie H. In vitro folding of inclusion body proteins. *Federation of American Societies for Experimental Biology*. 1996;10(1): 49 – 56. Available from: doi:10.1096/fasebj.10.1.8566547

132. Burgess RR. Refolding solubilized inclusion body proteins. In: Burgess RR, Deutscher MP (eds.) *Guide to protein purification*. Volume 463. Academic Press; 2009. p. 259 – 282.

133. Esposito D and Chatterjee DK. Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*. 2006;17(4): 353 – 358. Available from: doi:10.1016/j.copbio.2006.06.003

134. Lebendiker M and Danieli T. Production of Prone-to-Aggregate Proteins. *FEBS Letters*. 2013;588(2): 236 – 246. Available from: doi:10.1016/j.febslet.2013.10.044

135. Packman LC, Borges A and Perham RN. Amino acid sequence analysis of the lipoyl and peripheral subunit-binding domains in the lipoate acetyltransferase component of the pyruvate dehydrogenase complex from Bacillus stearothermophilus. *Biochemical Journal*. 1988;252(1): 79 – 86. Available from: doi:10.1042/bj2520079

136. Ishima R, Torchia DA, Lynch SM, Gronenborn AM and Louis JM. Solution Structure of the Mature HIV-1 Protease Monomer. *Journal of Biological Chemistry*. 2003;278(1): 43311 – 43319. Available from: doi:10.1074/jbc.M307549200

137. Hong P, Koza S and Bouvier ESP. Size-Exclusion Chromatography for the Analysis of Protein Biotherapeutics and their Aggregates. *Journal of Liquid Chromatography and Related Technologies*. 2012;35(20): 2923 – 2950. Available from: doi:10.1080/10826076.2012.743724

138. Sherry D, Worth R and Sayed Y. Two-Step Preparation of Highly Pure, Soluble HIV Protease from Inclusion Bodies Recombinantly Expressed in *Escherichia coli*. *Current Protocols in Protein Science.* 2020;100(1): e106. Available from: doi:10.1002/cpps.106

139. Yaron A, Carmel A and Katchalski-Katzir E. Intramolecularly Quenched Fluorogenic Substrates for Hydrolytic Enzymes. *Analytical Biochemistry*. 1979;95(1): 228 – 235. Available from: doi: 10.1016/0003-2697(79)90210-0

140. Nallamsetty S, Kapust RB, Tözsér J, Cherry S, Tropea JE, Copeland TD and Waugh DS. Efficient Site-specific Processing of Fusion Proteins by Tobacco Vein Mottling Virus Protease in vivo and in vitro. *Protein Expression and Purification*. 2004;38(1): 108 – 115. Available from: doi:10.1016/j.pep.2004.08.016

141. Waugh DS. An Overview of Enzymatic Reagents for the Removal of Affinity Tags. *Protein Expression and Purification*. 2011;80(2): 283 – 293. Available from: doi:10.1016/j.pep.2011.08.005

142. Huang L and Chen C. Understanding HIV-1 Protease Autoprocessing for Novel Therapeutic Development. *Future Medicinal Chemistry*. 2013;5(11): 1 – 10. Available from: doi:10.4155/fmc.13.89

143. Lindsten K, Uhlíková T, Konvalinka J, Masucci MG and Dantuma NP. Cell-based Fluorescence Assay for Human Immunodeficiency Virus Type 1 Protease Activity. *Antimicrobial Agents and Chemotherapy*. 2001;45(9): 2616 – 2622. Available from: doi:10.1128/AAC.45.9.2616-2622.2001

144. Wan M, Takagi M, Loh BN, Xu XZ and Imanaka T. Autoprocessing: An Essential Step for the Activation of HIV-1 Protease. *Biochemical Journal*. 1996;316(2): 569 – 573. Available from: doi:10.1042/bj3160569

