BINDING DETERMINANTS OF THE HIV-1 REV RNA TO THE POKEWEED ANTIVIRAL PROTEIN

KASS JOBST

A THESIS SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIRMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

APRIL 2015

Abstract

The pokeweed antiviral protein (PAP) belongs to a family of N-glycosidases called ribosome inactivating proteins (RIPs), which remove an adenine residue from the sarcin/ricin loop (SRL) of rRNA through a process called depurination. PAP also has antiviral activity against a variety of plant and animal viruses that may occur by depurinating the viral RNA. The mechanism by which PAP specifically targets viral RNA over cellular RNA is not well understood. It has recently been shown that PAP depurinates the ORF of Rev RNA of HIV-1 *in vivo* without causing toxicity. In this study, I identify the sequence and structure of Rev RNA that PAP preferentially interacts with, compare the PAP binding site in Rev to the sarcin/ricin loop, and attempt to describe the importance of RNA structure for PAP specificity when targeting viral RNA over cellular RNA. My results show that PAP binds to a short GGGAA sequence at the site of depurination of Rev RNA. Structural analysis reveals that the binding site is within a hairpin pentaloop. The pentaloop contains a pseudo GNRA loop structure that swings G and A residues out of the hairpin, allowing PAP access to these purines for binding and depurination. Mutation of the GGGAA sequence to AAAAA resulted in reduced binding to the RNA. These results show a new RNA structure PAP binds and depurinates other than the SRL and that the secondary structure and accessibility to a purine substrate on RNA is important for PAP binding. These data provide new insight into the specificity of PAP to target an RNA.

Acknowledgements

I would like to express gratitude to my mentor, Dr. Katalin Hudak, for accepting me into her lab and providing the best environment for my experiments and my development. Her enthusiasm for my topic and tremendous expertise has encouraged me to push the boundaries of my knowledge and skills as a researcher. I appreciated her boundless patience and encouragement throughout my degree. Thank you for your tutelage, advice, and guidance; you will continue to inspire me in my future career.

I would like to thank Dr. Andrew White for being my advisor and for providing advice on my research. Thank you to Dr. Audette and Dr. Kubiseski for being a part of my examining committee.

Thank you to my lab members: Alice Zhabokritsky, Mez Kutky, Rajita Karan, Gabriela Krivdova, Kira Neller, Alexander Klenov, Shira Ellion-Jourard, Luca Spremulli, Camille Diaz, and Ming Lim. Your friendship and support throughout my degree is valuable to me. Thank you to Dr. White's research team for the sharing equipment, skills, and ideas. Special thanks to Andrew Hogan for his summer project on my work and his involvement with optimizing binding assays and cloning.

Thank you to Karen Rethoret, Dr. Kubiseski, and Dr. Coe for training and access to confocal microscopes used during my project. Thank you to York undergrad lab technicians Maria Mazzurco, Irina Shuraloyva, and Helen Jung for their help and access to laboratory equipment.

Lastly, thank you to my family and my close friends. Your love and support has carried me through all aspects of my study and my life.

Table of contents

List of Tables

List of Figures

List of Common Abbreviations

aa	Amino acid
AIMV	Alfalfa mosaic virus
Amp	Ampicillin
ATP	Adenine triphosphate
GTP	Guanosine triphosphate
B_{MAX}	Maximum binding
BMV	Brome mosaic virus
BSA	Bovine serum albumin
CaMV	Cauliflower mosaic virus
cDNA	Complementary DNA
CMV	Cytomegalovirus
CPM	Counts per minute
DAPI	4',6-diamidino-2-phenylindole
dH2O	Distilled water
DMEM	Dulbecco's modified Eagle's Medium
DMS	Dimethyl sulfate
DMSO	Dimentylsulfoxide
ddNTP	Dideoxyribonucleotide triphosphate
dNTP	Deoxyribonucleotide triphosphate
DSR	Downstream reverse
DTT	Dithiothreitol
EDTA	Ethylene dinitrilotetraacetic acid
eEF	Eukaryotic elongation factor
EGTA	Ethylene glycol tetraacetic acid
Env	Envelope
EtBr	Ethedium bromide
EtOH	Ethanol
ERK	Extracellular signal-regulated kinases
FBS	Fetal bovine serum
FITC	Fluorescein isothiocyanate
FPLC	Fast protein liquid chromatography
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
HBV	Hepatitus B virus
HEK	Human embryonic kidney
Hela	Henrietta Lackscervical cancer cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV-1	Human immunodeficency virus I
HSV	Herpes simplex virus
HTLV-1	Human T-lymphotrophic virus I
IC_{50}	Half maximal inhibitory concentration
JNK	C-Jun N-terminal kinase
K_{Cat}	Catalytic rate constant
K_D	Dissociation binding constant
LB	Luria-Bertani broth
LTR	Long terminal repeat
MAPK	Mitogen activated protein kinase

Met	Methionine
MFE	Minimum free energy
MOPS	3-(N-morpholino) propanesulfonic acid
mRNA	Messenger RNA
Nef	Negative effector
NLS	Nuclear localization signal
NWB	North-western buffer
OD	Optical density
ORF	Open reading frame
PAP	Pokeweed antiviral protein
PAPc	C-terminal mutant of PAP
PAPx	Active site mutant of PAP
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCI	Phenol chloroform isoamyl alcohol
PCR	Polymerase chain reaction
PVX	Potato virus X
Rev	Regulator of virion expression
RIP	Ribosome inactivating protein
RRE	Rev response element
rRNA	Ribosomal RNA
RT	Reverse transcription
SRL	Sarcin/ricin loop
STE	Sodium Tris EDTA
Stx	Shiga toxin
SV40	Simian vacuolating virus 40
TAPS	N-tris(Hydroxymethyl)methyl-3-aminopropanesulfonic acid sodium-potassium salt
Tat	Transactivator of transcription
TAR	Tat responsive element
TBE	Tris borate EDTA
TE	Tris EDTA
TMV	Tobacco mosaic virus
tRNA	Transfer RNA
UV	Ultraviolet
Vif	Viral infectivity factor
Vpu	Viral protein U
Vpr	Viral protein R
YFP	Yellow fluorescent protein

1. Introduction

1.1. Ribosome inactivating proteins

Ribosome inactivating proteins (RIPs) are toxic proteins that inhibit protein synthesis at the level of the ribosome[1]. These proteins affect the ribosome by a well-defined RNA N-β-glycosidase (EC 3.2.2.22) reaction that depurinates a single adenine from an invariant GAGA sequence within the sarcin/ricin loop (SRL) of the large ribosomal RNA[1–3]. It was demonstrated that depurination of the SRL was able to inhibit translation by affecting the interaction of elongation factors EF-1 and EF-2 with the ribosome[4] (EF-G and EF-Tu of bacterial ribosomes[5]), which are involved in the binding and translocation of tRNAs during protein synthesis.

RIPs are sub-divided into three major groups: plant type I, type II and type III RIPs and Shiga and Shiga-like toxin (Stx)[6] (**Figure 1**). Plant type I RIPs are comprised of an enzymatically active monomeric peptide chain, averaging 29 kDa in size, often referred to as the A-chain. Type I RIPs can be isolated from *Phytolacca americana* (Pokeweed antiviral protein)[7], *Trichosanthes kirilowii* (trichosanthin)[8], and *Gelonium multiflorum* (gelonin)[9]. Type II RIPs are similar in structure to type I because they also contain the A-chain, but are covalently linked via a disulfide bond to a 35 kDa lectin protein B-chain. The lectin chain makes these proteins particularly toxic, unlike type I counterparts, because of their ability to bind cell-surface glycoproteins and activate endocytosis events to gain access to cellular ribosomes[2,10,11]. Examples of Type II RIPs are ricin (from *Ricinus communis*)[12], abrin (*Abrus precatorius*)[13], and ebulin (*Sambucus ebulus*)[14]. The Stx group of RIPs are also described in terms of an “A” and “B” chain

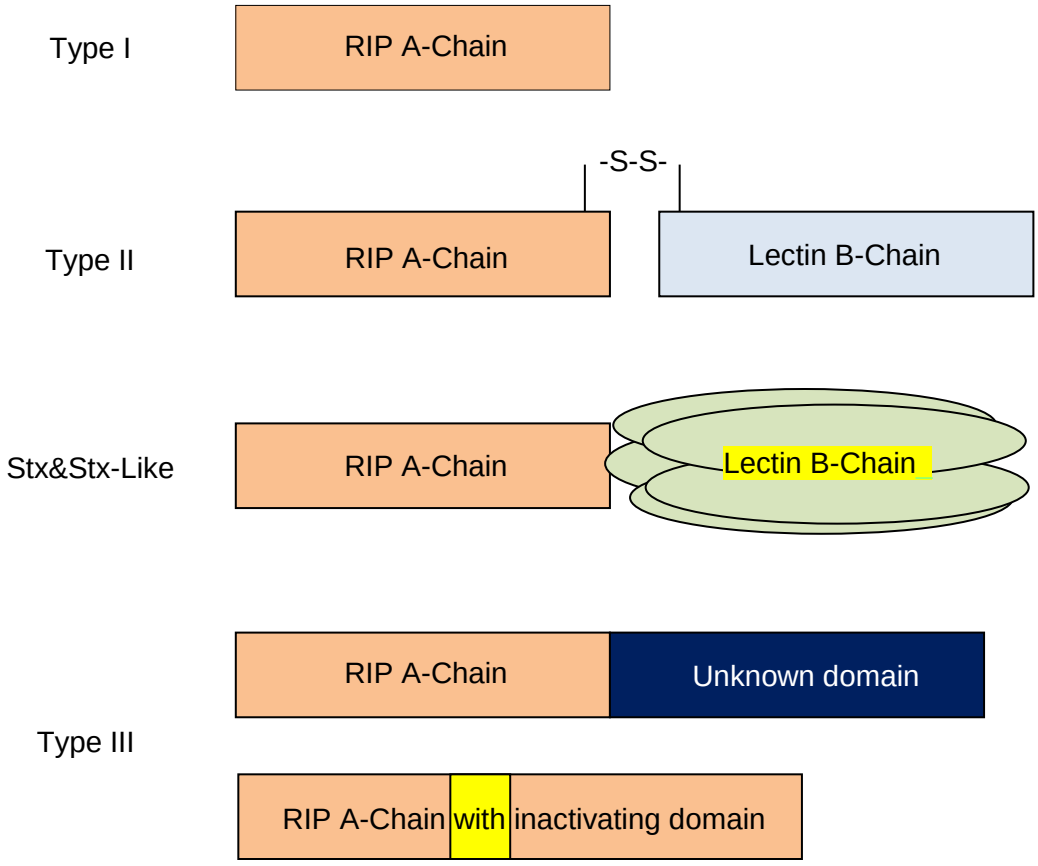


Figure 1 – Schematic overview of ribosome inactivating protein structural domains. All RIPs contain an enzymatically active A-chain with N-glycosidase activity. Type I RIPs only contain a single peptide chain with the enzymatic activity. Type II RIPs comprise the active A-chain disulphide bonded to a lectin B-chain. Stx and Stx-like RIPs have the active A-chain bound to a protein pentamer B-chain that also has carbohydrate binding properties. Type III RIPs are rare and have unique structures that do not fit in the above categories. For example, they may contain active A-chain bound to a long C-terminal domain with unknown function or an A-chain containing an inactivating domain than needs to be cleaved in order form the active enzyme.

organization like plant type II RIPs, but the overall structure of these proteins differs substantially. The Stx RIPs contain an enzymatically active A-chain equivalent to the plant type I and II, and the B-chain is a pentamer of 5 identical peptides that specifically bind to glycolipid globotriaosyleramide[15]. Less common structures of RIPs are also found in nature and are typically categorized as type III RIPs. For example, the JIP60 RIP (from *Hordeum vulgare*) contains an enzymatic A-chain with an extended 30 kDa C-terminus that plays no known function to the activity or specificity of the enzyme[16]. Another example is the maize RIP (from *Zea maize*), which requires proteolytic cleavage of 26 aa from the central domain of the peptide chain, yielding two non-covalently linked segments that form an active enzyme[17].

There are several interesting conceptions for the presence of RIPs in nature; however their role is not well understood. The most popular idea is that RIPs act in anti-herbivory such that predators or parasites avoid eating the plants; however this is less likely for plants containing type I RIPs because they are non-toxic to predators[18]. Scientific interest in these enzymes arose from unexpected N-glycosidase activity against RNA substrates other than the sarcin/ricin loop of ribosomes. Several lines of evidence show that RIPs are able to hinder various forms of viral infection in both plant[19–21] and mammalian[22–24] tissue by depurinating viral RNA directly, using a separate mechanism of action than damaging ribosomes. As a result, RIPs may play a role in plant defense against virus infection by directly damaging the viral RNA messages. While the primary function of these proteins is unknown, the fact that these proteins have antiviral activity may prove useful in both therapeutic applications in

medicine and the bioengineering of crops in agriculture. The antiviral activities, shown by type I RIPs, are of special interest because of their non-toxic properties.

1.2. Pokeweed antiviral protein

1.2.1. Overview

In 1925, Duggar and colleagues first described that plant extracts from *Phytolacca americana* (American pokeweed) could be used to inhibit the transmission of Tobacco mosaic virus (TMV) in tobacco plants[25]. In 1975, Irvin isolated and identified that the protein responsible for the antiviral activity in pokeweed was a type I RIP[7]. Since then, studies show that PAP exhibits antiviral activity against a number of other plant (*e.g.* Brome mosaic virus[26]) and animal viruses (*e.g.* Human immunodeficiency virus-1[27]). The antiviral activity of PAP was initially attributed to the ribosome-inactivating action in the pokeweed plant[28], but there is evidence to show that PAP can depurinate viral RNA directly[20,24,29]. Ribosomes stall at the site of depurination which results in translation arrest[30] and inhibition of protein synthesis of the viral message (**Figure 2**). In addition, PAP can be transiently expressed in the cytoplasm of animal cells without causing toxicity and express similar antiviral activity on human viruses[24,29]. Furthermore, several mutations of PAP show that the antiviral activity can be separated from RIP activity[31,32], and the depurination of the SRL is not always sufficient to cause toxicity in a cell[31]. Together, these results suggest that RIP activity alone does not describe the entire mechanism of viral inhibition by PAP. However, the exact mechanism by which PAP exhibits antiviral activity is not completely understood and is still under investigation. Much work has been done to define the structural requirements

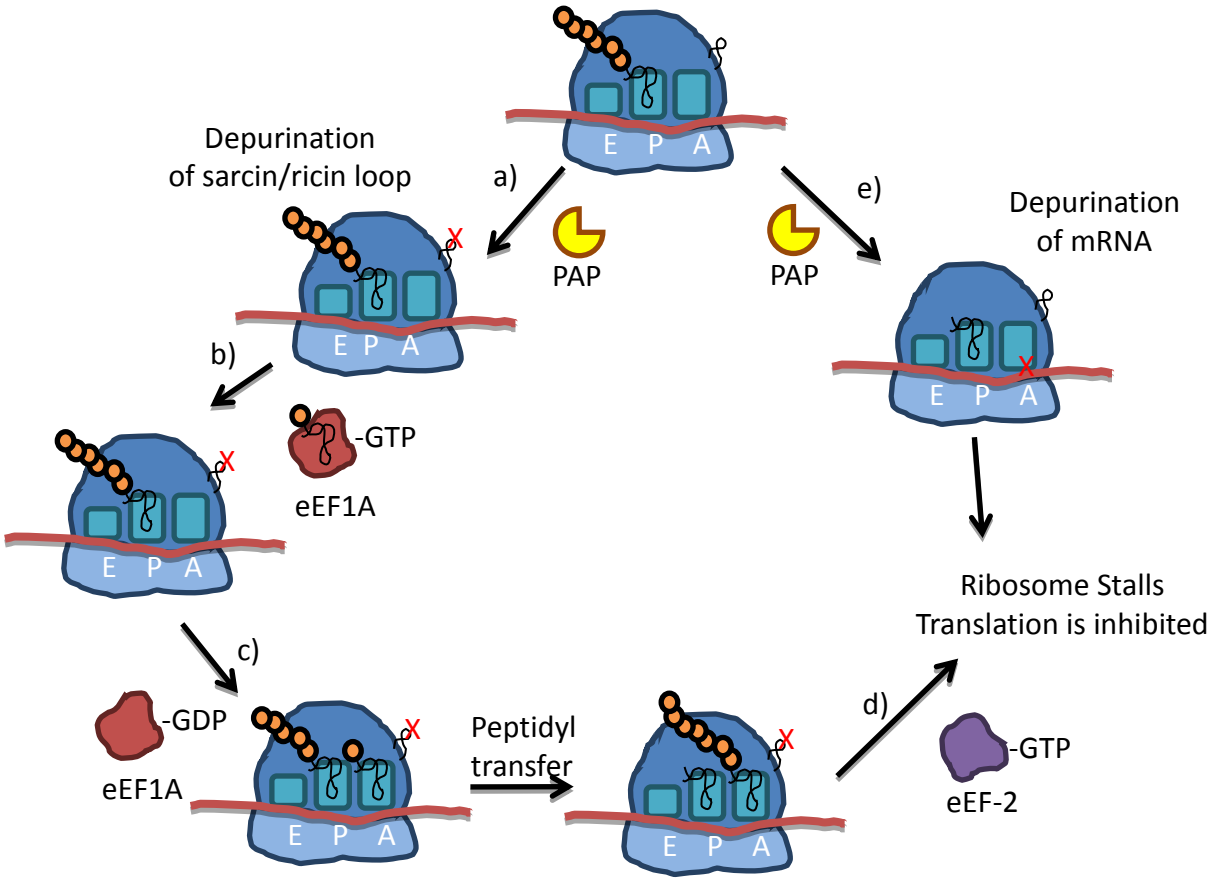


Figure 2 – Depurination of the sarcin/ricin loop or mRNA results in the stalling of translation. PAP has two mechanisms of inhibiting translation in the cell. **A)** RIPs commonly depurinate the conserved sarcin/ricin loop ribosomes which play a major role in the elongation reaction of protein translation. **B)** The eEF1A-tRNA-GTP ternary complex binds to the ribosome, bringing a charged tRNA to the A site. **C)** The eEF1A protein leaves and the ribosome will catalyze the peptidyl transfer of the polypeptide in the P site to the charged tRNA in the A site. **D)** The eEF2 will be unable to bind to the damaged SRL to catalyze the translocation step, thus stalling the ribosome. **E)** PAP has been shown to depurinate viral RNA directly. Translation of depurinated RNA also results in ribosome stalling because the ribosome can no longer recognize a codon for tRNA binding.

of PAP and related RIPs binding to ribosomes and the SRL of rRNA, but very little work has been done to show how PAP targets viral RNA. In this research project, I investigate the structural and sequential requirements of a viral RNA substrate in HIV-I required for PAP binding and gain further insight into how PAP can target one RNA over another.

1.2.2. Structure and enzymatic activity

The pokeweed antiviral protein is a 29 kDa, highly basic (theoretical pI of 8.9) protein isolated from *P. americana*, that falls into type I class of RIPs[33]. The mature PAP enzyme is 262 aa long, derived from a 313 precursor that is cleaved at both the N and C-termini[34,35]. PAP exhibits a definitive “RIP fold” structure, sharing 62 conserved amino acids amongst all RIPs[36]. Glu 176, Arg 179, Tyr 72 and Tyr 123 are key residues involved in the formation of PAP's catalytic site[37]. The crystal structure and schematic diagram of PAP is shown in **Figure 3**. Co-crystallization of PAP with an adenine analogue and structure fitting models of the SRL into the binding pocket have revealed that Glu 176 and Arg 179 residues are directly involved in the catalytic activity of the enzyme[38,39]. Arg 179 partially protonates N3 of the substrate adenine, which is expected to weaken the glycosidic N9-C1’ bond between the ribose sugar and nitrogenous base. Glu 176 both stabilizes an oxocarbenium transition state of the ribose sugar[40] and activates a water molecule for back attack of the C1’ carbon of the ribose sugar, causing cleavage of the N9-C1’ bond[41] (**Figure 4**). Mutation of either residue results in a significant reduction of enzymatic activity by about 1000 fold[42]. Tyr 72 and Tyr 123 are involved in orienting the substrate purine by forming an energetically favourable

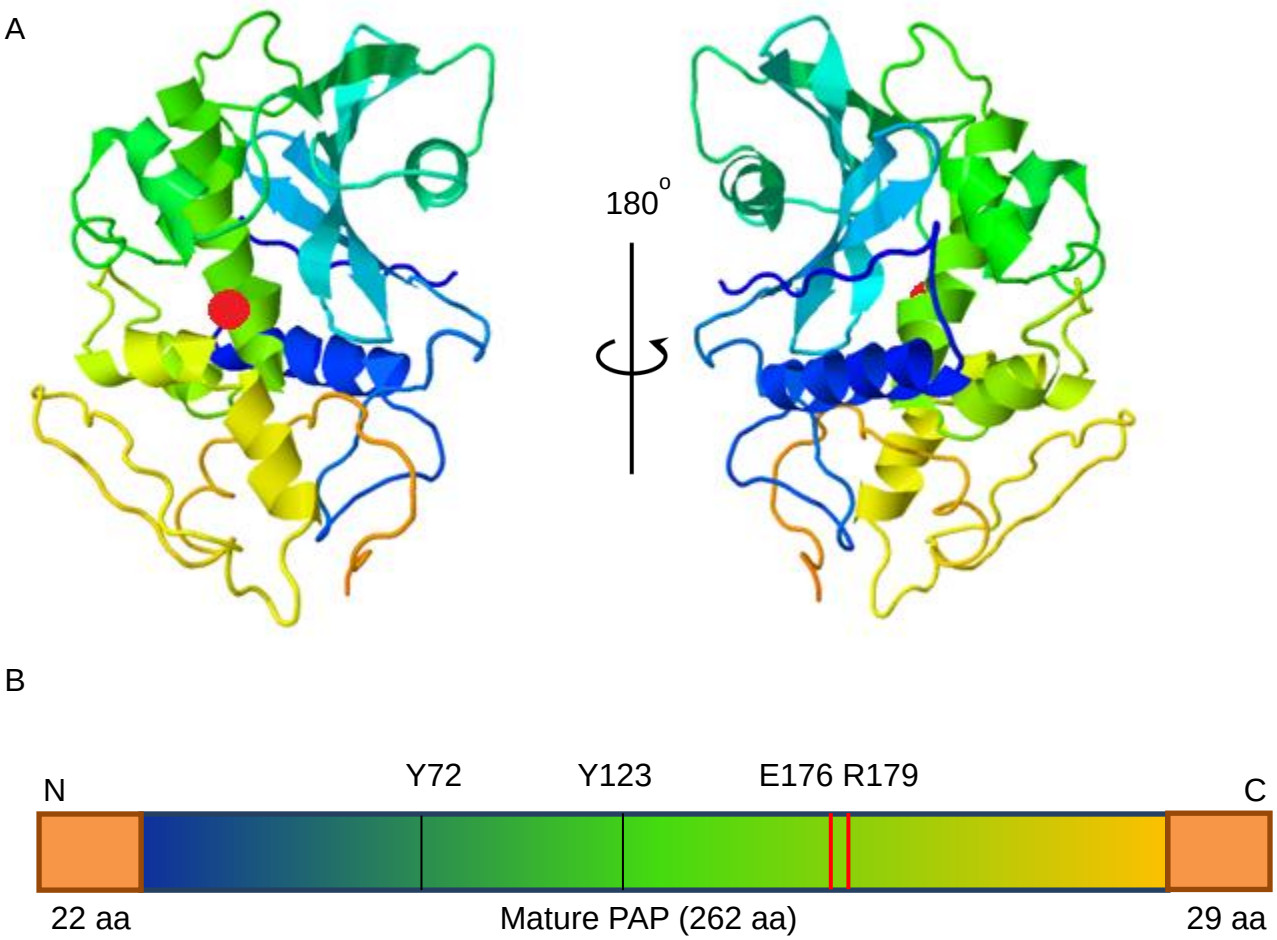


Figure 3 – Crystal structure of PAP and a schematic drawing of the active site residues. A) Ribbon diagram of 2.5 Angstrom crystal structure of PAP isolated from the spring leaves of *Phytolacca americana* (PDB: 1PAG)[33]. The red sphere designates the active site cleft where depurination occurs. **B)** The mature form of PAP is a 262 amino acid protein after post translational cleavage of a 22 and 29 amino acid at N and C terminal ends from the immature protein, respectively. The catalytic residues involved in depurination are glutamate 176 and arginine 179. Tyrosine 72 and 123 are important residues involved in sandwiching and orienting purine bases in the active site.

