Site-specific Incorporation of p-azido-L-phenylalanine for Photo-crosslinking Nucleic Acids

Gabriel Sullivan

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

A thesis submitted to
the University of Ottawa
in partial fulfillment of the requirements of the
Degree Master of Science Chemistry

Abstract

Current methods for studying RNA binding proteins (RBPs) combine the use of ultraviolet (UV) crosslinking and immunoprecipitation (CLIP) to analyze RNA-protein interactions. An underexplored alternative approach is using site-specific incorporation of photoactivatable non-canonical amino acids (ncAAs) to enhance the crosslinking efficiency of many CLIP protocols. This thesis describes the incorporation of the photo-crosslinking unnatural amino acid p-azido-L-phenylalanine (AzF) into the Hepatitis C Virus (HCV) non-structural protein 3 helicase (NS3h) for photo-crosslinking and *in vitro* analysis of the potential binding sites found within the HCV RNA genome. From the five potential sites identified from the NS3h crystal structure for AzF incorporation, two sites, E503AzF and Q580AzF, allowed for nucleic acid photo-crosslinking with fluorescently labelled DNA substrates. We further tested if these mutations adversely affected NS3h and binding activity through a molecular beacon helicase assay and fluorescence polarization methods. We found that E503AzF unexpectedly had a faster unwinding rate than wild type (WT) NS3h and managed to have a similar binding affinity to the tested DNA substrate. Finally, we found that there was a 5-fold increase in the photo-crosslinking efficiency of nucleic acids for E503AzF NS3h mutant compared to our WT NS3h at 254 nm UV light. We are currently working on methods for our CLIP-based protocol to ensure quality RNA footprint generation and purification from photo-crosslinked NS3h.

Other work contained in this thesis consists of using *Prevotella sp. P5-125* Cas13b (PspCas13b), a clustered regularly interspaced short palindromic repeats (CRISPR) RNA-targeting system, which has been previously shown to knockdown viral RNA and mRNA through designable guide CRISPR RNA (crRNA). Here we incorporated the photo-crosslinking ncAA AzF into PspCas13b to irreversibly bind the crRNA in an attempt to enhance knockdown efficiency

and longevity of viral and mRNA targets. We were able to design a crRNA that produced significant knockdown targeting the luciferase mRNA of a luciferase rennilla reporter system. When targeting an HCV subgenomic replicon luciferase reporter system, knockdown was not observed. Additionally, the WT PspCas13b had photo-crosslinking to the bound crRNA and requires further optimization for future use.

Statement of Work

All of the work presented in this thesis is my own and I take full responsibility for its contents. All experiments and analyses were performed by me. I had help purifying NS3h and all the AzF mutant proteins in Chapter 2 were from Noreen Ahmed and Christine Hum.

Acknowledgements

I would like to sincerely thank Dr. John Pezacki for all his insights, constructive guidance, and positive encouragement throughout my graduate degree. Your passion for science and discovery will stay with me throughout my future career. I would like to thank those who have trained me throughout my degree, Dr. Dave Prescott, Noreen Ahmed, and David Lefebvre. Your patience, enthusiasm and knowledge is what shaped me into the scientist that I am today.

I would like to thank Dr. Corrie daCosta, Dr. François-Xavier Campbell-Valois and Dr. John Pezacki for taking part in my Thesis Advisory Committee. You have all given me invaluable perspectives into my projects and propelled me to the completion of this thesis. I am extremely grateful to Dr. Roberto Chica and all those involved with organizing the Advanced Protein Engineering, Training, Internships, Courses, and Exhibition (APRENTICE) program for providing funding and amazing learning opportunities to further expand my knowledge base. I have met so many great people and friends through this program.

Thank you to Kaitlyn Morrill for always pushing my science in the right direction and keeping things piquant. Thank you to David Lefebvre for being an amazing lab mentor, Dungeon Master and friend. Your determination and joyful attitude through hardships truly speaks to the strength of your character. I would like to thank Christine Hum for all your invaluable insights relating to my projects. Even with a million things to do you would always have time to help. Your kindness, resilience and work ethic is truly something to behold. There really is no one else I'd rather manually syringe His-tag purify 10 different AzF mutants with. I would like to thank Salmaan Chummun for the numerous Tim Hortons coffees which kept me afloat and the countless shared laughs. Your humour and thoughtfulness know no bounds. Thank you to Jordan Pham for your all-around help in the lab and the great golf games and tennis matches. To my fellow lab

mates, Mariam Serhan, Tiffany Stern, Parrish Evers, Eryn Lundrigan, Spencer Uguccioni, Étienne Bélanger and the many others past and present, I truly wish to thank you for all the help you provided me during these years and for the many fun discussions.

I would like to thank Sandrine Legault for always taking the time to listen, even when sinking in Red Fluorescent Proteins. I would like to thank Louise Connell and Christian Tessier for being the best stop to and from the autoclave. Thank you for always taking the time to listen to my struggles and triumphs, and for providing me with new directions and experimental ideas - you always re-energized my days, like taking a breath of fresh air.

To my family, thank you for taking the time to listen to all the presentations I've done throughout my degree. It may have been a brief 30 minutes of your time, but your interest meant so much to me. Thank you for your encouragement and support. To Shelby Hagerman, you have been my backbone and lifeline throughout this experience. I don't think I would have been able to make it this far without your reassurance and love. Thank you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
4SU	4-thiouridine
6SG	6-thioguanine
AzF	p-azido-L-phenylalanine
bp	base pair
BzCas13b	*Bergeylla zoohelcum* Cas13b
pBpa	p-benzoyl-L-phenylalanine
Cas	CRISPR associated protein
cDNA	Complementary DNA
CRISPR	clustered regularly interspaced short palindromic repeats
Cy3	cyanine 3
Cy5	cyanine 5
CLIP	crosslinking and immunoprecipitation
DAA	direct-acting antiviral
DBCO	Dibenzocyclooctyne
dsDNA	double-stranded DNA
DR	direct repeat
ER	endoplasmic reticulum
HCV	hepatitis C virus
IRES	internal ribosome entry site
IVT	in vitro transcription
Kb	Kilobase
LVP	lipo-viro-particles
mRNA	messenger RNA
ncAA	noncanonical amino acid
NT	non-target
NTP	nucleoside triphosphate
NS3h	non-structural protein 3 helicase
nt	Nucleotide
PbuCas13b	*Prevotella buccea* Cas13b
PspCas13b	*Prevotella sp. P5-125* Cas13b
RBP	RNA binding protein
ssDNA	single-stranded DNA
TfmdPhe	4’-[3-(trifluoromethyl)-3H-diazirin-3-yl]-l-phenylalanine
tRNA	transfer RNA
AARS	aminoacyl tRNA synthetase
UV	ultraviolet
UTR	untranslated region
WT	wild type

Chapter 1: General Introduction

1.1 Hepatitis C Virus

In 2005, it was estimated that over 184 million people worldwide have chronic Hepatitis C Virus (HCV) infection [1]. HCV is a positive-sense single-stranded, enveloped RNA virus of the family Flaviviridae which has 7 major genotypes and nearly 100 different subtypes. Only 25-30% of patients who are acutely infected with HCV will show any sign of infection; however, the 80% majority of acutely infected individuals will develop chronic HCV infections [2,3]. Chronic HCV infections cause progressive hepatic fibrosis which can lead to cirrhosis, liver failure, and hepatocellular carcinoma which account for the majority of HCV mortality, resulting in 500,000 deaths per year globally [1,4]. The high level of genetic heterogeneity throughout HCV genotypes and subtypes complicates the development of a permanent treatment as well as the possibility of developing an effective vaccine [5–7]. Fortunately, there is hope for individuals with chronic HCV infections, as oral treatments of direct-acting antiviral (DAA) drugs which target essential HCV non-structural proteins involved in replication machinery are highly effective in curing HCV infections [3].

1.1.1 HCV Lifecycle

HCV virions exist associated with hepatocyte-derived lipoproteins known as lipo-viro-particles (LVP) which play a role in antibody evasion and aid in cellular entry [8–10]. The initiation of HCV LVP entry begins when apolipoproteins make contact with host cell low-density lipoprotein receptors and heparan sulfate proteoglycans [11,12]. This association leads to the HCV envelope glycoprotein E1/E2 heterodimer binding to host receptors, namely cluster of differentiation 81 (CD81) and scavenger receptor B1 (SCARB1) [13]. Interactions between CD81 and claudin-1 allow for clathrin-mediated endocytosis which transports the virus into the cell. Through interactions with CD81 and the acidic environment of the endosome, viral E1/E2 proteins

undergo conformational changes that allow for the fusion of the viral lipid envelope and the endosomal membrane, releasing the positive-sense viral genome into the cytosol [14,15].

Once in the cytosol, the 9.6kb RNA, single open reading frame genome will be translated at the endoplasmic reticulum (ER) into a 3000 amino acid polyprotein whereby host cell and viral proteases cleave the polyprotein into three structural proteins (Core nucleocapsid protein (C), E1 and E2) and seven non-structural proteins (p7 ion channel, NS2, NS3, NS4A, NS4B, NS5A, and NS5B). The expression of NS4B and NS5A allows for the formation of the membranous web, changing host ER membranes into single and double-membrane vesicles which compartmentalize viral replication from the surrounding cytosol. This allows it to go unnoticed by pattern recognition receptors avoiding activation of the innate immune system [16,17]. Inside these vesicles, the non-structural proteins NS3 to NS5B form the replication complex and begin viral genome replication [18]. The RNA-dependant RNA polymerase (RdRp), NS5B, amplifies both negative strand templates and positive strand genomes. Following replication by the RdRp, the duplexed RNA is unwound by the helicase domain of NS3. Nucleocapsid assembly and viral RNA packing are orchestrated through NS2 protein interactions with E1/E2, p7, NS3 and NS5A as well as C protein interactions with NS5A [19–21]. Hijacking the host's very low-density lipoprotein (VLDL) secretory pathway, the nascent virions fuse with apolipoprotein containing pre-VLDL particles to form LVPs which will be non-lytically released from the host cell [22].

1.1.2 Functional Secondary Structures of the HCV Genomic RNA

The HCV genome has extensive RNA secondary and tertiary structures essential for many steps in the HCV life cycle. The full HCV genome has up to twenty highly structured elements, many with roles in viral translation, replication, and infectivity [23,24]. The 5′ untranslated region (UTR) of the genomic RNA harbours the internal ribosome entry site (IRES) which mediates cap-

independent translation through direct binding of the 40S ribosomal subunit and initiation factor eIF3 [25–29]. The IRES also has two conserved sites for host microRNA-122 binding which stabilizes the IRES to adopt a functional fold improving replication and translation [30,31]. The sequestering of microRNA-122 also plays a key role in manipulating host lipid metabolism by increasing lipid synthesis and decreasing lipid export [32]. A series of short stem-loop motifs found across the HCV genome bind to the core protein and offer a significant effect on RNA encapsulation [33]. The 3′ UTR is essential for negative template synthesis as well as forming long-range interactions with the 5′ UTR, circularizing the genome to allow for efficient replication [34,35]. The HCV genome RNA structure is so finely tuned that it limits the length of double-stranded RNA helices to 7 bp or less to avoid host pattern recognition receptors in the cell [36].

1.2 Non-Structural Protein 3 (NS3)

NS3 is a 67 kDa multi-domain protein with an N-terminal serine protease domain and C-terminal helicase domain connected by a flexible linker (**Figure 1.1A**). The serine protease domain is 180 amino acids in length and belongs to the trypsin/chymotrypsin protease superfamily. It is essential for the cleavage of the genomic polypeptide at NS3/4A, NS4A/4B, NS4B/5A, and NS5A/5B junctions as well as the cutting of immunogenic proteins [37–40]. The helicase domain hydrolyses nucleoside triphosphates (NTPs) to fuel unwinding in a 3′-5′ direction. This helicase activity unwinds double-stranded RNA viral intermediates formed after duplication, displaces RNA binding proteins bound along the viral genome and is thought to assist in genomic RNA structure conversions [41–44]. Interactions between NS4A and NS3 have been shown to position the serine protease active site for enhanced peptide cleavage as well as stimulate the RNA unwinding activity [45,46]. Both NS3 domains can be expressed separately *in vitro* with their enzymatic activity

retained; however, both domains in full length NS3 have been shown to influence each other's activities [43,47].

1.2.1 NS3 Helicase

The NS3 helicase domain, NS3h, is part of the DEXH/D-box subgroup of helicases and belongs to the larger helicase superfamily 2 [41,43]. All superfamily 2 helicases have seven highly conserved motifs important for their ATPase functionality [48]. The NS3h has 3 subdomains and can be viewed in a "Y" formation (**Figure 1.1B**). Subdomains 1 and 2 are located at the top of the "Y" and contain all the conserved superfamily 2 helicase motifs, specifically motifs I, II, IV, V, and VI, which adopt similar RecA protein folds essential for ATP binding and hydrolysis [49]. ATP binding and hydrolysis occur in the cleft between both subdomains 1 and 2. NS3h has been shown to hydrolyze other NTPs with different efficiencies with its decreasing preference as follows: ATP > CTP > UTP/dTTP > GTP [42]. The specificity of NTP hydrolysis is determined through hydrogen bond interactions between NS3h and the Watson-Crick base paring area of the NTP [42]. The NS3h ATPase functionality is coordinated through a magnesium ion which reorients the nucleotide in the active site for faster hydrolysis [50].

NS3h harbours a nucleic acid binding cleft separating subdomains 1 and 2 from subdomain 3. The cleft mainly interacts with the phosphate backbone of the nucleic acids and has a binding site size of 7-8 nucleotides (nt) [51]. NS3h binds DNA and RNA at a low nanomolar K_d and is non-sequence specific, though NS3h does bind RNA with different affinities where poly(U) > > poly(A) > poly(G), poly(C) [51,52]. There are two functionally important residues in the cleft, W501 which base stacks against nucleic acid bases, and V432 positioned between bases on the 5′ end of bound nucleic acids which act as "bookends" preventing bound nucleotide strands from slipping

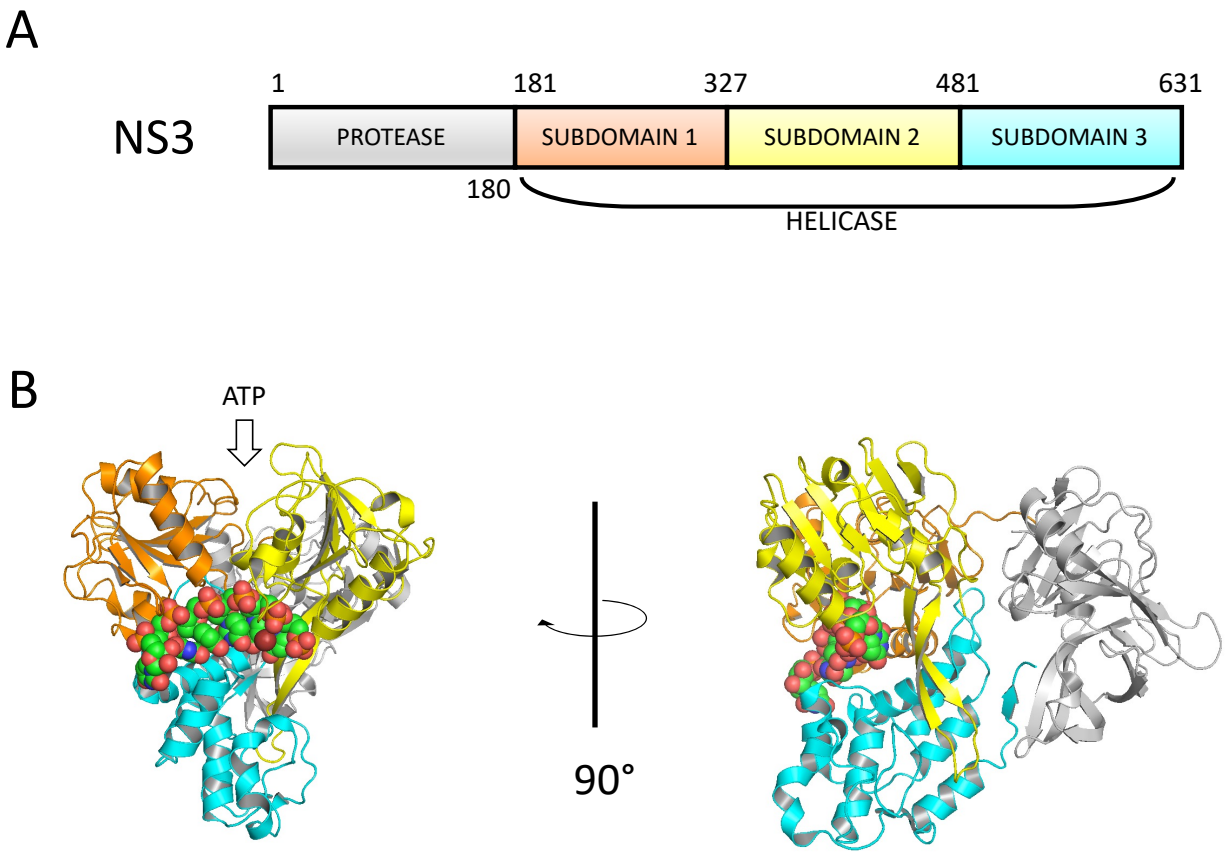


Figure 1.1 The Sequence and Structural Organization of NS3: (A) NS3 contains an N-terminal serine protease domain and a C-terminal helicase domain. (B) Structure of NS3 showing the architecture of the domains (PBD:3O8C). The protease domain can be seen in grey and is responsible for HCV polyprotein processing. Helicase subdomains 1, 2 and 3 are coloured orange, yellow, and blue, respectively. The bound RNA (green) can be seen in the binding cleft and is represented as spheres. ATP binding and hydrolysis occur between subdomains 1 and 2 allowing for translocation along bound DNA/RNA in a 3′-5′ direction. The helicase is responsible for unwinding RNA secondary structures.

back in the wrong direction. A mutation of either of these bookends results in a loss of DNA/RNA affinity and unwinding [53]. NS3h does not form higher order oligomers with itself, though multiple NS3h molecules can bind to the same strand increasing unwinding [51].

The helicase unwinding mechanism is extensively supported by the ratcheting inch-worm model [49,54–57]. First, NS3h binds nucleic acids in its open state where the bookends W501 and V432 anchor the strand with 5nt in between them (**Figure 1.2A**). When ATP is bound, there is a shift to the closed state where subdomains 1 and 3 shift towards subdomain 2 (**Figure 1.2B**) causing 1nt to slide outside the binding site no longer stacking with W501, leaving only 4nt between the bookend residues. V432 holds the 5′ end of the single-stranded DNA/RNA (ssDNA/ssRNA) in place at this stage. After ATP hydrolysis, W501 holds the 3′ end of the ssDNA/RNA while domain 2 moves forward into an open state allowing bookend V432 to slide down 1nt in the 3′ to 5′ direction. This movement is what directly exerts a force on the DNA junction, disrupting the hydrogen bonds and unwinding duplexed DNA/RNA. A schematic for NS3h unwinding through the inch-worm model can be seen in **Figure 1.2C**.

1.3 Methods to Study RNA Binding Protein Interactions

The study of RBP interactions is essential to gain a full understanding of cellular homeostasis as RBPs influence RNA processing and functions in the cytoplasm and nucleus [58]. There are two different approaches for studying RBP interactions. The first is an RNA-centric method where an RNA segment of interest is used to study the proteins bound to it. These methods often involve the use of tagged RNAs of interest which act as bait for RBPs. These RBP-RNA complexes are then purified and quantitative mass spectrometry is used to discover the identity of bound RBPs [59]. There are a vast array of RNA-centric methods which can be performed either *in vitro* or *in vivo* [60,61]. The second method for studying RBP interactions is a protein-centric method

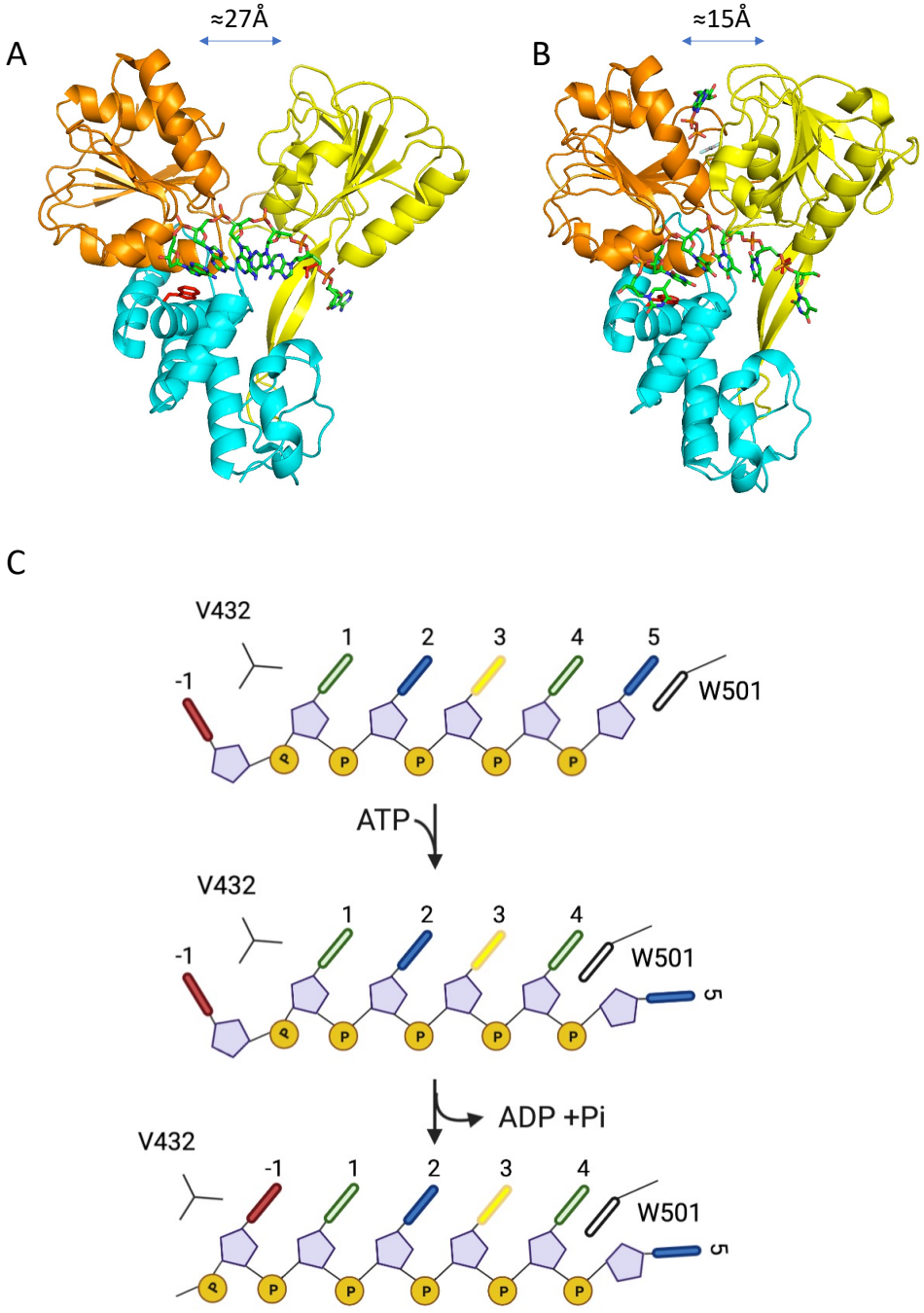


Figure 1.2 Inch-worm Model of NS3 Helicase Unwinding: (A) Structure of NS3h bound to ssDNA (PDB: 3KQH). Subdomains 1, 2 and 3 are coloured in orange, yellow, and blue, respectively. The stick structures of bound RNA are shown in green and bookends W501 and V432 are shown in red. (B) Movement of domains when bound to RNA and ATP mimic ADP-AlF^{-}_{4} (PBD: 3KQL). (C) Bookends W501 and V432 are essential for helicase activity. Individual DNA bases are shown in red, yellow, blue, or green. Adapted from Gu et al., 2010 [49].

where the RBP RNA interactions want to be discovered. These methods rely on UV irradiation which forms stable RNA-protein complexes and subsequent identification of RBP binding sites [62]. Methods that rely on UV crosslinking and purification of the RNA-protein complex of interest through immunoprecipitation are called crosslinking and immunoprecipitation CLIP methods [59]. Methods that use a similar protocol, but different purification techniques or crosslinkers are referred to as CLIP variants [63].

1.3.1 Fundamental Steps in the CLIP Methodologies

All CLIP-based methods and variants follow similar core steps in their protocols (**Figure 1.3**) [58,64–66]. First, cells are initially irradiated with UV light to crosslink the RBP of interest to the RNA. The cells are then lysed and a partial RNase digestion is performed on the cell lysate leaving a 5′-OH and 3′-phosphate on the bound RNA. The 3′ end of the RNA is then dephosphorylated using an alkaline phosphatase and an RNA sequencing adapter is ligated on. Next, the RBP-RNA complex of interest is purified using immunoprecipitation or affinity chromatography. The purified RBP-RNA complex is then electrophoresed on an SDS-PAGE gel and transferred onto a nitrocellulose membrane to separate any free RNAs to reduce background. The RBP-RNA complexes are excised from the nitrocellulose membrane and then the protein is fully digested using Proteinase K, freeing the bound RNA. Complementary DNA (cDNA) is constructed through reverse transcription of the isolated RNA fragments. The cDNA is purified and size selected before the final 5′ sequencing adapter is ligated onto the cDNA. This is amplified and sent for high-throughput sequencing to identify the RBP of interest binding locations.

