

2019-02-01

Utilization of plant Puf RNA-binding proteins to manipulate endogenous mRNA physiology

Wen, Xin

Wen, X. (2019). Utilization of plant Puf RNA-binding proteins to manipulate endogenous mRNA physiology (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca>. <http://hdl.handle.net/1880/109909>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology

by

Xin Wen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

FEBRUARY, 2019

© Xin Wen 2019

Abstract

RNA-binding proteins have an important role in the regulation of post-transcriptional gene expression. The Puf (Pumilio) family of RNA-binding proteins bind to RNAs in a sequence specific manner. The RNA-binding domain of canonical Puf proteins (the Pumilio homology domain, PUM-HD) consists of eight Puf repeats. Each Puf repeat binds to RNA in a modular fashion, so that one Puf repeat binds to a single nucleotide. The nucleotide binding recognition code of Puf repeats has been well characterized. This one repeat:one nucleotide recognition code has allowed for the alteration of Puf repeats so that the PUM-HD can bind specific RNA targets. These altered Puf repeats have been fused to effector domains so that fusion proteins can affect RNA physiology in the cell. The aim of this thesis was to fuse an Arabidopsis Puf RNA-binding protein with unique binding characteristics to two types of effector domains. These fusion proteins were expressed in onion epidermal cells to determine if they could alter the stability or translation of a reporter mRNA. This research provides foundational evidence that highlights the potential for the utility of plant Puf proteins to alter gene expression for applied purposes.

Acknowledgements

First of all, I would like to thank Dr. Douglas Muench for being an exceptional supervisor for me. His advice and support throughout this project helped me learn a lot about what I should do for research and communication. I really thank him about his encouragement and his attitude toward solving a problem. Whenever there was a problem in the experiment, he could always point me out to the possible approach to solve and let me think about the experiment more deeply. I feel very lucky to have such a good supervisor, and I hope to take whatever he taught me into my future work and studies.

I would also like to thank my lab mates, both the present and former, who helped me a lot through my graduate program. I would like to thank Chi for training me with the experiments needed for my project. I might ignore many important details for my experiments if he hadn't trained me. I would like to thank Mitch for his help with the poly(A) polymerase experiments, and advice for my experiments. I would also thank Dilini, Jeremy, Chris, Jo and Gill for making my graduate program enjoyable and memorable.

I would thank Dr. Hansen and Dr. Barrette-Ng as my supervisory committee members. They gave me a lot of help when I had trouble with my research. Only with their advice I could improve my research approaches in different methods. I would also thank Dr. Mayi Arcellana-Panlilio to be my external supervisor and her patience for listening to my thesis.

I would like to thank my friends for always supporting me. I always had a good time with them during my stay in Canada. I hope that wherever I would be in the future, I could always keep the friendship with them.

Lastly, I would like to thank my family. They provided me with a lot of support even though they could not be at my side. They were always there to listen and provided help and advice whenever I was frustrated, excited or nervous.

Table of Contents

Abstract	ii
Acknowledgements	iii
Table of Contents.....	v
List of Tables	vii
List of Figures and Illustrations	viii
List of Symbols, Abbreviations and Nomenclature	x
Chapter One: Introduction.....	1
1.1. Post-transcriptional regulation.....	2
1.2.1. RNA-binding domains	5
1.2.1.1. RNA Recognition Motif.....	5
1.2.1.2. K Homology domain.....	6
1.2.1.3. Pentatricopeptide repeat (PPR)	7
1.2.1.4. Zinc-finger (ZnF) domains.....	8
1.2.1.5. Double-stranded RNA binding motifs (dsRBDs).....	9
1.2.1.6. The Puf family RNA-binding proteins	9
1.3. Plant Puf RNA-binding domain.....	12
1.4 APUM23	12
1.5. Puf protein engineering.....	17
1.5.1. RNA-binding protein engineering	17
1.5.2. Engineering Puf RNA-binding proteins.....	18
1.5.3. Artificial site specific RNA endonucleases and the PIN-domain	19
1.5.4. Using the poly(A) polymerase, GLD-2, to increase mRNA stability and translation rate	21
1.6 Hypothesis and objectives of thesis	23
Chapter Two: Materials and Methods	25
2.1 PCR and molecular cloning.....	25
2.1.1 Cloning of APUM protein encoding constructs to observe their localization	25
2.1.2 Cloning of the APUM-PIN encoding constructs for in vivo expression	27
2.1.3. Cloning of constructs encoding the RNA targets of APUM-PIN	31
2.1.4 PAP-APUM with Flag tag cloning	31
2.2 Onion epidermal cell transfection using biolistic particle bombardment.....	35
2.3 Microscopic analysis.....	36

2.4 Western blotting	36
2.5 Quantitative PCR (qPCR) analysis	37
Chapter Three: Altering the expression patterns of target mRNAs <i>in vivo</i> by fusing RNase and poly(A) polymerase domains to APUM proteins	40
3.1 Introduction	40
3.2 Results	41
3.2.1 Subcellular localization of wild type and truncated APUM23	41
3.2.2 Effect of a recombinant tAPUM–RNase fusion protein on regulating the degradation of RFP target mRNAs in onion epidermal cells.	42
3.2.2.1 – Epifluorescence microscopy of RFP expression	44
3.2.2.2 – Western blot analysis of RFP expression RFP expression	47
3.2.2.3 qPCR analysis of RFP mRNA levels	53
3.2.3 Effect of a recombinant tAPUM–PAP fusion protein on augmenting the polyadenylation of RFP target mRNAs in onion epidermal cells	55
3.2.3.1 Determination of RFP mRNA Poly(A) length	57
3.2.3.2 Quantification of RFP and APUM-PAP levels by western blotting.....	57
3.3 Discussion	62
Chapter Four: Discussion and Future Direction	65
References	69
APPENDIX 1	74
APPENDIX 2	78
License Agreement for Figure 1.1	78
License Agreement for Figure 1.2	79
License Agreement for Figure 1.3	81
License Agreement for Figure 1.4	82
License Agreement for Figure 1.5	84
License Agreement for Figure 1.6	86
License Agreement for Figure 1.7	88

List of Tables

Table. 2.1: PCR primers used for cloning in order to determine the localization of truncated APUM23.	28
Table. 2.2: Primers/sequences used for cloning 3X Flag Tag-APUM-Linker-PIN with	30
Table. 2.3: Primers for cloning APUM targets.....	33
Table. 2.4: Primers for qPCR reactions.....	39
Table. 3.1: Quantified expression level of RFP and APUM-PIN.....	50
Table. 3.2: Protein expression levels of RFP from mRNA containing the APUM23 target or APUM2 target relative to the expression of the APUM-PIN expression.	51

List of Figures and Illustrations

Figure 1.1: Post-transcriptional gene regulation in a cell.....	3
Figure 1.2: Crystal structure of PUM homology domain.....	11
Figure 1.3: The primary structure of Puf proteins in Arabidopsis and rice.....	13
Figure 1.4: Consensus sequences for different APUM proteins.....	15
Figure 1.5: Structure of the APUM23-RNA interaction.	16
Figure 1.6: Amino acid combinations in TRMs of common Puf proteins in nature and their target nucleotides.	20
Figure 1.7: The ASRE binds and cleaves the target RNA.	22
Figure 2.1: Model of APUMs structure to determine the localization of APUM proteins.	26
Figure 2.2: Structural arrangement of the APUM-PIN constructs.	29
Figure 2.3: Structural arrangement of the APUM target mRNA.....	32
Figure 2.4: Structural arrangement of the Puf-GLD-2 constructs.	34
Figure 3.1: Localization of APUM protein-RFP fusion in onion epidermal cells by epifluorescence microscopy.	43
Figure 3.2: Strategy for using an APUM-PIN fusion protein to cleave specific mRNAs.	45
Figure 3.3: Expression level of RFP-target mRNA in onion epidermal cells by epifluorescence microscope.....	46
Figure 3.4: Expression level of RFP-APUM targets and APUM-PIN proteins.	49
Figure 3.5: Relative expression levels of RFP and APUM-PIN protein from co-expression experiments.....	52
Figure 3.6: Levels of RFP mRNA in cells expressing the RFP construct alone or in combination with tAPUM23-PIN.	54

Figure 3.7: Strategy for using APUM proteins fused to the GLD-2 poly(A) polymerase (PAP) to enhance polyadenylation of specific mRNAs.....	56
Figure 3.8: Poly(A) tail length of CFP-APUM 23 target.	58
Figure 3.9: Determine the PAP activity of PAP-APUM23-Flag and Flag-APUM23-PAP	60
Figure 3.10: Expression level of CFP when co-expressed with APUM proteins fused to PAP. ...	61

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
5' -UTR	5'- untranslated region
3' -UTR	3'- untranslated region
µg	Microgram
µL	Microlitre
µM	Micromolar
APUM	<i>Arabidopsis</i> Puf protein
ASRE	Artificial site-specific RNA endonucleases
BME	2-mercaptoethanol
CFP	Cyan Fluorescent Protein
C-terminal	Carboxyl-terminal
DmPUM	<i>Drosophila melanogaster</i> Pumilio
DNA	Deoxyribonucleic acid
dsRBDs	Double-stranded DNA-binding motifs
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GLD-2	Germ Line Development 2
FBF	<i>Caenorhabditis elegans</i> Pumilio/fem-3 binding factor
hnRNPs	Heterogeneous nuclear ribonucleoproteins
KH domain	K homology domain
mRNA	Messenger RNA
NABP	Nucleic acid binding protein domain
NMD	Nonsense-Mediated Decay
N-terminal	Amino-terminal
PAP	Poly(A) polymerase
PCR	Polymerase chain reaction
PIN	PilT N-terminus
PPR	Pentatricopeptide repeat
psi	Pounds per square inch
PTC	Premature termination codon
PTS1	Peroxisomal targeting sequence 1
Puf	Pumilio and FBF
PUM-HD	Pumilio homology domain
qPCR	Quantitative reverse transcription PCR
RBD	RNA-binding domain
RBP	RNA-binding protein
RFP	Red fluorescent protein
RNA	Ribonucleic Acid
RRM	RNA Recognition Motif
rRNA	Ribosomal RNA
SMG	Nonsense mediated mRNA decay factor
snRNA	small nuclear RNA

ssRNA	Single-Stranded RNA
TAE	Tris-acetate-EDTA
TFIIIA	Transcriptional Factor IIIA
TPR	Tetratricopeptide Repeat
TRM	Tripartite recognition motif
tRNA	Transfer RNA
w/v	weight by volume
ZnF	Zinc Finger
NLS	Nuclear localization sequence
NoLS	Nucleolar localization sequence

Chapter One: Introduction

RNA-binding proteins (RBPs) have an essential role in the regulation of gene expression. They function in pre-mRNA splicing, capping, and polyadenylation, and have important roles in mRNA decay, stability, localization and translation (Glisovic *et al.*, 2008; Yin *et al.*, 2013). The interaction between RNA-binding proteins and their RNA targets is mediated by single or multiple RNA-binding domains within the protein. Several types of RNA-binding domains have been identified, including the RNA Recognition Motif (RRM), K Homology (KH) domain, pentatricopeptide (PPR) domain and the Pumilio Homology domain (PUM-HD or Puf domain) (Wang *et al.*, 2002; Grishin, 2000; Oubridge *et al.*, 1994; Wickens *et al.*, 2002). PPR and Puf domains are unique, in that their binding to RNA is modular. In these two types of RNA-binding domain, individual polypeptide repeats within the domain can bind with specificity to a single nucleotide. This one-repeat/one-nucleotide binding characteristic has opened up opportunities for researchers to engineer the binding specificity of these proteins so that they can target specific RNA sequences in the cell. The amino acid binding codes of Puf proteins are better understood than those of PPR containing proteins, and this has allowed researchers to modify Puf repeats to target specific RNA sequences in the cell (Cheng *et al.*, 2016; Filipovska *et al.*, 2011). By generating artificial Puf domains to target specific RNA sequences, and appending functional effector domains to these engineered Pufs, it is possible to alter the physiology of target RNAs within the cell.

The *Arabidopsis* genome encodes 26 putative Puf family proteins (Tam *et al.*, 2010; Francischini and Quaggio, 2009). Although not well characterized, Pufs are known to have critical roles in post-transcriptional control of gene expression in *Arabidopsis*. An *Arabidopsis*

Puf protein, APUM23, was studied in this thesis research because its novel Puf repeat number and arrangement may provide a structural backbone that could be used to construct an engineered RNA-binding protein that can specifically target cellular RNAs.

1.1. Post-transcriptional regulation

Gene expression in eukaryotic cells is regulated at many levels. Transcriptional regulation has historically been the focus of many studies. However, post-transcriptional control is known to also have a major role in gene regulation (Hamid and Akgul, 2014). Post-transcriptional regulation can include pre-mRNA processing, RNA transport, decay, stability and translation (Figure 1.1) (Cheng *et al.*, 2016). These RNA regulatory processes have an important role in plants throughout development or in response to environmental stress (Gallie, 1993).

Each post-transcriptional regulation process is critical for preparing the transcripts to be translated. Capping (5 methyl-guanosine) allows an mRNA to be recognized by the eukaryotic translation initiation factor eIF4A, and also increases mRNA stability (Hamid and Akgul, 2014). Intron splicing is catalyzed by the spliceosome, and can contribute to the formation of multiple polypeptide variants through alternative splicing events (Jurica and Moore, 2003; Chen *et al.*, 2013). Polyadenylation is mediated by poly(A) polymerase, an event that is important in 3' end stability and in facilitating translation initiation through interactions with poly(A)-binding protein (Sanfacon *et al.*, 1990). RNA editing can also occur as a post-transcriptional regulation mechanism in mitochondria and chloroplasts, where editing is a necessary event to produce the proper open reading frame on the transcript (Szempruch *et al.*, 2015). Regulatory events that are required for mRNA export from the nucleus, and trafficking to regions of the cytosol also require the activity of RNA-binding proteins (Pfaff *et al.*, 2018). Removal of an mRNA from the cytosol

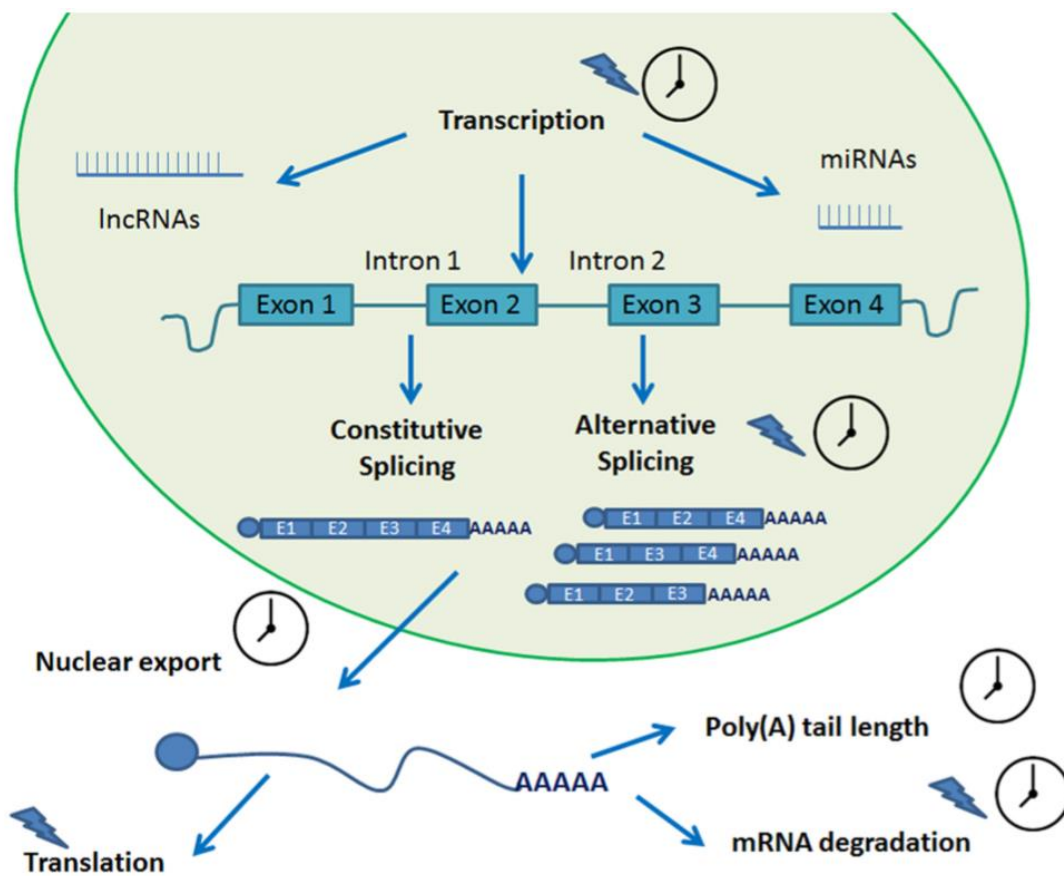


Figure 1.1: Post-transcriptional gene regulation in a cell. After transcription in the nucleus, mRNAs undergo modifications that control their fate, including 5' capping, splicing, polyadenylation, and nuclear export. Once in the cytosol, mature mRNAs can be further modified by changes in the poly(A) tail, or they can be turned over. mRNAs are also regulated by movement to different locations in the cell prior to translation (Romanowski and Yanovsky, 2015).

involves complex interactions with RNA-binding proteins. This process typically begins with the removal of the poly(A) tail followed by the degradation of mRNA. This degradation can be performed by different pathways, including 3' - 5' pathways, 5' - 3' pathways, and nonsense-mediated decay (NMD)(Belostotsky and Sieburth, 2009). In plants, if an mRNA possesses a long 3'UTR or a premature termination codon close to an exon-exon junction, NMD will be activated and the aberrant mRNA degraded by NMD core proteins (Matlin *et al.*, 2005). RNA binding proteins play important roles in each of these post-transcriptional control processes by binding directly to RNA. Their binding specificity and functional activity is usually mediated by interacting proteins (Schoning *et al.*, 2008). Therefore, understanding the mechanisms that allow RNA-binding proteins to interact with and regulate their target RNAs will allow for a better understanding of post-transcriptional gene regulation.

1.2. RNA-binding proteins

RNA-binding proteins can influence mRNA dynamics by directly binding to elements within mRNAs via their RNA-binding domains (RBDs) (Schoning *et al.*, 2008). RNA-binding proteins can interact with the coding region or 5'- or 3'-untranslated regions (UTR) of an mRNA (Hentze *et al.*, 2018). They bind to and regulate coding (mRNAs) and noncoding RNAs, such as transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), microRNA, and long-noncoding RNA (Wei and Wang, 2015). By regulating different stages of mRNA metabolism in post-transcriptional processes, RNA-binding proteins modulate many important biological activities by up-regulating or down-regulating gene expression. For example, CHLAMY1, interacts with mRNAs that contain a 7 to 16 unit UG nucleotide repeat in their 3'UTRs (Zhao *et al.*, 2004). Through the interaction with these mRNAs, CHLAMY1

regulates the translation of circadian proteins and functions in regulating the circadian process in plants (Romanowski and Yanovsky, 2015). RNA-binding proteins also have critical roles in the response of stressors. For instance, by destabilizing the mRNA secondary structure to facilitate translation, the RNA-binding protein CspA helps cells respond to low temperature (Liu *et al.*, 2013).

1.2.1. RNA-binding domains

RBPs recognize RNA targets through their RNA-binding domains. Depending on the type of RNA-binding domain, RNA-binding proteins can either bind to single-stranded RNA (ssRNA) or double stranded RNA (dsRNA), and recognize specific RNA nucleotide sequences and/or structured RNAs (Hogan *et al.*, 2008; Messias and Sattler, 2004). RNA-binding proteins can contain a single RNA-binding domain or multiple domains of different types. Several types of RNA-binding domain have been well characterized, although recent RNA-interactome studies have demonstrated that many other, poorly understood types of RNA-binding domains exist (Castello *et al.*, 2017).

1.2.1.1. RNA Recognition Motif

RRM proteins are the largest group of RBPs that recognize ssRNA (Schoning *et al.*, 2008). For example, about 500 human proteins were found to contain RRM domains (Chen and Varani, 2013). Furthermore, 196 proteins containing RRM domains have been identified in *Arabidopsis* (Zhao *et al.*, 2004). The RRM has tandem repeats of a short motif, containing four stranded antiparallel β -sheet and two α -helices, with a size of 80 – 90 amino acids (Chen and Varani, 2013). There is a large variation in the structures of different RRM proteins. Only two regions in the RRM,

octapeptide RNP1 and hexapeptide RNP2, are widely conserved and are critical for RNA interactions (Schoning *et al.*, 2008). When binding to an RNA, the Arg or Lys residue on RNP1 forms a salt bridge with the RNA phosphodiester backbone, and the two conserved residues on RNP1 and RNP2 form stacking interactions with the RNA nucleotides (Chen and Varani, 2013). The RRM has an important role in post-transcriptional regulation. For example, SR proteins, which contain one or two N-terminal RRM domains that can bind to exons, assist with assembly of the spliceosome on pre-mRNAs. Heterogeneous nuclear ribonucleoproteins (hnRNPs), another family proteins that contains RRM, is involved in repression of splicing (Matlin *et al.*, 2005).

Although RRM proteins are encoded widely by eukaryotic genomes, no RRM binding codes have been defined, and the binding specificity of this domain has not been strategically altered in the lab. This is due to the complexity of the RRM-RNA interaction (Wei and Wang, 2015). The binding affinity of an RRM is affected by many factors, including variability in its N- and C- terminal sequence and the presence of insertions in the arrangement of the $\beta\alpha\beta\alpha\beta$ region. The secondary structure of an RNA often affects the affinity of RRM binding (Chen and Varani, 2013). Therefore, the binding specificity cannot be determined by RRM sequence only, which makes it difficult to predict a set of binding codes for RRM protein.

1.2.1.2. K Homology domain

The KH RNA-binding domain is present in many RNA-binding proteins, and is often involved in transcriptional and translational regulation. It was first identified in hnRNPs that are involved in repression of splicing. Three KH motifs are found on hnRNPs, indicating the KH domains are critical for RNA-binding function in these proteins (Slomi *et al.*, 1994). A KH

domain is composed of a three-stranded β -sheet and three α -helices. The conserved RNP1 sequence located on the central β -sheet surface mediates the interaction with RNA. The residues can be arranged in two different versions, type I and type II, depending on how the three strands of the β -sheet are arranged. KH domains interact with RNA through a 'GXXG' loop which connect the two α -helices (Chen and Varani, 2013). The KH domain only recognizes four nucleotides in an RNA sequence and binds with low binding affinity (Wei and Wang, 2015). RNA recognition only depends on hydrogen bonding and shape complementarity (Chen and Varani, 2013). Thus, KH domains often bind RNA in cooperation with other KH domains, which is why multiple KH domains are often found in one single protein (Wei and Wang, 2015). Because of its binding mechanism, binding codes for KH domain are difficult to predict.