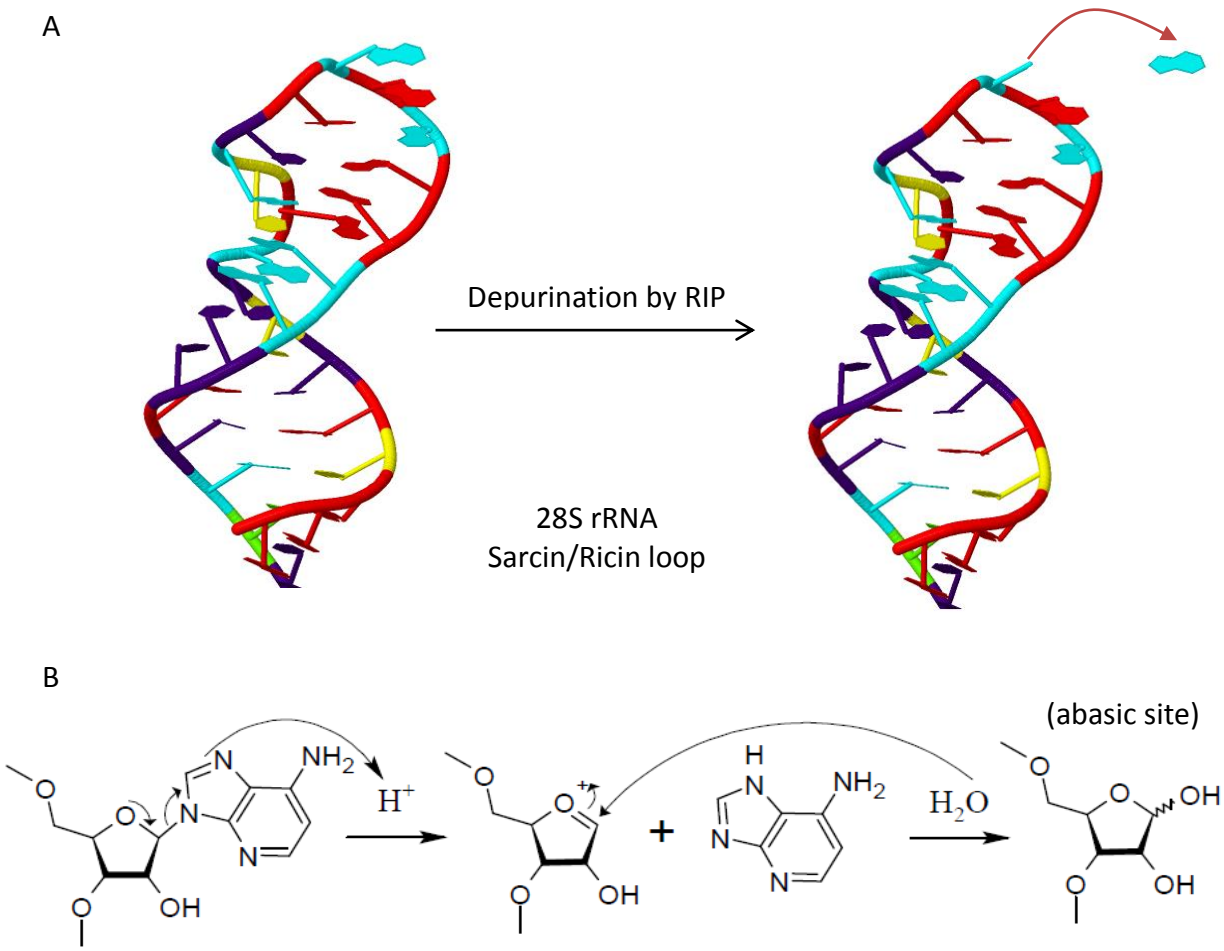
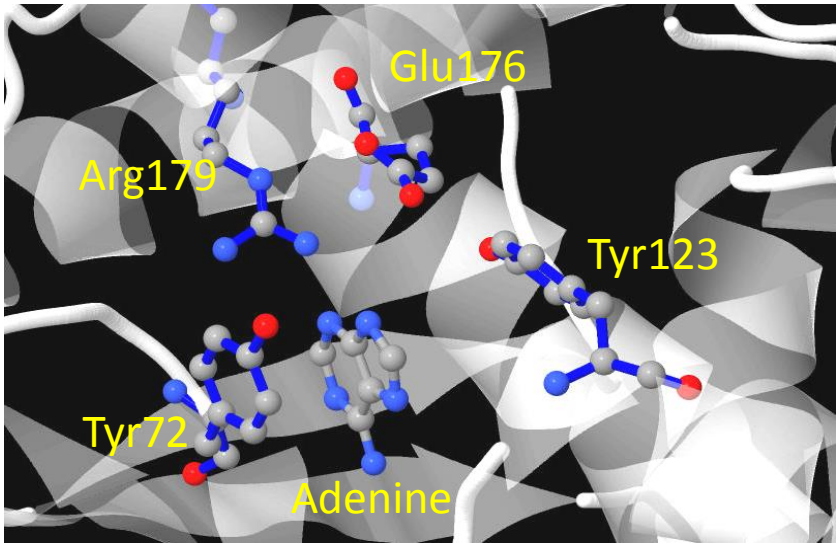


Figure 4 – Crystal structure of the 28S sarcin/ricin loop showing the result of RIP depurination and the reaction mechanism. A) 2.1 Angstrom crystal structure of a 29-nucleotide RNA containing the sarcin/ricin loop (SRL) of rat 28S rRNA (PDB: 430D)[81]. The conserved GAGA tetraloop loop contains tilted Watson-Crick base-pairng with an uncommon base-pairing between the first G and the second A that enhances the stability of the hairpin structure. A4323 is the target for RIP depurination, specifically because the adenine is external to the loop and exposed for attack. **B)** The mechanism of depurination is 2 step: protonation of N-7 (or N-3) of a purine leads to the weakening of the glycosidic N-9-C1' bond and formation of an oxocarbenium ion at the ribose sugar. Hydrolysis is completed with a nucleophilic attack of water on the ribose C1'. Glu 176 and Arg 178 of PAP are directly involved in stabilizing the oxocarbenium transition state and protonation the purine base.

A)



B)

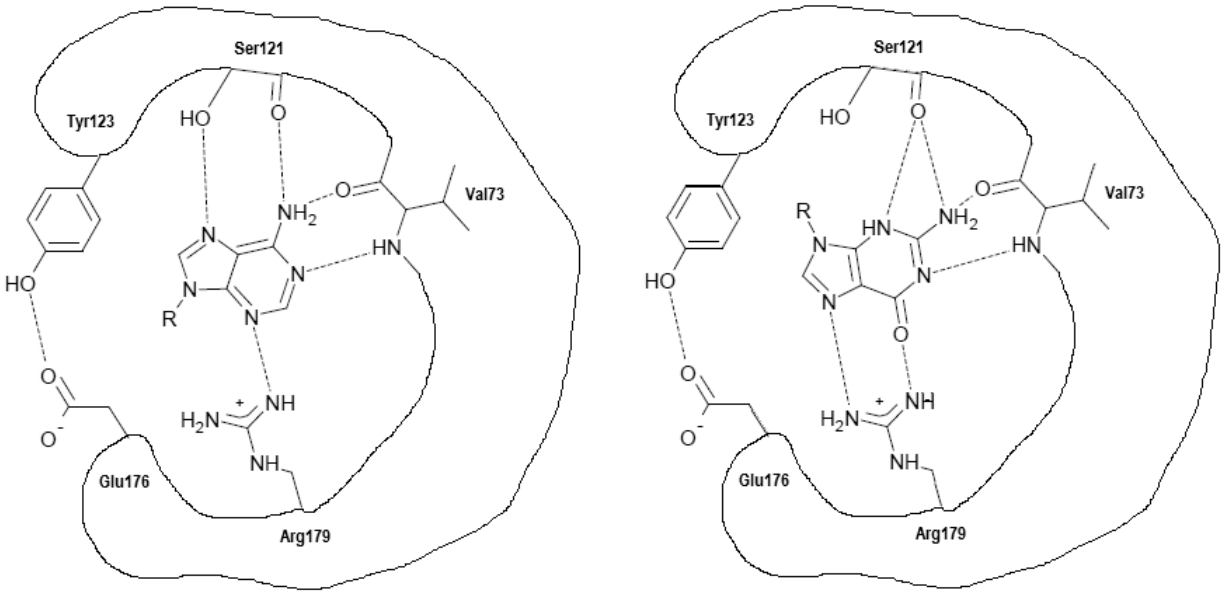


Figure 5 – Schematic diagrams of active site pocket residues involved in catalysis and substrate specificity of PAP A) Ribbon diagram of active site pocket of PAP in complex with adenine (PDB: 1QCI)[38]. The adenine purine is held between two tyrosine residues and enzymatic activity is dependent on E176 and R179. The pdb file was modeled using the JMol program (jmol.org). B) Several other hydrophilic residues stabilize binding to purines via hydrogen bonding. Guanine (right) is held in the active site at a 180º angle when compared to adenine (left).

pi-stacking conformation close to the active site residues[37] (**Figure 5a**). Based on mutational analysis, Tyr 72 is more important than Tyr 123 for enzymatic activity, but Tyr 123 is important for depurination of the SRL from intact ribosomes, potentially by interacting with protein residues on the ribosome[31]. Binding specificity to purines is mediated through additional hydrogen-bonding between the substrate ring and PAP active site residues[38,43] (**Figure 5b**). Arg 179, the carbonyl backbone of Val 73, and Ser 121 hydrogen bond to the N3, the external ring amine, and N7 of adenine, respectively[38]. Guanine can also fit the active site pocket, but is rotated 180^{o} such that Arg179, the carbonyl backbone of Val 73, and the carbonyl backbone of Ser 121 hydrogen bond to the C6 carbonyl oxygen, N1, and N3 of guanine instead[43]. The rotation of guanine in the active site results in shorter hydrogen bonds which is thought to increase the binding of the substrate; however PAP has greater catalytic activity for depurinating adenines from rRNA templates about 100x faster than guanine[43].

1.2.3. RIP activity

PAP has been extensively studied for its ability to depurinate the sarcin/ricin loop of ribosomes and de-proteinized rRNA[44]. PAP has broad range activity against many different sources of ribosomes, such as plants, animals, yeast, bacteria, and fungi[45]. Much of PAP's eclectic activity against ribosomes of differing species is due in part to PAP's interaction with a core ribosomal subunit, the L3 protein[46]. Mutational analysis shows that residues close to the active site, 69NN70 and 90FND92 of PAP, are directly involved in binding to the L3 protein on ribosomes, but play no role in the enzymatic activity of PAP[39]. In addition, PAP cannot depurinate ribosomes from yeast

cells containing a mutant of the L3 protein[47]. PAP depurinates an adenine nucleotide from the conserved GAGA tetraloop structure which in turn inhibits binding of elongation factors eEF1 and eEF2 to the ribosome (**Figure 2 a-d**).

Wild type PAP in *P. americana* is tightly controlled in its synthesis and localization in the cell because of its RIP activity. The majority of PAP synthesis is directed to the rough endoplasmic reticulum where it is shuttled to the golgi apparatus to be processed to a mature polypeptide, and finally exported out of the cell into the apoplastic space[48,49]. It is thought that most plant viruses use mechanical breakage of the cell wall to access the host cell, thus a popular hypothesis of PAP's function in pokeweed is to cause ribotoxic stress and apoptosis upon viral infection[18]. Cells infected with a virus by mechanical damage would allow influx of PAP into the cytoplasm, killing the cell by activating ribotoxic stress, and potentially protecting the rest of the organism from systemic infection.

PAP has RIP activity on mammalian ribosomes *in vitro* and *in vivo*, but when PAP is expressed in the cytoplasm of mammalian cells (like HEK 293T or Hela cells) using transient expression vectors, there are no apparent symptoms of toxicity within the cell lines[24,29]. PAP depurinates about 16% of the total 28S rRNA in 293T cells, but [S^{35}]-methionine global translation studies show that translation is not reduced by PAP expression[50]. Depurinated ribosomes trigger a ribotoxic stress response that activates a large family of mitogen activated protein kinases related to proliferation, differentiation, and apoptotic signalling[44]. PAP expression in 293T cells has been shown to activate the MAPK signalling peptides ERK1/2 and c-JNK[50] related to the above pathways, but did not affect global translation rate or cell viability. Interestingly, 293T cells co-transfected

with PAP and HIV-1 show marked reduction of viral particles produced and viral protein expression[24,29]. It is thought that PAP may preferentially target the viral RNA for depuriantion over ribosomal RNA but the mechanism by which PAP is able to do this is not understood.

1.2.4. Antiviral activity

PAP has been demonstrated to have broad range antiviral activity that can inhibit the transmission of many plant viruses other than TMV[19], including Potato virus X (PVX), Cauliflower mosaic virus (CaMV), Alfalfa mosaic virus (AIMV), and Brome mosaic virus (BMV)[51]. Viral RNAs that are pre-treated with PAP *in vitro* result in a reduced capability for the RNAs to be infectious. For example, PAP has been shown to depurinate the RNA3 of BMV, which impedes both the viral RNA replication and sub-genomic RNA transcription[20,26,30,52]. PAP antiviral activity against mammalian viruses has also been reported, including the Human immunodeficiency virus type I (HIV-I)[27], Human T-lymphotrophic virus type I (HTLV-I)[24], Herpes simplex virus (HSV)[23], Poliovirus[23] and Hepatitis B virus (HBV)[53]. *In vivo* studies have shown that PAP lowers the production of HIV-I and HTLV-I proteins, significantly inhibiting the life cycle of these viruses in mammalian tissue culture, without causing cytotoxicity[24,29,50]. Damaged viral mRNAs are not efficiently translated by ribosomes, resulting in a marked reduction in viral protein synthesis. Various therapeutic and agricultural applications of PAP have been explored to utilize PAP as an antiviral agent against human and crop viruses. For example, immunoconjugates of PAP have been synthesized against CD4 positive cells infected with 22 clinical strains of HIV-1, and was able to inhibit viral protein expression

at IC_{50} values in the nanomolar range[54]. Unfortunately, due to the highly immunogenic properties of the protein, the dosage given to rabbit and chimpanzee was ineffective after the first administration[55,56]. Current work on use of PAP as a therapeutic agent is directed toward reducing the immunogenic properties of the protein[44] or mutating the enzyme to be less toxic and more specific to virus or virus expressing cells[32]. PAP has also been expressed in tobacco plants, which show high resistance to a broad-range of viruses that the wild type tobacco was not[49]. While this would produce a crop that is highly resistant to common crop viruses, the expression of PAP resulted in a stunted phenotype which is an unfavourable trade-off for producing a transgenic crop. Interestingly, PAP's antiviral activity can be de-coupled from its RIP activity. The first 16 amino acids of the N-terminus and the last 26 amino acids of the C-terminus of the PAP protein were shown to be important for the recognition of ribosomes for depurination[31]. Deletion of the first 16aa or the last 26aa of mature PAP results in a reduced depurination of ribosomes and cell toxicity in transgenic tobacco. Other mutations have been made in the active site of PAP that are involved in the recognition of the L3 protein of ribosomes[39]. Of these mutations, only the C-terminal deletion retained antiviral activity against a Potato virus X infection of tobacco that was similar to wild type PAP with no observable stunting of the plant[32]. These findings indicate that PAP antiviral activity can be separated from its RIP activity and that both actions do not necessarily act at the same time.

1.2.5. RNA substrate specificity

As a RIP, PAP has been studied primarily for its ability to bind and depurinate the sarcin/ricin loop of the 28S rRNA, but there is not much known about how PAP binds and depurinates off-target RNAs. It is generally recognized that PAP is capable of depurinating purines from multiple RNA substrates such as rRNA, viral RNAs, and various mRNAs[57–59]. The standard substrate for RIP-RNA interaction is the sarcin/ricin loop, and the interaction of the loop in the active site of RIPs has been extensively studied[37,42,60–62] (**Figure 4**). The highly conserved SRL is a short 29 nucleotide stem-loop (GGGUGCUC**AGUACGAGAGGA**ACCGCACCC; bolded region are highly conserved nucleotides across different species[63]) that contains unique secondary structure directly involved in binding of translation elongation factors EF-1 and EF-2 to the ribosome[5,64,65]. Solution structure of the SRL features a bulged G nucleotide that reaches across the major groove and interacts with the phosphate backbone of the opposite strand which is stabilized by two non-canonical A-U/A-G basepairs and cross-stranded adenine base stacking. The SRL also features a stable GNRA-type tetraloop closed by a GC pair that caps the loop[60,63]. Tetraloop formation is generally considered important for RIP interaction because ricin requires this structure to depurinate the 28S rRNA. Mutating either the G or C nucleotide closing the loop severely impacts the ability of ricin chain-A (RTA) to depurinate the RNA, but this is not the case for PAP. PAP is able to depurinate the SRL with or without a closed tetraloop structure[62]. In addition, mutation of the tetraloop sequence from GAGA to GAAA does not affect PAP's ability to depurinate the RNA. The bulged G is also not required for PAP to recognize the stem loop because similar GAGA tetraloops of the 16S eukaryotic rRNA containing entirely

Watson-Crick base pairs can be targeted by PAP for depurination[62]. PAP depurination sites have been identified in the RNA genome of HIV-1[29,50,66], HTLV-1[24], and BMV[20,52]. No structural analyses of these depurination sites have been performed, so it is unknown what kind of structure PAP interacts within these RNA and whether the depurination sites resemble the SRL. It is obvious that PAP has lower RNA specificity because of its broad range N-glycosidase activity against many different RNA substrates, but PAP does not target all substrates equally or depurinate every adenine or guanine base within a given RNA chain. For example, PAP has been shown to depurinate the HIV-1 genome at three discrete locations[29,66,67], but not every purine of the RNA is removed by PAP. In addition, the entire mRNA of GAPDH is not targeted by PAP for depurination[24]. PAP expression in virally infected mammalian cells can be shown to target both the SRL and viral RNA[24], but only the viral protein translation is significantly impaired. One possibility is that viral proteins determine the cellular localization of PAP by targeting the enzyme to viral RNA rather than the ribosomal RNA, but PAP has many interactions that are specific to binding to the ribosome and translating mRNA that counter this. For example, PAP directly binds to the ribosome through the core L3 protein[46]. Another example is that PAP binds to the eukaryotic initiation factor eIF-4G[68], a protein involved in binding to the 5' end of capped mRNA and the assembly of initiating ribosomes. In addition, PAP can tightly bind to the m7G cap of mRNA directly and the binding is enhanced with the addition of eIF4G[58]. Since many viruses need to hijack the host cell translational system in order to replicate, it is possible that PAP interacts with the ribosome in order to gain access to the viral RNAs. The underlying question is how is it that PAP causes significant reduction of viral protein

synthesis and not global protein translation when there is so much evidence demonstrating PAP binding and depurinating the ribosome? One hypothesis is that PAP uses the interaction with the ribosome to gain access to all translating mRNA, including viral RNAs. PAP may specially depurinate viral RNA over cellular RNA by recognizing favourable RNA structures within the viral RNA. Many RNA viruses have a polycistronic genome, consisting of many genes that are regulated by highly complex RNA structures involved in replication, localization, translational stalling, or transcription enhancement[69,70]. For example, the HIV-1 genome encodes for 15 different proteins from single RNA. The HIV-1 genomic RNA contains many structures and elements that are essential for its life-cycle[71], and is typically more complex in structure than cellular mRNA of a mono-cistronic gene in a mammalian cell[72]. As such, it is possible that the highly complex nature of the HIV-1 RNA would provide many structural elements that PAP can target for depurination, more-so than the other cellular RNAs. It can be argued that of the numerous different RNA expressed in the cell, it is not unlikely that PAP can target many different cellular messages that contain a favorable RNA structure for depurination. However, the effect PAP has on viral replication is still much greater than the effect it has on global translation, even if PAP targets other cellular mRNAs as it does the viral RNAs.

1.3. Goal of project

The heterologous expression of PAP in virus infected cells inhibits virus production without affecting translation of cellular mRNAs. PAP shows a novel anti-retroviral activity against the Human immunodeficiency virus - 1 (HIV-1) when it is

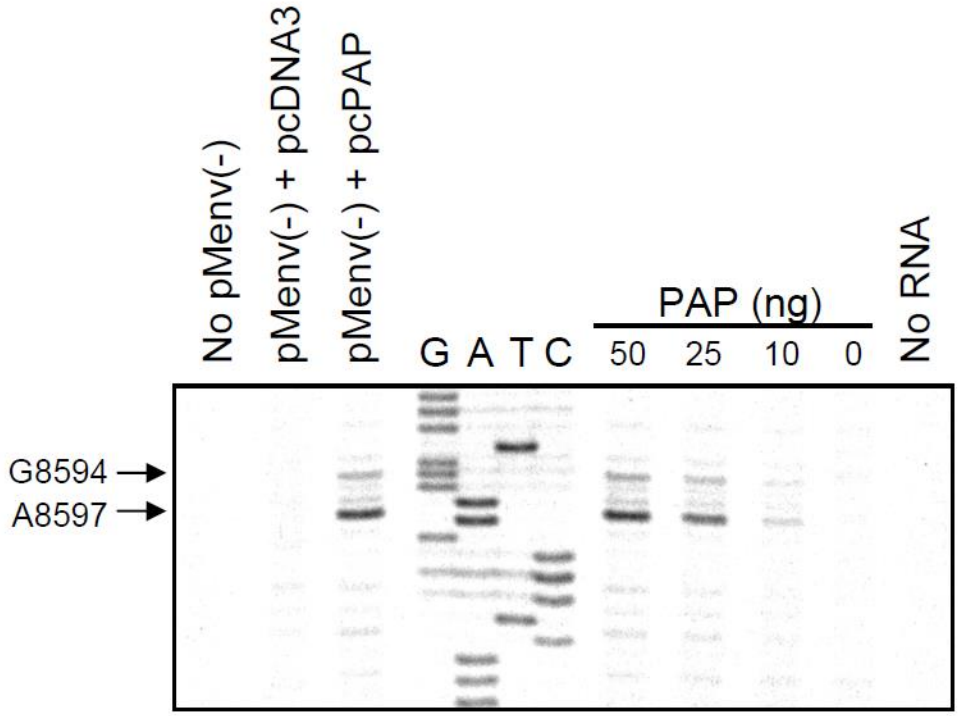


Figure 6 – PAP depurinates the Rev open reading frame of the HIV-1 2KB mRNA *in vivo.* 293T cells were co-transfected with pMenv$^{(-)}$ and pcPAP or pcDNA3. Cells were harvested 40 hours post transfection and 2 Kb RNAs were gel purified. 2Kb RNAs were also isolated from 293T cells transfected with only pMenv$^{(-)}$ and incubated with (50, 25, 10, 0 ng) of PAP *in vitro*. Primer extension was performed using an Rev specific radiolabeled reverse primer complementary to nucleotides 8648-8668 of the HIV-1 pMenv$^{(-)}$ proviral DNA (K03455.1). These data were obtained from previous work in our lab[29].