1.4 The Expanded Genetic Code

With the exception of selenocysteine and pyrrolysine, all proteins are synthesized using only 20 amino acids and 64 unique codons [67]. Therefore, the ability to site-specifically incorporate

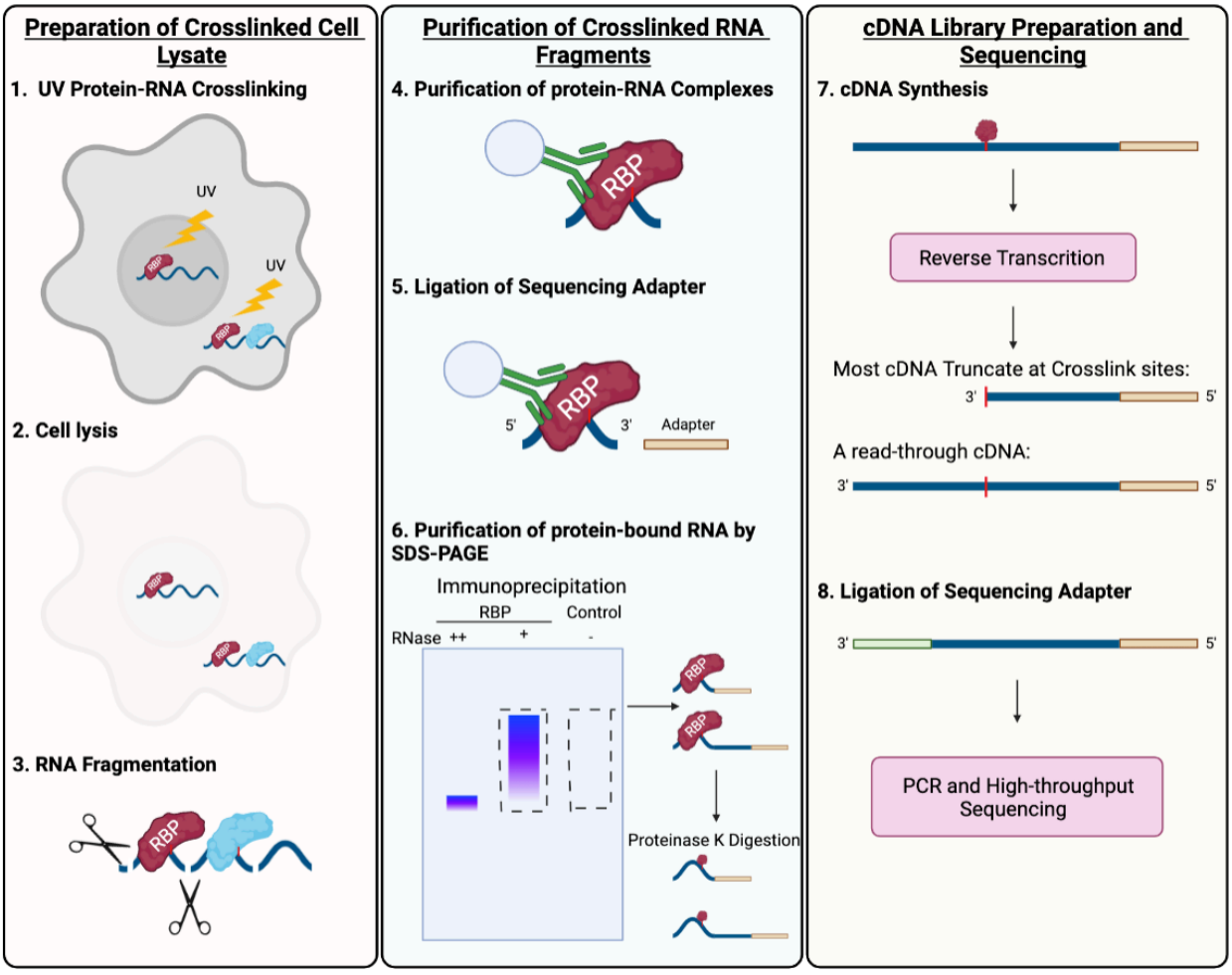


Figure 1.3 Common Workflow Found in CLIP Protocols: An overview of the core steps found in most CLIP variants to discover RBP binding interactions and locations. Adapted from Hafner et al., 2021.

new complex structures and functionalities in proteins provides unique tools for enzyme engineers to probe structure and function [68–70]. Over 200 non-canonical amino acids (ncAA) have been introduced into proteins to date with many distinctive reactive groups providing modifications such as optical probes, post-translational modification mimics, bioorthogonal chemistry probes, and photoactivatable groups [71,72]. Genetic code expansion techniques require the use of an orthogonal tRNA/aminoacyl tRNA synthetase pair (tRNA/AARS) to site-specifically incorporate ncAAs into the protein of interest. A two-step directed evolution selection scheme, using both positive and negative selection rounds, is used to design a tRNA/AARS which specifically aminoacylates their respective ncAA to the cognate tRNA [73]. The orthogonal tRNA is designed to recognize a stop codon as it is not recognized by any endogenous tRNA. The amber stop mutation, UAG, is most commonly used as it is found in only 9% of the *E.coli* genome and rarely terminates at essential genes [73]. The ncAA must not be recognized by any endogenous tRNA synthases and be metabolically stable in the cellular environment for efficient incorporation [73].

A general method for incorporating ncAAs into proteins is outlined in **Figure 1.4**. The gene encoding the protein of interest is mutated to the TAG amber stop mutation at the residue which is desired to be replaced by the ncAA. The orthogonal tRNA/AARS will produce mature tRNA conjugated with the ncAA. During translation, the amber stop codon will be recognized by the orthogonal tRNA and the ncAA will be incorporated into the polypeptide chain. Failure to incorporate the ncAA will lead to a prematurely terminated protein.

1.4.1 Photo-crosslinking ncAAs

There are three main families of photo-crosslinking ncAAs: aryl azides, diazirines, and aryl ketones. The most common aryl ketone ncAA is p-benzoyl-L-phenylalanine (pBpa) [74]. pBpa,

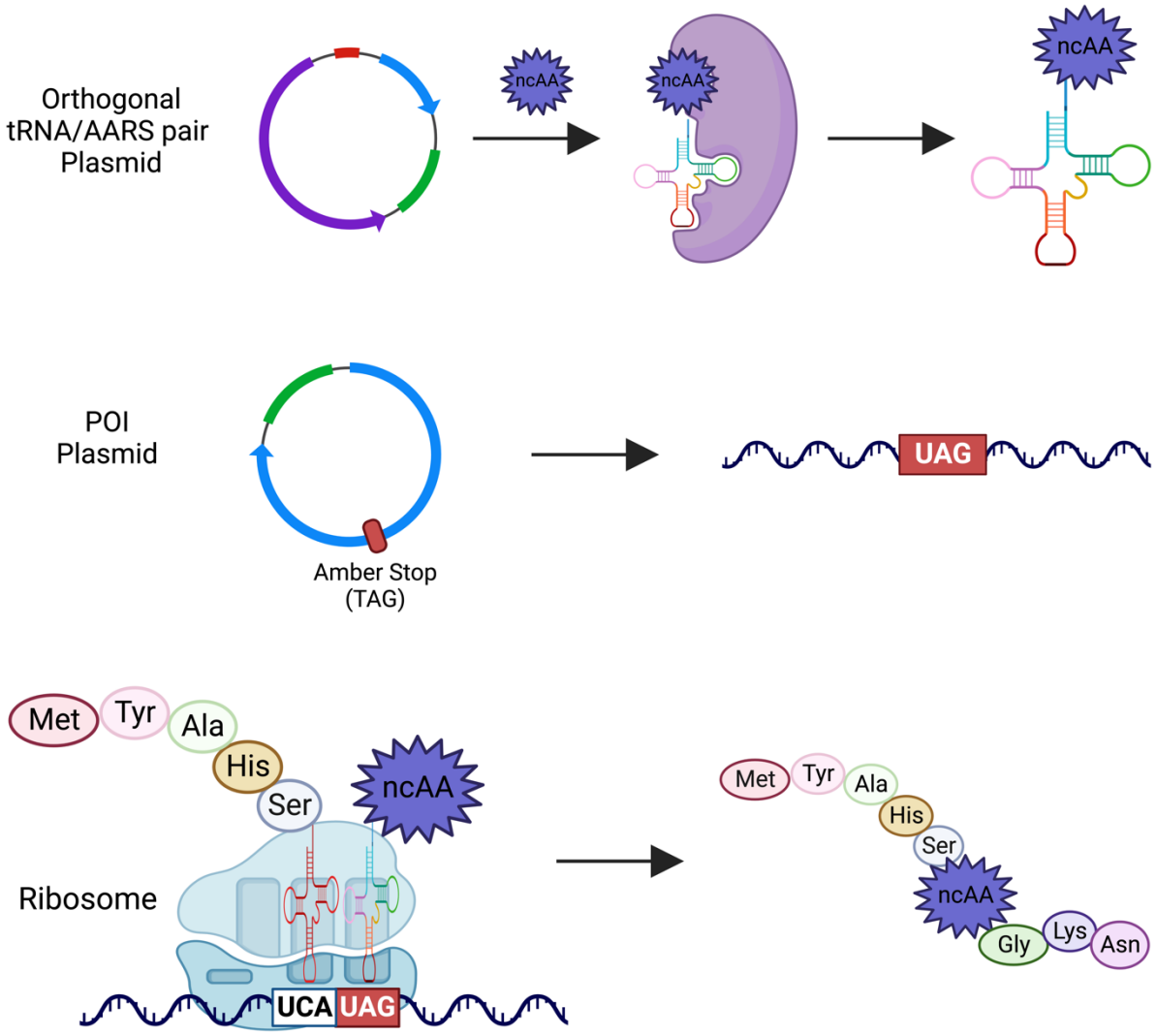


Figure 1.4 General Scheme for Incorporating ncAAs: The gene encoding the protein of interest is mutated to the TAG amber stop mutation at the residue which is desired to be replaced by the ncAA. The orthogonal tRNA/aminoacyl tRNA synthetase will produce mature tRNA linked with ncAA. During translation, the amber stop codon will be recognized by the orthogonal tRNA and the ncAA will be incorporated into the growing polypeptide chain.

the first photo-crosslinking ncAA was created in 1986 and was incorporated into synthetic peptides for photo labelling [75]. In 2002, with the development of amber stop codon repression and orthogonal tRNA/AARS pair, pBpa was now able to be site-specifically incorporated into full length proteins for mapping protein-protein interactions [76]. Since then, the incorporation of pBpa *in vivo* has led to the discovery of interactions between transcriptional activators and coactivators as well as interactions between chaperones and protein folding targets [77–79]. pBpa incorporated into *E.coli* catabolite activator protein has also shown photo-crosslinking capabilities towards nucleic acids.[80] When pBpa is activated through irradiation with 350-365 nm UV light it causes the formation of a diradical allowing strong reactivity towards C-H bonds (**Figure 1.5A**) [81].

Aryl Azides are another family of photo-crosslinking ncAAs and the most common of these is AzF which is a tyrosine analogue (**Figure 1.5B**). AzF was first incorporated into the genetic code of *E.coli* in 2002 by using the amber stop codon suppression and evolved tRNA/AARS pairs [82]. AzF has been incorporated into the human serotonin transporter to locate drug-binding sites through photo-crosslinking [83]. AzF has also been successfully incorporated into the tombusviruses viral suppressor protein p19 to photo-crosslink pre-miRNA and has been postulated to function as a therapeutic [84]. Irradiation with UV light under 310 nm causes the loss of N_2 and the formation of a highly reactive singlet nitrene which can directly react with C-H and N-H hydrogen [82,84]. This nitrene undergoes a ring expansion to form a less reactive didehydroazepine intermediate that has been shown to react readily with amines but not hydro-carbons (**Figure 1.5B**) [85]. Fluorination of the ring has been shown to slow the rate of ring expansion while keeping the longevity of the highly reactive nitrene [86].

AzF can be used for non-UV chemistries as well. The azide group can undergo strain-promoted alkyne-azide cycloaddition (SPAAC) with cyclooctynes [87]. This is a very important

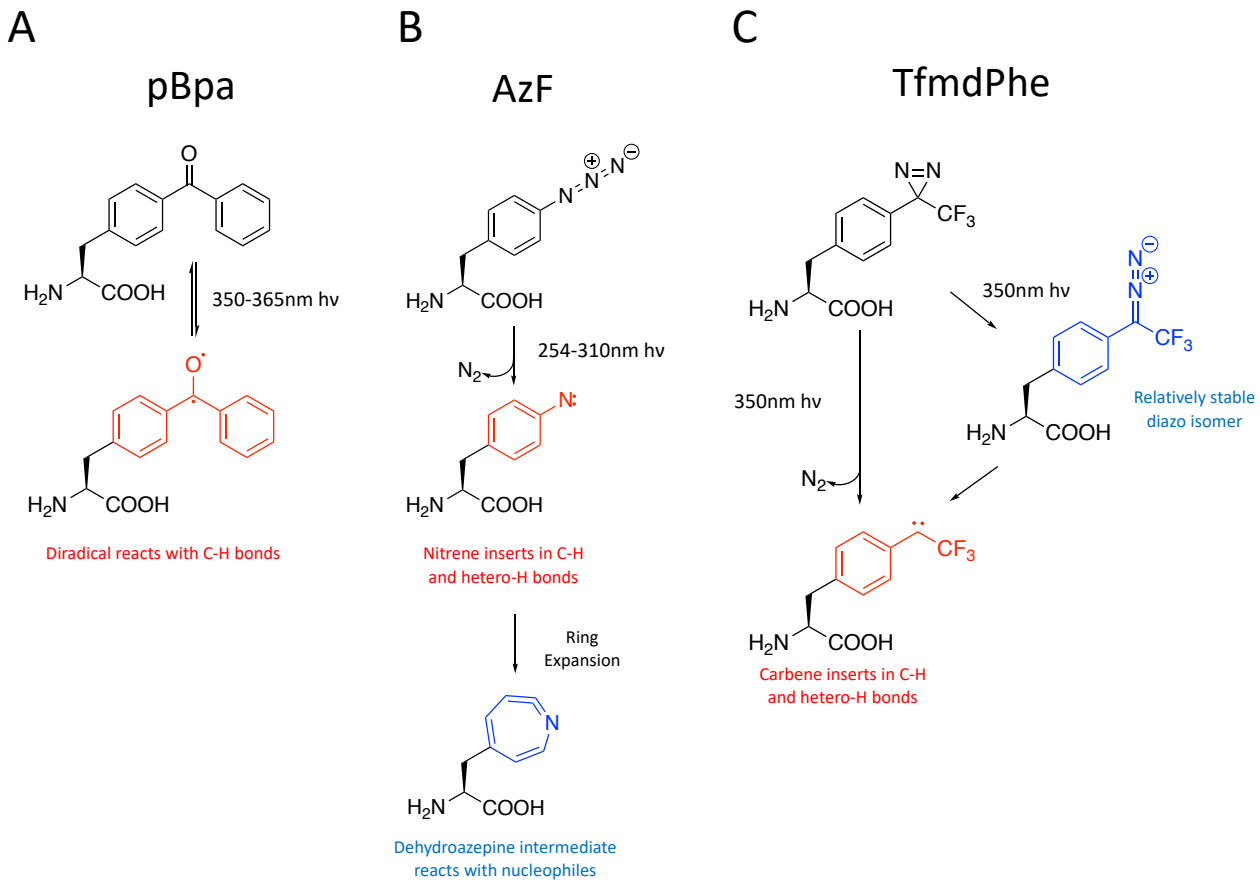


Figure 1.5 Reactivity of Three Common Photo-crosslinking ncAAs: Very reactive states are shown in red and low reactive states are shown in blue. (A) Upon UV irradiation, pBpa will reversibly generate a triplet ketyl biradical that will directly react with C-H bonds. (B) AzF under UV forms a singlet nitrene which quickly undergoes ring expansion to a dehydroazepine intermediate which will react with nucleophiles. (C) TfmdPhe when irradiated with UV generates a carbene and a stable diazo isomer.

bioorthogonal reaction that can label proteins *in vivo* with fluorophores, affinity reagents, peptides and more to allow for a multitude of downstream applications.

The next family of photo-crosslinking ncAAs is Diazirines. A commonly used diazirine for photo-crosslinking is 4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]-l-phenylalanine (TfmdPhe) (**Figure 1.5C**). TfmdPhe was first incorporated into the genetic code of *E.coli* in 2007 by using the amber stop codon suppression and evolved tRNA/AARS pairs [88]. Since then, diazirines have been widely used to trap the noncovalent interactions of proteins, lipids and cholesterol and many other biomolecules [89]. Diazirines have been introduced into human iron regulatory protein 1 and showed a 7-fold increase in photo-crosslinking efficiency compared to conventional protein-nucleic acid photo-crosslinking [90]. TfmdPhe and other diazirines are activated with 350 nm UV light which causes the loss of N_2 gas and carbene formation. Carbenes are highly reactive and will insert into C-H or O-H bonds (**Figure 1.5C**) [74]. TfmdPhe under UV light can undergo a rearrangement leading to the formation of a stable diazo isomer which can lead to reduced photo-crosslinking efficiency [91].

1.5 Clustered Regularly Interspaced Short Palindromic Repeats Systems

CRISPR and CRISPR-associated (Cas) proteins function as adaptive immune systems for bacteria and archaea which target invading foreign genetic elements from bacteriophages (**Figure 1.6**) [92–94]. The CRISPR system can compile and store sequence information from previous bacteriophage infections as spacers in the CRISPR array. Upon subsequent infections, the CRISPR array is transcribed into pre-crRNA which are processed into mature crRNA by endogenous RNases or Cas effector proteins with spacers and direct repeats (DR) important for Cas effector recognition [95]. The Cas effector bound to mature crRNA is active and searches for the complementary sequence to the spacer, hybridizing with the foreign genetic material and cleaving

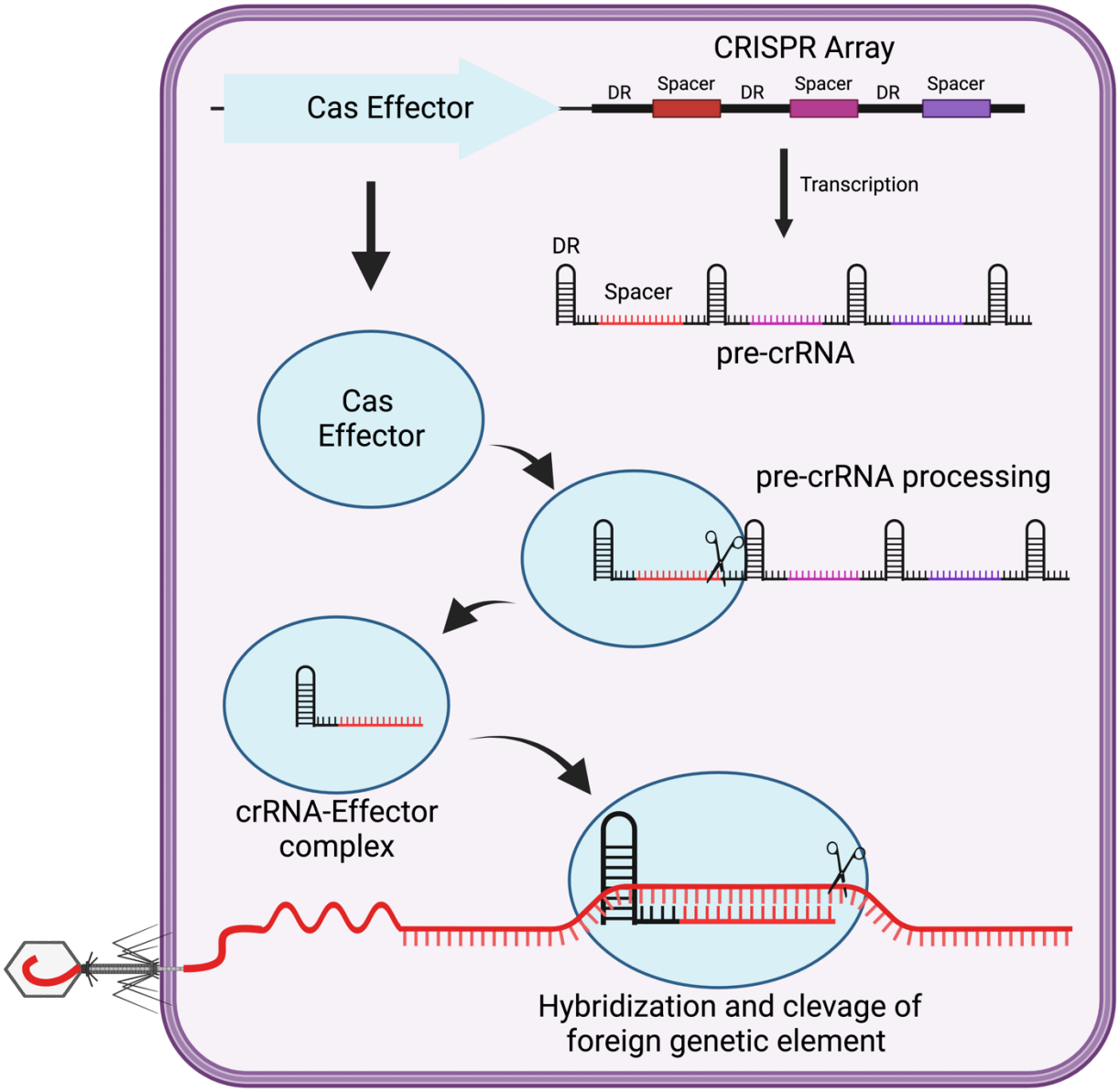


Figure 1.6 Targeting of Foreign Genetic Elements with Cas Effectors: During infection, the CRISPR array containing spacers sequences complementary to invading genetic materials will be translated as pre-crRNA. Cas effector proteins will be produced. The pre-CRNA is processed and Cas effectors bind and recognize the DR of the crRNA and will form an activated complex. The crRNA-Cas effector complex can search the cytosol for the invader. When the spacer hybridizes with the foreign genetic material it will cleave.

it [96,97]. CRISPR systems are characterized by their effector complexes such that Class 1 CRISPR systems have multiple Cas effectors which bind a CRISPR crRNA to coordinate nuclease activity while Class 2 systems use a single effector which uses crRNA as guides [98].

The most commonly known CRISPR Class 2 system is CRIPSR-Cas9. When bound to a crRNA and a transactivating crRNA, Cas9 can perform double-stranded breaks in DNA. Through pioneering work, the CRISPR-Cas9 system has been simplified to a single programmable guide RNA and Cas9 effector for DNA interference and targeting [99]. Due to the high degree of flexibility, this system is being used in many biomedical research applications [100–103]. Cas-9 has targeting restraints as a protospacer-adjacent motif is required for efficient cleavage of target DNA [99].

1.5.1 CRISPR-Cas13b

CRISPR Cas13 belongs in the CRISPR Class 2 Type VI systems. These systems are single enzyme effectors which use a crRNA guide to target and cleave single-stranded RNA. Type VI, initially named C2c2 systems, were first discovered through sequence database mining which found a new effector with two predicted higher eukaryotes and prokaryotes nucleotide-binding (HEPN) RNase domains [104]. This led to the discovery of a Cas13a effector which could be designed to bind and cleave RNA thus showing great potential for the development of RNA targeting tools [105]. Since this initial discovery, there have now been four different Cas13 subgroups (a-d) which all have two independent conserved HEPN domains – one involved in pre-crRNA processing and the other for catalyzing ssRNA cleavage. The Cas13b subgroup structures are different from the others harbouring a linear domain architecture with the HEPN domains at the N-and C- termini as well as the crRNA DRs located at the 3′ end [106,107]. Cas13b crRNAs have a DR size of 36nt with a poly-U stretch in the open loop region and conserved complementary sequences 5′-GUUG and CAAC-3′. Spacer lengths are 30nt, though larger and smaller spacers can

be designed with functional activity [106,108]. Cas13b has also been shown to have RNA targeting requirements as a double-sided protospacer flanking sequence which is required for efficient RNA knockdown.

1.6 Objectives

The first goal of the work in this thesis was to use the ncAA AzF to target nucleic acids to determine if it was applicable and could enhance photo-crosslinking in CLIP applications. We first identified five sites for AzF incorporation in NS3h to photo-crosslink the bound DNA. We then tested the photo-crosslinking efficiency of our AzF mutants against the conventional 254 nm UV irradiation used in CLIP. Next, we wanted to test if the AzF NS3h mutants had altered activities. We tested the unwinding activity of our photo-crosslinking mutants using a molecular beacon helicase assay and binding affinity using fluorescence polarization. Next, using our AzF NS3h photo-crosslinking mutants, we wished to determine the NS3h binding sites located in the extensive HCV RNA genome. To do this, we prepared and are currently working on a CLIP-based method to identify these NS3h binding locations.

The second section of this thesis focuses on using AzF to photo-crosslink crRNA bound to CRISPR Cas13b effectors to determine if the irreversible covalent bond would increase the longevity of the crRNA mediated mRNA and viral RNA knockdown. We first identified six sites for AzF incorporation in PspCas13b. We then tested all PspCas13b AzF mutants for their photo-crosslinking capabilities to bound crRNA. We also designed crRNAs targeting a luciferase gene in a luciferase renilla reporter plasmid and crRNAs targeting an HCV subgenomic replicon to test for efficient RNA knockdown.

1.7 References

1. Thrift AP, El-Serag HB, Kanwal F. Global epidemiology and burden of HCV infection and HCV-related disease. *Nat Rev Gastroenterol Hepatol*. 2017;14(2):122-132. doi:10.1038/nrgastro.2016.176
2. Bukh J. The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. *J Hepatol*. 2016;65(1):S2-S21. doi:10.1016/j.jhep.2016.07.035
3. Millman AJ, Nelson NP, Vellozzi C. Hepatitis C: Review of the Epidemiology, Clinical Care, and Continued Challenges in the Direct-Acting Antiviral Era. *Curr Epidemiol Reports*. Published online 2017. doi:10.1007/s40471-017-0108-x
4. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2095-2128. doi:10.1016/S0140-6736(12)61728-0
5. Forns X, Bukh J, Purcell RH. The challenge of developing a vaccine against hepatitis C virus. *J Hepatol*. 2002;37(5):684-695. doi:10.1016/S0168-8278(02)00308-2
6. Mikkelsen M, Bukh J. Current Status of a Hepatitis C Vaccine: Encouraging Results but Significant Challenges Ahead. *Curr Infect Dis Rep*. 2007;9(94-101).
7. Liang TJ. Current progress in development of hepatitis C virus vaccines. *Nat Med*. 2013;19(7):869-878. doi:10.1038/nm.3183
8. Piver E, Boyer A, Gaillard J, et al. Ultrastructural organisation of HCV from the bloodstream of infected patients revealed by electron microscopy after specific immunocapture. *Gut*. doi:10.1136/gutjnl-2016-311726
9. Wrensch F, Crouchet E, Ligat G, et al. Hepatitis C virus (HCV)-apolipoprotein interactions and immune evasion and their impact on HCV vaccine design. *Front Immunol*. 2018;9(JUN):1-9. doi:10.3389/fimmu.2018.01436
10. Andre P, Deforges S, Perret M, et al. Characterization of Low- and Very-Low-Density Hepatitis C Virus RNA-Containing Particles. *J Virol*. 2002;76(14):6919-6928. doi:10.1128/JVI.76.14.6919
11. Owen DM, Huang H, Ye J, Gale M. Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor. *Virology*. 2009;394(1):99-108. doi:10.1016/j.virol.2009.08.037
12. Xu Y, Martinez P, Séron K, et al. Characterization of Hepatitis C Virus Interaction with Heparan Sulfate Proteoglycans. *J Virol*. 2015;89(7):3846-3858. doi:10.1128/jvi.03647-14
13. Foung S, Dubuisson J, Pierce BG, et al. Role of Hepatitis C virus envelope Glycoprotein e1 in virus entry and Assembly. *Front Immunol*. 2018;9. doi:10.3389/fimmu.2018.01411
14. Koutsoudakis G, Kaul A, Steinmann E, et al. Characterization of the Early Steps of Hepatitis C Virus Infection by Using Luciferase Reporter Viruses. *J Virol*. 2006;80(11):5308-5320. doi:10.1128/jvi.02460-05
15. Sharma NR, Mateu G, Dreux M, Grakoui A, Cosset FL, Melikyan GB. Hepatitis C virus is primed by CD81 protein for low pH-dependent fusion. *J Biol Chem*. 2011;286(35):30361-30376. doi:10.1074/jbc.M111.263350
16. Neufeldt CJ, Joyce MA, Van Buuren N, et al. The Hepatitis C Virus-Induced Membranous Web and Associated Nuclear Transport Machinery Limit Access of Pattern Recognition Receptors to Viral Replication Sites. *PLoS Pathog*. 2016;12(2):1-28.