1.2.1.3. Pentatricopeptide repeat (PPR)

Pentatricopeptide repeat family proteins were first characterized in *Arabidopsis* and are abundant in plastids and mitochondria (Shikanai and Fujii, 2013). The PPR domain is a 35-amino acid motif, and one PPR protein often contains 2-30 motifs. The 35 amino acids in each motif form a hairpin of α -helices that is composed of four helical turns and a five-residue loop, and recognize one nucleotide in its RNA target (Yin *et al.*, 2013). PPR proteins are involved in several post-transcription process, such as RNA editing, mRNA maturation, and translation (Wei and Wang, 2015). These result in post-transcriptional regulation in many biological processes, including photosynthesis, respiration, chloroplast development, embryogenesis, cytoplasmic male sterility, isoprenoid biosynthesis and retrograde signaling (Shikanai and Fujii, 2013).

The PPR family proteins are divided into two subfamilies, which are P and PLS subfamily. The P subfamily is composed of 35-amino acid repeats (Yin *et al.*, 2013). It is

involved in many RNA maturation process in a cell, including RNA stabilization, splicing, intergenic cleavage, and translation. The PLS subfamily, which functions in RNA editing, contains repeats of 31-36 amino acids and extra domains at C terminal, which include three parts: long PPR-like motifs (L), short PPR-like motifs (S), and canonical PPR motifs (P) (Shikanai and Fujii, 2013; Yin *et al.*, 2013). A PPR motif forms a pair of antiparallel α -helices that recognizes specific RNA sequences (Shikanai and Fujii, 2013). Therefore, by coordinating PPR repeats, these proteins can be engineered to target RNA (Coquille *et al.*, 2014).

1.2.1.4. Zinc-finger (ZnF) domains

Abundant ZnF family proteins are found in eukaryotes. In humans, about 2% of proteins encoded by human genome are ZnF-containing proteins. Its functional domain, the zinc-finger domain, was first found in transcriptional factor IIIA (TFIIIA), which is the first transcriptional factor to be described (Klug, 2010). Although ZnF domains often interact with DNA sequences, there are also ZnF proteins that interact with RNA (Little *et al.*, 1999). Therefore, ZnF domains are considered as a good scaffold for constructing specific RNA-binding proteins because they are stable and modular (Blancafort *et al.*, 2004). A ZnF domain contains about 30 amino acids, including two cysteines and two histidines through which ZnF binds one or more Zn^{2+} in a structure (Matthews and Sunde, 2002). The ZnF domain contains a conserved position which contains a β -hairpin and an α -helix that are stabilized by Zn^{2+} coordination (Wei and Wang, 2015). ZnF domains can be categorized into different classes, including CCHH zinc fingers, CCCH zinc fingers, CCHC zinc knuckles and RanBP2-type ZFs, different classes of ZnF domains differ both in structure and functions (Chen and Varani, 2013; Matthews and Sunde, 2002). ZnF domains with different specific targets have been studied for medical applications,

including regulation of specific drug targets and disease genes. Vandevenne *et al.* (2014) engineered the ZnF domain to bind specific target sequences, which indicates ZF domain can be used as a scaffold for RNA-binding domain engineering.

1.2.1.5. Double-stranded RNA binding motifs (dsRBDs)

The dsRBD contains 65-70 amino acids, forming a $\alpha\beta\beta\beta\alpha$ fold and interacting with dsRNAs. It is involved in many post-transcriptional regulations, including RNA editing, miRNA biogenesis and function, and RNA export and localization. How the dsRBDs interact with their target RNAs remains unclear. However, previous research shows that many dsRBDs recognize the shape of A-form helix of dsRNAs (Chen and Varani, 2013). They directly interact with the 2'-hydroxyl groups of the ribose sugars, and interact with the non-bridging oxygen residues of the phosphodiester backbone directly or by water-mediated contacts, both using hydrogen bonding (Thomas and Beal, 2017). In addition to these parts in the RNAs, dsRBDs can also interact with the minor groove of helical or non-helical elements flanking the RNA duplex (Chen and Varani, 2013). Since the knowledge of the dsRBP-RNA interactions is still limited, the practical application of this dsRBP-RNA is also limited.

1.2.1.6. The Puf family RNA-binding proteins

Puf (Pumilio) family proteins are a group of RNA-binding proteins with a conserved structure that can bind to specific RNA sequence in a modular fashion. These RNA-binding proteins are named based on the first two members of Puf proteins, which are *Drosophila melanogaster* Pumilio (DmPUM) and *Caenorhabditis elegans* Pumilio/fem-3 binding factor (FBF) (Wang *et al.*, 2018; Lehmann and Nüsslein-Volhard, 1987; Zhang *et al.*, 1997). These

functions are performed by the Pumilio Homology Domain (PUM-HD), an RNA-binding domain that is composed of tandem Puf repeats that are conserved across eukaryotes. Puf domains fold into a crescent shape with the Puf repeats. Each repeat contains 35-39 amino acids, and recognizes one RNA nucleotide (Edwards *et al.*, 2001; Wang *et al.*, 2001). Most Puf proteins contain eight Puf repeats in their PUM-HD, and form a solenoid structure in a crescent shape (Figure 1.2)(Gupta *et al.*, 2009). The amino acids of a Puf repeat interact with a single nucleotide in the inner concave surface, and bind to an RNA in an anti-parallel manner (Filipovska *et al.*, 2011). Each Puf repeat is composed of three α -helices (two long and one short, Figure 1.2). A five amino acid motif located in the second α -helix determines the RNA-binding specificity. Specifically, the first, second and fifth amino acids are the sites that determine the RNA-binding specificity, which is called tripartite recognition motif (TRM). The amino acids at position 1 and 5 in a TRM can interact with the edge of the base through hydrogen bonds and Van der Waals forces, while the amino acid at position 2 can interact with two adjacent bases through base stacking interactions. The classical Puf domain has a conserved target sequence (UGUR) at the 5' end, and a common 8-nt target sequence for the PUM-HD is UGUAUAUA (Lu *et al.*, 2009).

Puf proteins play an important role in post-transcriptional regulation. They typically interact with the 3' UTR and regulate mRNA decay and translational repression (Miller and Olivas, 2011). In *Drosophila*, *Caenorhabditis elegans* and humans, Puf proteins are involved in germline stem cell identity maintenance (Zhang *et al.*, 1997; Wickens *et al.*, 2002; Forbes and Lehmann, 1998). In yeast, Puf proteins function in mitochondrial motility and biogenesis by interacting with the nuclear-encoded mRNAs that encode mitochondrial proteins (Abbasi *et al.*, 2011). Puf proteins also play critical roles in translation initiation, ribosomal RNA (rRNA) processing, and ribosome biogenesis, (Cheng *et al.*, 2016).

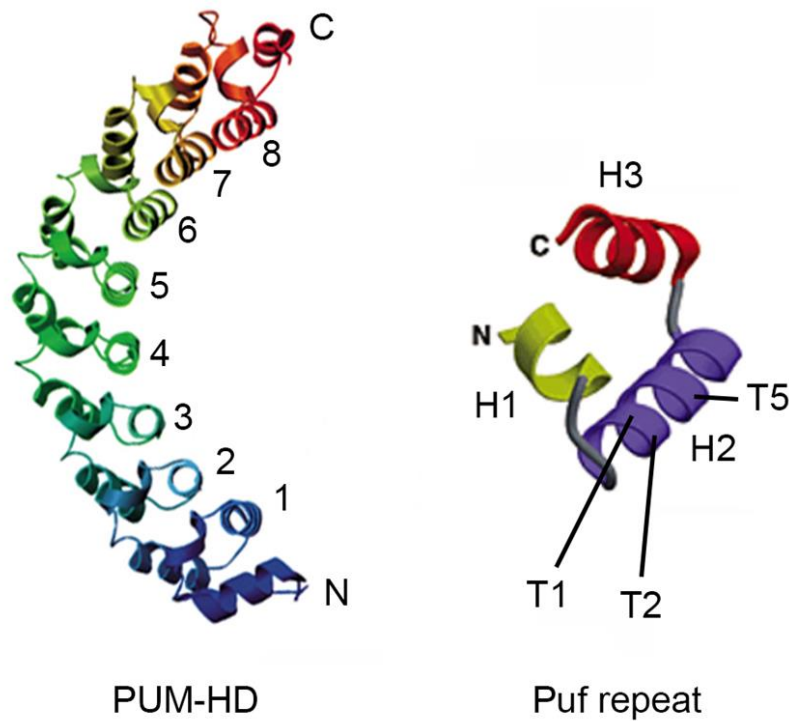


Figure 1.2: Crystal structure of PUM homology domain. The structure on the left is the crystal structure of a common *Pumilio* homology domain (PUM-HD). The PUM-HD typically contains eight Puf repeats. The structure on the right is the structure of a single Puf repeat which is composed of three α -helices. Each repeat is composed of three α -helices (H1, H2, and H3), the second of which usually binds to a single nucleotide in an RNA sequence. T1, T2 and T3 represent three amino acids in a five amino acid motif on the second α -helix. This region is named the tripartite recognition motif (TRM). TRM sequences are known to determine the nucleotide specificity of the Puf repeat (Gavis, 2001).

1.3. Plant Puf RNA-binding domain

Puf family proteins exist broadly in plants. Up to 26 and 19 members of Puf proteins have been characterized in Arabidopsis and rice, respectively (Figure 1.3) (Tam *et al.*, 2010). Like the Puf proteins in non-plant organisms, most Puf proteins in Arabidopsis and rice contain eight Puf repeats. Some of the Puf proteins contain different number of Puf repeats, and possess less or more predicted repeats (Figure 1.3). Puf proteins are involved in many stress responses and developmental processes in Arabidopsis, including defence responses, abiotic stress, seed dormancy, leaf development (Wang *et al.*, 2018). Most of the Arabidopsis Puf proteins (APUM) localize to punctate structures in the cytosol, some of which move between the nucleus and the cytoplasm (Tam *et al.*, 2010). In addition to the APUM proteins that localize in cytosol, APUM23 and APUM24, are localized in nucleolus (Tam *et al.*, 2010). Both of APUM23 and APUM24 contain a long polypeptide extension at the carboxyl-terminal (C-terminal) end, which contain putative lysine and arginine-rich nucleolar localization signal (Tam *et al.*, 2010).

1.4 APUM23

APUM23 is an Arabidopsis RNA-binding protein that plays a critical role in many biological functions, including leaf development, plant organ polarity, pre-rRNA processing, and regulation of ribosomal gene expression (Bao *et al.*, 2017). For example, APUM23 can regulate the expression of many genes that are involved in ABA and abiotic stress response through ABA signaling pathway (Cheng *et al.*, 2016). In addition, the 11-nt target sequence of wild type APUM23 is found in 18S rRNA sequence, supporting APUM23's function in 18S rRNA maturation, including rRNA biogenesis and ribosome assembly (Abbasi *et al.*, 2010; Zhang and Muench, 2015). APUM23 is unique, as it was the

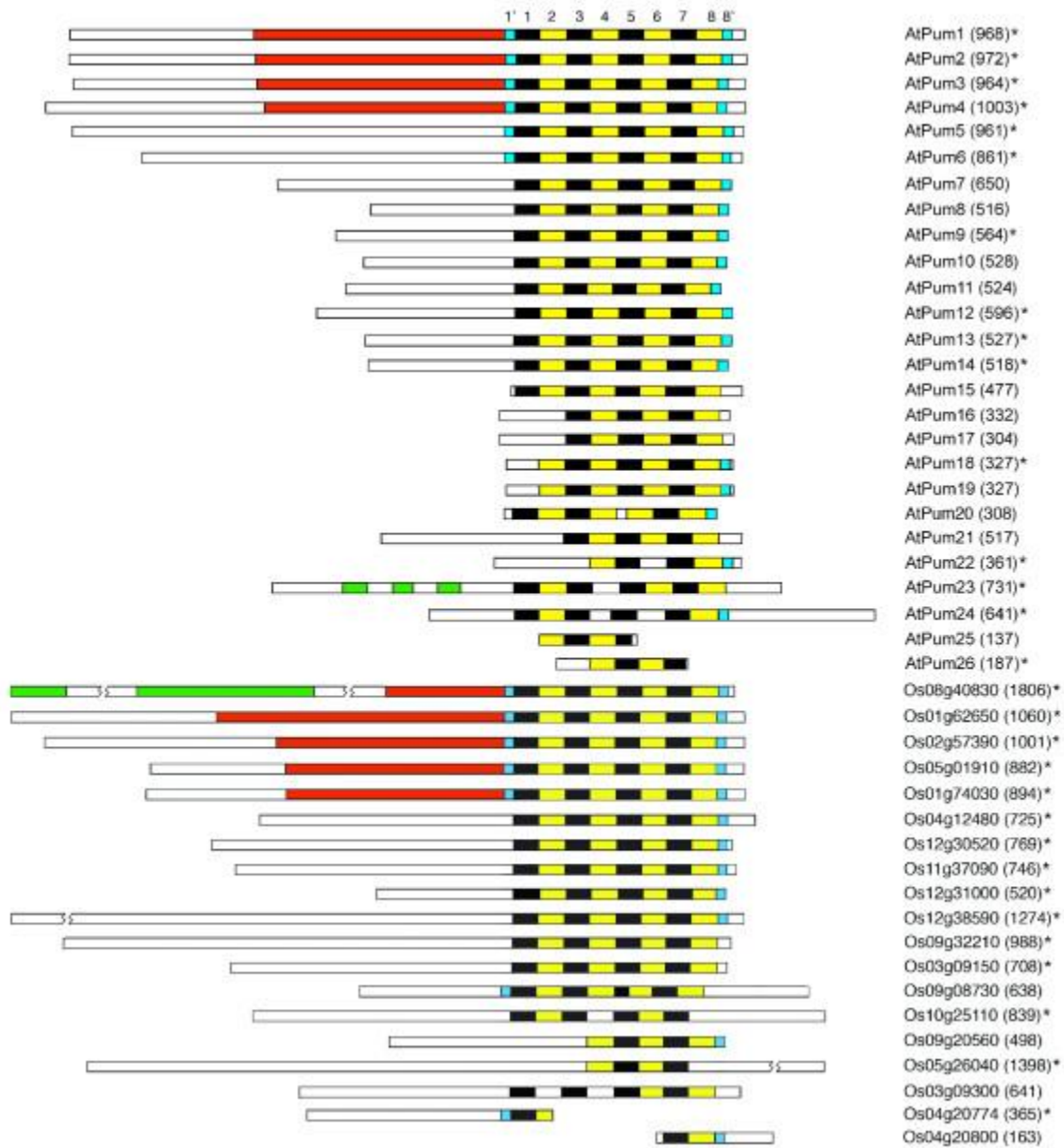


Figure 1.3: The primary structure of Puf proteins in Arabidopsis and rice. The Puf repeats are shown in black and yellow. The 1' and 8' pseudorepeats are shown in blue. In AtPUM23, additional Puf repeats are labelled in green, while the green regions in Os08g40830 represent domains of unknown functions. The red regions are the conserved nucleic acid binding protein domain (NABP) that are found in some plant Puf proteins. Sizes of predicted and known open reading frames are shown in brackets to the right of each protein. (Tam *et al.*, 2010).

first Puf protein to show natural cytosine binding by one of its TRMs (Figure 1.4) (Zhang and Muench, 2015).

APUM23 is an unusual Puf protein because it contains 11 Puf repeats in its PUM homology domain and its core target sequence is not UGUR like the classical Puf protein, but rather an 11-nucleotide sequence that lacks the UGUR core (GGAAUUGACGG) (Zhang and Muench, 2015; Bao *et al.*, 2017). A recent crystal structure of APUM23 with its consensus RNAs sequence was determined recently (Fig 1.5) (Bao *et al.*, 2017). This 11-nucleotide target sequence is bound to the Puf domain by hydrogen bonds and stacking interactions (Figure 1.5). Repeat 7 (R7) of the 11 Puf repeats is not involved in the interaction with the nucleotides, and the adenine that was thought to interact with this repeat is not recognized by the TRM on R7. Instead, it flips away from the inner concave where the nucleotides are bound to Puf repeats. (Figure 1.5C).

Unlike the classical Puf proteins that have their Puf repeats clustered in the C-terminal region, the Puf repeats of APUM23 are present throughout the centre region of the protein (Zhang and Muench, 2015). Since APUM23 interacts with a specific RNA sequence, it could be used as a scaffold for protein engineering in order to recognize specific RNA sequences. The 11-nt specific RNA recognition of APUM23 also makes it a good backbone because it may possess greater specificity compared to classical Puf proteins whose target sequence is eight nucleotides. In addition, because a cytosine binding TRM exists in APUM23, it is useful to study the structure of the cytosine binding TRM for Puf protein engineering.

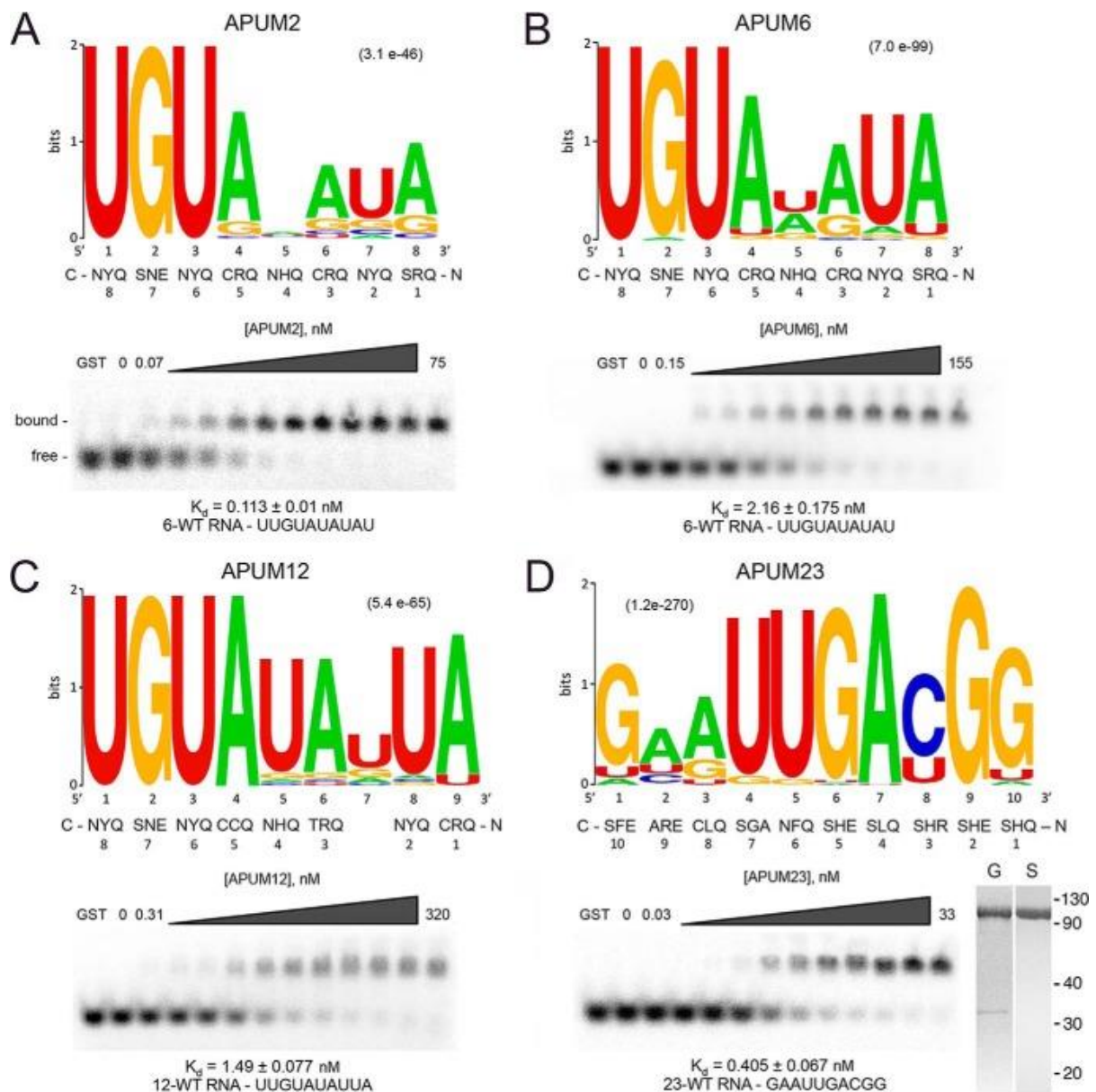


Figure 1.4 Consensus sequences for different APUM proteins. The figure shows the SELEX logo graphs of APUM2 (A), APUM6 (B), APUM12 (C), and APUM23 (D). APUM2 and APUM6 contain the same eight nucleotide consensus sequence as canonical APUM proteins. APUM12 contains a nine nucleotide consensus sequence, and the nucleotide at position 7 is likely extruded from the binding surface. APUM23 contains a unique 10-nucleotide target consensus sequence and natural cytosine binding residue (Zhang and Muench, 2015).

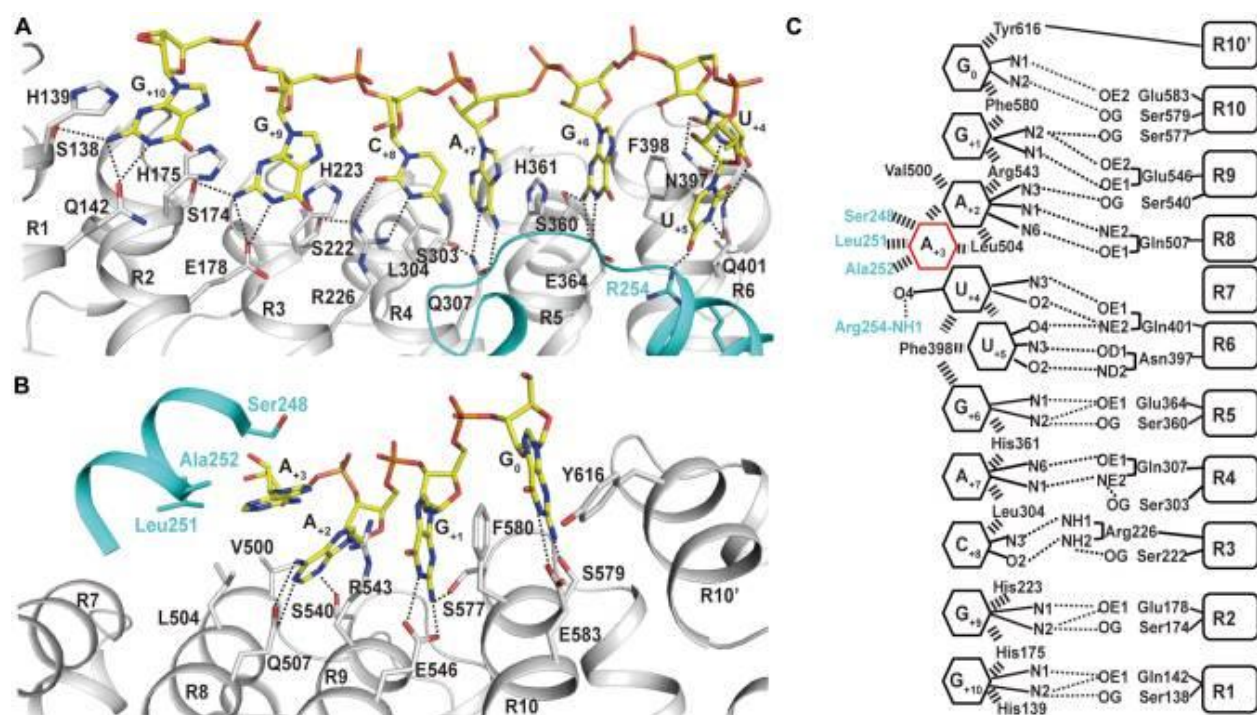


Figure 1.5: Structure of the APUM23-RNA interaction. (A, B) Crystal structure of the interaction between target RNA and the 11 Puf repeats of APUM23. (C) The interactions between APUM23 and target RNA. Hydrogen bonds are shown as dotted lines. The stacking interactions are shown as bars (Bao *et al.*, 2017).