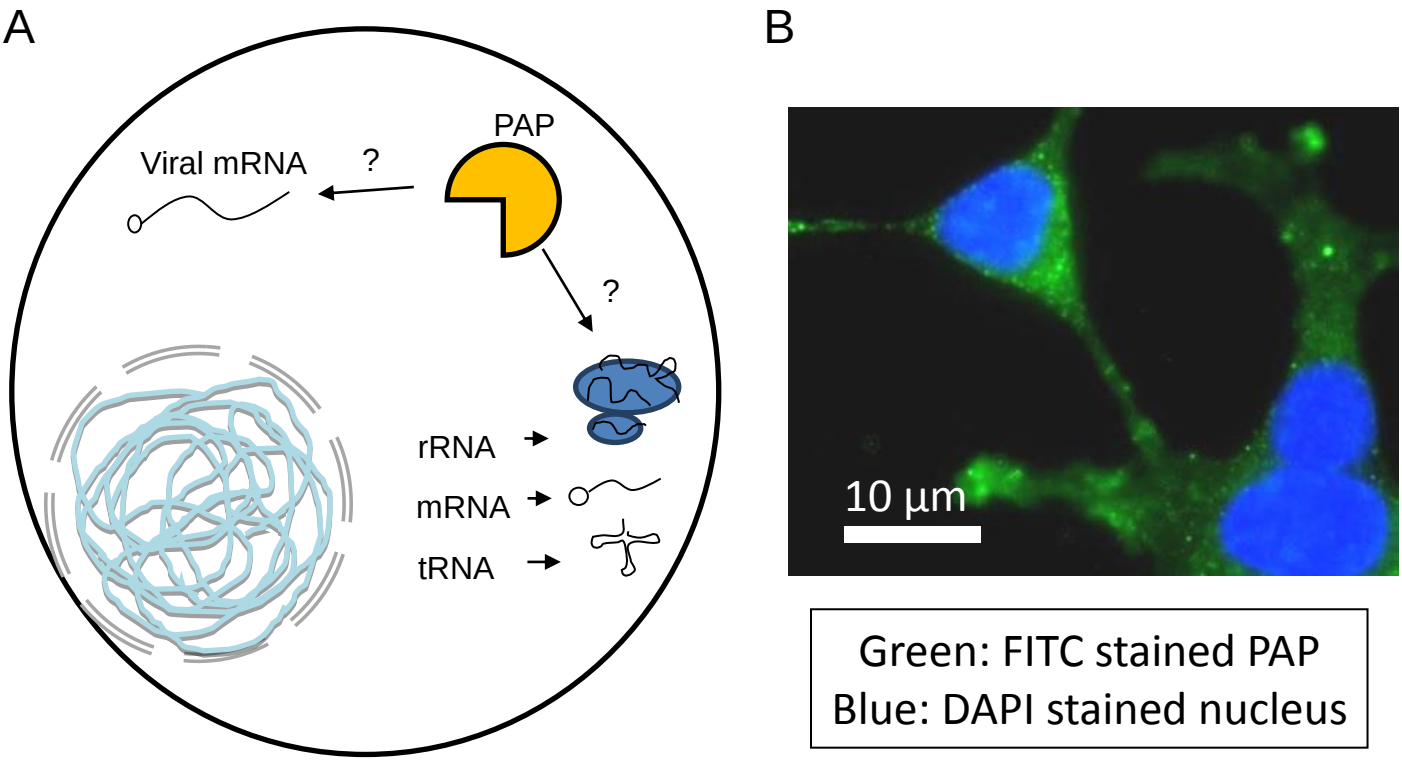


Figure 7– General model of a mammalian cell expressing PAP in the presence of viral RNA. A) PAP expressed in the cytoplasm of mammalian cells is not toxic but severely down-regulates the production of viral particles. PAP is in the presence of other cellular RNAs as well, so how PAP can specifically target viral RNA over cellular RNA is not well understood. **B)** 293T cells transfected with pcPAP were fixed and immunostained with mouse anti-flag antibody specific for the flag tagged PAP and rabbit anti-mouse IgG fused to FITC in order to visualize PAP in the cytoplasm. The chromosomal DNA was stained with DAPI and images were taken using a confocal microscope at 1000 x magnification under 380 and 485 nm light to visualize blue and green fluorescence. The white bar represents 10 µm length.

expressed in the cytoplasm of 293T cells, without causing toxicity[27,50]. More recently, PAP has been shown to depurinate regions of the mRNA for HIV-1 regulator of virion expression (Rev) *in vivo*[29] (**Figure 6**), which is a protein essential for HIV-1 transcripts to be exported from the nucleus for translation. At low concentrations, it is hypothesized that PAP is able to reduce the expression of viral proteins by depurinating viral RNA over cellular RNA, but it is not well understood how PAP can make this distinction. A schematic of the situation is shown in **Figure 7A** and **Figure 7B** shows the location of PAP in the cytoplasm of 293T cells. Although it is well known that PAP has a broad range of RNA targets, little work has been done to identify what sequence or structure of RNA PAP preferentially interacts with, besides the sarcin/ricin loop, and if this has any relationship to PAP specifically targeting viral RNA over cellular RNA. This research was conducted to gain insight into the mechanism by which PAP can target a viral RNA over a cellular RNA by observing how PAP binds to HIV-1 Rev RNA. Specifically, I attempt to distinguish the importance of either a conserved primary sequence or a secondary structure of RNA in the binding of PAP to Rev RNA. **I hypothesize that Rev RNA has distinct secondary structure that PAP specifically binds to that is similar to the SRL.**

In order to test this hypothesis, a combination of *in vitro* binding, footprinting, and structural assays of the interaction between PAP and Rev RNA were conducted. Binding assays were conducted to test for PAP's ability to bind to Rev RNA and if PAP can selectively bind one RNA over another. Wild type and mutant versions of Rev RNA were used as substrate for PAP binding and compared to representative cellular RNAs. A portion of the human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) RNA

was used as a negative control for binding because it was shown previously by our lab to not be depurinated by PAP[24]. The human arginine tRNA (Arg-tRNA) was used because of its unique secondary structure and that PAP is not known to bind to this RNA. The human SRL was used as a positive control for binding. All binding assays were done without a 5'-cap in hopes to observe PAP interaction with unmodified RNA alone. Dimethyl sulphate and 1-methyl-7-nitroisatoic anhydride footprinting assays were conducted in order to determine the location of PAP binding to the Rev RNA. Chemically modified nucleotides of folded RNA provide information of the location of single-stranded nucleotides around the PAP binding site and was used in combination with minimum free energy secondary structure folding programs to determine most likely secondary structure. 3D modeling of the final PAP binding site was conducted to predict the structure of RNA PAP interacts with. Identifying the RNA elements of PAP binding site in Rev RNA will help in understanding how PAP specifically binds to an RNA substrate and further elucidate the mechanism of how PAP targets viral RNA.

2. Materials and methods

2.1. Plasmids and cDNA genes

A brief description of each plasmid used throughout the thesis has been outlined below:

pBluescriptSK(-): plasmid contains lacZα gene under the control of the lac promotor/operon, and encodes the Amp resistance gene. The plasmid was used for cloning and identification of successful gene insertions after IPTG induction on media containing X-α-galactose.

pcDNA3: plasmid contains the cytomegalovirus (CMV), T7, and SP6 promoters and encodes ampicillin (Amp) resistance gene. The plasmid was used as a promoter control vector for transfection and was used as a vector backbone for other gene products.

pcYFP: plasmid encoding the yellow fluorescent protein downstream of the CMV promoter and encodes the Amp resistance gene. The plasmid was used in all mammalian cell transfections in order to test transfection efficiency.

pcPAP: plasmid contains the fully mature form of PAP gene downstream of the CMV promoter and encodes the Amp resistance gene. The plasmid was used to express PAP gene in mammalian tissue culture.

pGEM9z: plasmid contains lacZβ gene under the control of the lac promotor/operon, and encodes the Amp resistance gene. The plasmid was used to increase and equalize the total DNA mass of transfection reactions between transfections.

pMenv$^{(-)}$: plasmid contains the full 9.4Kb proviral HXB2 HIV-1 genome under control of the CMV and T7 promoters and the Amp resistance gene.

pcRev: plasmid contains 369 nucleotide HXB2 spliced Rev ORF under control of the CMV and T7 promoters and contains the Amp resistance gene. The gene is also flanked with an SP6 promoter designed to transcribe the reverse complement RNA for preparing RNA probes.

pcRev291: plasmid contains369 nucleotide HXB2 spliced Rev ORF with three guanidine residues between 291 and 293 nucleotides mutated to three adenine residues. The ORF is under control of the CMV promoter and the plasmid also contains the Amp resistance gene.

pcHSRL: pcDNA3 plasmid containing a 322 base pair fragment of the 28S human rRNA containing the sarcin/ricin loop domain. The RNA transcription is under control of the CMV and T7 promoters and the plasmid also contains the Amp resistance gene. The gene is also flanked with an SP6 promoter designed to transcribe the reverse complement RNA for preparing RNA probes.

H28SRNA: pcDNA3 plasmid containing a 500 base pair fragment of the 28S human rRNA containing the sarcin/ricin loop domain. The RNA transcription is under control of the CMV and T7 promoter and the plasmid also contains the Amp resistance gene.

Arg-tRNA: Short 120 base pair, mature form of the human argenine-tRNA used as a negative control for RNA binding. Prepared as a PCR product containing a T7 promoter in order to be transcribed to RNA for binding studies.

2.2. PAP isolation from *Phytolacca americana*

2.2.1. Ammonium sulphate fractionation

150 g of frozen pokeweed spring leaves were blended in 300 mL of pre-chilled grinding buffer (50 mM Tris HCl pH 7.4, 2 mM EDTA, 30 mM β-mercaptoethanol) and 3 g of polyvinylpolypyrrolidone until solution was a smooth slurry. The mixture was filtered through 4 layers of cheese cloth into a beaker on ice. While constantly stirring the filtrate on ice, 75.95 g of anhydrous ammonium sulphate $(NH_4)_2SO_4$ (to a concentration of 55% w/v) was added over a 15 minute period. The resulting solution was stirred for 3 hours at 4°C before centrifuging at 15,000 x g at 4°C for 15 minutes. The supernatant was stirred on ice while adding 64.97 g of $(NH_4)_2SO_4$ to a final concentration of 95% w/v. This mixture was left to precipitate overnight for 18 hours before centrifuging at 10,000 x g at 4°C for 20 minutes. Precipitated protein was retained and dissolved in 30 mL of Q-buffer (20 mM Tris HCl pH 7.5, 0.2 mM EDTA, 1mM β-mercaptoethanol) and dialyzed in 2 litres of Q-buffer across 14 kDa cellulose membranes (Spectrapore) for 2 days with 6 separate exchanges of buffer. The final protein extract was syringe filtered through a 0.45µm nylon membrane (Fisher Scientific) and degassed in a vacuum chamber at 0.8 bar for 20 minutes in preparation for FPLC.

2.2.2. Ion exchange chromatography

FPLC (Amersham Bioscience) of 30 mL of the protein extract was eluted through a 5 mL HiTrap Q HP anion exchange column (GE healthcare) pre-equilibrated with Q-buffer and was collected for cation exchange. Protein was subsequently passed through a 5 mL HiTrap SP HP cation exchange column (GE healthcare) pre-

equilibriated with S-buffer (10 mM MES pH 5.2, 0.2 mM EDTA, 1 mM β-mercaptoethanol) and washed with 5 column volumes of S-buffer. Positively charged proteins bound to the column were fractioned using a NaCl gradient from 0 to 350 mM, with the predominant PAP band eluting at 150 mM. Fractions containing PAP (analyzed by SDS-PAGE and silver staining) were collected and concentrated in 50 kDa concentration tubes (Amicon Ultra) and resuspended in 2 mL of PAP storage buffer (20 mM Tris HCl pH 7.5, 0.2 mM EDTA, 1 mM β-mercaptoethanol, 35% glycerol). All peak monitoring of protein fractions was conducted using the Unicorn 5.0 program.

2.2.3. PAP RNase activity assay

PAP protein isolates were analyzed for purity by SDS-PAGE and for RNase activity. A gradient of BSA, PAP, and RNase A protein were separated through a 12%acrylamide-SDS gel containing sheared torulla yeast RNA at 100 V for 1 hour. The gel was first coomassie stained in order to estimate protein size and purity relative to a control PAP isolation. RNases in the gel are activated by first washing off excess destain solution in 10 mM Tris HCl pH 7.0 twice, each wash for 5 minutes and then rocked for 30 minutes in 100 mM Tris HCl pH 7.0 at 50^{o}C. The RNase would digest the Torulla yeast RNA in the gel where protein has migrated. Finally, the gel was stained with RNA staining solution (10 mM Tris HCl pH 8.3, 0.2% methylene blue) for 5 minutes at room temperature, and destained in 10 mM Tris HCl pH 8.3 until desired contrast was achieved. RNase activity was visualized as an unstained region surrounding the protein band.

2.3. Plasmid transformation

All plasmids used in this study were transformed into DH5α *E. coli* bacteria for storage and amplification. Approximately 100 ng of plasmid DNA was added to 100 µL of competent cells and incubated on ice for 30 minutes. The bacteria were lightly mixed and then incubated in a 42^{o}C water bath for 1 minute and then back on ice for 10 minutes. Then, 300 µL of Luria-Bertani (LB) broth were mixed with the cells and incubated at 37^{o}C with shaking at 100 rpm for 1 hour. The culture was centrifuged at 5,000 x g for 10 seconds and the top 300 µL supernatant was discarded. The bacterial pellet was resuspended in the remaining supernatant and spread onto LB/Agar plates pre-coated in 20 µL of 100 µg/µL ampicillin (AMP). The culture plates were incubated at 37^{o}C for 18 hours and single colonies were selected for amplification of the plasmid and storage in glycerol stocks.

2.4. Generation of Rev and sarcin/ricin loop clones

2.4.1. Generation of pcHSRL

Primers were designed to amplify the human sarcin/ricin loop between nucleotides 4436 and 4757 from the 28S ribosomal RNA (accession number M11167) containing the HindIII restriction site on the forward primer (F-HindIII-HSRL-4456) and a BamHI restriction site on the reverse primer (R-BamHI-HSRL-4757). Polymerase chain reaction (PCR) was prepared as follows: to a pcr tube, 100 ng of KH004 plasmid, 200 ng of forward and reverse primer, 5 µL of 10 x thermopol buffer (200 mM Tris HCl pH 8.0, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton X-100) (New England Biolabs), 2 µL of 10 mM dNTPs, 2.5 units of Taq polymerase (New England Biolabs),

and 40 µL of milli-Q water was added. The PCR reaction was conducted with a preliminary 5 minute denaturation step at 95^{o}C for 5 minutes, followed by 30 amplification cycles, denaturing at 95^{o}C for 30 seconds, annealing at 65^{o}C for 20 seconds, and extending at 72^{o}C for 45 seconds. A final extension was carried out at 72^{o}C for 5 minutes. The reaction product was extracted in phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitated before running 1 µg on a 1% agarose gel next to a 1Kb ladder (New England Biolabs) to check for the correct size product. Next, 10 µg of the PCR product and pcDNA3 were digested using BamHI and HindIII restriction enzymes (New England Biolabs) in the following way: 10 µg of DNA was incubated with 40 units of BamHI and 40 units of HindIII enzyme in 1 x Buffer 2 (50 mM NaCl, 10 mM Tris HCl pH 7.9, 1 mM DTT, 10 mM $MgCl_2$) in a 50 µL volume for 2 hours at 37^{o}C. Next, 10 units of calf intestinal phosphatase (New England Biolabs) was added to the pcDNA3 digestion reaction and was incubated at 37^{o}C for an additional 2 hours. 15 µL of 5 x DNA loading dye (25% glycerol, 30 mM Tris HCl pH 8.0, 100 mM EDTA, 0.025% bromophenol blue) was added to each reaction and was separated on a 0.8% low melt grade agarose gel (Bio Shop) stained with ethidium bromide. Single bands of digested DNA were excised from the gel over a UV box and placed into a pre-weighed 1.5 mL microtube. Equal volumes of 2 x low melt buffer (40 mM Tris HCl pH 7.9, 20 mM EDTA) were added and incubated at 60^{o}C for 20 minutes to melt the agarose gel. Equal volumes of Tris buffered phenol (pH 8.0) to the total solution were added and vortexed for 2 minutes and centrifuged at 16,000 x g for 5 minutes at room temperature. 80% of the top supernatant was transferred to a new 1.5 mL tube containing equal volumes of Tris buffered phenol:chloroform:isoamyl alcohol 25: 24:1 (PCI) (pH 8.0), was vortexed

and was centrifuged at 16,000 x g for 5 minutes at room temperature. The top supernatant was transferred to a new 1.5 mL tube containing 2.5 volumes 95% ethanol and 1/10 volumes 3 M $NaC_2H_3O_2$ and was incubated at -20^oC for 18 hours. The DNA was pelleted by centrifuging at 16,000 x g for 30 minutes at 4^oC and washed with 70% ethanol before air drying at room temperature for 5 minutes and resuspending in 10 µL of milli-Q water. The digested sarcin/ricin loop and pcDNA3 vector were ligated in the following reaction: 2 µL of insert, 1 µL of vector backbone, 2 µL of 10 x ligation buffer (500 mM Tris HCl pH 7.5, 100 mM $MgCl_2$, 10 mM ATP, 100 mM DTT), 400 units of T4 DNA ligase (New England Biolabs), 14 µL of milli-Q water were combined into a 1.5 mL tube on ice and was incubated at 16^oC for 18 hours. After ligation, the reaction mixture is transformed into *E.coli* and plated onto ampicillin coated LB/agar plates following the procedure outlined in section **2.3**.

2.4.2. Generation of Rev ORF fragments

Primers were designed to amplify fragments of the Rev ORF in order to observe the binding interaction PAP has with separate regions of Rev. Fragment A is a 162 base pair RNA containing nucleotides 1 to 146 of the Rev ORF. The region was amplified using primers T7forward and RevORF146R with the pcRev plasmid as a template. Fragment B is a 175 base pair RNA containing nucleotides 89 to 264 of the Rev ORF. The region was amplified using primers T7RevORF89F and RevORF264R. Fragment C is a 153 base pair RNA containing nucleotides 198 to 351 of the Rev ORF. The region was amplified using primers T7RevORF198F and RevORF351R. Fragment AB is a 280 base pair RNA containing nucleotides 1 to 264 of the Rev ORF. The region was

amplified using primers T7forward and RevORF264R. Fragment BC is a 262 base pair RNA containing nucleotides 89 to 369 of the Rev ORF. The region was amplified using primers T7RevORF89F and RevORF351R. PCR was prepared as follows: to a pcr tube, 100 ng of pcRev plasmid, 200 ng of forward and reverse primer, 5 µL of 10 x thermopol buffer (200 mM Tris HCl pH 8.0, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton X-100) (New England Biolabs), 2 µL of 10 mM dNTPs, 2 µL of Taq polymerase, and 38 µL of milli-Q water was added. All PCR reactions for the above combinations were conducted with a preliminary 5 minute denaturation step at 95^{o}C for 5 minutes, followed by 30 amplification cycles, denaturation at 95^{o}C for 30 seconds, annealing at 55^{o}C for 30 seconds, and extension at 68^{o}C for 25 seconds. A final extension was carried out at 68^{o}C for 5 minutes. The final reaction was extracted with PCI and ethanol precipitated as described earlier. PCR reactions were checked on a 1% agarose gel for the correct size product before moving on to transcription reactions.

2.4.3. Generation Rev 291 mutant

Primers were designed to amplify and mutate the Rev ORF, specifically the 291-293 nucleotides, in order to observe a change in the binding interaction PAP has with this particular site. The first set of primers amplified the nucleotides 1 and 310 of the Rev ORF and the forward primer also contained the HindIII cut site (F-HindIII-Rev-1) and the reverse contained a CCC to UUU mutation at bases 291-293 (R-Rev-G291-293A). The second set of primers amplified the nucleotides 276-369 and 200 nucleotides of the pcDNA3 backbone and the forward primer contained a GGG to AAA mutation at bases 291-293 (F-Rev-G291-293A) and the reverse primer was specific to

the pcDNA3 backbone in pcRev (pcDSR). PCR was prepared as follows: to a pcr tube, 100 ng of pcRev plasmid, 200 ng of forward and reverse primer, 5 µL of 10 x Q5 reaction buffer (250 mM TAPS-HCl pH 9.3, 500 mM KCl, 20mM $MgCl_2$, 10 mM β-mercaptoethanol) (New England Biolabs), 2 µL of 10 mM dNTPs, 1 unit of Q5 high-fidelity DNA polymerase (New England Biolabs), and 38 µL of milli-Q water was added to a final volume of 50 µL. Both PCR reactions for the above combinations were conducted with a preliminary 5 minute denaturation step at 95^{o}C for 5 minutes, followed by 30 amplification cycles, denaturing at 95^{o}C for 30 seconds, annealing at 60^{o}C for 30 seconds, and extension at 72^{o}C for 30seconds. A final extension was carried out at 72^{o}C for 5 minutes. The final reaction was extracted with PCI and ethanol precipitated and then extracted from a 0.8% low melt gel as described earlier. An over-lapping PCR reaction was prepared as follows: to a pcr tube, 200 ng of the first PCR reaction, 200 ng of the second PCR reaction, 200 ng of F-HindIII-Rev-1 forward primer, 200 ngR-pcDNA3-DSR reverse primer, 5 µL of 10 x Q5 reaction buffer (250 mM TAPS-HCl pH 9.3, 500mM KCl, 20mM $MgCl_2$, 10 mM β-mercaptoethanol) (New England Biolabs), 2 µL of 10 mM dNTPs, 1 unit of Q5 high-fidelity DNA polymerase (New England Biolabs), and 28 µL of milli-Q water was added to a final volume of 50 µL. The over-lapping PCR reaction was conducted with a preliminary 5 minute denaturation step at 95^{o}C for 5 minutes, followed by 6initial amplification cycles, denaturing at 95^{o}C for 30 seconds, annealing at 65^{o}C for 45 seconds, and extending at 72^{o}C for 45 seconds and reducing the annealing temperature by 1^{o}C each round. This was followed by 25 amplification cycles, denaturation at 95^{o}C for 30 seconds, annealing at 59^{o}C for 45 seconds, and extension at 68^{o}C for 45 seconds. A final extension was carried out at 72^{o}C for 5

minutes. The final reaction was extracted with PCI and ethanol precipitated as described earlier. The overlapped product and pcDNA3 was digested with HindIII and XbaI in the following way: 10 µg of DNA was incubated with 40 units of HindIII and 40 units of XbaI enzyme in 1 x Buffer 2.1 (50 mM NaCl, 10 mM Tris HCl pH 7.9, 1 mM DTT, 10 mM $MgCl_2$, 100 µg/mL BSA) in a 50 µL volume for 2 hours at 37ºC. Next, 10 units of calf intestinal phosphatase (New England Biolabs) was added to the pcDNA3 digestion reaction and was incubated at 37ºC for an additional 2 hours. Digested DNA products were purified by low-melt agarose gel as described earlier. The digested 291 Rev mutant and pcDNA3 vector were ligated in the following reaction: 2 µL of insert, 1 µL of vector backbone, 2 µL of 10 x ligation buffer (500 mM Tris HCl pH 7.5, 100 mM $MgCl_2$, 10 mM ATP, 100 mM DTT), 400 units of T4 DNA ligase (New England Biolabs), 14 µL of milli-Q water were combined into a 1.5 mL tube on ice and was incubated at 16ºC for 18 hours. After ligation, the reaction mixture was transformed into *E.coli* and plated onto ampicillin coated LB/agar plates following the procedure outlined in **2.3**.

2.5. Cell culture

HEK 293T cells were cultured in growth media containing Dubelco's Modified Eagle Medium (DMEM), with 10% fetal bovine serum (FBS) and penicillin/streptomycin (10mg/mL). The cells were grown in 10cm polystyrene treated plates (Corning) and kept at 37ºC and 5% CO_2untill 80-90% confluent. Once cells had reached the optimal confluency, they were washed with 5 mL of sterile 1x phosphate buffered saline (PBS) (135 mM NaCl, 5 mM Na_2HPO_4, 2.5 mM KCl, 2 mM KH_2PO_4 pH 7.4) at room temperature. After aspirating the media, 1 mL of Trypsin solution (0.25% Trypsin, 0.1%

EDTA) was added to the plate for 1 minute at room temperature to release adherent cells from the plate. Then, 9 mL of growth media was added and cells were gently resuspended. These cells were seeded into a new culture dish as a 1:5 dilution in growth media and incubated for 48 hours before repeating the dilution.

2.6. Plasmid isolation via alkaline lysis

4 mL of LB broth containing 100 µg/mL of AMP was inoculated with a bacterial colony of *E.coli* containing the plasmid of interest and incubated at 200 rpm at 37°C for 18 hours. 1.5 mL of culture was aliquoted into a 1.5 mL microtube and centrifuged at 5,000 x g for 30 seconds. The top supernatant was discarded and an additional 1.5 mL of culture was added to the top of the pellet, centrifuged at 5,000 x g for 30 seconds, and the supernatant was discarded again. Next, 200 µL of ice cold solution I (100 mM Tris HCl pH 7.5, 10 mM EDTA, 0.4 mg/mL RNase A) was added to the pellet and mixed vigorously until homogenous. Then, 200 µL of solution II (200 mM NaOH, 1% SDS) was added and mixed gently at room temperature for 4 minutes. Next, 200 µL of ice cold solution III (3 M KCH_3CO_2, 2 mM acetic acid) was added, vortexed, and the sample was placed on ice for 10 minutes. The supernatant was transferred to a new tube, 420 µL of isopropanol was added, and the DNA precipitated for 30 minutes at room temperature. The precipitating DNA was centrifuged at 13,000 x g for 20 minutes at 4°C, decanted and air dried for 10 minutes before resuspending the pellet in 200 µL of 1 x TE. 200 µL of PCI was added, vortexed, and centrifuged for 5 minutes at 16,000 x g. The top aqueous phase was transferred to a new tube and 20 µL of 3 M NaH_3CO_2 pH 4.3 and 500 µL of 95% ethanol was added and precipitated at 4°C for 18 hours. The

precipitating solution was centrifuged at 16,000 x g for 5 minute, decanted, and the DNA pellet air dried for 5 minutes before resuspending in DEPC treated milli-Q water. Plasmid DNA was quantified using a UV spectrophotometer (Beckman).

