2016

Identification and characterization of RNA target sequences of plant RNA-binding proteins

Zhang, Chi

UNIVERSITY OF CALGARY

Identification and characterization of RNA target sequences of plant RNA-binding proteins

by

Chi Zhang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

MARCH, 2016

Abstract

Eukaryotic cells express a variety of RNA-binding proteins that have important roles in post-transcriptional control of gene expression, including RNA maturation, transport, stability/decay and translation. Here, *in vitro* RNA target selection was used in an effort to identify the consensus RNA target sequences for members of two groups of plant RNA-binding proteins: the PUF family of RNA-binding proteins, and two peroxisomal matrix proteins that have previously been shown to bind to RNA.

Plant genomes encode large families of PUF proteins and these show variability in their predicted Puf repeat number, organization, and amino acid sequence. RNA target identification using *in vitro* selection identified high confidence data for four Arabidopsis PUF (APUM) proteins. Three of these proteins bound to classical PUF RNA target sequences. However one, the nucleolar targeted APUM23, bound to a novel PUF RNA binding sequence. The APUM23 binding sequence is ten nucleotides in length, containing a centrally located UUGA core element, and a preferred cytosine at nucleotide position 8. These RNA sequence characteristics differ from those of other PUF proteins, as all natural PUFs studied to date bind to RNAs that contain a conserved UGU sequence at their 5' end and lack specificity for cytosine. The preferred 10-nucleotide sequence bound by APUM23 is present within the 18S rRNA sequence, supporting the known role of APUM23 in 18S rRNA maturation.

The RNA *in vitro* selection results for the peroxisomal matrix proteins malate dehydrogenase (pMDH) and multifunctional protein (pMFP) demonstrated that pMDH (but not pMFP) possesses sequence-specific RNA binding activity. pMDH bound with

high affinity to two RNA consensus sequences, one that consisted of a triple-repeat of a four nucleotide sequence, and the other sequence was a nine-nucleotide palindrome.

This research provides insight into the novelty of plant RNA-binding proteins and their target RNA sequences, and sheds light on the role of these two classes of proteins in post-transcriptional control of gene expression. It reveals that APUM23 could provide an advanced structural backbone for Puf repeat engineering and target-specific regulation of cellular RNAs. This work also builds upon existing evidence for authentic RNA binding activities for some metabolic enzymes.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, Dr. Doug Muench, for the continuous and immense support of my research and studies through my program. For his patience, consideration and care for me from both work and life activities. His guidance, communication and encouragement helped me in in the research and writing components of this thesis, especially during times when I felt lost and discouraged. I was lucky to have his help throughout the project.

Besides my supervisor, I would like to thank the rest of my thesis committee: Dr. Dave Hansen and Dr. Marcus Samuel for their insightful advice and stimulating discussions. I am grateful to Dr. Albrecht von Arnim (University of Tennessee) and Dr. James McGhee for their critical questions and comments as external examiners, which enhanced the quality of the thesis.

My sincere thanks also goes to Dr. Isabelle Barrette-Ng and Dr. Kenneth Ng, who provided me with guidance and resources in my protein preparation and crystallization work. Also, there is a long list of names of individuals who gave access to the laboratory and research facilities that were required for this work. Without their support, it would not be possible to conduct this research.

I thank my fellow labmates in Muench lab for keeping an organized and joyful work space, and for all the fun we have had in these years.

Last but not the least, I would like to thank my parents, as well as Lin and her family for providing unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them.

Table of Contents

Abstract ii
Acknowledgements iv
Table of Contents v
List of Tables vii
List of Figures viii
List of Symbols, Abbreviations and Nomenclature x

CHAPTER 1 – INTRODUCTION 1
1.1 General Introduction 1
1.2 Post-transcriptional regulation and regulatory elements 3
1.2.1 Regulatory cis-elements in the 5' UTR 3
1.2.2 Regulatory cis-elements in the 3' UTR 5
1.2.3 Regulation by sequence properties presented in the coding region 6
1.2.4 Cis-elements involved in RNA decay 7
1.3 The PUF family of RNA-binding proteins 9
1.3.1 PUF proteins and their RNA-binding features 9
1.3.2 PUF roles in translational repression, RNA decay and mRNA localization 15
1.3.3 Plant PUF proteins and their known functions 16
1.3.4 Practices and perspectives on engineering Puf repeats to achieve sequence-specific regulation of target RNAs 22
1.4 Peroxisomal matrix proteins and their possible role in post-transcriptional control 25
1.4.1 Plant peroxisomes and their biogenesis 25
1.4.2 The RNA- and microtubule- binding activity of peroxisomal multifunctional protein and peroxisomal malate dehydrogenase 26
1.5 Goals of this thesis 28

CHAPTER 2 - MATERIALS AND METHODS 30
2.1 Plant Growth Conditions 30
2.2 Molecular Cloning 30
2.2.1 Arabidopsis genomic DNA extraction 32
2.2.2 Arabidopsis total RNA extraction and reverse transcription 33
2.2.3 PCR and molecular cloning 34
2.3 Protein expression and purification 34
2.4 SELEX 36
2.5 EMSA 39
2.6 Structural Modeling of APUM23 40
2.7 Plant transformation 41
2.8 Agroinfiltration and transient expression 41
2.8.1 Agroinfiltration using Arabidopsis seedlings (AGRObest) 41
2.8.2 Agroinfiltration using Nicotiana benthamiana plants 42
2.8.3 Transient expression using biolistic particle bombardment 43
2.9 Microscopic analysis 43
2.10 Western blotting 44

2.11 Immunoprecipitation with UV or formaldehyde crosslinking.....45

CHAPTER 3 - DETERMINATION OF THE RNA TARGET SEQUENCES OF ARABIDOPSIS PUF PROTEINS.....47
3.1 Introduction.....47
3.2 Results.....50
3.2.1 Prediction of Puf repeat number and location in selected APUM Proteins.....50
3.2.2 SELEX analysis identified high confidence consensus RNA binding sequences for four APUM proteins.....59
3.2.3 Validation of the APUM2, 6, 12 and 23 RNA consensus sequences by EMSA.....68
3.2.4 Motif swapping of APUM23 Puf repeats to alter binding specificity.....74
3.2.5 Analysis of the APUM13 and 18 SELEX data.....81
3.3 Discussion.....82

CHAPTER 4 - PEROXISOMAL MDH BINDS TO RNA IN A SEQUENCE SPECIFIC MANNER.....89
4.1 Introduction.....89
4.2 Results.....90
4.2.1 SELEX identifies consensus RNA binding sequences for pMDH.....90
4.2.2 pMDH RNA mobility shift assays.....97
4.2.3 Identification of Arabidopsis mRNAs that contain the pMDH consensus sequences.....97
4.3 Discussion.....99

CHAPTER 5 - SUMMARY AND FUTURE DIRECTIONS.....105
5.1 Summary.....105
5.2 Future Directions.....109

REFERENCES.....113

APPENDIX 1 - CLONING STRATEGY, PRIMER, VECTOR CHOICE AND PURPOSE.....127

APPENDIX 2 – RAW SEQUENCES OF SELEX PRODUCTS.....133

APPENDIX 3 - SUMMARY OF SOME OF THE TRANSGENIC APPROACHES ATTEMPTED IN ORDER TO IDENTIFY *IN VIVO* TARGETS OF APUM23.....144

List of Tables

Table 2.1 Oligonucleotide primers used for APUM-HD and yeast Nop9 coding sequence amplification and cloning. 31

Table 3.1 Actual and relative apparent dissociation constants (K_d) of APUM2, APUM6 and APUM12 when bound to wild-type and nucleotide substituted RNA 69

Table 3.2 Actual and relative apparent dissociation constants of wild-type APUM23 protein when bound to wild-type and nucleotide substituted RNA 73

Table 3.3 Actual and relative apparent dissociation constants of mutant APUM23 proteins when bound to wild-type and nucleotide substituted RNA 78

Table 3.4 Actual and relative apparent dissociation constants of mutant APUM23 proteins when bound to wild-type and nucleotide substituted RNA 80

Table 4.1 Raw data from the pMFP SELEX experiment showing the nucleotide sequences of individually cloned SELEX products. 93

Table 4.2 Raw data from the pMDH SELEX experiment showing the nucleotide sequences of individually cloned SELEX products. 95

Table 4.3 Arabidopsis thaliana transcripts containing the SELEX-derived RNA consensus sequences of pMDH. 100

List of Figures

Figure 1.1 Ribbon model of the PUM-HD and single Puf repeat from Drosophila PUMILIO 10

Figure 1.2 Structures of the PUF-RNA complex. 14

Figure 1.3 The primary structure of plant PUF proteins 18

Figure 1.4 The process of ribosomal RNA maturation. 21

Figure 1.5 Working model for peroxisomal matrix protein synthesis and import 27

Figure 3.1 Diagrammatic summary of the SELEX procedure. 49

Figure 3.2 Predicted location and sequence of TRMs from the PUF proteins analyzed by SELEX. 51

Figure 3.3 Models of APUM23 referenced against the structure of human PUM1 54

Figure 3.4 The five-amino acid sequence motifs from each of the predicted APUM23 Puf repeats. 56

Figure 3.5 Sequence alignments of the APUM23 orthologs from representative species. 58

Figure 3.6 RNA enrichment profiles from SELEX experiments. 61

Figure 3.7 SELEX consensus RNA target sequences APUM2, 6, 12 and 23. 63

Figure 3.8 Diagrammatic representation of the APUM23 consensus sequence location in the 18S ribosomal RNA sequence. 66

Figure 3.9 Five-amino acid motif in Nop9 Puf repeats, representative EMSA and SELEX enrichment profile 67

Figure 3.10 RNA binding profiles for wild-type APUM proteins to wild-type (WT) and nucleotide-substituted RNAs 71

Figure 3.11 RNA binding analysis for wild-type and mutated APUM23 bound to wild-type and nucleotide-substituted RNAs 77

Figure 3.12 SELEX consensus RNA target sequences for APUM13 and 18 83

Figure 4.1 RNA enrichment profiles from the pMDH and pMFP SELEX experiments 92

Figure 4.2 pMDH consensus RNA binding sequence logo graphs 96

Figure 4.3 pMDH binds with high affinity to its consensus RNA binding sequences. 98

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
3D	three-dimensional
5’ ETS	5’ external transcribed sequence
aa	amino acid(s)
ADP	adenosine diphosphate
AtPRMT	Arabidopsis protein arginine methyltransferase
bp	base pair
BSA	bovine serum albumin
°C	degree Celsius
µCi	microcurie(s)
CTP	cytidine triphosphate
Da	dalton
kDa	kilodalton
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
cDNA	complementary DNA
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced GFP
EGTA	ethyleneglycol-bis(β-aminoethyl ether)-N,N’-tetraacetic acid
EMSA	electrophoretic mobility shift assay
ER	endoplasmic reticulum
F-actin	filamentous actin
FPLC	fast performance liquid chromatography
g	gram
g	unit of gravity
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GST	glutathione S-transferase
GTP	guanosine triphosphate
h	hour
HEPES	N-2-hydroxyethylpiperazine-N’-2-ethane sulfonic acid
HRP	horseradish peroxidase
IPTG	isopropyl-β-D-thiogalactoside
ITS1	internal transcribed sequence 1
kb	kilobase(s)
L	liter(s)
µL	microliter(s)

mL	milliliter(s)
M	molar
mM	milimolar
μM	micromolar
nM	nanomolar
pM	picomolar
MEME	Multiple Em for Motif Elicitation
MES	2-(N-morpholino)ethane sulfonic acid
min	minute
mol	mole(s)
MW	molecular weight
NAD^+	nicotinamide adenine dinucleotide
NADH	NAD^+ reduced
NP-40	Nonidet P-40
NRE	Nanos response element
nt	nucleotide
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
pMDH	peroxisomal malate dehydrogenase
pMFP	peroxisomal multifunctional protein
PMSF	phenylmethylsulfonyl fluoride
PUF	Pumilio and FBF
RNA	ribonucleic acid
mRNA	messenger RNA
miRNA	micro RNA
rRNA	ribosomal RNA
RNAi	RNA interference
RNase	ribonuclease
RNP	ribonucleoprotein
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcription PCR
s	second(s)
SDS	sodium dodecyl sulfate
SELEX	Systematic evolution of ligands by exponential enrichment
SEM	standard error of the mean
TCA	trichloroacetic acid
Tris	tris(hydroxymethyl)aminomethane
U	unit
UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet

V	volt
v	volume
w	weight

CHAPTER 1 – INTRODUCTION

1.1 General Introduction

Eukaryotic cells utilize multi-layered regulatory mechanisms to control gene expression. This ensures that genes are expressed with high fidelity in a spatial and temporal manner. Transcriptional gene regulatory mechanisms are well known and involve a wide range of transcription factors. Although the transcriptional on-and-off switch is a universal control mechanism for regulating gene expression, it can be limited in its ability to rapidly respond to cellular cues or to satisfy local demand at the subcellular level. Gene regulation at the post-transcriptional level provides an additional layer of control, and includes mechanisms that are involved in mRNA stability, localization and translational control. Central to post-transcriptional regulation are the activities of RNA-binding proteins. As mRNAs are being synthesized in the nucleus, the nascent RNA strand is immediately bound by numerous RNA-binding proteins (Singh et al. 2015). These RNA-protein interactions are required initially for RNA splicing, maturation and quality control. When mature mRNA exits the nucleus as a component of a ribonucleoprotein complex (RNP), the RNA-binding protein complement of this complex is remodeled and provides a specific signature to the mRNA to determine whether an mRNA is to be translated, localized or degraded. The different functional domains of an mRNA (the 5'UTR, coding region and 3'UTR) are bound by RNA-binding proteins that are components of regulatory RNPs (Singh et al. 2015). These RNPs are dynamic, and are frequently reassembling to meet the needs of the specific mRNA.

Eukaryotic genomes encode hundreds of RNA-binding proteins, each of which contains one or more RNA-binding domains. Several classes of RNA-binding domain

exist, including the RNA-recognition motif (RRM), the K-homology (KH) domain, the pentatricopeptide repeat (PPR) and the Pumilio Homology Domain (PUM-HD)(Kishore et al. 2010; Ambrosone et al. 2012; Gerstberger et al. 2014). However, there are numerous proteins that possess RNA binding activity but contain no recognizable RNA-binding domain. Two recent studies used *in vivo* UV-crosslinking and stringent washing conditions to identify authentic mammalian RNA-binding proteins (Baltz et al. 2012; Castello et al. 2012). These research groups identified numerous classes of RNA-binding proteins, including those that function in intermediary metabolism and lack any known RNA-binding domains. The so-called "moonlighting" enzymes bind to their cognate RNAs using largely uncharacterized mechanisms.

Understanding RNA-protein interactions is central to the study of post-transcriptional control. The research presented in this thesis is aimed at providing insight into the RNA binding specificities of two groups of RNA-binding proteins in plant cells. These are members of the PUF RNA-binding protein family, as well as peroxisomal matrix proteins that appear to also possess RNA binding activity in the cytosol. PUF proteins are found broadly among eukaryotic organisms and are known to have important roles in regulating gene expression at the post-transcriptional level. The literature related to plant PUFs is not extensive. Two peroxisomal matrix proteins, malate dehydrogenase (pMDH) and the multifunctional protein (pMFP), are thought to possess RNA-binding activity in the cytosol prior to their import into the peroxisome. This cytosolic activity is possible because peroxisomal matrix proteins are fully folded and sometime oligomerized in the cytosol prior to import. Similar to plant PUFs, little is known about the RNA binding activities of pMDH and pMFP. The literature review presented in this chapter will be

focused on post-transcriptional regulation mechanisms as they are related to PUF and peroxisomal matrix protein RNA binding activities.

1.2 Post-transcriptional regulation and regulatory elements

RNA binding proteins are components of the parts list that make up RNPs in the cell, and recognize *cis*-elements on mRNAs that lie in the three main regions of the mRNA; the 3'UTR, coding region and 5'UTR. These elements can be common to most mRNAs (e.g., the m^7G cap and poly(A) tail), or they can be less frequently represented elements that are bound by sequence-specific RNA-binding proteins. In contrast, some RNA-binding proteins bind in a sequence independent manner as a consequence of a processing action (such as splicing), or due to secondary or tertiary structures in the RNA. A short summary of RNA sequence elements and the proteins they interact with is presented in this section. A recent review provides a more in depth analysis of these interactions (Singh et al. 2015).

1.2.1 Regulatory cis-elements in the 5' UTR

Nuclear transcribed mRNAs have the feature of a 5' terminal m^7G cap structure. Along with the poly(A) tail, the m^7G cap has a central role in many post-transcriptional events, such as export from the nucleus, translation, and RNA decay, most of which are mediated by the cap-binding protein. In addition to eIF4E, plants have two other isoforms of cap binding protein, eIFiso4E and nCBP (Browning et al. 1992; Ruud et al. 1998). Combined with two isoforms of the scaffolding protein (eIF4G and eIFiso4G), there are two forms of eIF4F complexes in the plant cytoplasm (Browning 1996). eIFiso4F shows

higher affinity to hypermethylated caps than eIF4F, whereas the eIF4F is more sensitive to secondary structure in the 5’ UTR (Carberry and Goss 1991). Evidence has shown that nuclear eIF4E can facilitate the export of a subset of mRNAs of genes involved in the cell cycle via CRM1-mediated pathway in mammalian cells (Culjkovic et al. 2006; Culjkovic et al. 2007). Moreover, a common ~50 nucleotide, structurally conserved element was found in the 3’UTR of these mRNAs and is referred to as the eIF4E sensitivity element. This element is sufficient to localize eIF4E bound mRNAs to eIF4E nuclear bodies. This eIF4E-mediated preferential mRNA export mechanism was shown to be regulated by several growth-suppressive factors (Culjkovic et al. 2006; Culjkovic et al. 2007). Overexpression of eIF4E leads to the translational activation of a group of functionally related mRNAs that contains a hairpin structure in the 5’ UTR, and this helps with apoptotic rescue (Larsson et al. 2006; Mamane et al. 2007).

The length of the 5’ UTR varies and the 5’ UTR possesses a high degree of complexity in terms of nucleotide sequence and structure. It can be regarded as the most crucial region for translation initiation, since it provides a binding region for a majority of translation initiation factors. Additionally, the 43S pre-initiation complex needs to continue scanning the mRNA from distal 5’end until it comes across an appropriate start codon, which is a process regulated by the degree of RNA secondary structure in this region. Generally speaking, high GC content in the 5’ UTR, secondary structure, and an extended 5’ UTR length do not favor translation initiation, especially under stress (Branco-Price et al. 2005; Kawaguchi and Bailey-Serres 2005). The iron responsive element (IRE) is a well-known example of structured 5’ UTR element that mediates the translation inhibition. The typical IRE is folded in a loop-stem-bulge structure, which can be bound by the iron

responsive protein (IRP) to repress the translation of IRE-containing mRNAs by blocking access of the 43S translation pre-initiation complex (Hentze et al. 1987; Gray and Hentze 1994).

1.2.2 Regulatory cis-elements in the 3' UTR

Regulatory *cis*-acting elements are mostly clustered in the noncoding region of the mRNA due to the lower selection pressure of this region during evolution (Conne et al. 2000). While the 5' UTR is primarily involved in translational regulation, the 3' UTR participates in most of the mRNA metabolism processes. This includes 3' end cleavage and polyadenylation, export, localization and mRNA stability, mainly due to a simple reason that the lengthy 3' UTR can accommodate additional regulatory factors (Moore 2005). The length of the 3' UTR varies in different species and for different genes; a large fraction of genes use alternative cleavage and polyadenylation to generate alternative 3' UTR types (Tian et al. 2005; Jan et al. 2011; Mayr 2015). The direct impact of having isoforms that differ in the length at the 3' UTR is that the mRNA will be subjected to distinct types of regulation events due to the difference on the composition of *cis*-elements included in the 3' UTR region. In addition, the half-life and the translation rate of the isoforms will also show discrepancy (Mayr 2015).

One type of *cis*-acting element existing in the 3' UTR is the sequence that guides subcellular mRNA localization. For example, the 3' UTR of β-actin mRNA contains a 54 nt zipcode sequence that is recognized by the zipcode binding protein (ZBP1). The ZBP1 directs β-actin mRNA to the leading edge of motile fibroblast cells (Ross et al. 1997; Liu 2002; Oleynikov and Singer 2003). PUF family RNA-binding proteins coupled with the

Nanos response element (NRE) can mediate mRNA subcellular localization while translation is temporarily repressed. This type of translational repression has been reported in many organisms and occurs in response to developmental and cell differentiation processes, which will be introduced in section 1.3.2.

Translation initiation relies on mRNA circularization to form the closed-loop structure, which is mediated by the contact of eIF4G and the poly(A) binding protein 1 (PABP1)(Imataka et al. 1998). This cross-talk between 5' UTR and 3' end is often under regulation as well. Examples and details will be covered in the section 1.3.2.

1.2.3 Regulation by sequence properties presented in the coding region

There are examples of post-transcriptional regulation events mediated by *cis*-elements that are present within the coding region that can affect translation. For instance, the constitutive transport element (CTE) is often located in an intron. Some CTE-containing unspliced mRNAs can be bound by TAP1 (an antigen peptide transporter) along with the NTF2-like export factor 1 (NXT1), and exported from the nucleus via a non-CRM1 pathway (Jin et al. 2003). This promotes translation after the mRNA exits from the nucleus. Similar to TAP1, an isoform of Wilms tumour 1 protein also recognizes the CTE, and binds to CTE-containing mRNAs in cytoplasm to enhance their translation (Bor et al. 2006).

The exon-junction complex (EJC) is a large mRNP that preferentially binds to newly spliced transcripts at their exon-exon junction regions in the nucleus (Le Hir et al. 2000). The EJC core components Y14 (which contains an RNA-binding domain) and Magoh remain bound to the mRNA after the export until the first round of translation is

completed (Dostie and Dreyfuss 2002). The partner of Y14/Magoh (PYM) has been shown to positively regulate the translation of spliced mRNAs by recruiting the 40S ribosomal subunit, which suggests a role for the exon junction complex in translational regulation (Diem et al. 2007). The EJC was previously known for its role in nonsense-mediated decay (NMD), which is a quality control mechanism used to eliminate premature termination codon (PTC) containing transcripts (Chang et al. 2007). The pioneer round of translation will remove all EJC components deposited on the transcripts, unless a PTC exists. The PTC triggers the interaction of the NMD factor up-frameshift protein 1 (Upf1) with the EJC component Upf2 and Upf3b, and consequently Upf1 is phosphorylated after being recruited by EJC during the activation of NMD (Chang et al. 2007; Isken et al. 2008).

1.2.4 Cis-elements involved in RNA decay

Decay is also an important chapter in mRNA metabolism, in which a number of RBPs and associated factors, as well as the corresponding *cis*-acting elements, need to be coordinated. The degradation mechanisms used by eukaryotic mRNAs include the deadenylation-dependent mRNA decay, nonsense-mediated decay (NMD) and endonucleolytic degradation.

The addition of the poly(A) tail and the incorporation of the m^7G cap are hallmark of mRNA maturation events. These are also critical events that maintain mRNA stability, because the default mRNA decay process begins with shortening of the poly(A) tail. The exonucleolytic activity is carried out by at least two conserved deadenylation complexes. One of them is the Ccr4-Not complex, which is essential for cytoplasmic mRNA deadenylation in yeast and mammals (Zheng et al. 2008). Within this complex, Ccr4 serves

the catalytic function, and the scaffold protein Not is responsible for the associations with a number of other factors (Vindry et al. 2014). A second major mRNA deadenylation complex is formed by the PAN2 deadenylase and PAN3 auxiliary protein. The direct interaction of PAN3 with the poly(A)-binding protein, PABPC1, recruits the deadenylation complex to transcripts (Funakoshi et al. 2007). Upon initiation of deadenylation, the mRNA is subsequently subjected to the hydrolysis of the 5' cap structure and 5' to 3' degradation by Xrn1 exonuclease, or alternatively to the exosome complex involved in degradation in the 3' to 5' direction (Chen and Shyu 2011). The AU-rich elements (ARE, AUUUA pentamer), which mediate deadenylase recruitment, are one of the most abundantly conserved *cis*-elements found in the 3' UTR (Xie et al. 2005). Among many ARE-binding protein families, the CCCH type zinc finger domain containing RNA-binding protein Tristetraprolin (TTP) and its paralogs were shown to recruit many types of deadenylase via the direct interaction with Not1 (Clement et al. 2011; Sandler et al. 2011). Examples of PUF family protein induced decay of NRE containing mRNA via recruitment of the Ccr4-Pop2-Not complex will be introduced in section 1.3.2.

The degradation of mRNA does not always depend on the initiation of deadenylation. It could be initiated by endonucleolytic cleavage at the 3' UTR (Matsushita et al. 2009) or utilize other mechanisms to degrade the transcripts lacking a poly(A) tail. Histone mRNAs contain a conserved stem-loop in their 3' UTRs, which can be bound by SLBP and UPF1 proteins (Marzluff et al. 2008). The degradation of histone mRNAs is initiated by the oligouridylation at the 3' end, which relies on the SLBP and UPF1 to recruit the terminal uridylyl transferase. This added poly(U) tract is recognized by the heptameric

LSM1-7 complex and PAT-1, resulting in the recruitment of decapping machinery and then 5’ to 3’ degradation (Marzluff et al. 2008; Su et al. 2013).

1.3 The PUF family of RNA-binding proteins

1.3.1 PUF proteins and their RNA-binding features

PUF (Pumilio/FBF) RNA-binding proteins are a group of conserved eukaryotic proteins that typically bind to RNA in a sequence-specific manner and have essential roles in post-transcriptional control of gene expression (Quenault et al. 2011). PUF proteins vary in number among different organisms. Three PUF proteins are expressed in human cells, while twelve are present in *C. elegans*, six in *S. cerevisiae*, and two in Drosophila (Quenault et al. 2011). As a taxonomic group, plants possess the highest number of PUF proteins. Rice and the model plant, *Arabidopsis thaliana* (Arabidopsis), encode up to 19 and 26 PUF proteins, respectively (Francischini and Quaggio 2009; Tam et al. 2010; Abbasi et al. 2011). PUF proteins typically bind to RNA sequences that reside in the 3’ untranslated region (3’ UTR) of the target mRNA, and have a mechanistic role in regulating mRNA stability, translational activation/repression, and localization (Quenault et al. 2011).

The RNA-binding domain of PUF proteins, the highly conserved Pumilio Homology Domain (PUM-HD), is usually 300 to 400 amino acids in length and located in the C-terminal region of the protein. The PUM-HD typically contains eight tandem Puf repeats that fold into a crescent-shaped structure. Each Puf repeat is approximately 36 amino acids in length and folds into three α-helices (Figure 1.1). The second α-helix of each Puf repeat is exposed to the inner concave surface of the PUM-HD, and key amino acids in this helix contact a single RNA base. Thus, a PUM-HD that contains eight Puf

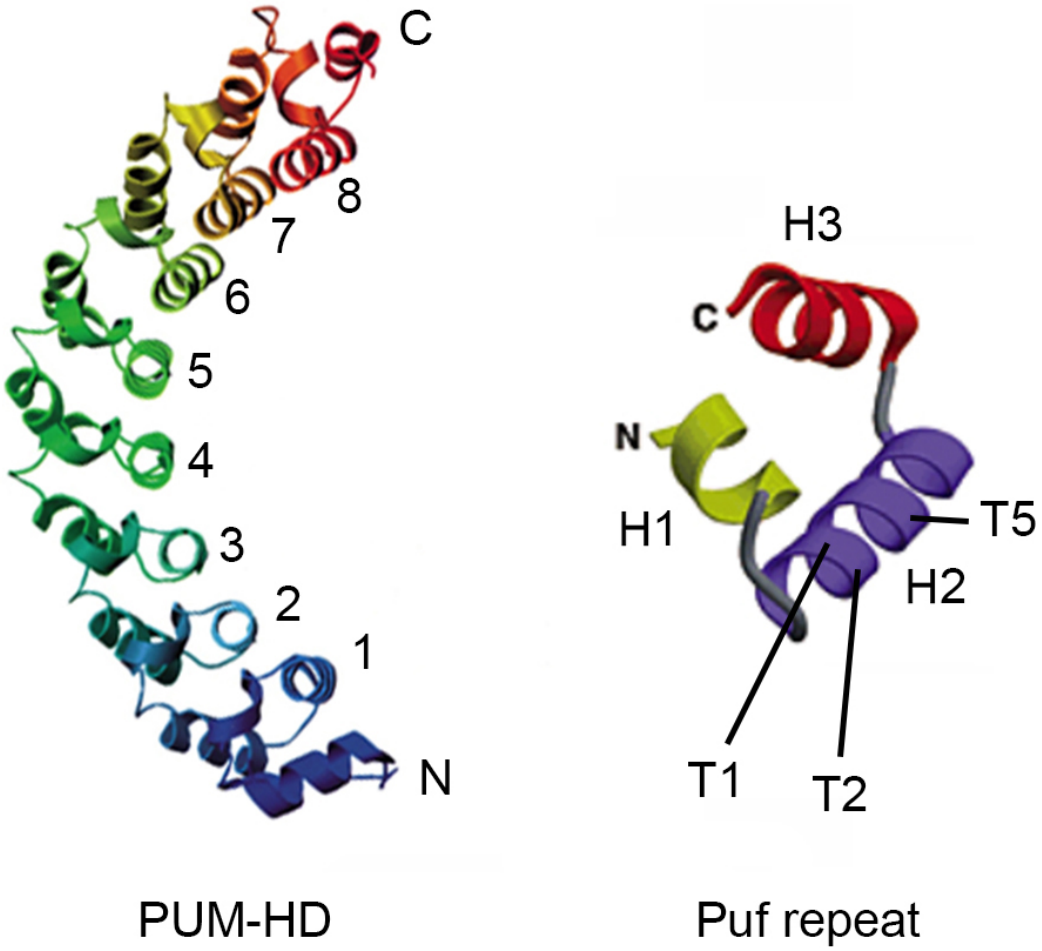


Figure 1.1 Ribbon model of the PUM-HD and single Puf repeat from Drosophila PUMILIO. Each of the eight Puf repeats is numbered and marked with a different color (left panel) (Modified from (Edwards et al. 2001). A single Puf repeat composed of three α-helices (H1, H2, H3; right panel), the second of which (H2) binds to a single nucleotide on the concave surface of the protein faces the inner groove. T1, T2 and T5 identify the location of the three amino acids (1, 2 and 5) that comprise the TRM in the five-amino acid motif located on H2. (Modified from Gavis 2001)

repeats binds to an RNA target consisting of eight nucleotides. This one-nucleotide:one-repeat binding pattern possessed by PUF proteins distinguishes them from other cytosolic RNA-binding proteins. The binding of the PUM-HD to RNAs is anti-parallel, with RNA bases 1 through 8 sequentially interacting with Puf repeats 8 through 1. There are, however, some atypical interactions between PUF proteins and their RNA targets. For example, the eight Puf repeats of *C. elegans* FBF-2 and yeast Puf4 bind to nine-nucleotide RNA targets, and do so by flipping away one nucleotide from the RNA binding surface (Miller et al. 2008). Yeast Puf3 possesses a binding pocket between Puf repeats 8 and 8' (a pseudo-Puf repeat) that accommodates a cytosine located upstream of the UGUR core at the -2 position (Zhu et al. 2009). All PUF proteins studied to date bind to RNA sequences that contain a UGUR (where R is a purine) at their 5' end. The nucleotide at the fifth position is often degenerate, and the final three nucleotides often demonstrate a preference for AUA (Zamore et al. 1999). Despite the apparent structural conservation of PUF proteins, a divergent PUF protein structure was recently identified. A nucleolar human PUF protein (Puf-A) and its yeast ortholog (Puf6) were shown to contain eleven Puf repeats arranged in an atypical manner and folded into a L-shaped structure rather than a usual crescent shape observed from other PUF proteins (Thomson et al. 2007; Abbasi et al. 2010; Qiu et al. 2014).

Within the second α-helix of Puf proteins are three key amino acids (the tripartite recognition motif, TRM) that participate in nucleotide interactions (Figure 1.1). These three amino acids are located at positions 1, 2 and 5 within a five-amino acid motif (positions 3 and 4 consist of hydrophobic amino acid spacers)(Campbell et al. 2014). The side chains of amino acids 1 and 5 contact with the Watson-Crick edge or the Hoogsteen

edge of the base through hydrogen bonding and/or van der Waals forces, whereas the side chain of amino acid at position 2 inserts into two adjacent bases to provide a stacking interaction (Figure 1.2). The nucleotide binding specificity of Puf repeats are largely determined by the amino acids at position 1 and 5. In the original RNA recognition code for Puf repeats, cysteine and glutamine at positions 1 and 5 bind adenine, asparagine and glutamine bind to uracil, and serine and glutamate to guanine (Wang et al. 2002). Puf repeats that naturally bind to cytosine have not yet been identified. However, a cytosine-recognizing TRM has been engineered by amino acid substitutions on a native Puf repeat backbone using a yeast three-hybrid screen (Dong et al. 2011; Filipovska et al. 2011). The specificity of TRMs is not always conclusive, as flexibility and promiscuity have been reported recently. A detailed study of PUF binding characteristics has provided an expanded repertoire of binding specificities that will be valuable for future engineering of PUF proteins customized for RNA target-specific regulation (Campbell et al. 2014).