1.5. Puf protein engineering

1.5.1. RNA-binding protein engineering

Since RNA-protein interactions are critical in post-transcriptional regulation, a custom RNA-binding protein that can bind to a desired target sequence would be useful in the application of many gene regulated processes (Nelles *et al.*, 2015). Because the life of an RNA is short, alteration of RNA targets does not cause a permanent change in expression at the genome level (Wei and Wang, 2015). Many attempts have been made to create artificial RBPs by combining a sequence specific RNA-binding domain with functional RNA effector domains, with a peptide linker between the two domains (Blancafort *et al.*, 2004). Since Puf and PPR RNA-binding domains determine RNA-binding specificity by the amino acids located in their tandem RNA-binding repeats, the key amino acids in these repeats can be engineered to alter the specificity of these RNA-binding domains. (Wei and Wang, 2015).

Puf or PPR domains are good candidates for engineering RNA-binding proteins because of their solenoid structure and one repeat:one nucleotide binding characteristics (Chen and Varani, 2013; Gully *et al.*, 2015). Therefore, the target sequences of Puf and PPR proteins are predictable, and their RNA binding repeats can be altered to bind different RNA sequences specifically. In PPR repeats, the 5th amino acid determines its specific nucleotide target (Yin *et al.*, 2013). In Puf proteins, three amino acids in the second α -helix determine the specific RNA targets of Puf proteins (Filipovska *et al.*, 2011). Therefore, the RNA binding specificity of Puf and PPR domains can be altered for each target nucleotide by modifying the amino acids that determine the binding specificity, which makes them good scaffolds to make engineered target-specific functional domains (Wei and Wang, 2015).

On the other hand, there are RNA-binding domains whose binding specificities are not well understood, including RRM, KH domains, ZnF domains and multi-subunit RNA-binding proteins. In RRM containing proteins, different RRM domains can participate in various functions, including protein-protein interactions and RNA-protein interactions. An RRM can also affect the affinity of RNA binding by another RRM in the same protein, which makes binding codes difficult to decipher (Kobayashi *et al.*, 2013). ZnF domains recognize RNAs by both the RNA structure and specific RNA sequences, and they were revealed to have two RNA recognition modes (Wei and Wang, 2015; Blancafort *et al.*, 2004). Therefore, simply changing the amino acids in the ZnF domain would not be able to alter the binding specificity as we expected. KH domains can function with either a single copy of KH domain or work in a cooperative way with multiple KH domains due to their low affinity (Wei and Wang, 2015). Therefore, the binding codes of KH domain are difficult to identify. Thus, for these RNA-binding domains we cannot alter the specificity of the binding domain by simply changing the amino acids in the RNA-binding domain (Blancafort *et al.*, 2004).

1.5.2. Engineering Puf RNA-binding proteins

As discussed above, Puf proteins contain solenoid RNA-binding domains and each Puf repeat recognizes a single RNA base. Through their RNA-binding domains, Puf RNA-binding domain can recognize specific single stranded RNA by its primary sequence. Furthermore, three amino acids (TRMs) in the second α -helix determines the binding specificity of the Puf domain (Filipovska *et al.*, 2011). Therefore, the specificity of Puf domains can be altered by modifying their TRMs. Canonical Puf protein contain eight Puf repeats that typically bind to an eight nucleotide sequence (Filipovska *et al.*, 2011). Adenine, guanine and uracil are the predominant

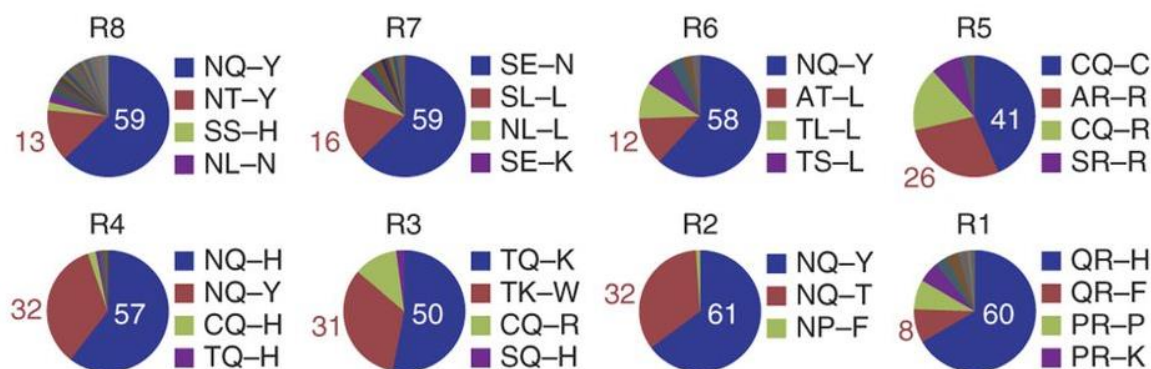
nucleotides that are recognized, each by several sets of TRMs in different Puf proteins. Figure 1.6A shows a summary of the amino acid combinations of the TRMs of the eight Puf repeats found in nature (Campbell *et al.*, 2014). These researchers analyzed different amino acid combinations, and summarized the TRMs that recognized adenine, guanine, cytosine and uracil. Figure 1.6B shows the amino acid combinations of TRM-7 and their target nucleotides. Additionally, a natural cytosine binding Puf repeat has been identified in APUM23 (Zhang and Muench, 2015), and the TRM (SHR) of this repeat matches that of a synthetic cytosine TRM identified in a yeast three-hybrid screen for cytosine binding TRMs (Filipovska *et al.*, 2011; Dong *et al.*, 2011). TRM code identity has allowed for prediction of binding specificity and design in engineered Puf domains (Campbell *et al.*, 2014).

Although canonical Puf protein bind to an 8-nt RNA sequence, by inserting additional Puf repeats into the canonical PUM-HD backbone, researchers have had some success in generating Puf RNA-binding domains that interact with longer RNA sequences. For example, by inserting extra Puf repeats close to the end of the repeat 5, the extended Puf proteins can recognize a 16 nucleotide RNA sequence, which greatly improves their specificity and improves the regulation of the Puf proteins (Filipovska *et al.*, 2011). Furthermore, Puf proteins were found to be able to invade complex RNA structures and bind to their target sequences, which means Pufs can be used to target double-stranded RNAs (Filipovska *et al.*, 2011).

1.5.3. Artificial site specific RNA endonucleases and the PIN-domain

There are no naturally occurring endonucleases that can cleave RNA in a sequence specific manner. The known endonucleases either recognize their targets by their structure (e.g., RNase H or RNase III) or cleave RNAs with single nucleotide specificity (e.g., RNase T1 or

A



B

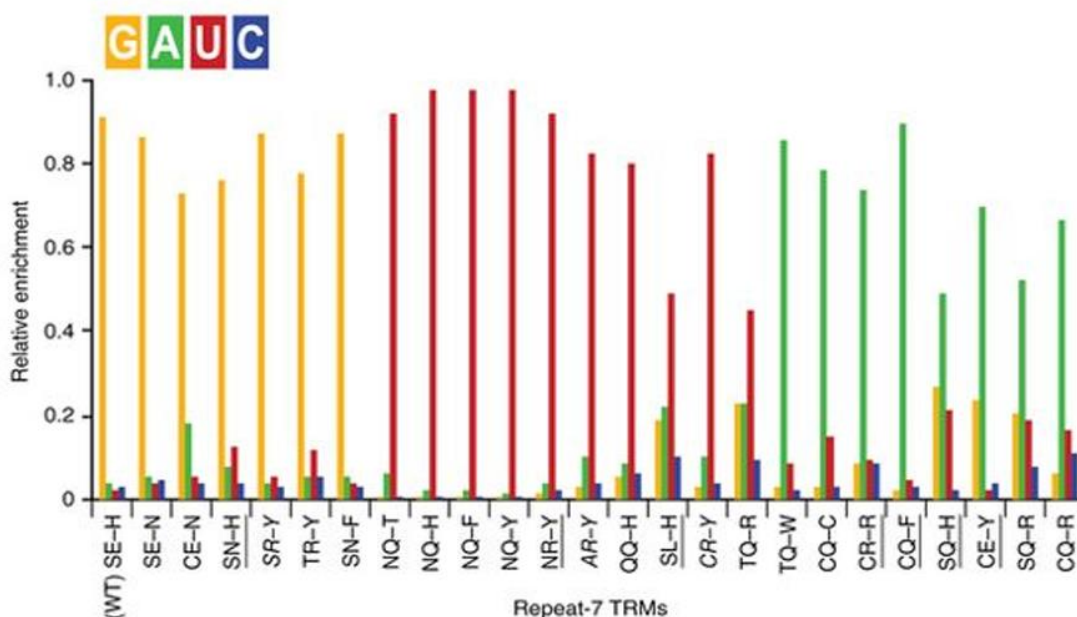


Figure 1.6: Amino acid combinations in TRMs of common Puf proteins in nature and their target nucleotides. A) Different amino acid combinations in each TRM of the eight Puf repeat positions in canonical Puf proteins and their enrichment in nature. Each pie chart represents one Puf repeat, and is summarized based on 94 Puf proteins. B) Target nucleotides of different TRM amino acid combinations at Repeat 7 (position +2 of the bound nucleotides). Yellow, green, red and blue bars in the graph represent guanine, adenine, uracil and cytosine, respectively (Campbell *et al.*, 2014).

RNase A) (Choudhury *et al.*, 2012). An artificial site specific endonuclease (ASRE) was created by fusing a recognition domain to a catalytic domain (Choudhury *et al.*, 2012). More specifically, an ASRE was formed by fusing a Puf protein that recognized a specific 8-nt RNA sequence to a PilT N-terminus (PIN) domain. The PIN domain is a non-specific endonuclease domain that is found in all three domains of life, and present in the SMG5 and SMG6 nonsense-mediated mRNA decay (NMD) pathway proteins (Glavan *et al.*, 2006). The PIN-domain is composed of about 180 amino acids residues (Takeshita *et al.*, 2007). The ASRE uses the PIN-domain from human SMG6, as it contains a well-defined structure and only requires a divalent metal cation for RNA cleavage (Choudhury *et al.*, 2012; Zhang *et al.*, 2014). When PIN was fused to the amino-terminus of the PUM-HD, the PUM-HD bound to its target sequence and PIN folded back and cleaved the RNA at the site upstream the binding site (Fig 1.7) (Choudhury *et al.*, 2012).

1.5.4. Using the poly(A) polymerase, GLD-2, to increase mRNA stability and translation rate

The poly(A) tail plays an important role in mRNA stability and translation regulation (Park *et al.*, 2016). An intact poly(A) tail is required for nuclear export and translation. When the poly(A) tail of an mRNA is shortened or removed, it results in translational repression and decay of the mRNA (Nakel *et al.*, 2015). Therefore, a poly(A) polymerase that sequentially adds one poly(A) tail nucleotide is critical for post-transcriptional regulation of an mRNA (Kwak *et al.*, 2004). GLD-2 protein is a poly(A) polymerase identified in *Caenorhabditis elegans*, and localizes in P-granules in the embryo (Kwak *et al.*, 2004). GLD-2 protein contains low activity by itself. However, when it binds to GLD-3 protein, GLD-2 activity increases (Nakel *et al.*, 2015). Although the KH domains of GLD3 do not seem to provide RNA-binding ability, the activation

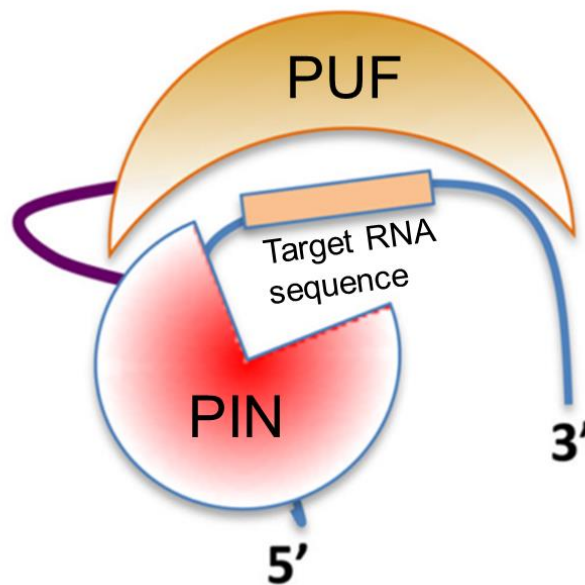


Figure 1.7: The ASRE binds and cleaves the target RNA. When the Puf domain of the ASRE recognizes the specific target sequence on an RNA, it binds to the target sequence in an anti-parallel fashion. Then, the PIN-domain fused downstream folds back and cleaves the target RNA. The cutting site is at the third or fourth base downstream of the binding site. (Choudhury *et al.*, 2012).

of GLD-2 poly(A) polymerase by GLD-3 might be due to enhancing its specific RNA-binding property (Nakel *et al.*, 2015). Kwok et al (2004) demonstrated that fusion of GLD-2 to the sequence specific RNA-binding protein, MS2, increased the length of the poly(A) tail on its target mRNA and enhanced its translation in *Xenopus* oocytes. More recently, a PUM-HD was fused to GLD-2 in a similar manner and resulted in poly(A) tail extension and increased translation (Cooke *et al.*, 2011).

1.6 Hypothesis and objectives of thesis

APUM23 contains 11 Puf repeats, which potentially provides a higher nucleotide specificity in its target RNAs compared to canonical Puf proteins that possess eight Puf repeats. Therefore, using a similar engineering strategy to that described in the previous section, APUM23 can be tested as a site-specific regulator of mRNA.

Hypothesis: Appending the RNA-binding domain of APUM23 to an effector domain (an RNase and a poly(A) polymerase) will result in the specific degradation or translational enhancement of a target mRNA.

The specific objectives of this thesis project were to:

1. Determine whether a truncated form of APUM23, with its nuclear/nucleolar localization signal removed, localizes to the cytoplasm.
2. Demonstrate whether truncated APUM23 fused to the PIN RNase can cleave its specific target mRNA *in vivo*.

3. Demonstrate whether truncated APUM23 fused to a poly(A) polymerase can enhance the expression of its specific target mRNA *in vivo*.

Chapter Two: Materials and Methods

2.1 PCR and molecular cloning

The constructs used were for *in vivo* expression of fluorescent protein fusions in onion (*Allium cepa*) epidermal cells from bulb scales. Phusion High Fidelity DNA polymerase (ThermoFisher) was for all PCR reactions. The template of PCR amplification was from various sources, including APUM23 coding sequence, APUM2 coding sequence and purchased target constructs. The primers for PCR amplification were synthesized by Integrated DNA Technologies (IDT). All PCR reactions were performed in thermal cycler (C1000 Touch, Bio-Rad). The DNA products amplified by PCR reactions were electrophoresed on 1% (encoding APUM proteins targets) or 2% (encoding APUM RNA targets) agarose gels in 1x TAE buffer, and then eluted using a gel extraction kit (QIAquick Gel Extraction Kit, QIAGEN). The eluted DNAs were then digested by restriction enzymes, and ligated with compatible digested vectors using a ratio of 1:3 to 1:5 (vector: insert) using T4 DNA ligase (Invitrogen). TOP 10 *E. coli* competent cells were transformed via heat-shock, and were grown on LB media (1% w/v NaCl, 1% w/v Tryptone, 0.5% w/v yeast extract, and 1.5% w/v agar) containing ampicillin (1 mg/L). Plasmid DNA was purified using DNA preparation kits (Qiagen QIAprep Spin Miniprep Kit for DNA templates and Qiagen Plasmid Midi Kit for DNA used for *in vivo* expression).

2.1.1 Cloning of APUM protein encoding constructs to observe their localization

Three DNA constructs were produced to demonstrate the subcellular localization of APUM proteins translationally fused to RFP. These were wild type APUM23-RFP, truncated APUM23-RFP (nuclear/nucleolar targeting signal removed), and APUM2 homology domain (HD)-RFP (Figure 2.1). Wild type APUM23 and truncated APUM23 were PCR amplified from

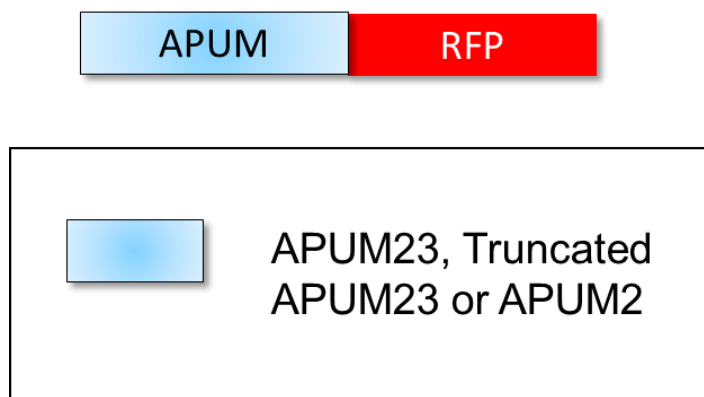


Figure 2.1: Diagram of APUM-RFP fusion structure to determine the localization of APUM proteins. Wild type APUM23, truncated APUM23 and APUM2 homology domain constructs were fused to RFP sequence.

the APUM23 template. The truncated APUM23 contained the amino acids 85 to 657 of the full length APUM23 and encoded all Puf RNA-binding repeats but lacked the NLS. The APUM2-HD was cloned from the APUM2 coding region template. The primer sequences (GenScript) used for the PCR reactions are listed in Table 2.1. The PCR products were inserted between the KpnI/NcoI sites of PRTL2 Δ NS/MCS-RFP for observation by epifluorescence microscopy.

2.1.2 Cloning of the APUM-PIN encoding constructs for *in vivo* expression

The APUM proteins used for the APUM-PIN mRNA targeting assays contained a 3X Flag epitope tag sequence at their amino-terminus and the PIN non-specific, RNA-binding sequence at the carboxyl-terminus. The cloning strategy for APUM23-PIN, APUM2-PIN is shown in Figure 2.2. A seven amino acid (7aa) linker (VDTGNGS) was used as a flexible sequence to couple the APUM and PIN domains, for a shorter linker (3aa) showed very low cleaving activity, while a longer linker (12aa) increased non-specific cleaving (Choudhury *et al.*, 2012) The PIN domain was derived from nucleotide sequence 1238 to 1421 of human SMG6 with a stop codon added downstream the PIN sequence (Glavan *et al.*, 2006).

Three cloning steps were used to build each of these constructs. The primers and synthetic DNA sequences used for PCR cloning are listed in Table 2.2. First, the DNA sequence encoding the linker-PIN region was chemically synthesized (Genscript) and cloned into the pQE30 (Qiagen) vector. Next, the APUM coding sequences were amplified by PCR, and the products were cloned into the BamHI and SalI sites of pQE30 upstream and in-frame of the Linker-PIN sequence. The APUM23-Linker-PIN and APUM2-Linker-PIN sequences were used as templates for the next PCR amplification step. Synthetic ds-DNA (gBlock Gene Fragments, IDT) were used for cloning to insert the 3X Flag tag (with start codon) upstream the APUM-Linker-PIN sequence. The G-block ds-DNA contained sequence that overlapped with the 5' end

Table. 2.1: PCR primers used for cloning to determine the localization of truncated APUM23.

Construct	Sequence
Wild type APUM23	5' primer GGAATTGGTACCATGGTTTCTGTTGGTTCTAAATCATTGC
	3' primer GAAATTCCATGGCAATTCTCATTTTATTTGAATGCCG
Truncated APUM23	5' primer GGAATTGGTACCATGAGGAAAGAGATTGATCCAGAGAC
	3' primer GAAATTCCATGGCCTTTGGAAAATTTGATTTGTTG
APUM2 homology domain	5' primer GGAATTGGTACCATGTTTGGATCTTCAATGCTTGAAG
	3' primer GAAATTCCATGGCCAAAGCCATCCTCCTCTCTCCAG

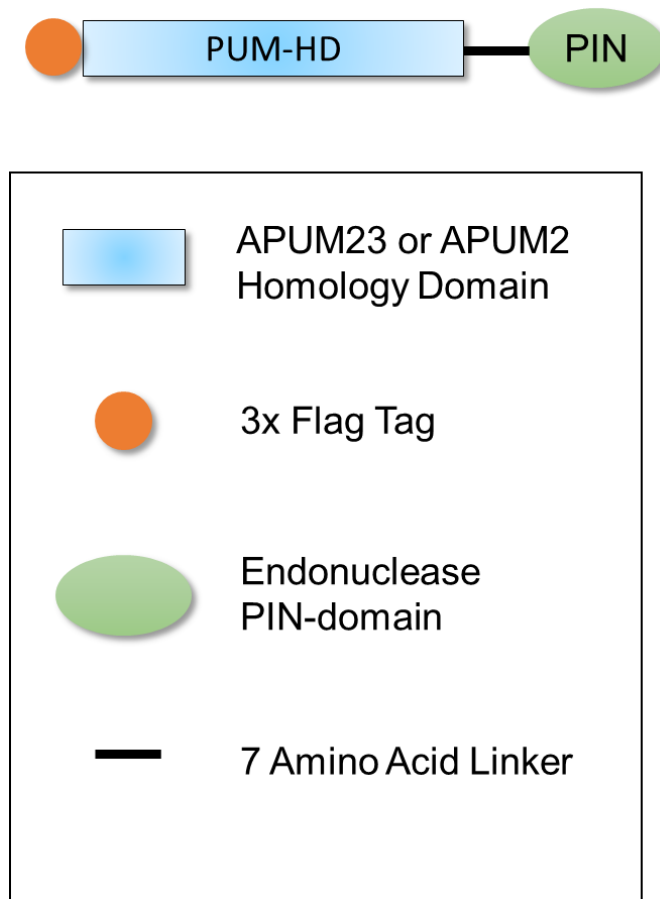


Figure 2.2: Structural arrangement of the APUM-PIN constructs. The 3X Flag tag was fused to the amino terminus of truncated APUM23 and the APUM2-HD. The PIN RNase was fused to the carboxyl-terminus of the APUM proteins through the linker peptide.

Table. 2.2: Primers/sequences used for cloning 3X Flag Tag-APUM-Linker-PIN.