2.7. Transient transfection

HEK 293T cells were seeded in DMEM at a 1:5 ratio onto 10cm polystyrene treated plates (Corning) and incubated at 37ºC and 5% CO_2 for 24 hours until 50-70% confluent. Cells were transfected with 15 µg of total DNA including 1 µg of pcYFP, to observe transfection efficiency, using the calcium phosphate method. Plasmid DNA was first diluted into 405 µL of dH_2O before 45 µL of 2.5 M $CaCl_2$ was gently added by pipetting. The mixture was incubated at room temperature for 5 minutes before adding dropwise to 450 µL of 2X HeBS buffer (350 mM NaCl, 60 mM HEPES pH 7.15, 1 mM Na_2HPO_4) with continuous, gentle vortexing. The mixture was incubated at room temperature for 20 minutes before dropwise addition to the tissue culture. The cells were incubated at 37ºC and 5% CO_2 for 18 hours before observing green fluorescence (520 nm) under blue light (485 nm) of cells expressing YFP for estimating transfection efficiency. The tissue culture was washed with 5 mL of PBS and 10 mL of fresh media was added to the cells. These cells were incubated for 24 hours at 37ºC and 5% CO_2 before harvesting.

2.8. Immuno-staining PAP in HEK 293T

HEK 293T cells were seeded as a 1:10 ratio of fully confluent cells into a 6 well polystyrene treated plates (Corning) containing an autoclaved 18 x 18 mm glass

coverslip. The culture was grown at 37ºC and 5% CO_2 for 24 hours until 30-50% confluent. Cells were transfected with 7.5 µg of total DNA using the calcium phosphate method described earlier. Plasmid DNA was first diluted into 202.5 µL of dH_2O before 22.5 µL of 2.5 M $CaCl_2$ was gently added by pipetting. The mixture was incubated at room temperature for 5 minutes before adding dropwise to 225 µL of 2X HeBS buffer with continuous, gentle vortexing. The mixture was left to incubate at room temperature for 20 minutes before dropwise addition to the tissue culture. The cells were incubated at 37ºC at 5% CO_2 for 18 hours before washing with 1 mL of PBS and adding 5 mL of fresh media. After 24 hours, cells were washed with 1 mL of PBS and 1 mL of Fixative solution was added (4% paraformaldehyde, 100 mM PIPEs pH 7.0, 2 mM $MgCl_2$, 1.25 mM EGTA) and left to incubate at room temperature for 20 minutes. The fixative solution was washed 3 times with 1 mL of PBS and 1 mL of permeabilizing solution was added (0.2% triton x-100, 1 x PBS, 10 mM $MgCl_2$) and left at room temperature for 10 minutes. The permeabilizing solution was washed 3 times with 1 mL of MgPBS (1 x PBS, 10 mM $MgCl_2$) and the glass cover slips were wicked dry before placing face down onto parafilm spotted with 50 µL of normal goat serum blocking solution (1 x PBS, 10 mM $MgCl_2$, 5% normal goat serum) for 30 minutes at room temperature. After blocking, the cover slip was placed on 50 µL of primary antibody for PAP (1:200, polyclonal rabbit anti-PAP) in protein block solution (3% BSA, 1 x PBS) for 4 hours at room temperature. Excess antibody was washed off using three washes of PBS and was then placed on 50 µL of fluorescein isothiocyanate (FITC) labelled secondary antibody for rabbit (1:500, monoclonal mouse-anti-rabbit) in protein block solution for 2 hours at room temperature. The coverslip was washed three times with MgPBS and wicked dry before immersing in

antifade solution (Life Technologies) containing 125 ng/mL of 4',6-diamidino-2-phenylindole (DAPI) and was sealed to a glass slide. The cells were stored in dark at 4^{o}C for at least 4 hours before imaging on a confocal microscope for DAPI and FITC fluorescence.

2.9. Harvesting cells

48 hours after transfection, cells were washed with 5 mL of ice cold 1 x PBS solution and resuspended in 1 mL of ice cold 1 x PBS before collecting into a 1.5 mL microtube. The cells were centrifuged at 9,000 x g for 4 minutes at 4^{o}C and the excess PBS was aspirated away. The pellet was resuspended with equal volume (usually between 150-200 µL) of cytoplasmic lysis buffer (25 mM HEPES pH 7.5, 2 mM EGTA, 1 mM DTT, 10% glycerol, 1% NP-40), vortexed for 1 minute and incubated on ice for 10 minutes. Insoluble cell debris was pelleted by centrifuging at 16,000 x g for 15 minutes at 4^{o}C and the supernatant was retained. An aliquot of the supernatant was retained for western blot analysis while the remainder was further purified for total RNA content. To a new 1.5 mL microtube ,100 µL of the supernatant was transferred, brought to 200 µL volume with 1 x TE, and 500 µL of RNAzol-RT (Bioshop) was added and vortexed for 2 minutes before incubation on ice for 10 minutes. The samples were centrifuged at 4^{o}C at 16,000 x g for 10 minutes. The top 80% of the supernatant was transferred to a new 1.5 mL tube containing 500 µL of isopropanol. The mixture precipitated at -20^{o}C for 2-18 hours before centrifuging for 20 minutes at 4^{o}C at 16,000 x g. The supernatant was decanted and the pellet was air dried for 5 minutes before resuspending with 200 µL of 1 x TE. The RNA solution was extracted first with 200 µL of Tris buffered phenol pH 8.0

and then with 200 µL of citrate buffered phenol:chloroform:Isoamyl alcohol (25:24:1) pH 4.5. The supernatant was transferred to a new 1.5 mL tube containing 20 µL of 3M sodium acetate and 500 µL of 100% ethanol. The RNA was precipitated at -80^{o}C for 2-18 hours before pelleting the RNA by centrifugation at 16,000 x g for 30 minutes at 4^{o}C. The pellet was washed with 70% ethanol and centrifuged again at 16,000 x g for 15 minutes at 4^{o}C. The supernatant was decanted and the pellet air dried for 5 minutes before resuspention in 20 µL of 0.5 x TE. RNA concentration was quantified using a spectrophotometer and the quality of the RNA was checked by running 1 µg on a 1% agarose gel and observing the 28S, 18S, and 5.8S, and 5S rRNA.

2.10. Western-blot analysis

The concentrations of the protein samples obtained from harvested 293T HEK cells were first calculated using a Bradford assay (OD_{595} nm) and aliquoted into 150 µg samples. Samples were then brought to 10 µL using lysis buffer and 10 µL of 2 x SDS loading buffer (125 mM Tris HCl pH 6.8, 20% glycerol, 1.3 M β-mercaptoethanol, 4% SDS, 0.25% bromophenol blue) was added. Samples were boiled for 5 minutes before they were loaded onto a 12% acrylamide gel and separated via SDS-PAGE in running buffer (192 mM glycine, 25 mM Tris HCl pH 8.3, 0.1% SDS) at 200 V for 30 minutes. After separation, the gel was placed onto nitrocellulose membrane (GE healthcare) pre-rinsed in cold western transfer buffer (160 mM glycine, 25 mM Tris HCl pH 8.3, 20% methanol, 0.02% SDS). The gel and membrane were then sandwiched between four pieces of filter paper (Fisher Scientific) which have also been pre-rinsed in cold western transfer buffer. The proteins were transferred to the nitrocellulose membrane using a

semi-dry gel transfer apparatus (BIO-RAD) for 40 minutes at 0.12 Amps per membrane. After transfer, the membrane was blocked in 5% milk in PBS-T (1 x PBS, 0.1% Tween 20) for 2 hours while rocking at room temperature. The blocking media was replaced with 5 mL of fresh 5% milk in PBS-T and both Rev antibody (1:100; monoclonal, NIH 7376) and β-actin antibody (1:5000; monoclonal, Sigma Aldrich F3022) was added to rock for 18 hours at 4^{o}C. The membrane was then washed with PBS-T for 10 minutes with rocking a total of three times before adding 5 mL of 5% milk in PBS-T and horse-radish peroxidase (HRP) conjugated secondary antibody (1:5000; monoclonal, Sigma Aldrich A9044) for 1.5 hours. The membrane was then washed with PBS-T for 10 minutes with rocking a total of three times and was blotted dry with a Kim Wipe. The Rev and β-actin proteins were visualized by adding 1.5 mL of chemiluminescence reagent (Perkin Elmer) for 1 minute at room temperature. The membrane was wrapped in cellophane and a clear blue x-ray film (Thermo Scientific) was placed against the blot in dark and left to expose for 10 minutes before developing.

2.11. Primer extension analysis

2.11.1. End Labeling Primers

Primers were end labelled with [γ^{32}P] for visualizing cDNA products after a reverse transcription reaction. Briefly, 200 ng of DNA primer (synthesized from Sigma Aldrich) was combined with 2 µL of 10 x polynucleotide kinase buffer (PNK) (700 mM Tris HCl pH 7.6, 100 mM $MgCl_2$, 50 mM DTT), 10 units of T4 PNK enzyme (New England Biolabs), 2 µL of [γ^{32}P]-ATP (Perkin Elmer), and brought to a final reaction volume of 20 µL with milli-Q water. The reaction mixture was incubated at 37^{o}C for 30 minutes before bringing the final reaction volume to 50 µL with milli-Q water and was

filtered through a centrifuge column (Thermo Scientific) containing 700 µL of G-50 Sephadex solution (GE Healthcare) at 2,600 x g for 2 minutes at room temperature. The flow through was collected and the radioactivity was calculated using a scintillation counter (Perkin Elmer 2910 TR).

2.11.2. Primer extension reaction

Either 20 µg of total RNA isolated from 293T or 8 µg of treated *in vitro* transcribed RNA was dissolved in 8 µL of DEPC water in preparation for reverse transcription. To this RNA, 2 µL of 5 x first strand buffer (FSB) (75 mM KCl, 50 mM Tris HCl pH 7.9, 10 mM DTT, 3 mM $MgCl_2$) and 1 x 10^6 cpm of γ-^{32}P radiolabeled primer was added and heated to 95^oC for 2 minutes and set on ice for 1 minute. The primer was left to anneal with the RNA template at room temperature for 20 minutes before adding 9 µL of RT master mix (2 µL 5 x FSB, 0.8 µL of 0.5 M DTT, 2 µL of 10 mM dNTP (New England Biolabs), 40 units of Murine RNase inhibitor (New England Biolabs), 5 units of Super Script III reverse transcriptase (SSIII) (Life Technologies), and 3.2 µL of DEPC treated H_2O), centrifuged to collect solution to bottom of the tube, and set to 50^oC for 1 hour to extend. The reaction was stopped by adding 10 µL of 2 x RNA loading buffer (95% formamide, 20 mM EDTA, 0.25% bromophenol blue, 0.25% xylene blue) and was heated to 95^oC for 5 minutes and chilled on ice for 2 minutes before loading onto a large denaturing 7M Urea/ 8% polyacrylamide electrophresis gel. A sequencing ladder was separated beside the primer extension reactions in order to determine the nucleotide locations of each termination. The cDNA products were separated at 30 watts for 3.5 hours, vacuum dried, and screened against a phosphor imaging screen

(Bio-Rad) to visualize the cDNA products. The phosphor screen was scanned using a Typhoon Trio (GE healthcare) phosphor imager.

2.11.3. Sequencing ladder

Between 7-14 µg of plasmid DNA was incubated with 0.2 M NaOH and 0.2 mM EDTA in a 20 µL reaction volume for 6 minutes at room temperature in order to denature the double stranded structure. Next, 2 µL of 3 M sodium acetate (pH 4.3) and 55 µL of 100% ethanol was added and incubated at -20^{o}C for 18 hours to stop the reaction and precipitate the DNA. The plasmid was pelleted by centrifugation at 16,000 x g for 30 minutes at 4^{o}C before decanting the supernatant and was washed with 100 µL of 70% ethanol. Another 15 minute 16,000 x g centrifugation at 4^{o}C was conducted, the supernatant was decanted and the pellet was air dried for 5 minutes. The pellet was resuspended with 7 µL of milli-Q water and 2 µL of reaction buffer (250 mM NaCl, 200 mM Tris HCl pH 7.5,100 mM $MgCl_2$) and 1 µL of 0.2 µg/µL of reverse primer was added. The reaction mixture was set into a 65^{o}C water bath for 2 minutes and left to cool to 30-37^{o}C over a period of 20 minutes for primer annealing, and finally on ice for 2 minutes. 5 µL of sequenase master mix was added to the reaction mixture (1 µL of 100 mM DTT, 0.4 µL of labeling mix (7.5 µM dGTP, 7.5 µM dTTP, 7.5 µM dCTP), 6.25 µCi of [α^{35}S] dATP (Perkin Elmer), 3.25 units of Sequenase DNA polymerase (USB), 1.75 µL of dilution buffer (10mM Tris HCl pH7.5, 5 mM DTT, 100 µM EDTA), and 1.6 µL of dH_2O), and was incubated for 5 minutes at room temperature. Immediately after, 3.5 µL of this mixture was transferred into four tubes containing 2.5 µL termination mixes of dideoxynucleotide analogues (ddATP, ddTTP, ddGTP, ddCTP) and were incubated at

37ºC for 5 minutes. Finally, 4 µL of RNA loading buffer was added to each tube and stored at -20ºC until ready to load onto a gel. Samples were heated to 95ºC for 5 minutes and then ice for 2 minutes before loading onto a 7M Urea / 8% acylamide gel.

2.12. Radio-labelled RNA for north-western blot

2.12.1. Amplification of DNA template

Forward and reverse primers specific to the sequence region of interest were designed in order to create a negative sense radio-labelled RNA for north-western blotting. The list of primers used in this study for preparing north-western probes, and corresponding promoter sequences, via PCR are found in **Table 1**. A single 369 base pair anti-sense probe complementary to the Rev ORF was used to detect all Rev fragments and mutant RNAs using theSP6forward and HXB2 Rev 1 F primers with the pcRev plasmid as a template. A 322 base pair anti-sense probe of the HSRL was prepared using SP6forward and T7forward primer with the pcHSRL plasmid as a template. Both PCR reactions were prepared under the same conditions as follows: to a pcr tube, 100 ng of plasmid, 200 ng of forward and reverse primer, 5 µL of 10 x thermopol buffer (200 mM Tris HCl pH 8.0, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 20 mM $MgSO_4$, 1% Triton X-100) (New England Biolabs), 2 µL of 10 mM dNTPs, 2 µL of Taq polymerase enzyme, and 38 µL of milli-Q water was added. The PCR reaction was conducted with a preliminary 5 minute denaturation step at 95ºC for 5 minutes, followed by 30 amplification cycles, denaturation at 95ºC for 30 seconds, annealing at 55ºC for 30 seconds, and extension at 68ºC for 25 seconds. A final extension was carried out at 68ºC for 5 minutes. All PCR products were checked for the correct size by separating 5

µL of the reaction mixture on a 1% agarose gel next to a 1 Kb DNA ladder. 50 µL of milli-Q water was mixed in with the reaction mixture and was extracted with 100 µL of PCI before precipitation with 10 µL of 3M sodium acetate (pH 4.3) and 250 µL of 100% ethanol. The DNA was pelleted by centrifugation at 16,000 x g for 30 minutes, washed with 70% ethanol, and was air dried for 5 minutes before resuspending in 20 µL of 0.5 x TE. Concentration of the PCR reaction was quantified using a spectrophotometer.

2.12.2. *In vitro* transcription of radio-labeled probe

To a 1.5 mL microtube, 10 µL of 5 x SP6 transcription buffer (200 mM Tris HCl pH 7.9, 30 mM $MgCl_2$, 50 mM DTT, 10 mM spermidine), 2 µL of 10 mM NTPs-CTP, 1 µL of 0.5 mM CTP, 5 µL of [α^{33}P]-CTP (Perkin Elmer), 2 µL of 0.5M DTT, 40 units of SP6 enzyme (New England Biolabs), 1 µg of PCR template, and 27 µL of milli-Q water. The reaction was incubated at 37^{o}C for 2.5 hours before adding 1 µL of DNase I (New England Biolabs) and incubated for an additional 30 minutes at 37^{o}C. The reaction was stopped by heat inactivation at 65^{o}C for 5 minutes and was filtered through a G50 column and counted for radioactivity as described in section **2.11.1**. Approximate [α^{33}P]-CTP incorporation per transcript was calculated by the total percent radioactivity in cpm of the final RNA transcript divided by the total cpm of the initial reaction mixture. To determine the approximate molar amount of RNA produced, this percentage was multiplied by the total moles of CTP (radioactive and non-radioactive) used in the reaction divided by the total number of cytosine nucleotides found in the transcript.

2.13. North-western blotting

A nitrocellulose membrane and filter paper were first soaked in 1 x RIP buffer (60 mM KCl, 10 mM Tris HCl pH 7.4, 10 mM $MgCl_2$) for 10 minutes before placing, nitrocellulose on top, in a 96 well dot blot apparatus (BIORAD). Vacuum was applied to dry the wells and closed before adding 200 µL of 1 x RIP buffer to each well. Vacuum was applied again to dry the wells before applying the PAP protein. Next, 8 replicates of a PAP and BSA gradient consisting of 4, 2, 1, 0.5, and 0 µg of protein in 100 µL of 1 x RIP buffer was added to each of the wells before applying vacuum. Another 200 µL of 1 x RIP buffer was passed through before closing the vacuum. The apparatus was disassembled and the nitrocellulose membrane was rocked in Ponceau staining solution (1% glacial acetic acid, 0.1% Ponceau stain) for 5 minutes at room temperature and destained in 1 x RIP buffer three times for 5 minutes each at room temperature in order to check for equal loading of protein. Each replicate was cut and placed into separate 100 mL hybridization tubes for each RNA probe. The membranes were blocked with 20 mL of Torula yeast blocking buffer (150 mM NaCl, 50 mM Tris HCl pH 7.8, 5 mg/mL Torula yeast RNA) at 37^{o}C for 1 hour. Excess block was washed two times with 50 mL of 10 mM Tris HCl pH 7.8 for 5 minutes at 37^{o}C each. Membranes were incubated with 5 x 10^5 cpm/mL of RNA probe in 10 mL of north-western buffer (NWB) (10 mM Tris HCl pH 7.8, 1 mM EDTA, 50 mM $NaCl_2$, 0.02% Ficoll, 0.02% BSA, 0.02% polyvinylpyrolidone) for 1 hour at room temperature. Excess probe was removed using four 30 minute washes in 10 mL of NWB at room temperature. The blot was air dried for 10 minutes before being screened against a phosphor imaging screen (Bio-Rad) to visualize the bound RNA. The phosphor screen were scanned using a PharosFX

phosphor imager and densitometry data was collected from blots using Quantity One © software.

2.13.1. *In vitro* transcription

In a 1.5 mL microtube, 10 µL of 5 x T7 transcription buffer (200 mM Tris HCl pH 7.9, 50 mM NaCl, 30 mM $MgCl_2$, 50 mM DTT, 10 mM spermidine), 2 µL of G+NTP mix (10 mM ATP, 10 mM CTP, 10 mM UTP, 15 mM GTP), 2 µL of 0.5M DTT, 2 µL of T7 enzyme, 1 µg of PCR template, and 33 µL of milli-Q water. The reaction was incubated at 37^{o}C for 3 hours before adding 2 units of DNase I (New England Biolabs) and incubated for an additional 30 minutes at 37^{o}C. Next, 50 µL of 1 x TE was added and the RNA was extracted with 100 µL of citrate buffered PCI (pH 4.3) before adding 10 µL of 3 M sodium acetate (pH 4.3) and 250 µL of 100% ethanol. The RNA was precipitated at -20^{o}C for 1 hour and pelleted by centrifugation at 16,000 x g for 30 minutes. The RNA pellet was washed with 70% ethanol and air dried for 5 minutes before suspending in 20 µL of DEPC treated water. RNA concentration was estimated using a spectrophotometer and 1 µg of RNA was loaded onto a 1% agarose gel to check the quality.

2.14. Electromobility shift assay

Before starting the reaction, radio-labeled RNA was further refined by gel-purification. Briefly, radio-labeled RNA was dissolved in 100 µL of 1 x RNA loading buffer (95% formamide, 20 mM EDTA, 0.25% bromophenol blue, 0.25% xylene blue) and was heated to 95^{o}C for 5 minutes and ice for 2 minutes. The RNA was loaded into

a single well 7 M Urea / 4.5% acrylamide denaturing gel and separated at 150 V for 30 minutes at room temperature. The gel was exposed to a phosphor screen for 5 minutes and scanned immediately to determine the location of the radiolabelled RNA. The RNA was excised from the gel and crushed into the bottom of a 1.5 mL microtube with 3 volumes (between 600 and 800 µL) of RNA elution buffer (0.3 M $NaC_2H_3O_2$, 5 mM EDTA, 10 mM Tris HCl pH 7.4, 0.1% SDS) and rocked at room temperature for 1 hour. The samples were centrifuged at 16,000 x g for 5 minutes at room temperature and the supernatant was syringe filtered through a 0.45 µM Nalgene filters (Thermo Scientific) into new 1.5 mL microtubes. The RNA was precipitated by adding 1 mL of 100% ethanol, 10 µg of glycogen, and 50 µL of 3 M sodium acetate (pH 4.3) and incubated at -20^{o}C for 18 hours. The RNA was pelleted by centrifugation at 16,000 x g for 20 minutes and was washed with 70% ethanol before air drying for 5 minutes and suspending in 20 µL of DEPC treated water. Radioactivity of final eluted RNA was determined from a 1 µL aliquot in a scintillation counter (Perkin Elmer 2910 TR). Theoretical yield was determined by calculating the total moles of CTP used in the reaction (estimated moles of labelled and un-labelled CTP) and multiplying by the average molecular weight of the four nucleotides in an RNA chain (1320 g/mol) and by the percent CTP incorporation calculated as the cpm of radioactivity of the precipitated RNA divided by the total cpm of radioactivity in the initial reaction mixture. Specific activity of the radiolabelled probe was calculated by the total radioactivity of the probe divided by the total mass.

Equimolar amounts of [α^{33}P]-CTP radiolabelled RNA were dissolved in 77 µL of 1 x RIP buffer and incubated at 65^{o}C for 5 minutes. RNA was then incubated at 30^{o}C to fold for 15 minutes before aliquoting 7 µL into ten different 14 µL aliquots of PAP protein

dilutions in 1.5 x binding buffer (7.5% glycerol, 15 mM HEPES pH 7.0, 90 mM KCl, 15 mM Tris HCl pH 7.4, 15 mM $MgCl_2$, 0.015% BSA) ranging from 5 µM to 0.04 µM and 0 µM and was incubated for an additional 10 minutes at 30^{o}C. The samples were then set on ice for 10 minutes before adding 5 µL of native loading dye (25% glycerol, 60 mM Tris HCl pH 6.8, 0.025% xylene blue, 0.025% bromophenol blue) and immediately loading onto a native 4.5% acrylamide gel and separated at 150 volts for 30 minutes at room temperature. The resulting gel was vacuum dried and screened against a phosphor imaging screen (Bio-Rad) to visualize the RNA. The phosphor screen was scanned using a PharosFX phosphor imager.