Individual PUF proteins are known to bind to numerous RNA targets in the cell. An early study that used co-immunoprecipitation of PUF proteins and their RNA targets followed by microarray analysis (RIP-Chip analysis), demonstrated that individual PUFs bound to distinct mRNA populations that encode proteins with common features are associated with function or subcellular localization (Gerber et al. 2004). For example, yeast Puf3 bound to approximately 200 mRNA targets that almost exclusively encoded mitochondrial proteins (Gerber et al. 2004). More recent approaches using *in vivo* UV-crosslinking of PUF proteins to their mRNA targets (crosslinking and immunoprecipitation, CLIP) identified PUF targets with higher confidence and determined their specific binding sites (Hafner et al. 2010). These studies expanded the types and

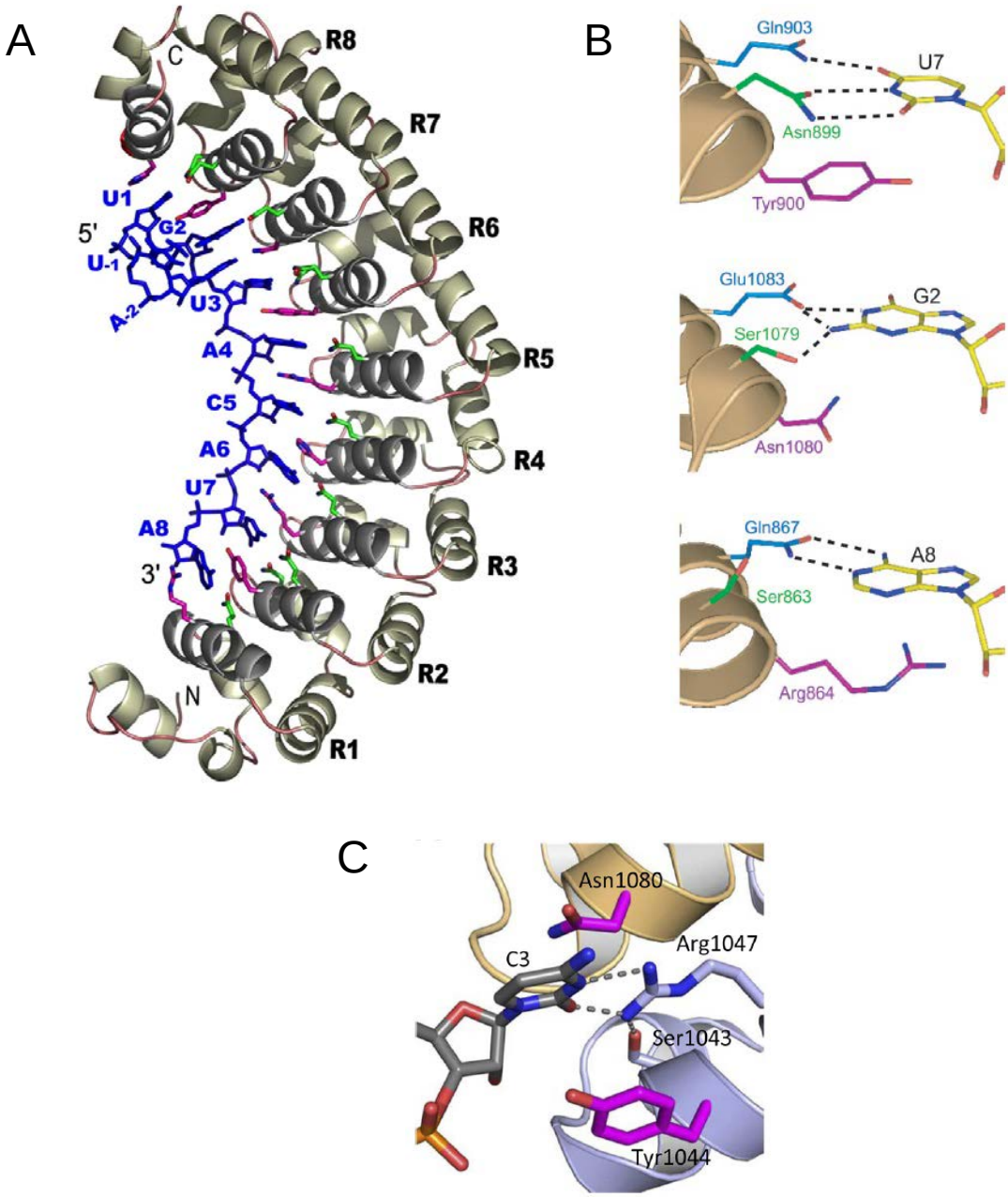

A
C
R8
R7
R6
R5
R4
R3
R2
R1
N
5'
3'
U1
G2
U-1
A-2
U3
A4
C5
A6
U7
A8
B
Gln903
U7
Asn899
Tyr900
Glu1083
G2
Ser1079
Asn1080
Gln867
A8
Ser863
Arg864
C
Asn1080
Arg1047
C3
Ser1043
Tyr1044

Figure 1.2 Structures of the PUF-RNA complex. (A) The structure of the PUM-HD from human PUM1 bound to the Nanos Response Element (NRE, AUUGUACAUA). Protein and RNA are shown in a ribbon and in a stick model, respectively. The RNA is aligned anti-parallel along the concave surface of the PUM-HD. Amino acid side chains that form hydrogen bonds and van der Waals for are shown in green and those that are invoved in base stacking are in magenta. (From Gupta et al. 2008) (B) Interactions of Puf repeat amino acid side chains with a uracil (top), guanine (middle), or adenine (bottom). The bases and interacting amino acid side chains are shown in a stick model (dark blue, nitrogen; red, oxygen; yellow, light blue, green, or purple, carbon). Hydrogen bonds are indicated with dashed lines. (From Cheong & Hall 2006). (C) Interaction of an engineered Puf repeat (SYxxR) with a cytosine. The bases and interacting amino acid side chains are shown as stick models (red, oxygen; blue, nitrogen; orange, phosphorus). Carbon atoms are colored gray in RNA. The amino acid side chains that interact with the edge of the base are colored light blue. The magenta side chains form stacking interactions with the RNA base. Hydrogen bonds are indicated with dashed lines. (From Dong et al. 2011)

number of specific PUF targets, and also provided greater insight into the characteristics of the sequences bound directly by PUF proteins. For instance, CLIP analysis of yeast Puf5 RNA targets demonstrated that Puf5 bound directly to RNA targets consisting of 9 or 10 nucleotides, with a lesser enrichment of 8, 11 and 12 nucleotide long sequences as well (Gerber et al. 2004; Wilinski et al. 2015).

1.3.2 PUF roles in translational repression, RNA decay and mRNA localization

A major regulatory role of PUF proteins is their ability to repress mRNA targets using multiple mechanisms. A classic example is that of Drosophila Pumilio. Pumilio is one of the founding PUF members, and represses the stability and translation of hunchback (*hb*) mRNA at the posterior end of oocytes (Wreden et al. 1997; Chagnovich and Lehmann 2001). The Nanos Response Elements (NRE1 and NRE2) are located within the 3' UTR of the *hb* mRNA and are directly bound by Pumilio. Pumilio is a component of a ternary complex that includes Nanos (Nos) and Brain Tumor (Brat) proteins. The Pumilio/Nos/Brat repression complex acts on *hb* mRNA in two ways. It interacts with the Ccr4-Pop2-Not deadenylase complex to shorten the poly(A) tail and trigger RNA decay, and this also influences translation rate (Wreden et al. 1997). Pumilio also recruits eIF4E-HP, a factor that represses translation by competing with eIF4E (cap binding protein) for the 5' cap on mRNAs, thereby reducing the translation rate (Cho et al. 2006). In addition, Pumilio can repress mRNAs that do not possess a poly(A) tail (Wreden et al. 1997; Chagnovich and Lehmann 2001). Yeast Puf5 provides another example of PUF-mediated mRNA repression. Puf5 was shown to bind directly to the Pop2 subunit of the mRNA deadenylase Ccr4-Pop2-Not complex (Goldstrohm et al. 2006). This Pop2-Puf5 interaction

also recruits decapping activator Dhh1 and decapping enzyme Dcp1 to the mRNA (Maillet and Collart 2002; Goldstrohm et al. 2006; Quenault et al. 2011). This assembly not only causes the repression of translation, but also the hydrolysis of the 5' cap. PUF repression of translation can also involve competitive binding of translation factors at the 5' cap. In Xenopus, repression of cyclin B1 mRNA involves the binding of Pumilio to its 3'UTR where it stabilizes the interaction of cytoplasmic polyadenylation element binding protein (CPEB) on the mRNA. CPEB then recruits a protein called Maskin, which then hinders the interaction between eIF4G and eIF4E, thereby repressing translation (Cao and Richter 2002).

Yeast Puf6 contributes to translational repression upon the asymmetric localization of *ASH1* mRNAs during the transport to yeast bud (Gu et al. 2004; Deng et al. 2008). Puf6 binds to an element located in the 3'UTR of *ASH1* mRNA and inhibits translation by interacting with the translation initiation factor eIF5B/Fun12 (Deng et al. 2008). This inhibition of translation also facilitates the localization of *ASH1* mRNA to the bud. This is because the impaired translation rate relieves competition between the translating ribosome and She2, a subunit of the locasome complex that recognizes the mRNA localization elements on the *ASH1* open reading frame (Paquin and Chartrand 2008; Shen et al. 2009). Once the *ASH1* mRNA arrives at the bud tip, Puf6 is phosphorylated by the CK2 kinase, which removes the repression of *ASH1* due to the release of Puf6 from the mRNA (Deng et al. 2008).

1.3.3 Plant PUF proteins and their known functions

The structure, binding properties, and function of Drosophila, *C. elegans*, yeast and

human PUF proteins have been examined in detail. However, only a small number of studies have been reported on plant PUF proteins. The Arabidopsis PUF family of proteins is extensive, consisting of up to 26 members (Francischini and Quaggio 2009; Abbasi et al. 2010; Tam et al. 2010). These proteins demonstrate considerable variability in their number of Puf repeats, the position of these repeats in the primary sequence, and in the identity of their putative TRMs (Figure 1.3, Tam et al. 2010). Over half of the Arabidopsis PUFs are predicted to possess eight tandemly arranged Puf repeats. Six of these proteins (APUM1 to APUM6) are conserved in the amino acids that comprise the TRM in each Puf repeat, when compared to those in human PUMILIO1 (PUM1). The remaining PUF proteins are predicted to contain three to ten Puf repeats, some of which contain large gaps between individual repeats (Francischini and Quaggio 2009; Abbasi et al. 2010; Tam et al. 2010). *In vivo* expression of several APUM proteins (APUM 7, 8, 9, 10, 12, 14, and 18) as fusions to fluorescent proteins indicates that these proteins localize to dynamic, punctate cytoplasmic structures (Tam et al. 2010). However, APUM23 and APUM24 (orthologs of yeast Nop9 and Puf6, respectively) localized to nucleoli (Abbasi et al. 2010; Tam et al. 2010). APUM23 and APUM24 deviate from the typical tandem eight Puf repeat arrangement in that they are predicted to possess additional repeats. Indeed, the recent crystal structure of Puf-A (the human ortholog of APUM24) revealed an atypical PUF structure that consisted of eleven Puf repeats (Qiu et al. 2014). This represents a newly identified nucleic acid binding fold that binds to RNA and DNA in a non-sequence specific manner. The unusual number and arrangement of predicted Puf repeats in APUM23 suggests that its structure is also atypical for a PUF protein. APUM23 is predicted to possess ten Puf repeats that are spread throughout the protein rather than clustered at the

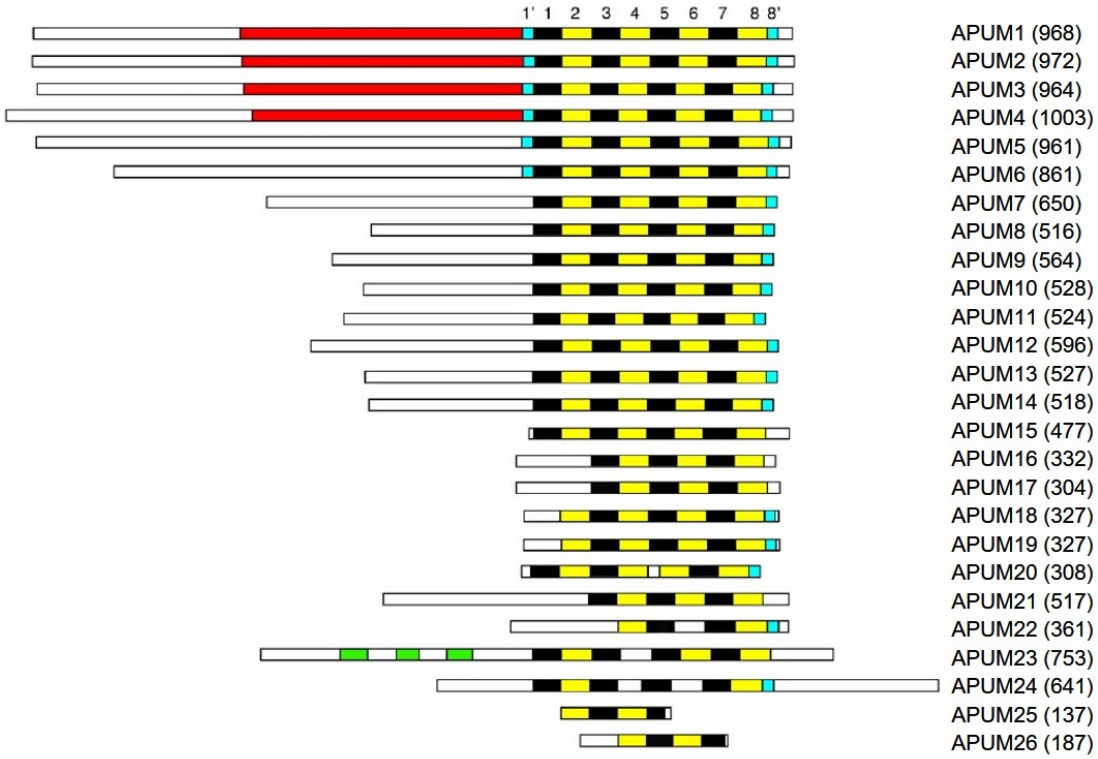


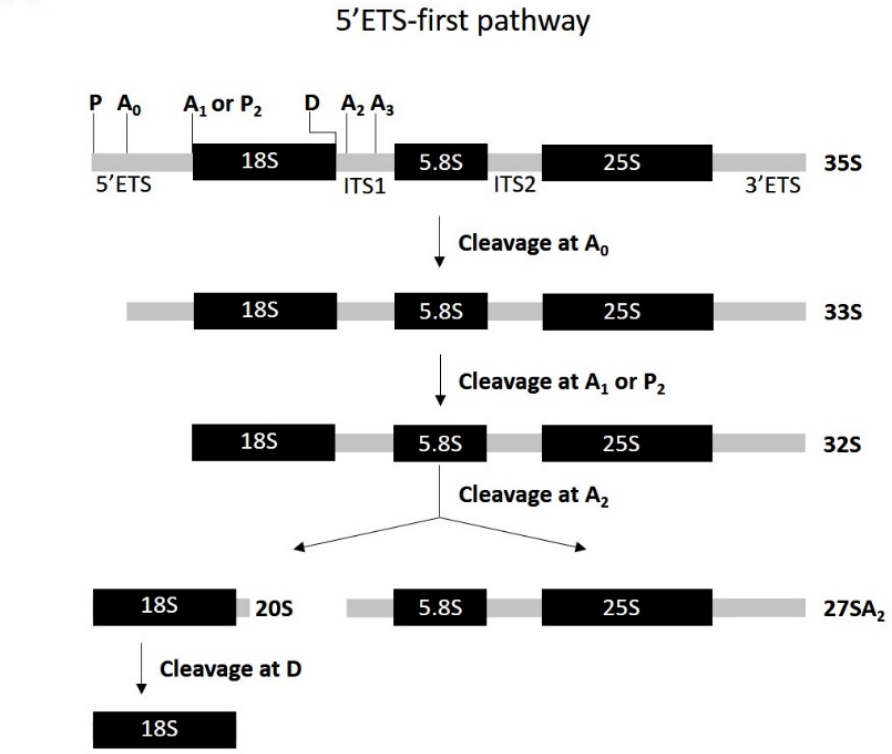
Figure 1.3 The primary structure of plant PUF proteins. The numbered yellow and black rectangles correspond to predicted Puf repeats in the PUM-HD of each protein. The 1' and 8' pseudorepeats are indicated in blue. A nucleic acid binding protein domain (NAPB) is present in four Arabidopsis PUF proteins. Three additional Puf repeats in APUM23 are shown in green. The number of amino acids in each PUF protein is indicated in parentheses. (Modified from Tam et al. 2010)

C-terminal region (Tam et al. 2010).

Endogenous RNA targets of plant PUF proteins have not been conclusively identified. However, APUM1 through APUM6 were shown to bind the typical UGUR-containing Nanos Response Element that is bound by Drosophila PUMILIO (Francischini and Quaggio 2009; Tam et al. 2010). Additionally, the APUM2 RNA consensus target sequence, as determined by a yeast three-hybrid screen, showed the typical eight-nucleotide PUF binding motif with a UGUA core and low specificity at nucleotide position 5 (Francischini and Quaggio 2009). Functional information on four Arabidopsis PUF proteins has been reported. APUM5 was shown to bind directly to cucumber mosaic virus RNA to suppress virus infection in Arabidopsis (Huh et al. 2013). APUM5 also appears to have a more general role in negatively regulating abiotic stress gene expression, and binds *in vitro* to the 3'UTRs of candidate gene mRNAs that contain UGUA sequences (Huh and Paek 2014). APUM9 and APUM11 are involved in reducing seed dormancy, possibly by regulating the translation efficiency of specific stored seed mRNAs upon imbibition (Xiang et al. 2014). APUM23 mutants display shoot and root developmental abnormalities, defects in pre-ribosomal RNA processing, and disrupted auxin homeostasis (Abbasi et al. 2010; Huang et al. 2014).

The rRNA processing role of APUM23 has been well characterized in a study that monitored the progress of pre-ribosomal maturation in an *apum23* mutant (Abbasi et al. 2010). A study on yeast Nop9 (the ortholog of APUM23) found similar defects during the ribosomal RNA processing due to the depletion of Nop9 protein (Thomson et al. 2007). rRNAs are processed in the nucleolus from the 35S pre-rRNA (Figure 1.4). Two alternative ribosomal RNA processing routes exist, one where the 5'ETS is removed first and the other

A


5'ETS-first pathway
P
A0
A1 or P2
D
A2
A3
18S
5.8S
25S
35S
5'ETS
ITS1
ITS2
3'ETS
Cleavage at A0
33S
Cleavage at A1 or P2
32S
Cleavage at A2
20S
27SA2
Cleavage at D

B

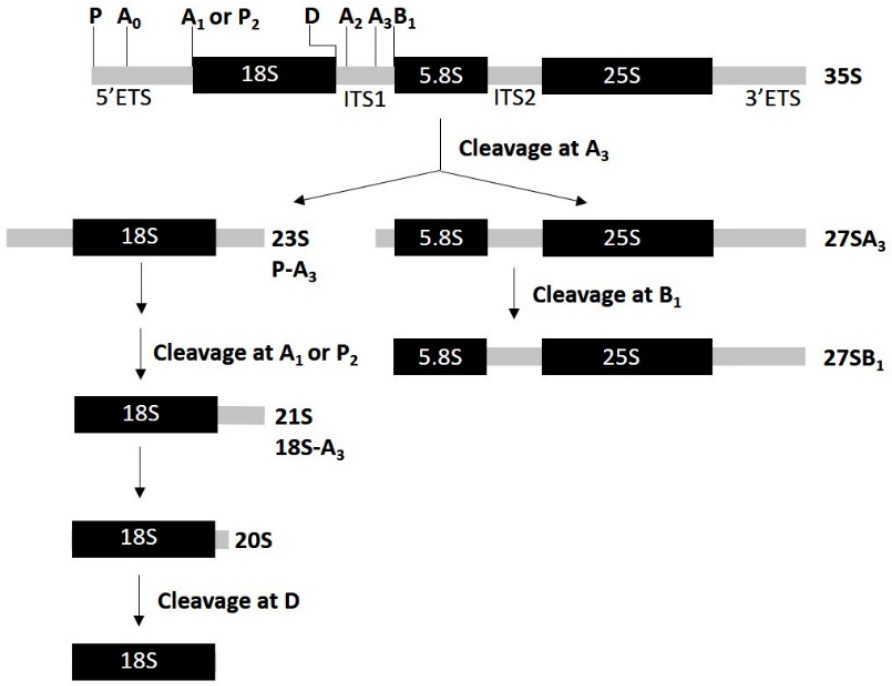

ITS1-first pathway
P
A0
A1 or P2
D
A2
A3
B1
18S
5.8S
25S
35S
5'ETS
ITS1
ITS2
3'ETS
Cleavage at A3
23S
P-A3
27SA3
Cleavage at B1
Cleavage at A1 or P2
27SB1
21S
18S-A3
20S
Cleavage at D

Figure 1.4 The process of ribosomal RNA maturation. (A) 5'ETS-first processing begins with the cleavage at A0 and A1 (or P) sites inside the 5'ETS region. The 18S region separates from 5.8S and 25S as in a 28S precursor later. This co-transcriptional processing is the major route in yeast. (B) ITS1-first processing is initiated by the separation of 18S containing precursor (23S P-A3 fragment) from the remaining portion. This processing route is used by most multicellular organisms. The P and P2 (P' and P1 are not shown) sites are nomenclatures used in plant studies, whereas A0 and A1 are used in yeast and animal studies.

where the ITS1 is cleaved first. These routes are both adopted in eukaryotes but with the bias (Hang et al. 2014; Weis et al. 2015)(Figure 1.4). Plants and mammals usually initiate processing by cleavage at A_3 in the ITS1 to produce the P-A_3 pre-rRNA (Figure 1.4B)(Mullineux and Lafontaine 2012; Hang et al. 2014; Henras et al. 2015). This is in contrast to yeast, which utilizes co-transcriptional processing to remove the 5' ETS from the nascent rRNA strand and yield 33S or 32S pre-rRNA (Figure 1.4A)(Osheim et al. 2004; Koš and Tollervey 2010; Fernández-Pevida et al. 2015). Mutation of the plant Apum23 gene resulted in an accumulation of a 2.6 kb intermediate product (P-A_3 fragment), suggesting that the immediate step to continue the processing of this 2.6 kb pre-rRNA (P-A_3) (which is the trimming of 5' ETS) was inhibited (Figure 1.4B)(Abbasi et al. 2010; Hang et al. 2014). The depletion of Nop9 in yeast led to the accumulation of unprocessed 35S pre-ribosomal RNA at high levels, as well as a 5' ETS-containing intermediate product (23S rRNA), which was the equivalent of the P-A_3 fragment in plants (Thomson et al. 2007; Abbasi et al. 2010). Yeast mainly relies on the 5' ETS-first pathway to process the 35S pre-rRNA (Figure 1.4A), therefore, the accumulation of the 35S pre-rRNA indicates that the removal of the 5' ETS has been abolished (Figure 1.4A). In summary, the accumulation of these precursors (35S, 23S or P-A_3) in both species is caused by the malfunction of the 5' ETS removal. This implies a likely role for APUM23 and Nop9 in 5'ETS processing.

1.3.4 Practices and perspectives on engineering Puf repeats to achieve sequence-specific regulation of target RNAs

Upon determination of the Puf repeat nucleotide recognition scheme and the

principals of the TRM specificity, the engineering of PUF proteins to target desired RNA sequences became an option. The simple shuffling and replacing of TRMs produced an artificial PUF that recognizes two tandem “UGUA” core sequence with higher affinity (Cheong and Hall 2006). The Puf repeat was also found to be stackable. An engineered PUM-HD consisting of more than eight Puf repeats provided expanded and possibly more specific binding specificity (Filipovska et al. 2011). A specific 13-nucleotide sequence appears once in ~70 million RNA bases, which is the size of the transcriptome in higher eukaryotic organisms. This means that an engineered PUF consisting of at least 13 Puf repeats has the potential to target any desired transcript with specificity.

Although an exclusive cytosine-binding Puf repeat has not yet been identified from native PUF proteins, the engineering of a TRM that prefers cytosine (Dong et al. 2011; Filipovska et al. 2011) completes the recognition code of Puf repeats for all four nucleotides. The capacity of Puf repeats to recognize all nucleotides allowed for the engineering of a recombinant PUF-based artificial splicing factor that made alternative splicing more successful and customizable (Yang Wang et al. 2009; Dong et al. 2011). Other similar applications that utilized the engineered PUF RNA recognition domain as a fusion with effector domains were successful as well. A fusion of an engineered PUF to poly(A) polymerase or deadenylase regulates the stability and translation of the desired target (Cooke et al. 2011). PUF fusion to an endonucleolytic domain enables PUF to function as an “RNA restriction enzyme”, which can be applied to gene silencing approaches (Choudhury et al. 2012). Introducing a designed PUF that binds to the 5’ UTR sequence of a target transcript serves as a barricade to the ribosome and represses translation (Cao et al. 2015). In addition to the construction of PUF fusion proteins with

catalytic functions, an engineered PUF was also applied to live-cell *in situ* imaging or RNA (Tilsner et al. 2009).

The specificity of engineered PUF proteins remains a concern, as off-target binding would be problematic (Wang et al. 2013; Abil and Zhao 2015). For example, vast sets of yeast transcripts are naturally under the regulation of Puf3 and Puf5. The promiscuous binding of nucleotides at certain Puf repeat positions creates a challenge for specific RNA binding. The relaxed specificity of some Puf repeats has been reported, as well as the alternative accommodation of sequences when PUF proteins bind to RNA (Yeming Wang et al. 2009; Lu and Hall 2011; Wilinski et al. 2015). Fortunately, studies on the fidelity of TRMs have made progress into the specificity code of PUFs (Campbell et al. 2014; Hall 2014). The binding strength of the TRM seems to be affected by the positional context of the nucleotides in the RNA (Campbell et al. 2014; Hall 2014). This suggests that substitution of only the TRM may not be able to achieve maximum specificity. One obvious solution for enhanced specificity would be achieved by introducing more Puf repeats to the binding domain in order to recognize a longer RNA sequence (Wang et al. 2013; Abil and Zhao 2015). However, this approach is based on the condition that each sequential nucleotide is bound by a Puf repeat, which is not a universal characteristic of native PUFs (Lu and Hall 2011). This may create a need to search a universal binder that maintains the protein architecture and exhibits tolerance to for sequence variation (Wang et al. 2013).

1.4 Peroxisomal matrix proteins and their possible role in post-transcriptional control

1.4.1 Plant peroxisomes and their biogenesis

Peroxisomes are highly dynamic cellular organelles that are bound by a single lipid bilayer. They have a number of functions, including the β-oxidation of fatty acids, the reduction of reactive oxygen species, and the glyoxylate cycle. In plants, additional functions for the peroxisome include the synthesis of a variety of hormones and signaling molecules (Brown and Baker 2003). The general model for peroxisome biogenesis appears to be one of semi-autonomy, where *de novo* biogenesis can occur from specialized regions of the endoplasmic reticulum (ER), and that peroxisomes can also grow and divide (Hu et al. 2012). In plants, a *de novo* pathway has not yet been established, and the ER membrane may serve only to provide selected membrane components to pre-existing peroxisomes. Peroxisomes lack their own genome and, therefore, must import their constituent proteins in a post-translational manner. Matrix proteins are imported from the cytosol in their fully folded - sometimes oligomerized state using pathways that are distinct from those chloroplast and mitochondrial protein import (Cross et al. 2015). Two co-dependent import pathways involve two recyclable peroxisomal matrix import receptors, Pex5 and Pex7. Matrix proteins that possess a C-terminal peroxisomal targeting sequence 1 (PTS1, a tri-peptide) interact with the Pex5 receptor, whereas matrix proteins that possess a PTS2 (an N-terminal nonapeptide that is cleaved after import) interact with the Pex7 receptor (Hu et al. 2012). Plant peroxisomes are dynamic structures in the cell, moving at rates of up to 7 microns/sec along actin filaments using myosin motors (Muench and Mullen 2003).

1.4.2 The RNA- and microtubule- binding activity of peroxisomal multifunctional protein and peroxisomal malate dehydrogenase

A biochemical screen was performed in our laboratory to identify proteins that possess microtubule and RNA binding activities, with the goal of identifying proteins that have a role in mRNA trafficking in the cell (Chuong et al. 2002). A protein that was identified in this screen was the peroxisomal multifunctional protein (pMFP), a protein that possessed microtubule binding activity and a high level of RNA binding activity. pMFP is a peroxisomal matrix protein with three enzyme activities involved in the β-oxidation of fatty acids in the peroxisome. It binds to microtubules *in vitro* and *in vivo*, and binds to RNA in a non-sequence specific manner (Chuong et al. 2002; Chuong et al. 2005). Interestingly, a large-scaled analysis of tubulin binding proteins identified five additional peroxisomal matrix proteins (Chuong et al. 2004). One of these proteins, peroxisomal MDH (pMDH), was later shown to bind to microtubules *in vivo* and bind to RNA in a non-sequence specific manner, similar to that of pMFP (Freeman and Muench, unpublished results). Coupled with live cell imaging of actin-mediated dynamic movements of peroxisomes that showed frequent interactions of these organelles with microtubules (Chuong et al. 2005), a working model was proposed whereby the microtubule and RNA binding activities of pMFP (and later pMDH) were implicated to have a post-transcriptional control over its own RNA (Muench and Park 2006). In this model, nascent pMFP interacts with microtubules in the cytosol where it binds to its own mRNA to repress its translation in an autoregulatory manner. When peroxisomes traveling along the actin filaments reach the vicinity of the microtubule and import pMFP, translational repression is relieved upon import of the peroxisomal matrix protein (Figure 1.5). This model was proposed to also

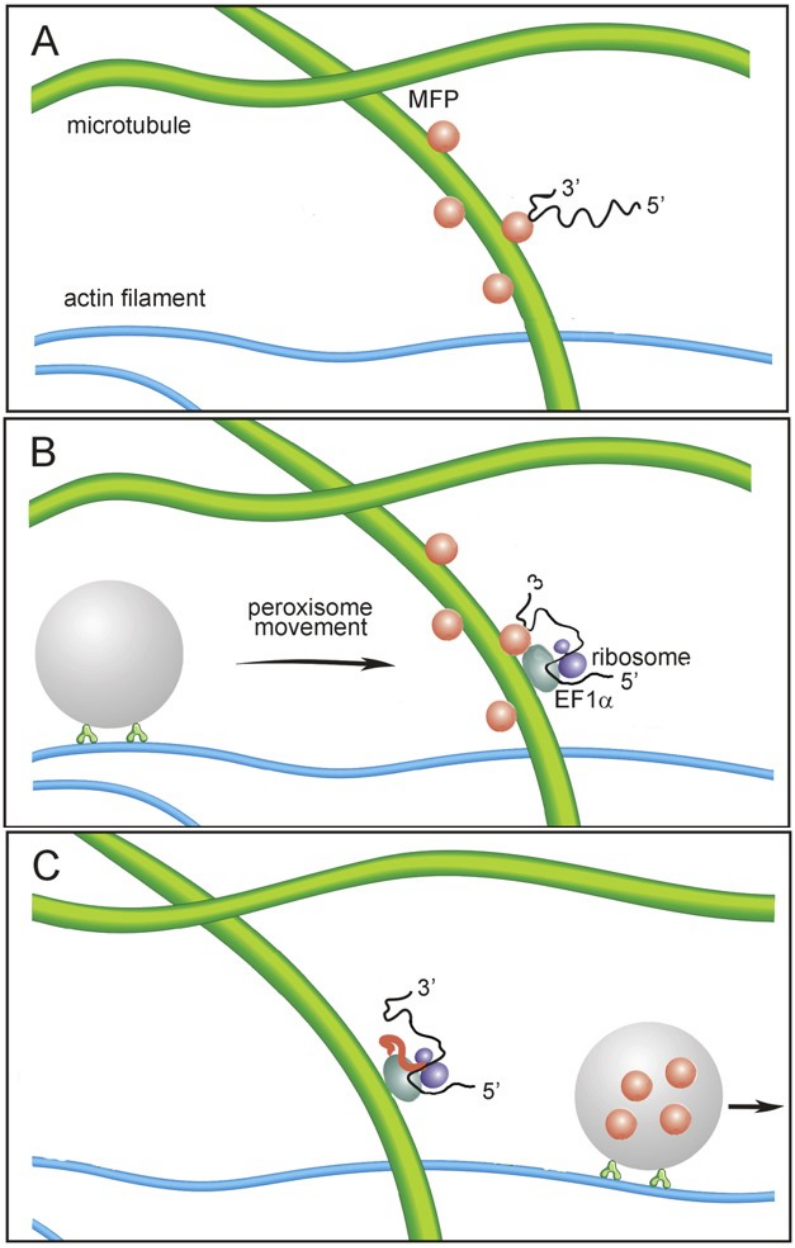


Figure 1.5 Working model for peroxisomal matrix protein synthesis and import. The model represents a possible role for the MT and RNA-binding activities of pMFP (and other peroxisomal matrix proteins) on its translational autoregulation and import into peroxisomes. (A) pMFP binds to MTs and its own mRNA. (B) This interaction functions to inhibit translation of nascent MFP protein in an autoregulatory manner. (C) As peroxisomes pause on or pass by MTs, MFP is imported into the peroxisome and translational repression is lifted. (from Muench and Park 2006)

apply to pMDH. Recent experiments involving pulse-chase labeling of total proteins and immunoprecipitation of pMDH in cells treated with cytoskeleton disrupting agents demonstrated that there was only subtle changes in the synthesis and peroxisomal import of pMDH, thereby providing refuting evidence for the cytoskeleton component of this model (Dahodwala and Muench, unpublished results). The identification of RNA targets that are bound by pMFP and pMDH *in vivo* would be valuable in providing insight into whether translational autoregulation is a component of the post-transcriptional control roles for these proteins, or whether they bind to other RNA and possess other roles in the cell.

The RNA binding activities of pMFP and pMDH adds to the list of metabolic enzymes that possess RNA binding 'moonlighting' activities (Cieśla 2006; Hentze and Preiss 2010; Scherrer et al. 2010). Two recent studies mammalian cell studies used *in vivo* UV crosslinking to covalently couple RNA-binding proteins to their cognate RNAs in an effort to characterize the RNA-interactome (Baltz et al. 2012; Castello et al. 2012). Importantly, these studies purified mRNA complexes from the UV crosslinked cells and washed under stringent denaturing conditions to remove non-specific interactions. In addition to numerous other metabolic enzymes, both studies identified cytosolic MDH as an authentic RNA-binding protein. The functional relevance of these studies links intermediary metabolism with the post-transcriptional regulation.

1.5 Goals of this thesis

The goal of this thesis was to identify RNA targets of two groups of RNA-binding proteins from Arabidopsis. These include members of the Arabidopsis PUF family of RNA-binding proteins, and peroxisomal MFP and MDH. Identification of the RNAs bound

by these proteins will assist in understanding the functional role of these proteins in post-transcriptional control of gene expression.