doi:10.1371/journal.ppat.1005428
17. Egger D, Wölk B, Gosert R, et al. Expression of Hepatitis C Virus Proteins Induces Distinct Membrane Alterations Including a Candidate Viral Replication Complex. *J Virol*. 2002;76(12):5974-5984. doi:10.1128/jvi.76.12.5974-5984.2002
18. Li H-C, Yang C-H, Lo S-Y, Ferrero DS. Hepatitis C Viral Replication Complex. *Viruses*. 2021;13:580. doi:10.3390/v13030520
19. Bartenschlager R, Penin F, Lohmann V, André P. Assembly of infectious hepatitis C virus particles. *Trends Microbiol*. 2010;19:95-103. doi:10.1016/j.tim.2010.11.005
20. Jirasko V, Montserret R, Appel N, et al. Structural and functional characterization of nonstructural protein 2 for its role in hepatitis C virus assembly. *J Biol Chem*. 2008;283(42):28546-28562. doi:10.1074/jbc.M803981200
21. Appel N, Zayas M, Miller S, et al. Essential role of domain III of nonstructural protein 5A for hepatitis C virus infectious particle assembly. *PLoS Pathog*. 2008;4(3). doi:10.1371/journal.ppat.1000035
22. Troels Scheel and Charles Rice. Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. *Mol Cell Biochem*. 2013;19(7):837-849. doi:10.1038/nm.3248.
23. Wan H, Adams RL, Lindenbach BD, Pyle AM. The In Vivo and In Vitro Architecture of the Hepatitis C Virus RNA Genome Uncovers Functional RNA Secondary and Tertiary Structures . *J Virol*. 2022;96(8). doi:10.1128/jvi.01946-21
24. Romero-López C, Berzal-Herranz A. The 5BSL3.2 functional RNA domain connects distant regions in the hepatitis C virus genome. *Front Microbiol*. 2017;8(OCT):1-16. doi:10.3389/fmicb.2017.02093
25. Pérard J, Leyrat C, Baudin F, Drouet E, Jamin M. Structure of the full-length HCV IRES in solution. *Nat Commun*. 2013;4. doi:10.1038/ncomms2611
26. Lytle JR, Wu L, Robertson HD. Domains on the hepatitis C virus internal ribosome entry site for 40s subunit binding. *Rna*. 2002;8(8):1045-1055. doi:10.1017/S1355838202029965
27. Quade N, Boehringer D, Leibundgut M, Van Den Heuvel J, Ban N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nat Commun*. 2015;6(May):1-9. doi:10.1038/ncomms8646
28. Méndez-Mancilla A, Wessels HH, Legut M, et al. Chemically modified guide RNAs enhance CRISPR-Cas13 knockdown in human cells. *Cell Chem Biol*. 2022;29(2):321-327.e4. doi:10.1016/j.chembiol.2021.07.011
29. Buratti E, Tisminetzky S, Zotti M, Baralle FE. *Functional Analysis of the Interaction between HCV 5'UTR and Putative Subunits of Eukaryotic Translation Initiation Factor EIF3*. Vol 26.; 1998.
30. Jopling CL, Schütz S, Sarnow P. Position-Dependent Function for a Tandem MicroRNA miR-122-Binding Site Located in the Hepatitis C Virus RNA Genome. *Cell Host Microbe*. 2008;4(1):77-85. doi:10.1016/j.chom.2008.05.013
31. Schult P, Roth H, Adams RL, et al. MicroRNA-122 amplifies hepatitis C virus translation by shaping the structure of the internal ribosomal entry site. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-05053-3
32. Wen J, Friedman JR. miR-122 regulates hepatic lipid metabolism and tumor suppression. *J Clin Invest*. 2012;122(8):2773-2776. doi:10.1172/JCI63966
33. Stewart H, Bingham RJ, White SJ, et al. Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome

packaging. *Sci Rep*. 2016;6(January):1-12. doi:10.1038/srep22952
34. Fricke M, Dünnes N, Zayas M, Bartenschlager R, Niepmann M, Marz M. Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. *Rna*. 2015;21(7):1219-1232. doi:10.1261/rna.049338.114
35. Shi ST, Lai MMC. HCV 5′ and 3′UTR: When Translation Meets Replication. *Hepat C Viruses Genomes Mol Biol*. Published online 2006:49-87. http://www.ncbi.nlm.nih.gov/pubmed/21250387
36. Rebecca L Adamsa, Nathan Pirakitikulrb and AMP. Functional RNA structures throughout the Hepatitis C Virus genome. *Rev del Col Am Cardiol*. 2018;72(23):2964-2979. doi:10.1016/j.coviro.2017.04.007.Functional
37. Bartenschlager R, Ahlborn-Laake L, Mous J, Jacobsen H. Nonstructural protein 3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *J Virol*. 1993;67(7):3835-3844. doi:10.1128/jvi.67.7.3835-3844.1993
38. Kolykhalov AA, Agapov E V, Rice CM. Specificity of the hepatitis C virus NS3 serine protease: effects of substitutions at the 3/4A, 4A/4B, 4B/5A, and 5A/5B cleavage sites on polyprotein processing. *J Virol*. 1994;68(11):7525-7533. doi:10.1128/jvi.68.11.7525-7533.1994
39. Li K, Foy E, Ferreon JC, et al. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proc Natl Acad Sci U S A*. 2005;102(8):2992-2997. doi:10.1073/pnas.0408824102
40. Li XD, Sun L, Seth RB, Pineda G, Chen ZJ. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc Natl Acad Sci U S A*. 2005;102(49):17717-17722. doi:10.1073/pnas.0508531102
41. Huang Z-S, Wang C-C, Wu H-N. HCV NS3 protein helicase domain assists RNA structure conversion. *FEBS Lett*. 2010;584(11):2356-2362. doi:https://doi.org/10.1016/j.febslet.2010.04.020
42. Belon CA, Frick DN. Fuel specificity of the hepatitis C virus NS3 helicase. *Mol Cell Biochem*. 2012;23(1):1-7. doi:10.1016/j.jmb.2009.03.059.Fuel
43. Frick DN. HCV Helicase: Structure, Function, and Inhibition. *Hepat C Viruses Genomes Mol Biol*. Published online 2006:207-244. http://www.ncbi.nlm.nih.gov/pubmed/21250378
44. Mukherjee S, Hanson AM, Shadrick WR, et al. Identification and analysis of hepatitis C virus NS3 helicase inhibitors using nucleic acid binding assays. *Nucleic Acids Res*. 2012;40(17):8607-8621. doi:10.1093/nar/gks623
45. Beran RKF, Pyle AM. Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase. *J Biol Chem*. 2008;283(44):29929-29937. doi:10.1074/jbc.M804065200
46. Beran RKF, Lindenbach BD, Pyle AM. The NS4A Protein of Hepatitis C Virus Promotes RNA-Coupled ATP Hydrolysis by the NS3 Helicase. *J Virol*. 2009;83(7):3268-3275. doi:10.1128/jvi.01849-08
47. McGivern DR, Masaki T, Lovell W, Hamlett C, Saalau-Bethell S, Graham B. Protease Inhibitors Block Multiple Functions of the NS3/4A Protease-Helicase during the Hepatitis C Virus Life Cycle. *J Virol*. 2015;89(10):5362-5370. doi:10.1128/jvi.03188-14
48. Kim JL, Morgenstern KA, Griffith JP, et al. Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: The crystal structure provides insights into the mode of unwinding. *Structure*. 1998;6(1):89-100. doi:10.1016/S0969-2126(98)00010-0
49. Gu M, Rice CM. Three conformational snapshots of the hepatitis C virus NS3 helicase

reveal a ratchet translocation mechanism. *Proc Natl Acad Sci U S A*. 2010;107(2):521-528. doi:10.1073/pnas.0913380107
50. David N. Frick, Sukalyani Banik and RSR. Role of divalent metal cations in ATP hydrolysis catalyzed by the hepatitis C virus NS3 helicase: Magnesium provides a bridge for ATP to fuel unwinding. *J Mol Biol*. 2007;365(4):1017-1032. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf
51. Raney KD, Sharma SD, Moustafa IM, Cameron CE. Hepatitis C virus non-structural protein 3 (HCV NS3): A multifunctional antiviral target. *J Biol Chem*. 2010;285(30):22725-22731. doi:10.1074/jbc.R110.125294
52. Gwack Y, Kim DW, Han JH, Choe J. Characterization of RNA Binding Activity and RNA Helicase Activity of the Hepatitis C Virus NS3 Protein. *Biochem Biophys Res Commun*. 1996;225(2):654-659. doi:10.1006/BBRC.1996.1225
53. Lin C, Kim JL. Structure-Based Mutagenesis Study of Hepatitis C Virus NS3 Helicase. *J Virol*. 1999;73(10):8798-8807. doi:10.1128/jvi.73.10.8798-8807.1999
54. Myong S, Bruno MM, Pyle AM, Ha T. Virus NS3 Helicase. *Science (80-)*. 2007;317(5837):513-516. doi:10.1126/science.1144130.Spring-Loaded
55. Zheng W, Tekpinar M. Structure-based simulations of the translocation mechanism of the Hepatitis C virus NS3 helicase along single-stranded nucleic acid. *Biophys J*. 2012;103(6):1343-1353. doi:10.1016/j.bpj.2012.08.026
56. Wei Cheng, Arunajadai SG, Moffitt JR, Jr. IT, Bustamante C. Single–Base Pair Unwinding and Asynchronous RNA Release by the Hepatitis C Virus NS3 Helicase. *Science (80-)*. 2011;333(6050):1746-1749. doi:10.1126/science.1206023.Single
57. Rajagopal V, Gurjar M, Levin MK, Patel SS. The protease domain increases the translocation stepping efficiency of the hepatitis c virus NS3-4A helicase. *J Biol Chem*. 2010;285(23):17821-17832. doi:10.1074/jbc.M110.114785
58. Hafner M, Katsantoni M, Köster T, et al. CLIP and complementary methods. *Nat Rev Methods Prim*. 2021;1(1). doi:10.1038/s43586-021-00018-1
59. Ramanathan M, Porter DF, Khavari PA, Alto P. Methods to study RNA-protein interactions. *Nat Methods*. 2019;16(3):225-234. doi:10.1038/s41592-019-0330-1.Methods
60. Gräwe C, Stelloo S, van Hout FAH, Vermeulen M. RNA-Centric Methods: Toward the Interactome of Specific RNA Transcripts. *Trends Biotechnol*. 2021;39(9):890-900. doi:10.1016/j.tibtech.2020.11.011
61. Gerber AP. Rna-centric approaches to profile the rna–protein interaction landscape on selected RNAs. *Non-coding RNA*. 2021;7(1):1-15. doi:10.3390/ncrna7010011
62. Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. CLIP Identifies Nova-Regulated RNA Networks in the Brain. *Science (80-)*. 2003;302(5648):1212-1215. doi:10.1126/science.1090095
63. Lee FCY, Ule J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*. 2018;69(3):354-369. doi:10.1016/j.molcel.2018.01.005
64. Ule J, Jensen K, Mele A, Darnell RB. CLIP: A method for identifying protein–RNA interaction sites in living cells. *Methods*. 2005;37(4):376-386. doi:10.1016/J.YMETH.2005.07.018
65. Lee FCY, Ule J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*. 2018;69(3):354-369. doi:10.1016/J.MOLCEL.2018.01.005
66. Lin C, Miles WO. SURVEY AND SUMMARY Beyond CLIP: advances and opportunities to measure RBP-RNA and RNA-RNA interactions. *Nucleic Acids Res*.

2019;47(11):5490-5501. doi:10.1093/nar/gkz295
67. Chin JW. Expanding and reprogramming the genetic code. *Nature*. 2017;550(7674):53-60. doi:10.1038/nature24031
68. Jianming Xie and Peter G. Schultz. A chemical toolkit for proteins — an expanded genetic code. *Annu Rev Biochem*. 2010;79:413-444.
69. Adhikari A, Bhattarai BR, Aryal A, et al. Reprogramming natural proteins using unnatural amino acids. *RSC Adv*. 2021;11(60):38126-38145. doi:10.1039/d1ra07028b
70. Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem*. 2010;79:413-444. doi:10.1146/annurev.biochem.052308.105824
71. Liu W, Brock A, Chen S, Chen S, Schultz PG. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods*. 2007;4(3):239-244. doi:10.1038/nmeth1016
72. Zhu HQ, Tang XL, Zheng RC, Zheng YG. Recent advancements in enzyme engineering via site-specific incorporation of unnatural amino acids. *World J Microbiol Biotechnol*. 2021;37(12):1-13. doi:10.1007/s11274-021-03177-1
73. Xie J, Schultz PG. An expanding genetic code. *Methods*. 2005;36(3):227-238. doi:10.1016/j.ymeth.2005.04.010
74. Tanaka Y, Bond Ab MR, Kohler JJ. Molecular BioSystems Photocrosslinkers illuminate interactions in living cellsw. *Mol Biosyst*. 2008;4:36. doi:10.1039/b803218a
75. Kauer JC, Erickson-Viitanen S, Wolfe HR, DeGrado WF. p-Benzoyl-L-phenylalanine, a new photoreactive amino acid. Photolabeling of calmodulin with a synthetic calmodulin-binding peptide. *J Biol Chem*. 1986;261(23):10695-10700. doi:10.1016/s0021-9258(18)67441-1
76. Chin JW, Schultz PG. In vivo photocrosslinking with unnatural amino acid mutagenesis. *ChemBioChem*. 2002;3(11):1135-1137. doi:10.1002/1439-7633(20021104)3:11<1135::AID-CBIC1135>3.0.CO;2-M
77. Jaya N, Garcia V, Vierling E. Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proc Natl Acad Sci U S A*. 2009;106(37):15604-15609. doi:10.1073/pnas.0902177106
78. Majmudar CY, Wang B, Lum JK, Hâkansson K, Mapp AK. A high-resolution interaction map of three transcriptional activation domains with a key coactivator from photo-cross-linking and multiplexed mass spectrometry. *Angew Chemie - Int Ed*. 2009;48(38):7021-7024. doi:10.1002/anie.200902669
79. Majmudar CY, Lee LW, Lancia JK, et al. Impact of Nonnatural Amino Acid Mutagenesis on the in Vivo Function and Binding Modes of a Transcriptional Activator. doi:10.1021/ja904378z
80. Lee HS, Dimla RD, Schultz PG. Protein–DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. *Bioorg Med Chem Lett*. 2009;19(17):5222-5224. doi:10.1016/J.BMCL.2009.07.011
81. Dormán G, Nakamura H, Pulsipher A, Prestwich GD. The Life of Pi Star: Exploring the Exciting and Forbidden Worlds of the Benzophenone Photophore. *Chem Rev*. 2016;116(24):15284-15398. doi:10.1021/acs.chemrev.6b00342
82. Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J Am Chem Soc*. 2002;124(31):9026-9027. doi:10.1021/ja027007w
83. Rannversson H, Andersen J, Sørensen L, et al. Genetically encoded photocrosslinkers

locate the high-affinity binding site of antidepressant drugs in the human serotonin transporter. *Nat Commun*. 2016;7. doi:10.1038/ncomms11261
84. Ahmed N, Foss D V., Powdrill MH, Pezacki JP. Site-Specific Cross-Linking of a p19 Viral Suppressor of RNA Silencing Protein and Its RNA Targets Using an Expanded Genetic Code. *Biochemistry*. 2019;58(33):3520-3526. doi:10.1021/acs.biochem.9b00428
85. Preston GW, Wilson AJ. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem Soc Rev*. 2013;42(8):3289-3301. doi:10.1039/c3cs35459h
86. Schnapp KA, Poe R, Leyva E, Soundararajan N, Platz MS. *Exploratory Photochemistry of Fluorinated Aryl Azides. Implications for the Design of Photoaffinity Labeling Reagents*. Vol 4.; 1993. https://pubs.acs.org/sharingguidelines
87. He Tian, Thomas P. Sakmar and TH. A Simple Method for Enhancing the Bioorthogonality of Cyclooctyne Reagent. *Chem Commun*. 2016;52(31):5451-5454. doi:10.1039/x0xx00000x
88. Tippmann EM, Liu W, Summerer D, Mack A V., Schultz PG. A genetically encoded diazirine photocrosslinker in Escherichia coli. *ChemBioChem*. 2007;8(18):2210-2214. doi:10.1002/cbic.200700460
89. Yu W, Baskin JM. Photoaffinity labeling approaches to elucidate lipid–protein interactions. *Curr Opin Chem Biol*. 2022;69:102173. doi:10.1016/J.CBPA.2022.102173
90. Tippmann EM, Liu W, Summerer D, Mack A V., Schultz PG. A genetically encoded diazirine photocrosslinker in Escherichia coli. *ChemBioChem*. 2007;8(18):2210-2214. doi:10.1002/cbic.200700460
91. Brunner J, Senn H, Richards FM. 3-Trifluoromethyl-3-phenyldiazirine. A new carbene generating group for photolabeling reagents. *J Biol Chem*. 1980;255(8):3313-3318. doi:10.1016/s0021-9258(19)85701-0
92. Palaz F, Kalkan AK, Can Ö, et al. CRISPR-Cas13 System as a Promising and Versatile Tool for Cancer Diagnosis, Therapy, and Research. *ACS Synth Biol*. 2021;10(6):1245-1267. doi:10.1021/acssynbio.1c00107
93. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*. 2005;60(2):174-182. doi:10.1007/s00239-004-0046-3
94. Bolotin A, Quinquis B, Sorokin A, Dusko Ehrlich S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005;151(8):2551-2561. doi:10.1099/mic.0.28048-0
95. Pyzocha NK, Chen S. Diverse Class 2 CRISPR-Cas Effector Proteins for Genome Engineering Applications. *ACS Chem Biol*. 2018;13(2):347-356. doi:10.1021/acschembio.7b00800.Diverse
96. Garrett SC. Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array. *Front Microbiol*. 2021;12. doi:10.3389/fmicb.2021.664299
97. Arroyo-Olarte RD, Bravo Rodríguez R, Morales-Ríos E. Genome editing in bacteria: Crispr-cas and beyond. *Microorganisms*. 2021;9(4). doi:10.3390/microorganisms9040844
98. Pickar-Oliver A, Gersbach CA. The next generation of CRISPR–Cas technologies and applications. *Nat Rev Mol Cell Biol*. 2019;20(8):490-507. doi:10.1038/s41580-019-0131-5
99. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. 2012;337(6096):816-821. doi:10.1126/science.1225829

100. Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/cas-mediated genome engineering. *Cell*. 2013;153(4):910-918. doi:10.1016/j.cell.2013.04.025
101. Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-04252-2
102. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143
103. Rees H, Liu D. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. 2018;19:770-788. doi:10.1038/s41576-018-0059-1.Base
104. Shmakov S, Abudayyeh OO, Makarova KS, et al. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell*. 2015;60(3):385-397. doi:10.1016/j.molcel.2015.10.008
105. Omar O. Abudayyeh, Jonathan S. Gootenberg, Silvana Konermann, Julia Joung, Ian M. Slaymaker, David B.T. Cox, Sergey Shmakov, Kira S. Makarova, Ekaterina Semenova, Leonid Minakhin, Konstantin Severinov, Aviv Rege, Eric S. Lander, Eugene V. Koonin FZ. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (80-)*. 2015;353(6299):1-23. doi:10.1126/science.aaf5573.C2c2
106. Smargon AA, Cox DBT, Pyzocha NK, et al. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell*. 2017;65(4):618-630.e7. doi:10.1016/j.molcel.2016.12.023
107. Slaymaker IM, Mesa P, Kellner MJ, et al. High-Resolution Structure of Cas13b and Biochemical Characterization of RNA Targeting and Cleavage. *Cell Rep*. 2019;26(13):3741-3751.e5. doi:10.1016/j.celrep.2019.02.094
108. Cox DBT, Gootenberg JS, Abudayyeh OO, et al. RNA editing with CRISPR-Cas13. *Science (80-)*. 2017;358(6366):1019-1027. doi:10.1126/science.aaq0180

Chapter 2: Incorporation of p-azido-L-phenylalanine into HCV NS3h to Test for CLIP Applicability and Determination of NS3h Binding Interactions

2.1 Abstract

To determine if the photo-crosslinking capabilities of the ncAA AzF could be used broadly to enhance CLIP-based protocols, we introduced five site-specific AzF mutations into the HCV NS3h. The sites chosen for AzF incorporation were E493, E503, Q580, T416, and K372. Each of these AzF sites targets a different element on nucleic acids for photo-crosslinking, namely either the phosphate backbone or the nucleobase. From there, we assessed the incorporation of AzF and photo-crosslinking proficiency of each AzF mutant when bound to nucleic acid substrates. Of the five AzF mutants created, only four were able to express in *E. coli*. Furthermore, mutants E503AzF and Q580AzF were able to photo-crosslink a DNA substrate using 302 nm UV light. Through fluorescence polarization and a molecular beacon helicase assay, we determined that mutant E503AzF had a higher nucleic acid unwinding rate and similar affinity for DNA substrate compared to the WT NS3h. We then compared the photo-crosslinking capabilities of AzF at 302 nm light versus the conventional 254 nm light commonly used in CLIP and found a 5-fold increase in photo-crosslinking efficiency for E503AzF when compared to WT NS3h at 254 nm light. Next, we set out to develop a CLIP-based experiment where we could investigate NS3h interactions with the RNA secondary structures formed within the HCV genome. Currently, we are working on developing a photo-crosslinking and RNA isolation protocol that can ensure quality RNA footprint generation for the identification of NS3h binding locations.

2.2 Introduction

In the early 1960s, it was found that the absorption of UV light could induce the crosslinking of DNA to proteins [1]. These crosslinks are formed when cellular proteins form a stable covalent bond upon exposure to endogenous and environmental stimuli, such as reactive oxygen species, UV light as well as carcinogens. Each of these methods has unique chemistries which allow the formation of a stable DNA-protein crosslink that can interfere with the progression of cellular DNA replication and transcription systems [2]. Upon exposure to UV light, purine and pyrimidines nucleobases form free radicals, which, if in close proximity to almost all amino acids react to form a stable covalent bond [3]. The CLIP methodology began to take shape in the early 2000s, where covalent photo-crosslinking allowed for stringent purification of RNA–protein complexes in mapping and understanding the complex interactions of RBP and spawned a significant amount of complementary methods [4].

The problem that many of these protocols face is the low photo-crosslinking efficiency, around 1%, between proteins and nucleic acids when irradiated at 254 nm light [5,6]. Also UV crosslinking in general, is less likely to capture weak RNA and protein interactions [7]. There have been alterations to the CLIP method such as photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation, where the photo-activatable nucleotide 4-thiouridine (4SU) or 6-thioguanine (6SG) is introduced into the cellular media where it is incorporated into nascent RNA transcripts. These labeled RNAs are excited in living cells and have excellent photo-crosslinking yields [8]. In a similar vein to introducing an unnatural element into the CLIP methodology, photo-crosslinking ncAAs have been theorized to be used in CLIP protocols to enhance the yield of photo-crosslinked RNA [5,6]. For example, the ncAA pBpa has been incorporated into *E. coli* catabolite activator protein to form covalent protein-DNA upon UV

irradiation [9]. Diazirines incorporated into human iron regulatory protein 1 showed a 7-fold increase in photo-crosslinking efficiency compared to conventional protein nucleic acid photo-crosslinking [6]. The applicability of photo-crosslinking aryl azides ncAAs, such as AzF, in CLIP-based experiments is still unknown. Previously, AzF has been successfully incorporated into the p19 viral suppressor to photo-crosslink and irreversibly sequester bound RNA [5]. Here we wished to assess the applicability of site-specific incorporation of AzF into CLIP-based protocols by evaluating the photo-crosslinking efficiency of AzF compared to irradiation with 254 nm UV light.

Using the AzF photo-crosslinking capabilities we wanted to investigate the interactions between NS3h and the HCV 9.6kb RNA genome, which forms extensive secondary and tertiary regulatory structures which are essential for its translation, replication and viral packaging [10–13]. The most famous of these structures is the 5′UTR IRES which is essential for HCV cap-independent translation upon entry into the cell by binding the ribosomal subunits to hijack the system for translation. There are extensive stem-loop structures found in the viral genomic RNA that interact with the viral capsid core proteins for viral assembly and multiple structures work in tandem to allow this to happen [14]. The 3′UTR is also known to play an essential role in HCV RNA replication [15,16]. It has been found that NS3h preferentially binds to the 3′ UTR of the HCV genome as well as aptamers resembling this sequence [17,18]. NS3h has been shown to promote intramolecular RNA strand annealing and facilitates RNA structure conversions [19]. It has also been proposed that the helicase might be actively involved in the initiation of RNA synthesis by resolving stem-loop structures found in the 3′UTR [20,21]. With this knowledge, we wanted to investigate where NS3h might also preferentially bind along the 9.6kb folded RNA genome.

In this current work, we describe the incorporation of the photo-crosslinking ncAA AzF into the HCV NS3h and assess its nucleic acid photo-crosslinking capabilities. We also used

fluorescence polarization and a molecular beacon helicase assay to determine if binding affinity or helicase functionality was lost due to AzF incorporation. Finally, we also sought to optimize a CLIP-based protocol to determine binding sites of NS3h on the HCV genome.

2.3 Results

2.3.1 Incorporation of AzF into Selected Sites in NS3h

The first step was to identify side chains in NS3h to introduce AzF which are in close proximity to RNA for photo-crosslinking. We examined the crystal structure of NS3h for potential sites of incorporation and a total of five sites were identified: E493, E503, Q580, T416, and K372 (**Figure 2.1 A**). Positions K372 and Q580 flank the entrance of the nucleic acid binding cleft, while E503 flanks the exit of the nucleic acid binding cleft, E493 is located directly inside the cleft and T416 is located above the bound strand pointing downward into the cleft. The structure of AzF was modeled onto each of these sites to further gain an initial understanding of the proximity of the AzF to the nucleic acid (**Supplemental Figure 2.1**). For mutations K372AzF and T416AzF, the reactive groups are both positioned near the phosphate backbone. Mutants E503AzF, E493AzF and Q580AzF have the AzF positioned to interact with the nucleobases.

We generated five constructs of NS3h protein with the TAG amber stop at each of the chosen sites. To determine if AzF was being incorporated in the chosen positions, the five constructs were expressed in *E.coli* with and without AzF supplemented in the media. The culture was then lysed and electrophoresed on an SDS-PAGE gel and imaged for equal loading. A His-tag western blot was performed on the lysates to probe for full length NS3h which expresses a C-terminal His-tag (**Figure 2.1 B**). Except for E493AzF, each NS3h AzF mutant was expressed when AzF was included in the growth media. We see little to no expression of the mutants in the absence of AzF in the growth media.

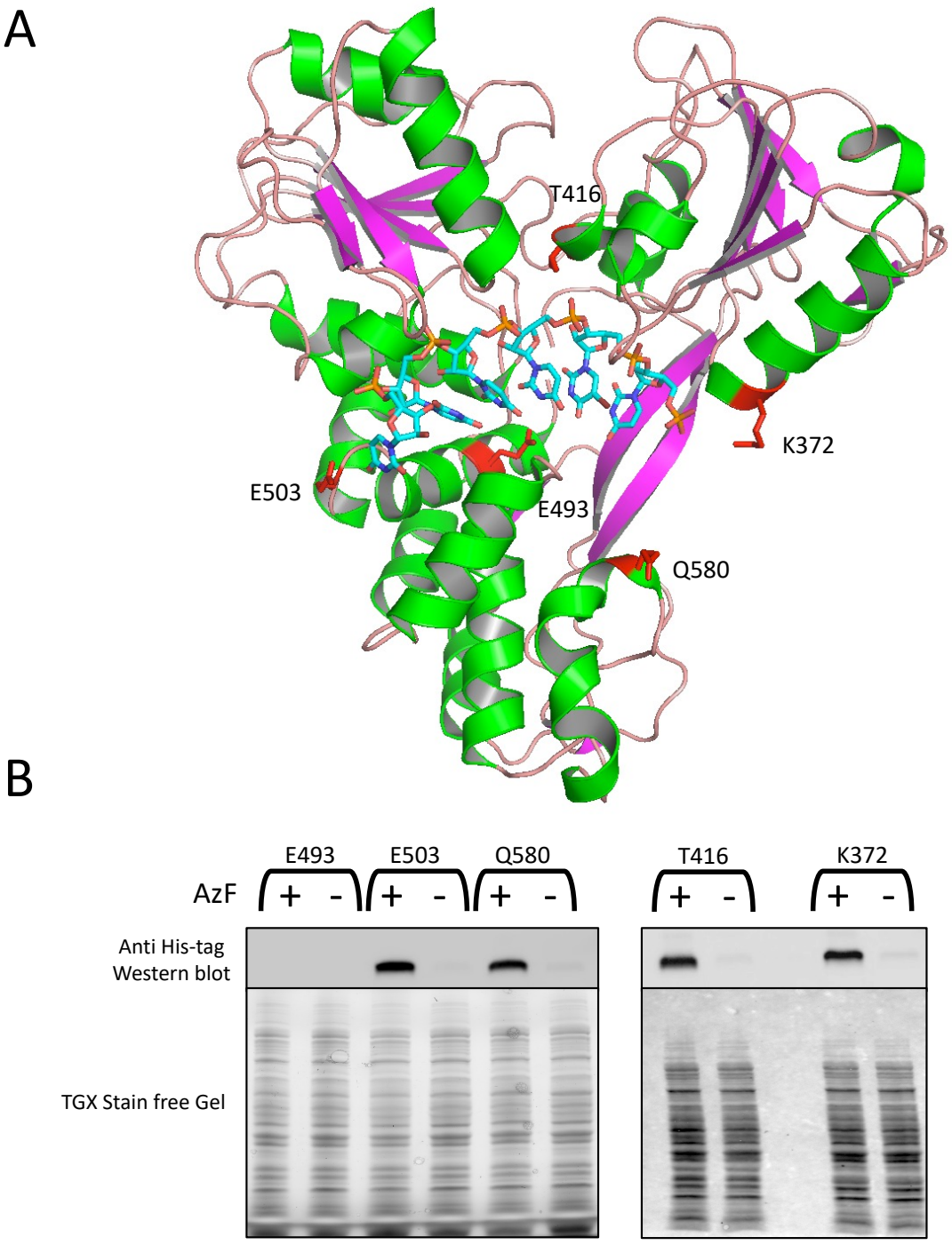


Figure 2.1 Site-specific Incorporation of AzF into NS3h: (A) The five sites chosen for AzF incorporation are modeled onto the NS3h crystal structure (PDB: 3O8C). The chosen sites are highlighted in red and bound RNA is shown in blue. (B) Western blot of the expression of all NS3h amber stop mutants. Mutants were expressed with 1 mM IPTG in the presence or absence of 1 mM AzF in the growth media. Cells were harvested 4 hours post-induction.