Amplified sequence	Sequence
Truncated APUM23	5' primer GGAATTGGATCCAGGAAAGAGATTGATCCAGAGAC
	3' primer GAAATTGTCGACCTTTGGAAAATTTGATTTGTTG
APUM2-HD cloned into PQE 30 vector with Linker-PIN domain	5' primer TTCCGGGGATCCTTTGGATCTTCAATGCTTGAAG
	3' primer GAAATTGTCGACCAAAGCCATCCTCCTCTCTCCAG
Truncated APUM23 G-block Flag primer	GCGGGAATTCCATGGCTGATTACAAAGATCATGATGG TGACTATAAGGACCACGACATCGATTACAAAGATGAT GATGATAAAGGAGGTGGAGGTGGAGGTGGAAGGAAA GAGATTGATCCAGAGAC
APUM2HD ds-DNA G-block Flag sequence	GCGGGAATTCCATGGCTGATTACAAAGATCATGATGG TGACTATAAGGACCACGACATCGATTACAAAGATGAT GATGATAAAGGAGGTGGAGGTGGAGGTGGATTTGGA TCTTCAATGCTTGAAG
Primers for amplification of 3X Flag-APUM-Linker-PIN	5' Flag PUM primer with NcoI site GCGGGAATTCCATGGCTGATTACA
	3' PIN sequence primer with NotI site GAAATTGCGGCCGCTTAGCAGCAGGACAGGATGAGAT C

of either truncated APUM23 or APUM2-HD. These were annealed with PIN sequence primer using the APUM-Linker-PIN plasmid as template. Then the APUM-Linker-PIN with the 3X Flag tag was used as template for amplification. The final PCR products were inserted in PRTL2 Δ NS/CFP vector between the NcoI and NotI sites, dropping out the sequence coding for CFP. These constructs were later used for the expression in onion epidermal cells.

2.1.3. Cloning of constructs encoding the RNA targets of APUM-PIN

The mRNA target of APUM-PIN encoded the red fluorescent protein (RFP) that contained a carboxyl-terminal, three amino acid peroxisomal targeting sequence (PTS1). This localized the protein product of this target mRNA to the peroxisome. The 3' untranslated region (UTR) of the RFP transcript contained the specific RNA binding site of either APUM23 or APUM2. The structural arrangement of the APUM target mRNA is shown in Figure 2.3. A PCR product containing the PTS1, stop codon and APUM RNA target sequence was amplified and cloned into the BglII and XbaI sites of PRTL2 Δ NS-RFP-MCS (Chuong *et al.*, 2005). The PCR primers are listed in Table 2.3.

2.1.4 PAP-APUM with Flag tag cloning

Since the GLD-2 poly(A) polymerase (PAP) needs a RNA binding specificity provided in order to activate its poly(A) polymerase activity, the Puf proteins provided the binding specificity (Nakel *et al.*, 2015). GLD-2 was used as a functional domain fused to truncated APUM23 or APUM2-HD to attempt to enhance the translation of their target mRNAs. The construct design is shown in Figure 2.4. GLD-2 was fused upstream or downstream the Puf protein, therefore, both fusion types were made to test which works more optimally for specific binding of the mRNA target and for the synthesis of the poly(A) tail. The constructs were

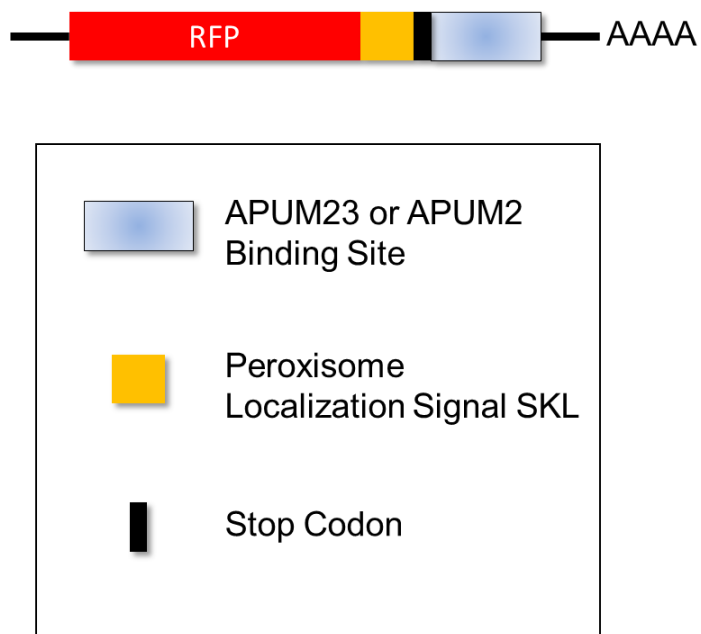


Figure 2.3: Structural arrangement of the APUM target mRNA. These constructs encoded RFP that contained a carboxyl-terminal peroxisomal targeting sequence (PTS1) and the specific RNA target of APUM23 and APUM2 in the 3' UTR.

Table. 2.3: Primers for cloning APUM targets.

PCR products	Sequence
APUM23 target	Sense GGAGGGAGAGATCTTCCAAGCTCTGATTCCGGTCGGGAATTG ACGGAGACCAAACGTCTAGAAGAGACGG
	Antisense CCGTCTCTTCTAGACGTTTGGTCTCCGTCAATTCCCGACCGGA ATCAGAGCTTGGAAGATCTCTCCCTCC
APUM2 target	Sense GGAGGGAGAGATCTTCCAAGCTCTGATTCCGGTCGGTGTATA TAAGACCAAACGTCTAGAAGAGACGG
	Antisense CCGTCTCTTCTAGACGTTTGGTCTTATATACACCGACCGGAAT CAGAGCTTGGAAGATCTCTCCCTCC

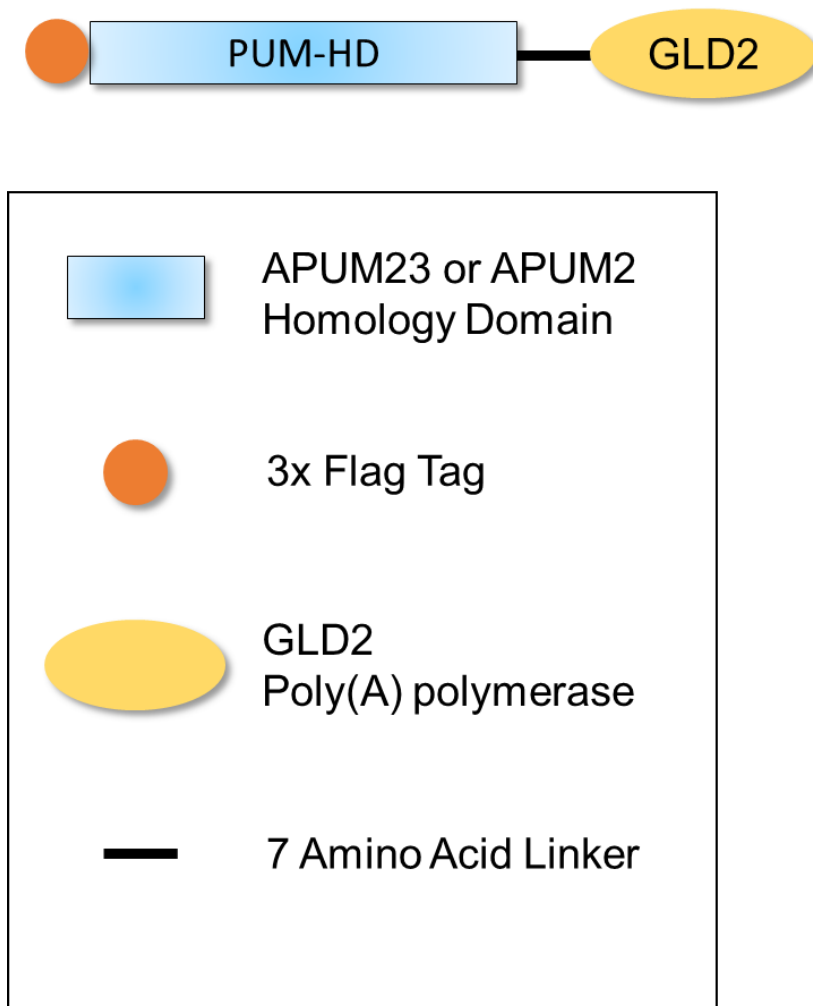


Figure 2.4: Structural arrangement of the Puf-GLD-2 constructs. GLD-2 was fused to APUM23 or APUM2 binding sites at either C- or N- terminus of the Puf sequence. The APUM23 or APUM2 binding sites were also fused with 3x Flag tag.

inserted in PRLT2ΔNS/CFP between NcoI and NotI sites, dropping out the CFP sequence. For this PAP effector domain expression system, the APUM23 and APUM2 triple-target consensus sequence of APUM23 or APUM2 was inserted into the 3'UTR region a CFP reporter vector (NcoI/NotI site of PRTL2dNS-CFP).

2.2 Onion epidermal cell transfection using biolistic particle bombardment

Onions (*Allium cepa*) were purchased from local grocer and stored in dark at 4°C. Onion scale segments were cut from the second and third layers of the onion bulb. The size of each segment was approximately 4 cm². The scale segments were then placed on chromatography paper soaked with ddH₂O in a petri dish with the epidermal surface facing up. Plasmid DNA used for bombardment was purified using a plasmid purification kit (Qiagen Midi Kit). Plasmid DNA was coated onto 1 μm gold particles by vigorously mixing (for 2 min) 17 μl of 2.5 M CaCl₂, 6 μL of 0.1 M spermidine, and plasmid DNA were added sequentially into 17 μl of 30 mg/ml gold particles suspended in 50% sterile glycerol. The RFP target plasmid amount used to coat gold particles was 1 μg or 31μg, or was mixed in ratios of RFP target plasmid:APUM-PIN plasmid at 1 μg:3 μg or 0.5 μg:1 μg. The DNA/gold particle mixture was then centrifuged for 3 sec after settling for 1 min. The pellet was washed with 70 μl of 70% ethanol followed by a 95% ethanol wash. The pellet was then re-suspended in 24 μl of 95% ethanol. The re-suspended gold particles were pipetted on to a microcarrier and air-dried. Particle bombardment (PDS-1000/He particle delivery system, Bio-Rad) used an 1100 psi rupture disk. After bombardment, the scale segments were inverted, and the petri dishes were sealed with parafilm in order to keep the environment moist. The scale segments were incubated overnight and imaged the following day.

2.3 Microscopic analysis

The epidermal layer was peeled from the bombarded onion scales and mounted on a glass slide with a drop of distilled water. After pipetting a few drops of distilled water onto the epidermal cells, a cover slip was gently applied to the peel. Peels were observed using an epifluorescence microscope (Leica DMR) with a Plan Fluotar 20x and a Plan Fluotar 40x objective lens and an RFP filter set. Images were captured by a CCD camera (Retica 1350 EX), and processed and analyzed using Velocity software (Perkin Elmer version 6.0).

2.4 Western blotting

Bombarded onion epidermal peels (18 peels of 4 cm²) were snap-frozen in liquid nitrogen and weighed. The frozen peels were crushed and ground in 2x crack buffer (10% w/v SDS, 50% v/v glycerol, 100 mM Tris pH 6.8, fresh 200 mM DTT, and fresh 0.05% 2-mercaptoethanol (BME)) in an ice-cold mortar. The denatured protein samples were heated in boiling water for 5 min followed by a 2 min centrifuge spin at 12,000 rpm. The protein samples were then electrophoresed through an SDS-PAGE gel (12% resolving gel) using 1X SDS PAGE running buffer (25 mM Tris base, 192 mM glycine and 0.1% w/v SDS). The proteins were transferred from the SDS gel to a nitrocellulose membrane (0.2 µm, Bio-Rad) in 1X Tris-glycine buffer (25 mM Tris base, 192 mM glycine, and 20% v/v methanol) at 60V constant voltage for 1 hr. The membrane was stained with Ponceau stain (0.5% w/v Ponceau S and 1% v/v acetic acid) for 5 min to observe protein loading levels. The membrane was blocked in TBST (10 mM Tris pH 7.5, 150 mM NaCl, and 0.1% v/v Tween 20 (Sigma)) with 3% (w/v) skim milk powder for 30 min. The primary antibodies, anti-RFP (polyclonal, Rockland) or anti-GFP (monoclonal, Thermo Fisher Scientific), were added at 1:1000 dilution and incubated for two hours on a shaker at room

temperature, followed with five washes (5 min each) in TBST. The secondary antibody (anti-rabbit and anti-mouse) was then added at a 1:3000 dilution and incubated for one hour on a shaker at room temperature, following by five washes (5 min each) in TBST. The chemiluminescent substrate (Thermo Scientific, Super SignalTM West Pico PLUS Chemiluminescent Substrate) was added to the membrane and incubated for 5 min. The membrane was imaged using an Amersham Imager 600 (GE Healthcare Life Science).

Densitometry was performed to quantify the band intensities obtained from the western blot analysis. Image J software (version 1.51J8) was used to measure band intensities. A rectangular box was drawn that surrounded all bands in each lane across the blot. Intensity profiles were produced, and a baseline was drawn across the profile to delimit the area under the peak for each band. Intensity values were measured and used in determining expression levels of RFP.

2.5 Quantitative PCR (qPCR) analysis

Total RNA from bombarded onion epidermal cells was extracted to quantify the level of target RFP mRNA. The onion scale peels were ground in liquid nitrogen and total RNA was isolated using a total RNA extraction kit (Spectrum, Sigma). A DNase step in purification procedure removed any contaminating DNA from the sample. The quality and concentration of the total RNA was determined by agarose gel electrophoresis and spectrophotometry (Nanodrop). M-MLV reverse transcriptase (Invitrogen) was used to synthesize cDNA from the mRNA template. All cDNA synthesis reactions were performed using a thermal cycler (C1000 Touch, Bio-Rad) and contained RNase inhibitor (Invitrogen). Two primers RFP mRNA primer sets were used to detect the RFP-target mRNAs. Actin and GAPDH primers sets were used for

internal control reactions. Primer sets are listed in Table 2.4. Power SYBR® Green PCR Master Mix (Life Technologies) was used in the qPCR reactions. Each treatment was replicated three times. The data was collected and analysed with Prism GraphPad Software (Version 8).

2.6 Poly(A) tail length assay

Total RNA was extracted from 18 onion epidermal peels that were bombarded with the APUM-PAP and CFP constructs. A poly(A) tail-length assay kit (USB Poly(A) Tail-Length Assay Kit, Affymetrix) was used to estimate the RFP mRNA poly(A) tail length. All reactions, including G/I tailing, reverse transcription and amplification, were performed using a thermal cycler (C1000 Touch, Bio-Rad). The amplified products were electrophoresed on 1% agarose gel in 1X TAE buffer to detect the poly(A) tail length.

Table. 2.4: Primers for qPCR reactions.

Primer pairs	Sequence
RFP-1	5' primer GCGAGTTCATCTACAAGGT
	3' primer CTCGTACTGTTCCACGATG
Actin	5' primer ATGCACCAAGAGCAGTATTC
	3' primer CATGCTCAATGGGGTACTT
GAPDH	5' primer CTTGTTGCTGTTAACGATCC
	3' primer TACAGCTACTGGTTTGTAC

Chapter Three: Altering the expression patterns of target mRNAs *in vivo* by fusing RNase and poly(A) polymerase domains to APUM proteins

3.1 Introduction

APUM23 is a 731 amino acid protein, and is one of two APUM proteins that localizes to the nucleolus. As predicted by amino acid sequence analysis and RNA consensus sequence binding, APUM23 contains ten putative Puf repeats that bind to the RNA consensus sequence GAAUUGACGG (Tam *et al.*, 2010; Zhang and Muench, 2015). These Puf repeats are not as tightly spaced as observed in typical Puf proteins, but rather span the polypeptide from amino acids 85 to 657. More recent crystal structure analysis showed that APUM23 contains 11 Puf repeats and prefers binding to the sequence GGAAUUGACGG (Bao *et al.*, 2017), one nucleotide longer than the previously determined consensus (Zhang and Muench, 2015). APUM23 contains a predicted nuclear localization sequence (NLS) at amino acid position 688 in the carboxyl-terminal polypeptide extension of this protein (Figure 1.3)(Tam *et al.*, 2010). Some modified NLSs with specific amino acid sequences can also serve as nucleolar localization sequences (NuLS) (Garcia-Bustos *et al.*, 1991). Therefore, the predicted NLS sequence in the carboxyl-terminal region of APUM23 may also function as a NuLS that localizes APUM23 to nucleoli. The novel consensus RNA binding sequence and crystal structure of APUM23 makes it an excellent candidate for fusion to functional domains that can target specific RNA sequences and alter their physiology (Zhang and Muench, 2015; Bao *et al.*, 2017).

In this chapter, experiments are described that were aimed at utilizing APUM23 as a polypeptide backbone to target specific mRNAs in the cell for degradation or enhanced translation. The first experiments demonstrated that a truncated APUM23 that had its predicted NLS/NuLS sequence removed would localize to the cytosol rather than its usual nucleolar localization. Next, the truncated APUM23 (and APUM2 control) was fused either to an RNase

(PIN endonuclease) or poly(A) polymerase (PAP) to determine whether these protein fusions could alter the stability and translation rate of an RFP target mRNA in living onion epidermal cells. The identification of an 11-nt RNA target sequence of APUM23 suggests that this protein possesses higher sequence specificity than the typical Puf proteins that recognize 8-nt RNA sequences. Previous research in the Muench laboratory showed that altering the Tripartite Recognition Motifs (TRMs) in selected Puf repeats of APUM23 could alter the specificity of APUM23 (Zhang and Muench, 2015). Therefore, APUM23 has the potential to be an excellent scaffold to build an engineered APUM23 protein fused to effector domains to target cellular mRNAs.

3.2 Results

3.2.1 Subcellular localization of wild type and truncated APUM23

As a prerequisite to using the APUM23 backbone to bind mRNAs in living cells, the NLS/NuLS must be removed so that the protein fusions are localized to the cytosol rather than to the nucleolus. A truncated version of APUM23 (tAPUM23) was designed to contain all of the predicted Puf repeats (amino acids 85 to 657 that covers all the Puf RNA-binding site in APUM23) but excluded the predicted NLS/NuLS sequences. The PUM-HD of APUM2 was used as a control APUM protein. APUM2 is a typical 8 Puf repeat-containing protein that binds with high affinity to the typical Puf RNA target sequence (UGUAUAUA), and does not contain a predicted NLS (Tam *et al.*, 2010).

The coding sequences of wild type APUM23, tAPUM23 (no predicted NLS/NoLS), and APUM2 were fused to a sequence that encoded the amino-terminal end of red fluorescent protein (RFP) in a plant expression plasmid. The APUM-RFP fusions were expressed in onion

epidermal cells by particle bombardment. The onion epidermal peel expression system is ideal for the transient expression, as these cells are large, relatively transparent, and contain a single layer of cells (Scott *et al.*, 1999). Onion bulbs can be purchased from the local grocer and allows for rapid, transient gene expression of fluorescently tagged proteins. Epifluorescence microscopy of onion epidermal cells produces only subtle autofluorescence using standard excitation wavelengths, and produces excellent image series due to the dynamic nature of these living cells.

Epifluorescence microscopy demonstrated that APUM2-RFP localized to the cytosol (Figure 1A), as was expected by the absence of a NLS in the primary amino acid sequence of this polypeptide. The truncated version of APUM23 (amino acids 85 to 657) fusion protein that lacked the NLS (tAPUM23-RFP) was also localized to the cytosol (Figure 3.1B). In contrast, wild type APUM23 fusion protein (APUM23-RFP) targeted to the nucleolus (Figure 3.1C, D), as was observed previously (Tam *et al.*, 2010). These results demonstrate that the truncated version of APUM23 that had the predicted carboxyl-terminal NLS/NuLS deleted, localized to the cytosol. Therefore, this truncated version of APUM23 (amino acids 85 to 657) was used in subsequent experiments (described below) that targeted cytosol mRNAs for degradation or translational enhancement.

3.2.2 Effect of a recombinant tAPUM–RNase fusion protein on regulating the degradation of RFP target mRNAs in onion epidermal cells.

The truncated APUM23 was fused to a non-specific RNA endonuclease in an effort to determine whether this protein fusion could specifically bind and degrade in onion epidermal cells an artificial mRNA that contained the corresponding APUM RNA binding sequence. The PIN domain of SMG6 was used as the non-specific RNase domain (Choudhury *et al.*, 2012) that was fused to the carboxyl-terminus of tAPUM23 and APUM2. The target mRNA used here encoded an RFP that possessed a

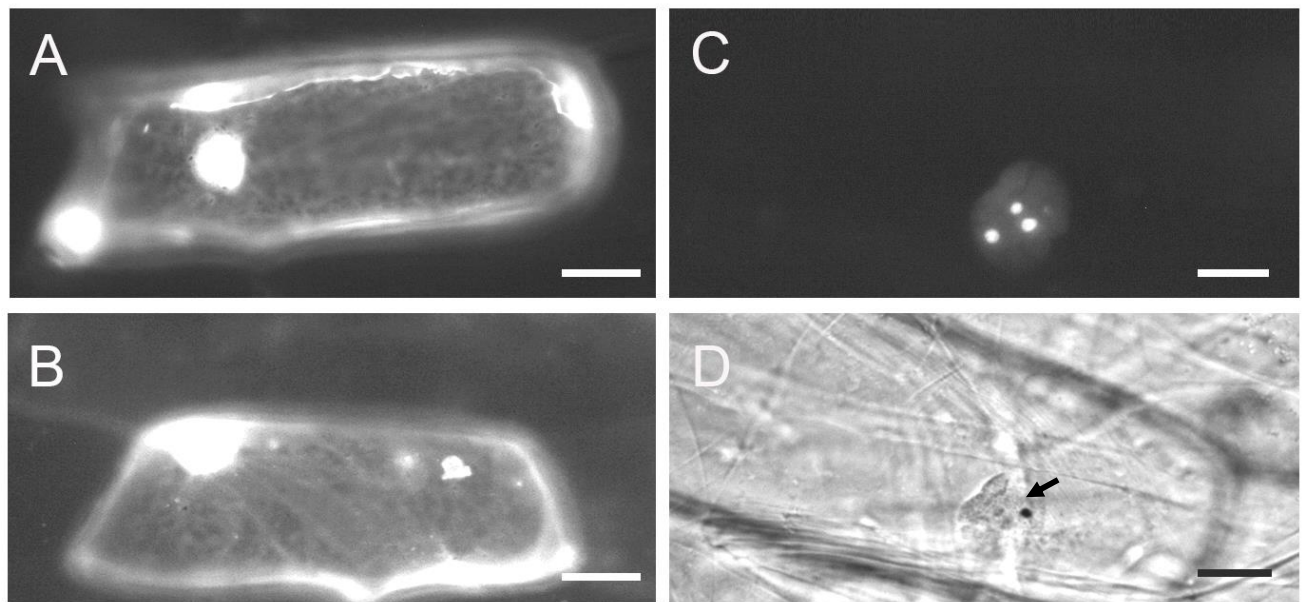


Figure 3.1: Localization of APUM protein-RFP fusion in onion epidermal cells by epifluorescence microscopy. A) The fluorescence localization APUM2 homology domain-RFP. B) The fluorescence localization of truncated tAPUM23-RFP. C) The fluorescence localization of wild type APUM23-RFP in the nucleus and prominent nucleoli D) Bright field image of the onion cell shown in C). The arrow points the nucleus, which contains a gold particle. A, B) Bar = 26 μm . C, D) Bar = 13 μm .

carboxyl- terminal, three amino acid (serine-lysine-leucine) peroxisomal targeting sequence (PTS1)(Wolins and Donaldson, 1997) that will direct the RFP to peroxisomes. This provided a more accurate level of RFP quantification when using epifluorescence microscopy compared to the more diffuse cytosolic localization of RFP. Within the 3' UTR of the mRNA that encoded the RFP reporter protein was either the APUM23 or APUM2 RNA consensus binding site (GGAAUUGACGG or UGUUAUAUA sequence, respectively). Hence, when an APUM-PIN protein bound to the 3' UTR in the corresponding RFP target mRNA, the PIN domain would have access to the mRNA and would cleave it. A previous study demonstrated that when a classical eight Puf repeat-containing Puf protein was fused at its carboxyl terminus to the PIN domain, this fusion protein showed specific cleavage of the RNA target *in vivo* (Choudhury *et al.*, 2012). In that study, the PIN domain folded back and cleaved the fourth nucleotide downstream the RNA binding site of Puf domain. A description of this target RNA degradation system is outlined in Figure 3.2.