145. Huang L and Chen C. Autoprocessing of Human Immunodeficiency Virus Type 1 Protease Miniprecursor Fusions in Mammalian Cells. *AIDS Research and Therapy*. 2010;7(27): 1 – 10. Available from: doi:10.1186/1742-6405-7-27

146. Huang L, Li Y and Chen C. Flexible Catalytic Site Conformations Implicated in Modulation of HIV-1 Autoprocessing Reactions. *Retrovirology*. 2011;8(79): 1 – 10. Available from: doi:10.1186/1742-4690-8-79

147. Morgner N, Schmidt C, Beilsten-Edmands V, Ebong IO, Patel NA, Clerico EM, Kirschke E, Daturpalli S, Jackson SE, David Agard et al. Hsp70 Forms Antiparallel Dimers Stabilised by Post-translational Modifications to Position Clients for Transfer to Hsp90. *Cell Reports*. 2015;11(5): 759 – 769. Available from: doi:10.1016/k.celrep.2015.03.063

148. Blanco R, Carrasco L and Ventoso I. Cell Killing by HIV-1 Protease. *Journal of Biological Chemistry*. 2003;278(1): 1086 – 1093. Available from: doi:10.1074/jbc.M205636200

149. Yang H, Nkeze J and Zhao RY. Effects of HIV-1 protease on cellular functions and their potential applications in antiretroviral therapy. *Cell and Bioscience*. 2012;2(32): 1 – 3. Available from: doi:10.1186/2045-3701-2-32

150. Chen X, Zaro J and Shen W. Fusion Protein Linkers: Property, Design and Functionality. *Advanced Drug Delivery Reviews*. 2013;65(10): 1357 – 1369. Available from: doi:10.1016/j.addr.2012.09.039

151. Li G, Huang Z, Zhang C, Dong B, Guo R, Yue H, Yan L and Xing X. Construction of a linker library with widely controllable flexibility for fusion protein design. *Applied Microbiology and Biotechnology*. 2016;100(1): 215 – 225. Available from: doi:10.1007/s00253-015-6985-3

152. Fu Q, Li Y, Zheng Z, Liu A, Yuan Z, Peng J, He J. Optimization of triple plasmids transfection into HEK293 cells mediated by polyethylenimine. *Journal of Biomedical Engineering*. 2015;32(1): 137 – 141. Available from: PMID: 25997281

153. Grieger JC, Soltys SM and Samulski RJ. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. *Molecular Therapy*. 2016;24(2): 287 – 297. Available from: doi:10.1038/mt.2015.187

154. Merten OW, Hebben, M and Bovolenta C. Production of lentiviral vectors. *Molecular Therapy – Methods and Clinical Development*. 2016:3(1):16017. Available from: doi:10.1038/mtm.2016.17

155. Kim TK and Eberwine JH. Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry*. 2010;397(8): 3173 – 3178. Available from: doi:10.1007/s00216-010-3821-6

156. Tien C, Huang L, Watanabe SM, Speidel JT, Carter CA and Chen C. Context-dependent autoprocessing of human immunodeficiency virus type 1 protease precursors. *PLoS ONE* . 2018;13(1): e0191372. Available from: doi:10.1371/journal.pone.0191372

157. Beninati S and Mukherjee AB. A novel transglutaminase-catalyzed posttranslational modification of HIV-1 aspartyl protease. *Biochemical and Biophysical Research Communications*. 1992;187(3): 1211 – 1218. Available from: doi: 10.1016/0006-291x(92)90432-k

158. Bhatwa A, Wang W, Hassan YI, Abraham N, Li XZ and Zhou T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in *Escherichia coli* and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*. 2021;9(1): 630551. Available from: doi:10.3389/fbioe.2021.630551

159. Rufer AC, Thiebach L, Baer, K, Klein HW and Hennig M. X-ray structure of glutathione *S*-transferase from *Schistosoma japonicum* in a new crystal form reveals flexibility of the substrate-binding site. *Acta Crystallographica Section F Structural Biology and Crystallization Communication*. 2005;61(3): 263-265. Available from: doi:10.1107/S1744309105004823