The objectives of this thesis were to:

1. Use an *in vitro* selection approach to identify the RNA target sequences of several representative Arabidopsis APUM proteins as well as pMFP and pMDH.
2. Use biochemical approaches to further analyze the RNA binding characteristics of these proteins with their cognate RNA sequences.
3. Identify the RNAs that selected Arabidopsis APUM proteins interact with *in vivo*.

CHAPTER 2 - MATERIALS AND METHODS

2.1 Plant Growth Conditions

Arabidopsis seeds were surface sterilized with 70% ethanol and washed in ddH_2O prior to 3 days of stratification at 4°C. Germination was either in soil or on 0.5X Murashige and Skoog media (MS, PhytoTechnologies Laboratory) on plates containing 1% sucrose and 0.7% phytoagar (PhytoTechnologies Laboratory). Arabidopsis plants were grown in the growth chamber (Percival, AR36LC8 or AR66LC8) with a diurnal program (16-hour day and 8-hour night at 22°C) *Nicotiana benthamiana* plants were grown in soil in a growth chamber (Conviron, CMP4030) at the University of Calgary Greenhouse. The diurnal program was set at 16-hour days at 25°C and 8-hour nights at 22°C.

2.2 Molecular Cloning

The clones made in this thesis were used for protein expression in *E. coli*, expression in transgenic Arabidopsis plants and transient expression in *Nicotiana benthamiana*. PCR amplification was from various template sources, including Arabidopsis genomic DNA, cDNA reverse-transcribed from Arabidopsis total RNA, company purchased cDNA clones purchased, and cDNA libraries from yeast (*Saccharomyces cerevisiae*, gift from Dr. Vanina Zaremberg). A few clones were synthesized by Gateway recombination using previously made pEntry clones made by other members of the lab, and there were several prokaryotic expression constructs were designed and synthesized by GenScript. The primers synthesized to construct plasmids for protein expression in prokaryotic system are listed in the Table 2.1. For the cloning for

Table 2.1 Oligonucleotide primers used for APUM-HD and yeast Nop9 coding sequence amplification and cloning.

APUM2-HD	Forward	5’- CGAGGAGGATCCTTTGGATCTTCAATGCTTGAAG -3’
	Reverse	5’- CGAGGAGTCGACTTACAAAGCCATCCTCCTCTCTCCAG -3’
APUM6-HD	Forward	5’- GAAATTGGATCCTATCCTGGATGGCAGCCACAA -3’
	Reverse	5’- GAAATTGTCGACTCATCTCCTCAATTCTTGGTTTTC -3’
APUM12-HD	Forward	5’- GAAATTGGATCCCTCACAATGAGTCTCAATAATCTG -3’
	Reverse	5’- GAAATTGTCGACTTACTTCTTCGAGCTAAGTGC -3’
APUM18	Forward	5’- GAAATTGGATCCATGGCAGTCGCTGATAATCC -3’
	Reverse	5’- GAAATTGTCGACTCAGCAACGAAGCCTAAATGA -3’
APUM23	Forward	5’- GAAATTGGATCCATGGGTGAACGAGGAAAGTC -3’
	Reverse	5’- GAAATTGTCGACTCAAATTCTCATTTTATTTGAATG -3’
APUM24	Forward	5’- GGGCCCCTGGGATCCATGTCTTCCAAAGGTCTGAAACC -3’
	Reverse	5’- GCCGCTCGAGTCGACTCATTCAGGTTTCTTGGTTGC -3’
Nop9	Forward	5’- GAAATTGGATCCATGGGAAAGACTAAAACAAGAGGC -3’
	Reverse	5’- GAAATTGTCGACTTATCTATAGTGCTTTTGCTTTTTGAAG -3’

other purposes, all details including the vector of choice and the structure are listed in Appendix 1.

2.2.1 Arabidopsis genomic DNA extraction

A segment of four-week fully expended rosette leaf material (30-50 mg) was harvested in a microcentrifuge tube, frozen in liquid nitrogen and stored at -80°C. The sample was snap frozen again in liquid nitrogen before grinding in the microcentrifuge tube using a micro-pestle attached to a hand drill. Grinding was continued after the addition of 300 µL genomic DNA Extraction Buffer (100 mM Tris-HCl 8.0, 20 mM EDTA, 500 mM NaCl, 1.5% SDS) to break up visible chunks of leaf tissue. The crude lysate was incubated at 55°C for 30 minutes with occasional inversion. 300 µL chloroform/alcohol mixture (80% chloroform, 4% isoamyl alcohol and 16% ethanol) was added and after the lysate cooled to room temperature. The content was mixed thoroughly by inverting the tube vigorously, and then allowed to stand for 10 minutes at room temperature for phase separation. The supernatant was removed after centrifuging at 13000 rpm for 10 minutes. An equal volume of isopropanol was added and mixed. An optional incubation at low temperature (-20°C) was performed for improved precipitation prior to the 10-minute centrifugation at 13000 rpm at 4°C. The DNA pellet was washed twice by 70% ethanol followed by a brief high speed centrifugation, then air dried before dissolving the pellet in 5 mM Tris-HCl 8.0.

2.2.2 Arabidopsis total RNA extraction and reverse transcription

A segment of four-week fully expended rosette leaf material (30-50 mg) was harvested in a microcentrifuge tube, frozen in liquid nitrogen and stored at -80°C. The sample was snap frozen again in liquid nitrogen before grinding in the microcentrifuge tube using a micro-pestle attached to a hand drill. Two to three minute of grinding after the addition of 300 µL TRIzol (Invitrogen) was necessary to break up the visible chunks of leaf tissue. 60 µL chloroform was added, followed by 15 seconds of vigorous shaking. The tube was allowed to stand for one minute before 10 minutes of centrifugation at 11000 rpm, 4°C. The aqueous phase (top) was transferred to a new tube and was optionally mixed by 300 µL chloroform for an additional extraction. The clear aqueous phase was removed after a 4 minute centrifugation at 11000 rpm, then mixed with 1.2 times volumes of isopropanol. Optional incubation at low temperature (-20°C) for 2 hours was performed prior to a 10-minute centrifugation at 13000 rpm, 4°C. The RNA pellet was washed once with cold 70% ethanol followed by 3-minute high speed centrifugation spin at 4°C. The pellet was air dried and dissolved in DEPC treated ddH_2O. The total RNA subjected to qPCR analysis was treated with TURBO DNase (Ambion) in its compatible 1X buffer at 37°C for 30 minutes. Inactivation of DNase activity was performed by adding EDTA to a final concentration of 15 mM and heating at 75°C for 10 minutes. The isopropanol precipitation was used to extract the total treated RNA. Reverse transcription was performed using M-MLV or SuperScript III following the manufacturer's procedures (Invitrogen). One µg of total RNA was transcribed in a 20 µL reaction with 40 units of RNaseOUT (Invitrogen) or 20 units of SUPERase·In RNase Inhibitor (Ambion).

2.2.3 PCR and molecular cloning

The PCR reactions used Taq (Lamda Biotech.), Phusion (Thermal Scientific) or Platinum Pfx (invitrogen) DNA polymerase. All reactions were performed in a thermocycler (PTC-200, MJ Research). Oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT). The DNA fragments amplified from PCR were electrophoresed on 0.8-1% agarose gel (Amresco) in 1X TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA), then eluted by QIAquick Gel Extraction Kit (QIAGEN) and restriction enzyme digested. The final DNA concentrations were determined by spectrophotometry (NanoDrop, ND-1000). In general, 200 ng of DNA with the proper molarity ratio of insert: vector (3:1 to 2:1) was used in either ligation (T4 DNA ligase, Invitrogen) or recombination (In-Fusion recombinase, Clontech). The transformation method (heat-shock or electroporation) and host competent cell were chosen differently according to the purposes. Positive transformants were grown on solid LB media (1% w/v NaCl, 1% w/v Tryptone, 0.5% w/v yeast extract and 1.5% w/v agar) containing the appropriate antibiotics. Plasmid isolation was performed using QIAprep Spin Miniprep Kit (QIAGEN). Plasmids from candidate clones were used as templates in a PCR screen to confirm the size of the amplification product prior to sequencing.

2.3 Protein expression and purification

All PUM-HD fragments were cloned into the *Bam*HI/*Sal*I sites of pGEX-6p-1 (GE Healthcare), and resulted in a translational fusion to the C-terminus of the GST tag. APUM23 coding region substitutions were synthesized and inserted into the corresponding region of the APUM23 coding region in pGEX-6p-1 (GenScript). The nucleotide

sequences used in these substitution experiments into the original APUM23 sequence were from Puf repeats encoded by APUM2. The host cells BL21 (DE3) or Rosetta (DE3) (Novagen) were used for protein expression. A 0.5% (v/v) overnight start culture was inoculated into fresh LB media and grown at 37°C until the OD_{600} value reached 0.6. The induction with 0.05-0.5 mM IPTG was for 3 h at 37°C or 17 h at 15-22 °C. Cells were pelleted and resuspended in lysis buffer (25 mM Tris-HCl pH 8.0, 200-300 mM NaCl, 3 mM DTT) with the addition of 1 mg/mL lysozyme and 0.17 mg/mL PMSF as well as a proper concentration of the protease inhibitor (cOmplete, Roche). A 15 minute lysozyme digestion was performed on ice before sonication. The supernatant obtained by centrifugation was passed through a gravity column loaded with Glutathione Sepharose 4B (GE Healthcare) resin (1/30 of the supernatant volume). The resin was immediately washed in 10 column volumes of lysis buffer and 3 times of 10 column volumes of wash buffer (25 mM Tris-HCl pH 8.0, 500 mM NaCl, 3 mM DTT) before elution with elution buffer (25 mM Tris-HCl pH8.0, 200 mM NaCl, 30mM L-Glutathione Reduced (Sigma) and 3 mM DTT). The elution was concentrated in a 30 kDa or 50 kDa nominal molecular weight limit centrifugal filter units (Millipore) at 4°C by centrifugation at 3000 xg. The concentrated protein solution was further purified by AKTA FPLC (GE Healthcare) using a Superdex 200 Increase 10/300 GL (GE Healthcare) size exclusion column and a running buffer containing 10 mM Tris pH 8.0 and 100 mM NaCl. The samples from the fractionation were examined by SDS-PAGE (Protean II unit, Bio-Rad) and pure fractions were pooled together and the protein qualified by Bradford assay. Purified protein products were stored in -80°C freezer with the addition of glycerol to 20% (v/v).

2.4 SELEX

SELEX was performed similarly to that described previously for the murine PUM2 protein with some modifications (White et al. 2001). A degenerate DNA oligonucleotide (5'- GGGATCCGAATTCCCGACT$(N)_{20}$GGAAGCTTACTCGAGCGC -3') was used to produce the starting random pool of RNA having a complexity of $1x10^{12}$. Twenty pmole of this oligonucleotide provided approximately 12-fold coverage of each unique sequence combination. PCR amplification of this oligonucleotide was achieved using the forward SELEX primer (5'-GGACC**TAATACGACTCACTATAGGG**ATCCGAATTCCCGACT -3') contained the T7 RNA polymerase binding site (bold) paired with the reverse SELEX primer (5'-GCGCTCGAGTAAGCTTCC -3'). After seven cycles of amplification (94°C for 35 s, 55°C for 35 s, 68°C for 45 s) using Platinum Pfx Polymerase (Invitrogen) and a final extension of 2 minutes at 68°C, the dsDNA was electrophoresed on 3% agarose gel (TopVision, Thermo Scientific) for 30 minutes at ~95V and the 79 bp fragment was eluted. RNA pools were transcribed from 1 µg of purified dsDNA template using 100 units of T7 RNA Polymerase-Plus (Ambion) in a reaction containing 2 mM NTPs and 33 nM [α32P]-UTP (3000Ci/mmol; PerkinElmer Life Sciences) for 2 hours at 37°C, followed by a digestion using 10 units of DNase (TURBO; Ambion) at 37°C for one hour. RNA was isolated by TRIzol (Invitrogen) and chloroform extractions followed by an isopropanol precipitation and dissolved in 50 µL reaction buffer (10 mM HEPES, pH 7.4, 1 mM EDTA, 50 mM KCl, 1 mM DTT, 0.01% (w/v) BSA and 0.01% (v/v) Tween 20), and then passed through an equilibrated spin column (Bio-Gel P-6; Bio-Rad). A parallel non-radioactive

trial was performed to provide an RNA concentration reference, which was quantified by spectrophotometry (model ND-1000; NanoDrop).

For each round of selection, RNAs were pre-cleared in 20 µL of GST-coupled Glutathione Sepharose 4B resin to reduce non-specific RNA binding prior to RNA enrichment. Two µL of precleared RNA was removed to determine RNA input by isotope counting using a scintillation counter (Beckman). Two nmol of GST tagged PUM-HD and 0.4 nmol of precleared RNA was mixed in 100 µL of reaction buffer containing 50 units of RNase inhibitor (SUPERase·In; Ambion). This 5:1 molar ratio of protein:RNA was necessary to initiate enrichment of the RNA in the first two rounds of SELEX, because the APUM proteins did not possess full activity. In the third and fourth rounds, the protein:RNA molar ratio added was equal, and in the fifth and subsequent rounds, the ratio was 1:5. Reducing the protein concentration in subsequent rounds provides more stringent selection, as has been performed elsewhere (White et al. 2001). The RNA enrichment profiles were normalized according to these ratios. After 30 minute incubation at room temperature, the RNA-protein complexes were mixed with 20 µL of Glutathione Sepharose 4B resin for 15 minutes at room temperature, with gentle resuspension of the matrix every 5 minutes. The resin was washed three times in 400 µL reaction buffer followed by elution of the protein-RNA complexes with 200 µL elution buffer (25 mM Tris-HCl pH 8.0, 200 mM NaCl, 30 mM reduced glutathione and 3 mM DTT). Bound RNA was released using 300 µL TRIzol followed by chloroform extraction and ethanol precipitation. The RNA pellet was dissolved in 19 µL DEPC-treated water. Two µL was removed for scintillation counting to calculate the percentage RNA bound by comparison to the input isotope count. The remaining RNA was reverse-transcribed using SuperScript III (Invitrogen) and 1.2 µM

SELEX reverse primer in a 40-µL reaction, and the cDNA product was precipitated by ethanol. The PCR-amplifications from cDNA templates for each SELEX cycle were performed in 200 µL reactions with 1.875 µM SELEX forward and reverse primers, 1 mM dNTPs, 1 mM $MgCl_2$ and 4 units of Platinum Pfx polymerase for 10 to 12 thermal cycles (94°C for 35 s, 55°C for 35 s, 68°C for 30 s) and a 3-minute final extension at 68°C. The resulting dsDNA was recovered by gel elution and used for the subsequent transcription. This SELEX cycle was repeated 9-12 times until no further enrichment was observed. The final dsDNA from the cycle with the highest level of enrichment was digested with EcoRI and HindIII and cloned into pBlueScript for Sanger sequencing. Only unique and reliable sequences from independent transformants were used for analysis using MEME (http://meme-suite.org/tools/meme)(Bailey and Elkan 1994) to obtain the consensus sequence, logo graph and the *E*-value. The number of sequenced clones that were used for MEME analysis were: 52 for APUM2, 60 for APUM6, 38 for APUM12, 62 for APUM13, 35 for APUM18, 112 for APUM23, 43 for APUM24, 49 for Nop9, 58 for pMFP and 43 for pMDH. The MEME distribution of the binding site was set with a default 'zero or one occurrence per sequence', and the length of the consensus sequence motif was set at 4-20 nucleotides. The MEME analysis failed to identify a valid consensus sequence from several raw sequences (16 from APUM2, 11 from APUM13, 15 from APUM18 and 7 from APUM23), which were not included in the generation of the logo graph. The refined logo graphs were generated by the web logo graph tool (http://weblogo.berkeley.edu/logo.cgi) using the populations validated by MEME. The log likelihood ratio (represented by the *E*-value) of the consensus sequence motifs produced by MEME for APUM24 (5.2 e), Nop9

(5.1 e^{-3}) and pMFP (2.1 e^{+7}) were very high. Therefore, these RNA-binding proteins were considered to bind the RNA sequence non-specifically.

2.5 EMSA

EMSAs were performed as described previously (Tam et al. 2010), with minor modifications. Recombinant APUM proteins were reconstituted in binding buffer (10 mM HEPES, pH 7.4, 1 mM EDTA, 50 mM KCl, 1 mM DTT, 0.01% Tween 20). Synthetic RNAs (Dharmacon) were radiolabelled using [γ^{32}P]-ATP (3000 Ci/mmol; PerkinElmer Life Sciences) and T4 polynucleotide kinase (Thermo Scientific). Binding reactions were in a 20 µL volume containing 0.01 nM (APUM2 trials) or 0.05 nM (APUM6, APUM12, APUM23 and mutated APUM23 protein trials) labelled RNA and 0.03 to 1024 nM range of gradient protein concentrations with the presence of 20 units RNase inhibitor (SUPERase·In; Ambion). The reactions were incubated at room temperature for 30 minutes and electrophoresed on a 6% non-denaturing acrylamide gel at 96 V for 20 minutes at 4°C. The gels were dried and exposed to a storage phosphor screen (Kodak), and the screen was scanned using a PhosphorImager (Molecular Imager FX; Bio-Rad). Densitometry was performed using Quantity One software (version 4.5.1; Bio-Rad) and ImageJ (1.47v, Wayne Rasband, National Institutes of Health), and the data was analyzed using Prism 5 (version 5.03; GraphPad). To determine the apparent dissociation constant and binding curve, the fractions of RNA bound to protein (the relative pixel intensity in the complex band divided by the sum of the pixel intensities in the complex band plus the free RNA band) were plotted against protein concentrations in a semi-log graph format. The curves were fitted in the equation of “One site - specific binding” defined by the Prism 5 software.

The average apparent dissociation constant values (± SEM) were derived from three or four trials, depending on the particular experiment.

The activity of the recombinant proteins was determined by EMSA using an excess amount of RNA in order to saturate the protein binding capacity. Fifty nM of the cognate radiolabelled RNA was incubated with 5 nM, 2 nM and 1 nM of protein in a 20 µL volume, and electrophoresed alongside 4 nM, 0.6 nM and 0.1 nM free RNA as a concentration standard. The absolute amount of RNA bound by protein of different concentrations was determined by comparing the pixel intensity of the complex band with the free RNA bands in the concentration standard. RNA concentration was plotted on a graph against protein concentration, and the slope of the line was measured to determine the percentage of active protein molecules.

2.6 Structural Modeling of APUM23

The SWISS-MODEL protein structure homology server (Biasini et al. 2014) was used to produce APUM23 models that guided the mapping of predicted Puf repeat locations and TRM identity, as described in the result. The classical PUF protein structure consisting of α-helical repeats and precisely positioned TRM amino acids provided the template for these models. Partial models were produced by analyzing two overlapping regions of the APUM23 polypeptide: amino acids 1–382 and 221–731. 94 and 67 templates were identified for these two sequences, respectively. The top 20 templates that were most similar in sequence to APUM23 were modeled for each region. Models that demonstrated the typical PUF concave structure, Puf repeat α-helical structure, and TRM positioning were used for repeat and TRM predictions.

2.7 Plant transformation

The expression constructs were introduced into *Agrobacterium tumefaciens* (GV3101 strain) by electroporation (Gene Pulser, Bio-Rad) using 2 mm electrode gap cuvette at 2500 volts/~0.5 second pulse setting. The wild-type Arabidopsis (Col-0) or one with desired genetic background were grown in soil under normal conditions and trimmed to increase bolting. The plants were transformed using the floral dip method, by submerging the above-ground part of plant in a 5% sucrose solution with resuspended *Agrobacterium* culture (OD_{600}=0.8) containing 0.05% Silwet L-77 (Sigma) for 2-3 seconds. The dipped plants were cultured normally and seeds were harvested. The seeds were surface-sterilized and placed on 0.5X MS medium plates containing 1% sucrose and 0.7% phytoagar and proper antibiotics or herbicide (Hygromycin B: 25 mg/L, Kanamycin: 20 mg/L, phosphinothricin: 15 mg/L). The seeds were stratified for 3 days at 4°C in the dark before 5-hour exposure under light at room temperature. Following a 48-72 hour dark incubation at room temperature, seedlings with elongated hypocotyl were transferred to soil and grown in the culture chamber. The putative transformants were subjected to PCR genotyping and western blotting test to confirm identity. The positive transformants were grown for two more generations to acquire homozygous lines.

2.8 Agroinfiltration and transient expression

2.8.1 Agroinfiltration using Arabidopsis seedlings (AGRObest)

Seeds were surface-sterilized and placed on 0.5X MS medium plates containing 1% sucrose and 0.7% phytoagar and stratified for 3 days before transfer to a culture chamber.

Agrobacterium cells containing desired construct were cultured in LB liquid medium at 28°C with appropriate antibiotics until the OD_{600} value reached 0.6. The cells were then pelleted and re-suspended in liquid media AB-MES (17.2 mM K_2HPO_4, 8.3 mM NaH_2PO_4, 18.7 mM NH_4Cl, 2 mM KCl, 1.25 mM $MgSO_4$, 100 µM $CaCl_2$, 10 µM $FeSO_4$, 50 mM MES, 2% glucose (w/v), pH 5.5) lacking antibiotics but with the addition of 200 µM acetosyringone, then cultured at 28°C for 12-16 hours. The cells were pelleted again and resuspended in ¼ MS and ½ AB-MES solution containing 200 µM acetosyringone (OD_{600}=0.02). Ten 3-day old seedlings were transferred into one well on a 6-well culture plate and co-cultivated with 1 mL *Agrobacterium* culture prepared earlier for 2 days growth under normal culture conditions. Co-cultivation medium was removed after infection and replaced by 1 mL ½ MS medium containing 100 µM Timentin and incubated for additional one more day before analysis. The expressing seedlings were examined by epifluorescence microscope and snap-frozen for use in western-blot analysis.

2.8.2 Agroinfiltration using Nicotiana benthamiana plants

The *N. benthamiana* plants were grown until 4 weeks old. The thick green leaves with 5-7 cm diameter were used for this method. *Agrobacterium* containing the desired construct was cultured in LB liquid medium at 28°C with appropriate antibiotics until the OD_{600} value reached 0.5-0.8. The cells were then pelleted down and resuspended in infiltration buffer (50 mM MES, 10 mM $MgCl_2$, 150 µM acetosyringone, pH 5.5) to make a OD_{600}=0.5 solution. The suspension was incubated in the dark for 3 hours at room temperature without shaking. A 1 mL volume needleless syringe was used to inject the solution into the abaxial side of the leaves. The infiltrated plants were grown under normal

conditions for 40 hours to reach maximum expression. Leaf tissue was collected and snap-frozen for further experiments.

2.8.3 Transient expression using biolistic particle bombardment

Fluorescent protein tagged proteins were transiently expressed in fava bean leaf epidermal cells by particle bombardment using a gene particle gun (PDS-1000, Bio-Rad). Young and healthy (5 cm in length) fava bean leaves were used. Ten µg plasmid DNA was coated onto 1 µg gold particles (1.2 µm) and mixed sequentially with 17 µL $CaCl_2$ (2.5 M), 6 µL spermidine (0.1 M) in a microfuge tube while vigorously vortexing. After 2 minutes of continuous vortexing, the gold particles were allowed to settle at the bottom of tube for 1 minute. The liquid was removed and the pellet was washed with chilled 70% ethanol and then 95% ethanol. The pellet was finally resuspended in 24 µL 95% ethanol by flicking the tube and then transferred to a plastic macrocarrier for air drying. Once dried, the macrocarrier was assembled on the particle gun where a 1200 psi force was applied to accelerate the particle to be embedded into the epidermal cells that were placed approximately 10 cm below the particles. The leaves were kept in dark at room temperature for 18 hours followed by microscopic examination.

2.9 Microscopic analysis

The mounting method varied depending on the material used for microscopic analysis. If Arabidopsis seedlings were used, the whole seedling was mounted on a glass slide (25mm x 75 mm x 1 mm) with a few drops of distilled water. A cover slip was applied on the top and used to crush the tissue gently. Cell culture or seedling tissue was mounted

with its storage buffer. Arabidopsis or fava bean epidermal cell layers were peeled from the mesophyll cells, and mounted with distilled water. All specimens were observed using epifluorescence microscpopy (Leica DMR) with a Plan Fluotar 20x and 40x or Plan Apo 63x (oil immersion) objective lens. The images were captured by a top mounted CCD camera (Retiga 1350 EX) and processed using Velocity software (Version 4.3.1, Improvision).

2.10 Western blotting

Leaf tissue was snap-frozen by liquid nitrogen and crushed into powder in a mortar and pestle or in a microfuge tube by plastic micro-pestles, depending on the amount of tissue used in the experiment. Crack buffer (2X, 100 mM Tris 6.8, 200 mM DTT, 4% (w/v) SDS, 20% (v/v) glycerol, 0.02% (w/v) bromophenol blue) was added to the powder before it thawed. The solubilized and denatured protein was boiled for 10 minutes in a water bath followed by a 2-minute centrifugation at 13000 rpm. The protein samples were electrophoresed on SDS-PAGE, then transferred onto a nitrocellulose membrane (0.2 μm, Bio-Rad) in Tris-glycine buffer (25 mM Tris base, 192 mM glycine) with 260 mA constant current for 1.5 hours at 4°C. The resulting membrane was blocked in 1x PBS buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) with 3% (w/v) skim milk powder. The primary or secondary antibody was added in a proper dilution ratio according to manufacturer's instruction and incubated on an orbital shaker at room temperature for at least 3 hours. The washes between the primary and the secondary antibodies and after the secondary antibody probing were using 1x PBS with 0.1% Tween-20 (Sigma). The chemiluminescent substrate for HRP (Thermo Scientific) was added onto

the membrane before exposing the membrane to a BIOMAX light film (Kodak) in an autoradiograhy cassette (Fisher Scientific).

2.11 Immunoprecipitation with UV or formaldehyde crosslinking

The UV crosslinking was performed in a crosslinker (Ultra-Lum) emitting a light source of 254 nm. The energy applied on leaf tissue or seedlings was 3 times of 400 mJ/cm^2. Incubation in 1% (w/v) paraformaldehyde (in 1x PBS buffer) was by vacuum infiltration for 15 minutes was also used to crosslink protein with protein or RNA in plant tissue. The formaldehyde crosslinking was stopped by adding 0.125 M glycine. Three times wash was performed using cold 1x PBS to eliminate the formaldehyde residue. The crosslinked tissue was snap frozen and ground prior to adding IP buffer (50 mM Tris 8.0, 150 mM NaCl, 0.1% (w/v) SDS, 0.5% (v/v) Triton-X100, 1 mM EDTA and protease inhibitor cocktail). The supernatant was obtained by 14000 rpm centrifugation for 2 minutes. The anti-GFP antibody/protein A agarose resin or anti-FLAG antibody conjugated agarose (Sigma) or GFP-Trap magnetic resin (Chromotek) was added to the supernatant and the mixture was subjected to the end-to-end rotation at 4°C for 3 hours. After discarding the supernatant, the resin was washed 3 times in wash buffer (25 mM Tris 8.0, 500 mM NaCl, 0.5% (v/v) Triton-X100, 1 mM EDTA. If RNA was to recovered, then the resin was resuspended in the digestion buffer (25 mM Tris pH 7.5, 50 mM NaCl, 12.5 mM EDTA, 1% (w/v) SDS) containing 1 mg/mL proteinase K (Roche) and 20 units SUPERase·In RNase Inhibitor per 100 µL volume. The reaction incubation was performed at either 42 °C for 30 minutes or 4 °C for overnight. The mixture was subjected to Trizol and chloroform extraction and

ethanol precipitation to obtain the RNA. The resulting RNA was reverse transcribed and analyzed by PCR.

CHAPTER 3 - DETERMINATION OF THE RNA TARGET SEQUENCES OF ARABIDOPSIS PUF PROTEINS

3.1 Introduction

Previous studies have determined that the Arabidopsis PUF family consists of up to 26 members (Francischini and Quaggio 2009; Abbasi et al. 2010; Tam et al. 2010). Although most PUF proteins studied to date possess eight clustered Puf repeats in their RNA binding domain, Arabidopsis PUFs exhibit significant variability in the number and arrangement of their Puf repeats as well as the identity of their tripartite recognition motifs (TRMs)(Francischini and Quaggio 2009; Tam et al. 2010; Abbasi et al. 2011). The nucleolar localized Arabidopsis PUF protein, APUM23, possesses a pattern of predicted Puf repeats that differs from those of all other APUM proteins. APUM23 is predicted to possess ten Puf repeats that are spaced throughout the protein and contains some unusual TRMs (Tam et al. 2010). The identification of the RNA targets of APUM proteins is important in understanding their functional diversity. *In vitro* techniques have been valuable in identifying the direct target sequences that are bound by some types of RNA binding proteins. The systematic evolution of ligands by exponential enrichment (SELEX) is one of these *in vitro* approaches.

SELEX is a biochemical technique that uses iterative rounds of selection from pools of random single-stranded DNA or RNA oligonucleotide sequences to determine binding specificity of target ligands or proteins (Ellington and Szostak 1990). Nucleic acids that bind with high affinity to the ligand or protein are enriched after multiple selection cycles. SELEX has become a powerful tool in the identification of RNA targets of RNA-binding

proteins, especially for those of sequence-specific RNA-binding proteins (see Figure 3.1 and Experimental Procedures for a description of the procedure used in the present study). SELEX provides an advantage over *in vivo* approaches in that it is an approach that is not constrained by limited concentrations of protein or RNA, or does not require the specific immunoprecipitation of the RNA-binding protein from complex mixtures of protein. Thus, the sequence results obtained by SELEX are reproducible, even when the target protein in cells is present in low abundance in cells.

This chapter reports the use of SELEX to identify the consensus RNA binding sequence of seven plant PUF proteins to determine if the variability of the RNA-binding domains of plant PUFs results in binding to variable RNA targets. The SELEX procedure produced high confidence, specific RNA binding for four out of the seven APUM proteins that were studied. Three proteins (APUM2, 6 and 12) preferred classical PUF UGUA core-containing RNA targets that were eight or nine nucleotides in length. In contrast, APUM23 bound specifically to a novel RNA target that was 10 nucleotides in length and possessed a non-canonical nucleotide sequence core. APUM23 also had a binding preference for cytosine at the eighth nucleotide position of the RNA target. The SELEX-derived consensus sequences were validated by electrophoretic mobility shift assays using wild type and base-substituted RNA. Swapping of the five-amino acid motif that contained the TRM confirmed the identity of three of the predicted Puf repeats of APUM23, including the cytosine-binding repeat. The 10-nucleotide APUM23 target sequence was shown to be present in the 18S rRNA sequence, thereby supporting the known functional role of APUM23 in 18S rRNA maturation. This work also reveals that APUM23, an ortholog of yeast Nop9, could provide an advanced structural backbone for Puf repeat engineering and

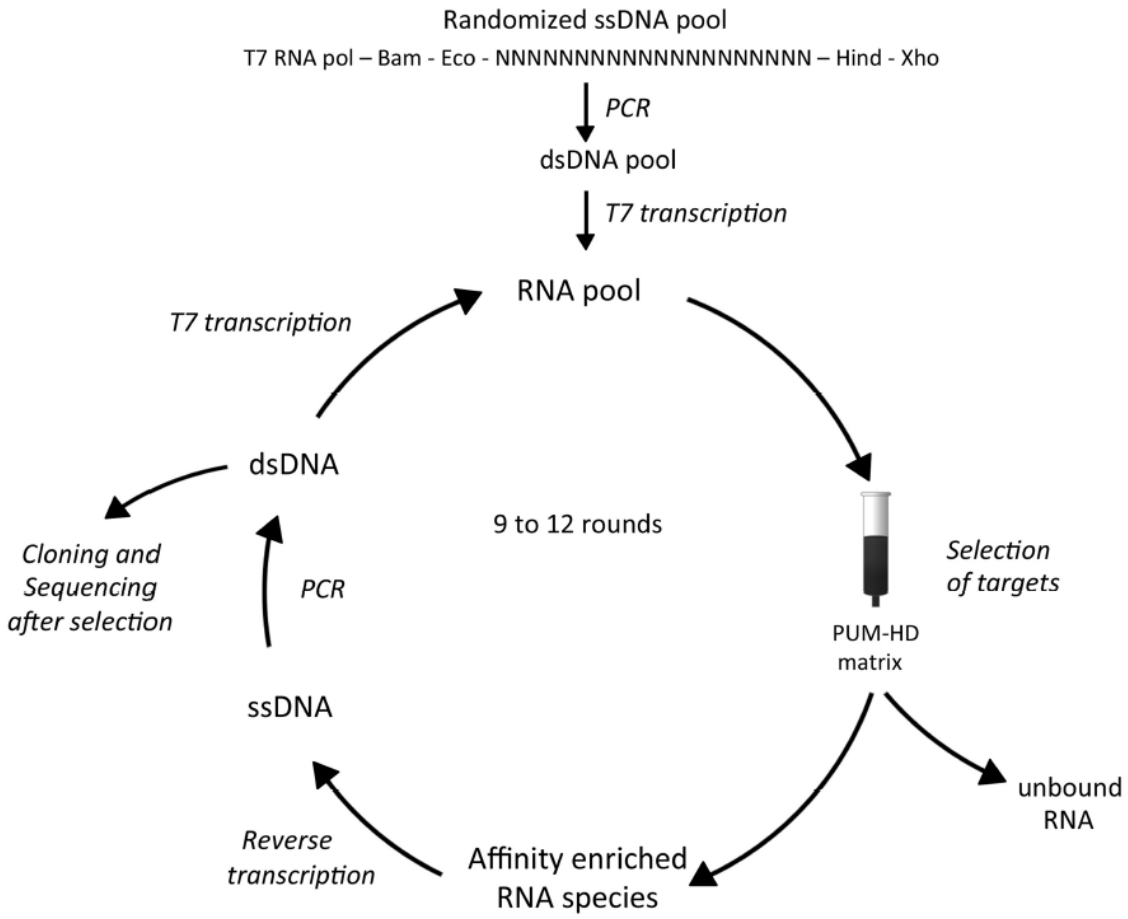


Figure 3.1 Diagrammatic summary of the SELEX procedure.

target-specific regulation of cellular RNAs.