3.2.2.1 – Epifluorescence microscopy of RFP expression

To first analyze the targeting and expression level of the RFP constructs alone or when expressed in combination with the APUM23-PIN or APUM2-PIN fusion constructs, transfected onion epidermal cells were imaged using epifluorescence microscopy. When both versions of the RFP mRNA (with either the APUM23 or APUM2 target sequence located in the 3' UTR of this reporter mRNA) were expressed individually, there was often a high level of RFP fluorescence observed in the peroxisomes of onion epidermal cells (Figure 3.3A, B). This expression level, however, varied from cell to cell, with some cells showing low peroxisome fluorescence. Next, the RFP mRNA constructs were co-expressed with the tAPUM23-PIN or APUM2-PIN

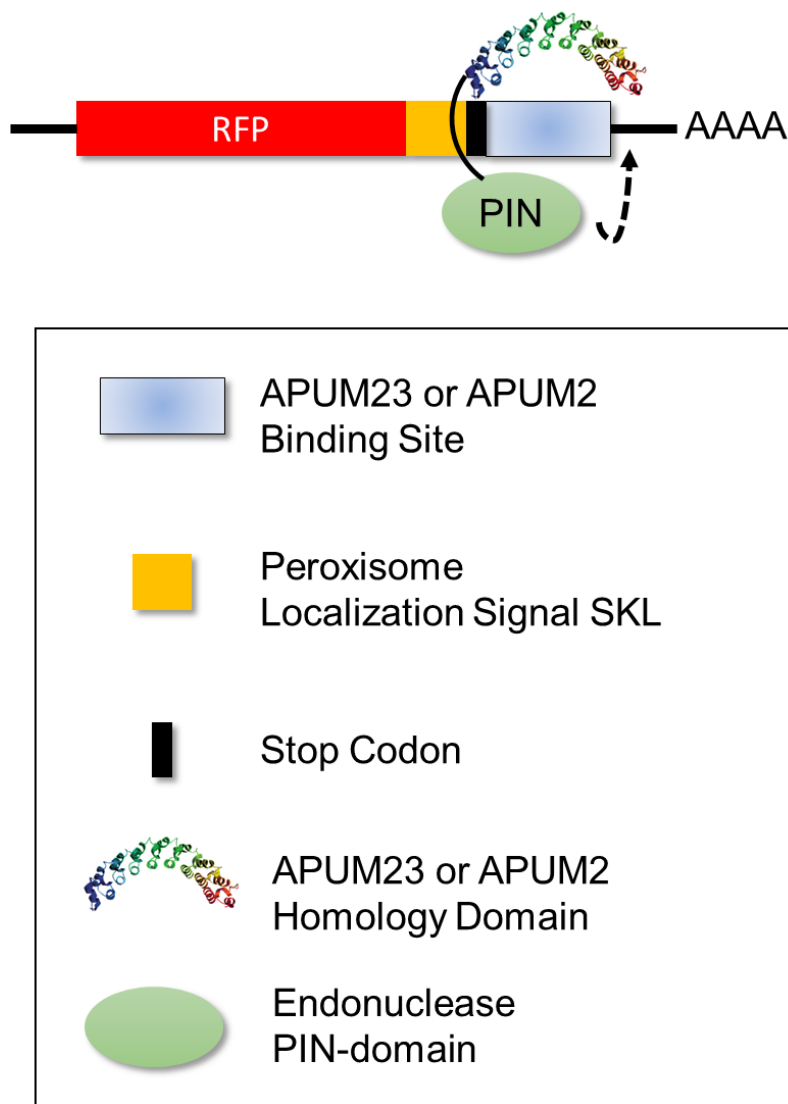


Figure 3.2: Strategy for using an APUM-PIN fusion protein to cleave specific mRNAs. tAPUM23 or the APUM2-HD were fused to the RNA endonuclease domain (PIN) to generate a fusion protein that can specifically bind and cleave the RFP mRNA target that contains the APUM23 or APUM2 binding sequence in its 3' UTR. Appended to the APUM-PIN protein was a 3X Flag Tag epitope. The RFP protein product possessed a PTS1 that targeted the protein to the peroxisome to enable fluorescence imaging expression analysis. The RFP mRNA constructs were expressed in the absence or presence of tAPUM23-PIN or APUM2-PIN in onion epidermal cells. If this system is functional *in vivo*, the target sequence in the RFP mRNA 3' UTR will be recognized by the APUM-PIN fusion protein in an anti-parallel fashion. The PIN domain will then fold back and cleave the phosphodiester backbone in the mRNA downstream of the binding site and render the RFP mRNA target unstable in the cytosol, resulting in turnover of RFP mRNA by the endogenous RNA degradation machinery in the cell.

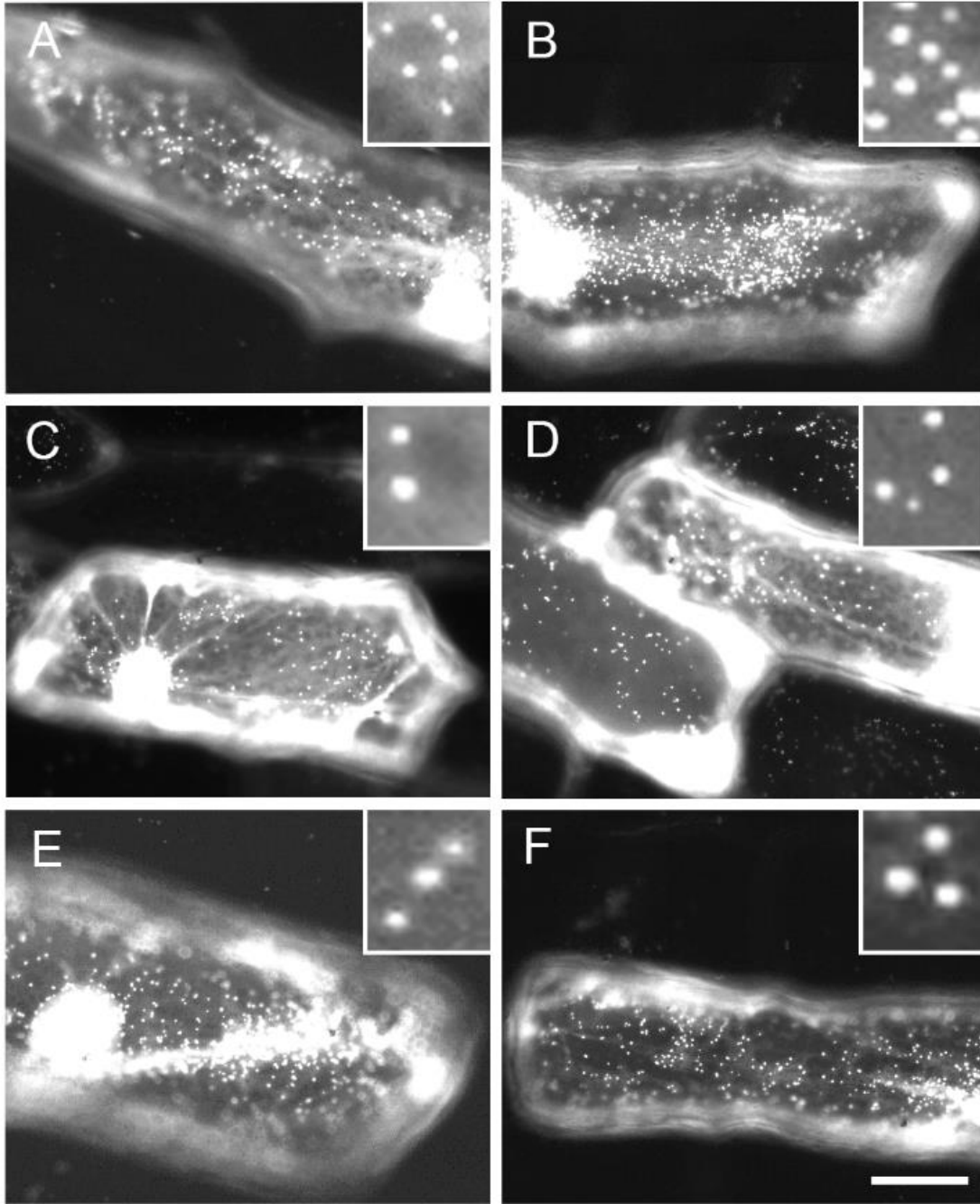


Figure 3.3: Expression level of RFP reporter protein in peroxisomes in onion epidermal cells by epifluorescence microscopy. Expression of A) only the RFP with APUM23 target sequence in the 3' UTR, B) only the RFP with APUM2 target sequence in the 3' UTR, C) RFP with APUM23 target sequence in the 3' UTR co-expressed with tAPUM23-PIN. D) RFP with APUM2 target sequence in the 3' UTR co-expressed with tAPUM23-PIN. E) RFP with APUM23 target sequence in the 3' UTR co-expressed with APUM2-PIN. F) RFP with APUM2 target sequence in the 3' UTR co-expressed with APUM2-PIN. Bar = 26 μ m.

constructs. In an attempt to ensure that the PIN endonuclease was present at a level sufficient to degrade a large proportion of the RFP mRNA, the APUM-PIN plasmid DNA was used in a molar excess amount compared to the RFP mRNA construct in the transfection assays. When tAPUM23-PIN (3µg DNA) was co-expressed with APUM23 target (1µg DNA) or the APUM2 target (1µg DNA), and when APUM2-PIN (3µg DNA) was co-expressed with APUM23 target (1µg) or the APUM2 target (1µg), onion epidermal cells showed high levels of peroxisome RFP fluorescence compared to the cells that expressed the RFP constructs alone (Figure 3.3C, D, E, F). Similar to the expression results with the RFP constructs alone, co-expression resulted in variable levels of peroxisome fluorescence in epidermal cells. These variable results were also consistently observed when the ratio of the plasmid DNA concentration was changed (e.g., when 1µg each of RFP DNA plasmid was co-expressed with APUM-PIN DNAs) or when incubation times were shortened or lengthened (8 hrs or 48 hrs, compared to 24 hrs).

3.2.2.2 – Western blot analysis of RFP expression RFP expression

The cell-to-cell variability observed using epi-fluorescence imaging made it challenging to determine whether there were significant differences in RFP expression in cells that co-expressed the RFP mRNA target and the corresponding APUM-PIN protein. Therefore, to more accurately quantify the expression level of the RFP and APUM-PIN proteins, protein extracts from transfected cells were quantified by western blotting. The western blots were probed with the RFP antibody and Flag antibody to quantify the expression level of RFP and APUM-PIN, respectively. Only cells that were co-transfected with the RFP and APUM-PIN plasmids provided meaningful data, since APUM-PIN expression levels served as an internal control to normalize the RFP protein expression levels.

Five western blot replicates were performed for each set of protein expression experiments. A representative replicate blot is shown in Figure 3.4A. Additional replicate data is shown in the Appendix section. The western blot using the anti-RFP antibody shown in Figure 3.4 (left panel) clearly showed a single RFP reporter protein band (~27 kD) in all lanes that contained protein from transfected cells. The blot probed with anti-Flag antibody (right panel) shows significant background signal in each of the lanes. However, intense signal was observed for the APUM2-PIN protein (lanes 1 and 2), with a weaker signal observed for the tAPUM23-PIN protein. To obtain quantitative data at the protein expression level, densitometry of the western blot bands was performed. Figure 3.4B shows the measurements from densitometry scanning of the bands observed in the representative western blots in Figure 3.4A. Table 3.1 summarizes the results of the densitometry scans for all of the replicated blots (Figure 3.4 and Supplementary Figure S1).

To normalize the expression of the RFP expression patterns relative to the APUM-PIN expression as an internal control, the measured level of RFP in each replicate was divided by the corresponding APUM-PIN level within samples from each replicate. The average ratio for each of the expression conditions was then determined for the five replicates in each treatment (Table 3.2). These average RFP expression levels were then plotted onto a histogram (Figure 3.5). This histogram shows that the co-expression of tAPUM23 with RFP mRNA that contained the APUM23 RNA binding sequence did not result in a decrease in RFP protein level when compared to the expression that was observed when tAPUM23 was co-expressed with the RFP mRNA that contained the non-corresponding APUM2 RNA binding sequence (Figure 3.5). This indicates that tAPUM23-PIN does not bind to its corresponding RNA target in RFP mRNA in vivo, or that binding takes place, but the RFP mRNA is not cleaved and degraded by the PIN RNase. When APUM2-PIN was co-expressed with RFP mRNA that contained the APUM2 RNA

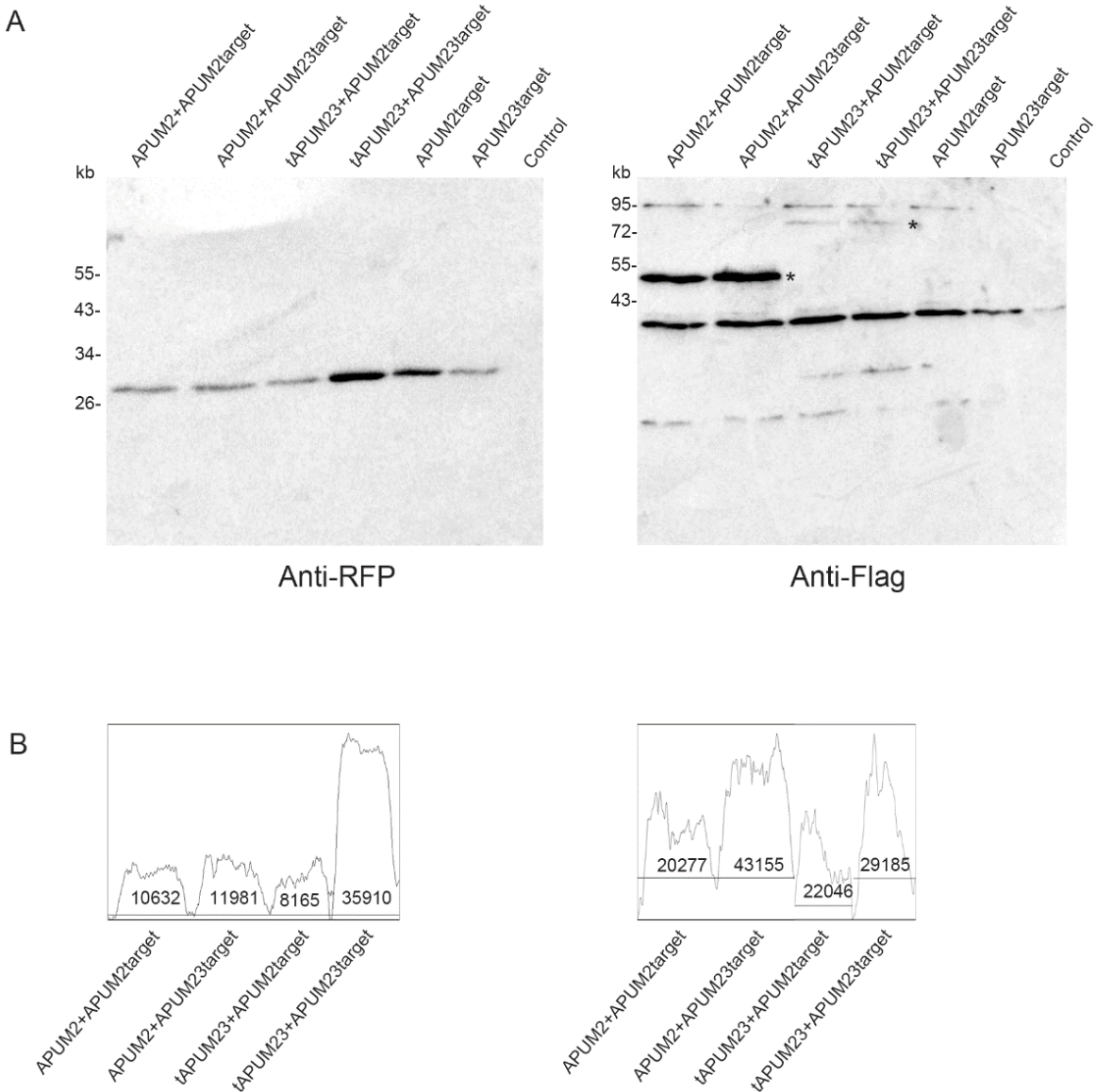


Figure 3.4: Expression level of RFP reporter and APUM-PIN proteins. A) Expression level of RFP (left) and APUM-PIN (right) analyzed by western blotting. RFP reporter protein was identified with anti-RFP antibody (left). APUM-PIN proteins were identified with anti-Flag antibody (right). tAPUM2-PIN (~52 kD) and APUM23-PIN (~75 kD), are denoted by asterisks (*) on the blot. The control lane contained non-transfected onion epidermal cell extract. B) Densitometric analysis of the expression level of RFP targets and APUM-PIN proteins. The densitometry peaks are aligned with the lanes on the western blots. To adjust for expression differences between samples, only the lanes that contained co-expressed (RFP and APUM-PIN) were analyzed by densitometry, as the APUM-PIN protein level served as the internal control for protein expression. The expression level is shown as the area of the curve.

Table. 3.1: Quantified expression level of RFP and APUM-PIN. Summary of the expression level of RFP reporter (identified by the RNA target sequence its mRNA contained) and APUM proteins by western blotting. Each treatment was replicated five times. The expression level of both APUM RFP targets and APUM proteins were shown as pixel quantity under the peak.

Treatment	Replicate	Expression level (area under the peaks)	
		RFP	APUM protein
tAPUM23-PIN + APUM23 target	1	35910.53	29185.86
	2	25549.15	29671.50
	3	9728.82	9758.82
	4	13135.58	16485.44
	5	12500.15	7723.67
tAPUM23-PIN + APUM2 target	1	8165.05	22046.93
	2	7845.08	7104.67
	3	21526.36	10510.84
	4	18000.77	22094.74
	5	26953.56	17567.45
APUM2-PIN + APUM23 target	1	11981.97	43155.00
	2	6452.15	28017.66
	3	1763.15	19301.3
	4	6633.20	50334.07
	5	240.22	2637.87
APUM2-PIN + APUM2 target	1	10632.89	20277.56
	2	13470.41	40819.71
	3	23603.95	69060.22
	4	1189.23	9933.34
	5	21321.80	56599.17

Table. 3.2: Protein expression levels of RFP from mRNA containing the APUM23 target or APUM2 target relative to the expression of the APUM-PIN expression as determined by western blot densitometry from Table 3.1. Summary of data of the relative expression levels of RFP (identified by the RNA target sequence its mRNA contained) to its corresponding APUM-PIN protein level. The relative expression levels were calculated by dividing the expression level of the RFP protein by the expression level of the corresponding APUM-PIN proteins. The combinations of APUM proteins and RFP expressed from the mRNA targets were shown as treatments in the table. The mean expression levels are mean values of the five replicates of the treatment.

Treatment	Relative expression level					Mean
	1	2	3	4	5	
tAPUM23-PIN + APUM23 target	1.230	0.861	0.997	0.797	1.618	1.101
tAPUM23-PIN + APUM2 target	0.370	1.104	2.048	0.815	1.534	1.174
APUM2HD-PIN + APUM23 target	0.278	0.230	0.091	0.132	0.091	0.164
APUM2HD-PIN + APUM2 target	0.524	0.330	0.342	0.127	0.377	0.340

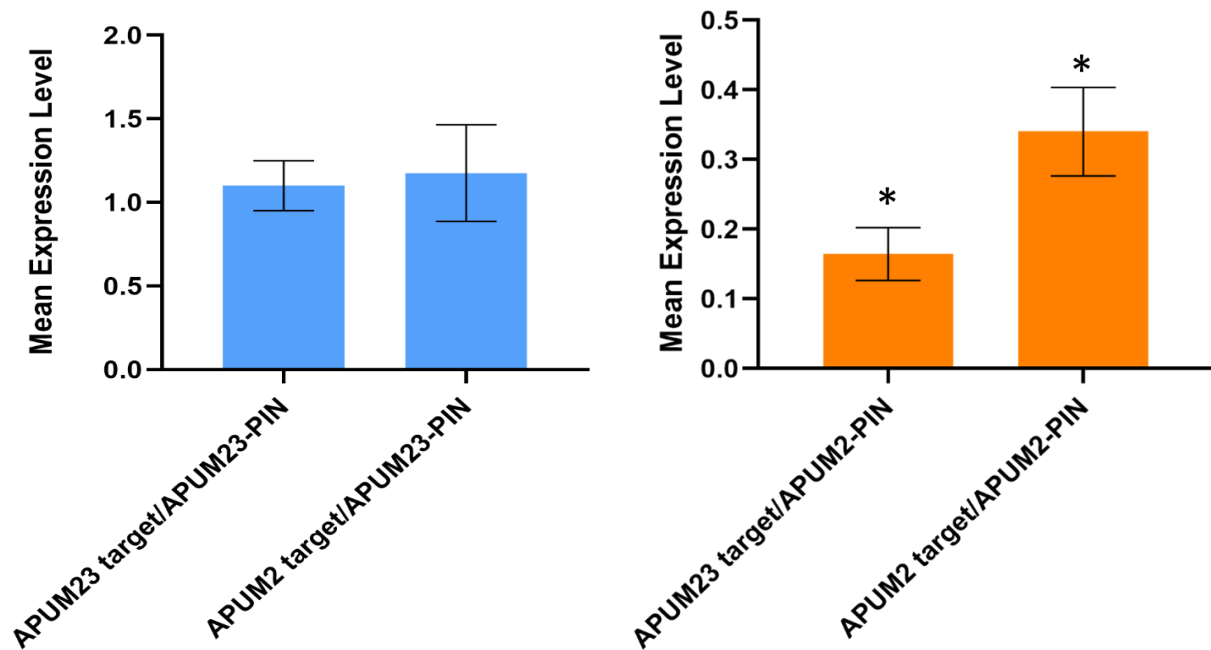


Figure 3.5: Relative expression levels of RFP and APUM-PIN protein from co-expression experiments. Mean expression level of RFP as determined by ratios of RFP and APUM-PIN protein based on densitometry of western blot band intensities. Particle bombardment was used to transfect onion epidermal cells with 3 μ g tAPUM23-PIN (left panel) or APUM2-PIN (right panel) and 1 μ g RFP plasmid containing the APUM23 target or APUM2 target in the 3' UTR of the transcribed RFP mRNA. Mean values \pm SEM were based on a value of n = 5 (p < 0.05).

binding site, there was an approximately 2-fold increase in RFP protein compared to the co-expression of APUM2-PIN with RFP mRNA that contained the APUM23 RNA binding site (Figure 3.5). This is opposite to what would be expected if APUM2-PIN successfully bound and cleaved its RNA target in RFP mRNA. This suggests that APUM2-PIN may bind to the RFP mRNA target that contains the APUM2 consensus binding site, but APUM2-PIN is not actively cleaving the mRNA, but rather may be stabilizing it instead. Overall, these results indicate that this APUM-PIN hybrid is not effective in binding specifically to its RNA target and cleaving it.