2.3.2 Photo-crosslinking of NS3h AzF Mutants to Nucleic Acids

The next step was to determine if the incorporated AzF was in close enough proximity to an element of the nucleic acid where, upon UV irradiation, it could form a stable covalent bond. To do this, we first purified NS3h WT and all four of the expressing AzF NS3h mutants using a His-tag protein purification system. We further purified our protein through size-exclusion fast protein liquid chromatography (FPLC). The protein purity can be seen in the chromatograms (**Supplemental Figure 2.2**). The fractions containing the pure NS3h were pooled and concentrated.

To assess the photo-crosslinking capabilities of all the mutants, purified mutants were incubated with a random 20-mer ssDNA substrate labelled with a Cy5 fluorophore on the 3′ end and a duplexed double-stranded DNA (dsDNA) substrate with a PolyT(20) loading region for NS3h to bind (**Figure 2.2 A**). To induce photo-crosslinking post-incubation, samples were irradiated with 302 nm light and band shifting was examined using denaturing SDS-PAGE gels **(Figure 2.2 B, D**). First, we observed a shift in mass from 52 kDa in the stain-free gel to 60kDa and 78kDa for NS3h photo-crosslinked to Cy5 ssDNA and dsDNA, respectively. In the Cy5 image of both SDS-PAGE gels, NS3h WT has a very faint band when photo-crosslinked to ssDNA and the dsDNA substrate. Mutants E503AzF and Q580AzF have intense bands for both DNA substrates. Using densitometry, we compared the Cy5 intensity of WT NS3h against E503AzF and Q580AzF (**Figure 2.2 C**). The Cy5 intensity of the bound ssDNA compared to WT is 4- and 2-fold greater for mutants E503AzF and Q580AzF, respectively. With the dsDNA substrate, we observed a larger photo-crosslinking efficiency compared to the WT with a near 10-fold and 5-fold Cy5 intensity increase for E503AzF and Q580AzF. For mutants T416AzF and K372AzF, we observed similar photo-crosslinking to that of WT under UV irradiation (**Figure 2.2 E**).

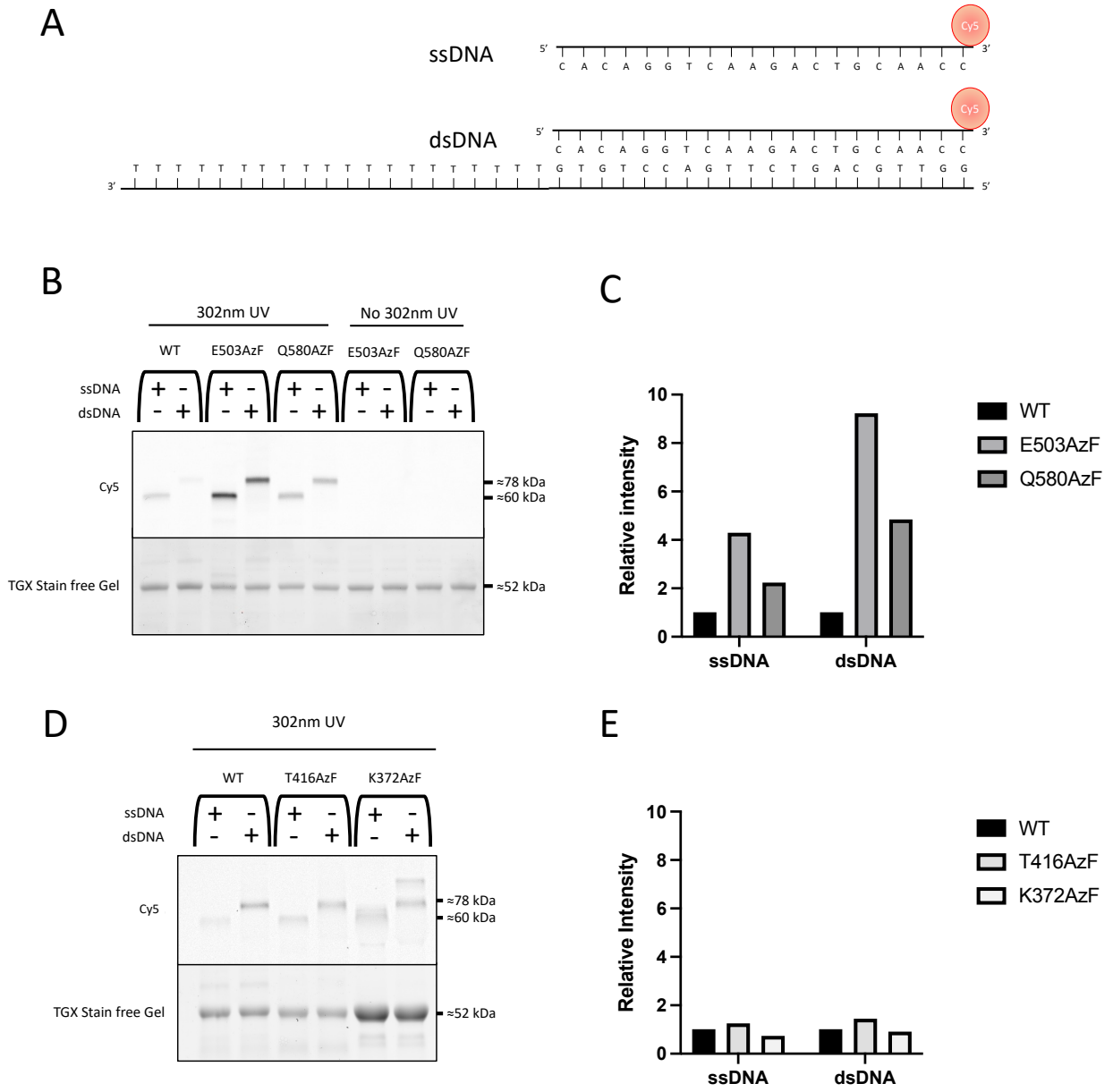


Figure 2.2 Mutants E503AzF and Q580AzF Demonstrate Enhanced Photo-crosslinking Activity: (A) Structure of 20-mer ssDNA and dsDNA Cy5 labelled substrates used in the photo-crosslinking assay. Cy5 is shown as a red circle. (B and D) Photo-crosslinking SDS-PAGE gels for mutants E503AzF, Q580AzF, T416AzF and K372AzF. 10 uM of purified NS3h was incubated with 250 nM of DNA substrate and irradiated with 302 nm UV light for 10 minutes. On the top is the Cy5 fluorescence scan and on the bottom is the stain-free image to confirm equal loading. (C and E) Densitometry analysis was performed by normalizing the Cy5 intensity with the protein abundance. The adjusted intensity was then compared relative to WT as a control.

2.3.3 Characterizing AzF for CLIP Applications

To determine if AzF incorporated into a protein could provide enhanced photo-crosslinking capabilities against the traditional 254 nm wavelength used in CLIP methods, we measured the photo-crosslinking ability of WT NS3h using 254 nm light and mutants E503AzF and Q580AzF at 302 nm light. We then ran the photo-crosslinked products on an SDS-PAGE gel and imaged for Cy5 fluorescence (**Figure 2.3 A**). We can see that with 10 minutes of 254 nm UV light exposure, there is little photo-crosslinking for WT NS3h; however, for both E503AzF and Q580AzF, we observe more intense bands. Densitometry analysis was used to compare the intensity of mutants E503AzF and Q580AzF bands at 302 nm irradiation to WT and at 254 nm light. We obtained a 4.5-fold increase in the intensity of the Cy5 signal for E503AzF. (**Figure 2.3 B**). We can also see banding patterns above the expected bands for mutant E503AzF.

Research determining the reactivity of didehydroazepine, an intermediate formed when AzF is irradiated with UV reacts very readily with guanosine monophosphate, adenosine and uridine. However, it was found that the didehydroazepine, did not react with cytidine at a measurable rate [22]. As AzF is being used to photo-crosslink nucleic acids with the end goal of CLIP applications, we wanted to ensure that there was no sequence bias to the nucleobase the incorporated AzF was able to photo-crosslink. To do this, we tested the photo-crosslinking ability of E503AzF and Q580AzF to three different Cy5 labelled DNA substrates: a random 20-mer ssDNA, a PolyT(15-mer) and a PolyC(15-mer) (**Supplementary Table 2.1**). Each substrate was incubated and photo-crosslinked to E503AzF, Q580AzF, or WT and analyzed on a denaturing SDS-PAGE gel. (**Figure 2.3 C**) We can see that there is no visible difference between the photo-crosslinking ability of AzF to any of these substrates as the band intensities are comparable. Again we observed banding patterns above the expected bands for both E503AzF and Q580AzF.

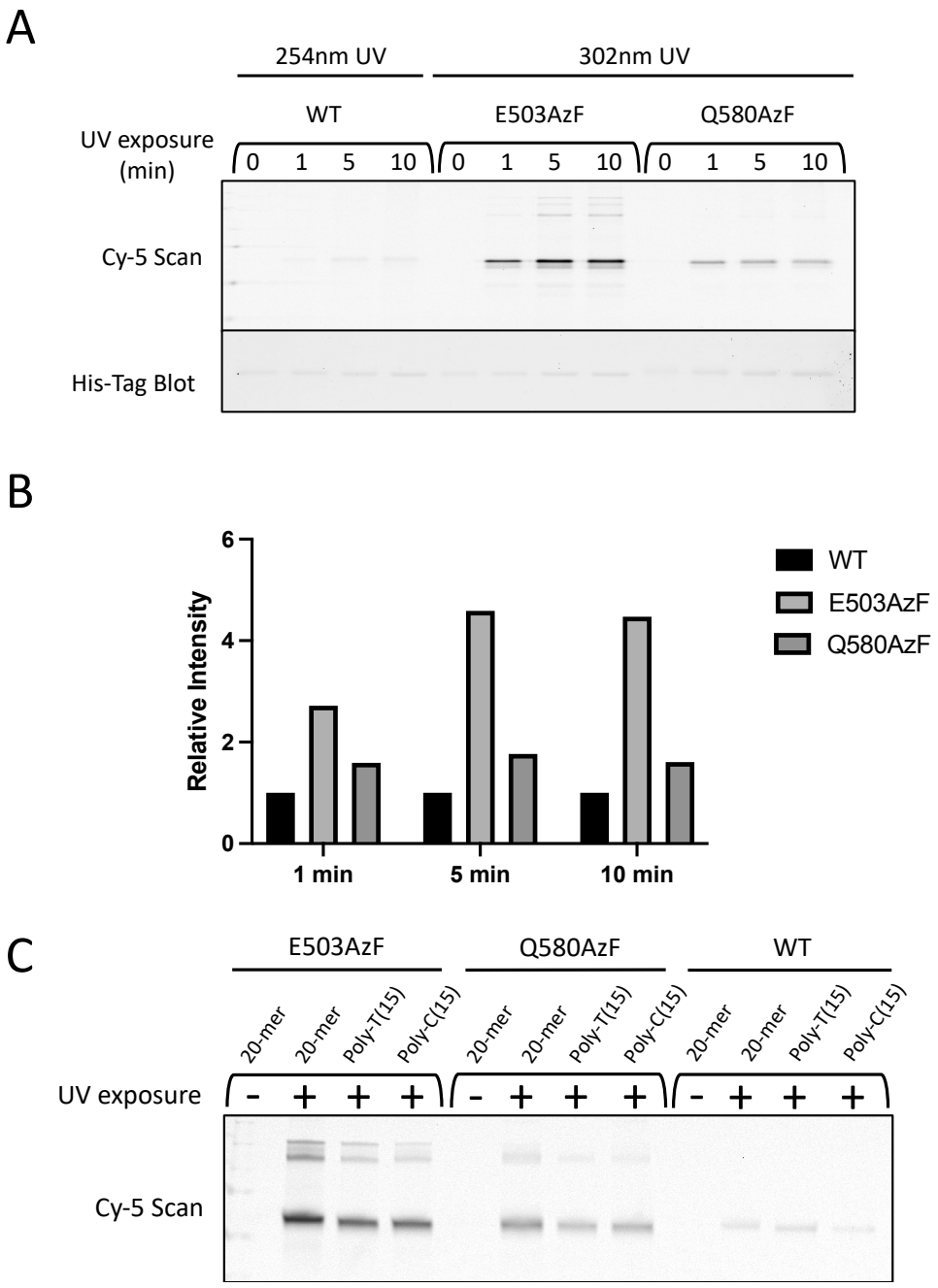


Figure 2.3 Assessing Photo-crosslinking Capabilities of NS3h AzF Mutants: (A) Photo-crosslinking SDS-PAGE gel comparing mutants E503AzF and Q580AzF against WT. 10 uM of purified NS3h was incubated with 250 nM of DNA substrate and irradiated with 302 nm or 254 nm UV light. (B) Densitometry analysis was performed by normalizing the Cy5 intensity against the protein abundance. The adjusted intensity of E503AzF and Q580AzF photo-crosslinked at 302 nm light was then compared relative to WT photo-crosslinked at 254 nm light. (C) Testing to see if there are any sequence biases to AzF photo-crosslinking. 10 uM of purified NS3h was incubated with 250 nM of a random 20-mer, PolyT(15-mer) and PolyC(15-mer) DNA substrate and irradiated with 302 nm UV light.

2.3.4 Measuring Helicase Activity for AzF Photo-crosslinking Mutants

Next, we needed to ensure that the introduction of AzF did not alter the nucleic acid binding or ATPase functions of NS3h. Having identified E503AzF and Q580AzF as the two constructs which could efficiently photo-crosslink a labelled nucleic acids substrate, we used a molecular beacon helicase assay to monitor the helicase activity of the enzyme (**Figure 2.4 A**). This assay uses a dual labelled DNA oligomer with a quencher on the 5′ and fluorophore on the 3′ annealed into a duplex, allowing the measurements of the fluorophore wavelength. As the helicase unwinds the duplex, the labelled strand will form a stable hairpin loop bringing the quencher and fluorophore in close proximity terminating the fluorescent signal.

As a negative control, an NS3h construct was created with the well-characterized loss of helicase activity W501A mutation. This mutant was His-tag purified and size-exclusion chromatography was used to obtain pure W501A. The unwinding rate of the WT NS3h was found to be 0.36 ± 0.02 nM/min and the most efficient photo-crosslinking mutant E503AzF had a faster unwinding rate of 0.80 ± 0.08 nM/min, more than a 2-fold increase compared to WT (**Figure 2.4 B**). Q580AzF had a reduced unwinding rate of 0.12 ± 0.02 nM/min in comparison to WT. An unwinding rate could not be obtained for the W501A mutant, though a small downward trend is observed.

Next, we wanted to consider how the binding affinity of the AzF mutants could be altered by the ncAA. To investigate this, fluorescence polarization was used to measure the difference in polarized light from NS3h when bound and unbound from DNA. The concentration of a Cy3 labelled random 15-mer was kept constant and NS3h was titrated in a 2-fold dilution (**Figure 2.5**). With an increasing amount of WT NS3h, we observe a clear increase in polarization of the Cy3 15-mer in a concentration dependant manner, where the amount of WT NS3h needed to bind half

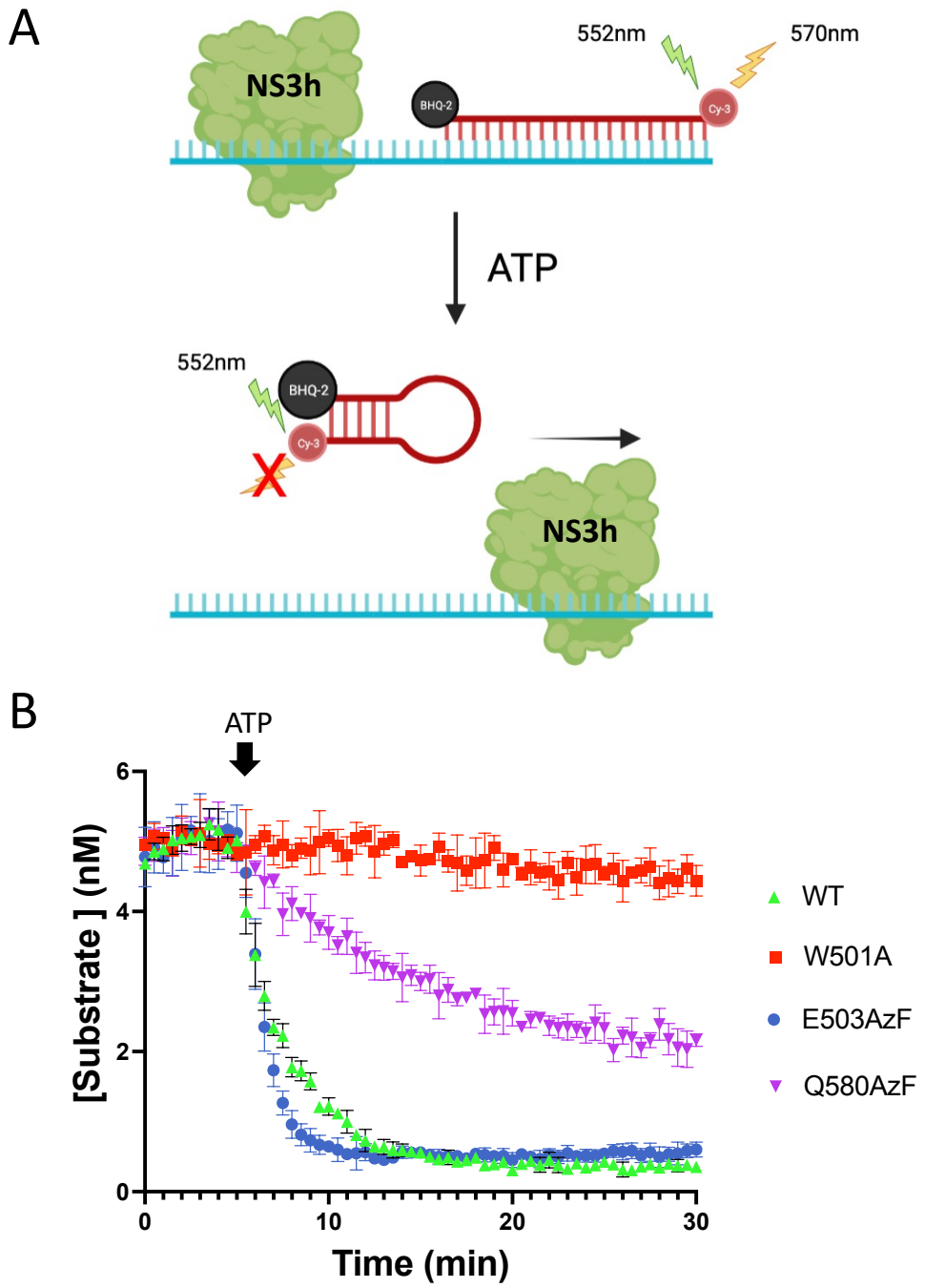


Figure 2.4 Mutant E503AzF has a Faster Unwinding Rate than WT NS3h: (A) Schematic of the molecular beacon helicase assay. Upon the addition of ATP, the helicase unwinds the substrate allowing the fluorophore to be brought near the quencher resulting in a loss of signal. (B) Molecular beacon helicase assay substrate unwound upon the addition of 0.5 mM ATP. Reactions contained 150 nM NS3h and 5 nM of a molecular beacon helicase assay nucleic acid substrate. All values represent the mean ± SD (n=3).

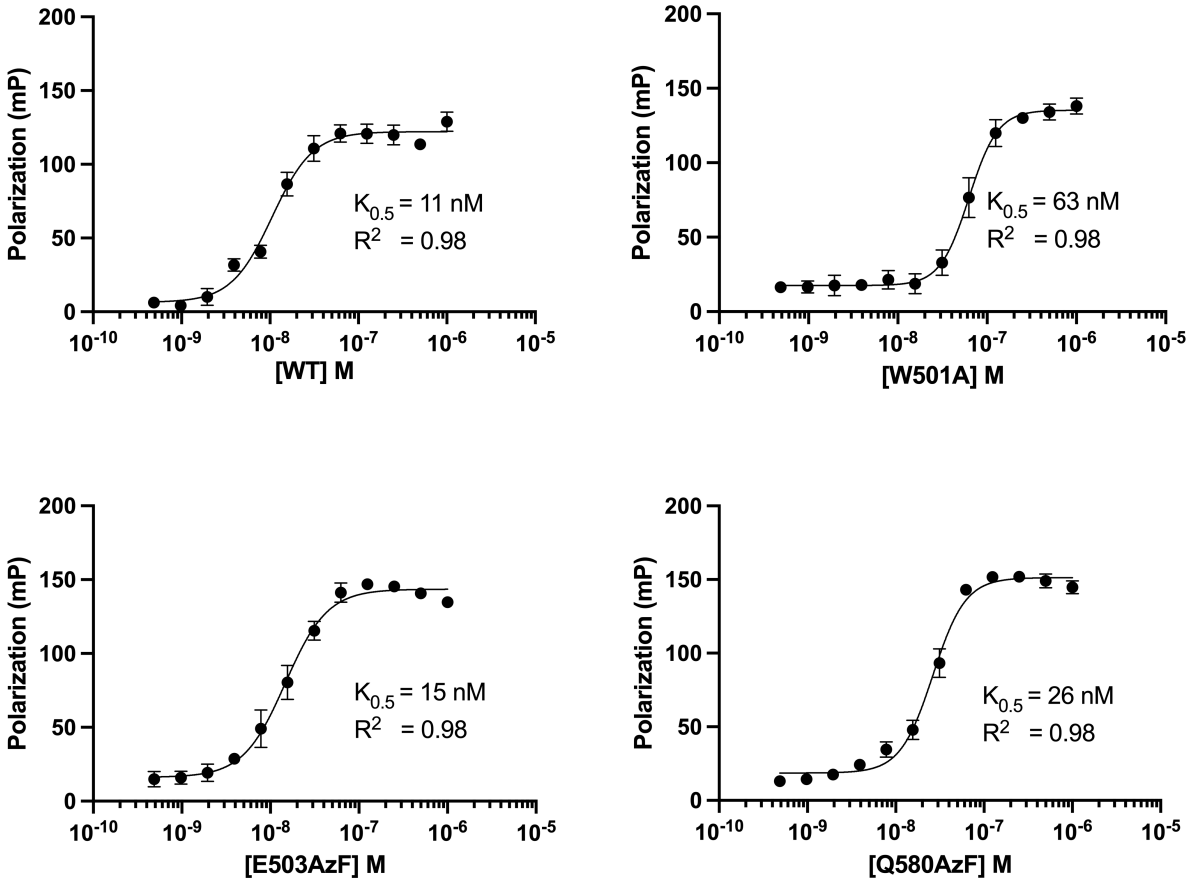


Figure 2.5 Mutant E503AzF Binds ssDNA with a Similar Magnitude to WT: (A) Binding plots were constructed using 2-fold serial diluted (0.48 - 1000 nM) NS3h protein and 5 nM Cy3 labelled DNA substrate. Fluorescence polarization was monitored by exciting at 535 nm and measuring parallel and perpendicular polarized light at 595 nm. Data were fitted to a concentration-response equation (four-parameter, variable slope). $K_{0.5}$ represents the amount of NS3h needed to bind half of the Cy3 DNA substrate. All values represent the mean ± SD (n=3).

the Cy3-15mer ($K_{0.5}$) was 11 ± 1.6 nM. E503AzF and Q580AzF had concentration-dependent curves with $K_{0.5}$ of 15 ± 1.8 nM and 26 ± 2.4 nM, respectively. The reduced binding mutant W501A had a $K_{0.5}$ of 63 ± 5.3 nM and plateaued at a higher concentration of around 125 nM, compared to the 31.2 nM plateau of WT.

With both the molecular beacon helicase assay and fluorescence polarization analysis showing a higher level of nucleic acid unwinding and a minimal change in nucleic acid affinity, we identified E503AzF to be the most suitable mutant to move forward with further CLIP-based applications to generate and isolate RNA binding locations for NS3h on the HCV RNA genome.