2.15. Filter-binding assay

Equal amounts of each [α^{33}P]-CTP radiolabelled RNA were dissolved in 1.5 mL of 1 x RIP buffer and were incubated at 65^{o}C for 5 minutes. RNA was then incubated at 30^{o}C to fold for 15 minutes before aliquoting150 µL into ten different 300 µL aliquots of PAP protein dilutions in 1.5 x binding buffer ranging from 15 µM to 0.1 µM and 0 µM and was incubated for an additional 10 minutes at 30^{o}C. The samples were then set on ice for 10 minutes before filtering through a nitrocellulose membrane. The membrane was washed three times with FB wash buffer (10 mM HEPES pH 8.0, 0.1% Tween-20, 1 mM EDTA) and dried for at least 30 minutes before placing into a 7 mL glass scintillation vial (Fischer Scientific) with 3 mL of scintillation fluid. The vials were measured for radioactivity using a scintillation counter (Perkin Elmer 2910 TR). Alternatively, equal amounts of each [α^{33}P]-CTP radiolabelled RNA were dissolved in 666 µL of 1 x RIP buffer and were incubated at 65^{o}C for 5 minutes. RNA was then incubated at 30^{o}C to

fold for 15 minutes before aliquoting 66.6 µL into 6 different 133.3 µL aliquots of PAP protein dilutions in 1.5 x binding buffer ranging from 15 µM to 0.4 µM and 0 µM and was incubated for an additional 10 minutes at 30^{o}C. The samples were then set on ice for 10 minutes before filtering through a nitrocellulose membrane and subsequently a nylon membrane. FB wash buffer (10 mM HEPES pH 8.0, 0.1% Tween-20, 1 mM EDTA) was filtered through three times and the membranes were dried for at least 5 minutes at room temperature before cross-linking the RNA with 1200 J/m^2 of UV light for 40 seconds. The membrane was incubated with 10 mL of prehybridization solution (0.5M sodium phosphate pH 7.2, 1 mM EDTA, pH 7.0, 7% SDS) at 60^{o}C for 2 hours. The solution was replaced with an additional 10 mL of prehybridization solution and 2 x 10^6 CPM of radiolabelled negative sense RNA probe was added and incubated for 18 hours at 60^{o}C. The membrane was washed with 10 mL of solution 1 (1 X SSC, 1% SDS) for 10 minutes at room temperature and 4 more times with solution 2 (0.5 x SCC, 0.1% SDS) for 10 minutes each at 60^{o}C. The membrane was then dried at room temperature for 20 minutes and screened against a phosphor screen. The phosphor screen was scanned using a PharosFX phosphor imager.

2.16. 1M7 and DMS Footprinting analysis

To a 1.5 mL microtube, 48 µg of RNA was dissolved in 200 µL 0.5 x TE and heated to 65^{o}C for 5 minutes. The RNA was set on ice for 2 minutes before adding 100 µL of 3.33 x RNA folding buffer and was set to 30^{o}C for 15 minutes to fold. After folding, 50 µL of RNA is aliquoted into six 1.5 mL microtubes. Into three of the tubes, 10 µg of PAP protein was added and the solutions were left to incubate at 30^{o}C for an additional

15 minutes. A set of PAP and RNA alone samples were treated with 7 µL of 50 mM1-methyl-7-nitrosatoic anhydride (1M7) in dimethyl sulphoxide (DMSO), 7 µL of 0.5 M dimethyl sulphate (DMS) in DMSO, or 7 µL of DMSO, and was incubated at room temperature for 4 minutes. The reaction was stopped by adding 50 µL of stop buffer (0.1 M DTT, 0.6 M $NaC_2H_3O_2$ pH 4.3) and 100 µL of citrate buffered PCI (pH 4.3), vortexed, and the supernatant was placed into a new 1.5 mL microtube containing 250 µL of 95% ethanol. The RNA was left to precipitate for 1 hour at -80^{o}C before centrifuging at 16,000 x g for 20 minutes, decanting and washing the pellet with ice cold 70% ethanol, and air drying for 5 minutes. The pellet was resuspended in 8 µL of DEPC treated water and primer extension was conducted as described in section **2.11.2**.

2.17. 1M7 and DMS structural analysis of PAP binding site in Rev

Strong bands produced from the primer extension reactions of Rev RNA with 1M7 or DMS treatment (section 1.16) were used to identify non-basepaired nucleotides in Rev. Nucleotides that were found to be reactive were used as constraints in both MFold[73] and ViennaRNA[74] folding algorithms by prohibiting the base-pairing of these nucleotides. Multiple minimum free energy structures of the RNA sequence, predicted using a loop-based energy model, were generated and 5 of the smallest Gibbs free energy (G) structures from each program were chosen to compare the possible structures the Rev ORF RNA folds. A small, 15 nucleotide hairpin was consistently found between all folded structures and was used for 3D modeling. The predicted secondary structure of the RNA hairpin was used as a constraint for predicting the 3D model of the RNA using RNAcomposer[75]. RNAcomposer builds 3D models of RNAs

based on their secondary structures using a library of solved x-ray crystal and NMR solvent structure data-base called RNA FRABASE. RNA FRABASE acts like a dictionary relating RNA secondary structure and tertiary structure elements to small stretches of RNA and RNAcomposer attempts to assemble these small stretches to predict how the tertiary structure would form.

Table 1 - List of primers used in this study

Primer name	Reaction	Sequence
REVORF146R	Fragment PCR	5’-CTTCTATTCCTTCGGGCCTGTCG-3’
REVORF264R	Fragment PCR	5’-AAGTCTCTCAAGCGGTGGTAGCTG-3’
REVORF351R	Fragment PCR	5’-CTATTCTTTAGTTCCTGACTCC-3’
T7REVORF198F	Fragment PCR	5’-GAGAGTAATACGACTCACTATAGGGAG ACTTATCTGGGACGATCTGCG-3’
T7REVORF89F	Fragment PCR	5’-GAGAGTAATACGACTCACTATAGGGA GACTATCAAAGCAACCCACCTCCC-3’
T7forward	Fragment PCR	5’-ATCATAATACGACTCACTATAGGGAGA-3’
SP6forward	North-western probe PCR	5'-ATCCATACGATTTAGGTGACACTATAG-3
GAPDH 300R	Fragment PCR	5’-TGGGATTTCCATTGATGACAAGC-3’
HindIIIHSRL4475F	Cloning	5’-CACACAAGCTTCAGAATTCACCAAGC TTGG-3’

BamHIHSRL5003R	Cloning	5’-CACACGGATCCAAACCCCGACCCAGAAGCAG-3’
HXB2 Rev R (8648)	Primer Extension	5’-CTTTAGTTCCTGACTCCAATACTGTAGGGAG-3’
APEX H	Primer Extension	5’-AGTCATAATCCCACAGATGGTAG-3’
F Rev G291-293A	Mutagenesis	5’-AGGGGGTAAAAAGCCCTCAAATATTGGTGGAATCTCC--3’
R Rev G291-293A	Mutagenesis	5’-GAGGGCTTTTTACCCCCTGCGTCCCAGAAGTTCC-3’
F HindIII REV 1	Cloning	5’-TACGACTCACTATAGGGAGACCCAAGCTTGGTACC-3’
R-pcDNA3-DSR	Cloning	5’-TTCCGGTAATTAGTATCGCTATGACATGCATGCC-3’
HXB2 Rev 1 F	North-western probe PCR	5’- GGAGACCCAAGCTTGGTACC-3’
F-T7-HARGT	Fragment PCR	5’-GTTAATACGACTCACTATAGGGCCAGTGGCGAATGGATAACGCG-3’
R-MatHARGT	Fragment PCR	5’-CGAGCCAGCCAGGAGTCGAACCTGG-3’

3. Results

3.1. PAP depurinates the Rev RNA *in vitro* in the same position as *in vivo*.

It is well known that PAP catalytically removes an adenine from the sarcin/ricin loop of eukaryotic and prokaryotic rRNA. PAP can also depurinate viral RNA, including recently the Rev ORF within the HIV-1 2KB RNAs *in vivo*. In this study, I sought to better understand how PAP targets RNA for depurination by investigating PAP's interaction with the HIV-1 Rev RNA substrate. In order to study PAP interaction with Rev RNA, I required an RNA substrate that would match the structural characteristics of the *in vivo* RNA that was small enough for *in vitro* binding assays. The Rev RNA had been previously cloned by our lab in pcRev. I hypothesized that if PAP could target an *in vitro* transcript of pcRev for depurination in the same position as *in vivo*, the structural characteristics of the RNA would likely be the same for PAP binding. In order to test my hypothesis, primer extension of PAP-treated *in vitro* transcribed Rev RNA encompassing the Rev ORF was conducted across the expected depurination site (**Figure 8**). Briefly, the *in vitro* RNA was prepared by using the T7 polymerase and a PCR template of pcRev. The *in vitro* transcribed Rev RNA was denatured and allowed to re-fold before treating with increasing amounts of PAP. Treated RNA was phenol extracted and precipitated before using a Rev-specific radiolabelled reverse primer complementary to nucleotides 8648-8668 of the HIV-1 pMenv$^{(-)}$ proviral DNA (K03455.1) for primer extension analysis. Depurination sites were observed at G291 and A294 from the *in vitro* transcribed Rev RNA. These sites align with the expected *in vivo* depurination sites, G8594 and G8597, shown previously[29] (**Figure 6**). The intensity of depurination of the *in vitro* depurination sites were concentration dependent on PAP.

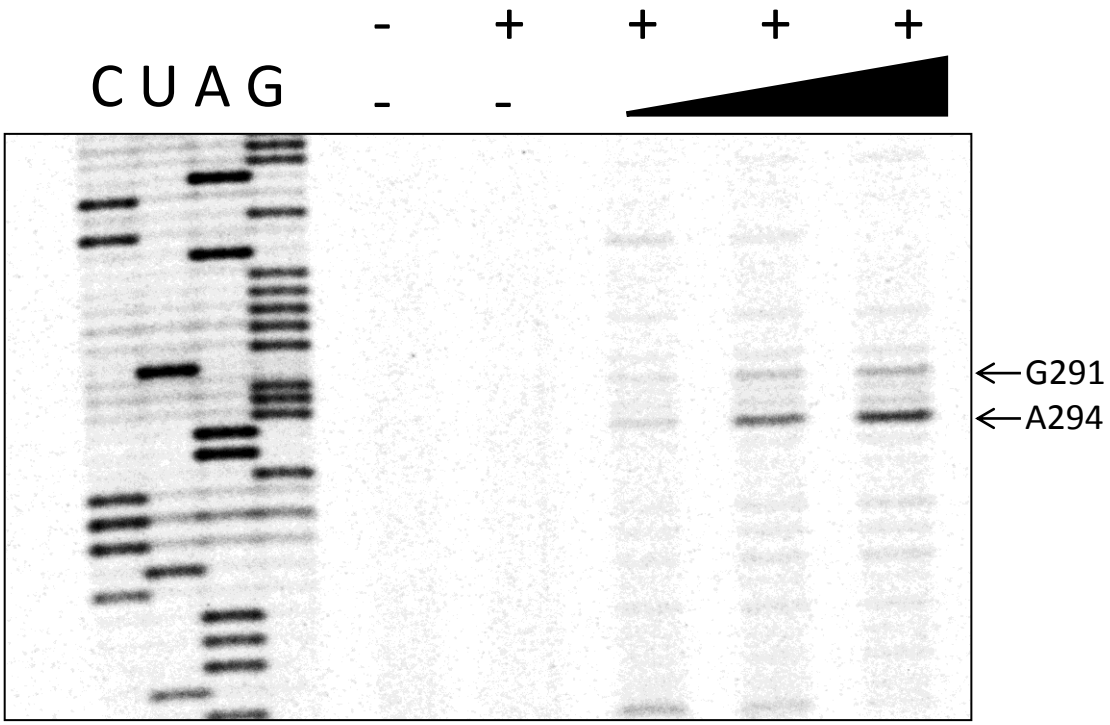


Figure 8 – PAP depurinates the Rev ORF *in vitro*. 8 µg of *in vitro* transcribed Rev ORF from pcRev was treated with 0, 1.25, 2.5, or 5 µg of PAP. Treated RNA was phenol extracted and precipitated before primer extension was performed using a reverse primer complementary to nucleotides 8648-8668 of the HIV-1 pMenv$^{(-)}$ proviral DNA (K03455.1). The cDNA products were separated through a 7M urea / 8% acrylamide gel and observed using a phosphorimager. Manual dideoxynucleotide sequencing was performed using the same primer as described above on the pMenv$^{(-)}$ proviral clone. Arrows indicate the positions of depurination.

Other fainter bands appeared up and downstream of the expected depurination site, but were not consistently observed from reaction to reaction. Other depurination sites may be the result of PAP depurination, but are regarded as un-stable structures or less active sites of PAP's enzymatic activity. The results suggest a strong structure or sequence that promotes PAP binding and catalysis near the 3' end of the Rev RNA in both *in vivo* and *in vitro*.

3.2. PAP binds to Rev and SRL RNA more tightly than GAPDH RNA or Arg-tRNA

In this study, I asked whether PAP could bind to Rev RNA more tightly than to other cellular RNAs. As a preliminary study, the focus was only on the ability of PAP to recognize naked RNA without a cap structure and without protein interacting with the RNA. I hypothesized that a particular sequence or structure of RNA is required for PAP to preferentially bind to some RNAs over others. It is well known that PAP can target RNA other than the sarcin/ricin loop *in vitro,* but it has never been studied what sequence or structure of RNA PAP preferentially binds. In order to test my hypothesis, I chose to do a north-western blot to identify two things: 1) does PAP bind to Rev RNA and 2) is PAP binding specific to only certain RNA substrates or does PAP bind each selected RNA equally. *In vitro* binding assays of PAP and the SRL had previously shown tight, saturatable binding to a 27-mer SRL RNA with a binding constant (K_D) of 0.2 +/- 0.08 pmol[39]. Therefore, I chose the SRL as a positive control for PAP binding, but as a longer 300 nucleotide sequence of the 28S rRNA containing the SRL. This was done in order to closely match the length of the Rev transcript and to control for non-

specific interactions of PAP binding due to their lengths. In order to test whether PAP can selectively bind to different RNA substrates, I chose a 300 nucleotide fragment of the GAPDH RNA to act as a negative control, because it was previously shown to have no depurination sites in its sequence. As an alternative control, I decided to test if PAP can bind to an *in vitro* transcribed arginine transfer RNA (Arg-tRNA). The Arg-tRNA is a small 95 nucleotide RNA molecule and is involved in carrying and transferring an arginine amino acid in protein synthesis. tRNAs have a highly conserved structure and sequence, so I wanted to test if PAP would interact with this RNA. Briefly, RNAs were radiolabelled during transcription using α^{33}P-CTP and incubated in excess with a nitrocellulose membrane blotted with increasing amounts of PAP. The relative amount of binding to each RNA was quantitatively compared by the intensity of radioactivity bound to equal amounts of PAP. The relative amount of radioactivity on each blot was counted by exposing the entire membrane to a phosphor screen and measuring the amount of energy stored on the screen by direct laser excitation (**Figure 9A**). Since the incorporation of α^{33}P-CTP is different for each RNA depending on the relative amount of cytosine nucleotides incorporated in the main RNA chain, the radioactivity of the blots was divided by the ratio of CTPs in GAPDH RNA, Rev RNA, SRL and Arg-tRNA by 8.8 : 5.2 : 4.4 : 1, respectively. As well, the background binding of RNAs to 0 µg of PAP was subtracted from each data point per blot. The final data were averaged across three trials and plotted on a bar graph, as shown in **Figure 9B**. The results show that PAP binding to all RNAs increased with more protein blotted. At all concentrations of protein, PAP appeared to bind significantly more ($p<0.05$, using Tukey's Studentized Range

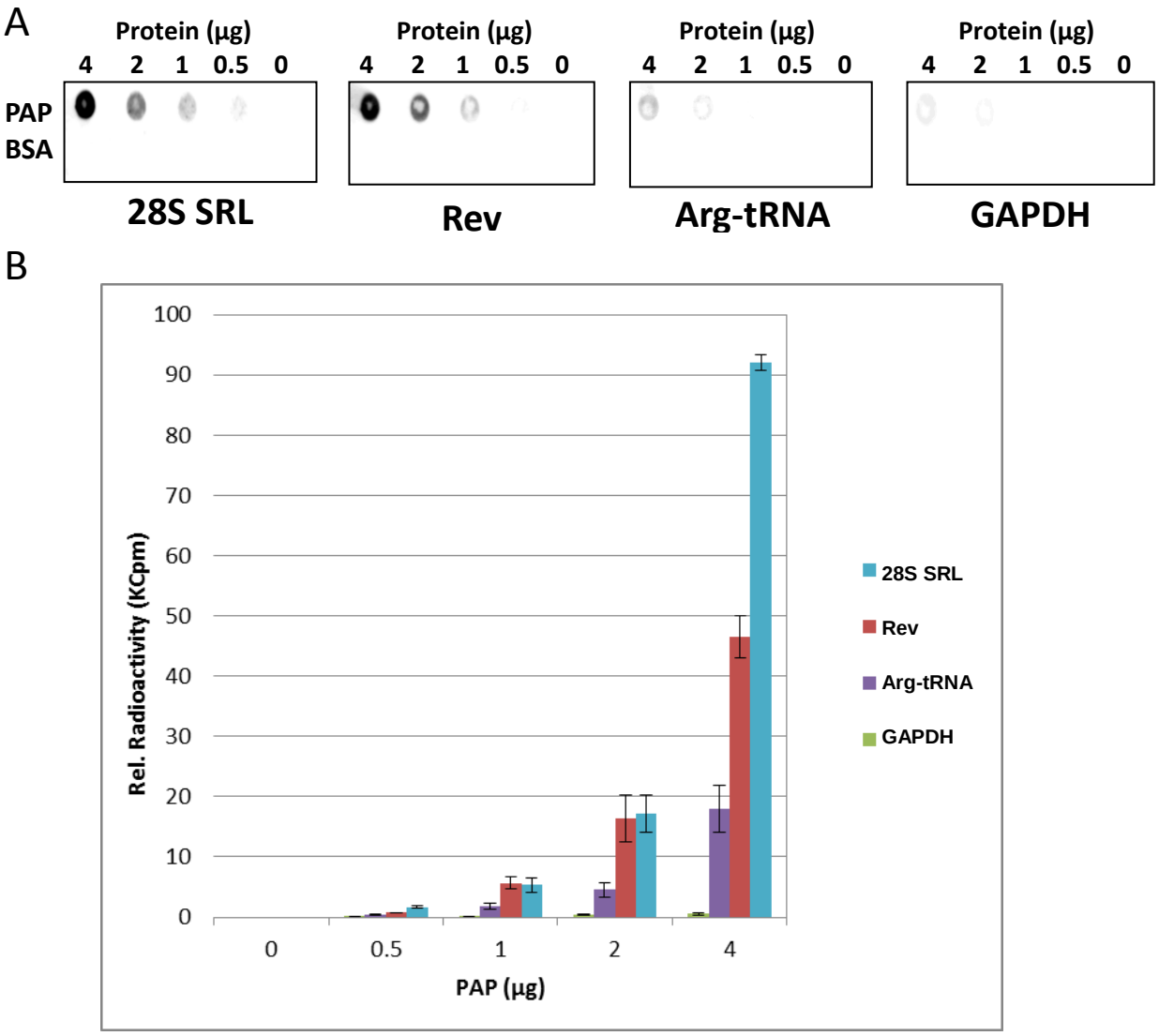


Figure 9 – PAP selectively binds to the 28S SRL and Rev RNA over the Arg-tRNA or GAPDH RNA.A) North-western blot analysis of PAP binding to *in vitro* transcribed SRL, Rev, GAPDH, and Arg-tRNA. PAP and BSA (4, 2, 1 and 0.5 µg) were blotted onto nitrocellulose and incubated with radiolabelled RNAs. Bound RNA was visualized with a phosphor imager. **B)** Average intensity of radioactivity (KCpm) from transcripts bound to PAP at 4, 2, 1 and 0.5 µg across three north-western blots. Relative intensity of radioactivity for each transcript was divided by the ratio of CTPs in a 8.8 : 5.2 : 4.4 : 1 ratio for GAPDH, Rev, SRL and Arg-tRNA, respectively. Densitometry data were collected using the Pharos FX Plus Molecular Imager and calculated using Quantity One. Error bars represent the standard error of the mean (SEM).

test for multiple comparisons of means) to the SRL and Rev RNA than GAPDH RNA and Arg-tRNA, and significantly more to the Arg-tRNA than GAPDH RNA. Only at 4 µg of PAP was there a difference in binding for all RNAs, where the binding to SRL was significantly more than Rev RNA ($p<0.05$). Additionally, BSA did not bind any of the prepared RNAs at any concentration equal to PAP, indicating that PAP binding was specific and the conditions were stringent enough to eliminate background binding.

To confirm the results of the north-western hybridization, an electromobility shift assay (EMSA) and a filter binding assay were conducted as shown in **Figure 10A** and **10B**, respectively. The EMSA was conducted using gel purified α^{33}P-CTP radiolabelled RNA from an *in vitro* transcription. Equimolar amounts of RNA were denatured at 95^{o}C and allowed to re-fold before incubating with increasing amounts of PAP. Following separation on a non-denaturing gel, the relative amount of binding of each RNA was qualitatively compared by the amount of PAP needed to shift 50% of the RNA band. About 156 nM of PAP was required to shift 50% of SRL RNA, 312 nM of PAP to shift Rev RNA, 625 nM of PAP to shift Arg-tRNA, and 1.25 µM of PAP to shift GAPDH RNA. All RNAs were visualized on a urea gel before conducting the EMSA for quality and purity, and in every case the RNA extracted produced a single band. The presence of multiple bands on the native gel is likely due to the alternate folding of the RNA. In the case of the Arg-tRNA, there were three conformations, each with their own binding shift. In this case, the average concentration of PAP required to bind 50% of the RNA among the three conformations was considered.

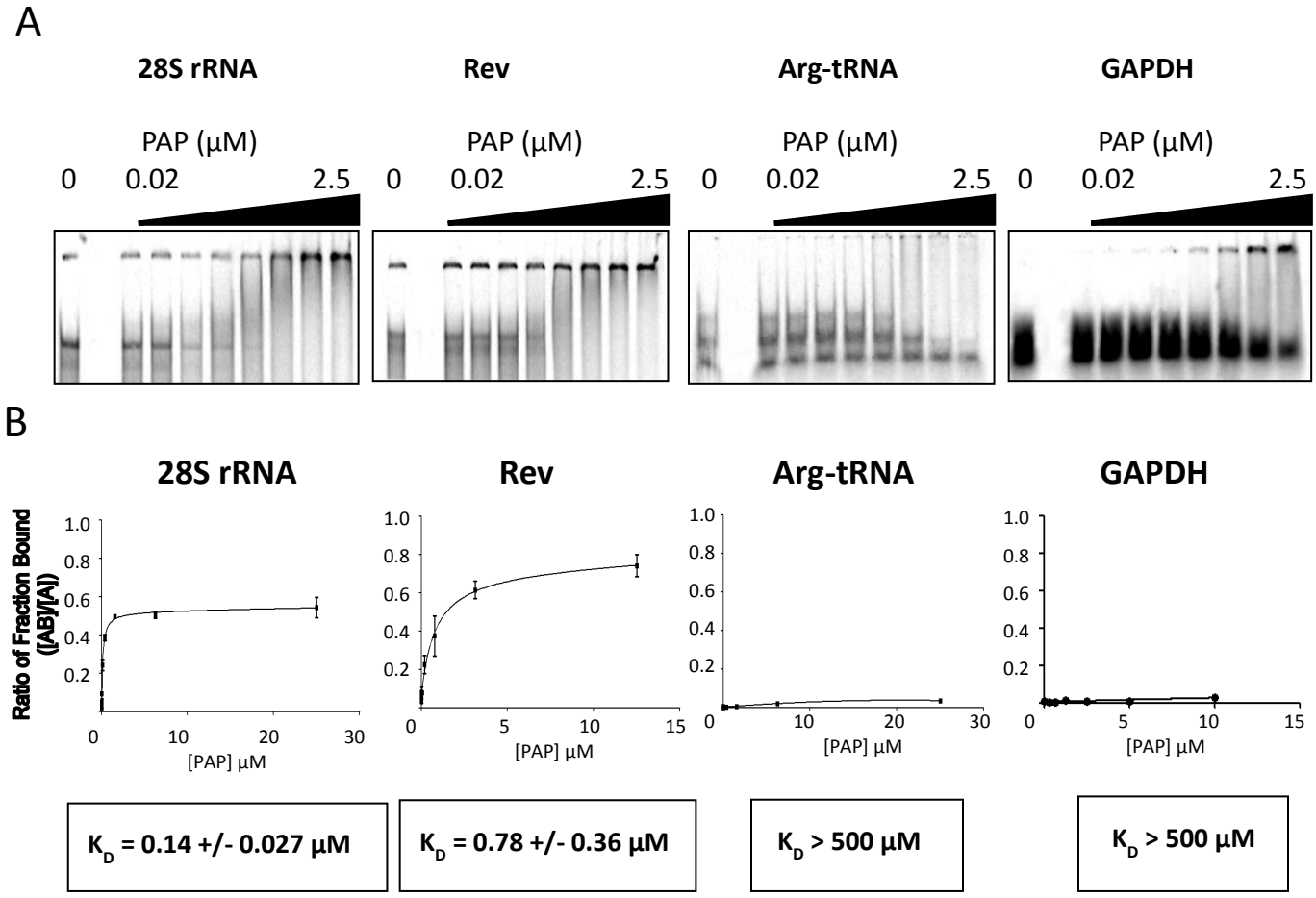


Figure 10 – PAP binds to the 28S rRNA and Rev RNA tighter than to Arg-tRNA and GAPDH RNA. Equimolar amounts of radiolabeled RNA transcripts of Rev, 28S rRNA, Arg-tRNA, and GAPDH were heat denatured and then refolded before aliquoting into a gradient of PAP. The mixture was incubated at 30ºC for 15 minutes before **A)** separation through a 4.5% native acrylamide gel or **B)** filtration through a nitrocellulose membrane. RNA bound to PAP was detected by either exposing to a phosphor screen or scintillation counting, respectively. The data points are the means of three replicates. Best fit curves and binding constants were calculated using a statistical program called Prism5. Error bars depict the standard error of the mean.