3.2.2.3 *qPCR analysis of RFP mRNA levels*

Coupling the PIN RNase domain to tAPUM23 or APUM2 is expected to result in the cleavage and subsequent degradation of the corresponding target RFP mRNA. Therefore, quantitative reverse transcription PCR (qPCR) was performed to quantify the RFP mRNA levels in epidermal peels expressing the RFP construct alone, or in combination with tAPUM23-PIN protein to determine if tAPUM23-PIN can cleave and lead to the degradation of the target mRNA. The mean Ct value of three qPCR replicates was measured for each treatment. The mean delta Ct (Δ Ct mean) was calculated using the following the formula: Δ Ct mean = $Ct_{\text{gene test mean}} - Ct_{\text{endogenous control mean}}$. The $Ct_{\text{gene test mean}}$ represents the mean Ct value of the samples that were amplified with the RFP primer pair. The $Ct_{\text{endogenous control mean}}$ represents the mean Ct values of the samples that were amplified with Actin and GAPDH primers as internal controls. The log transformation of Δ Ct mean was then calculated to determine the mRNA expression level (Figure 3.6). In Figure 3.6A, when the endogenous control is GAPDH (top panels), the first replicate (left panel) showed that tAPUM23-PIN appeared to result in a decrease in RFP mRNA when both are expressed in the cell, compared to the RFP mRNA expressed alone. However,

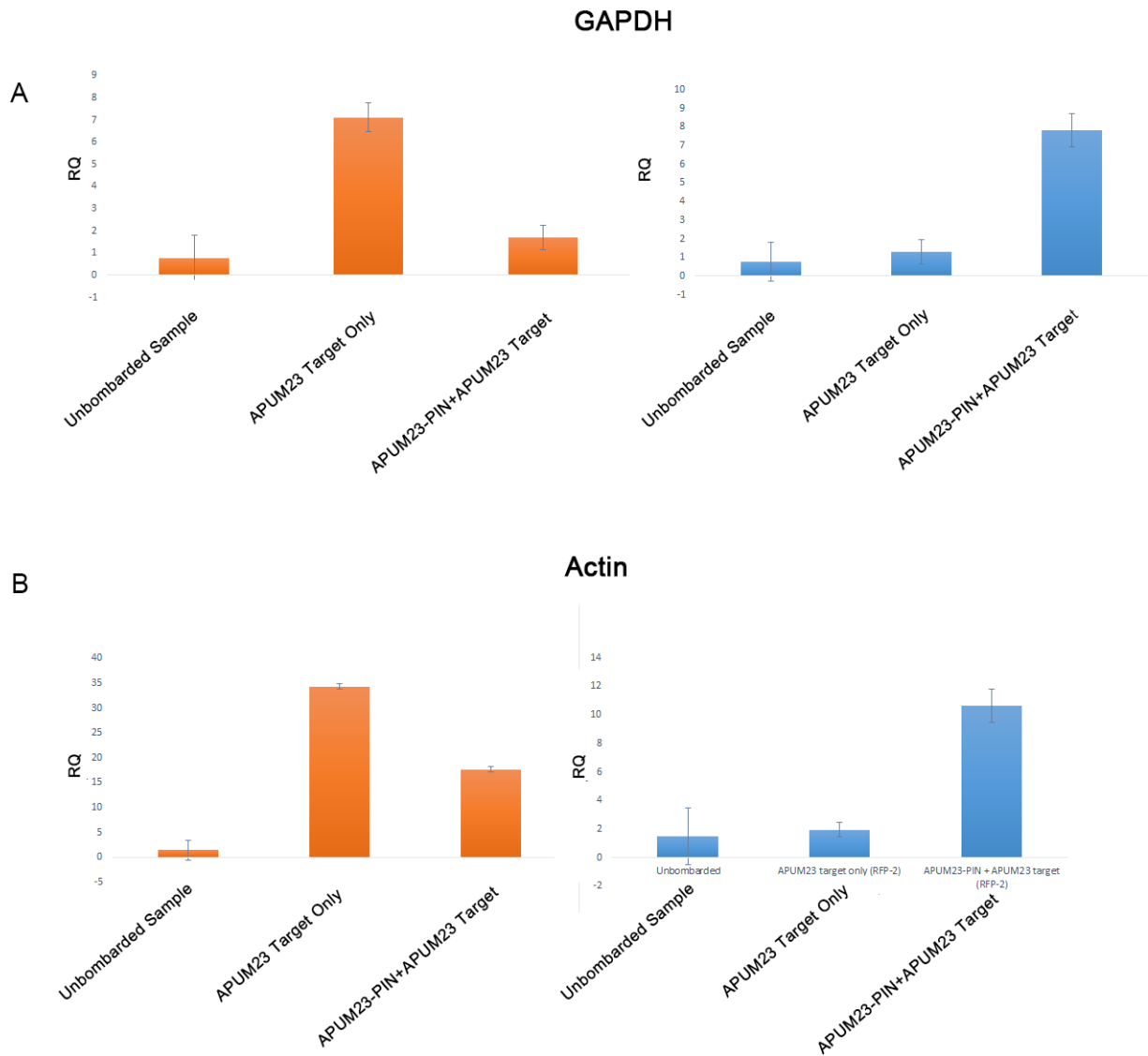


Figure 3.6: Levels of RFP mRNA in cells expressing the RFP construct alone or in combination with tAPUM23-PIN. qPCR results APUM23 The endogenous control mRNA levels were measured using primers that amplified A) GAPDH and B) actin mRNAs. The APUM23 target-RFP mRNA in the bar graphs with orange and blue bars were from biological replicate dataset 1 and 2, respectively. RQ represents relative quantification of the mRNA level. \pm SD were based on a value of n = 3 qPCR replicates.

replicate 2 (right panel) showed the complete opposite result, with co-expression resulting in a stabilization of the mRNA target. When actin primers were used as the internal control, a similar result was obtained (Figure 3.5B). These results indicate that the two biological replicates that were performed were not consistent, and that the experiment would need to be repeated with additional replicates being performed.

3.2.3 Effect of a recombinant tAPUM–PAP fusion protein on augmenting the polyadenylation of RFP target mRNAs in onion epidermal cells

Puf proteins have been fused to other protein domains in an effort to differentially regulate the status of target mRNAs. For example, Puf proteins were fused to GFP to be used for labeling the specific target mRNA of the Puf protein (Yoshimura *et al.*, 2012). They were also fused to tristetraprolin, a zinc-finger protein that promotes mRNA decay (Abil *et al.*, 2014), and CAF1, another post-transcriptional regulator that removes the poly(A) tail from mRNAs and represses translation (Cooke *et al.*, 2011). GLD-2 is a poly(A) polymerase that plays an important role in mRNA polyadenylation (Kwak *et al.*, 2004). Previous research showed that when GLD-2 was fused to a Puf protein, GLD-2 was activated and promoted translation of its target mRNAs because of enhanced polyadenylation (Cooke *et al.*, 2011). In an attempt to use APUM23 as a translational enhancer of its target mRNA, GLD-2 was fused to tAPUM23 and APUM2-HD at their amino- and carboxyl-terminal ends and co-expressed in onion epidermal cells with a CFP mRNA target that contained three tandem repeats of the APUM23 or APUM2 target consensus sequence. The APUM-PAP strategy to enhance polyadenylation and translation is shown at Figure 3.7. Poly(A) tail assays were used to determine whether CFP mRNA polyadenylation was enhanced by an interaction with APUM-PAP. Increased polyadenylation of CFP mRNA is expected to result in an increase in translation of CFP protein, as measured by western blots.

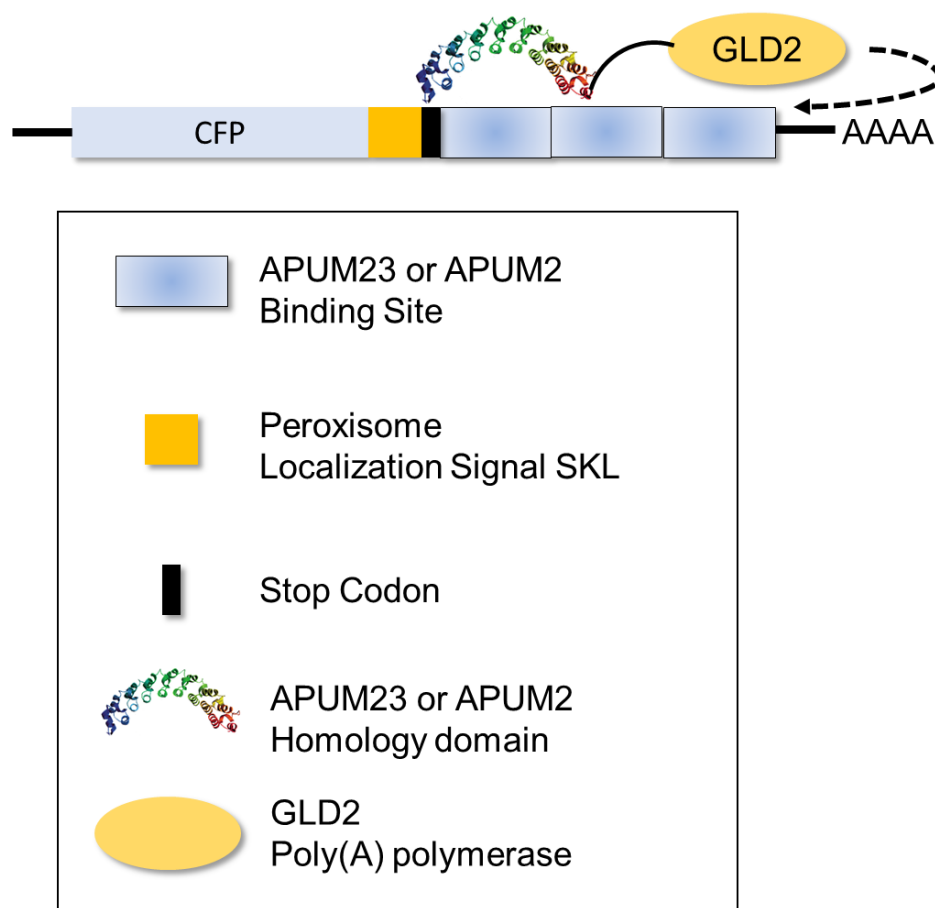


Figure 3.7: Strategy for using APUM proteins fused to the GLD-2 poly(A) polymerase (PAP) to enhance polyadenylation of specific mRNAs. tAPUM23 or APUM2-HD were fused to GLD-2 (PAP) and the N and C terminus. The co-expressed CFP mRNA contains three tandem copies of the APUM23 or APUM2 RNA target. The APUM-PAP fusion will bind to the mRNA target *in vivo* and this interaction will allow PAP to extend the poly(A) tail on the mRNA. The extended poly(A) tail should enhance translation through the increased binding of poly(A)-binding protein. Box sizes are not relative to actual sequence length.

3.2.3.1 Determination of RFP mRNA Poly(A) length

If PAP-tAPUM23 recognizes its specific CFP mRNA target, the PAP activity is expected to increase its poly(A) tail length. Figure 3.8A shows the strategy to determine the mRNA poly(A) tail length. Poly(A) tail assays involved the addition of a guanosine/inosine (G/I) tail to 3' end of all mRNAs. Specific amplification of the RFP mRNA poly(A) tail by PCR from a cDNA template used an RFP-specific upstream primer and a universal primer (poly(C)-TT) that annealed to the G/I tail to provide an estimate of poly(A) tail length. The size of the PCR product indicates the length of the poly(A) tail. CFP specific reverse and forward primers demonstrated that the cDNA and PCR reactions were successful. Figure 3.8B showed that CFP mRNA expression alone produced a strong PCR product that was just over 300 bp in length, with a smear of product above 500 bp. In cells co-expressing the CFP mRNA target and PAP-tAPUM23, the heavy PCR product above 300 bp became more diffuse and the smear of PCR product extended up to 1500 bp. This single replicate result suggests that PAP-tAPUM23 binds to its CFP mRNA target and extends the poly(A) tail.

3.2.3.2 Quantification of RFP and APUM-PAP levels by western blotting

To determine if the apparent increase in polyadenylation efficiency of the reporter mRNA enhanced the translation efficiency of the mRNA, quantitative western blotting was performed. The results of the initial expression assays used the RFP reporter that contained a single copy of the APUM23 target RNA consensus sequence. In this experiment control, RFP plasmids alone, and RFP plasmids co-expressed with the PAP-APUM or APUM- PAP constructs is shown in

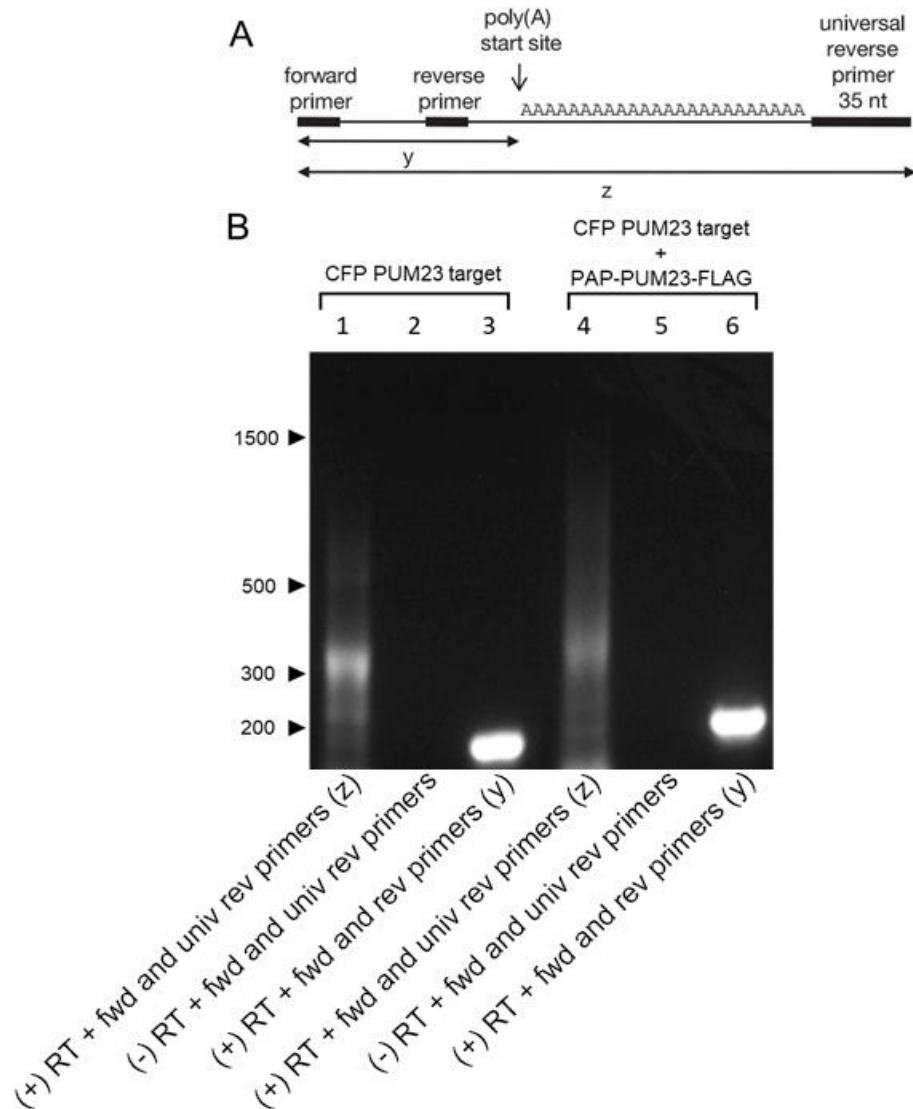


Figure 3.8: Poly(A) tail length of CFP-APUM 23 target. Particle bombardment was used to transfect onion epidermal cells with 0.5 μ g CFP-APUM23 target and 1 μ g PAP-APUM23. A) The primers used for amplifying cDNA derived from CFP-APUM23 target mRNA. The universal reverse primers were 35 nt primers designed downstream the poly(A) tail in order to test its length. B) Poly(A) tail length of CFP-APUM23 target with or without PAP-PUM23-Flag. Agarose gel was used to show the Poly(A) tail length. The smear on the gel showed the range of Poly(A) tail length of CFP-APUM23 target. The y and z represents the position of the product in A).

Figure 3.9. The blot indicates that the RFP vectors expressed well in all bombarded samples. The RFP band intensities indicated that RFP encoded by mRNA that contained the APUM23 consensus target sequence that was co-expressed with PAP-tAPUM23 visually showed more expression compared to the RFP encoded by mRNA that contained the APUM23 target expressed on its own (Figure 3.9). When the blot was probed with anti-Flag antibody (right panel), only the PAP-APUM23-Flag lane showed a band. This indicated that the other PAP-containing constructs did not express well. As a result, the PAP-APUM23-Flag protein was used for further analysis.

Next, the PAP-APUM23-Flag protein was co-expressed with CFP mRNA that contained the triple-tandem APUM23 target RNA or triple-tandem APUM2 target RNA sequence in the 3' UTR. Protein extracts from these bombardments were analysed by western blot, similar to the APUM-PIN experiments. The CFP protein was probed with the GFP antibody in this experiment (Figure 3.10). Similar to Figure 3.4, the anti-Flag probed western blot showed a weaker signal compared to that of the reporter protein. The CFP, APUM-PAP and PAP-APUM bands were quantified by densitometry and a histogram was plotted to indicate the relative differences in expression level of CFP mRNA containing the APUM23 target or APUM2 target when co-expressed with PAP-APUM23 (based on four replicates) (Figure 3.8B). The relative expression level was determined by dividing the RFP antibody densitometric value by the Flag antibody value from each sample. This data suggested that the CFP mRNA containing the APUM23 triple-target produced greater CFP expression compared to the CFP containing the APUM2 triple-target (Figure 3.10B). However, when a t-test was performed for significance ($p < 0.05$), the APUM23 triple target CFP treatment levels were not significantly different from the APUM2 triple-target CFP due to the variability observed in the CFP-APUM23 target data. Despite this

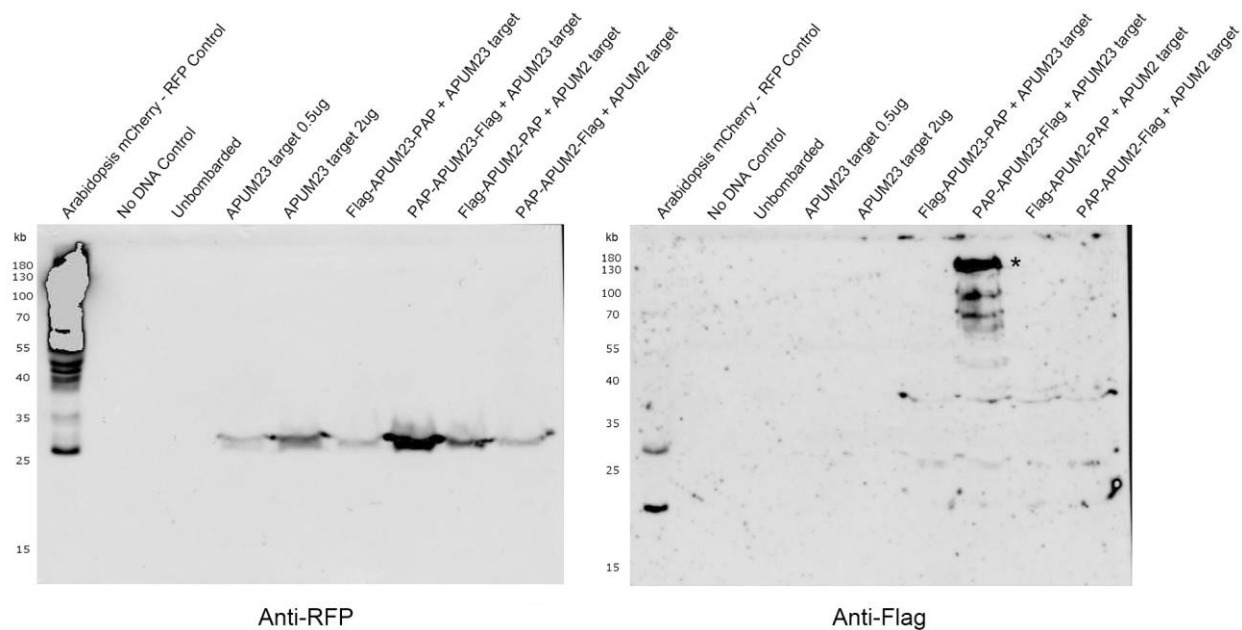


Figure 3.9: Expression level of RFP when co-expressed with/without APUM proteins fused to PAP. Particle bombardment was used to transfect onion epidermal cells with 1 μ g PAP-APUM23-Flag, PAP-APUM2-Flag, Flag-APUM23-PAP or Flag-APUM2-PAP with 0.5 μ g RFP-APUM23 target or RFP-APUM2 target to determine the construct for further experiments. Arabidopsis seedlings expressing mCherry (a variant of RFP) was used as a control to verify that the RFP antibody was binding to its target. The Flag protein signal is denoted with an asterisk in the right panel.