2.3.5 Optimizing a CLIP-based protocol

We next constructed and optimized an *in vitro* photo-crosslinking and affinity chromatography methodology to identify binding locations of NS3h along the genomic RNA of HCV. The schematic of the general protocol is shown in **Figure 2.5**. Briefly, we first generated full length Japanese Fulminant Hepatitis 1 (JFH-1) HCV genomic RNA through *in vitro* transcription (IVT) and folded it to ensure the proper formation of secondary structure. Next, we needed to photo-crosslink our E503AzF to bind to the IVT HCV genome using 302 nm light. For this step, we needed to find a buffer to fold the HCV genome and allow for efficient NS3h photo-crosslinking. We assessed two HCV RNA folding buffers that have been previously used in the identification of RNA secondary structures folding, which we identify as buffers A and B (**Supplemental Figure 2.3 A**) [14,23]. Since both buffers had a large ionic concentration of salts, we sought to test if the different folding buffers would have an adverse effect on NS3h binding and photo-crosslinking ability. To do this, we photo-crosslinked E503AzF to a random 20-mer labelled ssDNA substrate in both folding buffers A and B and ran this on an SDS-PAGE gel, and imaged

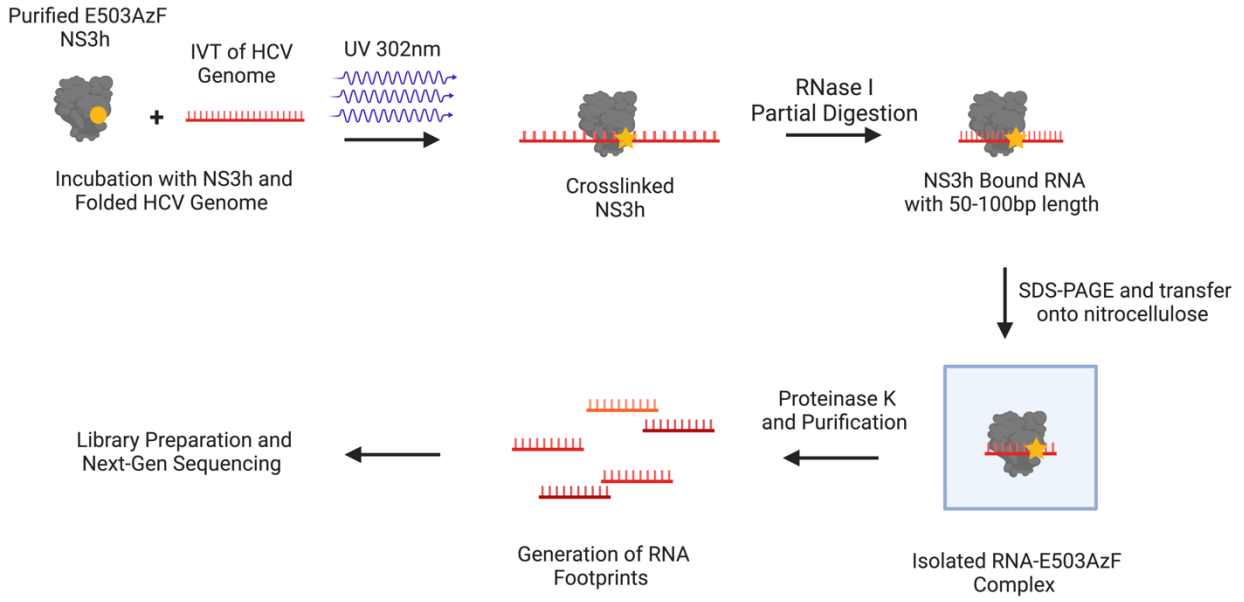


Figure 2.6 Generating and Optimizing a Protocol to Generate NS3h Binding Locations: Schematic of the general CLIP-based protocol. Full length JFH-1 HCV genomic RNA is folded to ensure the formation of the proper secondary structure before being introduced to E503AzF. The reaction is photo-crosslinked using 302 nm UV light. RNase I will be used to partially digest the bound RNA generating proper-sized fragments. The resulting E503AzF photo-crosslinked to HCV genomic RNA will be further separated from background RNA by isolating it using SDS-PAGE gel and transferring it onto a nitrocellulose membrane. The nitrocellulose is cut removing the E503AzF and HCV RNA complex and treated with proteinase K to digest E503AzF releasing the bound RNA. The RNA is then sent for library preparation and Next-Gen Sequencing.

for Cy5 photo-crosslinking (**Supplemental Figure 2.3 B**). We found that neither of the folding buffers allowed for photo-crosslinking. Upon further investigation, we found that with increasing concentrations of NaCl, there was a loss of photo-crosslinking between E503AzF and the labelled DNA substrate (**Supplemental Figure 2.3 B**).

The next step of the protocol was to use RNase to fragment the RNA bound to E503AzF, resulting in fragments bound in the range of 50-100 bp in length. This step needed to be optimized as the duration of digestion and quantity of RNase differs for each method. As well, determining the optimal fragmentation time is essential for downstream applications of the RNA. We fragmented using multiple dilutions of RNase I and ran the resulting products on a 12% urea denaturing gel (**Figure 2.6)**. It was found that a 1/10 dilution of RNase I at 37^{o}C for 20 minutes led to a large smear of fragmented RNA found within the desired range of 50-100 bp in length. The addition of undiluted RNase I led to the complete degradation of the genomic RNA. At much higher dilutions of 1/100,000 and 1/1,000,000, we have little to no degradation.

The final steps of this CLIP-based protocol are NS3h-RNA complex isolation and RNA purification (**Figure 2.5**). To do this, we need to first isolate the photo-crosslinked RNA-protein complex from the RNases and unbound RNA fragments. Instead of using immunoprecipitation, we decided to use the His-tag on E503AzF and perform affinity chromatography under denaturing conditions. The resulting E503AzF bound to HCV genomic RNA will be further removed from background RNA by isolating it using SDS-PAGE gel and transferring it onto a nitrocellulose membrane. From here we cut the nitrocellulose membrane to remove the E503AzF and HCV RNA complex. These nitrocellulose fragments are then treated with proteinase K to digest E503AzF and release the bound RNA. This RNA can be purified and sent for library preparation and Next-gen sequencing to identify the precise locations where our E503AzF is bound.

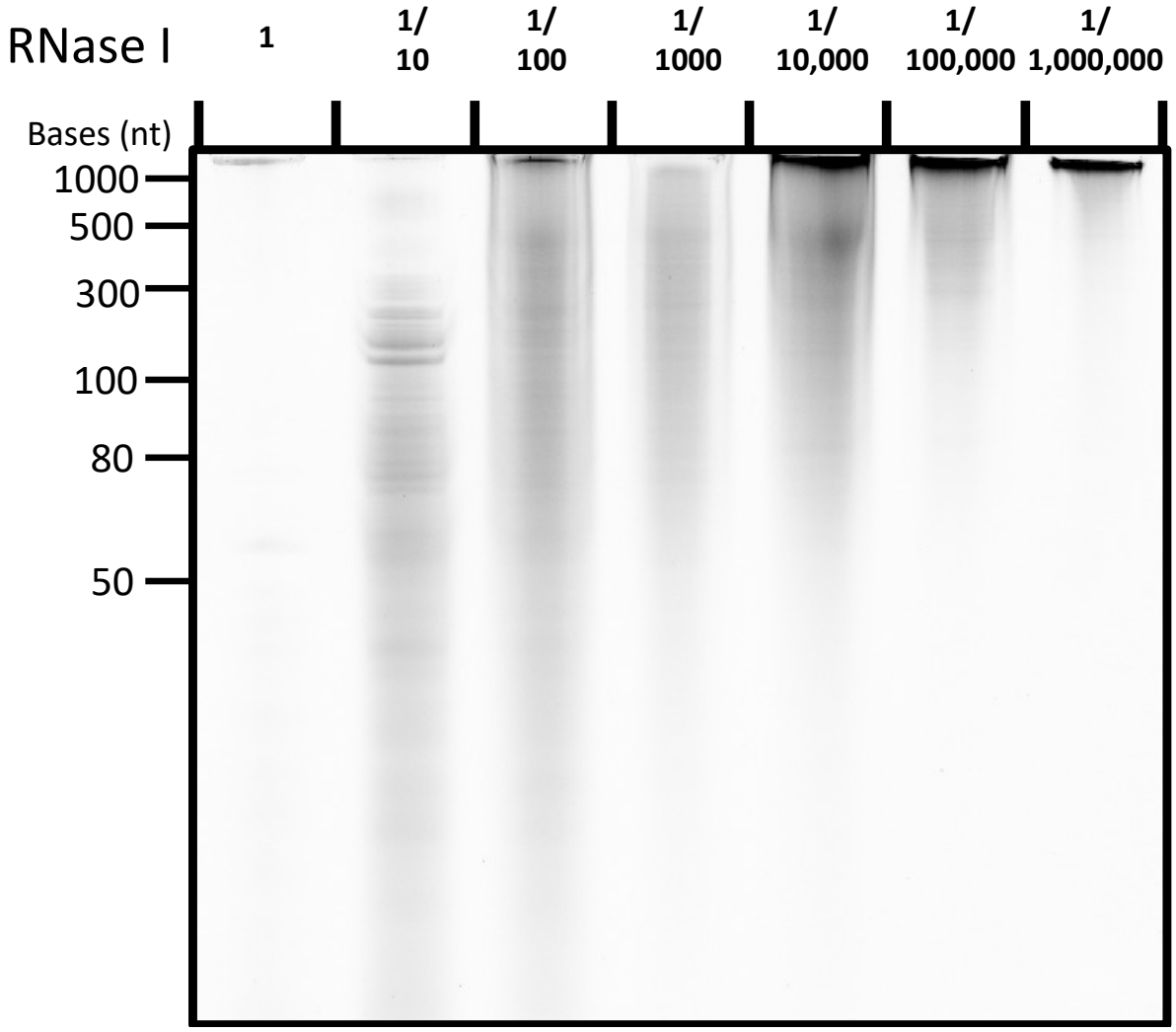


Figure 2.7 Optimizing RNase I Digestion with HCV Genomic RNA: Determining the ideal concentration of RNase I to obtain optimal fragment sizes. HCV genomic RNA was incubated with dilutions of RNase I at 37ºC for 20 minutes and the resulting fragments were run on a 12% urea denaturing gel and stained using SybrGold and imaged.

2.4 Discussion

CLIP methods are an important technique used to understand RBPs and their impact on cellular mechanics. Many of these protocols rely on the formation of free radicals in nucleobases upon 254 nm light absorption to crosslink with nearby residues; however, this process has low photo-crosslinking capabilities of around >1% [5,6,8]. Knowing this, we investigated if the photo-crosslinking ability of the ncAA AzF could be used to enhance CLIP-based protocols. Secondly, we were also interested in the role HCV NS3h plays in binding and unwinding the HCV genome as NS3h has been theorized to influence HCV genomic RNA secondary structure [24]. This made NS3h the perfect candidate for site-specific AzF incorporation to determine the interactions between NS3h and the HCV genome.

We began by analyzing an NS3h crystal structure bound with RNA (PDB: 3OX8) and found five sites in close proximity to the bound nucleic acid strand (E493, E503, T416, Q580 and K372). To do this, we used Pymol to mutate each of the chosen sites to an AzF residue to visually inspect the proximity of AzF to the bound nucleic acid strand. Except for Q580, each of the chosen sites is conserved across many NS3h genotypes [25]. K372 and T416 are located in motifs IV and V, E493 repels the nucleic acids from the back of the binding cleft, and E503 is located near the W501 bookend. Though four sites are conserved, E493 is the only known residue to have a functional role in helicase unwinding. The mutagenesis of LexA and RecA with the ncAA acridonylalanine has shown that elements such as conservation, hydrophobicity, or accessibility are not great predictors of tolerability towards unnatural amino acid mutagenesis [26].

E503AzF and Q580AzF were the only constructs that were able to photo-crosslink a nucleic acid substrate using the AzF (**Figure 2.2 A**). These sites were flanking both the entrance and exit of the nucleic acid binding cleft. The reasoning behind this is that it would allow the

nitrene intermediate to undergo addition reactions with a double bond, an active hydrogen insertion in a C-H or N-H, or the dedihydroazepine to be in close proximity to react with a nucleophile. We think, T416AzF and K372AzF were incapable of having AzF in a close enough proximity to the bound DNA strand to react with. The low photo-crosslinking effiency seen for T416AzF and K372AzF could also be explained by the positioning of the AzF against the phosphate backbone. This would allow AzF to react with phosphodiester bonds and a potential ribose group. Previous work has found ribose does not to react with the dedihydroazepine at a measurable rate [22]. As expected, we do observe photo-crosslinking between WT NS3h and the DNA substrates as 302 nm light will still be able to photo-crosslink DNA to proteins [27]. Though we did observe a 5- and 2-fold increase in the relative intensity of photo-crosslinked Cy5 ssDNA substrate for E503AzF and Q580Azf compared to that of the WT. These results were again recapitulated when we tested WT NS3h photo-crosslinking with 254 nm light against the E503AzF and Q580AzF mutants. We found that similarly, both E503AzF and Q580AzF photo-crosslinked ssDNA 4.5- and 1.8-fold more compared to WT NS3h, respectively. These results were not as high as the 7-fold increase in photo-crosslinking efficiency observed for diazirenes when compared to conventional 254 nm light [6]. However, these results indicate that site-specifically incorporated AzF in an optimal position does have the ability to photo-crosslink with higher yields than conventional 254 nm light irradiation. While the photo-crosslinking ability of the E503AzF and Q580AzF mutants are directly related to their proximity to the nucleic acid, there may be more optimal locations on NS3h where AzF incorporation could potentially lead to higher photo-crosslinking capabilities.

While investigating into the selectivity of AzF nucleic acid photo-crosslinking, we discovered an article which ran reactions with the dedihydroazepine intermediate AzF forms after UV irradiation [22]. Through quenching studies of dedihydroazepine in HEPES buffered solution,

they evaluated the reactivity of common biological materials. They found that guanosine monophosphate, adenosine, and uridine all had very quick reactions and cytidine had no measurable reaction rate. This was concerning as we want AzF to be used for CLIP and any technical bias in nucleobase sequence specificity could skew the final binding results after sequencing [4]. We decided to test the ability of E503AzF and Q580AzF to photo-crosslink with a DNA substrate consisting of fully cytidines, thymines and a random assortment of nucleobases (**Figure 2.3 C**). We found that there was no difference in photo-crosslinking efficiency between each of these different substrates, although these results could be explained by AzF crosslinking to a ribose group or the phosphate backbone instead of the nucleobases.

We observed strange banding patterns above the expected bands for both E503AzF and Q580AzF in the photo-crosslinking gels (**Figure 2.3 A and C**) and wanted to further investigate what they could represent. We thought that due to the mutations being close to the surface of the protein, they could potentially be photo-crosslinking to other NS3h monomers. It has been shown that two NS3h molecules can bind to the same DNA/RNA strand and that surface residues mediate protein-protein interactions between subdomains 2 and 3 of adjacent molecules [28]. To test this hypothesis, we performed a strained-promoted azide-alkyne cycloaddition (SPAAC) (**Supplemental Figure 2.4 A)** with our WT and AzF mutants E503AzF and Q580AzF with Cy5 tagged dibenzocyclooctyne (DBCO) (**Supplemental Figure 2.4 B**). We discovered that both E503AzF and Q580AzF readily reacted with the solvent-accessible DBCO. Crystal structures of multiple bound NS3h on a single strand of nucleic acid revealed that E503AzF is positioned to react with a neighbouring NS3h lysine and asparagine and Q580AzF was in close proximity to an aspartic acid of the NS3h next to it, and could potentially form a covalent bond (**Supplemental Figure 2.4 Cand D**). This is not surprising as photo-crosslinking ncAAs have been primarily

focused on preserving protein-protein interactions [29–31]. The photo-crosslinking gels (**Figure 2.3 A and C**) show the majority of photo-crosslinking was to the DNA substrate, so we decided that it would have a negligible impact on the proceeding CLIP experiments.

We needed to ensure that the working AzF NS3h mutants maintained as similar binding properties as possible compared to the WT NS3h. As a negative control, the bookend tryptophan 501 was mutated to an alanine, causing a loss of stacking interactions between it and the nucleobase. This mutation has been found to inhibit unwinding and severely reduce nucleic acid binding ability [28,32,33]. We tested the two photo-crosslinking mutants first using a molecular beacon helicase assay to determine if the AzF substitution had any effect on unwinding capabilities. (**Figure 2.4 B**) Surprisingly, we found that E503AzF had an unwinding rate of 0.80 ± 0.08 nM/min, a 2-fold increase in unwinding rate compared to the WT with an unwinding rate of 0.36 ± 0.02 nM/min. While this was unexpected, mutations of conserved residues such as E493 to a lysine or glutamine have shown both increased ATPase activity and unwinding activity [34]. As expected for W501A, there was no detectable unwinding rate upon the addition of ATP. We observed a lowered unwinding rate for Q580AzF of 0.12 ± 0.02 nM/min. Seeing these results we then sought to investigate and test the affinity of NS3h mutants for nucleic acids with fluorescence polarization.

The amount of WT NS3h needed to bind half the Cy3-15mer ($K_{0.5}$) was 11 ± 1.6 nM, which is in good accordance with what has been previously published for NS3h fluorescence polarization experiments [35]. Our best photo-crosslinking mutant, E503AzF, also had very similar concentration-dependant curves with $K_{0.5}$ of 15 ± 1.8 nM, which is in close range to that of the WT. Mutant Q580AzF had a $K_{0.5}$ of 26 ± 2.4 nM which was expected as we observed a reduced unwinding rate for this mutant. We think that the AzF moiety at the entrance could be interfering

with nucleic acid entry. This loss of nucleic acid binding affinity may have played a part in the 2-fold photo-crosslinking efficiency observed compared to the 5-fold increase for E503AzF (**Figure 2.3 B**). The reduced binding mutant had a $K_{0.5}$ of 63 ± 5.3 nM and plateaued at a significantly higher concentration of 125 nM (**Figure 2.4 B**). Taken together, E503AzF, with a 5-fold increase in photo-crosslinking capabilities and a similar affinity for nucleic acid substrate would be used for the next steps in the CLIP and identification of NS3h binding locations on the HCV genomic RNA.

To begin our CLIP methods we wanted to add NS3h directly into folded IVT genomic HCV RNA. After performing a photo-crosslinking gel of NS3h to a DNA substrate in different HCV folding buffers previously used to investigate secondary structure, we found that NS3h did not photo-crosslink in either buffer (**Supplemental Figure 2.3 A and B**). We realized that the increased ionic strength in both buffers was causing the loss of photo-crosslinking. NS3h is highly sensitive to ionic strength so much so that NaCl inhibits unwinding with an Ic_{50} of 15 mM and a doubling in NaCl concentration leads to K_d being increased 4 times [36,37]. To test this, we incrementally increased the concentration of NaCl in our binding buffer and observed a loss in photo-crosslinking (**Supplemental Figure 2.3 C**). To resolve this problem, we decided to simply fold the HCV RNA genome in Folding Buffer A, then photo-crosslink NS3h with the folded IVT HCV genome in a modified binding buffer containing an identical $MgCl_2$ concentration which is more crucial for RNA secondary and tertiary structure formation [38–40].

The next step we wished to optimize is the RNase I partial digestion. We chose RNase I because it was unbiased in nuclease sequence specificities and has previously been used in CLIP protocols [41–43]. Optimization of this step is crucial as it allows us to obtain proper fragmentation size for RNA which is crucial for the downstream cDNA synthesis and sequencing applications [44].

The results of our RNase partial degradation determined that a 1/10 RNase I dilution digested for 20 minutes at 37°C yielded fragments sizes of 50-100 bp (**Figure 2.6**). After partial digestion, performing a His-tag purification under denaturing conditions would work best in removing the RNase I and ensure that only RNA molecules that were covalently linked to protein were purified [45]. The final steps in our methods as seen in **Figure 2.5** are common in all CLIP-based methodologies. For our experiment, we chose to adapt a protocol that was created to locate binding interactions between brome mosaic virus capsid protein and the RNA genome [46]. They performed an SDS-PAGE and nitrocellulose membrane transfer to reduce the contamination of non-crosslinked RNA since the membrane has poor RNA binding capacity [43]. The nitrocellulose segments above the expected molecular weight of the protein containing the RNA-protein complex are extracted and the protein is digested with Proteinase K releasing the RNA crosslinked with a protein adduct. This RNA will be purified and prepared for downstream library preparation and sequencing protocols using polynucleotide kinase which will dephosphorylate the 3′ ends of the RNA and phosphorylate the 5′-OH.

2.5 Conclusion

Overall, with NS3h mutant E503AzF, we have provided proof that site-specific incorporating AzF into a position close to the nucleic acid increased photo-crosslinking capabilities when compared to that of the conventional 254 nm UV irradiation. Furthermore, we showed that the incorporated AzF in mutant E503AzF has little effect on the affinity and helicase functionality of NS3h. We hypothesize that this could be a useful tool for CLIP with low abundance RBPs or those which have a weaker RNA binding and thus a lower chance for photo-crosslinking. We also hope that the incorporation of AzF into RBPs could be implemented in future CLIP-based applications and studies. We are currently working on an *in vitro* CLIP method to determine NS3h binding locations on the highly structured HCV genomic RNA.

2.6 Materials and Methods

2.6.1 Amber Stop Codon Site-directed Mutagenesis

The coding sequence for HCV NS3h (genotype 1a) corresponding to amino acids 167−631 of NS3 with C- terminal His-tag was subjected to site-directed mutagenesis and amber stop TAG mutations were generated at positions: E503, Q580, E493, T416,and K372. This was performed in a reaction mixture of containing 20ng NS3h plasmid DNA, 1X xtremebuffer, 0.4 mM dNTPs, 0.3 uM amber stop mutagenesis primers (**Supplemental Table 2.2**), 0.02U/uL KOD polymerase and placed in the thermocycler 98ºC for 2 minutes, 93ºC for 30 seconds, 55ºC for 1 minute, 72ºC for 7 minutes. This full process was repeated a total of 18 times. Amber stop mutants were confirmed through Sanger sequencing performed at Genome Quebec (Montreal, QC, Canada).

2.6.2 Small Batch AzF Incorporation Screening

Small-scale 4mL LB cultures were inoculated with *E. coli* BL21 (DE3) cells transformed with NS3h plasmid as well as a plasmid encoding the orthogonal tRNA/tRNA synthetase pair, a gift from the lab of Dr. Peter Schultz. The small-scale cultures were grown to an OD of 0.6 and proteins were expressed with the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 1 mM AzF for a total of 4 hours at 20ºC. The cells were lysed using 1X Laemmli Buffer (Biorad #1610737EDU). The cell lysates were run on a 12% SDS-PAGE gel imaged using the Chemi Doc MP imager (Bio-rad) and transferred onto a PVDF membrane. This was followed by probing for the NS3h using the 6X-His Tag monoclonal Antibody (Thermo Scientific # MA1-21315-HRP).

2.6.3 Large Scale NS3h Purification

Competent *E. coli* BL21 (DE3) cells were transformed with NS3h plasmid as well as a plasmid encoding the orthogonal tRNA/tRNA synthetase pair, a gift from Dr. Peter Shultz (The Scripps Research Institute). Transformed E.coli were plated on spectinomycin/ampicillin plates and incubated overnight at 37ºC. A single colony was selected and grown for 4 hours in 2 ml of LB media with working concentrations of ampicillin (100 ug/mL) and spectinomycin (50 ug/mL). This culture was then used to inoculate 75 ml of LB with appropriate antibiotic selection and grown overnight. The media was spun down and used to inoculate 600ml of LB and grown until an optimal OD 0.6 was achieved. Protein expression was induced with the addition of 1 mM of IPTG and 1 mM AzF (MedChemExpress LLC #HY-16714) into the media. The proteins were expressed overnight at 18ºC. The NS3h proteins were purified from pelleted cells through a two-stage process.

First, the cells were resuspended with 1X binding buffer (50 nM Tris pH 8, 300 mM NaCl, and 10 mM imidazole pH 8) and sonicated on ice for 2 minutes at 50% amplitude. The lysate was then added into a HisTrap HP column (Sigma #GE17-5319-01) and washed with 1X wash buffer (50 nM Tris pH 8, 600 mM NaCl, and 60 mM imidazole pH 8) and eluted in a buffer consisting of 50 mM Tris (pH 8), 300 mM NaCl, and 250 mM imidazole. Next, size-exclusion chromatography was performed using an S75 column (GE Healthcare Life Sciences) in a filtered buffer containing 20 mM Tris (pH 8) and 150 mM NaCl. Fractions containing NS3h were pooled together and purified protein concentrations were determined by absorbance at 280 nm on a NanoDrop 2000 (Thermo Scientific) using the calculated extinction coefficient of 48,205 M−1 cm−1.[47] NS3h visualization was performed through western blotting using a six-His tag horseradish peroxidase-conjugated antibody.

2.6.4 Photo-crosslinking Gels and Analysis

Photo-crosslinking experiments were performed as previously shown in [5]. Briefly, 1 uM of purified NS3h protein (WT, E503AzF, Q580AzF, K372AzF, and T416AzF) was incubated with 250 nM of various fluorophore labelled DNA substrates (**Supplemental Table 2.1**) in a binding buffer consisting of 25 mM MOPS pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% Tween 20 unless otherwise indicated. Following a 45 minute incubation, the reactions were exposed to 1-10 minutes of UV light at 302 nm or 254 nm. The samples were run on a 12% SDS denaturing gel and imaged for Cy5 fluorescence using a ChemiDoc MP imaging system (Bio-Rad). Images were imported to ImageJ and densitometry was performed to determine the Cy5 intensity. Densitometry analysis was performed by normalizing the Cy5 intensity by the protein quantity in the stain free. The adjusted intensities were then compared relative to WT as a control.

2.6.5 Fluorescence Polarization

Fluorescence polarization binding assays were performed in a total volume of 100uL in 96 well, flat bottom, white plates. Purified NS3h and mutants were serially diluted using FP-buffer (25 mM MOPS pH 7.5, 1.25 mM MgCl2, 0.0025mg/ml BSA, 0.005% Tween 20) and combined with 5 nM Cy3 labelled DNA substrate (5'-Cy3- GGTAGTAATCCGCTC -3'). Fluorescence polarization was monitored with a SpectraMax i3 plate reader (Molecular Devices) by exciting at 535 nm and measuring parallel and perpendicular polarized light at 595 nm. Grating factors were calculated using measurements taken from wells with Cy3 labelled DNA alone. A total of three independent experiments were performed (n=3) and the data were fitted to a concentration-response equation (four-parameter, variable slope) using Graphpad- Prism.

2.6.6 Molecular Beacon Helicase assay

The helicase activity of NS3h was monitored using a molecular beacon helicase assay as described previously with the following changes [47]. Briefly, reactions were carried out at room temperature in 50uL reactions performed in triplicate, in 96-well black round bottom plates. 150 nM NS3h and 5 nM of molecular beacon helicase assay nucleic acid substrates were combined in a binding buffer containing 25 mM MOPS, pH 7.5, 1.25 mM MgCl2, 0.0025 mg/ml BSA, 0.005% Tween 20. Fluorescence was monitored using a SpectraMax i3 plate reader (Molecular Devices) every 30 seconds by exciting at 635 nm (9 nm slit width) and measuring the emission at 670 nm (15 nm slit width). After an initial 5 minute incubation, ATP was added to the reaction wells to a final concentration of 0.5 mM to initiate helicase activity. A total of three independent experiments were performed (n=3) and the data were fitted to (non-linear) plateau followed by one phase decay using Graphpad- Prism.

2.6.7 In Vitro Transcription of HCV

To get full length HCV genomic RNA, a restriction digest was performed for 3 hours with 5 ug of JFH-1 pDNA, 1X Cutsmart buffer and 20 U of Xbal (NEB #R0145S). Transcription and subsequent RNA purification was performed according to the MEGAscriptTM T7 transcription kit (ThermoFisher scientific #AM1334) kit protocol. The length and quality of the HCV IVT genomes were confirmed using 1% agarose gel electrophoresis.

2.6.8 RNase I digestion Tests

Purified IVT HCV genomic RNA was incubated with various dilutions of RNase I (ThermoFisher scientific #EN0601) ranging from 1U-0.001 U. The reaction was performed in 1X

modified binding buffer 25 mM MOPS pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, and 0.1% Tween 20. Reactions were incubated at 37ºC for 20 minutes and the resulting fragments were run on a 12% urea denaturing gel. An ssRNA low-range molecular ladder (NEB #N0364S) was run on the gel as well to determine fragmentation size. The gel was stained a 1X SYBR Gold (ThermoFisher scientific #S33102) and imaged using a ChemiDoc MP imaging system (Bio-Rad).

2.6.9 Protein labelling by SPAAC

1 uM NS3h WT and AzF mutants with 2 uM Cy5-DBCO were allowed to react at room temperature for 30 minutes. The reactions were combined with 1X Laemli SDS-PAGE loading buffer and heated at 95ºC for 10 minutes. The resulting labelling reactions were electrophoresed on a 12% SDS-PAGE gel and imaged for Cy5 using a ChemiDoc MP imaging system (Bio-Rad).