3.2 Results

3.2.1 Prediction of Puf repeat number and location in selected APUM Proteins

Selection of the APUM proteins to be analyzed in this study (APUM2, 6, 12, 13, 18, 23 and 24) was based on the conservation or variability in the number and position of their predicted Puf repeats, and on the TRM composition of their repeats (Figure 3.2). A previous report from our laboratory (Tam et al. 2010) predicted the Puf repeat locations and TRM compositions of each of these proteins from sequence alignments with other well characterized PUF proteins. APUM2 and APUM6 were predicted to possess a PUM-HD with the characteristic eight tandem Puf repeats that are tightly clustered at the C-terminal region of the proteins, and they share identical TRMs with human PUM1 (Edwards et al. 2001). APUM12 and APUM13 also appear to contain eight Puf repeats based on sequence alignments with human PUM1, and all of the TRMs of APUM12 are identical to those of yeast Puf4 (Miller et al. 2008; Yeming Wang et al. 2009). In contrast, most of the APUM13 TRMs are different from those that are present in human PUM1 (except in repeat 6 and 8) as was supported by structural modeling (Tam et al. 2010). APUM18 is the shortest Arabidopsis PUF protein chosen in this study. It was predicted to possess seven clustered Puf repeats and the absence of N-terminal variable region (Tam et al. 2010). The APUM23 and APUM24 Puf repeat and TRM predictions made by Tam et al. (2010) required further analysis because their Puf repeat patterns were not clustered in the C-terminal portion of the protein.

APUM23 is predicted to possess ten Puf repeats that are unevenly positioned

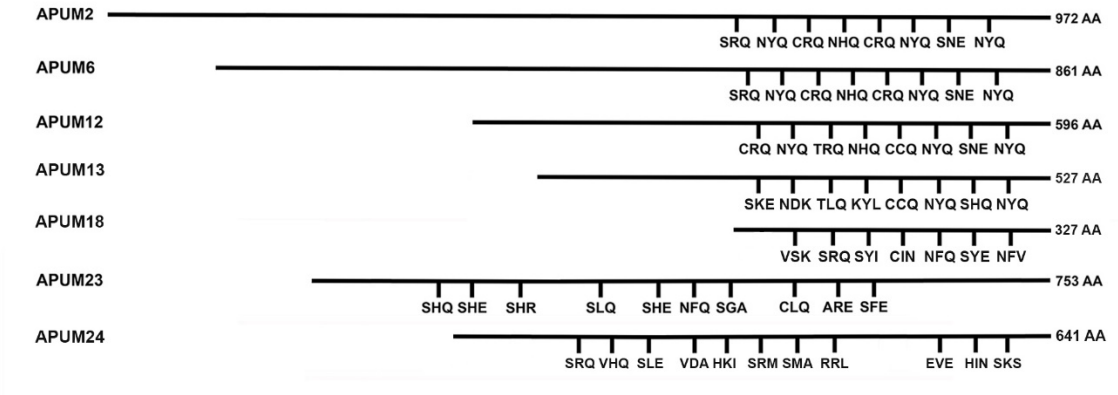


Figure 3.2 Predicted location and sequence of TRMs from the PUF proteins analyzed by SELEX. Location map of the TRMs within each of the predicted Puf repeats in the APUM proteins used in this study. The three-letter TRMs are indicated. The length of each protein (in amino acids, AA) is shown at the right of each line.

throughout the central region of the protein (Figure 3.2). Our prediction for the number and location of the APUM23 Puf repeats and the identity of their TRMs was based on three lines of evidence. First was their sequence identity to TRMs that are common in Puf repeats from other PUF proteins. Second was the positioning of the predicted TRMs in alpha-helical regions in a three-dimensional model of APUM23. Third was the level of conservation of the five-amino acid motif that contains the TRM sequence in plant orthologs of APUM23. Modelling the three-dimensional structure of the entire APUM23 protein sequence (see Section 2.6 for details on modelling), when referenced against the crystal structure of human PUM1, resulted in a distorted structure in a portion of the model. Therefore, two overlapping partial models were obtained that together modelled the entire protein. Twelve putative Puf repeats were identified in these models (Figures 3.3), many of which possessed previously identified TRMs located within the alpha-helix on the inner, concave face of the protein that contacts RNA (Wang et al. 2002; Campbell et al. 2014). The position of these TRMs on each of the modeled repeats were oriented in a typical Puf repeat fashion, with the TRMs (amino acids at positions 1, 2 and 5) exposed to the outer portion of the second alpha helix and the hydrophobic amino acids (positions 3 and 4) hidden on the inside of the helix. Repeat 3' was considered an unlikely Puf repeat candidate since it possesses an unusual predicted TRM (DNE) that is not conserved at this position in other species, including other members of the dicotyledonous class of plants (Figure 3.4 and 3.5). This repeat may be positioned away from the RNA binding surface *in vivo*, allowing repeats 3 and 4 to bind to their adjacent target bases. Putative Puf repeat 10' possesses a TRM (SRQ) (Figure 3.3B) that is present in other PUF proteins (Campbell et al. 2014). However, the presence of the 10' repeat at this position is not highly conserved

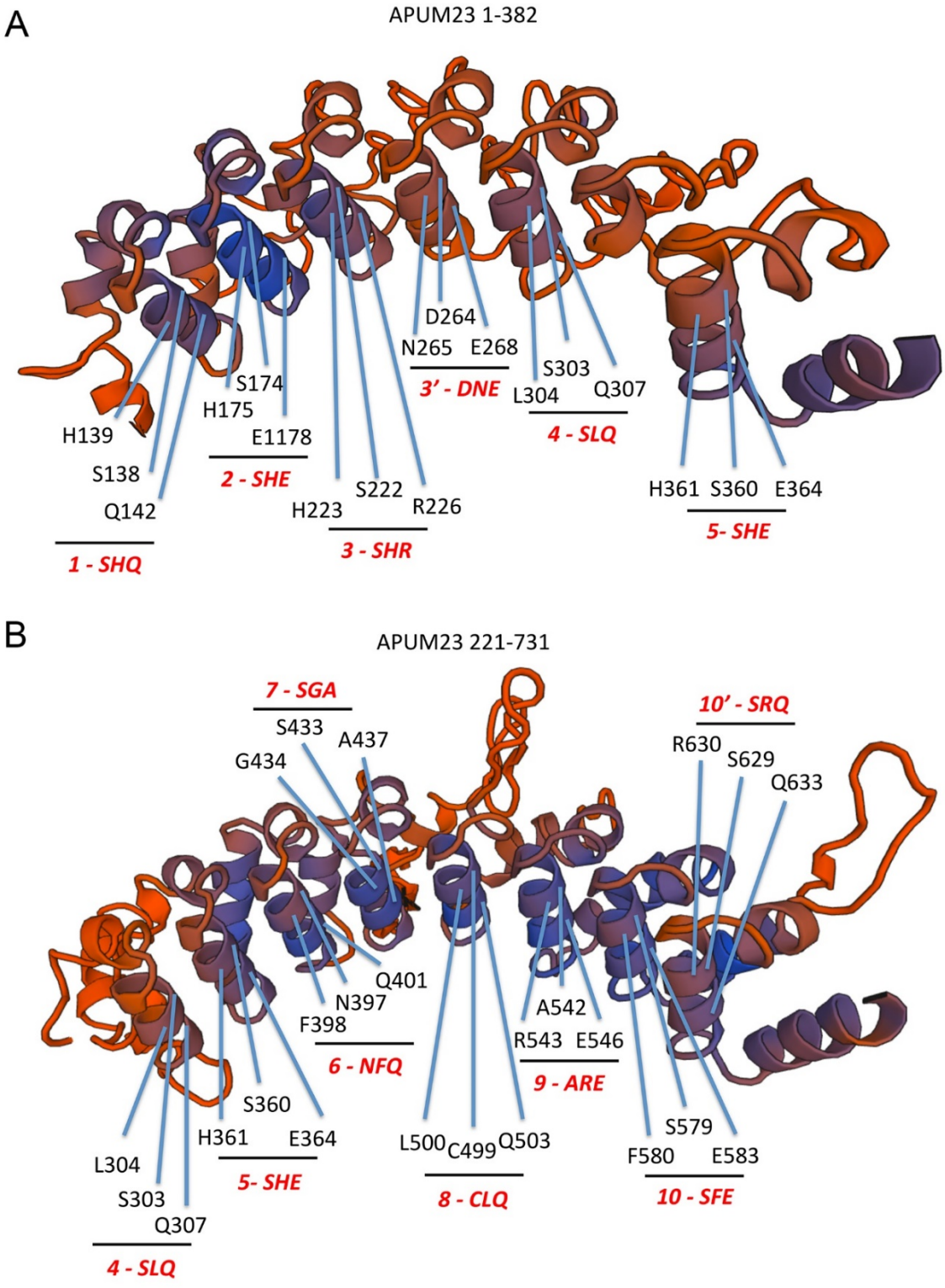

A
APUM23 1-382
H139
S138
Q142
1 - SHQ
S174
H175
E1178
2 - SHE
S222
H223
R226
3 - SHR
D264
N265
E268
3' - DNE
S303
L304
Q307
4 - SLQ
H361
S360
E364
5- SHE
B
APUM23 221-731
7 - SGA
S433
A437
G434
Q401
N397
F398
6 - NFQ
S360
H361
E364
5- SHE
L304
S303
Q307
4 - SLQ
L500
C499
Q503
8 - CLQ
10' - SRQ
R630
S629
Q633
A542
R543
E546
9 - ARE
S579
F580
E583
10 - SFE

Figure 3.3 Models of APUM23 referenced against the structure of human PUM1 (PDB ID: 1M8X). The APUM23 amino acid sequence was split into two overlapping parts, and each part was modeled independently using the Swiss-Model server. Ribbon diagrams of (A) the N-terminal region of APUM23 (amino acids 1 to 382) comprising putative Puf repeats 1, 2, 3, 3', 4 and 5 and (B) the C-terminal region (221 to 753) putative repeats 4, 5, 6, 7, 8, 9, 10 and 10' are shown. Based on their location in the second helix, the amino acids in the putative TRM (i.e., amino acids 1, 2 and 5 of the five-amino acid RNA binding motif on the inner facing helix) are identified. The putative repeats are numbered with their corresponding TRM (red). The three digit numbers (black) refer to the linear position of the specific amino acid (single letter amino acid code) within the protein. Blue lines point to individual amino acids on the inner facing alpha-helix.

A

APUM23

		12345
R1	(138)	SHVLQ
R2	(174)	SHVAE
R3	(222)	SHVLR
* R3′	(264)	DNLLE
R4	(303)	SLVLQ
R5	(360)	SHLVE
R6	(397)	NFVIQ
R7	(433)	SGVVA
R8	(503)	CLILQ
R9	(542)	ARVIE
R10	(579)	SFTVE
* R10′	(628)	SRPDQ

B

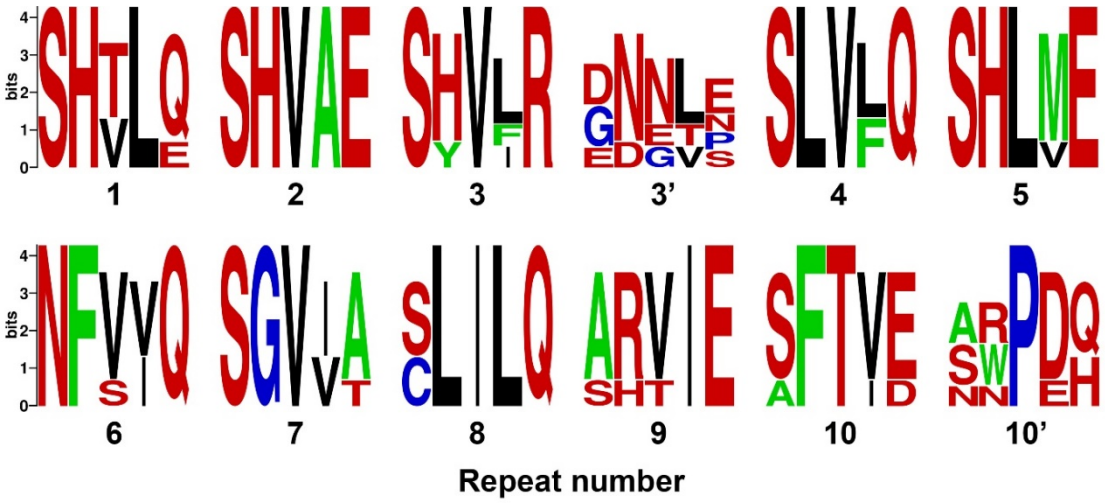
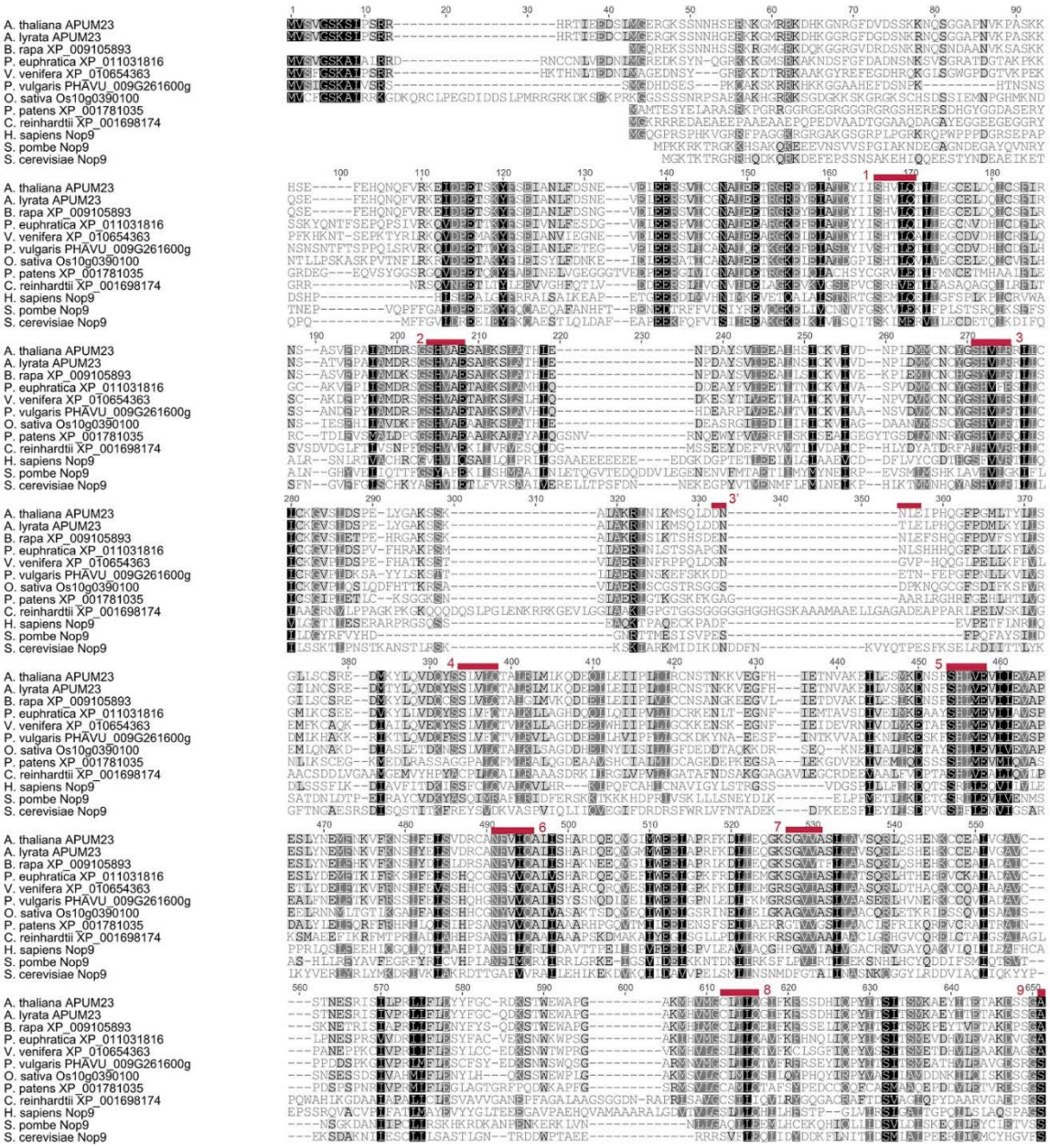


Figure 3.4 The five-amino acid sequence motifs from each of the predicted APUM23 Puf repeats. (A) The complete five-amino acid sequence motifs are shown. The numbers in parentheses refer to the position of the first amino acid from each motif in the APUM23 amino acid sequence. Shaded letters identify the amino acids at motif positions 1, 2, and 5 that are predicted to provide nucleotide contacts. Repeats 3' and 10' (*) are repeats that were identified in the APUM23 model as possible Puf repeats (Figure 3.3) but do not possess TRMs that are conserved within dicotyledonous plants and thus are not considered to be nucleotide-interacting Puf repeats. (B) Conservation of the five-amino acid motifs in the predicted Puf repeats of APUM23 orthologs. Logo graphs of the five-amino acid motifs involved in nucleotide binding from the predicted Puf repeats of APUM23 orthologs consisting of five dicotyledonous plant species shown in the alignment in Figure 3.5 (*Arabidopsis lyrata, Brassica rapa, Populus euphratica, Vitus venifera, and Phaseolus vulgaris*). The first, second, and fifth amino acids in each logo graph correspond to the predicted TRMs.



(Continued on next page)

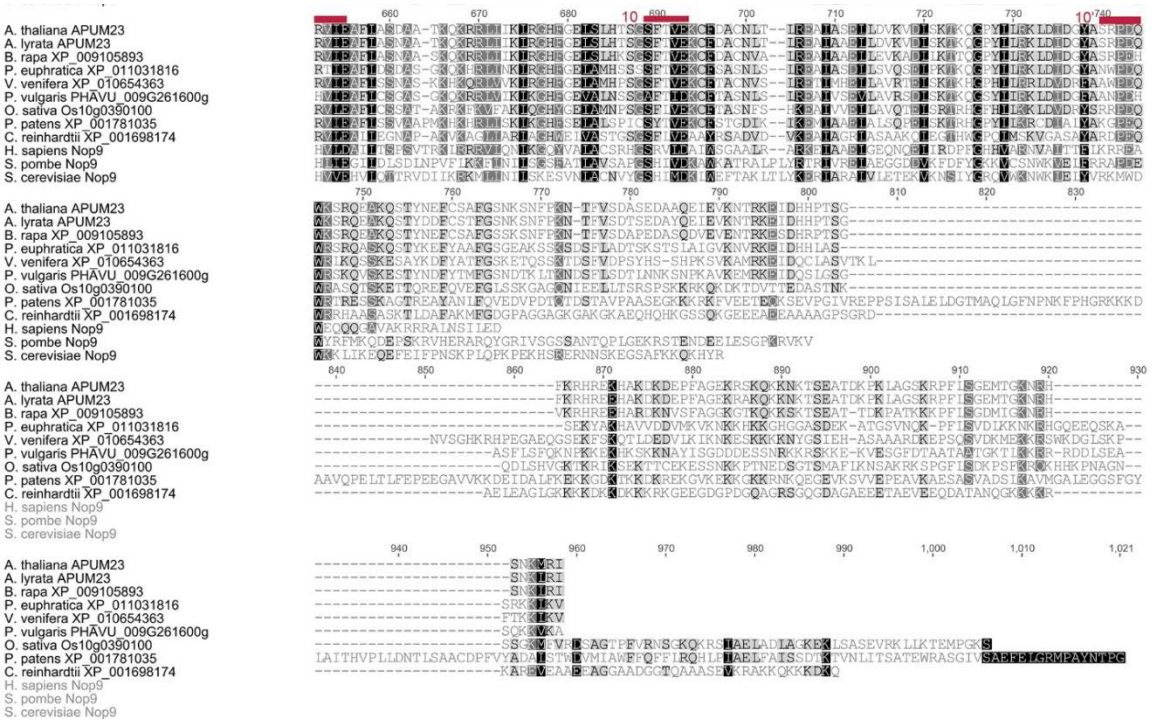


Figure 3.5 Sequence alignments of the APUM23 orthologs from representative species. Sequences orthologous to *Arabidopsis thaliana* APUM23 proteins are: the plant species, *Arabidopsis lyrata*, *Brassica rapa*, *Populus euphratica*, *Vitus venifera*, *Phaseolus vulgaris*, *Oryza sativa*; moss, *Physcomitrella patens*; algal, *Chlamydomonas reinhardtii*; human, *Homo sapiens*; yeast, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. The five-amino acid motifs within each of the predicted Puf repeats are identified by the red bars located above the sequence. The level of shading reflects the percentage of amino acid similarity at any position. The alignment was constructed using Geneious software (trial version 8.1.5).

within other dicotyledonous plants (Figure 3.4 and 3.5), and the combination of proline and aspartic acid at the third and fourth position (Figure 3.4A) of its predicted five-amino acid motif is atypical. Aside from repeats 3' and 10', the remaining repeats displayed high conservation in APUM23 orthologs from other species (Figures 3.4 and 3.5).

APUM24 is an ortholog of human Puf-A and yeast Puf6. The TRM predictions of APUM24 were originally based on Puf repeat sequence conservation (Tam et al. 2010) and modified recently based on the crystal structure that was determined for its human ortholog Puf-A (Qiu et al. 2014). The arrangement of the Puf repeats in APUM24 is predicted to consist of a cluster of eight centrally located Puf repeats and three clustered Puf repeats in C-terminal region (Figure 3.2). The TRMs in Puf-A (and the predicted TRMs of APUM24) are largely dissimilar from those other PUF proteins. Puf-A was shown to lack RNA binding specificity presumably due to the identity of its atypical TRMs (Qiu et al. 2014).

3.2.2 SELEX analysis identified high confidence consensus RNA binding sequences for four APUM proteins

SELEX analysis was performed to determine a consensus RNA binding sequence for the seven APUM proteins chosen for this study. Between six and twelve rounds of *in vitro* selection were carried out for the recombinant GST-APUM proteins (Figure 3.6). The round of selection that demonstrated the highest percentage of RNA binding was used for PCR amplification, cloning and sequencing. The RNA enrichment profiles indicated that the APUM2, 6, 12 and 23 experiments were successful in that there was a progressive increase in bound RNA at a high saturation level (>10%). The APUM13 and 18 enrichment profiles showed a low level of RNA binding at saturation (<1% and <3%, respectively,

Figure 3.6). This suggested that there was low activity of the recombinant protein, thus decreasing the confidence of the SELEX results for these proteins. The enrichment profile for APUM24 reached almost 40% after the third round of SELEX indicating that there was a non-specific enrichment of RNA (Figure 3.6). Indeed, sequencing of cloned products of the APUM24 SELEX run (Appendix 2) and analysis by MEME indicated that the results of this experiment did not generate a consensus RNA binding sequence for APUM24. Since high confidence SELEX results were obtained for APUM2, 6, 12 and 23, a more detailed analysis was performed on these proteins only. The SELEX results of APUM 13, 18 and 24 are discussed later in this chapter.

The consensus RNA binding sequences that were obtained for APUM2, 6, 12 and 23 were assembled into logo graphs (Figures 3.7, see Experimental Procedures for details on consensus sequence determination and logo graph assembly). The APUM2 and APUM6 consensus sequences consisted of eight nucleotides, whereas the APUM12 and APUM23 sequences consisted of nine and ten nucleotides, respectively. A 5' UGUA core was present in the APUM2, 6 and 12 RNA sequences. Subtle degeneracy was observed at nucleotide position 4 in the sequences of APUM2 and APUM6, whereas a higher degree of degeneracy was observed at position 5 in these sequences. Degeneracy at position 5 is common for many other PUF protein targets (Miller and Olivas 2011). APUM12 deviated from the classical one repeat-one nucleotide pattern for PUF binding, as it showed an insertion of an additional uracil near the 3' end of its consensus RNA sequence when compared to the target sequences of APUM2 and APUM6 (Figure 3.7). The APUM23 consensus sequence of 10 nucleotides matched the predicted number of TRMs in this protein (Figure 3.7D, Appendix 2). This RNA target lacked the typical 5' UGUA core,

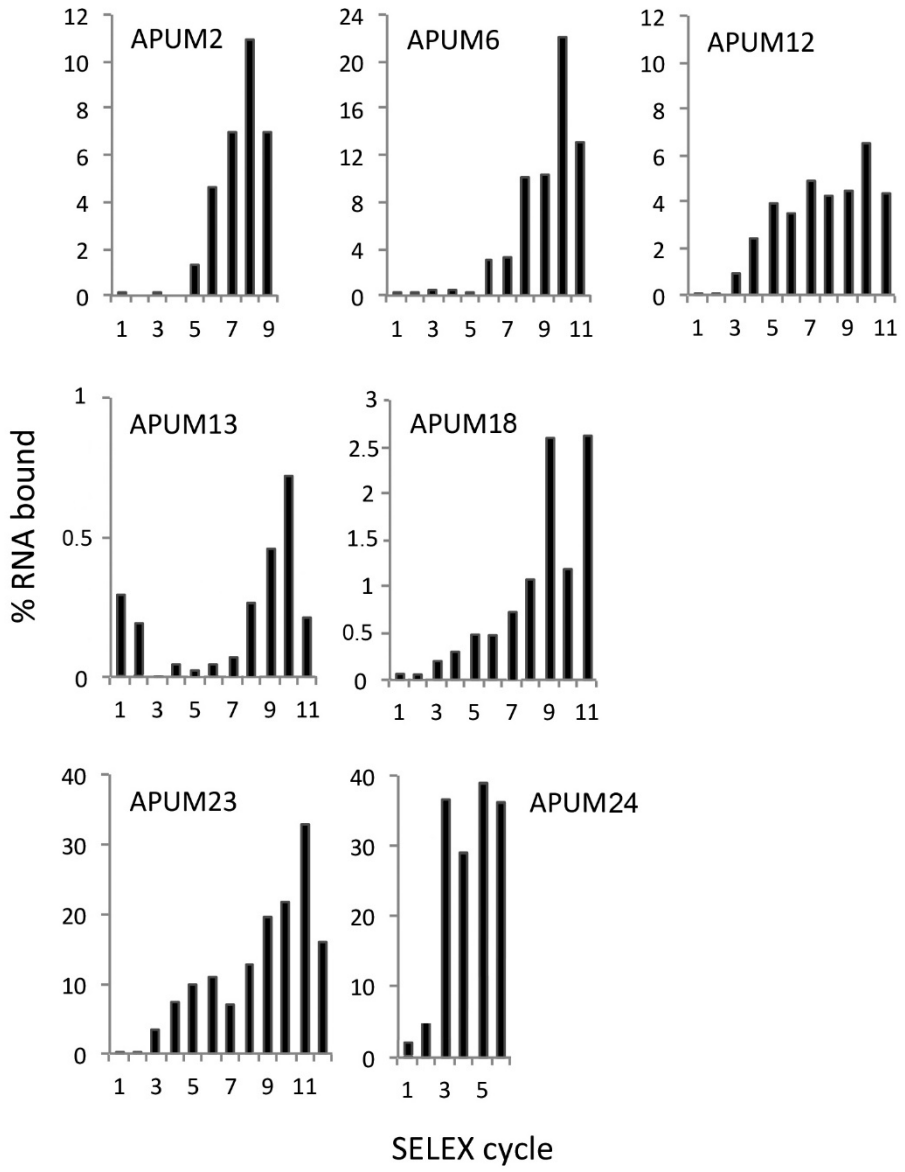


Figure 3.6 RNA enrichment profiles from SELEX experiments. SELEX assays were performed using the PUM-HDs of APUM 2, 6 and 12, and the full-length proteins of APUM18, 23, 24. The percentage of RNA that bound to the PUF-coupled matrix was determined after each cycle of enrichment. DNA resulting from the cycle that had the highest percentage of bound RNA was used for cloning and sequencing.

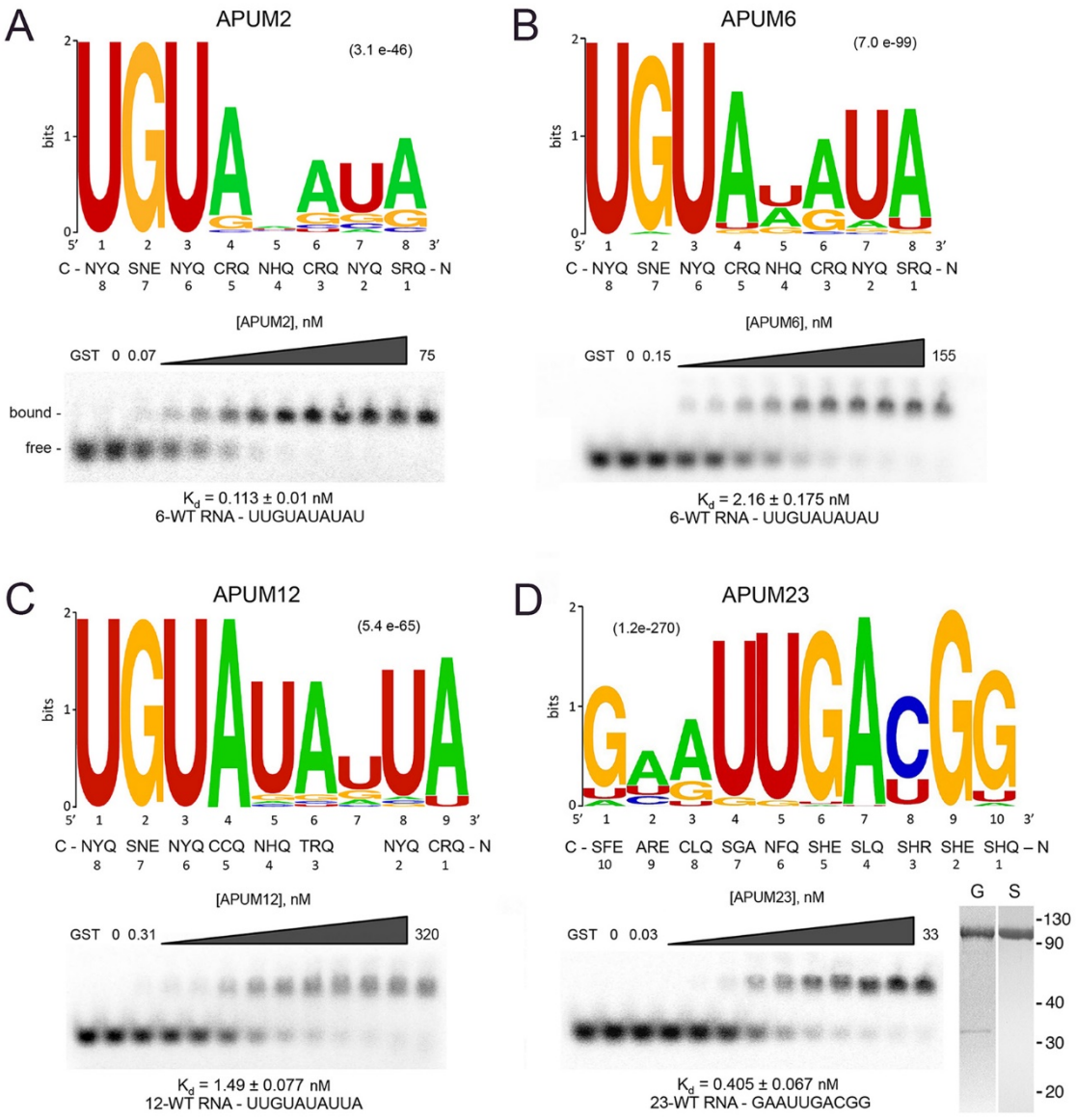

A
APUM2
(3.1 e-46)
bits
C - NYQ SNE NYQ CRQ NHQ CRQ NYQ SRQ - N
[APUM2], nM
GST 0 0.07 75
bound -
free -
K_d = 0.113 ± 0.01 nM
6-WT RNA - UUGUAUAUAU
B
APUM6
(7.0 e-99)
C - NYQ SNE NYQ CRQ NHQ CRQ NYQ SRQ - N
[APUM6], nM
GST 0 0.15 155
K_d = 2.16 ± 0.175 nM
6-WT RNA - UUGUAUAUAU
C
APUM12
(5.4 e-65)
C - NYQ SNE NYQ CCQ NHQ TRQ NYQ CRQ - N
[APUM12], nM
GST 0 0.31 320
K_d = 1.49 ± 0.077 nM
12-WT RNA - UUGUAUAUUA
D
APUM23
(1.2e-270)
C - SFE ARE CLQ SGA NFQ SHE SLQ SHR SHE SHQ – N
[APUM23], nM
GST 0 0.03 33
K_d = 0.405 ± 0.067 nM
23-WT RNA - GAAUUGACGG
G S
130
90
40
30
20

Figure 3.7 SELEX consensus RNA target sequences APUM2, 6, 12 and 23. Logo graphs of the RNA SELEX products for APUM2 (A), APUM6 (B), APUM12 (C), and APUM23 (D) are shown. The numbers below the letters in the logo graph refer to the nucleotide position in the consensus sequence (5' to 3' direction). The *E*-value of the consensus sequence is indicated in parentheses above the logo graph. The TRMs of the predicted Puf repeats are shown below their corresponding nucleotide. The uracil at nucleotide position 7 in the APUM12 logo graph (C) is not aligned with a Puf repeat, because this nucleotide is likely extruded from the RNA binding surface of the protein (see text). The Puf repeats are numbered in reverse, because the binding of Puf proteins to RNA is anti-parallel. Representative mobility shift assays of each APUM protein are shown below the corresponding logo graph. A GST sample served as a negative control. The lowest and highest concentration of protein in the exponential dilution series is indicated above each gel. Protein concentrations are corrected so that they reflect the concentration of active protein in each sample. The average apparent dissociation constant value (K_d) for the protein bound to the RNA consensus sequence is shown below each gel, as is the sequence of the cognate RNA used. The bottom right panel in D shows Coomassie Blue-stained lanes from SDS-PAGE gels that show GST-APUM23 protein that was purified using consecutive glutathione affinity chromatography alone (G) or both glutathione affinity chromatography and size exclusion chromatography steps (S). Gel marker lines are in kilodaltons.

but rather contained a conserved four nucleotide ‘core’ (UUGA) that was located centrally in the consensus sequence. It also preferred cytosine at nucleotide position 8 and showed strong binding to guanine at positions 1, 9 and 10. APUM2, 6, 12 and 23 showed high affinity binding to their cognate RNA probes in electrophoretic mobility shift assays (EMSAs), with apparent dissociation constants ranging from 0.113 nM to 2.16 nM (Figure 3.7).