160. Baer K, Al-Hasani H, Parvaresch S, Corona T, Rufer A, Nölle V, Bergschneider E and Klein HW. Dimerization-induced activation of soluble insulin/IGF-1 receptor kinases: an alternative mechanism of activation. *Biochemistry*. 2001;40(47):14268-14278. doi:10.1021/bi015588g

161. Leitner A, Dorn G and Allain FHT. Combining Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) Spectroscopy for Integrative Structural Biology of Protein–RNA Complexes. *Cold Spring Harbour Perspectives in Biology*. 2019;11(1): a032359. Available from: doi:10.1101/cshperspect.a032359

162. Ahmed M, Marchanka A and Carlomagno T. Structure of a protein-RNA complex by solid-state NMR spectroscopy. *Angewandte Chemie International Edition*. 2020;59(17): 6866 – 6873. Available from: doi:10.1002/anie.201915465

163. Gell DA, Kwan AH and Mackay JP. NMR spectroscopy in the analysis of protein-protein interactions, *Modern Magnetic Resonance*. 2017;1(1): 1 – 34. Available from: doi: 10.1007/978-3-319-28275-6_121-1

164. Hafner M, Katsantoni M, Köster T, Marks J, Mukherjee J, Staiger D, Ule J and Zavolan M. CLIP and complementary methods. *Nature Reviews Methods Primers*. 2021;1(20): 1 – 10. Available from: doi:10.1038/s43586-021-00018-1

165. Wang S, Zhang F, Mei M, Wang T, Yun Y, Yang S, Zhang G and Yi L. A split protease-*E. coli* ClpXP system quantifies protein–protein interactions in *Escherichia coli* cells. *Nature Communication Biology*. 2021;4(841): 1 – 7. Available from: doi: 10.1038/s42003-021-02374-w

166. Wehr MC, Laage R, Bolz U, Fischer TM, Grünewald S, Scheek S, Bach A, Nave KA and Rossner MJ. Monitoring regulated protein-protein interactions using split TEV. *Nature Methods*. 2006;3(1): 985 – 993. Available from: doi:10.1038/nmeth967

167. Hu CD, Chinenov Y and Kerppola TK. Visualization of Interactions among bZIP and Rel Family Proteins in Living Cells Using Bimolecular Fluorescence Complementation. *Molecular Cell*. 2002;9(4): 789 – 798. Available from: doi:10.1016/S1097-2765(02)00496-3

168. Ozawa T, Natori Y, sato M and Umezawa Y. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nature Methods*. 2007;4(1): 413 – 419. Available from: doi: 10.1038/nmeth1030

169. Yamada T, Yoshimura H, Inaguma A and Ozawa T. Visualization of nonengineered single mRNAs in living cells using genetically encoded fluorescent probes. *Analytical Chemistry*. 2011;83(14): 5708 – 5714. Available from: doi: 10.1021/ac2009405

170. Tilsner J, Linnik O, Christensen NM, Bell K, Roberts IM, Lacomme C and Oparka KJ. Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *The Plant Journal*. 2009;57(4): 758 – 770. Available from: doi: 10.1111/j.1365-313X.2008.03720.x

171. Huppertz I, Attig J, D'Ambrogio A, Easton LE, Sibley CR, Sugimoto Y, Tajnik M, König J and Ule J. iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods*. 2014;65(3): 274 – 287. Available from: doi: 10.1016/j.ymeth.2013.10.011

172. Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M and Ule J. Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biology*. 2012;13(8): R67. Available from: doi: 10.1186/gb-2012-13-8-r67

173. Nostrand ELV, Pratt GA, Shishkin AA, Gelboin-Burkhard C, Fang MY, Sunadaraman B, Blue SM, Bguyen TB, Surka C, Elkins K, Stanton R, Rigo F, Guttman M and Yeo GW. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nature Methods*. 2016;13(1): 508 – 514. Available from: doi: 10.1038/nmeth.3810