2.6.10 WT NS3h and E503AzF Photo-crosslinking and RNA Isolation

HCV genome folding was performed as previously described [23]. Briefly, purified full length JFH-1 RNA was heated to 90ºC for 10 seconds and placed on ice. The genome was then folded in a 1X folding buffer (50 mM HEPES pH 8.0, 5 mM $MgCl_2$, 200 mM potassium acetate) at a final concentration of 50ng/uL by heating at 65ºC for 5 minutes followed by slow cooling to room temperature. 10ug of full-length genomic RNA was incubated with 0.4 ug of E503AzF for 30 minutes in 1X modified binding buffer 25 mM MOPS pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, and 0.1% Tween 20 at a final volume of 200uL. This was irradiated with 302 nM light for 20 minutes. 0.1 U of RNAse I (ThermoFisher scientific #EN0601) was added to the photo-crosslinked RNA-NS3h and incubated at 37ºC for 20 minutes and then placed on ice.

Next, the RNA-NS3h complexes were isolated using Promega MagneHis™ (#V8500) following the manufacturer's protocol for denaturing conditions with some modifications. First,

the binding buffer was supplemented with urea to a final concentration of 8 M and NaCl to a final concentration of 500 mM before it was added directly into the RNA-NS3h mixture. After the addition of the denaturing binding buffer, 30 uL of MagneHis^{TM} Ni-particles are added to the mixture. The beads were placed on magnetic stands and were washed 3 times using MagneHis^{TM} Binding/Wash buffer. The NS3h-RNA complexes were eluted in 100 uL of Elution buffer. (4X) NuPAGE^{TM} LDS Sample Buffer (ThermoFisher scientific #NP0007) was added to the eluate and heated to 95ºC for 5 minutes. These samples were loaded onto a NuPAGE 4-12% Bis-Tris gel (ThermoFisher scientific #NP0326BOX) and run in 1X NuPAGE^{TM} MOPS SDS (ThermoFisher scientific #NP0001).

We further isolated the photo-crosslinked products as previously described in [46] with minor modifications. Briefly, the NS3h-RNA complexes were transferred onto a nitrocellulose membrane using a wet transfer apparatus at 100 V for 1 hour. The nitrocellulose was cut above the expected molecular weight of our protein using a fresh razor blade and transferred into a sterile 1.5 mL tube and rinsed with PBS. Proteinase K (ThermoFisher scientific #255530049), diluted to 1mg/ml in 1X PK buffer (50 mM Tris, 75 mM NaCl, 6.3 mM EDTA and 1% SDS), was added to the nitrocellulose fragments and digested for 1 hour at 37ºC with occasional mixing. The RNA fragments were purified using a Zymo RNA clean and concentrator (#R1019) following the manufacturer's protocol and eluted in 35 uL of elution buffer. To prepare the RNA for cDNA synthesis, 50U PNK (NEB #M0201L), 1X PNK buffer, and 10 mM ATP were added to the RNA and incubated at 37ºC for 30 minutes. The RNA was purified one final time using Zymo RNA clean and concentrator (#R1019) following the manufacturer's protocol.

2.7 References

1. Shetlar MD. Cross-Linking of Proteins to Nucleic Acids by Ultraviolet Light BT - Photochemical and Photobiological Reviews: Volume 5. In: Smith KC, ed. Springer US; 1980:105-197. doi:10.1007/978-1-4684-3641-9_4
2. Tretyakova NY, Groehler A, Ji S. DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Acc Chem Res*. 2015;48(6):1631-1644. doi:10.1021/acs.accounts.5b00056
3. Ramanathan M, Porter DF, Khavari PA, Alto P. Methods to study RNA-protein interactions. *Nat Methods*. 2019;16(3):225-234. doi:10.1038/s41592-019-0330-1.Methods
4. Hafner M, Katsantoni M, Köster T, et al. CLIP and complementary methods. *Nat Rev Methods Prim*. 2021;1(1). doi:10.1038/s43586-021-00018-1
5. Ahmed N, Foss D V., Powdrill MH, Pezacki JP. Site-Specific Cross-Linking of a p19 Viral Suppressor of RNA Silencing Protein and Its RNA Targets Using an Expanded Genetic Code. *Biochemistry*. 2019;58(33):3520-3526. doi:10.1021/acs.biochem.9b00428
6. Dziuba D, Hoffmann JE, Hentze MW, Schultz C. A Genetically Encoded Diazirine Analogue for RNA–Protein Photo-crosslinking. *ChemBioChem*. 2020;21(1-2):88-93. doi:10.1002/cbic.201900559
7. Ramanathan M, Porter DF, Khavari PA. Methods to study RNA–protein interactions. *Nat Methods*. doi:10.1038/s41592-019-0330-1
8. Spitzer J, Hafner M, Landthaler M, et al. PAR-CLIP (Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation): a Step-By-Step Protocol to the Transcriptome-Wide Identification of Binding Sites of RNA-Binding Proteins. *Methods Enzym*. Published online 2014. doi:10.1016/B978-0-12-420120-0.00008-6
9. Lee HS, Dimla RD, Schultz PG. Protein–DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. *Bioorg Med Chem Lett*. 2009;19(17):5222-5224. doi:10.1016/J.BMCL.2009.07.011
10. Romero-Lopez C, Barroso-DelJesus A, Garcia-Sacristán A, Briones C, Berzal-Herranz A. End-to-end crosstalk within the hepatitis C virus genome mediates the conformational switch of the 30X-tail region. *Nucleic Acids Res*. 2014;42(1):567-582. doi:10.1093/nar/gkt841
11. Pirakitikulr N, Kohlway A, Lindenbach BD, Pyle AM. The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures. *Mol Cell*. 2016;62(1):111-120. doi:10.1016/j.molcel.2016.01.024
12. Wan H, Adams RL, Lindenbach BD, Pyle AM. The In Vivo and In Vitro Architecture of the Hepatitis C Virus RNA Genome Uncovers Functional RNA Secondary and Tertiary Structures . *J Virol*. 2022;96(8). doi:10.1128/jvi.01946-21
13. Ali N, Wang C, Siddiqui A. Translation of hepatitis C virus genome. *Princess Tak Symp*. 1995;25:99-110.
14. Stewart H, Bingham RJ, White SJ, et al. Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Sci Rep*. 2016;6. doi:10.1038/srep22952
15. Kolykhalov AA, Mihalik K, Feinstone SM, Rice CM. Hepatitis C Virus-Encoded Enzymatic Activities and Conserved RNA Elements in the 3′ Nontranslated Region Are Essential for Virus Replication In Vivo. *J Virol*. 2000;74(4):2046-2051. doi:10.1128/jvi.74.4.2046-2051.2000

16. Yanagi M, St. Claire M, Emerson SU, Purcell RH, Bukh J. In vivo analysis of the 3′ untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. *Proc Natl Acad Sci U S A*. 1999;96(5):2291-2295. doi:10.1073/pnas.96.5.2291
17. Banerjee R, Dasgupta A. Specific Interaction of Hepatitis C Virus Protease/Helicase NS3 with the 3′-Terminal Sequences of Viral Positive- and Negative-Strand RNA. *J Virol*. 2001;75(4):1708-1721. doi:10.1128/jvi.75.4.1708-1721.2001
18. Nishikawa S, Nishikawa F, Fukuda K. In vitro selection of RNA aptamers against HCV-NS3 helicase and their structural similarity with 3'(+)UTR of HCV. *Nucleic Acids Res Suppl*. 2003;(3):241-242. doi:10.1093/nass/3.1.241
19. Huang Z-S, Wang C-C, Wu H-N. HCV NS3 protein helicase domain assists RNA structure conversion. *FEBS Lett*. 2010;584(11):2356-2362. doi:https://doi.org/10.1016/j.febslet.2010.04.020
20. Lindenbach BD, Rice CM. Unravelling hepatitis C virus replication from genome to function. *Nature*. 2005;436(7053):933-938. doi:10.1038/nature04077
21. Lohmann V. Hepatitis C Virus RNA Replication. In: Bartenschlager R, ed. *369*. Springer Berlin Heidelberg; 2013:167-198. doi:10.1007/978-3-642-27340-7_7
22. Rizk MS, Shi X, Platz MS. Lifetimes and reactivities of some 1,2-didehydroazepines commonly used in photoaffinity labeling experiments in aqueous solutions. *Biochemistry*. 2006;45(2):543-551. doi:10.1021/bi0516632
23. Mauger DM, Golden M, Yamane D, et al. Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc Natl Acad Sci U S A*. 2015;112(12):3692-3697. doi:10.1073/pnas.1416266112
24. Adams RL, Pirakitikulr N, Pyle AM. Functional RNA structures throughout the Hepatitis C Virus genome. *Curr Opin Virol*. 2017;(24):79-86. doi:10.1016/j.coviro.2017.04.007
25. Frick DN. HCV Helicase: Structure, Function, and Inhibition. *Hepat C Viruses Genomes Mol Biol*. Published online 2006:207-244. http://www.ncbi.nlm.nih.gov/pubmed/21250378
26. Hostetler ZM, Ferrie JJ, Bornstein MR, Sungwienwong I, Petersson EJ, Kohli RM. Systematic Evaluation of Soluble Protein Expression Using a Fluorescent Unnatural Amino Acid Reveals No Reliable Predictors of Tolerability. *ACS Chem Biol*. 2018;13(10):2855-2861. doi:10.1021/acschembio.8b00696
27. Müller S. *Nucleic Acids from A to Z*. Wiley-VCH; 2008. http://dnb.d-nb.de
28. Mackintosh SG, Lu JZ, Jordan JB, et al. Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus. *J Biol Chem*. 2006;281(6):3528-3535. doi:10.1074/jbc.M512100200
29. Young TS, Schultz PG. Beyond the canonical 20 amino acids: Expanding the genetic lexicon. *J Biol Chem*. 2010;285(15):11039-11044. doi:10.1074/jbc.R109.091306
30. Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem*. 2010;79:413-444. doi:10.1146/annurev.biochem.052308.105824
31. Preston GW, Wilson AJ. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem Soc Rev*. 2013;42(8):3289-3301. doi:10.1039/c3cs35459h
32. Preugschat F, Danger DP, Carter LH, Davis RG, Porter DJT. Kinetic analysis of the effects of mutagenesis of W501 and V432 of the hepatitis C virus NS3 helicase domain on ATPase and strand-separating activity. *Biochemistry*. 2000;39(17):5174-5183.

doi:10.1021/bi9923860
33. Lin C, Kim JL. Structure-Based Mutagenesis Study of Hepatitis C Virus NS3 Helicase. *J Virol*. 1999;73(10):8798-8807. doi:10.1128/jvi.73.10.8798-8807.1999
34. Frick DN, Rypma RS, Lam AMI, Frenz CM. Electrostatic analysis of the hepatitis C virus NS3 helicase reveals both active and allosteric site locations. *Nucleic Acids Res*. 2004;32(18):5519-5528. doi:10.1093/nar/gkh891
35. Mukherjee S, Hanson AM, Shadrick WR, et al. Identification and analysis of hepatitis C virus NS3 helicase inhibitors using nucleic acid binding assays. *Nucleic Acids Res*. 2012;40(17):8607-8621. doi:10.1093/nar/gks623
36. Levin MK, Patel SS. Helicase from hepatitis C virus, energetics of DNA binding. *J Biol Chem*. 2002;277(33):29377-29385. doi:10.1074/jbc.M112315200
37. Belon CA, Frick DN. Fuel specificity of the hepatitis C virus NS3 helicase. *Mol Cell Biochem*. 2012;23(1):1-7. doi:10.1016/j.jmb.2009.03.059.Fuel
38. Desirae Leipply, Draper DE. The dependence of RNA tertiary structure stability on Mg2+ concentration: interpretation of the Hill equation and coefficient. *Biochemistry*. 2010;49(9):1843-1853. doi:10.1021/bi902036j.The
39. Wu M, Tinoco I. RNA folding causes secondary structure rearrangement. *Proc Natl Acad Sci U S A*. 1998;95(20):11555-11560. doi:10.1073/pnas.95.20.11555
40. Fischer NM, Polěto MD, Steuer J, Van Der Spoel D. Influence of Na+ and Mg2+ ions on RNA structures studied with molecular dynamics simulations. *Nucleic Acids Res*. 2018;46(10):4872-4882. doi:10.1093/nar/gky221
41. Wang Z, Tollervey J, Briese M, Turner D, Ule J. CLIP: Construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. *Methods*. 2009;48(3):287-293. doi:10.1016/J.YMETH.2009.02.021
42. Grünberg S, Coxam B, Chen TH, et al. E. coli RNase i exhibits a strong Ca2+-dependent inherent double-stranded RNase activity. *Nucleic Acids Res*. 2021;49(9):5265-5277. doi:10.1093/nar/gkab284
43. Lee FCY, Ule J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol Cell*. 2018;69(3):354-369. doi:10.1016/j.molcel.2018.01.005
44. Garzia A, Meyer C, Morozov P, Sajek M, Tuschl T. Optimization of PAR-CLIP for transcriptome-wide identification of binding sites of RNA-binding proteins. *Methods*. 2017;118-119:24-40. doi:10.1016/j.ymeth.2016.10.007
45. Granneman S, Kudla G, Petfalski E, Tollervey D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A*. 2009;106(24):9613-9618. doi:10.1073/pnas.0901997106
46. Cheng Kao C, Chuang E, Ford J, Huang J, Podicheti R, Rusch DB. Mapping RNA Sequences that Contact Viral Capsid Proteins in Virions. 2017;(14). doi:10.21769/BioProtoc.2398
47. Ablenas CJ, Gidi Y, Powdrill MH, et al. Hepatitis C Virus Helicase Binding Activity Monitored through Site-Specific Labeling Using an Expanded Genetic Code. *ACS Infect Dis*. 2022;5:2118-2126. doi:10.1021/acsinfecdis.9b00220

Chapter 3: Photo-crosslinking of PspCas13b to Bound crRNA to Increase RNA Knockdown Longevity

3.1 Abstract

PspCas13b is an RNA-targeting CRISPR effector, which has been previously shown to knockdown viral RNA and mRNA through designable crRNAs. Here we incorporated the photo-crosslinking ncAA AzF into PspCas13b to form a stable covalent bond to irreversibly bind the crRNA. We wanted to test if the photo-crosslinked crRNA-PspCas13b complex could produce significant changes in knockdown longevity of viral RNA and mRNA targets. We were able to design crRNAs that produced an 88 ± 2% knockdown when targeting the mRNA of a luciferase renilla reporter system. When using PspCas13b to target an HCV subgenomic replicon, we did not observe any relevant knockdown. Out of six sites chosen for AzF photo-crosslinking, only I440AzF and F513AzF produced a 2.6- and 2.4-fold increase. However, there was photo-crosslinking between the crRNA and the WT PspCas13b which was unexpected and could be detrimental to our future applications.

3.2 Introduction

PspCas13b has shown to be the most resourceful Cas13b in mammalian cells due to its robust knockdown efficiency, lack of protospacer flanking sequence constraints, and low off-target effects [1–3]. Due to this flexibility, PspCas13b has become a powerful platform for RNA targeting and has become widely used for mammalian RNA interference and manipulation. A catalytically dead PspCas13b fused with an adenosine deaminase has allowed for select adenosine to inosine conversions on disease-relevant mRNA transcripts [4]. Cas13b RNA targeting capabilities are now being designed to destroy single-stranded genomic viral RNAs as an antiviral therapeutic strategy. PspCas13b has been programmed to target and knockdown SARS-COV-2 spike and nucleocapsid mRNA transcripts as well as suppress viral replication in mammalian cells while avoiding viral mutational escape [5]. It has also been used to target influenza A virus (IAV), causing a 22-fold reduction in viral RNA, and vesicular stomatitis virus (VSV) causing a 43.3-fold decrease in secreted VSV viral RNA [6].

PspCas13b has been thought of as a promising therapeutic technology with antiviral capacities [6–9]. However, increasing the efficiency and longevity of PspCas13b knockdown *in vivo* is essential for this to occur. One aspect which could be fine-tuned is the stability of Cas13b crRNAs in mammalian cells since they are rapidly degraded by endogenous nucleases and only have a half-life of seven hours [10]. It has been shown that the chemical modification of 2′-O-methylation on the 3′ end of Cas13b crRNAs increased their stability while also improving RNA targeting efficiency and half-life in human cells as well as extending targeting activity from 2 to 4 days [11]. Taking these results into account, we wanted to protect the crRNA from degradation by irreversibly binding it to their PspCas13b counterpart, potentially enhancing the knockdown longevity. To do this, we incorporated the photo-crosslinking ncAA AzF into PspCas13b targeting

the DR section of the crRNA as it is not directly involved in hybridizing with target RNAs and forms extensive interactions with the protein. We then wished to use this photo-crosslinked PspCas13b-crRNA complex to assess the RNA knockdown efficiency and longevity in mammalian cells. As functional crRNAs were needed to asses if knockdown could still occur for photo-crosslinked PspCas13b-crRNA complexes, we designed crRNAs targeting luciferase mRNA as well as a tricistronic HCV subgenomic replicon in hepatocellular carcinoma cells (HuH7-SGR-Luc). This HCV replicon produces all the non-structural proteins required for subgenomic replication (NS3-3′UTR) as well as a luciferase enzyme to quantify replicon levels (**Supplemental Figure 3.1**). We designed crRNAs targeting the NS3 gene of the subgenomic replicon hoping to see decreases in replication through a loss of luciferase signal. This replicon system has been widely used to monitor HCV replication [12–15].

In the current work, we describe the incorporation of the photo-crosslinking ncAA AzF into PspCas13b and screening for photo-crosslinking mutants. We also created crRNAs to target a luciferase renilla reporter system in HEK293Ts to monitor the knockdown efficiency of each crRNA. To test if we could obtain robust knockdown against viral RNA, we then designed crRNAs focusing on a tricistronic HCV subgenomic replicon.

3.3 Results

3.3.1 Incorporation of AzF into Selected Sites in PspCas13b

Since the crystal structure of PspCas13b has not yet been resolved, we needed to use Cas13b orthologs *Bergeylla zoohelcum* (BzCas13b) and *Prevotella buccea* (PbuCas13b) crystal structures to find sites for AzF incorporation into PspCas13b [16,17]. We first performed a multiple sequence alignment looking for areas of high conservation using 15 Cas13b orthologs (**Supplemental Table 3.1)**. Next, the crystal structure of PbuCas13b and BzCas13b were used to determine if these locations of high conservation were in close proximity to bound crRNA. Based on these findings, six residues were chosen from PbuCas13b and BzCas13b for their close proximity to the DR of the crRNA. The corresponding PspCas13b residues I440, N358, R477, L1048, F513 and Y626 were chosen for AzF incorporation. The results of this analysis can be seen in **Figure 3.1** and **Supplemental Figure 3.2**. We can see that for each of the six chosen sites, the tyrosine residue is highly conserved and thus we hoped that the incorporation of the AzF would not negatively affect binding, except for I440 which demonstrates histidine as the main conserved residue among the orthologs.

We generated six PspCas13b constructs with the amber stop codon at each of the chosen sites. To determine if AzF was being incorporated, the six constructs were expressed in *E.coli* with and without AzF supplemented in the media. The culture was lysed and electrophoresed on an SDS-PAGE gel and imaged for equal loading. Full-length PspCas13b was assessed by western blot (**Figure 3.1 B**). With the exception of L1048AzF, each PspCas13b AzF mutant was expressed when AzF was included in the growth media. We see little to no expression of the mutants in the absence of AzF in the growth media. We also observed a reduced expression of Y626AzF in comparison to the other mutants.

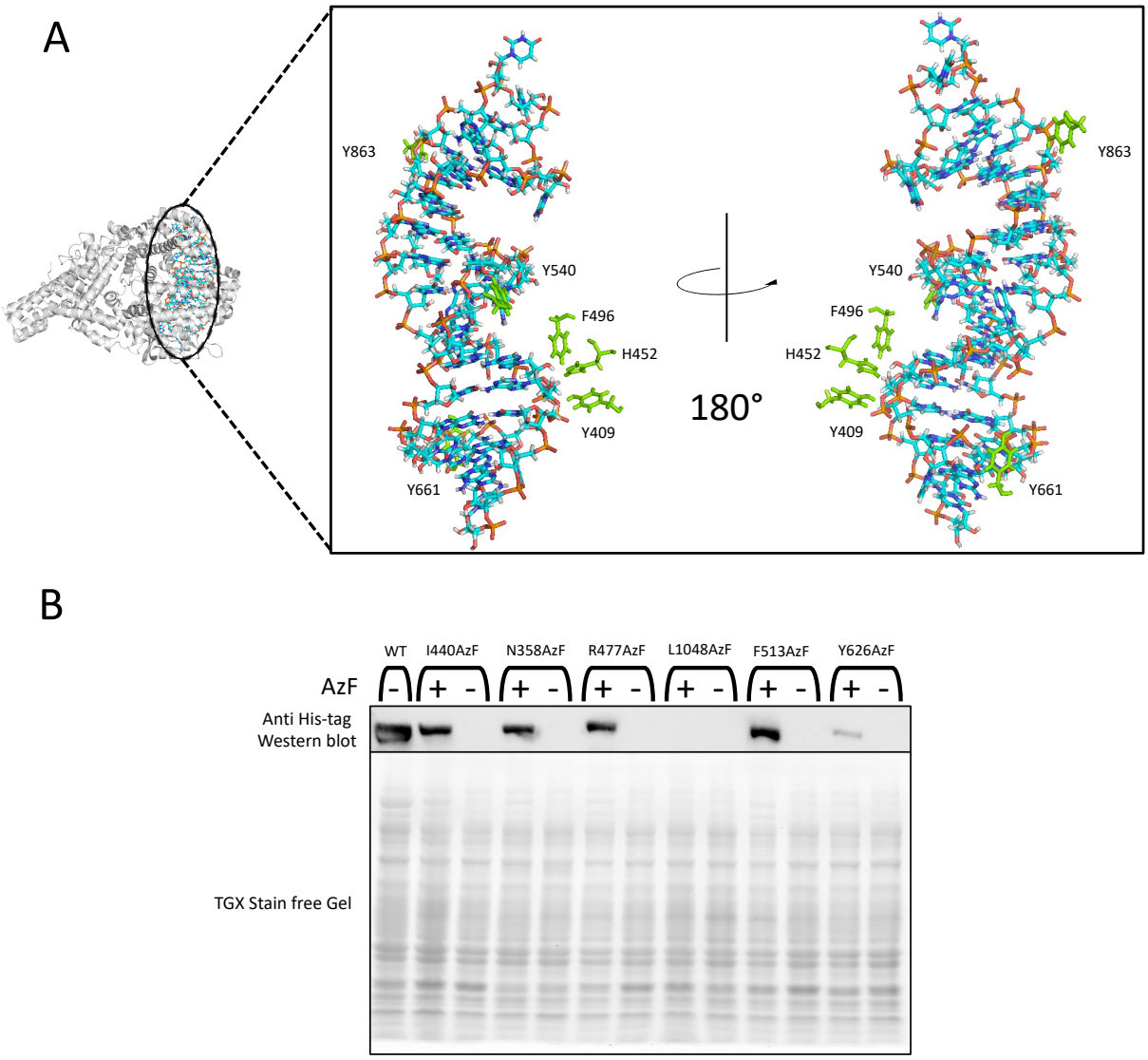


Figure 3.1 Site-specific Incorporation of AzF into PspCas13b: (A) The six sites chosen for AzF incorporation are modelled onto the PbuCas13b crystal bound to DR (PDB: 6DTD) structure as a reference to where they are expected to be. The chosen sites are highlighted in green, the bound crRNA DR is shown in blue. The full structure is shown in grey. (B) A western blot of the expression of all PspCas13b amber stop mutants. Mutants were expressed with 1 mM IPTG in the presence or absence of 1 mM AzF in the growth media. Cells were harvested 4 hours post-induction.

3.3.2 Photo-crosslinking of PspCas13b AzF Mutants to crRNAs

Next, we wanted to determine if the incorporated AzF was in close proximity to an element of the nucleic acid whereupon UV irradiation, could form a stable crRNA AzF covalent bond. To do this, we first purified WT PspCas13b and the five successful expressing AzF mutants by His-tag protein purification. To assess the photo-crosslinking capabilities of all the mutants, WT and mutant PspCas13b were incubated with tagged fluorescein crRNA. To induce photo-crosslinking post-incubation, samples were irradiated with 302 nm light and band shifting was examined using denaturing SDS-PAGE gels (**Figure 3.2 A**). First, we observed a shift in mass from 140kDa in the stain free gel to 160-170kDa for PspCas13b photo-crosslinked to labelled crRNA which can be observed in the overlayed image of the stain free and fluorescein scan. We did observe multiple bands in the fluorescein scan below the expected. The densitometry analysis showed the relative photo-crosslinking intensity of I440AzF and F513AzF compared to WT was 2.4- and 2-fold more, respectively. (**Figure 3.2 B**) Mutants N358AzF, R477AzF and Y626AzF have similar intensities to WT PspCas13b UV photo-crosslinking. We also noticed a noteworthy change in PspCas13b protein band intensities for UV irradiated PspCas13b samples compared to their non-UV counterparts. We noticed a large amount of non-AzF photo-crosslinking observed in the PspCas13b WT with the crRNA.

3.3.3 Testing crRNAs for Luciferase mRNA and HCV Subgenomic Replicon Knockdown

The next objective was to obtain working crRNAs in cell culture that would provide optimal knockdown efficiency. We designed five crRNAs (Luc 1-5) to target luciferase mRNA as well as a negative control non-targeting (NT) crRNA which will not target the luciferase mRNA. We transfected our designed crRNAs, WT PspCas13b and luciferase renilla reporter system in HEK293Ts. 24 hours post-transfection and luciferase and renilla were measured (**Figure 3.3A**).

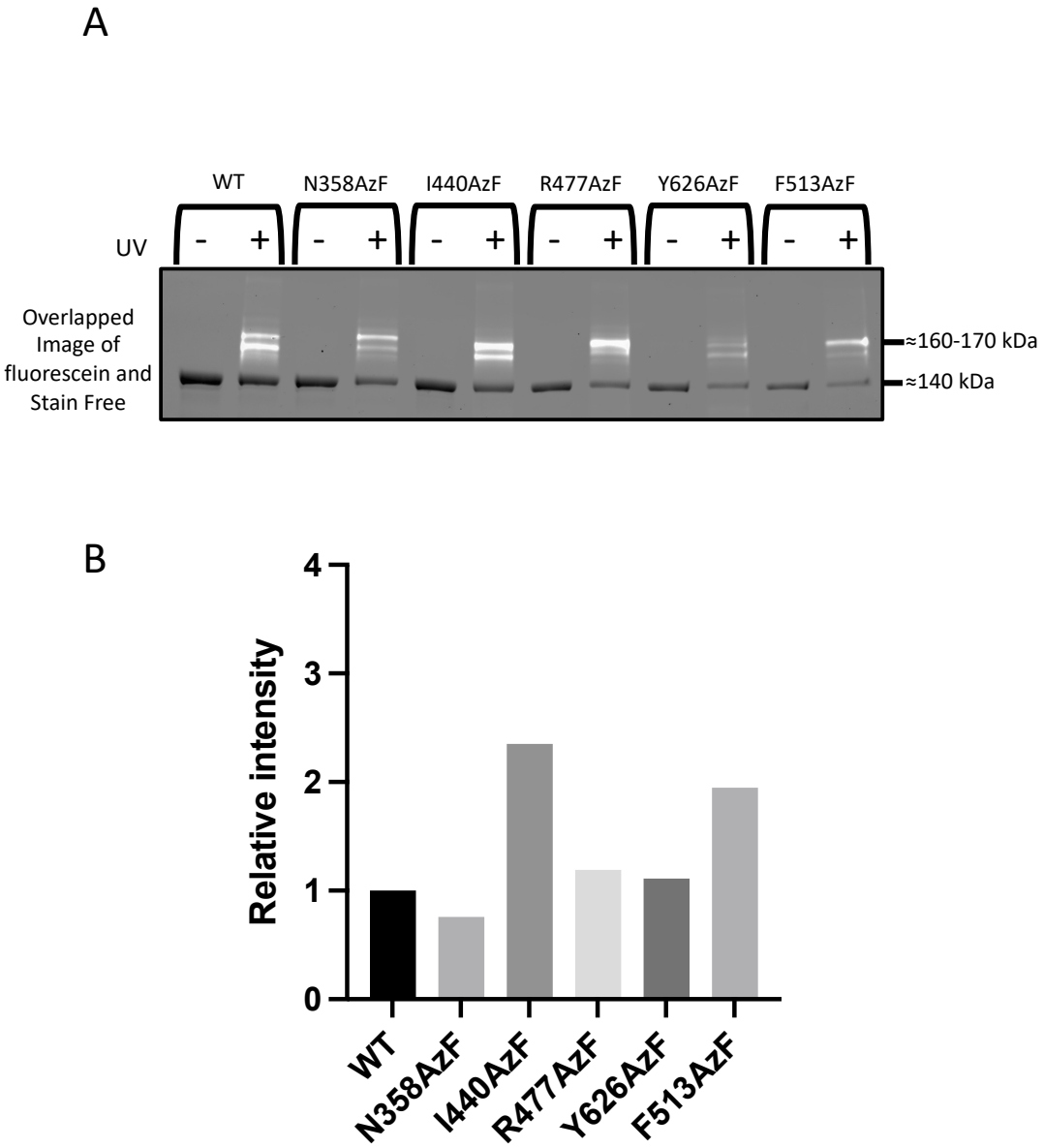


Figure 3.2 Assessing the Photo-crosslinking of AzF PspCas13b Mutants: (A) Photo-crosslinking SDS-PAGE gels for PspCas13b AzF mutants. Purified PspCas13b at 10 μM was incubated with 250 nm of crRNA substrate and irradiated with 302 nm light for 10 minutes. The gel is an overlapped image of the fluorescein scan and the stain free image. (B) Densitometry analysis was performed by normalizing the fluorescein intensity by the protein quantity of the non-UV controls. The adjusted intensities were then compared relative to WT as a control.