Nucleotides in the RNA consensus sequences of each APUM protein were aligned with a predicted Puf repeat that was represented by its TRM (Figure 3.7). The alignment of nucleotides with TRMs from APUM2 and APUM6 was of high confidence, since their TRMs and RNA consensus sequences are conserved with those of human PUM1. The APUM12 target was also quite conserved. However, the one additional nucleotide in its RNA could be attributable to the extrusion of one of the bases. Yeast Puf4 and APUM12 have identical TRMs and identical nine nucleotide RNA consensus sequence. In the Puf4-RNA complex, the uracil at nucleotide position 7 is flipped away from the binding surface (Miller et al. 2008; Yeming Wang et al. 2009). The conserved TRMs and target sequences of Puf4 suggest that a similar extrusion of uracil from the RNA-binding surface exists for APUM12, as is reflected in the TRM alignment shown in Figure 3.7C.

The alignment of the APUM23 Puf repeats with the nucleotides in its cognate RNA target is supported, in part, by the observation that cytosine at position 8 aligns with Puf repeat 3, a repeat that contains a predicted cytosine binding TRM (SHR, Figure 3.7D) (Dong et al. 2011; Filipovska et al. 2011). TRMs that are predicted to be natural cytosine binders are rare, and the cytosine binding preference by this TRM has only

been demonstrated in an engineered PUF protein (Dong et al. 2011). When considering the base-specifying amino acids at positions 1 and 5 in each TRM, eight of the ten TRM:nucleotide pairing assignments for APUM23 (TRMs 1 to 6, 8 and 10, Figure 3.7D) are conserved with previously observed nucleotide interactions for TRMs (Dong et al. 2011; Filipovska et al. 2011; Campbell et al. 2014). The nucleotide targets of the remaining two Puf repeats (repeat 7, SGA; repeat 9, ARE) have not been characterized elsewhere. Consistent with its role in 18S rRNA processing, the preferred APUM23 RNA binding sequence (GAAUUGACGG) is present at nucleotide position 1142 in the 18S rRNA sequence, suggesting a direct binding of APUM23 to the mature or unprocessed 18S rRNA (Figure 3.8).

The identification of an atypical RNA consensus sequence for APUM23 led us to determine whether an ortholog of APUM23 also showed specificity for the APUM23 RNA sequence. APUM23 and its *Saccharomyces cerevisiae* ortholog (Nop9) are nucleolar proteins that have a conserved role in 18S rRNA processing (Thomson et al. 2007; Abbasi et al. 2010). However, Nop9 shares only 20% amino acid identity (39% similarity) with APUM23 (Figure 3.5), and only three of the predicted TRMs present in APUM23 are conserved in Nop9 (Figure 3.9A). Nop9 showed very low binding affinity to the preferred ten-nucleotide APUM23 consensus sequence (K_d = 605 nM) (Figure 3.9B). We then performed a Nop9 SELEX experiment to determine its sequence preference. SELEX enrichment of RNA binding achieved maximal levels after only the third round of selection (Figure 3.9C). Sequence analysis of the Nop9 SELEX products (Appendix 2) and MEME analysis did not identify any RNA consensus sequence motif for Nop9, indicating that Nop9 lacks RNA sequence specificity *in vitro*.

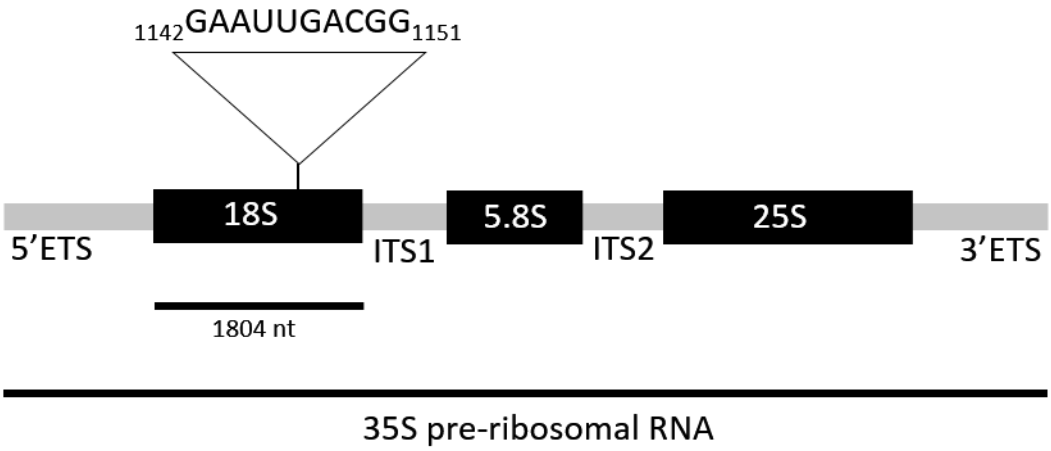


Figure 3.8 Diagrammatic representation of the APUM23 consensus sequence location in the 18S ribosomal RNA sequence.

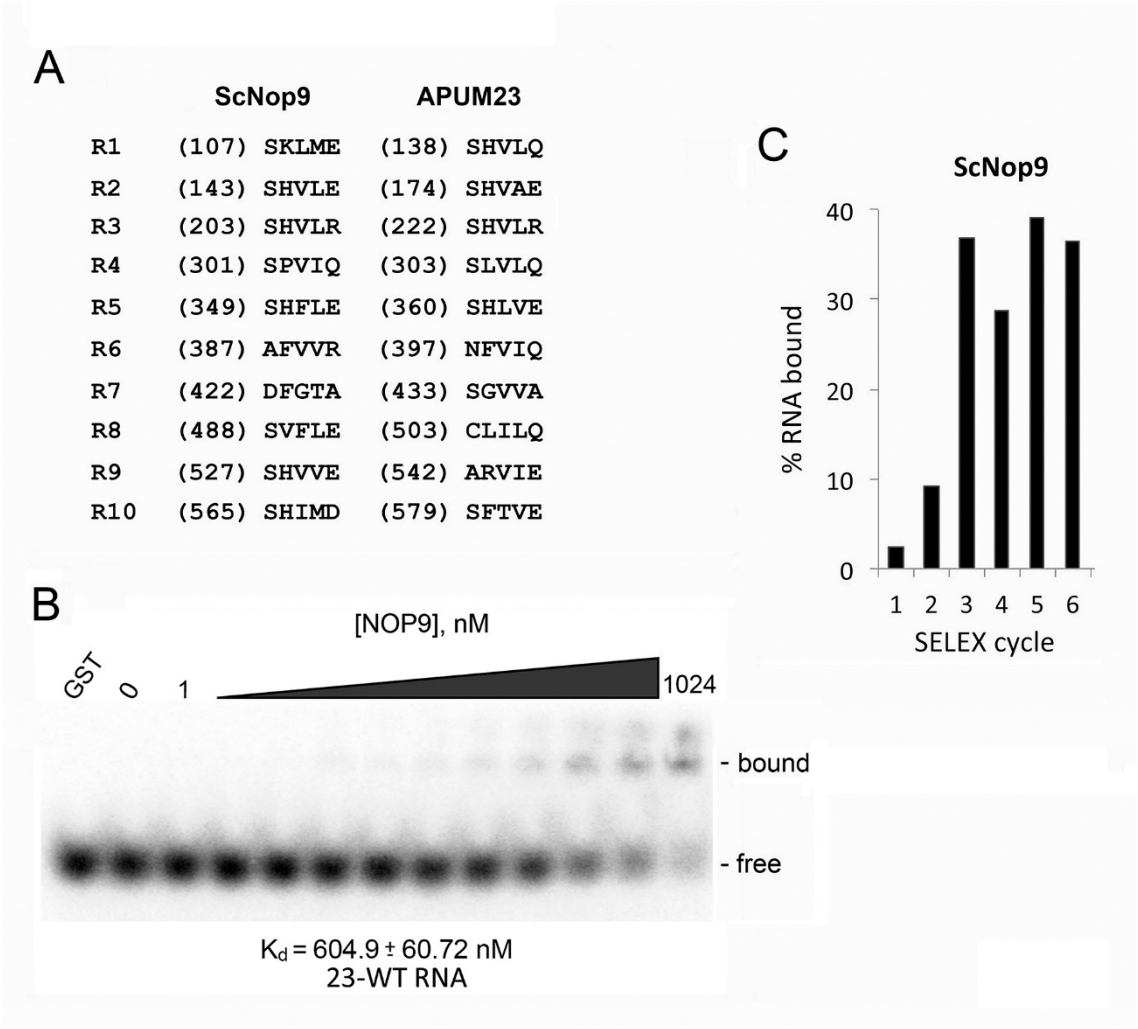


Figure 3.9 Five-amino acid motif in Nop9 Puf repeats, representative EMSA and SELEX enrichment profile. (A) comparison of the *S. cerevisiae* Nop9 and APUM23 five-amino acid RNA-binding motifs from their predicted Puf repeats. The position of the first amino acid in each motif within the corresponding Puf repeat is indicated in parentheses. (B) representative mobility shift gel for the Nop9 interaction with the preferred APUM23 RNA target sequence (23-WT RNA). The details are as described in Figure 3.7. (C) RNA enrichment profiles from the Nop9 SELEX experiment. The percentage of RNA that bound to the APUM-coupled matrix was determined after each cycle of enrichment.

3.2.3 Validation of the APUM2, 6, 12 and 23 RNA consensus sequences by EMSA

To validate the APUM2, 6, 12 and 23 RNA consensus sequences that were derived from the SELEX experiments, EMSAs using nucleotide substituted RNAs were performed. RNAs consisting of ten nucleotides were used in these assays, as this was the length of the longest consensus sequence. The sequence UGUAUAUA was used as the 'wild-type' cognate RNA sequence (6-WT) for APUM2 and APUM6 (Figure 3.7A and B), and was flanked on either end by uracil (Table 3.1). The central eight nucleotides matched the preferred consensus RNA sequence for APUM6. This sequence is present in the Nanos Response Element (NRE1) RNA that is bound by Drosophila Pumilio (Gupta et al. 2009). The cognate wild-type RNA for APUM12 (12-WT, Table 3.1) differed from the APUM6 wild-type RNA in that the last two nucleotides were swapped, matching the nine-nucleotide preferred consensus sequence for APUM12 (Figure 3.7C).

A substitution in the conserved UGUA core sequence (6-G2U) resulted in a large decrease in the affinity of APUM2, APUM6 and APUM12 to this RNA, to the point that the apparent dissociation constant could not be determined with the amount of protein used in our assays (Table 3.1). Substitutions at positions outside of the core (positions 5 and 8) showed smaller changes in binding affinity. A substitution at the variable nucleotide position 5 (6-U5A) resulted in a small increase in affinity to APUM2 and APUM6 (Table 3.1, Figure 3.10A and B), perhaps owing to subtle differences in binding in the SELEX versus EMSA assays. The APUM12 consensus sequence showed a strong preference for UA at its final two positions (Figure 3.7C). When these nucleotides were substituted with AU (i.e., the APUM6 wild-type probe, 6-WT) to conform to the typical 8-nucleotide PUF consensus sequence, there was a greater than four-fold drop in the binding affinity of

Table 3.1 Actual and relative apparent dissociation constants (K_d) of APUM2, APUM6 and APUM12 when bound to wild-type and nucleotide substituted RNA

Protein	RNA	RNA sequence	K_d (nM)	K_{rel}[1]
		1 2 3 4 5 6 7 8		
APUM2	6-WT	U**UGUA**UAUAU[2]	0.11 ± 0.01	1
	6-U5A	UUGUAAAUAU[3]	0.07 ± 0.01	0.64
	12-WT	UUGUAUAUUA	0.39 ± 0.05	3.4
	6-G2U	UUUUAUAUAU	>74.0[4]	>654
APUM6	6-WT	UUGUAUAUAU	2.16 ± 0.17	1
	6-U5A	UUGUAAAUAU	1.29 ± 0.23	0.60
	12-WT	UUGUAUAUUA	23.4 ± 3.60	11
	6-G2U	UUUUAUAUAU	>155	>71
APUM12	12-WT	UUGUAUAUUA	1.49 ± 0.08	1
	6-WT	UUGUAUAUAU	6.30 ± 0.69	4.2
	6-U5A	UUGUAAAUAU	13.8 ± 2.45	9.3
	6-G2U	UUUUAUAUAU	>318	>213

1 K_{rel} values represent the relative affinity of the specific APUM protein to wild-type or mutant RNAs.

2 The UGUA core of the wild-type RNA is highlighted with bold type. 6-WT and 12-WT refer to the SELEX consensus sequences for APUM6 and APUM12, respectively.

3 Substituted nucleotides are underlined.

4 The K_d value could not be determined because RNA binding was not observed for the maximal protein concentration used for this RNA-protein combination. Thus, the K_d value is shown as being greater than the highest concentration of protein used.

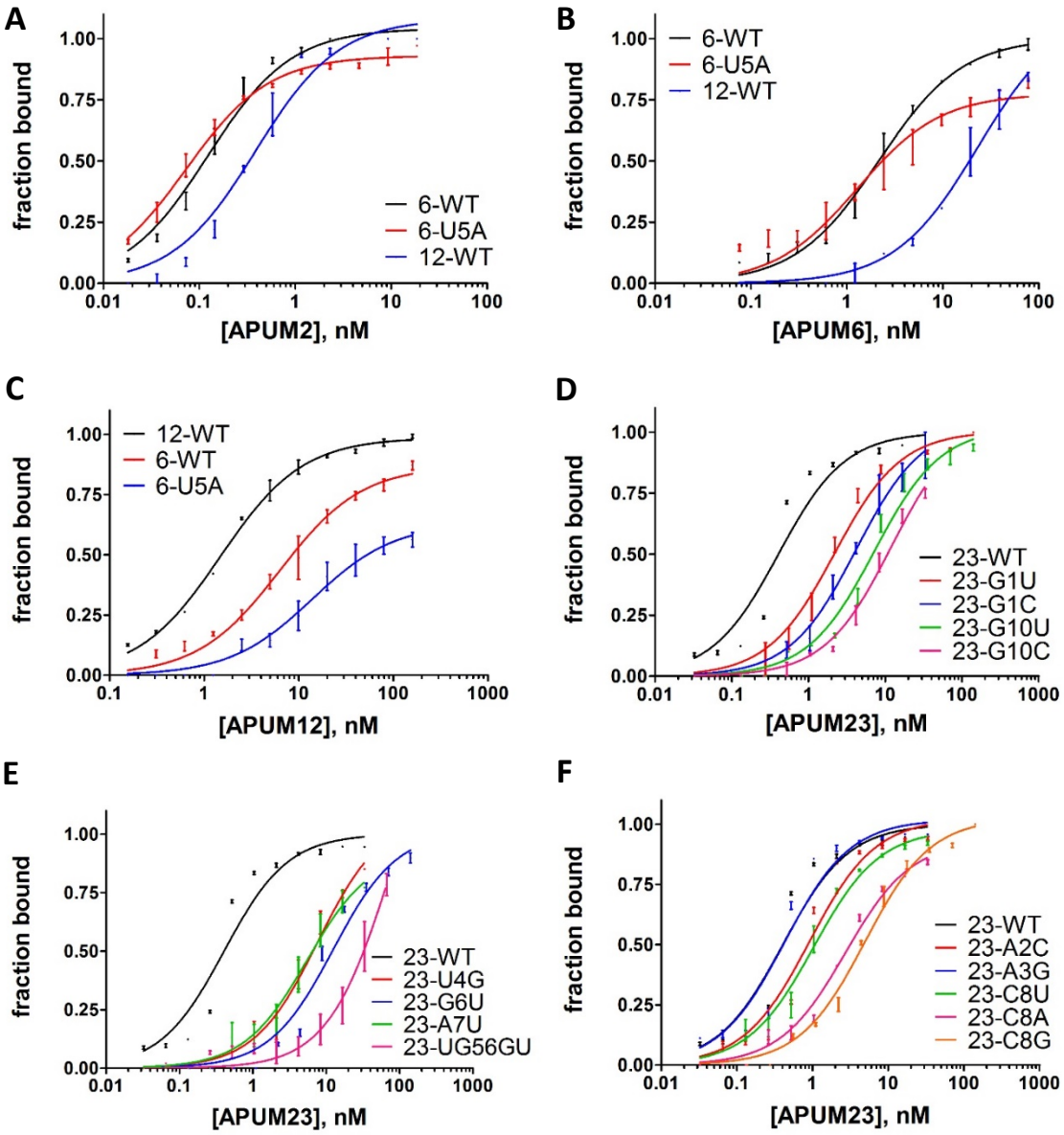

A
fraction bound
1.00
0.75
0.50
0.25
0.00
0.01
0.1
1
10
100
[APUM2], nM
6-WT
6-U5A
12-WT
B
fraction bound
[APUM6], nM
6-WT
6-U5A
12-WT
C
fraction bound
[APUM12], nM
12-WT
6-WT
6-U5A
D
fraction bound
1000
[APUM23], nM
23-WT
23-G1U
23-G1C
23-G10U
23-G10C
E
fraction bound
[APUM23], nM
23-WT
23-U4G
23-G6U
23-A7U
23-UG56GU
F
fraction bound
[APUM23], nM
23-WT
23-A2C
23-A3G
23-C8U
23-C8A
23-C8G

Figure 3.10 RNA binding profiles for wild-type APUM proteins to wild-type (WT) and nucleotide-substituted RNAs. (A) to (C) Binding curves of APUM2, 6 and 12 proteins to 6-WT, 6-U5A and 12-WT RNA oligonucleotides (RNA sequences are described in Table 3.1). (D) to (F) Binding curves of APUM23 to various nucleotide substituted RNAs (RNA sequences are described in Table 3.2). (D) Binding curves of APUM23 to 23-WT and nucleotide substituted RNAs at the consensus boundary positions. (E) APUM23 to 23-WT and nucleotide substituted RNAs within UUGA core. (F) APUM23 to 23-WT and nucleotide substituted RNAs at positions 2 and 3, as well as substitutions of the cytosine at nucleotide position 8. The fraction of RNA bound is shown on the vertical axis and the concentration of protein is shown on the horizontal axis. The curves were processed by Prism 5 to fit in the equation of "One site - specific binding". The vertical bars on the data points indicate the ± SEM value.

APUM12 to the modified RNA (Table 3.1, Figure 3.10C). An additional substitution at position 5 (6-U5A) in the RNA reduced the binding affinity of APUM12 another 2.2-fold (Table 3.1, Figure 3.10C). Overall, APUM 2, 6 and 12 are typical PUF proteins in that there are eight tandem Puf repeats predicted from its amino acid sequence and their target RNA possesses a conserved UGUA core. APUM12 differs in that it prefers 9 nucleotides rather than the typical 8 nucleotides.

Thirteen nucleotide-substituted RNAs were used in mobility shift experiments to validate the SELEX-derived consensus target sequence for APUM23. The binding affinity of APUM23 to these base-substituted RNAs showed a high correlation with the nucleotide composition of its RNA consensus sequence shown in Figure 3.7D. Single base substitutions within the UUGA core showed 14- to 31-fold decrease in APUM23 binding affinity compared to the wild-type RNA (Table 3.2, Figure 3.10E). A double substitution in the UUGA core (UG56GU) that mimicked the classical UGUA core showed close to 200-fold reduction in binding affinity to APUM23. RNAs with substitutions at positions outside the core also validated the APUM23 consensus RNA target sequence. For instance, the consensus sequence showed a greater preference for cytosine than uracil at position 8 (Figure 3.7D), and a uracil substitution at this position (C8U) resulted in a 2.5-fold reduction in affinity of the protein to this RNA (Table 3.2, Figure 3.10F). Adenine and guanine substitutions at this position (C8A and C8G) showed 6.7- and 12-fold reductions in affinity, supporting the lack of adenine or guanine at this position in the consensus sequence (Figure 3.7D). These data indicate that cytosine is the preferred nucleotide at position 8. Changes at nucleotide position 10 resulted in relatively high reductions in affinity (18- and 29-fold reduction), whereas changes at nucleotide

Table 3.2 Actual and relative apparent dissociation constants of wild-type APUM23 protein when bound to wild-type and nucleotide substituted RNA

Protein	RNA	RNA sequence	K_d (nM)	K_{rel}[1]
		1 2 3 4 5 6 7 8 910		
APUM23	WT	GAA**UUGA**CGG[2]	0.41 ± 0.07	1
	G1U	UAAUUGACGG[3]	2.18 ± 0.23	5.4
	G1C	CAAUUGACGG	4.19 ± 0.64	10
	A2C	GCAUUGACGG	0.91 ± 0.14	2.2
	A3G	GAGUUGACGG	0.43 ± 0.07	1.1
	U4G	GAAGUGACGG	7.68 ± 1.25	19
	UG56GU	GAAUGUACGG	77.4 ± 31.7	191
	G6U	GAAUUUACGG	12.4 ± 1.88	31
	A7U	GAAUUGUCGG	5.54 ± 0.97	14
	C8U	GAAUUGAUGG	1.03 ± 0.12	2.5
	C8A	GAAUUGAAGG	2.72 ± 0.26	6.7
	C8G	GAAUUGAGGG	4.94 ± 0.41	12
	G10U	GAAUUGACGU	7.16 ± 0.80	18
	G10C	GAAUUGACGC	11.9 ± 1.51	29
	Human PUM1	UGUAUAUA	>140[4]	>346

1 K_{rel} values represent the relative affinity of wild-type protein to wild-type or mutant RNAs.

2 The UUGA central core in the APUM23 wild-type RNA is highlighted with bold type.

3 Modified nucleotides are underlined. The Human PUM1 RNA is the Human PUMILIO1 consensus sequence.

4 K_d value could not be determined because RNA binding was not observed for the maximal protein concentration used for this RNA-protein combination. Thus, the K_d value is shown as being greater than the highest concentration of protein used.

positions 1, 2 and 3 showed a lower reduction in affinity (1.1- to 10-fold decrease). APUM23 did not possess a measurable apparent dissociation constant with the classical 8-nucleotide PUF target (UGUAUAUA, Human PUM1) due to its low affinity binding (Table 3.2). Coupled with the UG56GU results discussed above, this indicates that the UGU core, a component of the RNA targets of all other characterized Puf proteins, is not a preferred core in the RNA target of APUM23. Overall, these APUM23 mobility shift data validate the identification a newly described, 10-nucleotide PUF consensus RNA target sequence that contains a unique central UUGA core and a preference for cytosine at nucleotide position 8.

3.2.4 Motif swapping of APUM23 Puf repeats to alter binding specificity

In an attempt to determine whether the APUM23 Puf repeats could be modified to alter their base specificity, we engineered the protein by swapping the five amino acid RNA-binding motif in three of the predicted Puf repeats. Successfully altering base specificity would provide further support for the interaction of these repeats with their corresponding nucleotides shown in Figure 3.7D. We modified Puf repeat 1 to determine if it binds to the consensus sequence boundary at position 10, Puf repeat 3 to confirm binding to cytosine at nucleotide position 8, and Puf repeat 5 to confirm binding to the core guanine nucleotide at position 6. The five-amino acid motif sequence NYVIQ was chosen for substitution into each of these repeats so as to alter the binding preference of the substituted repeat to uracil. NYQ is a strong uracil binding TRM in other PUF proteins (Cheong and Hall 2006; Campbell et al. 2014), and its five-amino acid motif is present in two Puf repeats that are predicted to bind with uracil in the consensus RNAs of

APUM2 and APUM6 (Figure 3.7). A TRM with strong guanine binding preference (SNE) (Cheong and Hall 2006; Campbell et al. 2014) was also substituted into Puf repeat 3 to provide additional supporting evidence that this is the repeat that normally binds to cytosine at nucleotide position 8. SNVVE was used as the substituted motif, as this amino acid sequence was present at Puf repeat 7 in APUM2 and APUM12 and corresponded with guanine in the UGU core of their respective target RNAs. Substitutions in predicted Puf repeat 1 (aligns with a guanine at consensus target sequence position 10, Figure 3.7D) involved a switch from SHVLQ to NYVIQ (SHVLQ-1-NYVIQ) in an attempt to alter the binding specificity of this repeat to uracil (Figure 3.11A). In the nucleotide substitution assays described earlier, wild-type APUM23 showed an 18-fold decrease in affinity to G10U RNA compared to its affinity to its cognate wild-type RNA (Table 3.2). In contrast, the SHVLQ-1-NYVIQ substitution in Puf repeat 1 showed a 17.9-fold increase in affinity to the G10U RNA compared to wild-type RNA (K_{rel} = 0.056)(Table 3.3, Figure 3.11B). Two motif substitutions were made in predicted Puf repeat 3 (Figure 3.11A), the repeat that aligns with the preferred cytosine base at nucleotide position 8. The SHVLR-3-NYVIQ substitution mutant demonstrated a 4.4-fold increase in binding affinity to C8U RNA (K_{rel} = 0.23, Figure 3.11C, Table 3.3) compared to wild-type RNA, whereas the wild-type protein was shown to have a corresponding decrease in affinity (2.5-fold) to the C8U RNA over its cognate RNA (Table 3.2). The SHVLR-3-SNVVE mutant bound more than three-fold more tightly to its cognate RNA (C8G) than to wild-type RNA (Figure 3.11D, Table 3.3), whereas the wild-type protein had a 12- fold decrease in binding to C8G compared to wild type RNA (Table 3.2). A substitution in Puf repeat 5 (SHLVE-5-NYVIQ) showed enhanced binding to the G6U RNA over wild-type RNA (~1.5-fold) (Figure 3.11E, Table

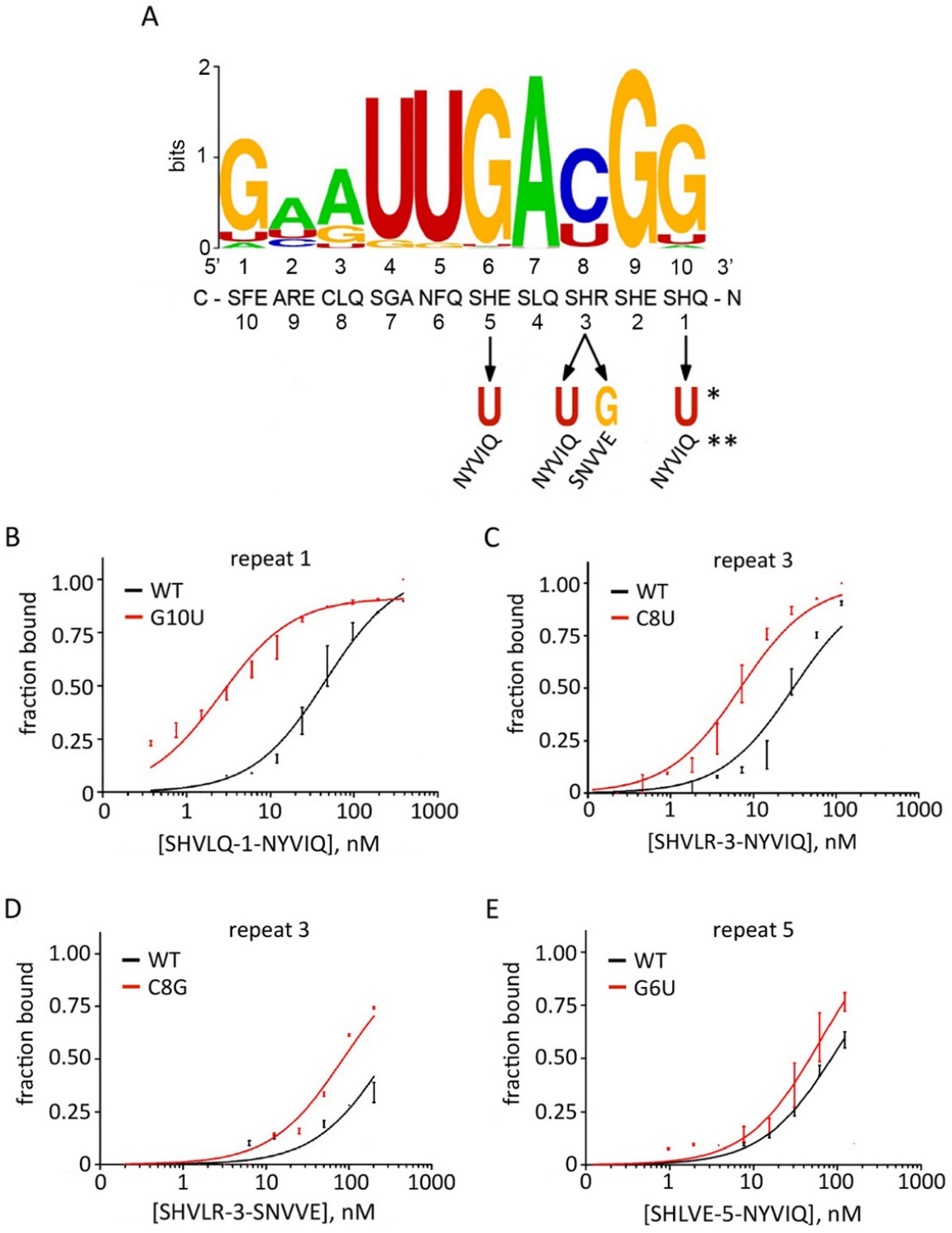

A
bits
5' 1 2 3 4 5 6 7 8 9 10 3'
C - SFE ARE CLQ SGA NFQ SHE SLQ SHR SHE SHQ - N
10 9 8 7 6 5 4 3 2 1
NYVIQ
NYVIQ SNVVE
NYVIQ **
B
repeat 1
WT
G10U
fraction bound
[SHVLQ-1-NYVIQ], nM
C
repeat 3
WT
C8U
fraction bound
[SHVLR-3-NYVIQ], nM
D
repeat 3
WT
C8G
fraction bound
[SHVLR-3-SNVVE], nM
E
repeat 5
WT
G6U
fraction bound
[SHLVE-5-NYVIQ], nM

Figure 3.11 RNA binding analysis for wild-type and mutated APUM23 bound to wild-type and nucleotide-substituted RNAs. A, logo map and predicted nucleotide preference for the five-amino acid motif swapping experiments. The labeling of nucleotide and TRM positions are as in Figure 3.7D. The predicted Puf repeats that were selected for mutagenesis are identified below the three-letter TRM of the specific Puf repeat. The expected nucleotide bound by the mutated Puf repeat (*) and the substituted five-amino acid motif in the Puf repeat (**) are shown. B–E, representative equilibrium binding data of substituted APUM23 Puf repeat motifs to their cognate (red lines) and wild-type (black lines) sequences. The fraction of RNA bound is shown on the vertical axis, and the concentration of mutant protein is shown on the horizontal axis.

Table 3.3 Actual and relative apparent dissociation constants of mutant APUM23 proteins when bound to wild-type and nucleotide substituted RNA

Protein	Substituted repeat[1]	RNA	RNA sequence	K_d (nM)	K_{rel}[2]
			1 2 3 4 5 6 7 8 910		
SHVLQ-1-NYVIQ	1	G10U	GAAUUGACGU[3]	2.55 ± 0.34	0.056
		WT	GAAUUGACGG	45.3 ± 5.95	1
SHVLR-3-NYVIQ	3	C8U	GAAUUGAUGG	6.95 ± 0.67	0.23
		WT	GAAUUGACGG	30.4 ± 3.49	1
SHVLR-3-SNVVE	3	C8G	GAAUUGAGGG	85.3 ± 5.81	0.30
		WT	GAAUUGACGG	282 ± 30.0	1
SHLVE-5-NYVIQ	5	G6U	GAAUUUACGG	62.2 ± 16.7	0.66
		WT	GAAUUGACGG	94.5 ± 15.5	1

1 Predicted Puf repeat that was mutated.

2 K_{rel} values represent the relative affinity of mutant proteins to wild-type or mutant RNAs.

3 Modified nucleotides are underlined.

3.3), while the wild-type protein bound 31-fold less tightly to G6U RNA compared to wild-type RNA (Table 3.3). The altered affinity of these motif-swapped proteins supports our predicted interactions between Puf repeat 1, 3 and 5 and nucleotides 10, 8 and 6, respectively. These motif-substituted proteins all had lower binding affinities for their cognate RNAs (Table 3.3) compared to the wild-type protein bound to its cognate RNA (Table 3.2). This was especially noticeable for the SHVLR-3-SNVVE and SHLVE-5-NYVIQ substitutions. This indicates that, although the substitutions altered the specificity of the repeat, they may have also altered the local structure at these positions thereby leading to a reduced binding affinity.

In addition to five-amino acid motif sequence swapping, substitutions to the TRM code only were made to four Puf repeats. Previous reports that describe the engineering of Puf repeats have generally modified the three amino acids associated with the TRM rather than modifying the entire five amino acid motif in a Puf repeat (Cheong and Hall 2006). Substitutions in predicted Puf repeat 1 of APUM23 involved a switch from the TRM code SHQ to SHR (SHxxQ-1-SHxxR) in an attempt to alter the binding specificity of this repeat to cytosine. This substitution showed a 1.7-fold decrease in affinity to the G10C RNA compared to wild-type RNA (Table 3.4). In the nucleotide substitution assays described earlier, wild-type APUM23 showed a 29-fold decrease in affinity to G10C RNA compared to its affinity to the cognate wild-type RNA (Table 3.2). A TRM substitution was also made in predicted Puf repeat 3, the cytosine binding repeat (Table 3.4). The SHxxR-3-CRxxQ substitution mutant demonstrated a 1.4-fold decrease in binding affinity to C8A RNA (Table 3.4) compared to wild-type RNA, whereas the wild-type protein was shown to have a decrease in affinity (6.7-fold) to the C8A RNA over its cognate RNA (Table 3.2).