174. Nostrand ELV, Shishkin A, Pratt GA, Nguyen TB and Yeo GW. Variation in single-nucleotide sensitivity of eCLIP derived from reverse transcription conditions. *Methods*. 2017;126(1): 29 – 37. Available from: doi: 10.1016/j.ymeth.2017.08.002

175. Goodwin S, McPherson JD and McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. 2016;17(1): 333 – 351. Available from: doi: 10.1038/nrg.2016.49

176. Shimomura O, Johnson FH and Saiga Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*. 1962;59(3): 223 – 239. Available from: doi:10.1002/jcp.1030590302

177. Chalfie M, Tu Y, Euskirchen G, Ward WW and Prasher DC. Green Fluorescent Protein as a Marker for Gene Expression. *Science*. 1994;263(5148): 802 – 805. Available from: doi: 10.1126/science.8303295

178. Gerdes HH and Kaether C. Green fluorescent protein applications in cell biology. *FEBS Letters*. 1996;389(1): 44 – 47. Available from: doi: 10.1016/0014-5793(96)00586-8

179. Zimmer M. Green Fluorescent Protein (GFP): Applications, Structure, and Related Photophysical Behavior. *Chemical Reviews*. 2002;102(3): 759 – 782. Available from: doi: 10.1021/cr010142r

180. Ward WW and Bokman SH. Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein. *Biochemistry*. 1982;21(19): 4535 – 4540. Available from: doi: 10.1021/bi00262a003

181. Ghosh I, Hamilton AD and Regan L. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *Journal of the Americal Chemical Society*. 2000;122(23): 5658 – 5659. Available from: doi:10.1021/ja994421w

182. Romei MG and Boxer SG. Split Green Fluorescent Proteins: Scope, Limitations, and Outlook. *Annual Review in Biophysics*. 2019;48(1): 19 – 44. Available from: doi: 10.1146/annurev-biophys-051013-022846

183. Foglieni C, Papin S, Salvadè A, Afroz T, Pinton S, Pedrioli G, Ulrich G, Polymenidou M and Paganetti P. Split GFP technologies to structurally characterize and quantify functional biomolecular interactions of FTD-related proteins. *Scientific Reports*. 2017;7(14013): 1 – 10. Available from: doi: 10.1038/s41598-017-14459-w

184. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M and Ragg T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*. 2006;7(3): 1 – 10. Available from: doi: 10.1186/1471-2199-7-3

185. Su R, Fan LH, Cao C, Wang L, Du Z, Cai Z, Ouyang YC, Wang Y, Zhou Q, Wu L *et al*. Global profiling of RNA-binding protein target sites by LACE-seq. *Nature Cell Biology*. 2021;23(1): 664 – 675. Available from: doi: 10.1038/s41556-021-00696-9

186. Ye X and Jankowsky E. High Throughput Approaches to Study RNA-Protein Interactions In Vitro. *Methods*. 2021;178(1): 3 – 10. Available from: doi: 10.1016/j.ymeth.2019.09.006

187. Guenther UP, Yandek LE, Niland CN, Campbell FE, Anderson D, Anderson VE, Harris ME and Jankowsky E. Hidden specificity in an apparently nonspecific RNA-binding protein. *Nature Letters*. 2013;502(1): 385 – 388. Available from: doi: 10.1038/nature12543

188. Jankowsky E and Harris ME. Mapping specificity landscapes of RNA-protein interactions by high throughput sequencing. *Methods*. 2017;118(119): 111 – 118. Available from: doi: 10.1016/j.ymeth.2017.03.002

189. Solovyov A, Vabret N, Arora KS, Snyder A, Funt SA, Bajorin DF, Rosenberg JE, Bhardwaj N, Ting DT and Greenbaum BD. Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. *Cell Reports*. 2018;23(1): 512 – 521. Available from: doi: 10.1016/j.celrep.2018.03.042