It is also worth noting that the 28S rRNA and Rev RNA samples likely contained complex aggregates of RNA that could not enter the gel. Only the functional RNA that PAP bound to in the gel was considered for the analysis. The filter binding assay followed the same protocol as the EMSA, however the amount of PAP used ranged from 20 μM to 20 nM and the final mixture was set on ice for 10 minutes before filtering through a nitrocellulose membrane. The relative amount of radioactivity bound to each of the membranes by PAP-RNA complexes was measured by scintillation counting. Each experiment was replicated 3 times and curves for the filter binding assay were generated using the statistical program Prism5. The binding constant (K_D) of each RNA to PAP was calculated as the concentration of PAP where 50% of the max binding capacity (B_{MAX}) was reached. The calculated K_D of PAP binding to Rev RNA and SRL was 0.14 +/- 0.027 μM and 0.78 +/- 0.36 μM, respectively. The K_D for PAP to Arg-tRNA and GAPDH RNA could not be calculated because binding could not be achieved under these conditions. These values correlate well with the EMSA data. In both experiments, PAP appeared to bind to the SRL and Rev RNA more tightly than GAPDH RNA and Arg-tRNA. In addition, PAP bound to the SRL 5 times more than Rev RNA, but were found to be not significantly different ($P = 0.15$, using one-tailed T-test). The binding capacity of both the 28S rRNA and Rev RNA did not reach perfect saturation. Based on the EMSA data, it is likely that only 50% of the total RNA was actively bound by PAP while the remaining RNA conformations were not. Interestingly, PAP bound to the SRL about 700 times less tightly than the 27-mer SRL tested previously. Since the size of the SRL construct I selected also includes about 300 nucleotides more of the 28S rRNA than the 27-mer, other structures may have obstructed or changed the conformation of

the RNA such that PAP did not interact as well with the sarcin/ricin loop. To the same effect, extraneous RNA structures in Rev RNA may be obstructing the full binding potential to PAP. These data suggest that Rev RNA contains a similar RNA sequence or structure that PAP can interact with based on binding affinity.

3.3. Footprinting and filter binding assay show PAP binds to 3'-end of Rev RNA

I next wanted to identify the exact region of Rev RNA bound by PAP and decided to conduct a footprinting assay, as shown in **Figure 11**. Footprinting was conducted using two nucleotide modifying reagents, dimethyl sulphate (DMS) and 1-methyl-7-nitroisatoic anhydride (1M7). DMS modifies the N7 of guanosines, N1 of adenosines, and N3 of cytidines, and 1M7 acylates a common ribose 2'-hydroxyl group of all nucleotides along the RNA chain. All modified nucleotides in a chain of RNA inhibit reverse transcriptase (RT) from reading through during cDNA polymerization, and cause the RT to terminate at these modified nucleotides. Depurination sites and strong secondary structure will also cause the RT to terminate, but are controlled for by conducting RT on RNA with or with PAP treatment. Nucleotides that are involved in secondary or tertiary structure, such as base-pairing or intermolecular interactions with a protein, are protected from modification. Footprinting was conducted by folding the *in vitro* transcribed RNA and incubating with PAP protein. Next, the RNA was treated with the modifying chemical for a short period of time before quenching. RNA was phenol extracted and precipitated before conducting primer extension using a specific radiolabeled reverse primer complementary to 3' end of Rev RNA (HXB2 Rev R).

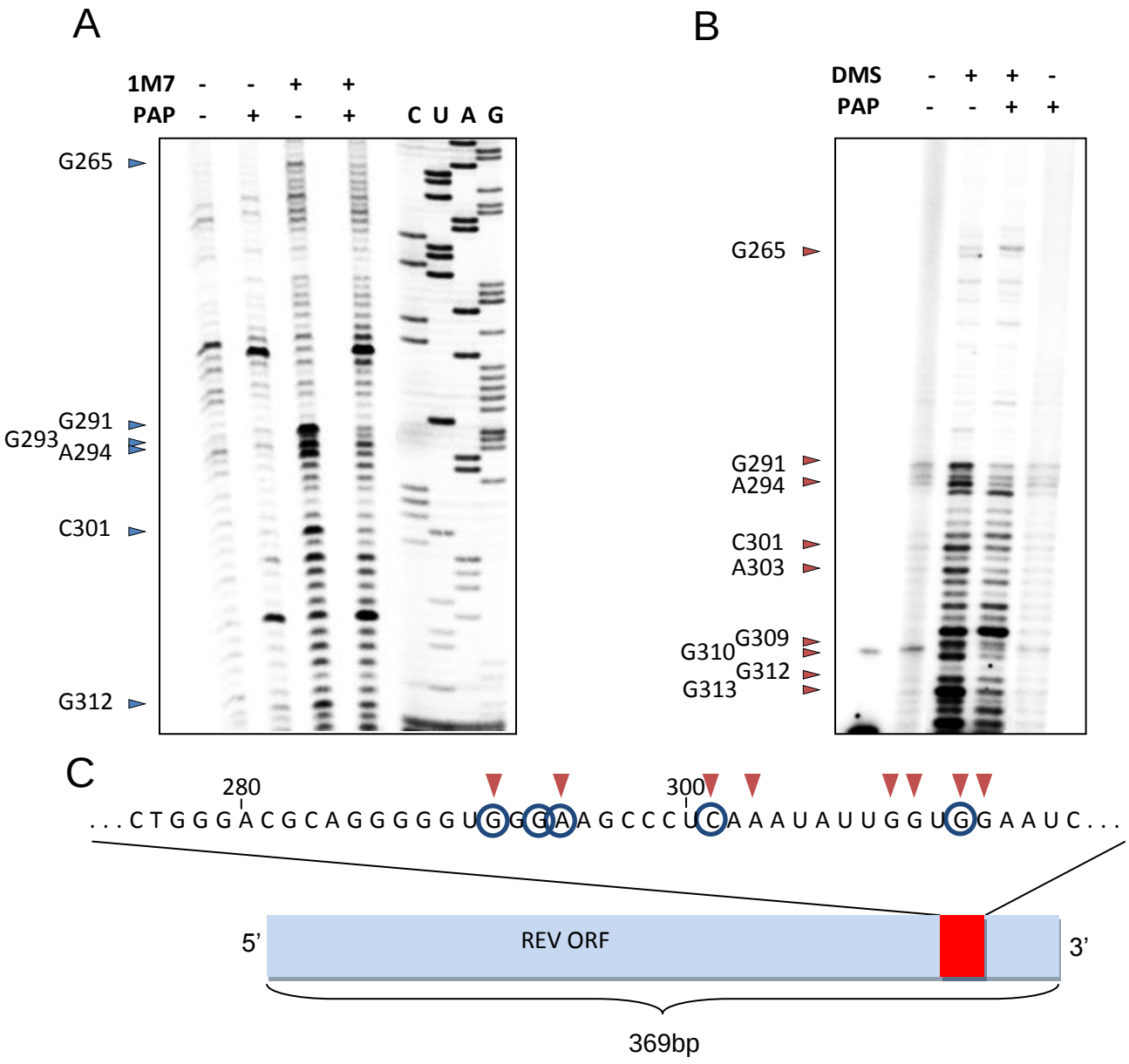


Figure 11 – Footprinting analysis shows a PAP binding site in the 3' region of the Rev RNA.

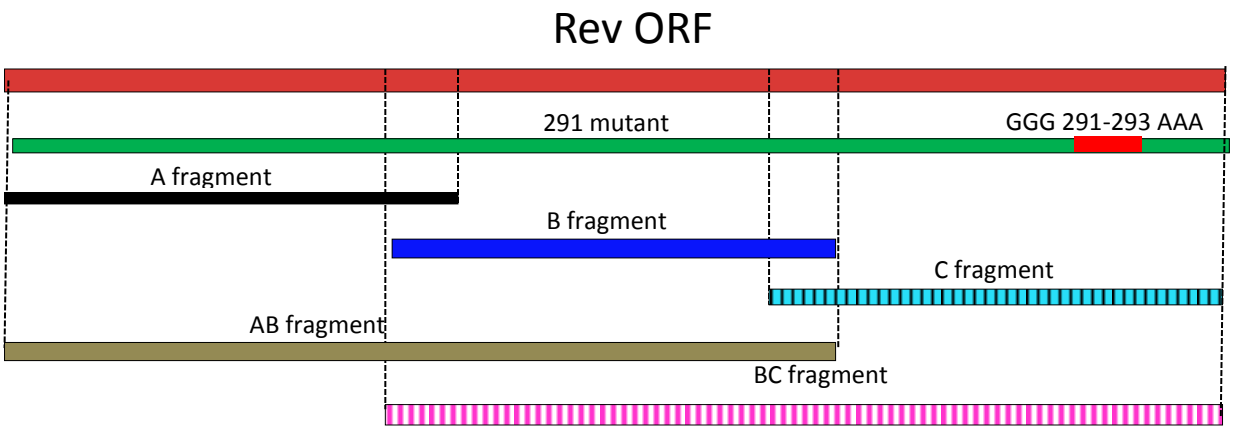
Figure 11 – Re-folded, *in vitro* transcripts of rev co-incubated with or without PAP before chemically treating with **A)** 1-methyl-7-nitroisatoic anhydride (1M7) or **B)** dimethyl sulphate (DMS). Treated RNAs were subsequently phenol extracted and precipitated prior to primer extension using a Rev specific radiolabeled reverse primer complementary to nucleotides 8648-8668 of the HIV-1 pMenv$^{(-)}$ proviral DNA (K03455.1). Manual dideoxynucleotide sequencing was performed on the pMenv$^{(-)}$ proviral clone using the same reverse primer. Termination bands on the gel are representative of either nucleotides modified by 1M7, DMS, or depurination sites. **C)** Pictorial representation of the rev RNA indicating the PAP binding site on the 3' end of the Rev ORF and the protected nucleotides (blue circles for 1M7 protection, and red triangles for DMS protection).

A Sanger sequencing ladder was separated adjacent to the primer extension products to determine the exact sequence of the bands. PAP dependent protection appeared at nucleotides G265, G291, A293, A294, C301, and G312 from 1M7 modification and G265, G291, A294, C301, A303, G309, G310, G312 and G313 for DMS modification (**Figure 11C**). The most consistent protection sites from both reactions were found between nucleotides G291-A294, in the same location of depurination. It is known that the PAP active site pocket is only large enough to occupy a di-nucleotide substrate[37], so the footprint is expected to be only a few nucleotides near the depurination. The protected nucleotides at G265, C301, G309, G310, G312, and G313 may also represent regions PAP interacts with Rev RNA *in vitro*, but do not align with any consistent or strong depurination site *in vitro* or *in vivo*. Interestingly, there were PAP-dependent RT terminations at A285 and A307 in the 1M7 PAP only control, but not in the DMS PAP only control. These terminations may be due to PAP depurination of the RNA. However, they were not consistent with previously shown reactions and were not present when the experiment was repeated using DMS. It is possible that Rev RNA has multiple conformations in solution that change in the presence of PAP, but these depurination are not consistent from reaction to reaction. The consistent PAP binding site at G291-A294 aligns exactly to the depurination sites shown *in vivo* and *in vitro,* so the region around this site was further investigated for PAP interaction.

My next objective was to determine whether the PAP binding site in the 3' region of Rev RNA is an element that can be separated from the ORF. In order to test this, five deletion mutations of the pcRev transcript (369 nt) were prepared via PCR and called A (nt 1-143), B (nt 80-245), C (nt 200-369), AB (nt 1-245), and BC (nt 80-369).

Since PAP was shown to depurinate nucelotides G291 and A294 and protect nucleotides around G291-A294, I decided to test the importance of the GGGA sequence for PAP binding. Three point mutations were generated in the full length Rev RNA to replace three guanine nucleotides (G291-G293) with adenines and was also tested for binding (**Figure 12A**). Equimolar amounts of *in vitro* transcripts of Rev ORF, A, B, C, AB, and AC deletion mutants and the 291-293 point mutant, were heat denatured and then refolded before aliquoting into a gradient of PAP. The mixture was allowed to incubate at 30^{o}C for 30 minutes before filtering through a nitrocellulose and subsequently a nylon membrane. RNA binding was detected by northern blotting using a radiolabeled antisense-Rev RNA. The relative amount of binding of each RNA was quantitatively compared by the intensity of radioactivity bound to equal amounts of PAP. Relative amount of radioactivity on each blot was counted by direct laser excitation and a ratio of the radioactivity on the nitrocellulose membrane over the total radioactivity on both the nylon and nitrocellulose membrane was used to calculate the specific binding. Curves and binding data were calculated using the statistical program Graphpad 5 (**Figure 12B**). Binding was compared using B_{Max} of each set because not enough data points were collected to determine an accurate K_D value. The calculated B_{Max} of PAP to Rev, A, B, C, AB, BC, and the 291 mutant was 0.85 +/- 0.38, 0.02 +/- 0.02, 0.04 +/- 0.03, 1.09 +/- 0.33, 0.12 +/- 0.25, 1.21 +/- 0.60, and 0.35 +/- 0.39, respectively. The B_{max} of PAP binding to Rev, BC, and C RNA was greater than to A, B, or AB mutants. As expected, PAP binding was confined to the 3'-end of the Rev RNA and specific binding to this region could be separated from the whole RNA. Furthermore, mutagenesis of G291-G293 to adenines resulted in reduced binding of PAP.

A



B

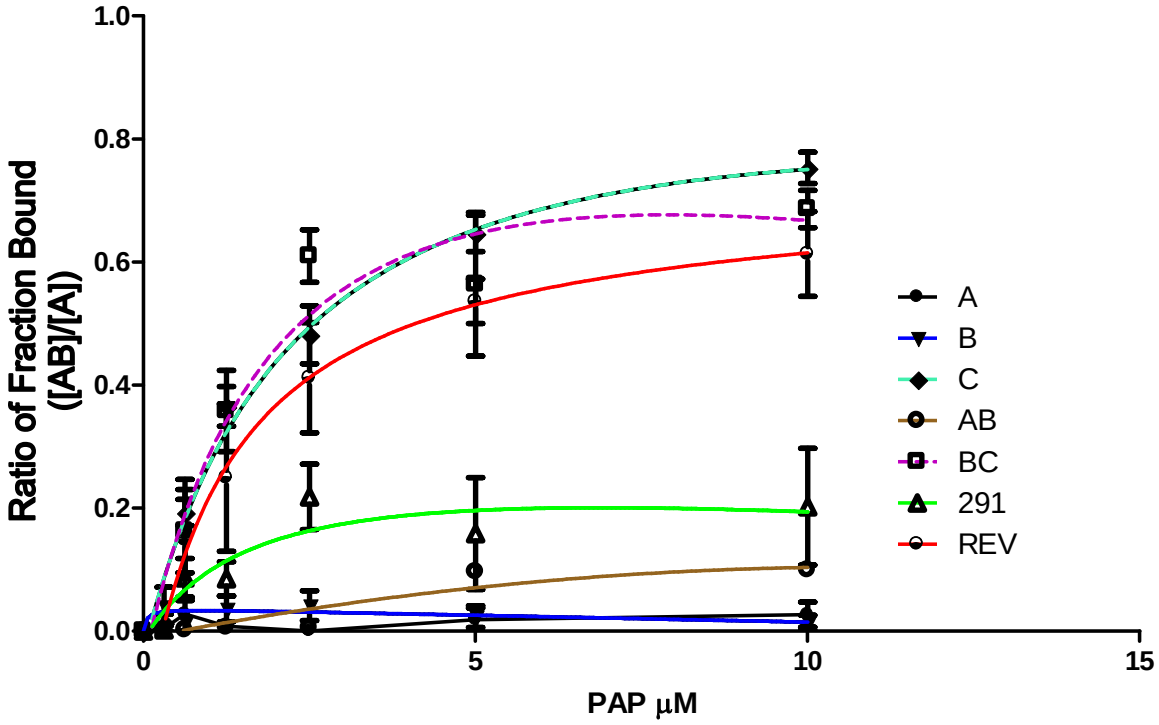


Figure 12 – PAP preferentially binds to the 3-end of the Rev ORF.

Figure 12 – A) Five deletion and one point mutation were prepared to confirm the location bound by PAP in the rev ORF. The Rev 291 mutant contains three guanosine nucleotides 291-293 mutated to three adenosine nucleotides on the site where PAP directly binds. The numbering refers to the nucleotide on the rev ORF starting from the 5' end. **B)** Equimolar amounts of *in vitro* transcripts of rev ORF, A, B, C, AB, AC, deletion mutants and the 291-293 point mutant, were heat denatured and then refolded before aliquoting into a gradient of PAP. The mixture was incubated at 30^{o}C for 15 minutes before immediately filtering through a nitrocellulose and nylon membrane. Bound Rev RNA was detected by northern using a 369 nt radiolabeled antisense-Rev RNA. The amount of binding of the probe was detected by phosphor screen imaging and densitometry. Densitometry data were collected using the Typhoon Trio Molecular Imager and calculated using Quantity One ®. The data are representative of 3 replicates. Best fit curves and binding constants were calculated using a statistical program called Prism5. Error bars represent the standard error of the mean.

3.4. Sequences surrounding *in vivo* PAP depurination sites in 28S rRNA, HTLV-1, and BMV RNA show no similarity

To get a better idea of what kind of RNA substrate PAP targets, I decided to compare the structure and sequence of known depurination sites by PAP. Several depurination sites have been identified on other viral RNA targets, including the Human T-Cell lymphotropic virus-1 (HTLV-1) and Brome mosaic virus (BMV). Since PAP can bind and depurinate a 27-mer RNA mimic of the sarcin/ricin loop, I thought that PAP would bind to a similarly sized RNA target around each depurination site (**Figure 13**). A range of 40 nucleotides around each depurination site was compared for sequence homology using the online alignment server BLAST, however no significant sequence similarity was evident from the test. I hypothesized that PAP may require a specific structure that is targeted for depurination and that it would contain similar elements as the SRL. Each assay involved a re-folding step that aimed to re-nature the most thermodynamically stable RNA structures. Therefore, minimum free energy (MFE) RNA structure prediction using thermodynamic-based parameters was used to fold each RNA sequence. ViennaRNA and MFold programs were used to fold each RNA, and it should be noted that the RNA structure generated included only secondary structures formed by canonical base pairs, as this algorithm does not predict non-canonical interactions or tertiary structures. The lowest energy structure from each program was compared and eight of the twelve folded RNAs formed hairpin loop structures such that the depurination site was within the loop structure. Closer inspection of these loop structures revealed only one other tetra-loop in the BMV RNA that closely matched the

Figure 13 - PAP depurination in the 28S rRNA, HTLV-1, or BMV RNAs sites show no sequence similarity. A sequence comparison between the binding site in rev to the known depurination sites in the 28S rRNA, HTLV-1 and BMV RNAs. Red letters are known depurinated nucleotides by PAP *in vivo* and underlined letters are nucleotides involved in PAP binding. Using the multiple alignment programs, BLAST and ClustalW, no consensus binding sequence near the PAP depurination site was present.

SRL, but without the G-bulge in its stem. The binding site in Rev RNA has a GGGAA penta-loop and a G-C rich 5 nucleotide stem loop.

3.5. Predicted structure of the PAP binding site in Rev RNA

In order to predict the structure of the PAP binding site in Rev RNA, I used the reactivity of the 1M7 and DMS chemicals on the 3' end of Rev as constraints for ViennaRNA and Mfold folding algorithms of the Rev RNA C fragment. Briefly, the footprinting analysis included a control for chemically treating the RNA alone without PAP interaction. Since 1M7 only acylates solvent accessible, single stranded RNA (at the 2-OH of ribose) and DMS only methylates nucleotide bases not undergoing hydrogen-bonding (N7 of guanosines, N1 of adenosines, and N3 of cytidines), only the conformationally flexible and single stranded RNA will be modified. Strong modifications of the RNA were constrained to be single stranded in predicting the RNA structure using both folding algorithms. Comparing various iterations of the folded C fragment, a short hairpin loop over the PAP binding site was consistently predicted and matched the loop structure predicted from the folding the 40 nucleotide strand in **Figure 13**. The PAP binding site stem-loop is shown in **Figure 14 A**. To get a better idea of how PAP binds to the loop element of the RNA I decided to predict the 3D loop structure using a 3D modeling program, RNAcomposer. RNAcomposer accepts secondary predicted structure, splits apart the major elements that make up the RNA, and uses the RNA FRABASE dictionary relating to known RNA tertiary elements from literature to predict the 3D structure of each element.

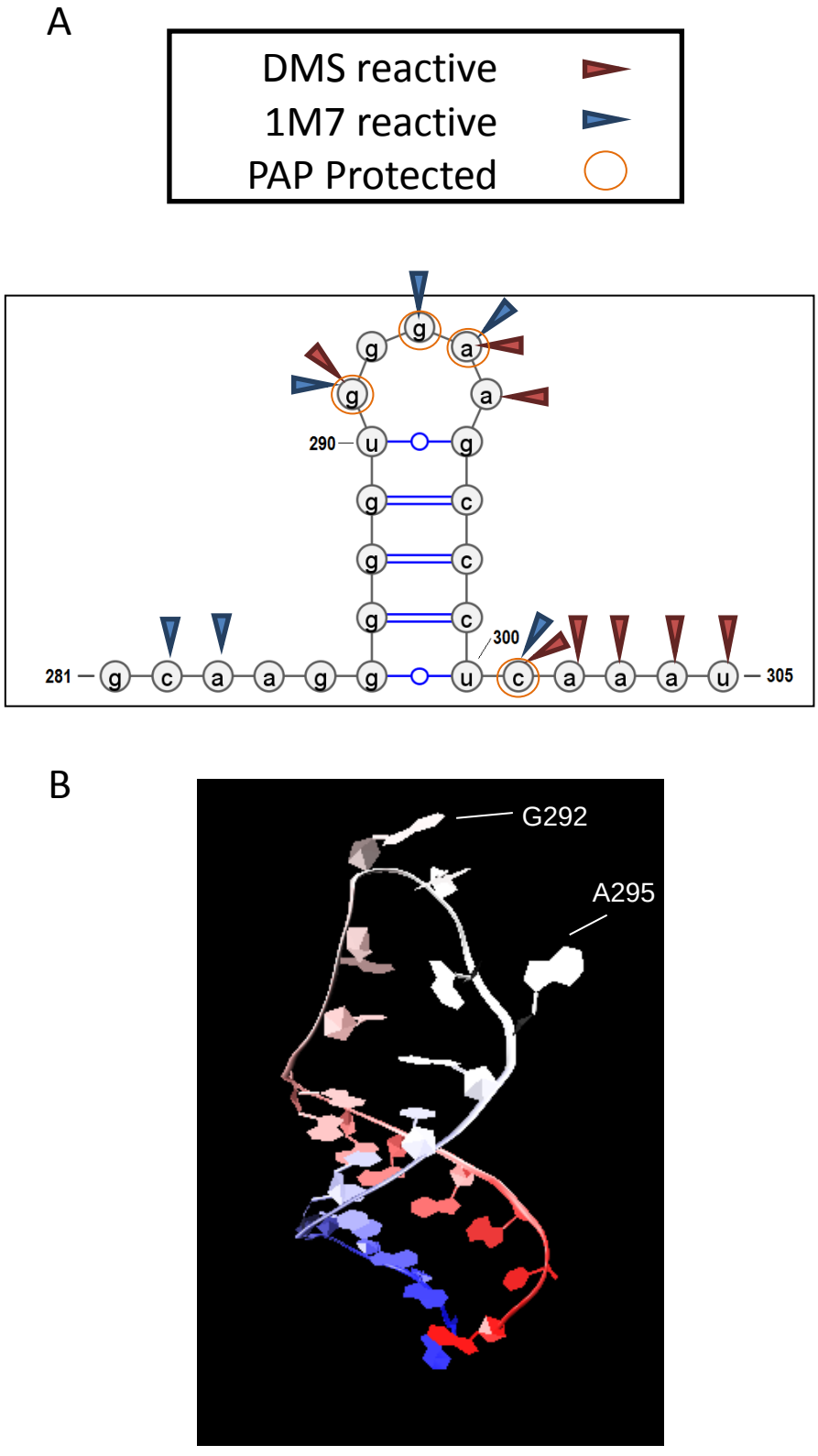
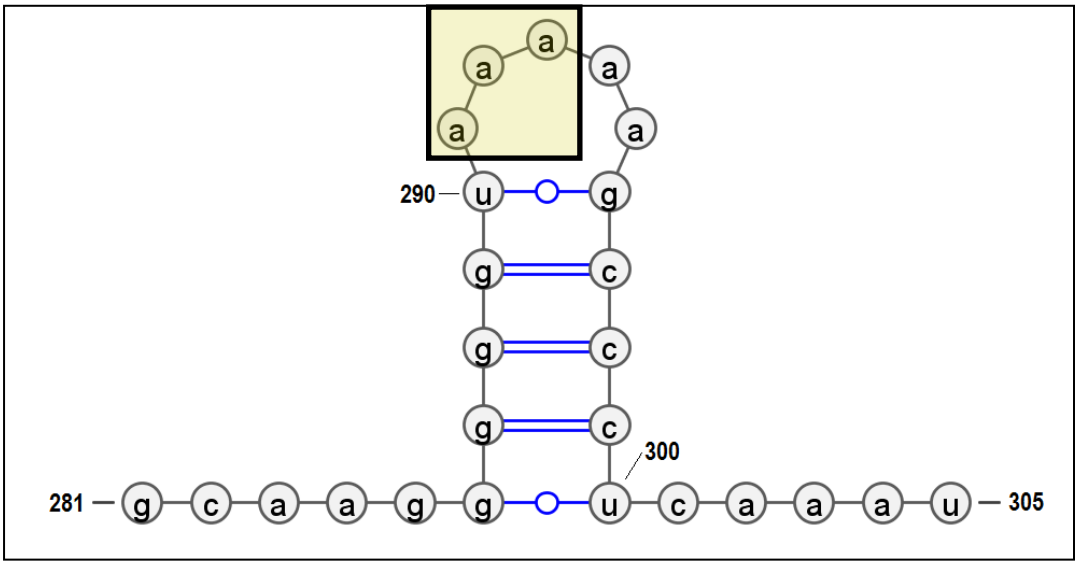


Figure 14 - Secondary and tertiary structure of PAP binding site in Rev RNA.