Table 3.4 Actual and relative apparent dissociation constants of mutant APUM23 proteins when bound to wild-type and nucleotide substituted RNA

Protein	Substituted repeat[1]	RNA	RNA sequence	K_d (nM)	K_{rel}[2]
			1 2 3 4 5 6 7 8 910		
SHxxQ-1-SHxxR[3]	1	G10C	GAAUUGACGC[4]	4.29 ± 0.34	1.7
		WT	GAAUUGACGG	2.55 ± 0.19	1
SHxxR-3-CRxxQ	3	C8A	GAAUUGAAGG	2.44 ± 0.31	1.4
		WT	GAAUUGACGG	1.70 ± 0.07	1
SLxxQ-4-NYxxQ	4	A7U	GAAUUGUCGG	6.58 ± 1.11	2.0
		WT	GAAUUGACGG	3.33 ± 0.20	1
SFxxE-10-SHxxR	10	G1C	CAAUUGACGG	64.8 ± 9.12	44
		A2C	GCAUUGACGG	3.88 ± 0.33	2.6
		WT	GAAUUGACGG	1.47 ± 0.18	1

1 Predicted Puf repeat that was mutated.

2 K_{rel} values represent the relative affinity of mutant proteins to wild-type or mutant RNAs.

3 xx refers to the amino acid positions 3 and 4 that are not modified in the mutant protein.

4 Modified nucleotides are underlined.

A TRM substitution in Puf repeat 4 (SLxxQ-4-NYxxQ) showed weaker binding to the A7U RNA over wild-type RNA (2-fold reduced affinity) (Table 3.4), while the wild-type protein bound 14-fold less tightly to A7U RNA compared to wild-type RNA (Table 3.2). Finally, the TRM substitution in Puf repeat 10 (SFxxE-10-SHxxR) showed significantly weaker binding to the G1C RNA over wild-type RNA (44-fold weaker) (Table 3.4), while the wild-type protein bound 10-fold less tightly to G1C RNA compared to wild-type RNA (Table 3.2). In these TRM substituted APUM23 proteins, each demonstrated a higher affinity to wild-type RNA than their cognate substituted RNA. However, the difference in binding affinity using wild type and substituted RNAs was reduced using the TRM-mutated proteins. These results, and the results presented in Table 3.3, suggest that the effective engineering of APUM23 will require swapping the entire five-amino acid motif (rather than the TRM only) in order to change the nucleotide preference of the Puf repeat.

3.2.5 Analysis of the APUM13 and 18 SELEX data

The low percentage of RNA enrichment that was observed in the SELEX analysis for APUM13 and APUM18 (Figure 3.6) suggested that these proteins possessed low RNA-binding activity. However, SELEX product cloning and sequence analysis was sufficient to produce consensus RNA binding logo graphs. Despite the low confidence level of these two SELEX results, some information can be gleaned from the consensus sequences since they do resemble typical PUF target sequences, particularly APUM13. The predicted TRM composition of Puf repeats 1, 2, 3 and 4 in APUM13 deviated significantly from the conventional PUF proteins, while the repeats 5, 6, 7 and 8 are common to many PUFs (compare the APUM13 TRMs to those of APUM2, 6 and 12 in Figure 3.2). The last four

predicted Puf repeats of APUM13 align with the typical UGUA core (Figure 3.12), whereas the first four Puf repeats show a preference for nucleotides that are typical of PUF proteins at these positions (U/AAUA, Figure 3.12). This suggests that the novel TRMs in Puf repeats 1, 2, 3 and 4 possess a specificity code for nucleotides that are similar to those bound by TRMs located at the same position in conventional PUF proteins. APUM18 also contains atypical TRMs (Puf repeats 1, 3, 4 and 7) and the SELEX-derived consensus sequence does not indicate a high specificity for nucleotides that correspond to these Puf repeats. In contrast, Puf repeats 2, 5 and 6 possess common Puf repeats and prefer nucleotide targets that are expected as described for other PUF proteins (Campbell et al. 2014). The confidence of the APUM18 consensus sequence was low due to the limited number of sequences used to generate the logo graph (Appendix 2). In addition, mobility shift experiments using recombinant APUM13 and APUM18 with the 6-WT RNA probe did not demonstrate any probe shift using a protein concentration range that was used for the other PUF proteins, likely due to low RNA binding activities of the proteins (data not shown), consistent with the low RNA enrichment data in the SELEX experiments.

3.3 Discussion

The SELEX-derived consensus RNA binding sequences of six APUM proteins together with the results that demonstrated non-specific binding of APUM24 provide insight into the diversity of RNA sequences bound by plant PUF-like proteins. The APUM2 and APUM6 SELEX-derived consensus RNA binding sequences were typical PUF target sequences, and were consistent in that the TRMs that they possesses are identical to human PUM1. APUM12 possesses eight Puf repeats with identical TRMs to that of yeast Puf4.

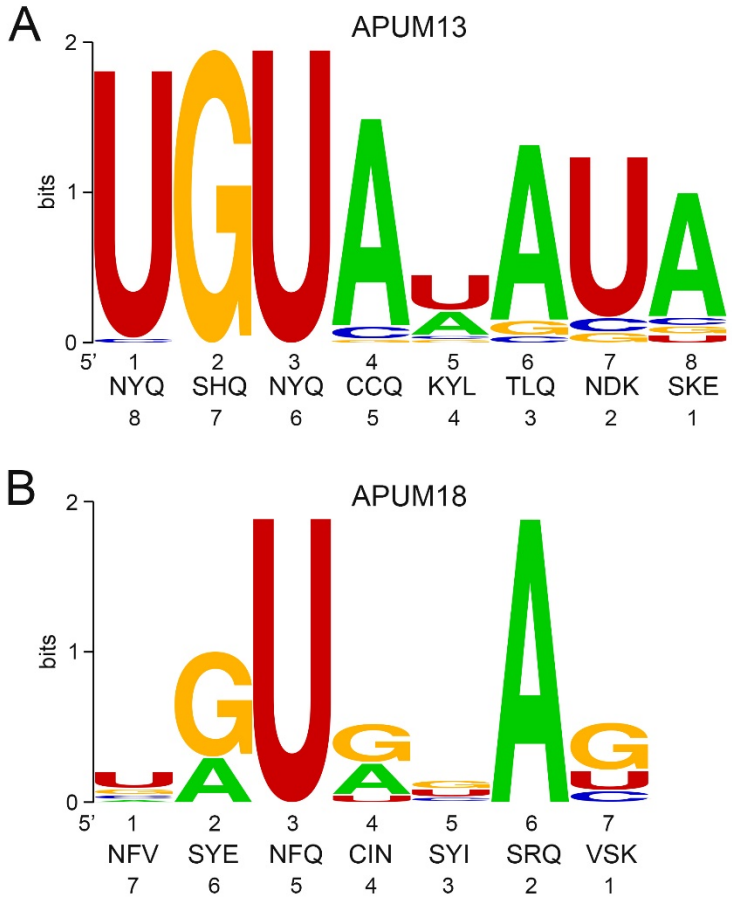


Figure 3.12 SELEX consensus RNA target sequences for APUM13 and 18. Logo graphs of the RNA SELEX products for APUM13 (A) and APUM18 (B) are shown. Details of the figure are as described in Figure 3.7. E-value of APUM13 and 18 consensuses are 3.3 e^{-38} and 3.7 e^{-18}, respectively.

Its consensus sequence was identical to that of Puf4, with an additional uracil located at a similar position to the Puf4 target (Figure 3.7C, Miller et al. 2008; Wang et al. 2009). This strongly suggests that base extrusion occurs between APUM12 and its RNA target. Extrusion of bases from the binding surface is a well-recognized feature of Puf proteins, and may provide a mechanism that allows for greater diversity in binding (Edwards 2015). The consensus binding sequence of APUM13, although of lower confidence, was the typical eight nucleotides in length. The TRMs in Puf repeats 5, 6, 7 and 8 in APUM13 are identical to those in APUM2 and APUM6, and this was consistent with its binding to the UGUA core sequence at the 5' end of the sequence (Figure 3.12). The unusual TRMs in Puf repeats 1, 2, 3 and 4 of APUM13 had nucleotide binding specificities that were typical for classical PUF proteins (U/AAUA). Although this was surprising, a structural model of the APUM13 PUM-HD reported previously by the Muench laboratory (Tam et al. 2010) showed that Puf repeats 2, 3 and 4 of APUM13 (repeat 1 could not be modeled because of low homology to the template) shared abundant hydrogen bonds and van der Waals forces with the UAU nucleotides at positions 5, 6 and 7. The bond distances within these interactions were of the typical length; however, the base stacking interactions were absent in the model (Tam et al. 2010). If indeed these base stacking interactions are absent *in vitro*, this may explain why APUM13 has a reduced affinity for its preferred RNA target as shown in the SELEX and mobility shift experiments.

Of particular interest was the identification of a novel PUF consensus sequence for APUM23. This sequence was 10 nucleotides in length, contained an atypical core element (UUGA) that was located centrally in the sequence, and preferred cytosine at nucleotide position 8. All PUF RNA target sequences studied to date contained a UGU core located

at the 5’ end of the sequence (Quenault et al. 2011). The evidence for a central UUGA core in the APUM23 consensus sequence and for the presence of a cytosine binding Puf repeat is supported by EMSA analysis that used nucleotide substituted RNAs as well as swapped RNA binding motifs. In addition to showing a preference for cytosine, Puf repeat 3 (contains a TRM code of (SHR) had a weaker preference for uracil (Figure. 3.7D and Table 3.2). This weaker preference for uracil was also observed for an engineered cytosine-binder that had a TRM code of SYR (Dong et al. 2011). Thus, our SELEX and EMSA results that showed that Puf repeat 3 prefers cytosine, coupled with the reports of an engineered cytosine binding repeat with the same TRM, indicate that repeat 3 in APUM23 has natural cytosine binding characteristics.

The identification of an RNA consensus target sequence for APUM23 provides insight into its functional roles in rRNA processing. Knock-out mutants of APUM23 have a partial defect in 35S ribosomal RNA processing that results in a small accumulation of a nonprocessed version of 18S rRNA (Abbasi et al. 2010). *Apum23* mutants displayed slower growth, shorter roots, and smaller serrated and pointed leaves (Abbasi et al. 2010; Huang et al. 2014). The *S. cerevisiae* ortholog (Nop9) has a similar 18S rRNA processing role; however, *nop9* mutants do not survive (Thomson et al. 2007). Interestingly, the 18S rRNA sequence of Arabidopsis contains a single APUM23 consensus sequence (GAAUUGACGG) at nucleotide position 1142 (Figure 3.8), as does the corresponding region in *S. cerevisiae* 18S rRNA. The frequency that this 10-nucleotide sequence is expected to appear randomly in RNA is approximately once per 1000 kb. This provides supporting evidence that APUM23 binds to the 10-nucleotide sequence in 18S rRNA. The nucleolar localization of APUM23 and Nop9 proteins also supports a direct binding to 18S

rRNA in its rRNA processing role. However, SELEX analysis of the yeast ortholog of APUM23 (Nop9) showed that this protein has no recognizable binding specificity toward RNA, and it binds to the 10-nucleotide APUM23 target with much lower affinity than APUM23. Despite having limited amino acid sequence and predicted TRM conservation, APUM23 partially complements the lethal yeast *nop9* mutant (Huang et al. 2014) suggesting that these proteins recognize similar RNA targets in their conserved rRNA processing role. Perhaps APUM23 and Nop9 both interact with 18S rRNA, but using different binding mechanisms. Nop9 might require interactions with binding partners *in vivo*, in a similar manner to that proposed for another atypical nucleolar PUF protein, yeast Puf6 (Qiu et al. 2014).

The future identification of *in vivo* targets of APUM23 will confirm that APUM23 binds to its preferred SELEX-derived RNA sequence that is present in 18S rRNA. As well, it could identify whether cytosolic mRNAs are also bound by APUM23. Also, the identification of APUM23 binding partners will determine whether APUM23 assists in recruiting RNA processing machinery, similar to the roles of some other PUF proteins. Approaches such as *in vivo* UV crosslinking and immunoprecipitation (CLIP, Ule et al. 2005) and co-immunoprecipitation of interacting proteins could be used as strategies to further characterize the function of APUM23. In this thesis research, a significant amount of effort was directed toward producing stable transgenic Arabidopsis lines that express tagged fusions of several APUM proteins (including APUM23) in order to facilitate the purification of their endogenous RNA and protein interactors. These APUM23 fusions included GFP, 6xHis-FLAG, or TAP tags under the control of the CaMV 35S constitutive promoter. However, the stable expression of these fusion proteins was never achieved at

levels sufficient for analysis. Some of the expression constructs were driven by the endogenous promoters of the selected *APUM* genes, and these never produced detectable levels of expression (Appendix 3). The 35S::GFP-APUM23 fusions showed expression in transgenic Arabidopsis plants. However, the GFP fluorescence was unusually transient and the protein appeared to be unstable *in vivo*, as was shown also in immunoprecipitation experiments with APUM23 fusions to GFP and FLAG-HIS epitope tags (Appendix 3). Perhaps APUM23 is sensitive to some form of degradation mechanism when overexpressed *in vivo*. A more effective approach to capture and preserve the *in vivo* interactions of APUM23 is necessary.

The novel RNA consensus sequence identified for APUM23, as well as its atypical arrangement of Puf repeats, suggests that it has a unique three-dimensional structure when bound to RNA. The modeled structures of APUM23 (Figure 3.3) assisted in predicting the identity of the Puf repeats in this protein. However, the structure of the protein *in vivo* is likely quite complex and differs from the classical PUF protein structure because of the presence of relatively large gaps between some of the predicted repeats (Figure 3.2). Co-crystal structures of APUM23 bound to its cognate RNA will identify the nucleotide-amino acid interactions that occur, identify the precise number of functional Puf repeats, and determine whether extrusion of bases is a component of its binding to RNA. Attempts were made in this thesis research to obtain a co-crystal structure of APUM23 bound to its cognate RNA. The recombinant GST-APUM23 (full length APUM23) was purified using glutathione affinity purification followed by gel filtration chromatography to a concentration of over 10 mg/mL. However, rapid degradation and aggregation of the purified protein prevented its incorporation in a crystal screen.

APUM23, a protein with uniquely dispersed Puf repeats, provides a newly identified backbone for a sequence-specific RNA binding protein. The RNA-binding motif swapping experiments provided evidence for specific binding of predicted Puf repeats 1, 3, and 5 and demonstrated that the APUM23 Puf repeats could be engineered to recognize different RNA targets. Also, the 10-nucleotide RNA-binding consensus sequence that is preferred by APUM23 may provide a greater range of specificity than the classical eight-repeat PUFs. These RNA binding characteristics demonstrate the potential for engineering APUM23 so that it can bind to specific cellular RNA targets to modulate the physiology and metabolism of these RNAs, as has been achieved for other PUF proteins (Filipovska et al. 2011; Wang et al. 2013; Abil and Zhao 2015).

CHAPTER 4 - PEROXISOMAL MDH BINDS TO RNA IN A SEQUENCE SPECIFIC MANNER

4.1 Introduction

The mechanisms responsible for peroxisomal matrix protein import are fundamentally different from those responsible for import into the chloroplast and mitochondrion. Peroxisomal matrix proteins are imported fully folded and sometimes in an oligomerized state from the cytosol using a transient pore import mechanism that is still not fully understood (Hettema et al. 2014). This unique import mechanism raises the possibility that peroxisomal matrix proteins possess activities in the cytosol, perhaps prior to their import into the peroxisome. pMDH and pMFP were observed to possess microtubule (MT) and RNA binding activity *in vitro*, and fluorescent protein fusions of these proteins label peroxisomes and microtubules when expressed in plant epidermal cells (Chuong et al. 2002; Chuong et al. 2005; Chuong et al. 2004; Freeman and Muench, unpublished data). This led to the development of a working model where these proteins have translational autoregulatory activities in the cytosol (Figure 1.5). The reported RNA binding activity of these proteins have not been well characterized. However, support for the RNA binding activity of other metabolic enzymes has been well documented (Hentze and Preiss 2010; Scherrer et al. 2010). Recent evidence using *in vivo* approaches to identify authentic RNA-binding proteins in mammalian cells have demonstrated that a considerable number of mRBPs do not contain previously annotated RNA-binding domain (Baltz et al. 2012; Castello et al. 2012; Ascano et al. 2013). Amongst these proteins with previously unknown RNA-binding domains, were a number of intermediary metabolic enzymes,

including cytosolic MDH and hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (HADHB, trifunctional protein)(Castello et al. 2012). These homologs of pMDH and pMFP provide supporting evidence for their RNA binding activity in the cytosol of plant cells.

To provide insight into whether pMFP and pMDH possess sequence-specific RNA-binding ability, SELEX was used to identify their possible RNA-binding consensus sequences. The results of these experiments demonstrated that pMDH, but not pMFP, bound to specific RNAs. RNA consensus sequences of pMDH were identified; one was a palindrome and the other was a triple repeat of four nucleotides. The binding behavior of pMDH to its SELEX-derived RNA consensus sequences were characterized in a series of EMSA experiments with or without the presence of its biochemical co-substrates. Arabidopsis cDNA sequence searches determined that an mRNA that encodes an F-box protein contained both of these pMDH consensus sequences, perhaps linking the RNA binding activity of pMDH to targeted protein degradation.

4.2 Results

4.2.1 SELEX identifies consensus RNA binding sequences for pMDH

SELEX was performed as outlined in Chapter 3 using recombinant pMDH (with its N-terminal peroxisomal targeting signal removed) and full length pMFP. The pMDH SELEX experiment resulted in a typical sequence-specific RNA enrichment profile that was similar to those observed for APUM2, 6, 12 and 23 (Chapter 3)(Zhang and Muench 2015). (Zhang and Muench 2015)(Zhang and Muench 2015)The pMDH profile showed a gradual increase in RNA enrichment that reached a maximum of 18% at the tenth cycle

(Figure 4.1). In contrast, the pMFP enrichment profile increased rapidly, reaching 20% at cycle 3 and peaking at 37% at cycle 5 (Figure 4.1). This pMFP profile is typical for proteins that possess non-specific RNA binding activity, similar to that of APUM24 and Nop9 (Figure 3.6 and 3.9). The pMDH and pMFP SELEX products from the tenth cycle were analyzed by DNA sequencing (Table 4.1 and 4.2). Consistent with the pMFP RNA enrichment profile (Figure 4.1), MEME analysis of the pMFP bound sequences did not produce any RNA consensus sequence (Table 4.1). In contrast, the pMDH SELEX sequencing identified three consensus sequences. Consensus sequence #1 consisted of three tandemly arranged direct repeats of an AUGG oligoribonucleotide sequence. The level of conservation of adenine at the first position in each of the four-nucleotide subunits was not same, indicating that degeneracy exists at this position (Figure 4.2). There were 11 out of 43 sequences that contained a variant of this consensus sequence. Consensus sequence #2 was a nine-nucleotide palindromic sequence that consists of two opposite AUUG sequences aligned head-to-head using an adenine as the axis (Figure 4.2). This consensus was generated from 9 different SELEX cloned products that all contained the exact GUUAAAUUG sequence. The third consensus is similar to the single subunit sequence that sequence is present in consensus sequence #1, but with weaker preference for guanine at position 3 (Figure 4.2). The consensus #3 appeared in high number of SELEX products (25/43, including dimers of the subunit), which suggests that clones bearing such a four-nucleotide primer sequence were sufficient to survive selection.

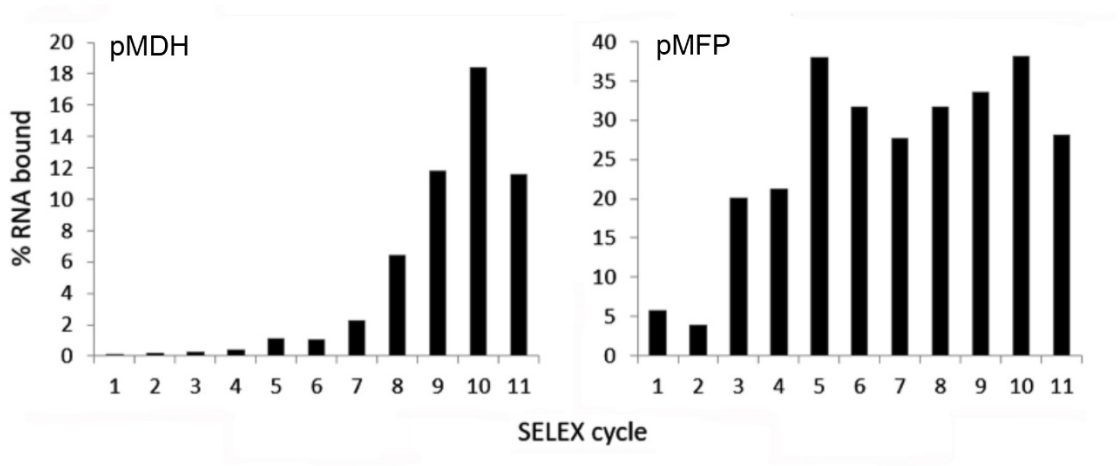


Figure 4.1 RNA enrichment profiles from the pMDH and pMFP SELEX experiments. SELEX assays were performed using 6xHis-tag containing recombinant proteins. The percentage of RNA that bound to the protein-coupled matrix was determined after each cycle of enrichment. DNA from the cycle sample that possessed the highest level of enrichment was used for cloning and sequencing.

Table 4.1 Raw data from the pMFP SELEX experiment showing the nucleotide sequences of individually cloned SELEX products.

pMFP SELEX products	Forward nucleotide sequence
A-1	TAATCCGGCCATAGCCTTGT
A-2	GGGGGTCACTGTACCGAAAT
A-3	TGAATTGGCTAAGTCCGGT
A-4	TCTAATATGTGACCAAGCC
A-5	CTGGTAGTATGGTACGTGGT
A-6	AATGTTCGTACCGTATCGTG
A-7	CATAGGAAAAGCACAATGGT
A-8	TTTGGGGTGGGTCAGGTAGG
A-9	TTATAGTTTACTTGGTCACG
A-10	AATGGGCGGAGGTGGGCTGG
A-11	GGTATAGATAACGTGTGGCC
A-12	TAGGGGGCTAGGCGTGGGGG
A-13	GTGTGGATGATATCGTGTGT
A-14	ACTTACCGGGTTACCGGGTGGGTG
A-15	GGTTCCGGCTAACGGATGGT
A-16	CAGGATGGACATGCTATTAT
A-17	GTGGTGTTGTGGACTGTGCC
A-18	ATAGGGGAAGCGGTGGGTGG
A-19	ATAATTGTGTGTGACGTCGT
A-20	GAGACGTCATTATACGAGGT
A-21	TACAAATGTGGGCCATGTAT
A-22	ACGATTAACATTGCCGGCCT
A-23	GATAGTAGCCAGGCCTTTGT
A-24	AGGGATAGGTGGAGGTGGGT
A-25	TAGGGTGTATGGTGGTGGTA
A-26	TGGGAGGGCGGGATTGGGCC
A-27	ATATCGGGACGTGGTGGGTG
A-28	GATGACGTAATGCACTTTGT
A-29	ACATTGAACAGTTCGGCATT
A-30	GACTTCCTTTAGTCTTTGGT
A-31	TGACATGGAAGGATGGTGGT
A-32	GTGAGTGAATACGTACATGG
A-33	ATTGCTGGACCGCAGCACGT
A-34	CGTGGGTCGGGTTGGGCGGG
A-35	CAGTTGTGGGGTGGGAGGGC
A-36	TGTATACGACATACGTCGAT
A-37	GGCTGAACCGCACCTTTTCT
A-38	AGTTAGCCAAATATGGCCTG
A-39	GTTGATTTGTTATGTCGTGT
A-40	TACTGAGCGGCACTGCAATA
A-41	TACAATAGTGATGGCATTAT
A-42	GGGGGGCGGGTCGAGGGTAG
A-43	GTGTTGACGTGTGTGATTGT
A-44	TGACCGAACGTGGGATGTAT
A-45	ATCACTAAGACACGAATGCT
A-46	TGGGGGGTGGGGAGGTCGTG
A-47	GATATAATGTGTCCCTTCCG
A-48	ATTGTGGGCATGGCCAGTGT

pMFP SELEX products	Forward nucleotide sequence
A-49	TCGAACTGAGAACGTGTTGT
A-50	ATATTATTGATGTGGGACGT
A-51	GCCAGCGGTTGTACGGTTAT
A-52	AGCAGACCAGCCTGACATGT
A-53	ATGGGGTGGCAGATGTGGAG
A-54	TTGCGACACAGATTACCACG
A-55	TTGGTATAATGGAGGTTGGT
A-56	TTTGGGTGGGGGATGGGCGG
A-57	AGACCGGGATGGGTGTGGTG
A-58	GATGAACCTGGTCCGATGTG

Table 4.2 Raw data from the pMDH SELEX experiment showing the nucleotide sequences of individually cloned SELEX products.

pMDH SELEX products	Forward nucleotide sequence
M-1	TAAGACCGGCAGGCAGTACG
M-2	ACAACGGAGTGGGATGGGCG
M-3	GAGGTCGTTAAATTGATGGT
M-4	GATCGGAGGTACGCCCGTCC
M-5	GCTCCGAAGGTGCGCCCGTC
M-6	GCGGTCGTTAAATTGAATGT
M-7	GTAAGTCCGACAGGCTAGCG
M-8	GATGGTTGGATGGGAGGGCA
M-9	TGAGTATGGCAGGAGGTGGG
M-10	GCGGTCGTTAAATTGTGTGT
M-11	TTAATGGTTGGATGGGAGGG
M-12	CGGGTAACGGAGGAGGCGTA
M-13	ATAACATGGATGGATGGGTG
M-14	TAAGAACATGGTTGGTTGGT
M-15	GTAAGACCGACAGGCAAGGG
M-16	GGGAGGATGGATGGGGAGTGGGGT
M-17	ATGGATGGATGGGAGGGGCA
M-18	AATCGGAGGTGTGCCCGTCC
M-19	TCAGGGAAGGGAAGGGTAGG
M-20	AAACAATGGATGGATGGGTG
M-21	GGGCAATAGGAGGAGGGAGG
M-22	AGGGTGGACACGTGAGGGTG
M-23	GTAAGACCGGCAGGCAGGGG
M-24	AATGGATGGGTTATGGTGTG
M-25	CGTGTATGGCAGGAGGTGGG
M-26	TATCGGAGGTGTGCCCGTCC
M-27	TGAGAAGTGTACTAGACCCG
M-28	GAGGGAAGGGGAGGGAGTGG
M-29	TGAGTATGGCAGGAGGCGGG
M-30	GAGGTCGTTAAATTGTCTGT
M-31	GCGGTCGTTAAATTGTTGTT
M-32	GTAAGACCGGCAGGCAAGCG
M-33	TAAGTCCGGCAGGCAGTATG
M-34	GATCGGAGGTGTGCCCGTCC
M-35	TATGGATGGGTTATGGTGTG
M-36	GGTCATAGGTGGAGGTAGGG
M-37	GCGGTCGTTAAATTGAGGTT
M-38	GAGGTCGTTAAATTGATTGT
M-39	AACAATGGATGGATGGGCGG
M-40	GAGGTCGTTAAATTGTGTGT
M-41	GAGGTCGTTAAATTGTTGTT
M-42	ACAAATGATGGATGGTTGGT
M-43	TAGGGGAAGGGAAGGGGTAG

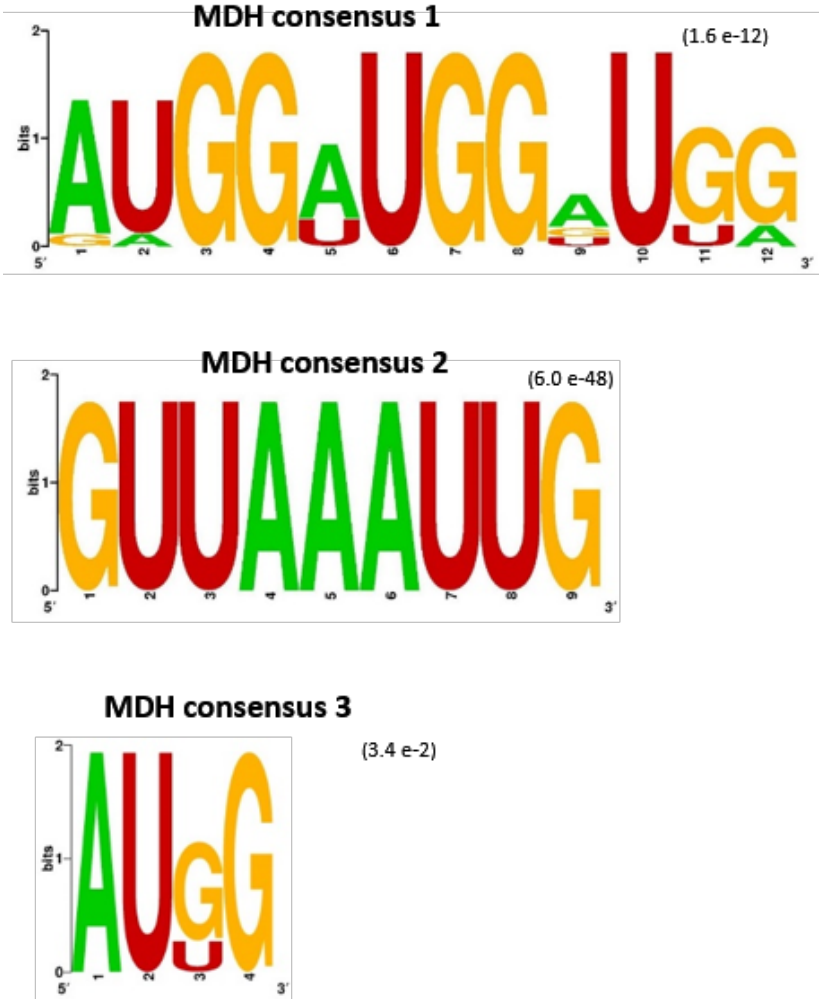


Figure 4.2 pMDH consensus RNA binding sequence logo graphs. SELEX-derived consensus RNA target sequences for pMDH. Numbers below the letters in the logo graph refer to the nucleotide position in the RNA sequence (5' to 3' direction). The *E*-value of the consensus sequence is indicated above the logo graph. The frequency of each consensus (or similar) presenting in the population are: consensus #1 was present in 11/43 sequences; consensus #2 was presented in 9/43 sequences; consensus #3 presented in 25/43 sequences.

4.2.2 pMDH RNA mobility shift assays

Previously, pMDH was shown to bind to the 3'UTR of prolamine seed storage protein mRNA in a UV-crosslinking assay (Freeman and Muench, unpublished data). However, it was unclear if it bound to such an RNA in the sequence-specific manner. The RNA consensus sequences that were obtained from the SELEX analysis were used in mobility shift experiments with increasing concentrations of pMDH to determine the binding affinity of the protein to the two RNA types. The binding affinity to the 12 nucleotide preferred sequence from RNA consensus #1 and a 13-nucleotide RNA that contained consensus sequence #2 was high, with apparent dissociation constants of 43 and 31 nM, respectively (Figure 4.3A, B, C). A UV-crosslinking assay using an AU-rich RNA oligo (UUUAUAUA) was shown not to bind to pMDH (data not shown).

Other dehydrogenases have been identified as RNA-binding proteins, and the RNA-binding site has been proposed to be the binding site for the nucleotide co-substrates NAD^+ and NADH (Cieśla 2006). For example, the RNA binding activity of GAPDH was blocked by the addition of excess amounts of NAD^+ and NADH (Nagy and Rigby 1995). A mobility shift experiment was performed using these nucleotides at concentrations up to 10-times excess the amount of pMDH protein, and demonstrated that the addition of nucleotide did not compete with binding of pMDH with consensus sequence #1 (Figure 4.3D).

4.2.3 Identification of Arabidopsis mRNAs that contain the pMDH consensus sequences

The identification of two relatively long (12 and 9 nucleotides) consensus RNA binding sequences for pMDH allowed for a BLAST search to determine the identity of

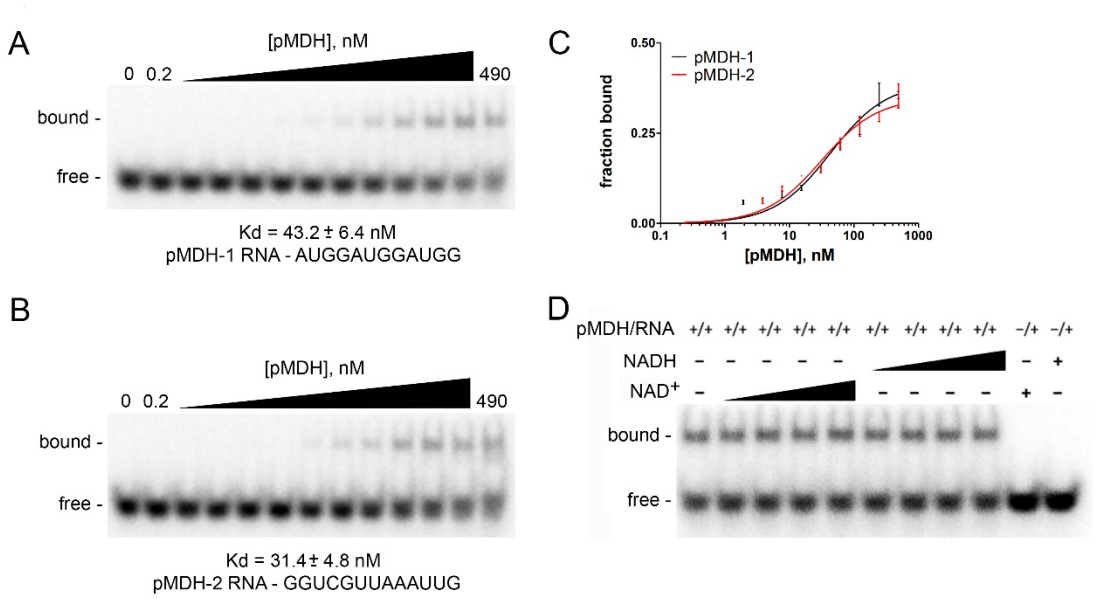


Figure 4.3 pMDH binds with high affinity to its consensus RNA binding sequences. Representative mobility shift assays of each pMDH using its RNA consensus sequence #1 (A) or a 13 nucleotide containing sequence #2 (B) as determined by SELEX. The lowest and highest protein concentration in the exponential dilution series is shown above the gel. Protein concentrations were corrected so that they reflect the active concentration of protein in each sample. The average apparent dissociation constant value (Kd) for the protein bound to its RNA consensus sequence is shown, as is the sequence of the RNA used in the assay. (C) representative equilibrium binding data of pMDH to consensus #1 (black line) and consensus #2 (red line) sequence. The fraction of RNA bound is shown on the vertical axis, and the concentration of protein is shown on the horizontal axis. (D) Electrophoresis mobility shift assay on pMDH and its consensus RNA sequence #1 in the presence of NADH or NAD+. The protein and RNA input was 4 μM and 1 nM, respectively, except in the last two reactions where there was no protein added. The concentration series of the NADH and NAD+ were 50 nM, 500 nM, 5 μM and 50 μM.