190. Landers CC, Rabeler CA, Ferrari EK, D'Alessandro LR, Kang DD, Malisa J, Bashir SM and Carone DM. Ectopic expression of pericentric HSATII RNA results in nuclear RNA accumulation, MeCP2 recruitment, and cell division defects. *Chromosoma*. 2021;130(1): 75 – 90. Available from: doi: 10.1007/s00412-021-00753-0

191. Price RL and Jerome WG (eds). *Basics of Confocal Microscopy*. New York (USA). Springer Science. 2011. Available from: doi: 10.1007/978-0-387-78175-4 [Accessed 8th April 2022]

192. Wintgens, JP, Rossner MJ, and Wehr MC. Characterizing Dynamic Protein-Protein Interactions Using the Genetically Encoded Split Biosensor Assay Technique Split TEV. *Methods in molecular biology.* 1596(1): 219–238. Available from: doi: 10.1007/978-1-4939-6940-1_14

193. Alonso-Gardón M and Estévez R. Split-Tobacco Etch Virus (Split-TEV) Method in G Protein-Coupled Receptor Interacting Proteins. In: Sofia Aires M. Martins and Duarte Miguel F. Prazeres (eds.). *G Protein-Coupled Receptor Screening Assays: Methods and Protocols, Methods in Molecular Biology*. Volume 2268. Switzerland. Springer Nature. 2021.

194. Wehr MC, Reinecke L, Botvinnik A and Rossner MJ. Analysis of transient phosphorylation-dependent protein-protein interactions in living mammalian cells using split-TEV. *BMC Biotechnology*. 2008;8(55): 1 – 10. Available from: doi: 10.1186/1472-6750-8-55

195. Fan X, Li X, Zhou Y, Mei M, Liu P, Zhao J, Peng W, Jiang ZB, Yang S, Iverson BL Zhang G and Yi L. Quantitative Analysis of the Substrate Specificity of Human Rhinovirus 3C Protease and Exploration of Its Substrate Recognition Mechanisms. *ACS Chemical Biology*. 2020;15(1): 63 – 73. Available from: doi: 10.1021/acschembio.9b00539

196. Kapust RB, Tözsér J, Copeland TD and Waugh DS. The P1' specificity of tobacco etch virus protease. *Biochemical and Biophysical Research Communications*. 2002;294(5): 949 – 955. Available from: doi:10.1016/S0006-291X(02)00574-0

197. Dolberg TB, Meger AT, Boucher JD, Corcoran WK, Schauer EE, Prybutok AN, Raman S and Leonard JN. Computation-guided optimization of split protein systems. *Nature Chemical Biology*. 2021;17(1): 531 – 539. Available from: doi: 10.1038/s41589-020-00729-8

198. Nunn CM, Jeeves M, Cliff MJ, Urquhart GT, George RR, Chao LH, Tscuchia Y and Djordjevic S. Crystal Structure of Tobacco Etch Virus Protease Shows the Protein C Terminus Bound within the Active Site. *Journal of Molecular Biology*. 2005;250(1): 145 – 155. Available from: doi: 10.1016/j.jmb.2005.04.013

199. Raran-Kurussi S, Cherry S, Zhang D and Waugh DS. Removal of Affinity Tags with TEV Protease. *Methods in Molecular Biology*. 2017;1586(1): 221 – 230. Available from: doi: 10.1007/978-1-4939-6887-9_14

200. Nam H, Hwang BJ, Choi DY, Shin S and Choi M. Tobacco etch virus (TEV) protease with multiple mutations to improve solubility and reduce self-cleavage exhibits enhanced enzymatic activity. *FEBS Open Biology*. 2020;10(4): 619 – 626. Available from: doi: 10.1002/2211-5463.12828

201. Wei L, Cai X, Qi Z, Rong L, Cheng B and Fan J. In vivo and in vitro characterization of TEV protease mutants. *Protein Expression and Purification*. 2012;83(2): 157 – 163. Available from: doi:10.1016/j.pep.2012.03.011