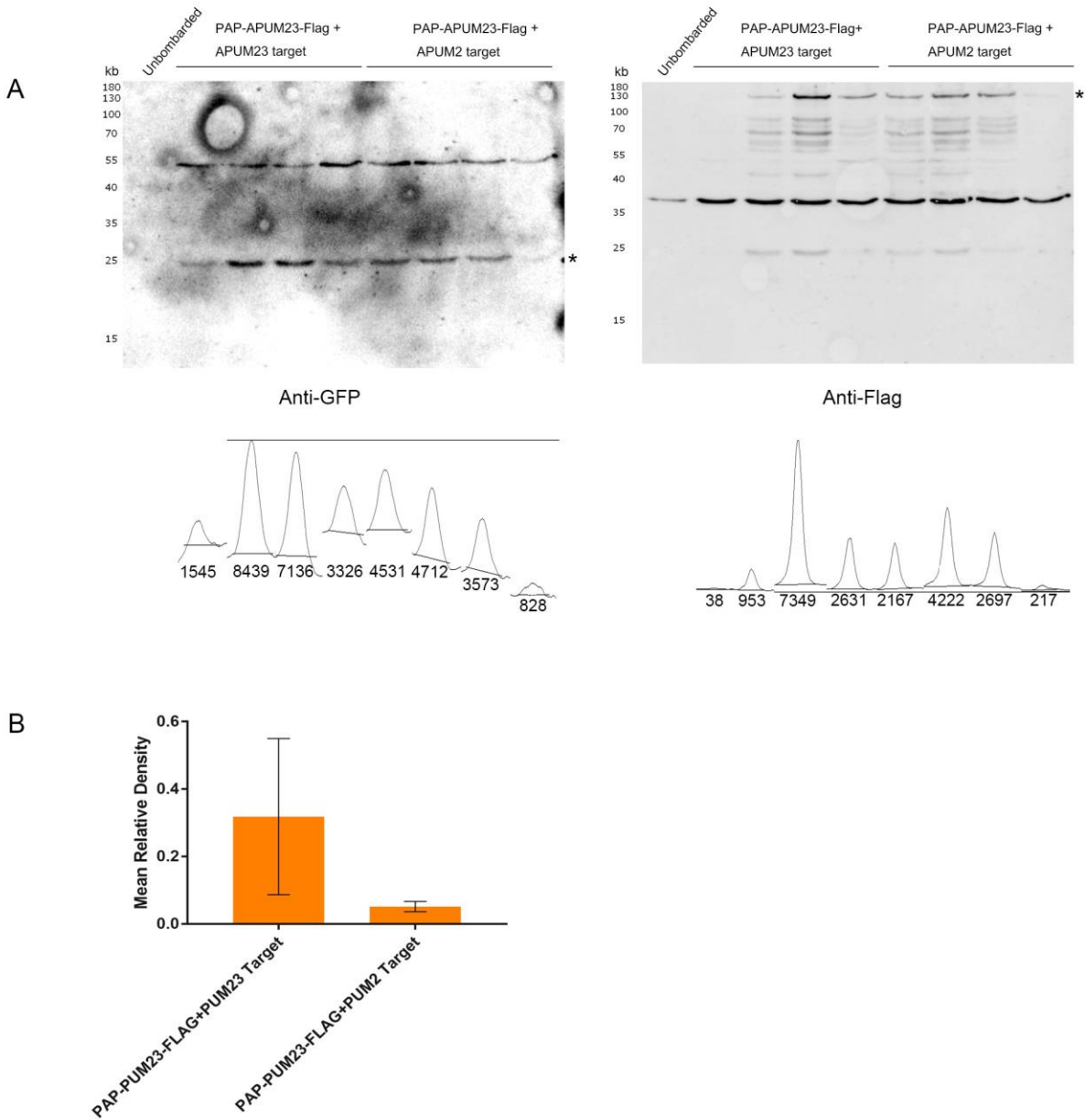


Figure 3.10: Expression level of CFP when co-expressed with or without PAP-APUM23 A) Particle bombardment was used to transfect onion epidermal cells with 1 μ g PAP-APUM23-Flag and 0.5 μ g plasmid encoding CFP mRNA containing the APUM23 or APUM2 triple target. CFP-APUM targets were quantified from western blots (four replicates shown) with GFP antibody (recognizes CFP) and PAP-APUM23 with Flag antibody. B) Mean relative expression level of CFP translated from mRNA containing the APUM23 RNA target or the APUM2 RNA target when co-expressed with PAP-tAPUM23. Mean values \pm SEM were based on a value of $n = 4$ western blotting results.

lack of significance, there is a trend toward higher expression when PAP-APUM23 is co-expressed with mRNA containing its consensus target sequence. Coupled with the evidence showing an extended poly(A) tail on CFP mRNA when co-expressed with PAP-APUM23 (Figure 3.8), this suggests that PAP-APUM23 can recognize its specific RNA target and function as a sequence specific poly(A) polymerase, at least at a low level as shown here.

3.3 Discussion

In this chapter, experiments were described that characterized the functional activity of APUM23 fused to PIN and GLD-2 to modify the degradation or translation of target mRNAs that encoded fluorescent proteins. Full-length APUM23, truncated APUM23 lacking the nuclear localization signal (NLS), and the APUM2 homology domain (HD) were cloned and fused to RFP. Subcellular localization analysis demonstrated that truncated APUM23-RFP was localized to the cytosol, a similar location as APUM2-RFP (Figure 3.3). This was in contrast to the nucleolar localization of the wild type, full-length APUM23 fused to RFP. This observation demonstrated that this truncated APUM23 could be used in subsequent effector domain fusion experiments that could target cytosolic mRNAs.

Truncated tAPUM23 and APUM2-HD were then fused to the amino-terminus of the PIN RNase domain to determine whether these fusion proteins could bind and degrade RFP mRNA that contained their consensus RNA binding sequence in the cytosol of onion epidermal cells. A peroxisomal targeting sequence was appended to the RFP sequence with the reasoning that RFP localized to peroxisomes would provide a better indication of overall fluorescence when visualized by epifluorescence microscopy. However, due to the variability of RFP expression between transfected cells, this approach did not provide informative data on the cellular

expression of RFP. Alternative approaches (western blotting and qPCR) were then used in an attempt to provide a more accurate measure of RFP protein levels and RFP mRNA abundance. The western blot analysis did not show statistically significant differences in RFP levels between control (RFP mRNA with APUM2 target) and specific (RFP mRNA with APUM23 target). However, when APUM2HD-PIN was expressed with RFP mRNA containing the APUM2 target, there was a significant increase in RFP levels. This observation was not expected, but it may be that the fused PIN RNase activity was not active *in vivo* and can stabilize the RFP mRNA when coupled to APUM2-HD.

In an attempt to use the APUM23 backbone to increase translation of its target mRNA, the PAP-tAPUM23 construct was co-expressed with CFP containing a triple-repeat of the APUM23 target consensus sequence. This was in an effort to increase the poly(A) tail length to enhance translation initiation and protein synthesis. This approach was used previously to successfully stimulate translation of a reporter gene (Kwak *et al.*, 2004). Western blot analysis indicated that PAP-tAPUM23 was the only APUM construct that expressed at high levels (Figure 3.9). There was a subtle increase in polyadenylation of CFP mRNA when co-expressed with PAP-tAPUM23 (Figure 3.10), and CFP protein also trended upward. These results suggested that the PAP domain was functional and that tAPUM23 bound to its target consensus sequence in the 3' UTR of CFP mRNA. RFP mRNA levels, as measured by qPCR, showed extremely variable results between the two replicates, indicating that this experiment was not consistent and needs to be repeated to achieve more reliable results.

Overall, these results provide some evidence that supports the hypothesis that APUM23 can be used to bind and modulate the activities of target mRNAs in the cell. The results presented here must be validated with further replicates and optimization of the expression

system. Previous studies that used effector domain fusions to Puf RNA-binding domains showed a large effect on mRNA metabolism. Canonical Puf RNA-binding domains from human and yeast have been used to alter mRNAs using an RNase (Choudhury *et al.*, 2012), a translational activator or inhibitor (Cooke *et al.*, 2011), and splicing activator or inhibitor (Wang *et al.*, 2009) as effector domains. GFP has also been fused to the Puf domain to track the movement and localization of mRNAs in living cells (Ozawa *et al.*, 2007). Future approaches for using Puf proteins to regulate RNA include: RNA editing, multiple alternative splicing, RNA trafficking and transcription elongation (Wang *et al.*, 2013). Included in these approaches can be alteration of the binding specificity code in Puf domains to bind specific endogenous sequences in the cell. Indeed, Puf engineering has resulted in Puf RNA-binding domains that contain swapped Puf repeats (Wang *et al.* FEBS J. 2013, 280: 3755). Canonical Puf proteins contain eight Puf repeats and recognize eight nucleotides. In contrast, APUM23 contains 11 Puf repeats and has an 11 nucleotide consensus sequence, and some of the Puf repeats have been altered to bind different RNAs (Zhang and Muench, 2015; Bao *et al.*, 2017). As a result of this increased sequence specificity and the preliminary evidence presented in this thesis, APUM23 has the potential to be engineered to regulate cellular RNAs with high specificity. However, a deeper analysis of the expression assays presented here will need to be performed to demonstrate that consistent and statistically significant results can be obtained.

Acknowledgements: I thank Mitchell Alberts for his assistance with the experiments related to the poly(A) polymerase (PAP) fusion experiments.

Chapter Four: Discussion and Future Direction

Puf RNA-binding proteins contain modular Puf repeats, each of which binds to single nucleotides in RNA. The nucleotide binding specificity of a Puf repeat is determined by three amino acids (tripartite recognition motifs, TRMs) that interact with each nucleotide, and have led to the identification of Puf nucleotide binding codes (Zhang and Muench, 2015; Wei and Wang, 2015). APUM23 is a unique sequence-specific RNA-binding protein and a member of the Puf family. APUM23 has possesses 11 Puf repeats and a three-dimensional structure that is unique among Puf proteins in eukaryotes (Zhang and Muench, 2015; Bao *et al.*, 2017). The enhanced binding surface of APUM23 (11 repeats versus 8 repeats of canonical Pufs) indicated that RNA-binding domain of this protein would provide an excellent scaffold to fuse with effector domains and used as an RNA-specific post-transcriptional regulator. The modular binding nature of several Puf repeats in APUM23 as shown through domain swapping experiments (Zhang and Muench, 2015; Bao *et al.*, 2017) indicates the potential of this protein to be altered to bind a range of RNA targets in the cell. In this study, APUM23 was fused to two effector domains, a non-specific RNA endonuclease and an RNase (PIN) and poly(A) polymerase (GLD-2), in an attempt to produce sequence-specific post-transcriptional regulators. The research presented in this thesis provides the first evidence demonstrating that APUM23 works as a sequence-specific scaffold to direct the activities of effector domains and regulate the physiology of mRNAs in the cell.

The results of this thesis research were presented in Chapter 3. These results provided insight into ability of APUM23 to bind to target sequences *in vivo*, and the potential of APUM23 to mediate the activities of the effector domains (PIN and GLD-2) on mRNA. The research strategy utilized a well-characterized plant cell expression system, the onion epidermal cell. This

provided a simple platform to transiently express constructs alone or in combination, allowing for expression to be visualized by epi-fluorescence microscopy and biochemically. This system has caveats that were apparent in the quality of the results obtained here. There was inconsistency in the number of cells that transiently expressed the constructs, and in the level of expression from cell-to-cell in the same epidermal segment. As a result, the data showed variability between replicates, and generated data that produced limited statistically significant results. However, the trends that were observed, especially for the PAP-tAPUM23 results, indicate that APUM23-effector interactions are functional in the cell. This research provides a foundation for future experiments that will provide more confidence in the results shown here, and will also involve experiments not performed here

Future work will involve a more in-depth analysis of the experiments reported here. New experiments are also planned that will provide additional insight on APUM23 binding specificity to *in vivo* targets, and activity of the appended effector domains on the target mRNAs.

Additional replicates and approaches that expand on the experimental plan presented here –

Since the results of the experiments show inconsistency between different replicates or a limited number of replicates, further replicates on the western blot quantification, qPCR and poly(A) tail length will provide more reliable results and allow us to make more firm conclusions about APUM23-effector activities. The treatments can be expanded to include poly(A) tail length assays on CFP mRNAs that contained both consensus RNA targets (APUM23 and APUM2 targets) for PAP-APUM23 and PAP-APUM2 proteins. If the onion epidermal cell system continues to provide variable results, transfection of tobacco leaf cells will be performed using an *Agrobacterium tumefaciens* vector system to deliver the gene constructs into cells. This

approach has been used effectively for quantitative transient gene expression in numerous studies (Chen et al., 2013, Adv Tech Biol Med).

In vivo and in vitro protein assays to determine functional activities of APUM and effector domains:

- a) Functional RNA binding activity of the APUM-poly(A) polymerase fusions will be tested by mobility shift assays using synthetic fluorescent RNA. The fusion proteins will either be recombinantly expressed in *E. coli* or purified by FLAG immunoprecipitation of the recombinant protein expressed in onion epidermal cells. Mobility shift experiments using the purified APUM-effector fusion protein will involve incubating the protein with short RNA probes that contain either the APUM23 or APUM2 consensus binding sequence. Additionally, the purified proteins can be tested for effector domain activities by calculating the binding affinity of APUM-GLD-2 and the target RNAs. These will include an RNase assay for PIN domain activity by incubating the fusion protein with longer fluorescent RNA containing the consensus sequence and determining whether cleavage of that RNA can occur *in vitro*. Polyadenylation extension assays can also be performed on short synthetic RNA that possesses a poly(A) tail. If these activities are non-functional, the linker domain that was used to connect PIN and GLD-2 to APUM will be modified in length and sequence composition in an effort to optimize the activity of the effector domains.
- b) Identification of the mRNA targets of specific APUM-effector proteins will be performed to determine the specificity of the fusion proteins *in vivo*. The recombinant APUM-effector proteins will be immunoprecipitated from transiently expressing onion epidermal

cells using the anti-Flag antibody. Associated mRNAs will be extracted and sequenced (RNA-seq) to determine whether the RFP/CFP mRNAs are enriched in the bound fraction.

- c) APUM23 will be engineered by altering the Puf domains so that it can target alternate RNA sequences. This will be based on the domain swapping experiments performed previously in the lab (Zhang 2015). The onion epidermal expression system will be used for this analysis.

If these APUM23-effector domains are proven to be effective post-transcriptional regulators *in vivo*, APUM23 would provide a new backbone for the targeting and manipulation of endogenous mRNAs, with potentially a higher specificity than other canonical Puf backbones. Using APUM23 as a scaffold to make RNA sequence specific functional domains could be a useful tool for agricultural and medical research, and even a therapeutic tool. Since this technology can be adapted for targeting of RNAs in any cell type and would be adaptable for cytosolic, mitochondrial and chloroplastic RNA regulation.

References

- Abbasi, N., Kim, H. B., Park, N. I., Kim, H. S., Kim, Y. K., Park, Y. I. and Choi, S. B. (2010) APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis. *The Plant Journal*, 64, 960-976.
- Abbasi, N., Park, Y.-I. and Choi, S.-B. (2011) Pumilio Puf domain RNA-binding proteins in Arabidopsis. *Plant Signaling & Behavior*, 6, 364-368.
- Abil, Z., Denard, C. A. and Zhao, H. (2014) Modular assembly of designer PUF proteins for specific post-transcriptional regulation of endogenous RNA. *Journal of Biological Engineering*, 8, 7.
- Bao, H., Wang, N., Wang, C., Jiang, Y., Liu, J., Xu, L., Wu, J. and Shi, Y. (2017) Structural basis for the specific recognition of 18S rRNA by APUM23. *Nucleic Acids Research*, 45, 12005-12014.
- Belostotsky, D. A. and Sieburth, L. E. (2009) Kill the messenger: mRNA decay and plant development. *Current Opinion in Plant Biology*, 12, 96-102.
- Blancafort, P., Segal, D. J. and Barbas, C. F., 3rd (2004) Designing transcription factor architectures for drug discovery. *Molecular Pharmacology*, 66, 1361-1371.
- Campbell, Z. T., Valley, C. T. and Wickens, M. (2014) A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nature Structural & Molecular Biology*, 21, 732-738.
- Castello, A., Frese, C. K., Fischer, B., Jarvelin, A. I., Horos, R., Alleaume, A. M., Foehr, S., Curk, T., Krijgsveld, J. and Hentze, M. W. (2017) Identification of RNA-binding domains of RNA-binding proteins in cultured cells on a system-wide scale with RBDmap. *Nature Protocols*, 12, 2447-2464.
- Chen, T., Cui, P., Chen, H., Ali, S., Zhang, S. and Xiong, L. (2013) A KH-domain RNA-binding protein interacts with FIERY2/CTD phosphatase-like 1 and splicing factors and is important for pre-mRNA splicing in Arabidopsis. *PLoS Genetics*, 9, e1003875.
- Chen, Y. and Varani, G. (2013) Engineering RNA-binding proteins for biology. *The FEBS Journal*. 280, 3734-3754.
- Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., Castleden, I., Song, Y., Song, B., Huang, J., Liu, X., Xu, X., Lim, B. L., Bond, C. S., Yiu, S. M. and Small, I. (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *The Plant Journal*, 85, 532-547.
- Choudhury, R., Tsai, Y. S., Dominguez, D., Wang, Y. and Wang, Z. (2012) Engineering RNA endonucleases with customized sequence specificities. *Nature Communications*, 3, 1147.
- Chuong, S. D., Park, N. I., Freeman, M. C., Mullen, R. T. and Muench, D. G. (2005) The peroxisomal multifunctional protein interacts with cortical microtubules in plant cells. *BMC Cell Biology*, 6, 40.

- Cooke, A., Prigge, A., Opperman, L. and Wickens, M. (2011) Targeted translational regulation using the PUF protein family scaffold. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 15870-15875.
- Coquille, S., Filipovska, A., Chia, T., Rajappa, L., Lingford, J. P., Razif, M. F., Thore, S. and Rackham, O. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nature Communications*, 5, 5729.
- Dong, S., Wang, Y., Cassidy-Amstutz, C., Lu, G., Bigler, R., Jezyk, M. R., Li, C., Hall, T. M. and Wang, Z. (2011) Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *The Journal of Biological Chemistry*, 286, 26732-26742.
- Edwards, T. A., Pyle, S. E., Wharton, R. P. and Aggarwal, A. K. (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs. *Cell*, 105, 281-289.
- Filipovska, A., Razif, M. F., Nygard, K. K. and Rackham, O. (2011) A universal code for RNA recognition by PUF proteins. *Nature Chemical Biology*, 7, 425-427.
- Forbes, A. and Lehmann, R. (1998) Nanos and Pumilio have critical roles in the development and function of *Drosophila* germline stem cells. *Development*, 125, 679-690.
- Francischini, C. W. and Quaggio, R. B. (2009) Molecular characterization of *Arabidopsis thaliana* PUF proteins--binding specificity and target candidates. *The FEBS Journal*, 276, 5456-5470.
- Gallie, D. R. (1993) Posttranscriptional regulation of gene expression in plants. *Annual Review of Plant Biology*, 44, 77-105.
- Garcia-Bustos, J., Heitman, J. and Hall, M. N. (1991) Nuclear protein localization. *Biochimica et Biophysica Acta*, 1071, 83-101.
- Gavis, E. R. (2001) Over the rainbow to translational control. *Nature Structural Biology*, 8, 387-389.
- Glavan, F., Behm-Ansmant, I., Izaurralde, E. and Conti, E. (2006) Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *The EMBO Journal*, 25, 5117-5125.
- Glisovic, T., Bachorik, J. L., Yong, J. and Dreyfuss, G. (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*, 582, 1977-1986.
- Grishin, N. V. (2000) KH domain: one motif, two folds. *Nucleic Acids Research*, 29, 638-643.
- Gully, B. S., Cowieson, N., Stanley, W. A., Shearston, K., Small, I. D., Barkan, A. and Bond, C. S. (2015) The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Research*, 43, 1918-1926.
- Gupta, Y. K., Lee, T. H., Edwards, T. A., Escalante, C. R., Kadyrova, L. Y., Wharton, R. P. and Aggarwal, A. K. (2009) Co-occupancy of two Pumilio molecules on a single hunchback NRE. *RNA*, 15, 1029-1035.
- Hamid, S. M. and Akgul, B. (2014) Master regulators of posttranscriptional gene expression are

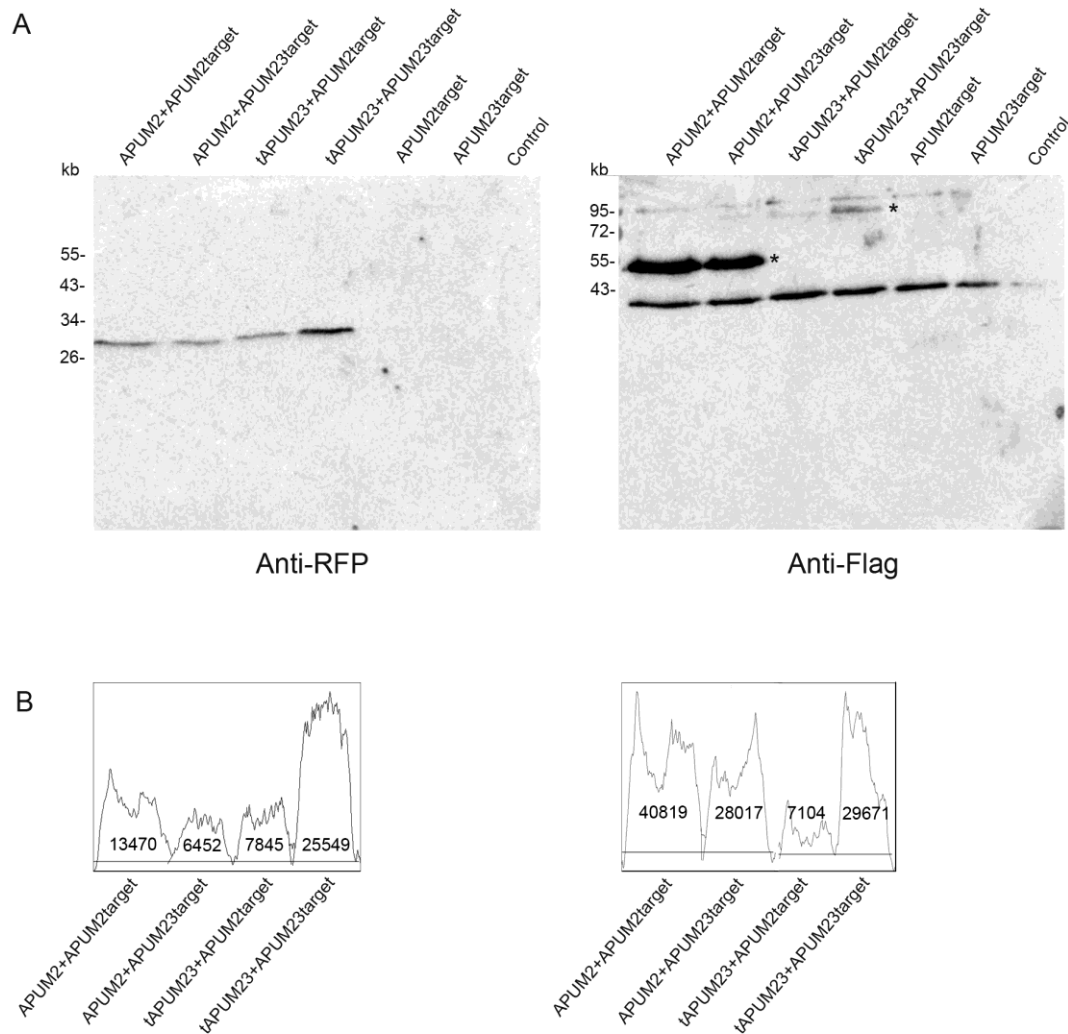
- subject to regulation. *Methods in Molecular Biology*, 1107, 303-310.
- Hentze, M. W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nature Reviews. Molecular Cell Biology*, 19, 327-341.
- Hogan, D. J., Riordan, D. P., Gerber, A. P., Herschlag, D. and Brown, P. O. (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biology*, 6, e255.
- Jurica, M. S. and Moore, M. J. (2003) Pre-mRNA Splicing: Awash in a Sea of Proteins. *Molecular Cell*, 12, 5-14.
- Klug, A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry*, 79, 213-231.
- Kobayashi, A., Kanaba, T., Satoh, R., Fujiwara, T., Ito, Y., Sugiura, R. and Mishima, M. (2013) Structure of the second RRM domain of Nrd1, a fission yeast MAPK target RNA binding protein, and implication for its RNA recognition and regulation. *Biochemical and Biophysical Research Communications*, 437, 12-17.
- Kwak, J. E., Wang, L., Ballantyne, S., Kimble, J. and Wickens, M. (2004) Mammalian GLD-2 homologs are poly(A) polymerases. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 4407-4412.
- Lehmann, R. and Nüsslein-Volhard, C. (1987) Involvement of the pumilio gene in the transport of an abdominal signal in the *Drosophila* embryo. *Nature*, 329, 167-170.
- Little, M., Holmes, G. and Walsh, P. (1999) WT1: what has the last decade told us? *BioEssays: news and reviews in molecular, cellular and developmental biology*, 21, 191-202.
- Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M. and Yan, J. (2013) Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. *Scientific Reports*, 3, 2054.
- Lu, G., Dolgner, S. J. and Hall, T. M. (2009) Understanding and engineering RNA sequence specificity of PUF proteins. *Current Opinion in Structural Biology*, 19, 110-115.
- Matlin, A. J., Clark, F. and Smith, C. W. (2005) Understanding alternative splicing: towards a cellular code. *Nature Reviews. Molecular Cell Biology*, 6, 386-398.
- Matthews, J. M. and Sunde, M. (2002) Zinc Fingers—Folds for Many Occasions. *IUBMB Life*, 54, 351-355.
- Messias, A. C. and Sattler, M. (2004) Structural Basis of Single-Stranded RNA Recognition. *Accounts of Chemical Research*, 37, 279-287.
- Miller, M. A. and Olivas, W. M. (2011) Roles of Puf proteins in mRNA degradation and translation. *Wiley interdisciplinary reviews. RNA*, 2, 471-492.
- Nakel, H., Bonneau, F., Eckmann, C. R. and Conti, E. (2015) Structural basis for the activation of the *C. elegans* noncanonical cytoplasmic poly(A)-polymerase GLD-2 by GLD-3. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 8614-8619.