Arabidopsis mRNAs that contain one or both of these sequences. The random chance that consensus sequence #1 is found in an mRNA sequence is greater than once in 16,000 kb, and greater than once in 250 kb for consensus sequence #2. Six mRNAs were determined to contain the preferred consensus sequence #1, including an F-box family protein (AT3G23955) and a WUSCHEL-interacting protein (AT3G15880) (Table 4.3). There were 48 mRNAs that contained consensus sequence #2 (Table 4.3). These mRNAs code for a broad range of proteins. Interestingly, there were three mRNAs that encoded different F-box proteins that contained consensus sequence #2. Remarkably, one of these F-box encoding mRNAs also contained consensus sequence #1, an extremely rare event that suggests that this mRNA could be an authentic target of pMDH. Also interesting is the presence of consensus sequence #2 in an mRNA that encodes a WUSCHEL-related homeobox protein (Table 4.3). The presence of the consensus sequences in an mRNA that encodes a WUSCHEL-interacting protein as well as a WUSCHEL protein may indicate a coordinated regulation of these mRNAs. WUSCHEL is a homeobox gene that regulates stem cell pools (Bäurle and Laux 2005).

4.3 Discussion

Evidence is presented in this chapter for the sequence specific RNA binding activity of pMDH. The RNA binding consensus sequences that were determined by SELEX analysis differed in that consensus sequence #1 was a twelve-nucleotide sequence that consists of a triple-direct repeat of an AUGG unit, whereas the consensus sequence #2 was a nine-nucleotide palindrome sequence (GUUAAAUUG). The repeated nature of these sequences may be related to the assembly of pMDH in the cytosol. For sequence #2, an

Table 4.3 Arabidopsis thaliana transcripts containing the SELEX-derived RNA consensus sequences of pMDH.

Consensus Sequence	mRNA Accession Number (NCBI)	Gene Number (TAIR)	Description	Location of the Consensus Sequence
AUGGAUGGAUGG (6)	NM_001203031	AT3G23955	F-box family protein	Coding region
	NM_001202976	AT3G15880	WUS-interacting protein 2	Coding region
	NM_001125683	AT5G02390	DUO1-activated unknown 1	Coding region
	NM_102424.2	AT1G26620	uncharacterized protein	Coding region
	NM_105603.2	AT1G69360	uncharacterized protein	Coding region
	NM_111367.2	AT3G04950	uncharacterized protein	Coding region
GUUAAAUUG (48)	NM_001203649	AT5G60170	RNA binding (RRM/RBD/RNP motifs) family protein	3' UTR
	NM_001125978	AT5G57290	60S acidic ribosomal protein P3-2	3' UTR
	NM_001203356	AT5G10710	Cenp-O domain-containing protein	3' UTR
	NM_119693.4	AT4G35270	plant regulator RWP-RK family protein	3' UTR
	NM_111207.2	AT3G03360	putative F-box protein	3' UTR
	NM_001202622	AT2G18850	SET domain-containing protein	3' UTR
	NM_124424.3	AT5G50420	O-fucosyltransferase family protein	3' UTR
	NM_122325.2	AT5G24190	lipase class 3-related protein	3' UTR
	NM_121531.5	AT5G15270	RNA-binding KH domain-containing protein	3' UTR
	NM_121522.3	AT5G15180	peroxidase 56	3' UTR
	NM_128948.3	AT2G33880	WUSCHEL-related homeobox 9	3' UTR
	NM_102270.4	AT1G24240	ribosomal protein L19 family protein	3' UTR
	NM_202038.3	AT1G05900	endonuclease III 2	3' UTR
	NM_001203031	AT3G23955	F-box family protein	3' UTR
	NM_001202993	AT3G18100	myb domain protein 4r1	5' UTR, Coding region
	NM_001202949	AT3G13340	transducin/WD40 domain-containing protein	5' UTR
	NM_001202783	AT2G39435	phosphatidylinositol N-acetyglucosaminlytransferase subunit P-like protein	5' UTR
	NM_001198520	AT1G80420	DNA-repair protein XRCC1	5' UTR
	NM_001160907	AT1G28710	nucleotide-diphospho-sugar transferase family protein	5' UTR
	NM_001036831	AT5G19485	transferases/nucleotidyltransferases	5' UTR
	NM_001203168	AT3G55450	PBS1-like 1 kinase	Coding region
	NM_001203006	AT3G19720	dynamin-like protein ARC5	Coding region
	NM_001202937	AT3G12020	kinesin motor protein-like protein	Coding region
	NM_001198507	AT1G79280	nuclear pore anchor	Coding region

Consensus Sequence	mRNA Accession Number (NCBI)	Gene Number (TAIR)	Description	Location of the Consensus Sequence
GUUAAAUUG	NM_104911.2	AT1G62310	transcription factor jumonji domain-containing protein	Coding region
	NM_001198329	AT1G58025	DNA-binding bromodomain-containing protein	Coding region
	NM_001198126	AT1G21610	wound-responsive family protein	Coding region
	NM_118644.6	AT4G25120	helicase SRS2-like protein	Coding region
	NM_105391.4	AT1G67220	histone acetyltransferase HAC2	Coding region
	NM_001161259	AT5G22791	F-box family protein	Coding region
	NM_001126005	AT5G62100	BCL-2-associated athanogene 2	Coding region
	NM_121385.4	AT5G13820	Telomere repeat-binding protein 4	Coding region
	NM_115408.3	AT3G55510	REBELOTE protein	Coding region
	NM_128973.3	AT2G34220	maternal effect embryo arrest 20	Coding region
	NM_001124144	AT1G77950	agamous-like 67 protein	Coding region
	NM_104973.3	AT1G62950	putative leucine-rich repeat transmembrane protein kinase	Coding region
	NM_102448.3	AT1G26840	origin recognition complex subunit 6	Coding region
	NM_101786.2	AT1G19290	pentatricopeptide repeat-containing protein	Coding region
	NM_100364.3	AT1G04860	ubiquitin-specific protease 2	Coding region
	NM_202336.2	AT1G61640	ABC1 domain-containing kinase	Coding region
	NM_001035824.3	AT3G60410	uncharacterized protein	3’ UTR
	NM_111273.5	AT3G04020	uncharacterized protein	3’ UTR
	NM_001203024	AT3G23255	uncharacterized protein	3’ UTR
	NM_101404.3	AT1G15350	uncharacterized protein	Coding region
	NM_117893.4	AT4G17840	uncharacterized protein	Coding region
	NM_001161069	AT2G30120	uncharacterized protein	Coding region
	NM_001197969	AT1G03290	uncharacterized protein	Coding region
	NM_121876.5	AT5G18710	uncharacterized protein	Coding region

RNA binding domain on each pMDH subunit could work together to bind one half of the consensus sequence in an inverted fashion due to the opposite orientation of the subunits of pMDH in the dimerized state in the cytosol (Hall et al. 1992). The binding of pMDH to the triple repeat found in consensus sequence #1 could occur through an oligomerization of pMDH where more than one dimer interacts with the target RNA at one time. The Rossmann fold that is associated with the nucleotide binding site is present in MDH enzymes (Chen et al. 2001), so it was assumed that pre-incubation of pMDH with excess amounts of NAD^+ and NADH prior to RNA addition would inhibit its RNA binding activity, as it has for other dehydrogenases (Cieśla 2006). Mobility shift experiments did not provide evidence for the presence of RNA binding activity in the Rossmann fold (Figure 4.3D). This result may indicate the presence of an additional RNA binding domain on the protein. Alternatively, the 10-fold preincubation of NAD^+ and NADH may not have provided sufficient inhibition at that concentration. Also, only one of the RNA consensus sequences was used in these experiments. The presence of two completely different consensus sequences indicates that there are either two RNA binding domains in the protein or, more likely, a conformational change in the RNA binding site alters the sequence specificity. Additional mobility shift experiments using elevated levels of NAD^+ and NADH could be used, as could the use of consensus sequence #1. This may provide more insight on the RNA binding characteristics of this protein.

BLAST searches for these RNA consensus sequences in the Arabidopsis transcriptome identified 6 mRNAs that contained consensus sequence #1 and 48 mRNAs that contained consensus sequence #2. These consensus sequences could be found in these mRNAs by chance. However, one mRNA that encodes an F-box family member contains

both consensus sequences, and these are about 800 nucleotides apart in the F-box mRNA. The link to pMDH binding to an F-box mRNA and regulating its physiology is speculative. Perhaps the post-transcriptional regulation of this F-box protein by cytosolic pMDH functions is in a specific protein degradation process. F-box proteins are a major component of E3 ubiquitin ligases, and the F-box protein contributes to the specificity of the E3 ligase for its substrate (von Arnim 2001; Skaar et al. 2013). Ubiquitination of the peroxisomal matrix protein import receptor, Pex5, is an important determinant of whether Pex5 is recycled or degraded (Platta et al. 2007). If indeed the mRNA that encodes this F-box protein is bound by pMDH, and if Pex5 is a ubiquitination substrate for this F-box protein, then pMDH may target the F-box mRNA for degradation or translational repression as a mechanism to regulate its import into the peroxisome. When cytosolic pMDH (and other peroxisomal matrix proteins) levels are low and the import of matrix proteins may not be required at a high level, the F-box mRNA target of pMDH is stable and translated. This would then lead to increased degradation of Pex5 by the proteasome. When pMDH levels are high, there is an increased binding of pMDH on the F-box mRNA resulting in its degradation or translational repression, and this leads to an increase in Pex5 receptor levels and an increase in import rate. Further investigation on the *in vivo* targets of pMDH is necessary to confirm binding to this F-box encoding mRNA.

The RNA binding activity of pMDH is supported by recent RNA interactome studies that identified mammalian cytosolic MDH as having RNA binding activity *in vivo* (Baltz et al. 2012; Castello et al. 2012). The role of the putative RNA binding activities of pMDH and pMFP *in vivo* are not known. If not functioning to bind to their own mRNAs to autoregulate translation or to F-box mRNA to regulate ubiquitination of Pex5 (pMDH),

they might have some completely undetermined and unrelated role in RNA metabolism in the cell. Perhaps the MT and RNA binding activities work together to localize specific mRNAs by anchoring them onto MT populations that are localized to specific regions of the cell, such as on the inner surface of the plasma membrane where MTs are abundant. We determined that pMDH possesses specific RNA binding activity. However, additional experimentation is required to identify the authentic targets of pMDH in the cell. Clues related to the cytosolic activities of pMFP and pMDH will require the identification of endogenous RNA targets using advanced techniques such as *in viv*o UV-crosslinking coupled with mRNA purification and/or immunoprecipitation.

CHAPTER 5 - SUMMARY AND FUTURE DIRECTIONS

5.1 Summary

SELEX was used in this study to determine the identity of RNA target sequences for APUM and peroxisomal matrix proteins. High confidence RNA consensus sequences were identified for four APUM proteins, and included conventional PUF target sequences for three of these proteins (APUM2, 6 and 12). A novel RNA sequence was identified for APUM23 that breaks current PUF RNA target binding rules in that it binds to a central UUGA core sequence and prefers cytosine at Puf repeat 8. In addition, SELEX analysis demonstrated that APUM24 binds to RNA in a non-specific binding manner, similar to its human ortholog Puf-A (Qiu et al. 2014). The SELEX data for the variant PUF proteins APUM13 and APUM18, although not as high in confidence, provided insight into their RNA binding sequence characteristics. Despite having several variant TRMs, APUM13 bound to a typical RNA sequence, suggesting a new specificity code for these TRMs. The APUM18 consensus RNA binding sequences consisted of seven nucleotides, which matched the number of its predicted Puf repeat. There was a significant amount of degeneracy in the APUM18 RNA sequence at positions that possessed unusual TRMs. This suggests that APUM18 binds to RNA with reduced sequence specificity. Perhaps APUM18 requires interactions with other proteins *in vivo* in order to bind specifically to RNAs.

The RNA binding specificity of APUM23 was studied in greatest depth here because of its novel consensus RNA binding sequence. Thirteen nucleotide-substituted RNAs were used in mobility shift experiments to validate the SELEX-derived consensus target sequence for APUM23. The binding affinity of APUM23 to these base-substituted

RNAs showed a high correlation with the nucleotide composition of its RNA consensus sequence shown in the logo graph. Base substitutions within the UUGA core showed a drastic decrease in APUM23 binding affinity compared to the wild-type RNA, as did testing sequences that possessed the classical PUF UGUA core. We predict that APUM23 possesses ten Puf repeats that are not clustered within the C-terminal region, but rather are unevenly positioned throughout the central region of the protein based on TRM code identification and structural modeling analysis. Swapping of the five amino acid motifs in three Puf repeats of APUM23 confirmed the identity of these repeats (including the cytosine-binding repeat) and demonstrated that APUM23 could be engineered to modify its nucleotide specificity. However, amino acid swapping of the three amino acid TRM code did not change the specificity of the Puf repeats tested, suggesting that engineering of the APUM23 backbone will require modification of the entire five-amino acid motif in each Puf repeat.

Previously, the cytoskeleton associated peroxisomal proteins pMDH and pMFP have been shown to possess RNA binding activity (Chuong et al. 2002; Freeman and Muench, unpublished). SELEX was performed in an effort to determine the potential RNA target sequences of these proteins. The pMFP SELEX RNA enrichment profile showed a rapid enrichment of bound RNA in the early cycles, indicating that this protein possessed non-specific RNA binding activity *in vitro*. This non-specific RNA binding activity was confirmed by DNA sequencing analysis of the pMFP cloned SELEX products. The pMDH SELEX experiment yielded three consensus sequences by the analysis using MEME. The preferred consensus sequence #1 consisted of three tandemly arranged direct repeats of an AUGG primer sequence (AUGGAUGGAUGG). Consensus sequence #2 was a nine-

nucleotide palindromic sequence (GUUAAAUUG). The third consensus was a short four-nucleotide sequence (AUGG), which was similar to the primer sequence that appeared in the consensus #1. The repetitive and palindromic consensus RNA sequences #1 and #2 could be due to the dimeric nature of pMDH and the ability of each subunit to bind to a nucleotide sequence, or to multimerization of pMDH in the cytosol. The inability of nucleotide co-substrates (NAD^+ and NADH) of pMDH to impede binding of RNA suggests that the nucleotide binding domain of pMDH does not function in RNA binding, as was shown for other dehydrogenases that possess RNA binding activity (Cieśla 2006).

BLAST searches of the Arabidopsis transcriptome for the presence of the two pMDH RNA consensus RNA sequences identified six mRNAs that contained consensus sequence #1 and over 40 mRNAs that contained consensus sequence #2. These consensus sequences could be found in these mRNAs by chance. However, one mRNA that encodes an F-box family member contained both consensus sequences, and these were about 800 nucleotides apart in the F-box mRNA. Although a link to pMDH binding to and regulating an F-box mRNA is speculative, degradation of peroxin proteins that are involved in peroxisomal protein import has been proposed as being mechanistically similar to the ER associated degradation (ERAD) pathway, and involves ubiquitylation and AAA-type ATPase activity (Smith and Aitchison 2013). Perhaps the post-transcriptional regulation of this F-box protein by cytosolic pMDH functions in the specific degradation of peroxins via the proteasome, where binding of pMDH to peroxin mRNA could provide a means to regulate import of matrix proteins via the degradation of this mRNA.

Overall, this research provides insight into the novelty of plant RNA-binding proteins and their target RNA sequences, and sheds light on the role of these two classes

of proteins in post-transcriptional control of gene expression. It reveals that APUM23 could provide an advanced structural backbone for Puf repeat engineering and target-specific regulation of cellular RNAs. This work also builds upon existing evidence for the authentic RNA binding activities and identity of target RNA sequences for some metabolic enzymes.

The SELEX data presented here provides valuable insight into plant RNA-binding protein target preferences. The advantages of using a SELEX approach include the ability to control the concentration and quality of RNA and RNA-binding protein, as well as providing a highly complex RNA population to be screened. Other, more recently developed *in vitro* approaches also have merit. These include on-chip technologies, such as the RNA-MITOMI (RNA–mechanically induced trapping of molecular interactions) and the RNAcompete method (Ray et al. 2009; Martin et al. 2012; Ray et al. 2013). The RNA-MaP (RNA on a massively parallel array) approach is based on the Illumina high through-put sequencing platform and combines target sequence screening from a highly complex RNA pool with *in situ* affinity quantitative measurement (Buenrostro et al. 2014). This approach has great potential for application in future systemic investigation on RNA-protein interactions.

The identification of target sequences of RNA-binding proteins has benefitted from *in vitro* approaches, especially in the case of proteins such as PUFs that interact with their RNA target bases in a modular fashion. However, *in vivo* RNA target identification approaches determine the authentic cellular target sequences. The yeast three-hybrid system screens for interactions in a cellular environment, but is limited in that it is a heterologous system that lacks the endogenous cellular milieu, including other protein factors that might be required for sequence-specific binding of an RNA-binding protein.

The endogenous RNA targets of yeast PUFs were the first to be systemically examined using the RIP-Chip method (Gerber et al. 2004). However, these experiments are also limited in that there is potential for off-target binding during the experimental processing steps. This problem can be overcome by using a crosslinking IP approach (CLIP and its variants) that is now broadly used (Hafner et al. 2010; Wilinski et al. 2015). An exciting new approach involves RNA target labelling *in vivo*, an approach that also gains insight into how the RNA-binding protein can transiently interact with RNA targets without exerting a regulatory effect (Lapointe et al. 2015).

5.2 Future Directions

Only a small number of studies have been reported on the characterization of plant PUF proteins (Abbasi et al. 2010; Tam et al. 2010; Abbasi et al. 2011; Huh et al. 2013; Huang et al. 2014; Huh and Paek 2014; Xiang et al. 2014). The Arabidopsis PUF family is extensive, consisting of up to 26 members (Francischini and Quaggio 2009; Abbasi et al. 2010; Tam et al. 2010). It is reasonable to speculate that plants need a variety of PUF proteins to achieve more sophisticated post-transcriptional regulation when the plant is exposed to rapid changes in environmental conditions. Functional data has been reported for only four Arabidopsis PUF proteins (Abbasi et al. 2010; Huh et al. 2013; Huang et al. 2014; Huh and Paek 2014; Xiang et al. 2014). Similar to the limited knowledge about plant PUM proteins, the RNA binding activities of pMDH and pMFP are largely uncharacterized. The identification of consensus RNA binding sequences for pMDH is one of the few examples of RNA target sequence determination for metabolic enzymes that

possess RNA binding activity. Similar fundamental experiments can be planned for future characterization of both of these proteins, as described below.

Approaches using immunoprecipitation of RNA-binding proteins that have been UV crosslinked to their cognate RNAs *in vivo* (CLIP and its variants, Ule et al. 2005) are among the most powerful approaches to identify authentic RNA targets of RNA-binding proteins. Research performed in our laboratory has demonstrated that selected APUM proteins as well as pMDH and pMFP can be UV crosslinked to RNA. Thus a CLIP strategy should be effective for these two groups of proteins. However, this procedure largely relies on the successful production of transgenic plants expressing epitope-tagged RNA-binding proteins or antibodies raised directly against the target RNA-binding protein. We have managed to produce multiple versions of antibodies against several Arabidopsis PUFs, pMDH and pMFP proteins, and have constructed transgenic lines that express these proteins fused to different tags. Efforts are in place to utilize these resources to identify endogenous RNA targets of these proteins. Not only will these results identify consensus RNA binding sequences for these proteins, they will also identify the function of the bound RNAs and provide insight into their function.

For instance, we have shown that APUM6 targets to the outer endoplasmic reticulum membrane, and is hypothesized to bind to mRNA that encode secreted proteins. A CLIP experiment could validate this hypothesis. CLIP would also determine whether APUM23 binds directly to the 18S rRNA sequence, and whether it also binds to mRNAs in the cytosol. Interestingly, 17 Arabidopsis mRNAs possess the preferred APUM23 target sequence (GAAUUGACGG), and only one of these mRNA types shows an increase in abundance in an *apum23* mutant (Abbasi et al. 2010). This mRNA encodes the auxin

transporter, MDM4/PGP4, a protein that appears to function in auxin uptake involved in basipetal transport of auxin at the root tip (Geisler and Murphy 2006; Terasaka et al. 2010). *apum23* mutants demonstrated a mislocalized accumulation of auxin away from the root tip (Abbasi et al. 2010), possibly caused by elevated MDR4/PGP4 levels. If indeed the MDR4/PGP4 mRNA is a direct target of APUM23, the stability of this mRNA may be negatively regulated by APUM23. In support of this role for APUM23, endogenous expression and overexpression of MDR4/PGP4 were both shown to be regulated at the post-transcriptional level (Terasaka et al. 2010). Thus, this data suggests a role for APUM23 binding to a cytosolic mRNA in the root. This would be in a similar manner to the nucleolar localized Puf6 that functions in rRNA processing and in the localization of Ash1 mRNA to the bud tip in yeast cells (Müller et al. 2011; Qiu et al. 2014). Identification of the *in vivo* RNA targets of pMDH would also be valuable in that it would determine whether the mRNA that encodes the F-box protein and contains both pMDH consensus sequences is indeed a target of pMDH. These CLIP results would lead to additional experiments, including the effects of knock-down or overexpression of the RNA-binding protein on transcript stability or translation. This would provide insight on the molecular mode of action of the APUM or pMDH/pMFP proteins.

Determining the co-crystal structure of APUM23 and pMDH would be important in understanding how these proteins bind to RNA. A co-crystal structure was important in understanding the unique RNA binding pattern of human Puf-A (Qiu et al. 2014), an ortholog of Arabidopsis APUM24. Puf-A folds into an L-shaped structure with 11 Puf repeats in two subdomains and binds structured RNA and DNA in a non-sequence-specific manner (Qiu et al. 2014). Puf-A and APUM24 are also nucleolar targeted PUF proteins

(Tam et al. 2010; Chang et al. 2011), and Puf-A is known to function in rRNA processing, although at a different step from yeast Nop9 and APUM23 (Thomson et al. 2007; Abbasi et al. 2010; Qiu et al. 2014). Although APUM23 and Puf-A have different RNA binding characteristics and low conservation in their amino acid sequences, the Puf-A structure demonstrates that a non-classical PUF protein can form a unique structure. Also, since nothing is known about the RNA binding mechanism for MDH proteins, a co-crystal structure of pMDH and its cognate RNA would be immensely valuable. It would determine whether pMDH binds to RNA at the nucleotide binding site, and whether it has the ability to bind to two distinct RNA sequences at different locations on the protein. This would be consistent with the SELEX data that identified two RNA consensus sequence types.

One of the exciting outcomes of this thesis research was the identification of the novel and specific APUM23 RNA binding sequence. This, coupled with the evidence that Puf motif swapping is possible for this protein, suggests that APUM23 provides an advanced structural backbone for PUF repeat engineering and target-specific regulation of cellular RNAs. The ten-nucleotide target RNA sequence bound by APUM23 provides a larger target sequence to engineer compared to the more typical eight-nucleotide recognition sequences of conventional PUM proteins. Co-crystal structures of APUM23 bound to its cognate RNA will identify the nucleotide-amino acid interactions that occur, and identify the precise number of functional Puf repeats. The binding affinity of APUM23 to RNA is high (<1 nM) and there is limited degeneracy in its target consensus sequence. These features could make the APUM23 backbone a valuable tool for regulating the metabolism of cellular RNAs of interest when the backbone is fused to effector domains, such as a translational enhancer or RNase.

REFERENCES

Abbasi N, Kim HB, Park N, Kim H-S, Kim Y-K, Park Y-I, Choi S-B. 2010. APUM23, a nucleolar Puf domain protein, is involved in pre-ribosomal RNA processing and normal growth patterning in Arabidopsis. Plant J. 64:960–976.

Abbasi N, Park Y-I, Choi S-B. 2011. Pumilio Puf domain RNA-binding proteins in Arabidopsis. Plant Signal. Behav. 6:364–368.

Abil Z, Zhao H. 2015. Engineering reprogrammable RNA-binding proteins for study and manipulation of the transcriptome. Mol. BioSyst. 11:2658–2665.

Ambrosone A, Costa A, Leone A, Grillo S. 2012. Beyond transcription: RNA-binding proteins as emerging regulators of plant response to environmental constraints. Plant Sci. 182:12–18.

von Arnim AG. 2001. A hitchhiker's guide to the proteasome. Sci. STKE 2001:pe2.

Ascano M, Gerstberger S, Tuschl T. 2013. Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. Curr Opin Genet Dev 23:20–28.

Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2:28–36.

Baltz AG, Munschauer M, Schwanhäusser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, et al. 2012. The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. Mol. Cell 46:674–690.

Bäurle I, Laux T. 2005. Regulation of WUSCHEL transcription in the stem cell niche of the Arabidopsis shoot meristem. Plant Cell 17:2271–80.

Biasini M, Bienert S, Waterhouse a., Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic

Acids Res. 42:W252–W258.

Bor YC, Swartz J, Morrison A, Rekosh D, Ladomery M, Hammarskjöld ML. 2006. The Wilms' tumor 1 (WT1) gene (+KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. Genes Dev. 20:1597–1608.

Branco-Price C, Kawaguchi R, Ferreira RB, Bailey-Serres J. 2005. Genome-wide analysis of transcript abundance and translation in arabidopsis seedlings subjected to oxygen deprivation. Ann. Bot. 96:647–660.

Brown L-A, Baker A. 2003. Peroxisome biogenesis and the role of protein import. J. Cell. Mol. Med. 7:388–400.

Browning KS. 1996. The plant translational apparatus. Plant Mol. Biol. 32:107–44.

Browning KS, Webster C, Roberts JKM, Ravel JM. 1992. Identification of an isozyme form of protein synthesis initiation factor 4F in plants. J. Biol. Chem. 267:10096–10100.

Campbell ZT, Valley CT, Wickens M. 2014. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. Nat. Struct. Mol. Biol. 21:732–738.

Cao J, Arha M, Sudrik C, Mukherjee a., Wu X, Kane RS. 2015. A universal strategy for regulating mRNA translation in prokaryotic and eukaryotic cells. Nucleic Acids Res. 43:4353–4362.

Cao Q, Richter JD. 2002. Dissolution of the maskin-eIF4E complex by cytoplasmic polyadenylation and poly(A)-binding protein controls cyclin B1 mRNA translation and oocyte maturation. EMBO J. 21:3852–62.

Carberry SE, Goss DJ. 1991. Wheat germ initiation factors 4F and (iso)4F interact differently with oligoribonucleotide analogues of rabbit alpha-globin mRNA. Biochemistry 30:4542–4545.

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. Cell 149:1393–1406.

Chagnovich D, Lehmann R. 2001. Poly(A)-independent regulation of maternal hunchback translation in the Drosophila embryo. Proc. Natl. Acad. Sci. U. S. A. 98:11359–11364.

Chang H-Y, Fan C-C, Chu P-C, Hong B-E, Lee HJ, Chang M-S. 2011. hPuf-A/KIAA0020 modulates PARP-1 cleavage upon genotoxic stress. Cancer Res. 71:1126–1134.

Chang Y-F, Imam JS, Wilkinson MF. 2007. The Nonsense-Mediated Decay RNA Surveillance Pathway. Annu. Rev. Biochem. 76:51–74.

Chen C-YA, Shyu A-B. 2011. Mechanisms of deadenylation-dependent decay. Wiley Interdiscip. Rev. RNA 2:167–183.

Chen J, Walter S, Horwich a L, Smith DL. 2001. Folding of malate dehydrogenase inside the GroEL-GroES cavity. Nat. Struct. Biol. 8:721–728.

Cheong C-G, Hall TMT. 2006. Engineering RNA sequence specificity of Pumilio repeats. Proc. Natl. Acad. Sci. U. S. A. 103:13635–9.

Cho PF, Gamberi C, Cho-Park YA, Cho-Park IB, Lasko P, Sonenberg N. 2006. Cap-Dependent Translational Inhibition Establishes Two Opposing Morphogen Gradients in Drosophila Embryos. Curr. Biol. 16:2035–2041.

Choudhury R, Tsai YS, Dominguez D, Wang Y, Wang Z. 2012. Engineering RNA endonucleases with customized sequence specificities. Nat. Commun. 3:1147.

Chuong SDX, Good AG, Taylor GJ, Freeman MC, Moorhead GBG, Muench DG. 2004. Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells. Mol. Cell. Proteomics 3:970–983.

Chuong SDX, Mullen RT, Muench DG. 2002. Identification of a rice RNA- and microtubule-binding protein as the multifunctional protein, a peroxisomal enzyme involved in the ??-oxidation of fatty acids. J. Biol. Chem. 277:2419–2429.

Chuong SDX, Park N-I, Freeman MC, Mullen RT, Muench DG. 2005. The peroxisomal multifunctional protein interacts with cortical microtubules in plant cells. BMC Cell Biol. 6:40.

Cieśla J. 2006. Metabolic enzymes that bind RNA: Yet another level of cellular regulatory network? Acta Biochim. Pol. 53:11–32.

Clement SL, Scheckel C, Stoecklin G, Lykke-Andersen J. 2011. Phosphorylation of Tristetraprolin by MK2 Impairs AU-Rich Element mRNA Decay by Preventing Deadenylase Recruitment. Mol. Cell. Biol. 31:256–266.

Conne B, Stutz a, Vassalli JD. 2000. The 3' untranslated region of messenger RNA: A molecular "hotspot" for pathology? Nat. Med. 6:637–41.

Cooke A, Prigge A, Opperman L, Wickens M. 2011. Targeted translational regulation using the PUF protein family scaffold. Proc. Natl. Acad. Sci. U. S. A. 108:15870–5.

Cross LL, Ebeed HT, Baker A. 2015. Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. Biochim. Biophys. Acta - Mol. Cell Res.

Culjkovic B, Topisirovic I, Borden K. 2007. Controlling gene expression through RNA regulons. Cell Cycle 6:65–69.

Culjkovic B, Topisirovic I, Skrabanek L, Ruiz-Gutierrez M, Borden KLB. 2006. eIF4E is a central node of an RNA regulon that governs cellular proliferation. J. Cell Biol. 175:415–26.

Deng Y, Singer RH, Gu W. 2008. Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. Genes Dev. 22:1037–50.

Diem MD, Chan CC, Younis I, Dreyfuss G. 2007. PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. Nat. Struct. & Mol. Biol. 14:1173–1179.

Dong S, Wang Y, Cassidy-Amstutz C, Lu G, Bigler R, Jezyk MR, Li C, Hall TMT, Wang Z. 2011. Specific and Modular Binding Code for Cytosine Recognition in Pumilio/FBF (PUF) RNA-binding Domains. J. Biol. Chem. 286:26732–26742.

Dostie J, Dreyfuss G. 2002. Translation is required to remove Y14 from mRNAs in the cytoplasm. Curr. Biol. 12:1060–1067.

Edwards T a, Pyle SE, Wharton RP, Aggarwal AK. 2001. Structure of Pumilio Reveals Similarity between RNA and Peptide Binding Motifs. Cell 105:281–289.

Edwards TA. 2015. Bespoke RNA recognition by Pumilios. Biochem. Soc. Trans. 43:801–6.

Ellington AD, Szostak JW. 1990. *In vitro* selection of RNA molecules that bind specific ligands. Nature 346:818–22.

Fernández-Pevida A, Kressler D, de la Cruz J. 2015. Processing of preribosomal RNA in *Saccharomyces cerevisiae*. Wiley Interdiscip. Rev. RNA 6:191–209.

Filipovska A, Razif MFM, Nygård KK a, Rackham O. 2011. A universal code for RNA recognition by PUF proteins. Nat. Chem. Biol. 7:425–427.

Francischini CW, Quaggio RB. 2009. Molecular characterization of Arabidopsis thaliana PUF proteins - binding specificity and target candidates. FEBS J. 276:5456–5470.

Funakoshi Y, Doi Y, Hosoda N, Uchida N, Osawa M, Shimada I, Tsujimoto M, Suzuki T, Katada T, Hoshino SI. 2007. Mechanism of mRNA deadenylation: Evidence for a molecular interplay between translation termination factor eRF3 and mRNA deadenylases. Genes Dev. 21:3135–3148.

Gavis ER. 2001. Over the rainbow to translational control. Nat. Struct. Biol. 8:387–389.

Geisler M, Murphy AS. 2006. The ABC of auxin transport: the role of p-glycoproteins in plant development. FEBS Lett. 580:1094–102.

Gerber AP, Herschlag D, Brown PO. 2004. Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. PLoS Biol. 2:e79.

Gerstberger S, Hafner M, Tuschl T. 2014. A census of human RNA-binding proteins. Nat. Rev. Genet. 15:829–845.

Goldstrohm AC, Hook B a, Seay DJ, Wickens M. 2006. PUF proteins bind Pop2p to regulate messenger RNAs. Nat. Struct. & Mol. Biol. 13:533–539.

Gray NK, Hentze MW. 1994. Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. EMBO J. 13:3882–91.

Gu W, Deng Y, Zenklusen D, Singer RH. 2004. A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. Genes Dev. 18:1452–65.

Gupta YK, Lee TH, Edwards T a, Escalante CR, Kadyrova LY, Wharton RP, Aggarwal AK. 2009. Co-occupancy of two Pumilio molecules on a single hunchback NRE. RNA 15:1029–1035.

Gupta YK, Nair DT, Wharton RP, Aggarwal AK. 2008. Structures of Human Pumilio with Noncognate RNAs Reveal Molecular Mechanisms for Binding Promiscuity. Structure 16:549–557.

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp A-C, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–41.

Hall MD, Levitt DG, Banaszak LJ. 1992. Crystal structure of Escherichia coli malate

dehydrogenase. A complex of the apoenzyme and citrate at 1.87 A resolution. J. Mol. Biol. 226:867–82.

Hall TMT. 2014. Expanding the RNA-recognition code of PUF proteins. Nat. Struct. Mol. Biol. 21:653–655.