202. Rorrer K, Parks TD, Scheffler B, Bevan M and Dougherty WG. Autocatalytic activity of the tobacco etch virus NIa proteinase in viral and foreign protein sequences. *Journal of General Virology*. 1992;73(4): 775 – 783. Available from: doi: 10.1099/0022-1317-73-4-775

203. Burmakina S, Geng Y, Chen Y and Fan QR. Heterodimeric coiled-coil interactions of human GABAB receptor. *PNAS*. 2014;111(19): 6958 – 6963. Available from: doi: 10.1073/pnas.1400081111

204. Wintgens JP, Wichert SP, Popovic L, Rossner MJ and Wehr MC. Monitoring activities of receptor tyrosine kinases using a universal adapter in genetically encoded split TEV assays. *Cellular and Molecular Life Sciences*. 2019;76(1): 1185 – 1199. Available from: doi: 10.1007/s00018-018-03003-2

205. Proteintech. *mCherry polyclonal antibody*. Available from: ptglab.com/products/mCherry-Antibody-26765-1-AP.htm [Accessed April 8th, 2022]

206. Fabrini R, De Luca A, Stella L, Mei G, Orioni B, Ciccone S, Federici G, Lo Bello M and Ricci G. Monomer-dimer equilibrium in glutathione transferases: a critical re-examination. *Biochemistry*. 2009;48(43): 10473 – 10482. Available from: doi: 10.1021/bi901238t

207. Maru Y, Afar DE, Witte ON and Shibuya M. The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. *The Journal of Biological Chemistry*. 1996;271(26): 15353 – 15357. Available from: doi: 10.1074/jbc.271.26.15353

208. Yadav DK and Lukavsky PJ. NMR solution structure determination of large RNA-protein complexes. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2016;97(1): 57 – 81. Available from: doi:10.1016/j.pnmrs.2016.10.001

209. Carlomagno T. Present and future of NMR for RNA–protein complexes: A perspective of integrated structural biology. *Journal of Magnetic Resonance*. 2014;241(1): 126 – 136. Available from: doi:10.1016/j.jmr.2013.10.007

210. Phan J, Zdanov A, Evdokimov AG, Tropea JE, Howard P, Kapust RB, Li M, Wlodawer A and Waugh DS. Structural Basis for the Substrate Specificity of Tobacco Etch Virus Protease. *Protein Structure and Folding*. 2002;277(52): 50564 – 50572. Available from: doi:10.1074/jbc.M207224200

211. Buttle DJ and Mort JS. Cysteine proteases. *Encyclopedia of Biological Chemistry.* 2013;1(1): 589 – 592. Available from: doi: 10.1016/B978-0-12-378630-2.00009-8

212. Salisbury CM, Maly DJ and Ellman JA. Peptide Microarrays for the Determination of Protease Substrate Specificity. *Journal of the Americal Chemical Society*. 2002;124(50): 14868 – 14870. Available from: doi: 10.1021/ja027477q

213. Plyasova AA and Zhdanov DD. Alternative Splicing of Human Telomerase Reverse Transcriptase (hTERT) and Its Implications in Physiological and Pathological Processes. *Biomedicines*. 2021;9(1): 526 – 536. Available from: doi: 10.3390/biomedicines9050526

214. Xi L and Cech TR. Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic Acids Research*. 2014;42(13): 8565 – 8577. Available from: doi: 10.1093/nar/gku560

215. Jan R and Chaudhry GES. Understanding Apoptosis and Apoptotic Pathways Targeted Cancer Therapeutics. *Advanced Pharmaceutical Bulletin*. 2019;9(2): 205 – 218. Available from: doi: 10.15171/apb.2019.024