- Nelles, D. A., Fang, M. Y., Aigner, S. and Yeo, G. W. (2015) Applications of Cas9 as an RNA-programmed RNA-binding protein. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 37, 732-739.
- Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. and Nagai, K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, 372, 432-438.
- Park, J. E., Yi, H., Kim, Y., Chang, H. and Kim, V. N. (2016) Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Molecular Cell*, 62, 462-471.
- Pfaff, C., Ehrnsberger, H. F., Flores-Tornero, M., Sorensen, B. B., Schubert, T., Langst, G., Griesenbeck, J., Sprunck, S., Grasser, M. and Grasser, K. D. (2018) ALY RNA-Binding Proteins Are Required for Nucleocytoplasmic mRNA Transport and Modulate Plant Growth and Development. *Plant Physiology*, 177, 226-240.
- Romanowski, A. and Yanovsky, M. J. (2015) Circadian rhythms and post-transcriptional regulation in higher plants. *Frontiers in Plant Science*, 6, 437.
- Sanfacon, H., Brodmann, P. and Hohn, T. (1990) A dissection of the cauliflower mosaic virus polyadenylation signal. *Genes & Development*, 5, 141-149.
- Schoning, J. C., Streitner, C., Meyer, I. M., Gao, Y. and Staiger, D. (2008) Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis. *Nucleic Acids Research*, 36, 6977-6987.
- Scott, A., Wyatt, S., Tsou, P. L., Rebertson, D. and Allen, N. S. (1999) Model System for Plant Cell Biology: GFP Imaging in Living Onion Epidermal Cells. *BioTechniques*, 26, 1125-1132.
- Shikanai, T. and Fujii, S. (2013) Function of PPR proteins in plastid gene expression. *RNA Biology*, 10, 1446-1456.
- Slomi, H., Choi, M., Slomi, M. C., Nussbaum, R. L. and Dreyfuss, G. (1994) Essential Role for KH Domains in RNA Binding: Impaired RNA Binding by a Mutation in the KH Domain of FMR1 That Causes Fragile X Syndrome. *Cell*, 77, 33-39.
- Szemprich, A. J., Choudhury, R., Wang, Z. and Hajduk, S. L. (2015) In vivo analysis of trypanosome mitochondrial RNA function by artificial site-specific RNA endonuclease-mediated knockdown. *RNA*, 21, 1781-1789.
- Takeshita, D., Zenno, S., Lee, W. C., Saigo, K. and Tanokura, M. (2007) Crystal structure of the PIN domain of human telomerase-associated protein EST1A. *Proteins*, 68, 980-989.
- Tam, P. P., Barrette-Ng, I. H., Simon, D. M., Tam, M. W., Ang, A. L. and Muench, D. G. (2010) The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. *BMC Plant Biology*, 10, 44.
- Thomas, J. M. and Beal, P. A. (2017) How do ADARs bind RNA? New protein-RNA structures illuminate substrate recognition by the RNA editing ADARs. *BioEssays: news and reviews in molecular, cellular and developmental biology*, 39.

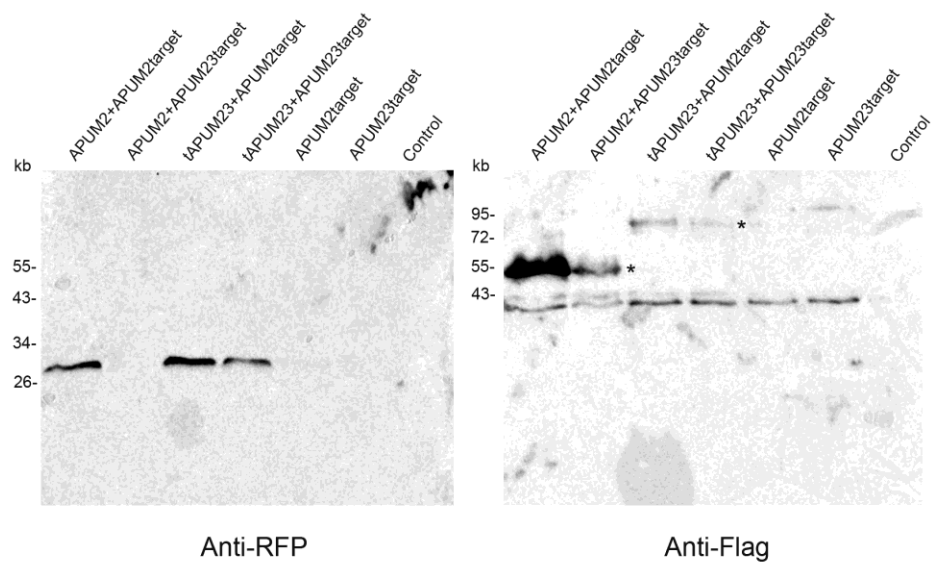
- Wang, M., Oge, L., Perez-Garcia, M. D., Hamama, L. and Sakr, S. (2018) The PUF Protein Family: Overview on PUF RNA Targets, Biological Functions, and Post Transcriptional Regulation. *International Journal of Molecular Sciences*, 19.
- Wang, X., McLachlan, J., Zamore, P. D. and Hall, T. M. T. (2002) Modular Recognition of RNA by a Human Pumilio-Homology Domain. *Cell*, 110, 501-512.
- Wang, X., Zamore, P. D. and Hall, T. M. (2001) Crystal Structure of a Pumilio Homology Domain. *Molecular Cell*, 7, 855-865.
- Wei, H. and Wang, Z. (2015) Engineering RNA-binding proteins with diverse activities. *Wiley Interdisciplinary Reviews. RNA*, 6, 597-613.
- Wickens, M., Bernstein, D. S., Kimble, J. and Parker, R. (2002) A PUF family portrait: 3'UTR regulation as a way of life. *Trends in Genetics*, 18, 150-157.
- Wolins, N. E. and Donaldson, R. P. (1997) Binding of the peroxisomal targeting sequence SKL is specified by a low-affinity site in castor bean glyoxysomal membranes. A domain next to the SKL binds to a high-affinity site. *Plant Physiology*, 113, 943-949.
- Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H. W., Wang, J., Zhu, J. K., Shi, Y. and Yan, N. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, 504, 168-171.
- Yoshimura, H., Inaguma, A., Yamada, T. and Ozawa, T. (2012) Fluorescent probes for imaging endogenous beta-actin mRNA in living cells using fluorescent protein-tagged pumilio. *ACS Chemical Biology*, 7, 999-1005.
- Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J. and Wickens, M. (1997) A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature*, 390, 477-484.
- Zhang, C. and Muench, D. G. (2015) A Nucleolar PUF RNA-binding Protein with Specificity for a Unique RNA Sequence. *The Journal of Biological Chemistry*, 290, 30108-30118.
- Zhang, W., Wang, Y., Dong, S., Choudhury, R., Jin, Y. and Wang, Z. (2014) Treatment of type 1 myotonic dystrophy by engineering site-specific RNA endonucleases that target (CUG)(n) repeats. *Molecular Therapy*, 22, 312-320.
- Zhao, B., Schneid, C., Iliev, D., Schmidt, E. M., Wagner, V., Wollnik, F. and Mittag, M. (2004) The circadian RNA-binding protein CHLAMY 1 represents a novel type heteromer of RNA recognition motif and lysine homology domain-containing subunits. *Eukaryotic Cell*, 3, 815-825.

APPENDIX 1

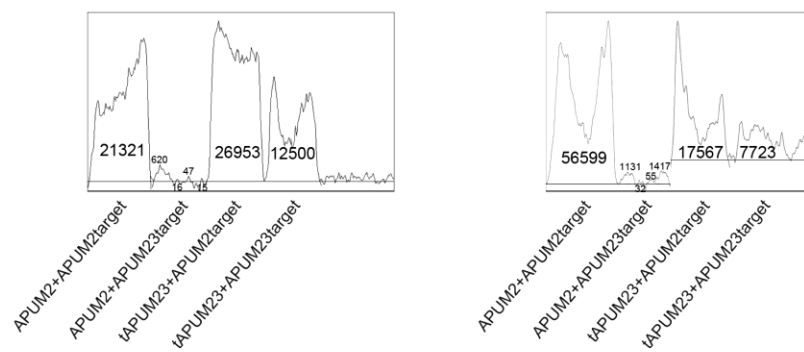
Supplementary Figure S1: Expression level of RFP-APUM targets and APUM-PIN proteins. Figure 3.4 shows one of the five replicates of Western blotting. Here shows the other four Western blots and their analysis. Each figure in this supplementary figure shows one replicate of the Western blots. A) The expression level of RFP-APUM targets (left) and APUM-PIN proteins (right) analyzed by western blotting. RFP-APUM targets were analyzed with RFP antibody (1:1000). APUM-PIN proteins were analyzed with Flag antibody (1:1000). APUM-PIN proteins, including APUM2-PIN and APUM23-PIN, were denoted by an asterisk (*) on the blot. B) Analysis of the expression level of RFP-APUM targets and APUM-PIN proteins. Peaks are co-related to the lanes shown on the western blots. The expression level are shown as the area of the curve.



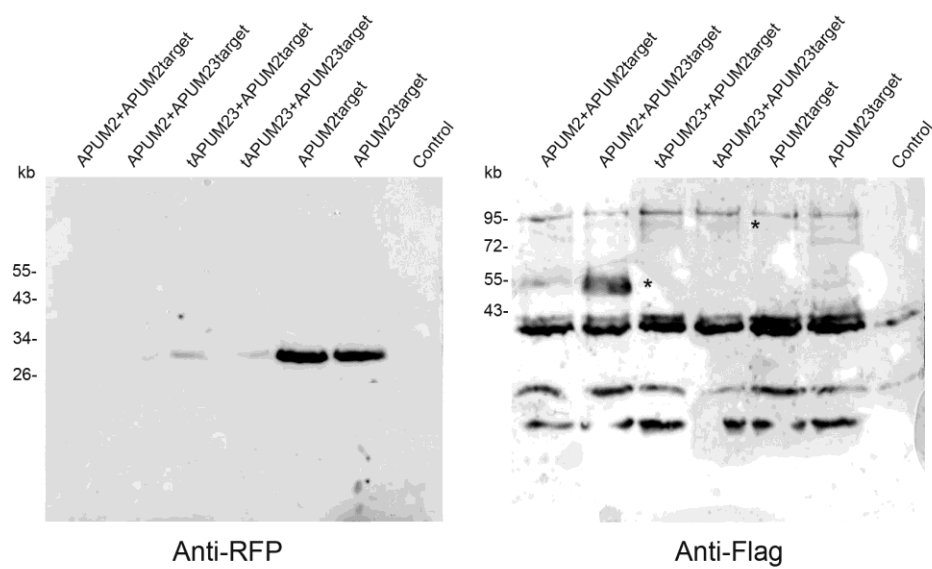
A



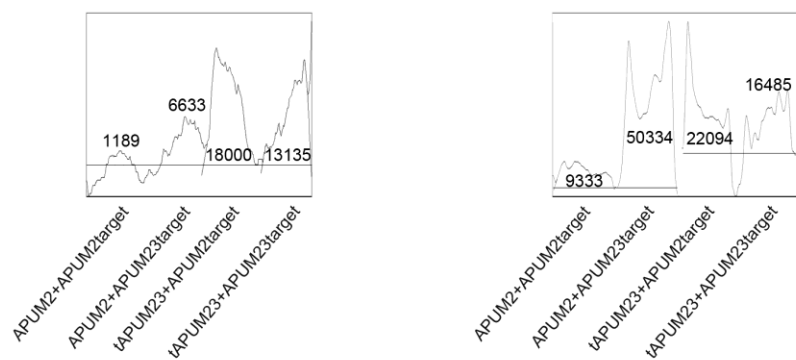
B

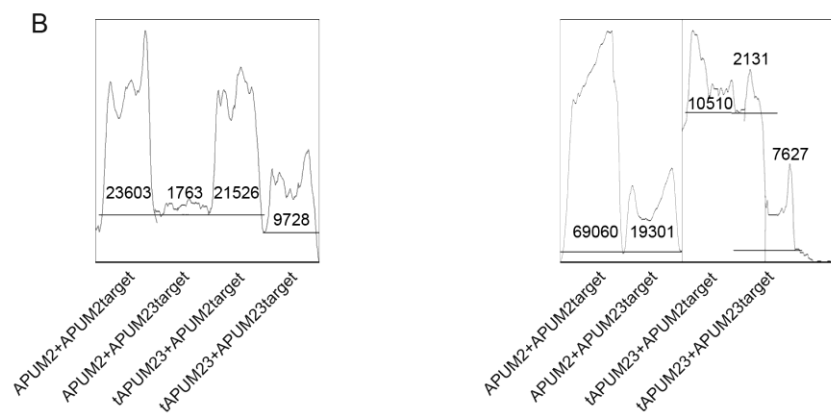
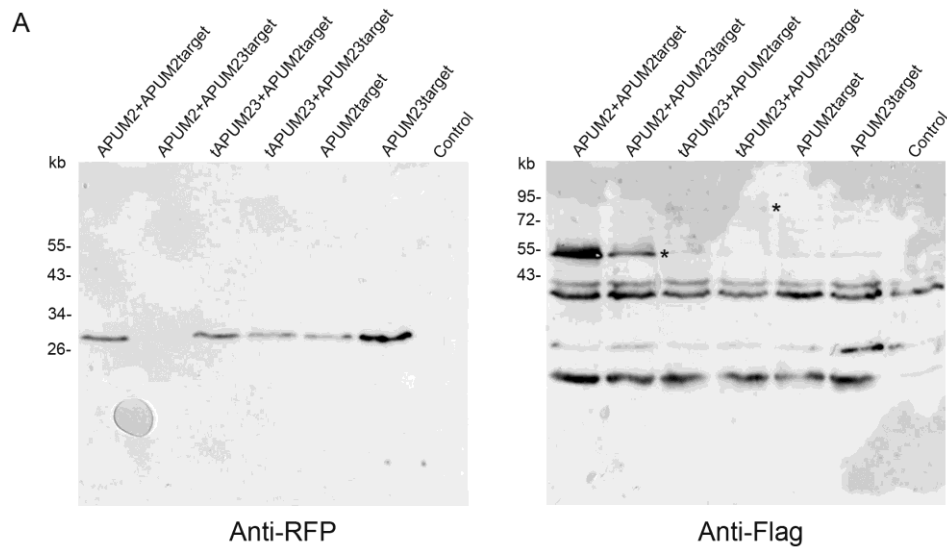


A



B





APPENDIX 2

License Agreement for Figure 1.1

© 2015 Romanowski and Yanovsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

License Agreement for Figure 1.2

License Details

This Agreement between University of Calgary -- Xin Wen ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4520350269352
License date	Feb 01, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Structural & Molecular Biology
Licensed Content Title	Over the rainbow to translational control
Licensed Content Author	Elizabeth R. Gavis
Licensed Content Date	May 1, 2001
Licensed Content Volume	8
Licensed Content Issue	5
Requestor type	academic/university or research institute
Format	print and electronic
Type of Use	Thesis/Dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	No
Circulation/distribution	<501
Author of this Springer Nature content	No
Title	Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology
Institution name	University of Calgary
Expected presentation date	Feb 2019
Portions	Figure 2
Requestor Location	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4 Canada Attn: Xin Wen

Billing Type

Invoice

Billing Address

University of Calgary
2500 University Dr NW

Calgary, AB T2N 1N4
Canada

Attn: Xin Wen

Total

0.00 USD

License Agreement for Figure 1.3

© Tam et al; licensee BioMed Central Ltd. 2010

This article is published under license to BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.


License Agreement for Figure 1.4



Note: Copyright.com supplies permissions but not the copyrighted content itself.

1 2 3
PAYMENT REVIEW **CONFIRMATION**

Step 3: Order Confirmation

 [Print order information:](#)
includes order confirmation,
terms and conditions, and
citation information

Thank you for your order! A confirmation for your order will be sent to your account email address. If you have questions about your order, you can call us 24 hrs/day, M-F at +1.855.239.3415 Toll Free, or write to us at info@copyright.com. This is not an invoice.

Confirmation Number: 11787889

Order Date: 02/01/2019


If you paid by credit card, your order will be finalized and your card will be charged within 24 hours. If you choose to be invoiced, you can change or cancel your order until the invoice is generated.

Payment Information

Xin Wen
University of Calgary
xin.wen2@ucalgary.ca
+1 (587) 889-5253
Payment Method: n/a

Order Details

[Journal of biological chemistry](#)

- **Order detail ID:** 71790473
- **Order License Id:** 4520341222806
- **ISSN:** 1083-351X
- **Publication Type:** e-Journal
- **Volume:**
- **Issue:**
- **Start page:**
- **Publisher:** AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BI
- **Author/Editor:** AMERICAN SOCIETY FOR BIOCHEMISTRY & MOLECULAR BIOL
- **Permission Status:**  **Granted**
- **Permission type:** Republish or display content

- **Type of use:** Thesis/Dissertation

-

-

Requestor type	Academic institution
Format	Print, Electronic
Portion	chart/graph/table/figure
Number of charts/graphs/tables/figures	1
The requesting person/organization	University of Calgary
Title or numeric reference of the portion(s)	Figure 4
Title of the article or chapter the portion is from	A Nucleolar PUF RNA-binding Protein with Specificity for a Unique RNA Sequence
Editor of portion(s)	N/A
Author of portion(s)	Chi Zhang and Douglas G. Muench
Volume of serial or monograph	290
Issue, if republishing an article from a serial	50
Page range of portion	30108-30118
Publication date of portion	Dec 11, 2015
Rights for	Main product
Duration of use	Life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Canada
In the following language(s)	Original language of publication
With incidental promotional use	no
Lifetime unit quantity of new product	Up to 499
Title	Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology
Institution name	University of Calgary
Expected presentation date	Feb 2019
Note: This item will be invoiced or charged separately through CCC's RightsLink service. More info \$ 0.00	
This is not an invoice.	
Total order items: 1	
Order Total: 0.00 USD	

License Agreement for Figure 1.5

License Details

This Agreement between University of Calgary -- Xin Wen ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	4520350159451
License date	Feb 01, 2019
Licensed Content Publisher	Oxford University Press
Licensed Content Publication	Nucleic Acids Research
Licensed Content Title	Structural basis for the specific recognition of 18S rRNA by APUM23
Licensed Content Author	Bao, Hongyu; Wang, Na
Licensed Content Date	Oct 3, 2017
Type of Use	Thesis/Dissertation
Format	print and electronic
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology
Publisher of your work	University of Calgary
Expected publication date	Feb 2019
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Title	Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology
Institution name	University of Calgary
Expected presentation date	Feb 2019
Portions	Figure 2
Requestor Location	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4 Canada Attn: Xin Wen

Publisher Tax ID	GB125506730
Billing Type	Invoice
Billing Address	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4 Canada Attn: Xin Wen

License Agreement for Figure 1.6

License Details

This Agreement between University of Calgary -- Xin Wen ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4520350501124
License date	Feb 01, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Structural & Molecular Biology
Licensed Content Title	A protein-RNA specificity code enables targeted activation of an endogenous human transcript
Licensed Content Author	Zachary T Campbell, Cary T Valley, Marvin Wickens
Licensed Content Date	Jun 29, 2014
Licensed Content Volume	21
Licensed Content Issue	8
Requestor type	academic/university or research institute
Format	print and electronic
Type of Use	Thesis/Dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	No
Will you be translating?	No
Circulation/distribution	<501
Author of this Springer Nature content	No
Title	Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology
Institution name	University of Calgary
Expected presentation date	Feb 2019
Portions	Figure 1b, Figure 2d
Requestor Location	University of Calgary

	2500 University Dr NW
	Calgary, AB T2N 1N4
	Canada
	Attn: Xin Wen
Billing Type	Invoice
Billing Address	University of Calgary
	2500 University Dr NW
	Calgary, AB T2N 1N4
	Canada
	Attn: Xin Wen
Total	0.0 USD

License Agreement for Figure 1.7

License Details

This Agreement between University of Calgary -- Xin Wen ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4520350679623
License date	Feb 01, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Communications
Licensed Content Title	Engineering RNA endonucleases with customized sequence specificities
Licensed Content Author	Rajarshi Choudhury, Yihsuan S. Tsai, Daniel Dominguez, Yang Wang, Zefeng Wang
Licensed Content Date	Oct 23, 2012
Licensed Content Volume	3
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	No
Will you be translating?	No
Circulation/distribution	<501
Author of this Springer Nature content	No
Title	Utilization of Plant Puf RNA-binding Proteins to Manipulate Endogenous mRNA Physiology
Institution name	University of Calgary
Expected presentation date	Feb 2019
Portions	Figure 1
Requestor Location	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4

Canada

Attn: Xin Wen

Invoice

University of Calgary
2500 University Dr NW

Calgary, AB T2N 1N4
Canada

Attn: Xin Wen

Billing Type

Billing Address

Total

0.0 USD