Figure 14 - A) Secondary structure of a short hairpin loop found at the 3' end of full length Rev RNA predicted by ViennaRNA and MFold. The blue arrows represent the 1M7 reactive nucleotides, red arrows DMS reactive nucleotides, and circled nucleotides are PAP protected footprints. **B)** 3D modeling of the loop structure in Rev was predicted using a combination of MFE secondary structure folding from MFold and ViennaRNA and 3D element prediction program RNAcomposer. RNAcomposer uses an online FRABASE database that splits elements of experimentally determined RNA structures and assembles these elements based on the secondary structure of a given sequence.

As a final step, the program assembles the elements into a single 3D model and structure refinement algorithms are applied to help with aligning favourable base-pairing and stacking interactions. Since PAP's active site only has room to bind a purine, I hypothesized that the stem loop must swing out the G291 and A294 nucleotides for PAP to gain access. The predicted 3D model is shown in **Figure 14B**. G292, G293 and A295 appear to swing out of the loop while G291 and A294 are folded into the loop structure. A pseudo GNRA loop structure appears to form between G291 and A294 nucleotides, where G291 and A294 are within close proximity to each other to base-pair and G293 can potentially stack with A294. I expected PAP to bind to the hairpin loop in a similar fashion as the SRL, so this pseudo-GNRA formation may explain how PAP consistently targets this site. Since PAP does not bind to the 291-293 mutant of Rev RNA, similar 3D modeling was conducted on the mutated stem-loop in order to see the difference in the loop structure (**Figure 15**). The GNRA sequence is not present in the given AAAAA loop sequence due to a lack of a guanosine base to form a G-A non-canonical base pairing. The 3D model shows that A292 is swung out of the structure while the remaining loop nucleotides are oriented towards the center. A291 and A293 as well as A294 and A296 appear to be oriented to base stack, which push away A292 and A295 from the central loop. Unlike A292, A295 is pointed in the direction of the RNA backbone.

A



B

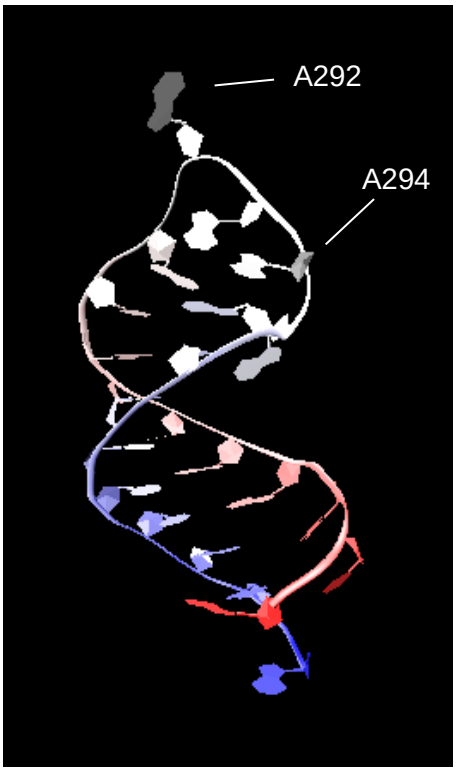


Figure 15 - Secondary and tertiary structure of mutant PAP binding site in Rev RNA.

Figure 15 - A) Secondary structure of the mutated hairpin loop found at 3' end of full length Rev RNA predicted by ViennaRNA and Mfold. Boxed nucleotides are the mutated bases **B)** 3D modeling of the mutant loop was also predicted using a combination of secondary structures from Mfold and ViennaRNA with the RNAcomposer modeling algorithm.

4. Discussion

4.1. Overview

The purpose of this project was to gain insight into how PAP can target a viral RNA over a cellular RNA by observing how PAP binds to HIV-1 Rev RNA. Specifically, I attempted to distinguish the importance of either a conserved primary sequence or a secondary structure of RNA in the binding of PAP to Rev RNA. PAP is a ribosome inactivating protein that removes an adenine from the SRL of rRNA. This process results in translation inhibition, which is toxic for a cell. However, there are many examples of PAP directly affecting plant and animal viruses that is separate from its RIP activity. For example, PAP can depurinate the RNA of BMV and HIV-1, which greatly inhibits their protein expression and life cycle in the host cell. 293T cells expressing PAP show no symptoms of toxicity or reduction in global translation and yet co-expression of PAP with HIV-1 shows a significant reduction of viral protein expression. The lack of toxicity also suggests that PAP does not extensively depurinate cellular RNAs. It is not known if PAP preferentially targets viral RNA over other cellular RNA. I sought to discover the sequence or structure in a viral RNA that PAP targets to further elucidate the RNA substrates that PAP preferentially binds and depurinates.

4.2. Preparation of Rev RNA for PAP binding

I first wanted to determine whether PAP could specifically bind to HIV-1 Rev RNA in comparison to other cellular RNAs. In order to test this, I chose to conduct various *in vitro* binding assays of PAP to Rev RNA. However, *in vitro* RNAs are not subject to the same conditions as they are in a cell because they are not capped and

poly-A tailed, they lack RNA binding proteins normally found in the cell, and they may be folded or localized differently *in vivo* that would result in an conformationally different RNA. In addition, the Rev RNA chosen in this experiment contains only the ORF, minus the 5' and 3' non-coding regions originating from the HIV-1 spliced mRNA. I thought that if PAP could depurinate *in vitro* transcribed Rev ORF similarly to the *in vivo* Rev RNA, the primary elements required for PAP recognition of the Rev RNA would be intact and PAP binding would be retained. In all binding and enzymatic assays, the *in vitro* RNA was heat denatured and re-folded before treatment with PAP. Re-folding promotes the formation of a single homogenous mixture of thermodynamically folded RNA which is important for selecting the most stable structures and for the reproducibility of the experiment. My results show that there is a PAP dependent depurination of a G291 and A294 of the Rev RNA *in vitro*. These depurination sites agree with reported *in vivo* depurination sites at G8594 and A8597.

4.3. PAP binding to RNAs

The results of my northwestern, EMSA, and filter binding assay show that PAP binds to Rev and SRL RNA significantly more than GAPDH RNA or Arg-tRNA and that binding to Rev RNA is about 5 times less than the SRL. These results indicate that: 1) PAP can selectively bind to different RNA substrates based on RNA sequence or structure alone, and 2) PAP can specifically bind to Rev RNA to a similar level as the SRL. These findings are interesting because it suggests that PAP may utilize the sequence or structure of RNA to direct its enzymatic specificity. As a RIP, PAP has

been extensively studied for its ability to bind and depurinate ribosomes. PAP has been shown to bind to the L3 of ribosomes and is presumably used as a mechanism to access the SRL for depurination from intact ribosomes[39,46,47]. However, PAP has antiviral activity that is separate from its RIP activity[32,34], and PAP can directly target viral RNA for depurination[20,24,29]. These findings prompted the investigation of alternate mechanisms by which PAP can specifically target viral RNA. It was shown that PAP can bind to the m7G cap of mRNA by holding the cap in its active site[58,76] and PAP binds the eukaryotic initiation factor eIF4G and iso-eIF4G[68]. Using these interactions, PAP could potentially bind to all capped mRNAs and indirectly to uncapped RNAs containing IRES using the eIF4G protein. Unfortunately, these findings do not separate how PAP may specifically interact with viral RNAs over cellular RNAs. My findings suggest that PAP can utilize the primary or secondary structure of RNA to specifically target one RNA over another. PAP potentially accesses all translating RNA by binding to ribosomes or mRNA but selectively depurinates RNA based on a sequence or structure.

Based on my findings, PAP binds to the SRL with similar affinity as the Rev RNA. If PAP binds to both a viral RNA and an rRNA to equal amounts, this would suggest that binding selectivity of PAP to unmodified RNA is not enough to determine its antiviral activity. To illustrate, approximately 80% of the total RNA in a given cell is ribosomal RNA, 15% is tRNA, and only about 4% is mRNA[77]. If PAP binds to both viral RNA and rRNA equally and has equivalent enzymatic activity on each substrate, PAP would depurinate more ribosomal RNA than viral RNA due to its greater availability. However, PAP has a greater effect on viral protein expression than it does on global translation rates when expressed in mammalian cells, so the scenario is more complex than this.

Viral RNAs contain many complex secondary structures required for their lifecycle that typical eukaryotic cellular mRNAs do not have. For example, HIV-1 contains a strong Tar element required for binding a trans-acting transcription activator (Tat) that phosphorylates RNA polymerase during the transcription of the HIV-1 genome. The HIV-1 gRNA also contains a psi packaging signal for localizing the RNA genome to an assembly site in the cell and several secondary structures between poly-protein regions to slow down translation rates and increase folding time of each protein[71]. PAP may have a greater effect on viral RNA than cellular RNA because there are more substrates for depurination per molecule of RNA. For example, The 28S rRNA is only targeted by PAP at the SRL on ribosomes, but HIV-1 has been shown to be depurinated within the ORF of Rev[29] and data from our lab show depurination in Tat and Vif[66,67]. The fact that PAP has been shown to bind to the ribosome does not mean that the SRL is the only substrate that PAP can depurinate. Viral mRNAs also use the translational system of the host to synthesize their proteins, so PAP may access viral RNA at a translation complex. PAP may target viral RNA for depurination more often because viral RNAs contain more structural elements PAP can target than the cellular RNAs.

It should be noted that the binding constants obtained for the SRL RNA in my work are much larger than what is reported for the 27 nt loop in the literature[37], which correlates to lower binding affinity. This is likely due to the fact that the RNA substrates used in this experiment are much larger. Due to the flexibility of the RNA at larger sizes, more than one conformation of the RNA tends to form, so specific bind sites may not always be present. As seen in the EMSA data, not all of the RNA fold into the same conformation under native conditions such that multiple bands appears in the gel. As

such, the true binding affinity for PAP to Rev RNA may be masked by the long RNA strand, especially if a small binding element like the SRL is targeted.

As expected, PAP does not bind to the GAPDH RNA. It has previously been shown that GAPDH is not depurinated by PAP[24], so it makes sense that PAP does not bind. I also wanted to test PAP binding to Arg-tRNA for three reasons: 1) tRNA makes up about 15% of the total RNA in cells and I wanted to see if PAP could interact with these RNAs[77], 2) tRNAs have very unique structure, 3) HIV-1 utilizes tRNA for synthesizing a cDNA intermediate of its genome, a major step in the viral life cycle that PAP may target[71]. The results show that PAP does not bind to Arg-tRNA. This may be because the *in vitro* transcribed tRNA does not correctly mimic other tRNA found in nature. The Arg-tRNA lacks proper methylation and pseudouridine modifications within the structure which are important for folding[78]. The presence of multiple bands in the EMSA native gel supports the idea that Arg-tRNA is not folded correctly. Although the tRNA used here is not the same as the native form, the significance of these data is that PAP does not bind any RNA randomly.

4.4. The importance of sequence or structure for PAP binding to RNA

Since PAP was found to bind to the Rev RNA to a similar degree as the SRL, I want to identify what it was about the primary or secondary structure of the RNA needed for PAP interaction. I first searched the literature for known *in vivo* depurinations of PAP in different viral RNAs (including the depurination site in HIV-1 Rev, HTLV-1 Gag, and BMV RNA3) and compared them using BLAST and ClustalW sequence alignment

programs. I could not find sequence homology at those sites, suggesting that PAP does not recognize sequence for its enzymatic activity. It is thought that the unique structure of the SRL tetraloop is essential for most RIP activity. However, PAP has been shown to bind and depurinate various mutants of the SRL, including mutations that abolish the tetraloop structure and loop sequence (from GAGA to GAAA)[62]. Mutation of the loop sequence to UUUU completely abolishes binding[39]. Unfortunately, exhaustive mutations of the SRL have not been published regarding how PAP binds and depurinates different hairpin structures. Since PAP requires either an adenine or guanine for binding in its active site, I hypothesize that PAP is dependent on access to a purine residue on RNA to bind and that the secondary structure of the RNA is important for placing purine bases in an accessible form for PAP's N-glycosidase activity. PAP's ability to depurinate a broad range of RNA substrates begs the question, what kind of RNA structures can PAP bind and depurinate? Understanding what kind of RNA structures PAP binds to could be used to identify potential targets for PAP activity.

4.5. The PAP binding site in Rev RNA

My footprinting data show that there is likely a binding site in the 3' end of Rev RNA that correlates well with the depurination site shown previously *in vivo* for HIV-1 RNA. Other PAP dependent binding sites also appear 3' of the binding site, but do not align with any depurination sites. It is possible that through interacting with the nucleotides G291-A294, PAP also binds to nearby residues. PAP is a positively charged protein[34], so it is not unlikely for it to interact with the negatively charged backbone of

any RNA non-specifically. In fact, at higher concentrations of PAP, there were no termination bands present after primer extension, presumably due to PAP coating the entire RNA with protein resulting in no modification of the Rev RNA by DMS or 1M7 (data not shown). This assay could be repeated using a gradient of PAP to tweeze apart specific versus non-specific binding. However, the consistent PAP dependent footprint between G291 and A294 in both the DMS and 1M7 treatments and the fact that the protection overlaps with the depurination sites, indicates that PAP likely binds here. The PAP binding site is expected to be small because the active site is only large enough to hold two purine bases[33]. There are residues near the active site that are involved in recognizing the ribosome[31], but binding to RNA alone is almost entirely dependent on the active site residues binding to a purine[2,38,39]. Interestingly, there is no literature showing footprinting data of PAP binding to the SRL. This may be because the catalytic activity of PAP is faster on the SRL than the binding site in Rev RNA due to the purine substrate difference. For example, depurination of adenines from the rRNA has been shown to be about 100x faster than guanines[43]. Comparing the SRL to the PAP binding site in Rev, the purine substrate for PAP in the SRL is a single adenine while the PAP binding site in Rev is both an adenine and guanine. The G291 of Rev RNA is bound and depurinated by PAP, indicating that the nitrogenous base must sit in the active site. Crystal structures of PAP with guanine shows that the purine sits in the binding site more tightly than adenine and is oriented 180^{o} away from the catalytic residues of PAP[43]. It is possible that PAP binds to the guanine residue in the Rev RNA and sits on the hairpin loop for longer than it would the SRL. My data show that PAP binding to Rev and the SRL are similar, but the extended length of the RNA substrates may not

accurately reflect the affinity PAP has for the small binding site in Rev. PAP has previously been shown to bind to the m7G cap of mRNA at nM concentrations[58]. It was suggested that this provides a mechanism by which PAP can access an mRNA for depurination. Since PAP does not depurinate guanines as quickly as adenines, PAP could bind to the exposed G nucleotide in Rev RNA in such a way that it may remain bound to the RNA template for longer and increase the chance of depurination of nearby sites. PAP does depurinate the G291, but to a less extent that A294. It is possible that PAP's lower enzymatic activity on G nucleotides, but tighter binding, may result in PAP lingering on the RNA longer.

Strong stops of the reverse transcriptase at the PAP binding site made it difficult to observe primer extension beyond the 3'end of the RNA. This is usually indicative of a strong secondary structure that the RT has trouble polymerizing through. Footprinting analysis throughout the remaining Rev RNA was not necessary because the filter binding assay of the Rev fragments show that PAP only binds to the 3' end of Rev RNA. In addition, mutation of the PAP binding site 291GGGA294 to AAAA resulted in a reduction in the binding capacity for the full length Rev RNA. These results suggest that PAP binding to Rev RNA is dependent on an element in the 3' end of Rev only, and specifically to the nucleotides G291-A294.

Modeling the secondary structure of the C fragment of Rev RNA in MFold and ViennaRNA revealed a short, consistent 15nt hairpin on the PAP binding site. The predicted secondary structure correlates well with the DMS and 1M7 chemical data because the reactive nucleotides were found on single stranded and loop structure bases. The stem loop consists of 3 Watson-Crick G-C pairs, two non-canonical G-U

basepairs, and a GGGAA pentaloop. 3D modeling of the hairpin loop using RNAcomposer shows that G292, G293, and A295 swing out of the loop structure while G291 and A294 are turned into the loop. This is the first pentaloop structure that PAP has been shown to interact with. I believe that this structure may be able to form a pseudo-GNRA structure, similar to what PAP binds to on the SRL. The SRL contains a GNRA tetraloop which is composed of a G to A non-canonical base pair, a purine (R) to adenine pi stack and an R to G non-canonical basepair[60]. These loops are typically closed by a Watson-crick basepair at their bases. **Figure 16A** shows a typical GNRA tetraloop structure. The hairpin pentaloop shows G291 and A294 in close proximity that may form a similar G291-A294 sugar/hoogsteen (sheared) basepair and an A293 to A294 base-stack just like a typical GRNA tetraloop. However, I refer to this as a pseudo-GNRA structure because the hairpin loop consists of 5 nucleotides, which may not allow these base-pair or pi stack interactions to occur. However, the additional nucleotide in the loop would allow for more flexibility in the backbone and the GNRA sequence requirement for the formation of the tetraloop may be all that PAP requires for binding and depurinating. For example, the SRL tetraloop was mutated such that the Watson-Crick base-pair closing the loop opened, resulting in a hexaloop, rather than a tetraloop. PAP was still able to depurinate the mutated SRL, regardless of the formation of the tetraloop[62]. The predicted 3D model of the PAP binding site in Rev does not correlate well with the chemical data, because G291, G293 and A294 should be exposed to chemical modification. It is likely that the pentaloop structure is flexible enough to shift loop-nucleotide interactions and the predicted 3D model may not entirely represent what the loop structure looks like in solution. As such, it can be argued that a

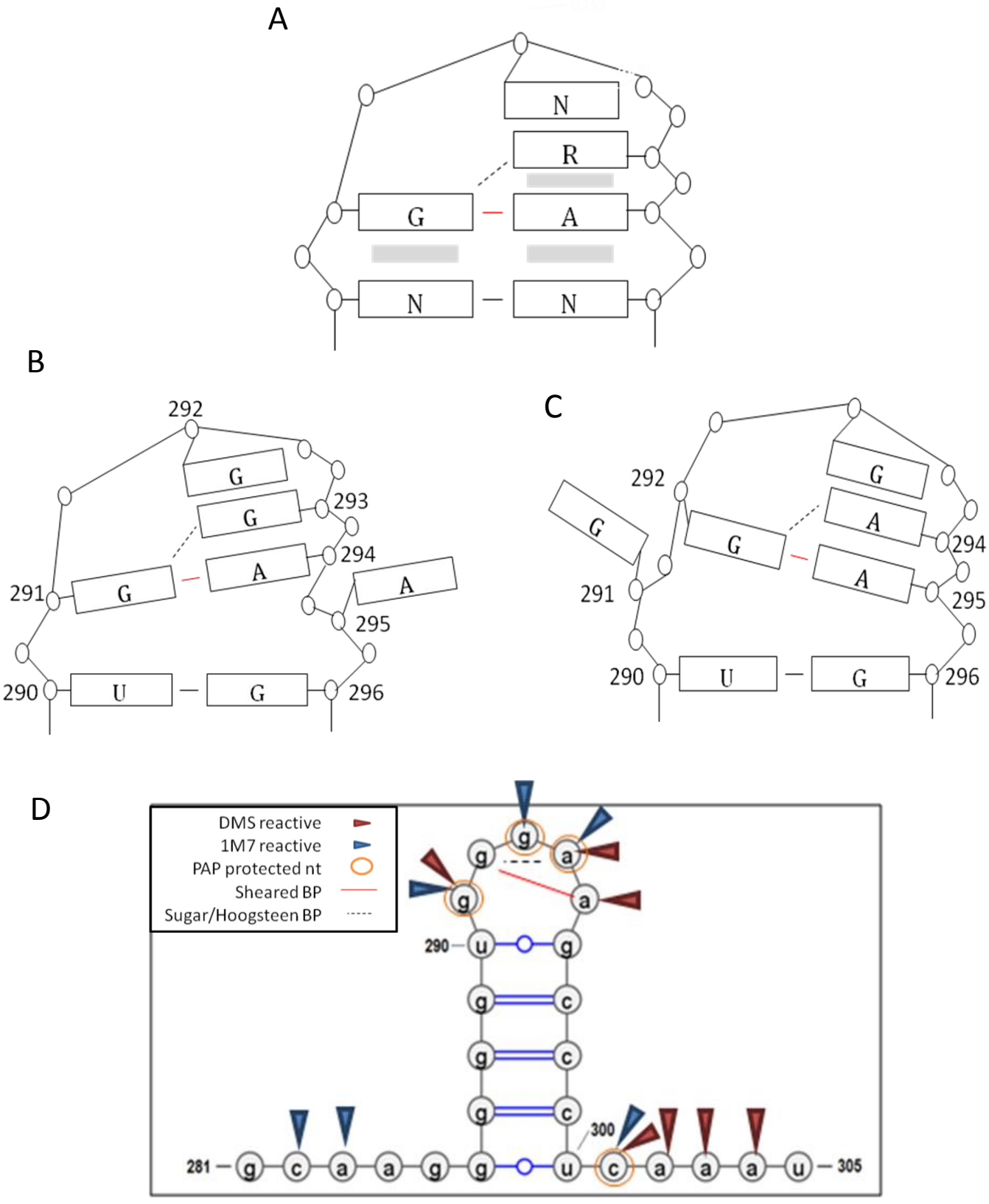


Figure 16 – PAP binding site in Rev may form a flexible pseudo-GNRA structure.

Figure 16 - A) Schematic of a GNRA tetraloop structure (N is A, C, G, or U; R is A or G). G and A nucleotides will basepair using the Sugar-Hoogsteen edges (red solid line) and the 2'OH of G will hydrogen bond the Hoogsteen edge of the purine (dashed line). The interaction is stabilized by pi stacking of the purine to adenine and the loops are typically closed by a Watson-Crick basepair. The 3D model of the PAP binding site in Rev RNA positions G291 and A294 in close proximity to basepair. A pseudo-GRNA formation may occur between **B)** G291 to A294 or **C)** G292 and G295. **D)** Chemical analysis of 15nt hairpin loop and the predicted base pairs. The blue arrows represent the 1M7 reactive nucleotides, red arrows DMS reactive nucleotides and circled nucleotides are PAP protected footprints.

similar pseudo-GNRA structure can be formed through G292 base-pairing with A295, A294 pi stacking with A295, and A294 base pairing with G292. In this scenario, G291, G293, and A294 would be swung out of the loop, making them accessible to modification by 1M7 and DMS. In addition, the reactivity of A295 with DMS and not 1M7 may be explained by a non-canonical base-pair interaction with G291 which may hide the 2'OH of the ribose, but not the N1 of adenine, from chemical probing. NMR or X-ray crystallography would reveal the most stable conformation of the loop. The two possible GNRA loops are illustrated in **Figure 16B and C**. Similar to how PAP depurinates the mutated SRL that abolished the closed tetraloop but retained the GAGA sequence[62], PAP may not require a perfect tetraloop formation to bind and depurinate the pentaloop in Rev RNA shown in my results.