216. Walsh JG, Cullen SP, Sheridan C and Martin SJ. Executioner caspase-3 and caspase-7 are functionally distinct proteases. *PNAS Biochemistry*. 2008;105(35): 12815 – 12819. Available from: doi: 10.1073/pnas.0707715105

217. Slee EA, Adrain C and Martin SJ. Executioner Caspase-3, -6, and -7 Perform Distinct, Non-redundant Roles during the Demolition Phase of Apoptosis. *Journal of Biological Chemistry*. 2001;276(10): 7320 – 7326. Available from: doi: 10.1074/jbc.M008363200

218. Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P. *Molecular biology of the cell (4th ed): a problems approach*. New York: Garland Science. 2002

219. Fulda S and Debatin KM. Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. *Oncogene*. 2006;25(1): 4798 – 4811. Available from: doi: 10.1038/sj.onc.1209608

220. Ponder KG and Boise LH. The prodomain of caspase-3 regulates its own removal and caspase activation. *Cell Death Discovery*. 2019;5(56): 1 – 10. Available from: doi: 10.1038/s41420-019-0142-1

221. Tenev T, Marani M, McNeish I and Lemoine NR. Pro-caspase-3 overexpression sensitises ovarian cancer cells to proteasome inhibitors. *Cell Death and Differentiation*. 2001;8(1): 256 – 264. Available from: doi: 10.1038/sj.cdd.4400808

222. Strebel A, Harr T, Bachmann F, Wernli M and Erb P. Green fluorescent protein as a novel tool to measure apoptosis and necrosis. *Cytometry*. 2001;43(2): 126 – 133. Available from: doi: 10.1002/1097-0320(20010201)43:2<126::aid-cyto1027>3.0.co;2-j

223. Yoon S, Park SJ, Han JH, Kang JH, Kin JH, Lee J, Park S, Shin HJ, Kim K, Yun M and Chwae YJ. Caspase-dependent cell death-associated release of nucleosome and damage-associated molecular patterns. *Cell Death and Disease*. 2014;5(1): e1494. Available from: doi: 10.1038/cddis.2014.450

224. Suzuki Y, Nakabayashi Y and Takahashi R. Ubiquitin-protein ligase activity of X-linked inhibitor of apoptosis protein promotes proteasomal degradation of caspase-3 and enhances its anti-apoptotic effect in Fas-induced cell death. *PNAS*. 2001;98(15): 8662 – 8667. Available from: doi: 10.1073/pnas.161506698

225. Choi YE, Butterworth M, Malladi S, Duckett CS, Cohen GM and Bratton SB. The E3 Ubiquitin Ligase cIAP1 Binds and Ubiquitinates Caspase-3 and -7 via Unique Mechanisms at Distinct Steps in Their Processing. *Journal of Biological Chemistry*. 2009;284(19): 12772 – 12782. Available from: doi: 10.1074/jbc.M807550200

226. Morgan CW, Diaz JE, Zeitlin SG, Gray DC and Wells JA. Engineered cellular gene-replacement platform for selective and inducible proteolytic profiling. *PNAS*. 2015;112(27): 8344 – 8349. Available from: doi: 10.1073/pnas.1504141112

227. Wang Y, Cheetham AG, Angacian G, Su H, Xie L and Cui H. Peptide–drug conjugates as effective prodrug strategies for targeted delivery. *Advanced Drug Delivery Reviews*. 2017;110-111(1): 112 – 126. Available from: doi: 10.1016/j.addr.2016.06.015

228. Greasley R, Khabazhaitajer M and Rosario DJ. A profile of enzalutamide for the treatment of advanced castration resistant prostate cancer. *Cancer Management and Research*. 2015;7(1): 153 – 164. Available from: doi: 10.2147/CMAR.S50585

229. Ryan CJ, Smith MR, de Bono JS, Molina A, Logothetis CJ, de Souza P, Fizazi K, Mainwaring P, Piulats JM, NG Siobhan *et al*. Randomized Phase 3 Trial of Abiraterone