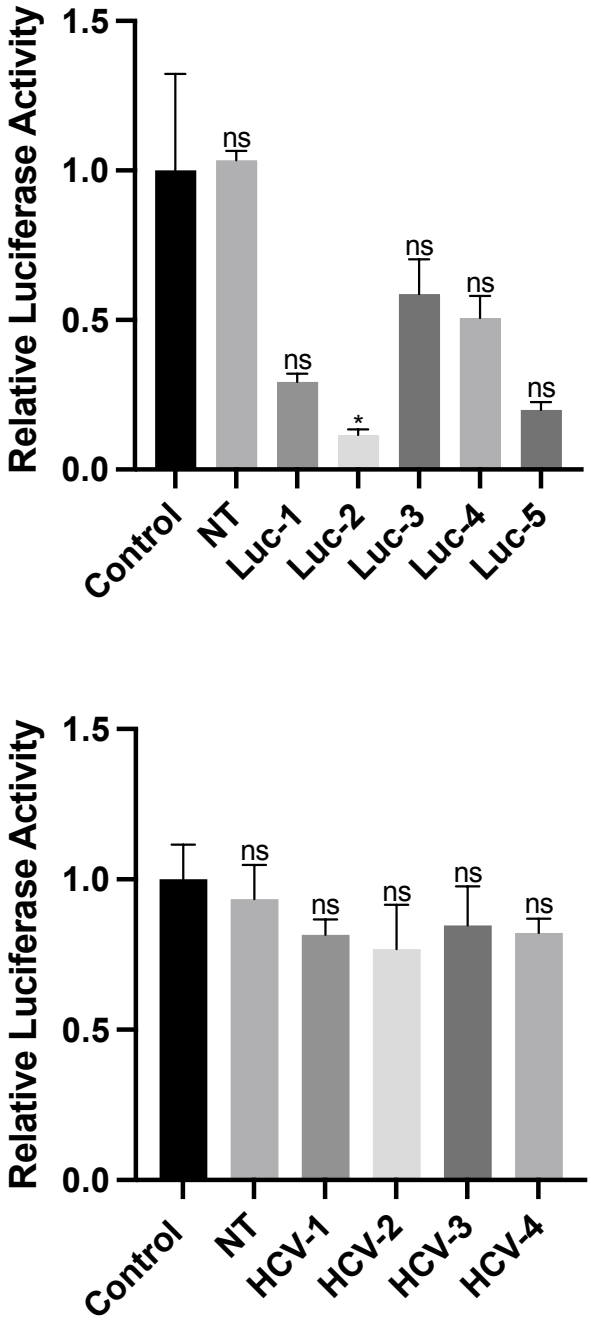


Figure 3.3 Using WT PspCas13b for RNA Knockdown in Mammalian Cells: (A) WT PspCas13b knocks down luciferase signal in HEK293Ts. For luciferase knockdown, HEK293Ts were transfected with cloned or empty crRNA plasmid, WT PspCas13b mammalian expression plasmid, and a luciferase renilla reporter system. Cells were lysed 24 hours post-transfection and analyzed on a luminometer. (n=3) (B) WT PspCas13b knockdown of HCV subgenomic replicon in HuH7-SGR-Luc. For replicon knockdown, HuH7-SGR-Luc were transfected with cloned crRNA plasmid (HCV 1-4 and NT) and WT PspCas13b mammalian expression plasmid. Cells were lysed 24 hours post-transfection and analyzed on a luminometer. (n=3) Significance was assessed using two-tailed, unpaired student's t-test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

We observed a decrease in luciferase signal for all designed crRNAs. The best performing crRNAs were Luc-2 and Luc-5 in which we observed a 88 ± 2% decrease and an 80 ± 3% decrease in luciferase signal, respectively. The worst-performing crRNA was Luc-3 with a modest decrease of 58 ± 10%. The non-targeting crRNA (NT) provided no significant changes in the knockdown of the luciferase signal. The control was an empty crRNA backbone vector, WT PspCas13b plasmid, and the luciferase renilla reporter.

We next set our sights on trying to target the HCV subgenomic replicon luciferase construct found in the HuH7-SGR-Luc cell line. We designed four crRNAs (HCV 1-4) to target the NS3 section of the subgenomic replicon as well as a negative control non-targeting crRNA. We transfected our target crRNAs with the WT PspCas13b into HuH7-SGR-Luc. We lysed 24 hours after to measure the luciferase signal and normalized it to the protein concentration (**Figure 3.3B**). We did not observe any significant decreases in luciferase signal for crRNAs HCV 1-4. We do see a decrease in luciferase signal for crRNA HCV-1 and HCV-4 which has a loss of over 18 ± 5% and 17 ± 5%, respectively.

3.4 Discussion

Knowing that Cas13b knockdown longevity is linked to the stability and half-life of the crRNA[11], we wanted to test if protecting crRNA through site-specific photo-crosslinking the DR section of crRNA to PspCas13b using AzF would lead to increased longevity of mRNA knockdown. As we did not have the crystal structure of PspCas13b, we had to work with the solved crystal structures of orthologs PbuCas13b and BzCas13b to locate sites for AzF incorporation. We still decided to work with PspCas13b as it has one of the highest RNA knockdown efficiencies and lacks a protospacer flanking sequence in mammalian cells out of many Cas13b orthologs [4]. To work around this limitation, we first performed a sequence alignment of 15 Cas13b orthologs and looked at areas of high conservation and located these on the PbuCas13b and BzCas13b crystal structures. We then chose sites that were both visually near the crRNA and showed they could tolerate the substitution of a tyrosine moiety (**Supplemental Figure 3.2**). We found six sites for AzF incorporation, namely I440, N358, R477, L1048, F513 and Y626, that were in areas of high conservation located in the binding pocket near the DR. Mutations N358, R477 and I440 are all located in the lid domain which caps the 3′ of the DR crRNA. Y626, I847, and F513 are all found in helical domains 1 and 2 and are important for bound crRNA coordination. The F513 is close to a flipped nucleobase and the I440, N358, R477, L1048, F513 and Y626 are all directed toward the phosphate backbone (**Figure 3.1A**).

The photo-crosslinking gel showed PspCas13b-crRNA fluorescein bands for all mutants and WT (**Figure 3.2A**). When we normalized the intensity of the fluorescein to the no UV control, we were able to see that there was a 2.6- and 2-fold increase in photo-crosslinked crRNA for both I440AzF and F513AzF, respectively (**Figure 3.2B**). However, we still witnessed a large amount of photo-crosslinking in the WT PspCas13b which was not what we had expected. In the previous

chapter, we only observed small amounts of non-AzF photo-crosslinking with WT NS3h. We had not fully considered how the extensive interactions between the larger bound crRNA and the free spacer section of the crRNA would affect photo-crosslinking. It has been shown that nucleobases located in the DR of crRNAs bound in both PbuCas13b and BzCas13b have direct interactions with many charged residues [16,18]. The nucleobases in the spacer section of the crRNA bound in BzCas13b also form extensive interactions with multiple domains of the protein [18].

We did not test to see if the photo-crosslinking had any detrimental effect on WT PspCas13b crRNA binding and RNA nuclease activity. Although we speculate that the photo-crosslinking to the bound crRNA would most likely be detrimental to the hybridization between the crRNA and target RNAs as well as the subsequent cleavage. If the covalent bond formation to PspCas13b were to occur between a nucleic acid in the 30 bp spacer section of the crRNA, it could reduce or eliminate activity as WT PspCas13b is intolerant to single and double mismatches in the crRNA spacer and target RNA [4].

We think further optimization and investigation into AzF as a method to photo-crosslink crRNAs bound by WT PspCas13b is needed so we can reduce this amount of photo-crosslinking. First, we would need to find an optimal UV irradiation time to determine maximized levels of AzF photo-crosslinking while lowering the photo-crosslinking seen in WT PspCas13b. We could also employ the use of UV bandpass filters while irradiating samples to reduce the amount of high-energy light which could potentially be helpful. If AzF is not optimal for this application, we could also look to using the ncAA photo-crosslinking family of diazines, which uses 350 nm light for excitation which could be used as an alternative to AzF [19]. We think that at this wavelength it would reduce potential damages to WT PspCas13b and lower the photo-crosslinking as nucleic acids are less likely to absorb 365 nm UV radiation [20,21].

We did design working crRNAs which provided a robust luciferase knockdown (**Figure 3.3 A**). We obtained an 88 ± 2% decrease of luciferase signal with crRNA Luc-2 which mimics the average 92% luciferase knockdown using PspCas13b previously reported [4]. We do see a small decrease in luciferase activity when targeting the HCV subgenomic replicon with crRNAs HCV-1 and HCV-4 but only an 18% and 17 ± 5% decrease, respectively (**Figure 3.3B**). Comparing our replicon knockdown to the robust 88% knockdown achieved with our luciferase renilla reporter and previous work with PspCas13b effectively knocking down influenza A virus and vesicular stomatitis virus, we don't think that we are getting knockdown of the replicon [6]. There could be RNA secondary structures and dsRNA elements blocking accessibility to the target chosen for hybridization which has been shown to reduce knockdown [22]. Additionally, the HCV RNA may not be accessible for PspCas13b to target to begin with. Non-structural proteins, as well as viral plus-strand RNA, have all been shown to be contained in the membranous web in Huh7 cells harboring the HCV subgenomic replicons. Under normal HCV infections, the membranous web is used to hide the RNA genome to limit host cell immune activation which we think may make it harder to target the viral RNA [23,24].

3.5 Conclusion

Ultimately, the results of this section show further optimization is needed for AzF photo-crosslinking crRNAs to PspCas13b as we obtained non-AzF photo-crosslinking in WT PspCas13b. We speculate that this may have disadvantageous effects on PspCas13b's ability to hybridize with target RNAs and subsequent nuclease activity. We were able to obtain a functional crRNA targeting luciferase mRNA in mammalian cells which led to an efficient knockdown of luciferase signal. However, when designed crRNAs targeting an HCV subgenomic replicon, we were unable to obtain working crRNAs that provided any knockdown. We think a new approach could be investigated if crRNAs and PspCas13b are to be covalently crosslinked to protect the crRNA from degradation.

3.6 Materials and Methods

3.6.1 Amber stop Codon Site-directed Mutagenesis

The coding sequence for PspCas13b found with N-Terminal His-tag (Addgene #115219) was subjected to site-directed mutagenesis and amber stop TAG mutations were generated at positions: I440, N358, R477, L1048, F513 and Y626. This was performed in a reaction mixture of containing 20ng PspCas13b bacterial expression plasmid DNA, 1X xtremebuffer, 0.4 mM dNTPs, 0.3 uM amber stop mutagenesis primers (**Supplemental Table 3.2**), 0.02 U/uL KOD polymerase and placed in a thermocycler at 98ºC for 2 minutes, 93ºC for 30 seconds, 55ºC for 1 minute, 72ºC for 7 minutes. This full process was repeated a total of 18 times. Amber stop mutants were confirmed through Sanger sequencing performed at Genome Quebec (Montreal, QC, Canada).

3.6.2 Small Batch PspCas13b AzF Incorporation Screening

Small-scale 4mL LB cultures were inoculated with *E. coli* BL21 (DE3) cells transformed with PspCas13b bacterial expression plasmid (Addgene #115219) as well as a plasmid encoding the orthogonal tRNA/tRNA synthetase pair, a gift from the lab of Dr. Peter Schultz. The small cultures were grown to an OD of 0.6 and proteins were expressed with the addition of 1 mM IPTG and 1 mM AzF (MedChemExpress LLC #HY-16714) for a total of 4 hours at 20ºC. The cells were lysed using 1X Laemmli Buffer (Biorad #1610737EDU). The cell lysate was electrophoresed on a 12% SDS-PAGE gel and imaged using the Chemi Doc MP imager (Bio-rad) before being transferred onto PVDF membranes for western blot analysis. This was followed by imaging using a primary PspCas13b antibody (Diagenode #C15200250) and secondary HRP goat anti-mouse (Jackson Immuno-research #115-035-062).

3.6.3 Large Scale PspCas13b Expression and Purification

Competent *E. coli* BL21 (DE3) cells were transformed with PspCas13b bacterial expression plasmid (Addgene #115219) as well as a plasmid encoding the orthogonal tRNA/tRNA synthetase pair, a gift from Dr. Peter Shultz (The Scripps Research Institute). Transformed *E.coli* were plated on spectinomycin/ampicillin plates and incubated overnight at 37ºC. A single colony was selected and grown for 4 hours in 2 ml of LB media with working concentrations of ampicillin (100µg/mL) and spectinomycin (50µg/mL). This culture was then used to inoculate 75 mL of LB with appropriate antibiotic selection and grown overnight. The media was spun down and used to inoculate 600 mL of LB and grown until an optimal OD 0.6 was achieved. Protein expression was induced with the addition of 1 mM of IPTG and 1 mM AzF (MedChemExpress LLC #HY-16714) to the media. The proteins expressed overnight at 18ºC. The PspCas13b proteins were purified from pelleted cells through a two-stage process.

First, the cells were resuspended with 1X binding buffer (50 nM Tris pH 8, 300 mM NaCl, and 10 mM imidazole pH 8) and sonicated on ice for 2 minutes at 50% amplitude. The lysate was then added into a HisTrap HP column (Sigma #GE17-5319-01) and washed with 1X wash buffer (50 nM Tris pH 8, 600 mM NaCl, and 60 mM imidazole pH 8) and eluted in a buffer consisting of 50 mM Tris (pH 8), 300 mM NaCl, and 250 mM imidazole. Eluted proteins were concentrated using Amicon Ultra-15 Centrifugal filter units (Sigma-Aldrich #UFC910024). Purified protein concentrations were determined by absorbance at 280 nm on a NanoDrop 2000 (Thermo Scientific). To confirm PspCas13b purification full-length protein was visualized through western blotting using a primary PspCas13b antibody (Diagenode #C15200250) and secondary HRP goat anti-mouse (Jackson Immuno-research #115-035-062).

3.6.4 Photo-crosslinking Gels and Analysis

Photo-crosslinking experiments were performed as previously shown in [25]. Briefly, purified PspCas13b protein (WT, N358AzF, I440AzF, R477AzF, F513AzF, and Y626AzF) at 1μM were incubated with 250 nM of fluorescein labelled crRNA substrate (5′ FAM AGUCGUG GCAGGAUAUCGAAUGUUCUUGUAGUUGUGGAAGGUCCAGUUUUGAGGGGCUAUU ACAAC 3′**)** in a buffer consisting of 25 mM Tris-HCl pH 7.5, 200 mM NaCl and 2 mM $MgCl_2$. Following a 30 minute incubation, the reactions were exposed to 10 minutes of light at 302 nm. The samples were electrophoresed on a 12% SDS denaturing gel and imaged for fluorescein fluorescence using a ChemiDoc MP imaging system (Bio-Rad). Images were imported to ImageJ and densitometry was performed to determine the fluorescein intensity. Densitometry analysis was performed by normalizing the fluorescein intensity by the protein quantity in the stain free. The adjusted intensities were then compared relative to WT PspCas13b as a control.

3.6.5 Cloning PspCas13b crRNAs into Expression Plasmid for Cell Culture Assays

The crRNA oligomers were phosphorylated and annealed in a reaction mixture containing 10 uM of forward and reverse primer (**Supplemental Table 3.3**) in 1X T4 ligation buffer and 10U of T4 Polynucleotide Kinase (NEB #M0201S). The oligomers were then placed in a thermocycler and ran at 37ºC for 30 min; 95ºC for 5 min before ramping down to 25ºC by lowering 5ºC every minute. The duplexed oligomers were then cloned into the PspCas13b crRNA backbone (Addgene #103854). The reaction mixture contained 1ng of plasmid crRNA plasmid, 10 uM duplexed oligomers, 1X Cutsmart Buffer (NEB), 500 uM ATP, 4U of BbsI (NEB# R0539S), and 60U of T4 ligase (NEB# M0202S) to a final volume of 20 uL in water. The cloning reactions were placed in a thermocycler and run as follows: 37ºC for 5 min and 21ºC for 5 min. This cycle was repeated six

times before heating to 60ºC for 5min in a heat block and cooling to 10ºC. Successful cloning of the crRNA backbone was confirmed through Sanger Sequencing performed at Genome Quebec (Montreal, QC, Canada).

3.6.6 Cell Culture

Hepatocellular carcinoma cells containing the HCV subgenomic replicon-luciferase construct (HuH7-SGR-Luc) were grown in DMEM media (ThermoFischer scientific # 11995065) supplemented with FBS (10%, Wisent Cat. # 080-450), 100 nM non-essential amino acids (ThermoFischer scientific #11140076) and 500 µg/mL G418-Sulfate antibiotic, Geneticin (ThermoFischer scientific. #11995065). HEK293Ts were grown in DMEM media (ThermoFischer scientific # 11995065) supplemented with FBS (10%, Wisent Cat. # 080-450). Cells were grown at 37ºC at 5% CO_2.

3.6.7 Luciferase mRNA Knockdown

For luciferase knockdown, HEK293Ts were seeded at 100,000 cells/well in a 24-well plate and transfected with 800 ng of crRNA PspCas13b expression plasmid (Addgene #103854), 400 ng PspCas13b mammalian expression plasmid (Addgene #103862) and 25 ng of psiCHECKTM-2 (Promega #C8021) using Lipofectamine 2000 (ThermoFischer scientific # 11668030) following the manufacturer's protocol. Cells were lysed 24 hours post-transfection in 150 µL of passive lysis buffer (Promega Cat.# E1941) and analyzed using a Lmax luminometer microplate reader (Molecular Devices). A luciferase solution (25 mM Glycylglycine, 15 mM K_2PO_4 (pH 8.0), 4 mM EGTA, 2 mM ATP, 1 mM DTT, 15 mM $MgSO_4$, 0.1 mM CoA, 75 µM luciferin) was added to the cell lysate and the signal was measured. Following this, a renilla solution (1.1 M NaCl, 2.2 mM

Na_2EDTA, 220 mM K_2PO_4 (pH 5.1), 0.44 g/L BSA, 1.3 mM NaN_3, 1.43 uM coelenterazine) was added to the wells to measure renilla signal. Relative chemiluminescence was achieved by normalizing the luciferase signal to that of the renilla. Three biological replicate were performed and significance was assessed using two-tailed, unpaired student's t-test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

3.6.8 HCV Subgenomic Replicon Knockdown

HuH7-SGR-Luc cells were seeded at 200,000 cells/well in a 12-well plate and transfected with 1.2 µg of cloned crRNA PspCas13b expression plasmid (Addgene #103854) and 600 ng PspCas13b mammalian expression plasmid (Addgene #103862) using Lipofectamine 2000 (ThermoFischer scientific #11668030) following the manufacturer's protocol. Cells were lysed 24 hours post-transfection in 300 µL of passive lysis buffer (Promega Cat.# E1941) and analyzed using a Lmax luminometer microplate reader (Molecular Devices). A luciferase solution (25 mM Glycylglycine, 15 mM K_2PO_4 (pH 8.0), 4mM EGTA, 2 mM ATP, 1 mM DTT, 15 mM $MgSO_4$, 0.1 mM CoA, 75 µM luciferin) was added cell lysate and the signal was measured. Protein concentrations were determined using a Pierce™ BCA Protein Assay Kit (ThermoFischer Scientific #23227). Relative chemiluminescence was achieved through normalizing luciferase signal to protein abundance. Three biological replicate were performed and significance was assessed using two-tailed, unpaired student's t-test, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

3.7 References

1. Tost J, Shukla A, Gaj T, et al. Programmable System of Cas13-Mediated RNA Modification and Its Biological and Biomedical Applications. *Front Cell Dev Biol*. 2021;9. doi:10.3389/fcell.2021.677587
2. Abudayyeh OO, Gootenberg JS, Essletzbichler P, et al. RNA targeting with CRISPR-Cas13. *Nature*. 2017;550(7675):280-284. doi:10.1038/nature24049
3. Ai Y, Liang D, Wilusz JE. CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. *Nucleic Acids Res*. 2022;50(11):e65-e65. doi:10.1093/nar/gkac159
4. Cox DBT, Gootenberg JS, Abudayyeh OO, et al. RNA editing with CRISPR-Cas13. *Science (80-)*. 2017;358(6366):1019-1027. doi:10.1126/science.aaq0180
5. Fareh M, Zhao W, Hu W, et al. Reprogrammed CRISPR-Cas13b suppresses SARS-CoV-2 replication and circumvents its mutational escape through mismatch tolerance. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-24577-9
6. Freije CA, Myhrvold C, Boehm CK, et al. Programmable Inhibition and Detection of RNA Viruses Using Cas13. *Mol Cell*. 2019;76(5):826-837.e11. doi:10.1016/j.molcel.2019.09.013
7. Palaz F, Kalkan AK, Can Ö, et al. CRISPR-Cas13 System as a Promising and Versatile Tool for Cancer Diagnosis, Therapy, and Research. *ACS Synth Biol*. 2021;10(6):1245-1267. doi:10.1021/acssynbio.1c00107
8. Baddeley HJE, Isalan M. The Application of CRISPR/Cas Systems for Antiviral Therapy. doi:10.3389/fgeed.2021.745559
9. Bagchi R, Tinker-Kulberg R, Salehin M, et al. Polyvalent guide RNAs for CRISPR antivirals. *Biophys J*. 2022;121(3):422a. doi:10.1016/j.bpj.2021.11.666
10. Saifullah, Sakari M, Suzuki T, Yano S, Tsukahara T. Effective rna knockdown using crispr-cas13a and molecular targeting of the eml4-alk transcript in h3122 lung cancer cells. *Int J Mol Sci*. 2020;21(23):1-17. doi:10.3390/ijms21238904
11. Méndez-Mancilla A, Wessels HH, Legut M, et al. Chemically modified guide RNAs enhance CRISPR-Cas13 knockdown in human cells. *Cell Chem Biol*. 2022;29(2):321-327.e4. doi:10.1016/j.chembiol.2021.07.011
12. Rakic B, Sagan SM, Noestheden M, et al. Peroxisome proliferator-activated receptor α antagonism inhibits hepatitis C virus replication. *Chem Biol*. 2006;13(1):23-30. doi:10.1016/j.chembiol.2005.10.006
13. Rakić B, Brûlotte M, Rouleau Y, Bélanger S, Pezacki JP. Bleomycin is a potent small-molecule inhibitor of hepatitis C virus replication. *ChemBioChem*. 2006;7(9):1330-1333. doi:10.1002/cbic.200600180
14. Supekova L, Supek F, Lee J, et al. Identification of human kinases involved in hepatitis C virus replication by small interference RNA library screening. *J Biol Chem*. 2008;283(1):29-36. doi:10.1074/jbc.M703988200
15. Pezacki JP, Sagan SM, Tonary AM, et al. Transcriptional profiling of the effects of 25-hydroxycholesterol on human hepatocyte metabolism and the antiviral state it conveys against the hepatitis C virus. *BMC Chem Biol*. 2009;9. doi:10.1186/1472-6769-9-2
16. Slaymaker IM, Mesa P, Kellner MJ, et al. High-Resolution Structure of Cas13b and Biochemical Characterization of RNA Targeting and Cleavage. *Cell Rep*. 2019;26(13):3741-3751.e5. doi:10.1016/j.celrep.2019.02.094

17. Zhang B, Ye W, Ye Y, et al. Structural insights into Cas13b-guided CRISPR RNA maturation and recognition. *Cell Res*. 2018;28(12):1198-1201. doi:10.1038/s41422-018-0109-4
18. Zhang B, Ye W, Ye Y, et al. Structural insights into Cas13b-guided CRISPR RNA maturation and recognition. *Cell Res*. doi:10.1038/s41422-018-0109-4
19. Dziuba D, Hoffmann JE, Hentze MW, Schultz C. A Genetically Encoded Diazirine Analogue for RNA–Protein Photo-crosslinking. *ChemBioChem*. 2020;21(1-2):88-93. doi:10.1002/cbic.201900559
20. Klak M, Gomółka M, Dobrzański T, et al. Irradiation with 365 nm and 405 nm wavelength shows differences in DNA damage of swine pancreatic islets. *PLoS One*. 2020;15(6 June):1-18. doi:10.1371/journal.pone.0235052
21. Mackinnon AL, Taunton J. Target Identification by Diazirine Photo-Cross-linking and Click Chemistry. *Curr Protoc Chem Biol*. 2009;1:55-73. doi:10.1002/9780470559277.ch090167
22. Bandaru S, Higashide Tsuji M, Shimizu Y, et al. Structure-based design of gRnA for Cas13. *Sci Rep*. 2020;10:11610. doi:10.1038/s41598-020-68459-4
23. Gosert R, Egger D, Lohmann V, et al. Identification of the Hepatitis C Virus RNA Replication Complex in Huh-7 Cells Harboring Subgenomic Replicons. *J Virol*. 2003;77(9):5487-5492. doi:10.1128/jvi.77.9.5487-5492.2003
24. Neufeldt CJ, Joyce MA, Van Buuren N, et al. The Hepatitis C Virus-Induced Membranous Web and Associated Nuclear Transport Machinery Limit Access of Pattern Recognition Receptors to Viral Replication Sites. *PLoS Pathog*. 2016;12(2):1-28. doi:10.1371/journal.ppat.1005428
25. Ahmed N, Foss D V., Powdrill MH, Pezacki JP. Site-Specific Cross-Linking of a p19 Viral Suppressor of RNA Silencing Protein and Its RNA Targets Using an Expanded Genetic Code. *Biochemistry*. 2019;58(33):3520-3526. doi:10.1021/acs.biochem.9b00428