4.6. Mutagenesis of the PAP binding site in Rev RNA

I wanted to test the significance of the nucleotides that PAP directly binds to in the Rev RNA by mutating 291GGG293 to AAA. Filter binding assay of PAP to the mutant Rev resulted in a much lower binding capacity, suggesting that PAP may have a lower affinity for the mutated binding site. Structural analysis of the mutant loop shows that the pseudo-GNRA formation is not possible because of the missing guanines. It is possible that since the pseudo-GNRA structure is not possible with this mutant, PAP cannot access the purine substrate to bind. However, based on the 3D model of the mutated loop, A292 appears prominently swung out of the loop structure, so it is not understood how PAP cannot bind this RNA. The catalytic activity of PAP on the wild-type loop vs the mutant loop would have to be tested in order to determine whether the

GGGAA sequence or structure is important for PAP activity as well as binding. This can be done by treating wild-type and mutant Rev RNA with PAP and subsequently treating the depurinated nucleotides with aniline. Aniline is a chemical that cleaves RNA at a depurinated base, which can be observed on an agarose gel. If PAP depurinates the mutated RNA as well as the wild-type, it can be concluded that PAP is not dependent on GGGAA sequence for depurination, but rather access to a purine base at the end of a hairpin loop. The difference in binding may suggest that PAP binds tighter to hairpin loops with guanosine nucleotides, since PAP binds more tightly to guanine in the active site. A competition assay can be conducted to see if the GGGAA wt loop exchanges with PAP faster than the AAAAA mutant. Alternatively, if PAP does not depurinate the mutant RNA, then PAP may be dependent on the pseudo-GNRA formation for binding.

4.7. Conclusions

PAP has broad range antiviral activity that has potential therapeutic and agricultural applications. Studying how PAP targets a viral RNA substrate will help in understanding the specificity of PAP in a cell. Specifically, I attempted to distinguish the importance of either a conserved primary sequence or a secondary structure of RNA in the binding of PAP to Rev RNA. My results show that PAP can specifically bind to Rev RNA and that PAP has differential binding to different RNA substrates. This means that PAP has dependence on either the secondary structure or sequence of naked RNA for binding. My results show that PAP can bind to the Rev RNA at the 3' end to a short 15 nt hairpin loop. This is the first pentaloop structure that PAP has been shown to

specifically bind. The loop formation may form a pseudo-GNRA structure, synonymous to the SRL tetraloop structure. Sequence analysis of the depurination sites in viral RNA by PAP reveals no sequence homology. Mutation of the PAP binding site in Rev results in a significant reduction in binding capacity for the RNA. The mutation removes the pseudo-GNRA structure of the wild type, but exposed adenines in the loop may still be targeted by PAP for depurination. My results suggest that PAP requires access to a purine nucleotide to bind and secondary structure of the RNA is important for facilitating this binding. In addition, PAP binding to a single RNA hairpin in Rev RNA is not enough to describe PAP's selection of viral RNA over the SRL. However, viral RNA are much more complex in structure than most cellular RNA in eukaryotes and viruses like HIV-1 have been shown to have more than one depurination site per molecule of RNA. Viral RNA may contain more secondary structures that PAP can target than PAP can depurinate. The complete secondary structure of the HIV-1 genome has been reported[71]. Understanding the substrates that PAP preferentially binds to would assist in predicting the substrates of depurination on viral RNA.

4.8. Future goals

So far, this research has discovered a new RNA structure that PAP binds to *in vitro*. However, it is unknown whether this structure is stable *in vivo* and can be specifically targeted by PAP *in vivo*. In order to test this, the 15 nt loop structure can be cloned upstream of the luciferase ORF in a mammalian expression vector. Transfection of this construct in 293T is expected to transcribe an RNA that is subsequently

translated into luciferase. If the stem loop structure is stable *in vivo*, then co-expression of PAP and the luciferase construct in the cytoplasm of 293T cells would result in PAP depurinating the luciferase mRNA. The placement of the loop upstream of the luciferase ORF should place the depurination site before the first ATG methionine and hence stall the ribosome from translating the enzyme. The amount of RNA that is depurinated can be correlated to an activity assay for luciferase bioluminescence in luciferin. A positive control can be made by placing the SRL in place of the Rev hairpin. A negative control for a loop that does not get depurinated is the SRL with the GAGA tetraloop mutated to UUUU. If the assay works well, several unique loop structures can be tested to observe a greater number of RNA substrates. For example, PAP can be tested for its ability to depurinate different combinations of tetra, penta, or hexaloop sequences at the end of the hairpin. Depurination can be tested on these RNAs by primer extension. This would tell us if PAP can inhibit translation by depurination of the RNA or by simply binding to the RNA and blocking the ribosome.

This research asks the question, what does PAP specifically bind to? PAP has never been exhaustively tested for the different RNA substrates that it can bind and depurinate. Binding assays and enzyme kinetics should be conducted on various 20-30 nt hairpin loops in order to determine PAP's preferred substrate. Surface plasmon resonance can be used to detect accurate binding constants and HPLC for adenine release can be used to calculate the enzyme activity of PAP against these hairpin loops. Determining the structure of RNA that PAP preferentially binds to would help in finding targets of PAP depurination or help in designing mutations of RNA that abolish PAP activity.

PAP expression in 293T cells has a significant effect on the expression of viral proteins and is non-toxic to the cell. The mechanism by which PAP can specifically inhibit viral protein translation over global translation is not understood. One possibility is that PAP may depurinate viral RNAs more than rRNA. Knowing whether PAP preferentially targets the viral RNA over cellular RNA would suggest that the mechanism of PAP inhibiting virus expression is primarily through depurination of the viral RNA. Alternatively, if depurination is the same or less than cellular RNA, then the effect of depurination on viral RNA is far greater than on rRNA or the effect of depurinating the rRNA has a greater impact on the virus life cycle. In order to test this, a northern blot can be conducted to test for the ratio of rRNA and HIV-1 RNA that is depurinated by PAP. In this assay, 293T cells would be transfected with either PAP or the proviral clone for HIV-1 co-transfected with PAP. After 48 hours, total RNA would be collected and split into two aliquots. One aliquot would be treated with aniline and the other would not. Aniline will cleave the RNA at a depurinated base. Total RNA will be separated through a denaturing urea / acrylamide gel and subsequently transferred to a nylon membrane. The RNA would be cross-linked to the nylon membrane and then hybridized with antisense RNA probes for either HIV-1 RNA or rRNA. The probe for the HIV-1 RNA would be specific to a region within the 2KB class of spliced genome in order to observe the 9, 4 and 2KB classes of RNA. With PAP treatment, only the 2KB class of RNA is present in the cytoplasm due to the altered spliced ratio of HIV-1 RNA caused by reduced expression of Rev. Rev expression is required for pulling out 9 and 4KB classes of HIV-1 RNA from the nucleus[79,80]. The probe for the 28S rRNA would be specific for the 3' end of the RNA, downstream of the SRL. Aniline treatment would

result in cleavage of both RNAs because both are targeted by PAP for depurination. Cleavage products should be detected on the gel as a smaller sized band. The relative amount of depurination can be assessed by the ratio of the cleaved RNA band and the total band intensity of the non-treated RNA.

5. References

1. Stirpe, F. Ribosome-inactivating proteins. *Toxicon* **44,** 371–83 (2004).

2. Endo, Y., Mitsui, K., Motizuki, M. & Tsurugi, K. The Mechanism of Action of Ricin and Related Toxic Lectins on Eukaryotic Ribosomes. *J Biol Chem* **262,** 5908–5912 (1987).

3. Endo, Y. & Tsurugi, K. The RNA N-glycosidase activity of ricin A-chain. *Nucleic Acids Symp. Ser.* **263,** 139–42 (1988).

4. Obrig, T. G., Irvin, J. D. & Hardesty, B. The Effect of an Antiviral Peptide on the Ribosomal Reactions of the Peptide Elongation Enzymes, EF-I and EF-II. *Arch Biochem Biophys.* **155,** 278–289 (1973).

5. García-Ortega, L., Álvarez-García, E., Gavilanes, J. G., Martínez-del-Pozo, Á.´ & Joseph, S. Cleavage of the sarcin-ricin loop of 23S rRNA differentially affects EF-G and EF-Tu binding. *Nucleic Acids Res.* **38,** 4108–4119 (2010).

6. Shang, C., Peumans, W. & Van Damme, E. J. M. in *Ribosome-inactivating Proteins: Ricin and Related Proteins* (eds. Stirpe, F. & Lappi, D. A.) 11–27 (John Wiley & Sons, Ltd., 2014). doi:10.1002/9781118847237.ch2

7. Irvin, J. Purification and partial characterization of the antiviral protein from Phytolacca americana which inhibits eukaryotic protein synthesis. *Arch Biochem Biophys.* **196,** 522–528 (1975).

8. Yeung, H. W., Poon, S. P., Ng, T. B. & Li, W. W. Isolation and Characterization of An Immunosuppressive Protein From Trichosanthes Kirilowii Root Tubers. *Immunopharmacol. Immunotoxicol.* **9,** 25–46 (1987).

9. Stirpe, F., Olsnes, S. & Pihl, A. Gelonin, a new inhibitor of protein synthesis, nontoxic to intact cells. Isolation, characterization, and preparation of cytotoxic complexes with concanavalin A. *J Biol Chem.* **255,** 6947–53 (1980).

10. Sandvig, K., Skotland, T., Van Deurs, B. & Klokk, T. I. Retrograde transport of protein toxins through the Golgi apparatus. *Histochem. Cell Biol.* **140,** 317–326 (2013).

11. Agapov, I. I. *et al.* The role of structural domains in RIP II toxin model membrane binding. *FEBS Lett.* **402,** 91–3 (1997).

12. Lin, J. Y., Tserng, K. Y., Chen, C. C., Lin, L. T. & Tung, T. C. Abrin and ricin: new anti-tumour substances. *Nature* **227,** 292–3 (1970).

13. Olsnes, S. The History of Ricin, Abrin and related Toxins. *Toxicon* **44,** 361–370 (2004).

14. Jiménez, P. *et al.* Ebulin from Dwarf Elder (Sambucus ebulus L.): A Mini-Review. *Toxins (Basel)* **7,** 648–658 (2015).

15. Karen, M. G. *et al.* Shiga Toxin Binding to Glycolipids and Glycans. *PLoS One* **7,** 10.1371/journal.pone.0030368 (2011).

16. Chaudhry, B. *et al.* The Barley 60 kDa Jasmonate-Induced Protein (JIP60) is a Novel Ribosome-Inactivating Protein. *Plant J.* **6,** 815–24 (1994).

17. Hey, T. D., Hartley, M. & Walsh, T. A. Maize Ribosome-Inactivating Protein (b-32) (Homologs in Related Species, Effects on Maize Ribosomes, and Modulation of Activity by Pro-Peptide Deletions). *Plant Physiol.* **107,** 1323–32 (1995).

18. Nielsen, K. & Boston, R. S. Ribosome-Inactivating Proteins: A Plant Perspective. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52,** 785–816 (2001).

19. Drugar, B. M. & Armstrong, J. K. The Effect of Treating the Virus of Tobacco Mosaic With the Juices of Various Plants. *Ann. Missouri Bot. Gard.* **12,** 359–366 (1925).

20. Karran, R. a. & Hudak, K. a. Depurination of Brome mosaic virus RNA3 inhibits its packaging into virus particles. *Nucleic Acids Res.* **39,** 7209–7222 (2011).

21. Taylor, S. *et al.* Correlation between the activities of five ribosome-inactivating proteins in depurination of tobacco ribosomes and inhibition of tobacco mosaic virus infection. *Plant J.* **5,** 827–835 (1994).

22. Lee-Huanga, S. *et al.* A new class of anti-HIV agents: GAP31, DAPs 30 and 32. *FEBS Lett.* **291,** 139–144 (1991).

23. Foà-Tomasi, L., Campadelli-Fiume, G., Barbieri, L. & Stirpe, F. Effect of ribosome-inactivating proteins on virus-infected cells. Inhibition of virus multiplication and of protein synthesis. *Arch. Virol.* **71,** 323–32 (1982).

24. Mansouri, S., Choudhary, G., Sarzala, P. M., Ratner, L. & Hudak, K. a. Suppression of human T-cell leukemia virus I gene expression by pokeweed antiviral protein. *J. Biol. Chem.* **284,** 31453–31462 (2009).

25. Duggar, B. M. & Armstrong, J. K. The Effect of Treating Virus of Tobacco Mosaic with Juice of Various Plants. *Ann Mol Bot Gard* **12,** 359–366 (1925).

26. Picard, D., Kao, C. C. & Hudak, K. a. Pokeweed antiviral protein inhibits brome mosaic virus replication in plant cells. *J. Biol. Chem.* **280,** 20069–20075 (2005).

27. Zarling, J. M. *et al.* Inhibition of HIV replication by pokeweed antiviral protein targeted to CD4+ cells by monoclonal antibodies. *Nature* **347,** 92–95 (1990).

28. Tourlakis, M. E., Karran, R. A., Desouza, L., Siu, K. W. M. & Hudak, K. A. Homodimerization of pokeweed antiviral protein as a mechanism to limit depurination of pokeweed ribosomes. *Mol. Plant Pathol.* **11,** 757–767 (2010).

29. Zhaborkritsky, A., Mansouri, S. & Hudak, K. Pokeweed antiviral protein alters splicing of HIV-1 RNAs, resulting in reduced virus production. *RNA* **20,** 1238–47 (2014).

30. Hudak, K. a, Wang, P. & Tumer, N. E. A novel mechanism for inhibition of translation by pokeweed antiviral protein: depurination of the capped RNA template. *RNA* **6,** 369–380 (2000).

31. Hudak, K. a. *et al.* Generation of pokeweed antiviral protein mutations in Saccharomyces cerevisiae: Evidence that ribosome depurination is not sufficient for cytotoxicity. *Nucleic Acids Res.* **32,** 4244–4256 (2004).

32. Tumer, N. E., Hwang, D. J. & Bonness, M. C-terminal deletion mutant of pokeweed antiviral protein inhibits viral infection but does not depurinate host ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 3866–71 (1997).

33. Arthur F. Monzingo, Edward J. Collins, Stephen R. Ernst, James D. Irvin, J. D. R. The 2·5 Å Structure of Pokeweed Antiviral Protein. *J. Mol. Biol.* **233,** 705–715 (1993).

34. Hur, Y. *et al.* Isolation and characterization of pokeweed antiviral protein mutations in Saccharomyces cerevisiae: identification of residues important for toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 8448–52 (1995).

35. Lin, Q., Chen, Z., Antoniw, J. & White, R. Isolation and characterization of a cDNA clone encoding the anti-viral protein from Phytolacca americana. *Plant Mol Biol.* **17,** 609–14 (1991).

36. Di Maro, A., Citores, L., Russo, R., Iglesias, R. & Ferreras, J. M. Sequence comparison and phylogenetic analysis by the Maximum Likelihood method of ribosome-inactivating proteins from angiosperms. *Plant Mol. Biol.* **85,** 575–588 (2014).

37. Rajamohan, F., Mao, C. & Uckun, F. M. Binding Interactions between the Active Center Cleft of Recombinant Pokeweed Antiviral Protein and the a-Sarcin/Ricin Stem Loop of Ribosomal RNA. *J. Biol. Chem.* **276,** 24075–24081 (2001).

38. Kurinov, I. V., Myers, D. E., Irvin, J. D. & Uckun, F. M. X-ray crystallographic analysis of the structural basis for the interactions of pokeweed antiviral protein with its active site inhibitor and ribosomal RNA substrate analogs. *Protein Sci.* **8,** 1765–1772 (1999).

39. Rajamohan, F., Ozer, Z., Mao, C. & Uckun, F. M. Active center cleft residues of pokeweed antiviral protein mediate its high-affinity binding to the ribosomal protein L3. *Biochemistry* **40,** 9104–9114 (2001).

40. Chen, X.-Y., Berti, P. J. & Schramm, V. L. Transition-State Analysis for Depurination of DNA by Ricin A-Chain. *J. Am. Chem. Soc.* **122,** 6527–6534 (2000).

41. Roday, S. *et al.* Inhibition of Ricin A-Chain with Pyrrolidine Mimics of the Oxacarbenium Ion Transition State. *Biochemistry* **43,** 4923–4933 (2004).

42. Rajamohan, F., Pugmire, M. J., Kurinov, I. V. & Uckun, F. M. Modeling and alanine scanning mutagenesis studies of recombinant pokeweed antiviral protein. *J. Biol. Chem.* **275,** 3382–3390 (2000).

43. Kurinov, I. V., Rajamohan, F., Venkatachalam, T. K. & Uckun, F. M. X-ray crystallographic analysis of the structural basis for the interaction of pokeweed antiviral protein with guanine residues of ribosomal RNA. *Protein Sci.* **8,** 2399–2405 (1999).

44. Domashevskiy, A. & Goss, D. Pokeweed Antiviral Protein, a Ribosome Inactivating Protein: Activity, Inhibition and Prospects. *Toxins (Basel).* **7,** 274–298 (2015).

45. Barbieri, L., Battelli, M. G. & Stirpe, F. Ribosome-inactivating proteins from plants. *Biochim. Biophys. Acta - Rev. Biomembr.* **1154,** 237–282 (1993).

46. Hudak, K. a., Dinman, J. D. & Tumer, N. E. Pokeweed antiviral protein accesses ribosomes by binding to L3. *J. Biol. Chem.* **274,** 3859–3864 (1999).

47. Di, R. & Tumer, N. E. Expression of a truncated form of ribosomal protein L3 confers resistance to pokeweed antiviral protein and the Fusarium mycotoxin deoxynivalenol. *Mol. Plant. Microbe. Interact.* **18,** 762–770 (2005).

48. Ready, M. P., Brown, D. T. & Robertus, J. D. Extracellular localization of pokeweed antiviral protein. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 5053–5056 (1986).

49. Lodge, J. K., Kaniewski, W. K. & Tumer, N. E. Broad-spectrum virus resistance in transgenic plants expressing pokeweed antiviral protein. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 7089–7093 (1993).

50. Mansouri, S., Kutky, M. & Hudak, K. Pokeweed antiviral protein increases HIV-1 particle infectivity by activating the cellular mitogen activated protein kinase pathway. *PLoS One* **7,** e36369 (2012).

51. Chen, Z. C., White, R. F., Antoniw, J. F. & Lin, Q. Effect of pokeweed antiviral protein (PAP) on the infection of plant viruses. *Plant Pathol.* **40,** 612–620 (1991).

52. Karran, R. a. & Hudak, K. a. Depurination within the intergenic region of Brome mosaic virus RNA3 inhibits viral replication in vitro and in vivo. *Nucleic Acids Res.* **36,** 7230–7239 (2008).

53. He, Y., Guo, C., Pan, Y., Peng, C. & Weng, Z. Inhibition of hepatitis B virus replication by pokeweed antiviral protein in vitro. *J. Gastroenterol.* **14,** 1592–1597 (2008).

54. Erice, a. *et al.* Anti-human immunodeficiency virus type 1 activity of an anti-CD4 immunoconjugate containing pokeweed antiviral protein. *Antimicrob. Agents Chemother.* **37,** 835–838 (1993).

55. Ramakrishnan, S. & Houston, L. L. Immunological and BiologicalStabilityof Immunotoxins in Vivoas Studiedby the Clearance of Disulfide-linkedPokeweedAntiviralProtein-Antibody Conjugatesfrom Blood1. **45,** 2031–2036 (1985).

56. Uckun, M., Turner, N., Irvin, D., Myers, E. & Fuller, G. B. Pharmacokinetic Features, Immunogenicity, and Toxicity of B43(anti-CD19)-Pokeweed Antiviral Protien Immunotoxin in Cynomolgus Monkeys. *Clin. Cancer Res.* **3,** 325–337 (1997).

57. Vivanco, J. M. & Tumer, N. E. Translation Inhibition of Capped and Uncapped Viral RNAs Mediated by Ribosome-Inactivating Proteins. *Phytopathology* **93,** 588–595 (2003).

58. Hudak, K. a, Bauman, J. D. & Tumer, N. E. Pokeweed antiviral protein binds to the cap structure of eukaryotic mRNA and depurinates the mRNA downstream of the cap. *RNA* **8,** 1148–1159 (2002).

59. Barbieri, L. *et al.* Polynucleotide:adenosine glycosidase activity of ribosome-inactivating proteins: Effect on DNA, RNA and poly(A). *Nucleic Acids Res.* **25,** 518–522 (1997).

60. Seggerson, K. & Moore, P. B. Structure and Stability of Variants of the Sarcin-Ricin Loop of 28S rRNA: NMR Studies of the Prokaryotic SRL and a Functional Mutant. *RNA* **4,** 1203–1215 (1998).

61. Mansouri, S., Nourollahzadeh, E. & Hudak, K. a. Pokeweed antiviral protein depurinates the sarcin/ricin loop of the rRNA prior to binding of aminoacyl-tRNA to the ribosomal A-site. *RNA* **12,** 1683–1692 (2006).

62. Marchant, A. & Hartley, M. R. The Action of Pokeweed Antiviral Protein and Ricin A-chain on Mutants in the α-Sarcin Loop of Escherichia coli 23S Ribosomal RNA. *J. Mol. Biol.* **254,** 848–855 (1995).

63. Szewczak, A. A., Moore, P. B., Chang, Y. L. & Wool, I. G. The Conformation of the Sarcin/Ricin Loop from 28S Ribosomal RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 9581–9585 (1993).

64. Gessner, S. L. & Irvin, J. D. Inhibition of Elongation factor 2-Dependent Translocation by Pokeweed Antiviral Protein and Ricin. *J. Biol. Chem.* **255,** 3251–3253 (1980).

65. Di, R. & Tumer, N. Pokeweed Antiviral Protein: Its Cytotoxicity Mechanism and Applications in Plant Disease Resistance. *Toxins (Basel).* **7,** 755–772 (2015).

66. Kutky, M. & Hudak, K. Regulation of HIV-1 transcription by pokeweed antiviral protein. (2012).

67. Krivdova, G. *(*2015).

68. Wang, M. & Hudak, K. a. A novel interaction of pokeweed antiviral protein with translation initiation factors 4G and iso4G: A potential indirect mechanism to access viral RNAs. *Nucleic Acids Res.* **34,** 1174–1181 (2006).

69. Hiscox, J. A. RNA viruses: hijacking the dynamic nucleolus. *Nat. Rev. Microbiol.* **5,** 119–127 (2007).

70. Vidalain, P. O. & Tangy, F. Virus-host protein interactions in RNA viruses. *Microbes Infect.* **12,** 1134–1143 (2010).

71. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460,** 711–6 (2009).

72. Davis, M., Sagan, S. M., Pezacki, J. P., Evans, D. J. & Simmonds, P. Bioinformatic and physical characterizations of genome-scale ordered RNA structure in mammalian RNA viruses. *J. Virol.* **82,** 11824–36 (2008).

73. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31,** 3406–3415 (2003).

74. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6,** 26 (2011).

75. Popenda, M. *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Res.* **40,** 1–12 (2012).

76. Baldwin, A. E., Khan, M. a., Tumer, N. E., Goss, D. J. & Friedland, D. E. Characterization of pokeweed antiviral protein binding to mRNA cap analogs: Competition with nucleotides and enhancement by translation initiation factor iso4G. *Biochim. Biophys. Acta* **1789,** 109–116 (2009).

77. Wendisch, V. F. *et al.* Isolation of Escherichia coli mRNA and Comparison of Expression Using mRNA and Total RNA on DNA Microarrays. *Anal. Biochem.* **290,** 205–213 (2001).

78. Kim, S. & Sussman, J. The general structure of transfer RNA molecules. *Proc. Nat. Acad. Sci. USA* **71,** 4970–4974 (1974).

79. Ocwieja, K. E. *et al.* Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.* **40,** 10345–55 (2012).

80. Frankel, A. & Young, J. HIV-1: fifteen proteins and an RNA. *Annu. Rev. Biochem.* **67,** 1–25 (1998).

81. Correll, C. C. *et al.* Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13436–13441 (1998).