Acetate in Men with Metastatic Castration-Resistant Prostate Cancer and No Prior Chemotherapy. *New England Journal of Medicine*. 2014;368(2): 138 – 148. Available from: doi: 10.1056/NEJMoa1209096

230. Apte MS and Cooper JP. Life and cancer without telomerase: ALT and other strategies for making sure ends (don't) meet. *Critical Reviews in Biochemistry and Molecular Biology*. 2017;52(1): 57 – 73. Available from: doi: 10.1080/10409238.2016.1260090

231. Zondagh J, Williams A, Achilonu I, Dirr HW and Sayed Y. Overexpression, Purification and Functional Characterisation of Wild-Type HIV-1 Subtype C Protease and Two Variants Using a Thioredoxin and His-Tag Protein Fusion System. *The Protein Journal*. 2018;37(1): 369 – 379. Available from: doi: 10.1007/s10930-018-9779-5

232. Eche S and Gordon ML. Recombinant expression of HIV-1 protease using soluble fusion tags in Escherichia coli: A vital tool for functional characterization of HIV-1 protease. *Virus Research*. 2021;295(1): 198289. Available from: doi: 10.1016/j.virusres.2020.198289

233. Schermelleh L, Ferrand A, Huser T, Eggeling C, Sauer M, Biehlmaier O and Drummen GPC. Super-resolution microscopy demystified. *Nature Cell Biology*. 2019;21(1): 72 – 84. Available from: doi: 10.1038/s41556-018-0251-8

234. Williams DJ, Puhl HL and Ikeda SR. Rapid Modification of Proteins Using a Rapamycin-Inducible Tobacco Etch Virus Protease System. *PLoS One*. 2009;4(10): e7474. Available from: doi: 10.1371/journal.pone.0007474

235. Rhyu MS. Telomeres, Telomerase, and Immortality. *Journal of the National Cancer Institute*. 1995;87(12): 884 – 894. Available from: doi: 10.1093/jnci/87.12.884

236. Camacho-Soto K, Castillo-Montoya J, Tye B, Ogunleye LO and Ghosh I. Small Molecule Gated Split-Tyrosine Phosphatases and Orthogonal Split-Tyrosine Kinases. *Journal of the American Chemical Society*. 2014;136(49): 17078 – 17086. Available from: doi: 10.1021/ja5080745

237. Chakrabandhu K and Hueber AO. Fas Versatile Signaling and Beyond: Pivotal Role of Tyrosine Phosphorylation in Context-Dependent Signaling and Diseases. *Frontiers in Immunology*. 2016;7(429): 1 – 10. Available from: doi: 10.3389/fimmu.2016.00429

238. Kratz F, Müller IA, Ryppa C and Warnecke A. Prodrug Strategies in Anticancer Chemotherapy. *ChemMedChem*. 2008;3(1): 20 – 53. Available from: doi: 10.1002/cmdc.200700159

239. Chelur DS and Chalfie M. Targeted cell killing by reconstituted caspases. *PNAS*. 2007;104(7): 2283 – 2288. Available from: doi: 10.1073/pnas.0610877104

240. Waldman AD, Fritz JM and Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nature Reviews Immunology*. 2020;20(1): 651 – 668. Available from: doi: 10.1038/s41577-020-0306-5

241. Bajrachaya R, Song JG, Back SY and Han H. Recent Advancements in Non-Invasive Formulations for Protein Drug Delivery. *Computational and Structural Biotechnology Journal*. 2019;17(1): 1290 – 1308. Available from: doi: 10.1016/j.csbj.2019.09.004

242. Sung YK and Kim SW. Recent advances in the development of gene delivery systems. *Biomaterials Research*. 2019;23(8): 1 – 10. Available from: doi: 10.1186/s40824-019-0156-z

243. Nachtergaele S and He C. The emerging biology of RNA post-transcriptional modifications. *RNA Biology*. 2017;14(2): 156 – 163. Available from: doi: 10.1080/15476286.2016.1267096