Chapter 4: Future Directions

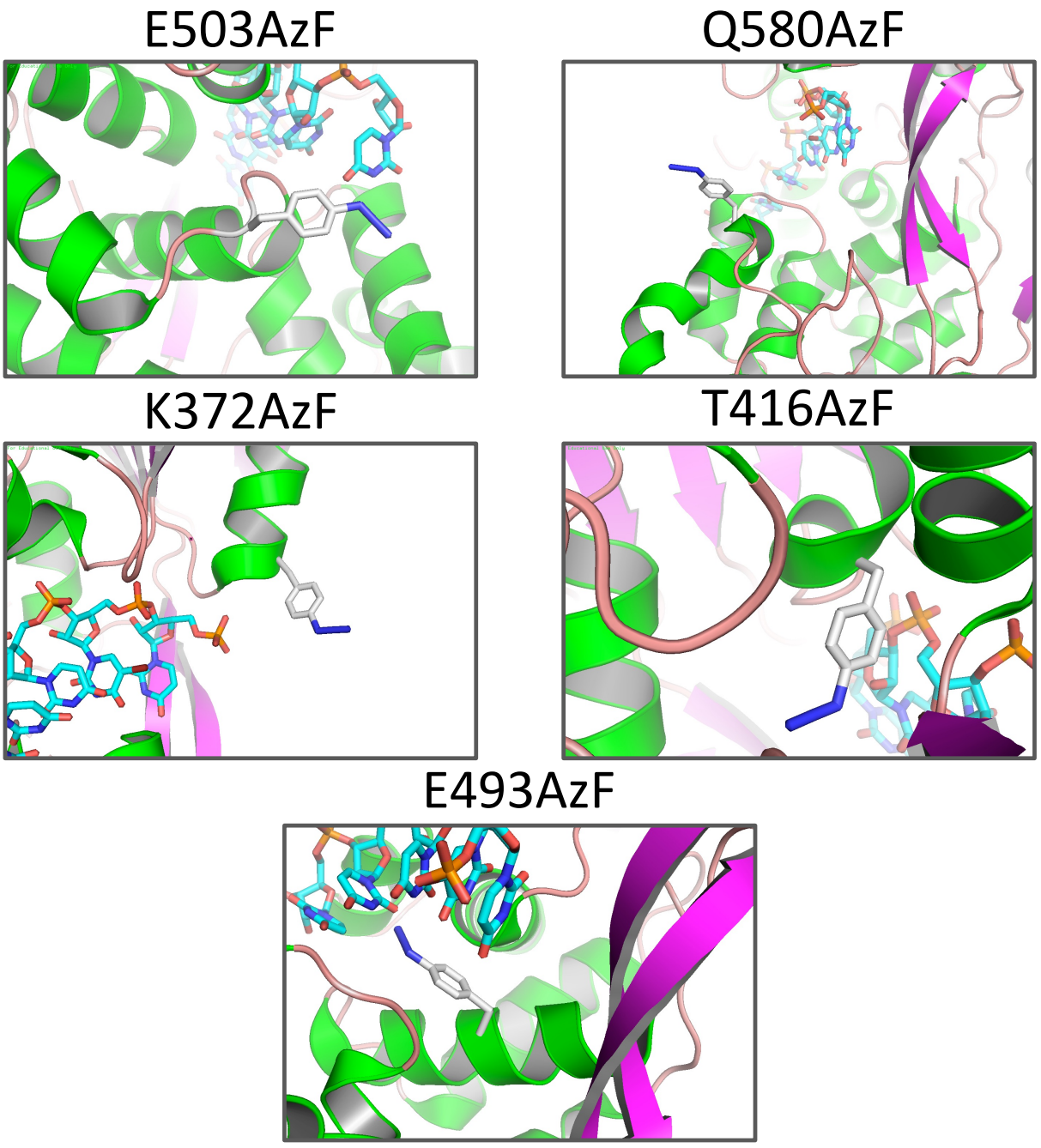
In regard to using AzF in different CLIP methodologies, we have demonstrated that photo-crosslinking E503AzF and Q580AzF yielded an increase in photo-crosslinked nucleic acid compared to conventional CLIP 254 nm light. The incorporation of AzF into RBPs is not applicable in many circumstances as it does require significant additional steps to ensure that it has not negatively affected the RBP's binding affinity and testing to ensure AzF is in close enough proximity to photo-crosslink bound nucleic acids. Another limitation of this method is the lack of crystal structures available for RBPs and more specifically, a crystal structure complexed with target RNA. Finding sites for AzF incorporation is very much an inspection of the crystal structure and amino acids in close proximity to the RNA. Despite this, we think that AzF could be used for CLIP methodologies for low abundance RBPs and those which have low affinities for their RNA targets and would benefit from this additional photo-crosslinking boost. We have shown that DNA is photo-crosslinked using AzF suggesting that incorporation of AzF into DNA binding proteins for photo-crosslinking and binding location mapping could also be performed, again with all the same limitations mentioned above.

Identifying NS3h binding sites located in the HCV genome through a CLIP-based protocol would provide a wealth of information on NS3h's role in binding the secondary structure of the HCV genome. This information could give additional insights into the changes from replication to translation in the HCV life cycle by the smoothening of secondary structures expected to be performed by NS3h. All that remains to be performed is to optimize the CLIP-based protocol and ensure the proper size of the fragmented RNA can be isolated and sent for library-preparation, sequencing and subsequent analysis.

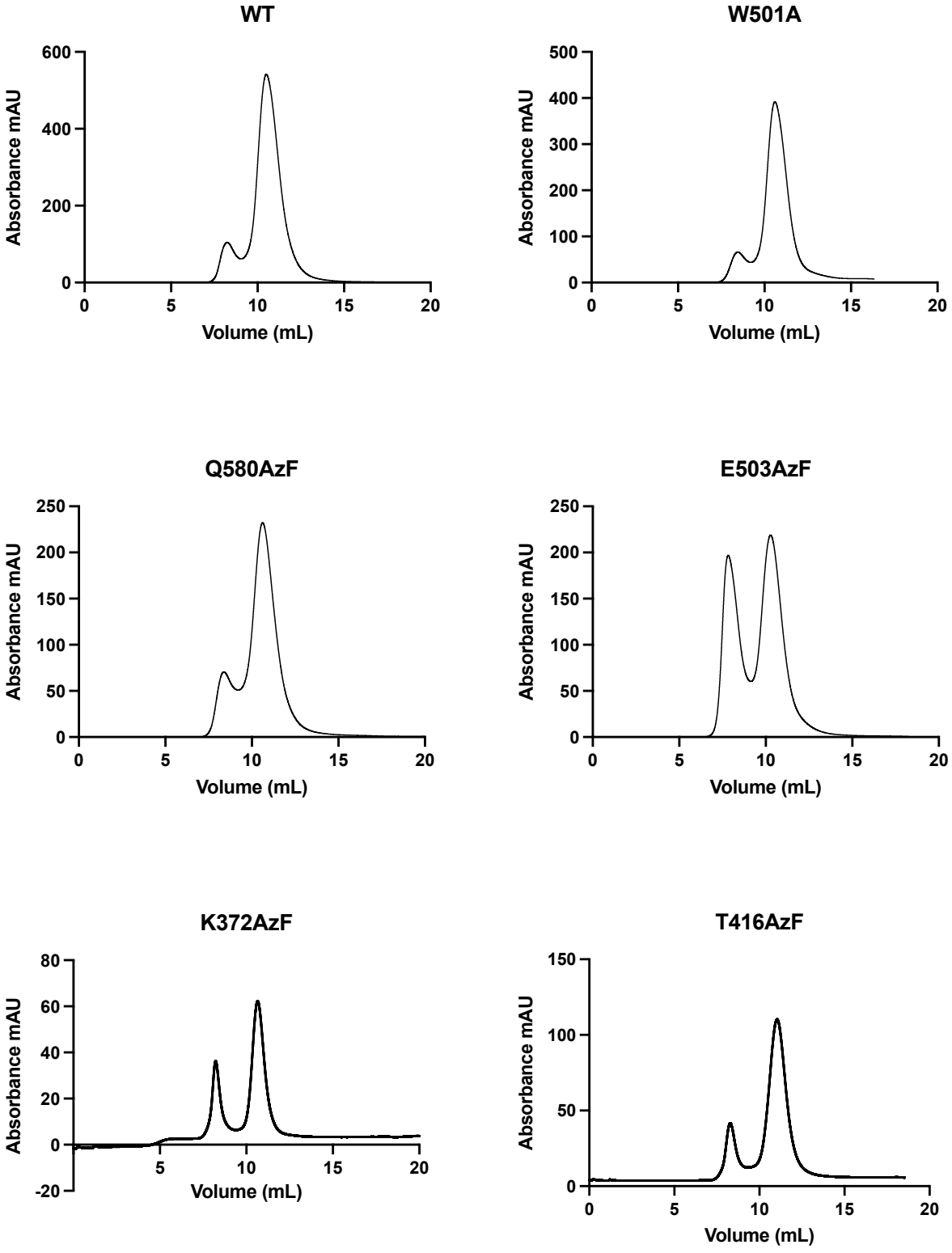
We have demonstrated that using AzF to photo-crosslink crRNAs to PspCas13b produced photo-crosslinking between WT PspCas13b and a bound crRNA. We need to further optimize AzF

photo-crosslinking if we wish to continue investigating the possibility of increasing the longevity of PspCas13b RNA knockdown through the covalent linking of crRNAs. We think that the use of different photo-crosslinking ncAAs such as benzophenone and diazirines could be explored as alternatives to AzF as they are excited at higher wavelengths in comparison. This would potentially mitigate the photo-crosslinking observed for the WT. We could also look to use photoreactive nucleotide analogs such as 4SU and 6SG to be incorporated into the DR section of the crRNA for the formation of a covalent bond to PspCas13b as both of the analogs are excited with 365 nm light as well. Photo-crosslinking crRNA to PspCas13b could have negative effects. PspCas13b affinity for crRNAs and nuclease activity and would need to be tested. Increasing the longevity of PspCas13b mediated RNA knockdown in mammalian cells could be important for its potential uses as a potent antiviral platform to target a variety of RNA viruses.

Appendix A – Supplemental Figures and Tables for Chapter 2



Supplemental Figure 2.1 Modelling AzF Mutations onto the NS3h Crystal Structure: The AzF structure was isolated from PDB: 5XHF and superimposed at all five residues chosen for incorporation into the NS3h crystal structure (PDB 3O8C).



Supplemental Figure 2.2 Size-Exclusion FPLC Chromatograms of the Purification of WT NS3h and all Mutants: The relative absorbance at 280 nm is shown as a black line.

Supplemental Table 2.1 Oligomers used for NS3h Photo-crosslinking.

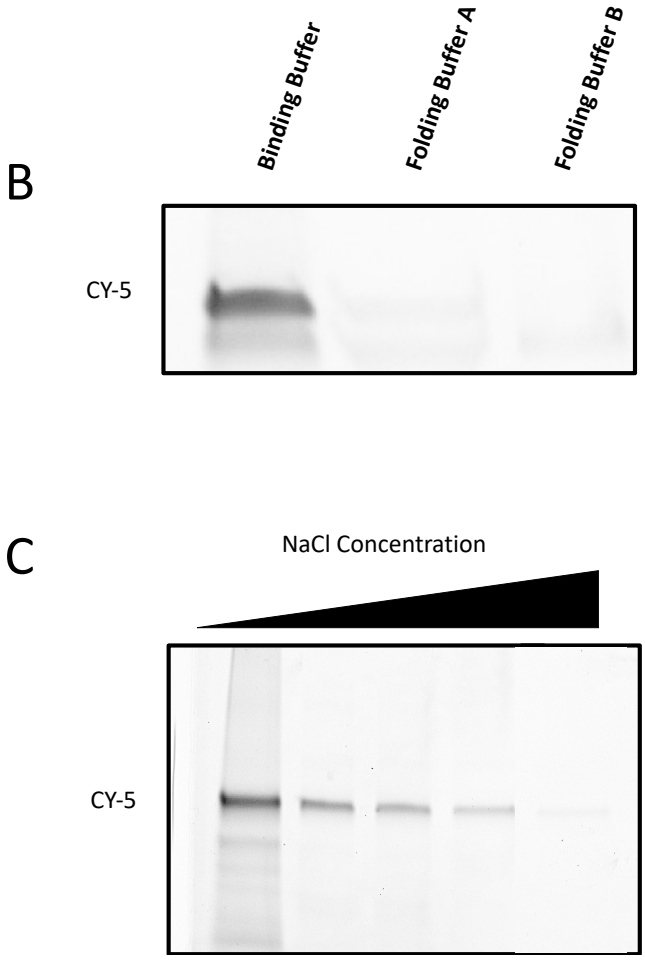
Photo-crosslinking Substrates	Sequence
Random 20-mer	5′-CACAGGTCAAGACTGCAACC-Cy5-3′
40-mer Complementary to 20-mer with poly(20)$_T$	5′-TTTTTTTTTTTTTTTTTTTTGGTTGCAGTCTTGACCTGTG-3′
PolyT(15-mer)	5′- Cy-TTTTTTTTTTTTTT-3′
PolyC(15-mer)	5′- Cy-CCCCCCCCCCCCCCC-3′

A

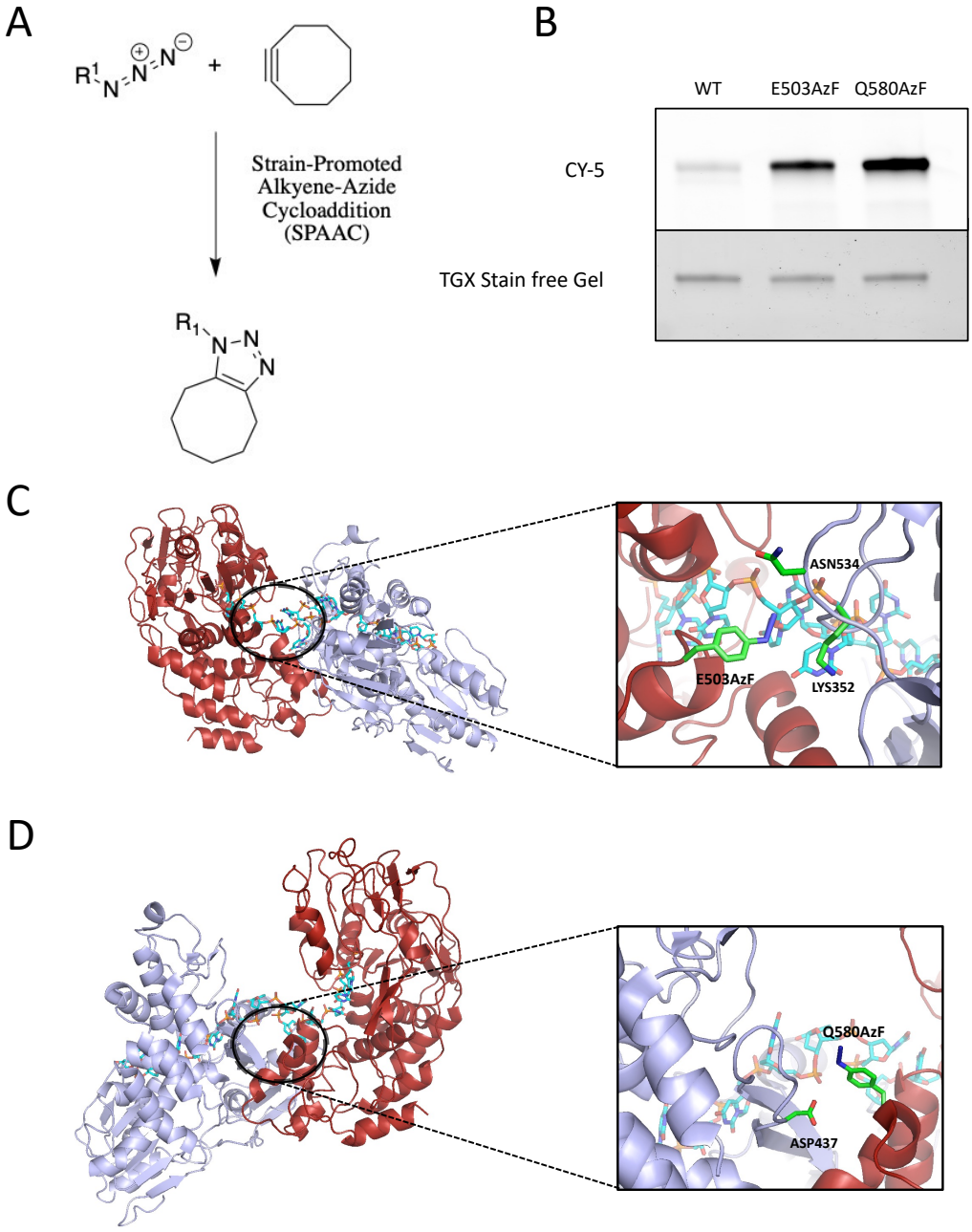
Folding Buffer A Composition (Mauger et al. 2015)
50mM HEPES (pH 8.0), 5mM $MgCl_2$, 200mM potassium acetate

Folding Buffer B Composition (Stewart et al. 2016)
330mM HEPES (pH 8.0), 20mM $MgCl_2$, 330mM NaCl

Binding Buffer Composition (Albenas et al. 2020)
25 mM MOPS pH 7.5, 10 mM NaCl, 3 mM MgCl2, and 0.1% Tween 20



Supplemental Figure 2.3 Ionic Concentrations in HCV Genome Folding Buffers are too High for E503AzF NS3h Photo-crosslinking: (A) Composition of the two selected buffers for RNA folding and the binding buffer used in all photo-crosslinking gels. (B) Photo-crosslinking SDS-PAGE gel with the binding buffer was substituted for either folding buffer A or B. 10 uM purified E503AzF was incubated with 250 nM of ssDNA substrate and irradiated with 302 nm light. (C) NaCl concentration affects photo-crosslinking efficiency. 10 uM purified E503AzF was incubated with 250 nM of ssDNA substrate in binding buffer with increasing NaCl concentrations (10 - 250 mM). The photo-crosslinking reactions were irradiated with 302 nm light and ran on an SDS-PAGE gel and imaged for Cy5.



Supplemental Figure 2.4 AzF Locations are Solvent Accessible and Near Adjacent NS3h Molecules: (A) Overview of the SPAAC reaction between an azido group and strained cyclooctene leads to stable triazoles. (B) Fluorescent labelling of NS3h WT and AzF mutants using Cy5-DBCO. SPAAC reaction between 1 uM NS3h and 2 uM Cy5-DBCO was incubated for 30 minutes and ran on an SDS-PAGE gel and imaged for Cy5. (C+D). Proposed crosslinking of E503AzF and Q580AzF to neighbouring residues of NS3h molecule bound to the same substrate (PDB: 2F55). The AzF structure was isolated from PDB: 5XHF and superimposed on positions E503 and Q580 on the red NS3h molecule. The blue NS3h molecule has the neighbouring residues in close proximity to AzF shown in green.

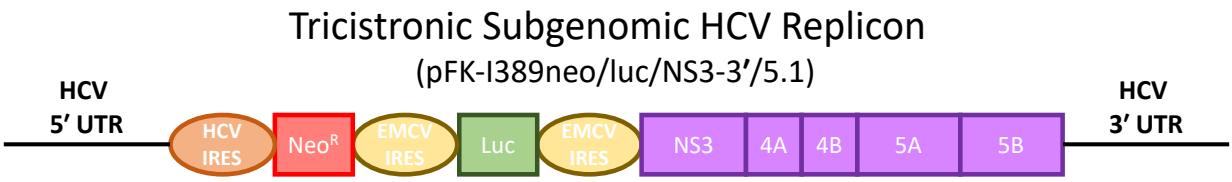
Supplemental Table 2.2 Oligomers used for NS3h Site-directed Mutagenesis.

TAG Mutation Location	Forward Primer	Reverse Primer
NS3h K372	5'-AGCTCGTCGCACTTCTACTTTGAGTGGCAGAAG-3'	5'-CTTCTGCCACTCAAAGTAGAAGTGCGACGAGCT-3'
NS3h T416	5'-GAAGTCGCCGGTAAAGCCCTACATGAGAGCATCGGTCGA-3'	5'-TCGACCGATGCTCTCATGTAGGGCTTTACCGGCGACTTC-3'
NS3h E493	5'-CGTCATAGCACTAACAGAGGACGGACGAGTCG-3'	5'-CGACTCGTCCGTCCTCTGTTAGTGCTATGACG-3'
NS3h E503	5'-GGCGTGAGCTAATACCAAGCACAGCCCGCG-3'	5'-CGCGGGCTGTGCTTGGTATTAGCTCACGCC-3'
NS3h Q580	5'-AACACTTCCACATCTAGTCCCACGATGGGGG-3'	5'-CCCCCATCGTGGGACTAGATGTGGAAGTGTT-3'

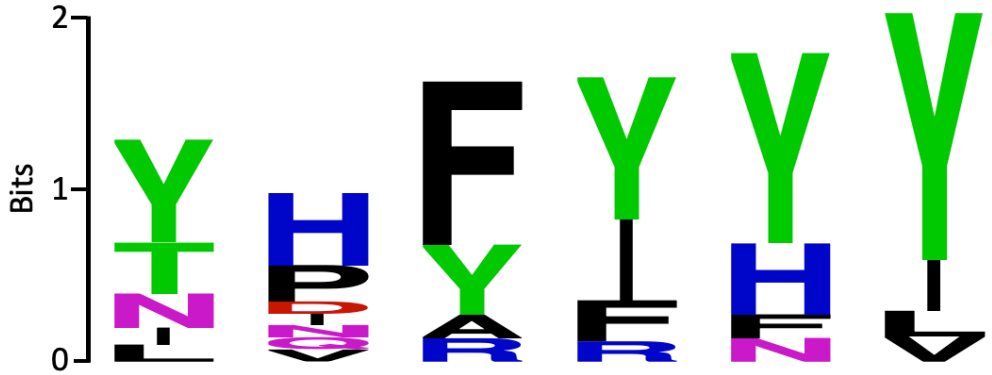
Appendix B – Supplemental Figures and Tables for Chapter 3

Supplemental Table 3.1 Cas13b Accessions Chosen for PspCas13b Ortholog Multiple Sequence Alignment: PspCas13b is highlighted in red, PbuCas13b is highlighted in green and BzCas13b is highlighted in blue.

Species	Cas13 Accession
BzoCas13b Bergeyella zoohelcum	WP_002664492
PinCas13b Prevotella intermedia	WP_036860899
PbuCas13b Prevotella buccae	WP_004343973
AspCas13b Alistipes sp. ZOR0009	WP_047447901
PsmCas13b Prevotella sp. MA2016	WP_036929175
RanCas13b Riemerella anatipestifer	WP_004919755
PauCas13b Prevotella aurantiaca	WP_025000926
PsaCas13b Prevotella saccharolytica	WP_051522484
Pin2Cas13b Prevotella intermedia	WP_061868553
CcaCas13b Capnocytophaga canimorsus	WP_013997271
PguCas13b Porphyromonas gulae	WP_039434803
PspCas13b Prevotella sp. P5-125	WP_044065294
FbrCas13b Flavobacterium branchiophilum	WP_014084666
PgiCas13b Porphyromonas gingivalis	WP_053444417
Pin3Cas13b Prevotella intermedia	WP_050955369



Supplemental Figure 3.1 HCV Subgenomic Replicon Model used for PspCas13b Mediated Knockdown: HuH7-SGR-Luc encodes all the non-structural proteins (NS3-NS5B) required for replication (purple) as well as a luciferase reporter (green) which are both translated using the encephalomyocarditis virus EMCV IRES (yellow). HuH7-SGR-Luc also encodes a neomycin resistance gene (red) which is used as a selection marker and is translated using the HCV IRES (orange).



Alignment Position	590	656	747	803	948	1268
PbuCas13b Residue	Y409	H452	F496	Y540	Y661	Y863
BzuCas13b Residue	T466	P518	Y597	I653	H779	Y919
PspCas13b Residue	N387	I440	R477	F513	Y626	I847

Supplemental Figure 3.2 Sequence Logos of Multiple Sequence Alignment Demonstrates PspCas13b AzF Sites have Conserved Tyrosine: 15 Cas13b orthologs were aligned using NCBI constraint-based multiple alignment tool. The six residues were chosen from analysis of PbuCas13b and BzCas13b structures looking at conserved residues in close proximity to bound crRNA. The corresponding PspCas13b residues can be seen. Sequence logos for each residue shows high tolerance for a tyrosine moiety.

Supplemental Table 3.2 Oligomers used for Pspcas13b Site-directed Mutagenesis.

TAG Mutation Location	Forward Primer	Reverse Primer
Cas13b I847	5′-CTGCTGGCGTTTCTCTACTGCCGGTTGCTGCG-3′	5′-CGCAGCAACCGGCAGTAGAGAAACGCCAGCAG-3′
Cas13b Y629	5′-CTGCATGATCCGCTAGTTCAGGCCGGTGATC-3′	5′-GATCACCGGCCTGAACTAGCGGATCATGCAG-3′
Cas13b N387	5′-GTCTGCCGAAGCCCTACAGGGGCTGCTCG-3′	5′-CGAGCAGCCCCTGTAGGGCTTCGGCAGAC-3′
Cas13b R477	5′-TTCCAGGGTGCTCATCTAGCAGCTGGGGATTGTC-3′	5′-GACAATCCCCAGCTGCTAGATGAGCACCCTGGAA-3′
Cas13b F513	5′-TCTTTCTGCATGGCCTGCTACAGTCTCTTGTACCGG-3′	5′-CCGGTACAAGAGACTGTAGCAGGCCATGCAGAAAGA-3′
Cas13b I440	5′CTCGACCTTGTTGTTTTCCAGCTAGTAGTGTGTGTAGGTGTCCAC-3′	5′GTGGACACCTACACACACTACTAGCTGGAAAACAACAAGGTCGAG-3′

Supplemental Table 3.3 List of Oligomers Cloned into PspCas13b crRNA Expression Plasmid.

Target crRNA	Forward Primer	Reverse Primer
Luc-1	5′-CACCGGCCCTTCTTAATGTTCTTAGCATCGGCCAT-3′	5′-CAACATGGCCGATGCTAAGAACATTAAGAAGGGCC-3′
Luc-2	5′-CACCGAGACATCTCGAAGTACTCGGCATAGGTGAT-3′	5′-CAACATCACCTATGCCGAGTACTTCGAGATGTCTC-3′
Luc-3	5′-CACCGCTGCAGGCCCTTCTTAGACACGAACACCAC-3′	5′-CAACGTGGTGTTCGTGTCTAAGAAGGGCCTGCAGC-3′
Luc-4	5′-CACCGCACCCGAAAGCCGCAAATCAGGTAGCCCAG-3′	5′-CAACCTGGGCTACCTGATTTGCGGCTTTCGGGTGC-3′
Luc-5	5′-CACCGCCGGACATAATCATAGGGCCGCGCACACAC-3′	5′-CAACGTGTGTGCGCGGCCCTATGATTATGTCCGGC-3′
Non-Target crRNA	5′-CACCGAGTCGTGGCAGGATATCGAATGTTCTTGTA-3′	5′-CAACTACAAGAACATTCGATATCCTGCCACGACTC-3
HCV-1	5′-CACCGGGTAATATGCTACAGCATTGAGTCCGAGGC-3′	5′-CAACGCCTCGGACTCAATGCTGTAGCATATTACCC-3′
HCV-2	5′-CACCGCCCTACCAGTCCTGCCTCGCCGCTGCGAGC-3′	5′-CAACGCTCGCAGCGGCGAGGCAGGACTGGTAGGGC-3′
HCV-3	5′-CACCGTCTGGGACAAGAAATGGGCGTCTATGTGGG-3′	5′-CAACCCCACATAGACGCCCATTTCTTGTCCCAGAC-3′
HCV- 4	5′-CACCGCATGCCATGATGTATTTGGTTATGGGGTGT-3′	5′ CAACACACCCCATAACCAAATACATCATGGCATGC-3′