Hang R, Liu C, Ahmad A, Zhang Y, Lu F, Cao X. 2014. *Arabidopsis* protein arginine methyltransferase 3 is required for ribosome biogenesis by affecting precursor ribosomal RNA processing. Proc. Natl. Acad. Sci. 111:16190–16195.

Henras AK, Plisson-Chastang C, O'Donohue M-F, Chakraborty A, Gleizes P-E. 2015. An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA 6:225–242.

Hentze MW, Preiss T. 2010. The REM phase of gene regulation. Trends Biochem. Sci. 35:423–426.

Hentze MW, Rouault T a, Caughman SW, Dancis a, Harford JB, Klausner RD. 1987. A cis-acting element is necessary and sufficient for translational regulation of human ferritin expression in response to iron. Proc. Natl. Acad. Sci. U. S. A. 84:6730–4.

Hettema EH, Erdmann R, van der Klei I, Veenhuis M. 2014. Evolving models for peroxisome biogenesis. Curr. Opin. Cell Biol. 29:25–30.

Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. 19:6860–6869.

Hu J, Baker A, Bartel B, Linka N, Mullen RT, Reumann S, Zolman BK. 2012. Plant Peroxisomes: Biogenesis and Function. Plant Cell 24:2279–2303.

Huang T, Kerstetter R a., Irish VF. 2014. APUM23, a PUF family protein, functions in leaf development and organ polarity in Arabidopsis. J. Exp. Bot. 65:1181–1191.

Huh SU, Kim MJ, Paek K-H. 2013. Arabidopsis Pumilio protein APUM5 suppresses

Cucumber mosaic virus infection via direct binding of viral RNAs. Proc. Natl. Acad. Sci. U. S. A. 110:779–84.

Huh SU, Paek K-H. 2014. APUM5, encoding a Pumilio RNA binding protein, negatively regulates abiotic stress responsive gene expression. BMC Plant Biol. 14:75.

Imataka H, Gradi a, Sonenberg N. 1998. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. EMBO J. 17:7480–9.

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JWB, Maquat LE. 2008. Upf1 Phosphorylation Triggers Translational Repression during Nonsense-Mediated mRNA Decay. Cell 133:314–327.

Jan CH, Friedman RC, Ruby JG, Bartel DP. 2011. Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. Nature 469:97–101.

Jin L, Guzik BW, Bor Y, Rekosh D, Hammarskjöld M-L. 2003. Tap and NXT promote translation of unspliced mRNA. Genes Dev. 17:3075–86.

Kawaguchi R, Bailey-Serres J. 2005. mRNA sequence features that contribute to translational regulation in Arabidopsis. Nucleic Acids Res. 33:955–965.

Kishore S, Luber S, Zavolan M. 2010. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. Brief. Funct. Genomics 9:391–404.

Koš M, Tollervey D. 2010. Yeast Pre-rRNA Processing and Modification Occur Cotranscriptionally. Mol. Cell 37:809–820.

Larsson O, Perlman DM, Fan D, Reilly CS, Peterson M, Dahlgren C, Liang Z, Li S, Polunovsky V a., Wahlestedt C, et al. 2006. Apoptosis resistance downstream of eIF4E: Posttranscriptional activation of an anti-apoptotic transcript carrying a consensus hairpin structure. Nucleic Acids Res. 34:4375–4386.

Liu G. 2002. Interactions of Elongation Factor 1alpha with F-Actin and beta -Actin mRNA: Implications for Anchoring mRNA in Cell Protrusions. Mol. Biol. Cell 13:579–592.

Lu G, Hall TMT. 2011. Alternate modes of cognate RNA recognition by human PUMILIO proteins. Structure 19:361–367.

Maillet L, Collart M a. 2002. Interaction between Not1p, a component of the Ccr4-Not complex, a global regulator of transcription, and Dhh1p, a putative RNA helicase. J. Biol. Chem. 277:2835–2842.

Mamane Y, Petroulakis E, Martineau Y, Sato T-A, Larsson O, Rajasekhar VK, Sonenberg N. 2007. Epigenetic activation of a subset of mRNAs by eIF4E explains its effects on cell proliferation. PLoS One 2:e242.

Marzluff WF, Wagner EJ, Duronio RJ. 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. 9:843–854.

Matsushita K, Takeuchi O, Standley DM, Kumagai Y, Kawagoe T, Miyake T, Satoh T, Kato H, Tsujimura T, Nakamura H, et al. 2009. Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. Nature 458:1185–1190.

Mayr C. 2015. Evolution and Biological Roles of Alternative 3’ UTRs. Trends Cell Biol. xx:1–11.

Miller M a, Olivas WM. 2011. Roles of Puf proteins in mRNA degradation and translation. Wiley Interdiscip. Rev. RNA 2:471–92.

Miller MT, Higgin JJ, Hall TMT. 2008. Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. Nat. Struct. Mol. Biol. 15:397–402.

Moore MJ. 2005. From birth to death: the complex lives of eukaryotic mRNAs. Science 309:1514–1518.

Muench DG, Mullen RT. 2003. Peroxisome dynamics in plant cells: A role for the cytoskeleton. Plant Sci. 164:307–315.

Muench DG, Park N-I. 2006. Messages on the move: the role of the cytoskeleton in mRNA localization and translation in plant cellsThis review is one of a selection of papers published in the Special Issue on Plant Cell Biology. Can. J. Bot. 84:572–580.

Müller M, Heym RG, Mayer A, Kramer K, Schmid M, Cramer P, Urlaub H, Jansen R-P, Niessing D. 2011. A cytoplasmic complex mediates specific mRNA recognition and localization in yeast. PLoS Biol. 9:e1000611.

Mullineux S-T, Lafontaine DLJ. 2012. Mapping the cleavage sites on mammalian pre-rRNAs: Where do we stand? Biochimie 94:1521–1532.

Nagy E, Rigby WFC. 1995. Glyceraldehyde-3-phosphate dehydrogenase selectively binds AU-rich RNA in the NAD(+)-binding region (Rossmann fold). J Biol Chem 270:2755–63.

Oleynikov Y, Singer RH. 2003. Real-time visualization of ZBP1 association with beta-actin mRNA during transcription and localization. Curr. Biol. 13:199–207.

Osheim YN, French SL, Keck KM, Champion E a., Spasov K, Dragon F, Baserga SJ, Beyer AL. 2004. Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in *Saccharomyces cerevisiae*. Mol. Cell 16:943–954.

Paquin N, Chartrand P. 2008. Local regulation of mRNA translation: new insights from the bud. Trends Cell Biol. 18:105–111.

Platta HW, El Magraoui F, Schlee D, Grunau S, Girzalsky W, Erdmann R. 2007. Ubiquitination of the peroxisomal import receptor Pex5p is required for its recycling. J. Cell Biol. 177:197–204.

Qiu C, McCann KL, Wine RN, Baserga SJ, Hall TMT. 2014. A divergent Pumilio repeat protein family for pre-rRNA processing and mRNA localization. Proc. Natl. Acad.

Sci. 111:18554–18559.

Quenault T, Lithgow T, Traven A. 2011. PUF proteins: repression, activation and mRNA localization. Trends Cell Biol. 21:104–112.

Ross AF, Oleynikov Y, Kislauskis EH, Taneja KL, Singer RH. 1997. Characterization of a beta-actin mRNA zipcode-binding protein. Mol. Cell. Biol. 17:2158–2165.

Ruud K a, Kuhlow C, Goss DJ, Browning KS. 1998. Identification and Characterization of a Novel Cap-binding Protein from Arabidopsis thaliana. J. Biol. Chem. 273:10325–10330.

Sandler H, Kreth J, Timmers HTM, Stoecklin G. 2011. Not1 mediates recruitment of the deadenylase Caf1 to mRNAs targeted for degradation by tristetraprolin. Nucleic Acids Res. 39:4373–86.

Scherrer T, Mittal N, Janga SC, Gerber AP. 2010. A Screen for RNA-Binding Proteins in Yeast Indicates Dual Functions for Many Enzymes. PLoS One 5:e15499.

Shen Z, Paquin N, Forget A, Chartrand P. 2009. Nuclear Shuttling of She2p Couples ASH1 mRNA Localization to its Translational Repression by Recruiting Loc1p and Puf6p. Mol. Biol. Cell 20:2265–2275.

Singh G, Pratt G, Yeo GW, Moore MJ. 2015. The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annu. Rev. Biochem. 84:150317182619002.

Skaar JR, Pagan JK, Pagano M. 2013. Mechanisms and function of substrate recruitment by F-box proteins. Nat. Rev. Mol. Cell Biol. 14:369–381.

Smith JJ, Aitchison JD. 2013. Peroxisomes take shape. Nat. Rev. Mol. Cell Biol. 14:803–17.

Su W, Slepenkov S V, Slevin MK, Lyons SM, Ziemniak M, Kowalska J, Darzynkiewicz E, Jemielity J, Marzluff WF, Rhoads RE. 2013. mRNAs containing the histone 3’ stem-loop are degraded primarily by decapping mediated by oligouridylation of the

3' end. Rna 19:1–16.

Tam PPC, Barrette-Ng IH, Simon DM, Tam MWC, Ang AL, Muench DG. 2010. The Puf family of RNA-binding proteins in plants: phylogeny, structural modeling, activity and subcellular localization. BMC Plant Biol. 10:44.

Terasaka K, Blakeslee JJ, Titapiwatanakun B, Peer W a, Bandyopadhyay A, Makam SN, Lee OR, Richards EL, Murphy AS, Sato F, et al. 2010. PGP4 , an ATP Binding Cassette P-Glycoprotein , Catalyzes Auxin Transport in Arabidopsis thaliana Roots. Mol. Cell. Biol. 17:1–18.

Thomson E, Rappsilber J, Tollervey D. 2007. Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast. RNA 13:2165–74.

Tian B, Hu J, Zhang H, Lutz CS. 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res. 33:201–212.

Tilsner J, Linnik O, Christensen NM, Bell K, Roberts IM, Lacomme C, Oparka KJ. 2009. Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. Plant J. 57:758–770.

Ule J, Jensen K, Mele A, Darnell RB. 2005. CLIP: a method for identifying protein-RNA interaction sites in living cells. Methods 37:376–86.

Vindry C, Vo Ngoc L, Kruys V, Gueydan C. 2014. RNA-binding protein-mediated post-transcriptional controls of gene expression: Integration of molecular mechanisms at the 3′ end of mRNAs? Biochem. Pharmacol. 89:431–440.

Wang X, McLachlan J, Zamore PD, Hall TMT. 2002. Modular recognition of RNA by a human pumilio-homology domain. Cell 110:501–12.

Wang Y, Cheong C-G, Hall TMT, Wang Z. 2009. Engineering splicing factors with designed specificities. Nat. Methods 6:825–830.

Wang Y, Opperman L, Wickens M, Hall TMT. 2009. Structural basis for specific

recognition of multiple mRNA targets by a PUF regulatory protein. Proc. Natl. Acad. Sci. U. S. A. 106:20186–20191.

Wang Y, Wang Z, Tanaka Hall TM. 2013. Engineered proteins with Pumilio/ *fem-3* mRNA binding factor scaffold to manipulate RNA metabolism. FEBS J. 280:3755–3767.

Weis BL, Kovacevic J, Missbach S, Schleiff E. 2015. Plant-Speci fi c Features of Ribosome Biogenesis. Trends Plant Sci. 20:729–740.

White EK, Moore-Jarrett T, Ruley HE. 2001. PUM2, a novel murine puf protein, and its consensus RNA-binding site. RNA 7:1855–1866.

Wilinski D, Qiu C, Lapointe CP, Nevil M, Campbell ZT, Tanaka Hall TM, Wickens M. 2015. RNA regulatory networks diversified through curvature of the PUF protein scaffold. Nat. Commun. 6:8213.

Wreden C, Verrotti a C, Schisa J a, Lieberfarb ME, Strickland S. 1997. Nanos and pumilio establish embryonic polarity in Drosophila by promoting posterior deadenylation of hunchback mRNA. Development 124:3015–3023.

Xiang Y, Nakabayashi K, Ding J, He F, Bentsink L, Soppe WJJ. 2014. REDUCED DORMANCY5 Encodes a Protein Phosphatase 2C That Is Required for Seed Dormancy in Arabidopsis. Plant Cell 26:4362–4375.

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. 2005. Systematic discovery of regulatory motifs in human promoters and 3’ UTRs by comparison of several mammals. Nature 434:338–345.

Zamore PD, Bartel DP, Lehmann R, Williamson JR. 1999. The PUMILIO-RNA interaction: A single RNA-binding domain monomer recognizes a bipartite target sequence. Biochemistry 38:596–604.

Zhang C, Muench DG. 2015. A Nucleolar PUF RNA-binding Protein with Specificity for a Unique RNA Sequence. J. Biol. Chem. 290:30108–30118.

Zheng D, Ezzeddine N, Chen C-Y a, Zhu W, He X, Shyu A-B. 2008. Deadenylation is prerequisite for P-body formation and mRNA decay in mammalian cells. J. Cell Biol. 182:89–101.

Zhu D, Stumpf CR, Krahn JM, Wickens M, Hall TMT. 2009. A 5’ cytosine binding pocket in Puf3p specifies regulation of mitochondrial mRNAs. Proc. Natl. Acad. Sci. U. S. A. 106:20192–20197.

APPENDIX 1 - CLONING STRATEGY, PRIMER, VECTOR CHOICE AND PURPOSE

Experimental purpose	Glycerol stock number	Product	Vector and tag	Primer name and sequence	Template used
Prokaryotic protein expression (SELEX and EMSA)	956	APUM13 HD (last 340 AAs)	pUC57	Synthesized by GenScript	
	957	APUM13 HD (last 340 AAs)	pGEX6p GST-gene	Subcloned from #956	
	960	APUM12 HD (last 339 AAs)	pGEX6p GST-gene	(13-Jun-2011) APUM12_BamHI_258_5': GAAATTGGATCCCTCACAATGAGTCTCAATAATCTG APUM12_STOP_SalI_3': GAAATTGTCGACTTACTTCTTCGAGCTAAGTGC	cDNA
	964	APUM18 coding sequence (327 AAs)	pGEX6p GST-gene	(13-Jun-2011) APUM18_BamHI_1_5': GAAATTGGATCCATGGCAGTCGCTGATAATCC APUM18_STOP_SalI_3': GAAATTGTCGACTCAGCAACGAAGCCTAAATGA	cDNA
	968	APUM6 HD (last 366 AAs)	pGEX6p GST-gene	(27-Sep-2011) APUM6_BamHI_496_5': GAAATTGGATCCCCTGGATGGCAGCCACAA APUM6_STOP_SalI_3': GAAATTGTCGACTCATCTCCTCAATTCTTGGTTTTC	cDNA
	976	APUM23 HD (last 454 AAs)	pGEX6p GST-gene	(02-Mar-2012) APUM23_BamHI_278_5': GAAATTGGATCCCAGTATAGCAGTTTGGTCTTACAG APUM23_STOP_SalI_3': GAAATTGTCGACTCAAATTCTCATTTTATTTGAATG	Riken cDNA clone #671
	977	APUM23 coding sequence (731 AAs)	pGEX6p GST-gene	(02-Apr-2012) APUM23_BamHI_1-5': GAAATTGGATCCATGGGTGAACGAGGAAAGTC APUM23_STOP_SalI_3': GAAATTGTCGACTCAAATTCTCATTTTATTTGAATG	Riken cDNA clone #671
	1012	APUM23 coding sequence (731 AAs)	pQE30 6xHis-gene	Subcloned from #977	

	1037	APUM23 coding sequence (753 AAs)	pGEX6p GST-gene	(06-Dec-2013) APUM23(753)-BamHI-1-5: GAAATTGGATCCATGGTTTCTGTTGGTTCTAAATCA APUM23_STOP_SalI_3': GAAATTGTCGACTCAAATTCTCATTTTATTTGAATG	Riken cDNA clone #671
	1056	ScNop9 coding sequence (666 AAs)	pGEX6p GST-gene	(23-Dec-2013) Nop9_BamHI_1-5': GAAATTGGATCCATGGGAAAGACTAAAACAAGAGGC Nop9_STOP_SalI-3': GAAATTGTCGACTTATCTATAGTGCTTTTGCTTTTTGAAG	Yeast genomic DNA
	1061	*apum23* SHxxQ-1-SHxxR	pGEX6p GST-gene	Synthesized by GenScript	
	1062	*apum23* SHxxR-3-CRxxQ	pGEX6p GST-gene	Synthesized by GenScript	
	1063	*apum23* SLxxQ-4-NYxxQ	pGEX6p GST-gene	Synthesized by GenScript	
	1064	*apum23* SFxxE-10-SHxxR	pGEX6p GST-gene	Synthesized by GenScript	
	1067	*apum23* SHVLQ-1-NYVIQ	pGEX6p GST-gene	Synthesized by GenScript	
	1068	*apum23* SHVLR-3-NYVIQ	pGEX6p GST-gene	Synthesized by GenScript	
	1069	*apum23* SHVLR-3-SNVVE	pGEX6p GST-gene	Synthesized by GenScript	
	1070	*apum23* SHLVE-5-NYVIQ	pGEX6p GST-gene	Synthesized by GenScript	
	1071	*apum23* SFTVE-10-NYVIQ	pGEX6p GST-gene	Synthesized by GenScript	
	1078	*apum23* SRPD-10'-NYVIQ	pGEX6p GST-gene	Synthesized by GenScript	
	1074	APUM24 coding sequence (641 AAs)	pGEX6p GST-gene	(10-Jun-2014) InfusionAPUM24_BamHI_1-5': GGGCCCCTGGGATCCATGTCTTCCAAAGGTCTGAAACC InfusionAPUM24_STOP_SalI-3':	Riken cDNA clone #679

				GCCGCTCGAGTCGACTCATTCAGGTTTCTTGGTTGC	
	1079	HsNop9 coding sequence (636 AAs)	pGEX6p GST-gene	(20-Mar-2015) HsNop9-Infusion-BamHI-1-5: GGGCCCCTGGGATCCATGGGGCAGGGTCCGC HsNop9-Infusion-SalI-3: GCCGCTCGAGTCGACTCAGTCTTCAAGTATGGAGTTCAATGC	MCF7 cell line cDNA
Prokaryotic protein expression for antigen	972	APUM13 AAs 1-239	pRSETα 6xHis-gene	Subcloned from #937	
	1018	APUM6 AAs 50-311	pQE30 6xHis-gene	(13-May-2013) APUM6 50 to 310 5': GAGCGGGGATCCGGCGGCTCAGTTTACAATAAC APUM6 50 to 310 3': GAGCGGGTCGACTTATGTAAAACCGGAAGACTGAAGTAC	#999
	1019	APUM6 AAs 50-512	pQE30 6xHis-gene	(13-May-2013) APUM6 50 to 310 5': GAGCGGGGATCCGGCGGCTCAGTTTACAATAAC APUM6 50-510 3': GAGCGGGTCGACTTATCGAGGGCCATTGGCACCTTC	#999
Transgenic plant construction, endogenous promoter driven	950	APUM6 endo promoter and gene	pCXGFP rsGFP-gene	Multiple primers synthesized in May-Jun-2011	gDNA
	951	APUM12 endo promoter and gene	pCXGFP rsGFP-gene	Multiple primers synthesized in May-Jun-2011	gDNA
	952	APUM13 endo promoter and gene	pCXGFP rsGFP-gene	Multiple primers synthesized in May-Jun-2011	gDNA
	958	APUM6 promoter	pCXGFP rsGFP-gene	(16-May-2011) Pum6 promoter 5' Bam Pum6 promoter 3' Bam: GGCACAGGATCCTTCTCCACAAAGAAGGGA	gDNA
	959	APUM12 promoter	pCXGFP rsGFP-gene	(23-May-2011) Pum12 promoter 5' Bam Pum12 promoter 3' Bam: GGCACAGGATCCCTCGAACTCGATACTATC	gDNA

	975	APUM2 endo promoter and gene	pCXGFP rsGFP-gene	(13-Jun-2011) APUM2_promo_5'BamHI: GGCACAGGATCCTTAAGTAATGTACAGCTTAACTAGGGTGG APUM2_coding_3'BamHI: GGCACAGGATCCCGCCATCTGAGGCTGGGT	gDNA
Transgenic plant construction, inducible promoter driven	997	pTA7002B: modified pTA7002	pTA7002-EGFP-MCS-STOP	MCS: FseI, HpaI, ClaI	
	996	pTA7002C: modified pTA7002	pTA7002-MCS-EGFP-STOP	MCS: XhoI, FseI, HpaI, ClaI	
	1000	pTA7002D: modified pTA7002	pTA7002-2xProteinA-3C-6xHis-9xMyc-MCS-STOP	MCS: FseI, HpaI, ClaI	
	1001	pTA7002E: modified pTA7002	pTA7002-MCS-9xMyc-6xHis-3C-2xProteinA-STOP	MCS: XhoI, FseI, HpaI, ClaI	
	1002	APUM6 genomic DNA driven by inducible promoter	pTA7002	Multiple primers synthesized in Sep-2012 to Apr-2013	gDNA
	999	APUM6 coding sequence driven by inducible promoter	pTA7002B EGFP-gene	Multiple primers synthesized in Sep-2012 to Apr-2013	cDNA clone #699
	998	APUM6 coding sequence driven by inducible promoter	pTA7002C gene-EGFP	Multiple primers synthesized in Sep-2012 to Apr-2013	cDNA clone #699
	1010	APUM6 coding sequence driven by inducible promoter	pTA7002D TAP-gene	Multiple primers synthesized in Sep-2012 to Apr-2013	cDNA clone #699
	1011	APUM6 coding sequence driven by inducible promoter	pTA7002E gene-TAP	Multiple primers synthesized in Sep-2012 to Apr-2013	cDNA clone #699

Transgenic plant construction, His/FLAG or EGFP fused to APUM23	1036	nHF-APUM23 coding sequence (753 AAs)	pCR8 6xHis-FLAG-gene	Multiple primers synthesized in Aug-2013 to Feb-2014	Riken cDNA clone #679
	1052	nHF-APUM23 coding sequence (753 AAs)	pH2GW7 6xHis-FLAG-gene	LR Gateway cloning from 1036	
	1057	APUM23 coding sequence (753 AAs)-cFH	pENTR gene-FLAG-6xHis	Multiple primers synthesized in Aug-2013 to Feb-2014	Riken cDNA clone #679
	1059	APUM23 coding sequence (753 AAs)-cFH	pH2GW7 gene-FLAG-6xHis	LR Gateway cloning from 1057	
	1076	APUM23 coding sequence (753 AAs)	pK7FWG2 gene-EGFP	LR Gateway cloning from 767	
	1041	EGFP-cFH	pCR8 EGFP-FLAG-6xHis	Multiple primers synthesized in Aug-2013 to Feb-2014	
	1043	EGFP-cFH	pH2GW7 EGFP-FLAG-6xHis	LR Gateway cloning from 1041	
Transgenic plant construction and transient expression	905	APUM22 coding sequence	pRTL2 gene-RFP	19-Aug-2009 Pum22-RFP 5'KpnI: GAGAGAGGTACCATGACTCATGAACGCGGC Pum22-RFP 3' NcoI: GAGAGACCATGGAATTATTTCCTTCTCAAACAATTCG	gDNA
	941	APUM6 coding sequence	pK7FWG2 Gene-GFP	LR Gateway cloning from #699	
	942	APUM13 coding sequence	pK7WGF2 GFP-gene	LR Gateway cloning from #828	

APPENDIX 2 – RAW SEQUENCES OF SELEX PRODUCTS

Raw data from the APUM2 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products.

APUM2 SELEX products	Reverse complementary sequence data
2-1	AAGCTTCCTGATCCCCCCGGGGTCCGATAGTCGGGAATTC*
2-2	AAGCTTCCCCACAAACCGGGGTTACAGAGTCGGGAATTC
2-3	AAGCTTCCGGCCACACCACGCGGGGCCCAGTCGGGAATTC
2-4	AAGCTTCCGACACGCAGAAGTCTAACTCAGTCGGGAATTC
2-5	AAGCTTCCAGGTCCCGCCACTACATATCAGTCGGGAATTC
2-6	AAGCTTCCCGTGCTCACGCCATATCTACAGTCGGGAATTC
2-7	AAGCTTCCAGCCGCCGAAACATATATACAGTCGGGAATTC
2-8	AAGCTTCCCCAGAGCGGGTTCATTTACAAGTCGGGAATTC
2-9	AAGCTTCCACACGCCGAAATCTATGTACAGTCGGGAATTC
2-10	AAGCTTCCAACGCACCGAAACTATATACAGTCGGGAATTC
2-11	AAGCTTCCGAATCTCAAGCTATATATACAGTCGGGAATTC
2-12	AAGCTTCCATCCGCCCATCGCCCCCCCCAGTCGGGAATTC
2-13	AAGCTTCCTGGCCACTCAGCTATGTACAAGTCGGGAATTC
2-14	AAGCTTCCACGCGCCGAAAAATATATACAGTCGGGAATTC
2-15	AAGCTTCCGCGCCGAATCGCTTATTTACAGTCGGGAATTC
2-16	AAGCTTCCTACACCCCTAAAGAGAGTACAGTCGGGAATTC
2-17	AAGCTTCCGCGCACCGATATATATATACAGTCGGGAATTC
2-18	AAGCTTCCAGCTCACTACGACCCCGCCCAGTCGGGAATTC
2-19	AAGCTTCCGCACACCGAACAATATTTACAGTCGGGAATTC
2-20	AAGCTTCCCCACAGACCGGGGCATACAGAGTCGGGAATTC
2-21	AAGCTTCCGAACCAAGACGAGCCTATACAGTCGGGAATTC
2-22	AAGCTTCCCGCAGCCCCCCGTGCCTCCCAGTCGGGAATTC
2-23	AAGCTTCCCCGAAAACTCCCTGTCATTTAGTCGGGAATTC
2-24	AAGCTTCCCGCACACGCGGGGCTCGCCGAGTCGGGAATTC
2-25	AAGCTTCCACCCCCCAACGCCGGTCTCCAGTCGGGAATTC
2-26	AAGCTTCCAAGGCCACACAGCTATATACAGTCGGGAATTC
2-27	AAGCTTCCGCGCCCGAAACTTATGTACAGTCGGGAATTC
2-28	AAGCTTCCTCCGCCGAGGCCCTCGGCCAAGTCGGGAATTC
2-29	AAGCTTCCCCTACGCCAACGTATTTACAAGTCGGGAATTC
2-30	AAGCTTCCGACGAACGCACAATATATACAGTCGGGAATTC
2-31	AAGCTTCCGTCTAACCGCGGCCCACTCAAGTCGGGAATTC
2-32	AAGCTTCCACAGGCCACGCGATATGTACAGTCGGGAATTC
2-33	AAGCTTCCATCGCACGCTTATATTTACAAGTCGGGAATTC
2-34	AAGCTTCCGGAGGCCGAACATATTTACAAGTCGGGAATTC
2-35	AAGCTTCCGCGTCCCGCACACTAAGCTCAGTCGGGAATTC
2-36	AAGCTTCCCCCGCGTCCGTCATAACTTCAGTCGGGAATTC
2-37	AAGCTTCCTGCGAGGAACTAATATTTACAGTCGGGAATTC
2-37	AAGCTTCCCCGCTGACCGCTCAAGGCCAAGTCGGGAATTC
2-39	AAGCTTCCGCACGCCGAATAGTATATACAGTCGGGAATTC

2-40	AAGCTTCCCCTCGGTCACAATATTTACAAGTCGGGAATTC
2-41	AAGCTTCCCGCGAACACCAATATTTACAAGTCGGGAATTC
2-42	AAGCTTCCCCTGGCCGCCCCCCACGTCCAGTCGGGAATTC
2-43	AAGCTTCCCCCCCTCCCCCTAAGCGCCCAGTCGGGAATTC
2-44	AAGCTTCCTTACATCCCCGCCGGACCACAGTCGGGAATTC
2-45	AAGCTTCCGAAGCCCGGAGAATATGTACAGTCGGGAATTC
2-46	AAGCTTCCGCCCCAAGCTACCTGGAGCGAGTCGGGAATTC
2-47	AAGCTTCCGAACCTGACCGCTCAAACACAGTCGGGAATTC
2-48	AAGCTTCCACAGGCCGAATGATATTTACAGTCGGGAATTC
2-49	AAGCTTCCCAGACGCCCCTTCGTCCCGGAGTCGGGAATTC
2-50	AAGCTTCCCCCCCTGCCCCCAGCTCGCAAGTCGGGAATTC
2-51	AAGCTTCCGCGCCCCGAAATCTATGTACAGTCGGGAATTC
2-52	AAGCTTCCCCCCACGAGCCCCAGGATCCAGTCGGGAATTC
*blue and green sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	
34 out of 52 sequences (65.4%) contain the UGU core. The chance that UGU appears in a random 20 nucleotide sequence is 28.1%.	
36 sequences were represented in the logo graph constructed by MEME. 8/36 (22.2%) of the sequences contain the primary 8-mer consensus (UGUAUAUA)	

Raw data from the APUM6 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products

APUM6 SELEX products	**Reverse complementary sequence data**
6-1	AAGCTTCCCTACGCAAGCAACTATTTACAGTCGGGAATTC*
6-2	AAGCTTCCACCTACCAAGCATTATTTACAGTCGGGAATTC
6-3	AAGCTTCCTCAGCACGCACAGTATATACAGTCGGGAATTC
6-4	AAGCTTCCTACACAAAAACTACATACACAGTCGGGAATTC
6-5	AAGCTTCCGCCACACGAAACCTATTTACAGTCGGGAATTC
6-6	AAGCTTCCCTACAAGAGTGCCTATTTACAGTCGGGAATTC
6-7	AAGCTTCCAACGAGCACAACATATATACAGTCGGGAATTC
6-8	AAGCTTCCTCTACAACACTTCTACCTACAGTCGGGAATTC
6-9	AAGCTTCCATGCTCTCGGTATTATATACAGTCGGGAATTC
6-10	AAGCTTCCACATAAAGGACATTATCTACAGTCGGGAATTC
6-11	AAGCTTCCGCCACACAAACGCTATATACAGTCGGGAATTC
6-12	AAGCTTCCAGAACAACATCACTATAAACAGTCGGGAATTC
6-13	AAGCTTCCATATACATACAAACACCTACAGTCGGGAATTC
6-14	AAGCTTCCGCCTACGAACACTATTTACAAGTCGGGAATTC
6-15	AAGCTTCCTACACTATACACATTTGACAAGTCGGGAATTC
6-16	AAGCTTCCACCAACACGAACTTATTTACAGTCGGGAATTC
6-17	AAGCTTCCATACAAACGCGCCTATTTACAGTCGGGAATTC
6-18	AAGCTTCCTACACTACGCAACTATTTACAGTCGGGAATTC
6-19	AAGCTTCCACACTACCTACCCTATCTACAGTCGGGAATTC
6-20	AAGCTTCCATCACACTATACATCCTCACAGTCGGGAATTC
6-21	AAGCTTCCTCACAGGAACGCTTATTTACAGTCGGGAATTC
6-22	AAGCTTCCTACACATCAAACATCCTTACAGTCGGGAATTC
6-23	AAGCTTCCTATGGCCATACATATCCTACAGTCGGGAATTC

6-24	AAGCTTCCATGCAACTCAGCTATACTACAGTCGGGAATTC
6-25	AAGCTTCCTACAACGTGACAGTATCTACAGTCGGGAATTC
6-26	AAGCTTCCACCATGCACATCTACTTACAAGTCGGGAATTC
6-27	AAGCTTCCCGGCAATAATTATATACAGCAGTCGGGAATTC
6-28	AAGCTTCCATATACATACACAAAGAAACAGTCGGGAATTC
6-29	AAGCTTCCGCCCTCAAACACCTATTTACAGTCGGGAATTC
6-30	AAGCTTCCACACACACGAACATATTTACAGTCGGGAATTC
6-31	AAGCTTCCGCCTACGAGACCCTATTTACAGTCGGGAATTC
6-32	AAGCTTCCAAAACACTGTCTACACTACAAGTCGGGAATTC
6-33	AAGCTTCCATACATCGCACATTACATACAGTCGGGAATTC
6-34	AAGCTTCCCACATATATACAACCACTACAGTCGGGAATTC
6-35	AAGCTTCCAGCACACCGAACTTATATACAGTCGGGAATTC
6-36	AAGCTTCCAGCCTACTGAACTTATTTACAGTCGGGAATTC
6-37	AAGCTTCCACCACACCAAGCCTATATACAGTCGGGAATTC
6-38	AAGCTTCCTTGCACGAGACGCTATATACAGTCGGGAATTC
6-39	AAGCTTCCGCCTACTAGAACATATATACAGTCGGGAATTC
6-40	AAGCTTCCACCTACACGAACTTATATACAGTCGGGAATTC
6-41	AAGCTTCCGCTACTAACACATATCCTACAGTCGGGAATTC
6-42	AAGCTTCCGCCACTCACACCTATTTACAAGTCGGGAATTC
6-43	AAGCTTCCACACACCAGAGCCTATATACAGTCGGGAATTC
6-44	AAGCTTCCACTCACTCAAACATATATACAGTCGGGAATTC
6-45	AAGCTTCCTACATACACACAACATTCACAGTCGGGAATTC
6-46	AAGCTTCCTACACCAAGCTCCTATTTACAGTCGGGAATTC
6-47	AAGCTTCCACACACCACACGTTATATACAGTCGGGAATTC
6-48	AAGCTTCCAGAGAGCTGAACTATCATACAGTCGGGAATTC
6-49	AAGCTTCCACTCACACGAACTTATTTACAGTCGGGAATTC
6-50	AAGCTTCCACCACTCAAACCTATTTACAAGTCGGGAATTC
6-51	AAGCTTCCTACACTCCACACCTATATACAGTCGGGAATTC
6-52	AAGCTTCCAATCAACATATCGTACATACAGTCGGGAATTC
6-53	AAGCTTCCACATTAGTACACAATCTTACAGTCGGGAATTC
6-54	AAGCTTCCAAACACTCTACATTACCTACAGTCGGGAATTC
6-55	AAGCTTCCAGCGAACGAAACCTATTTACAGTCGGGAATTC
6-56	AAGCTTCCACTCATTTACCGATATATACAGTCGGGAATTC
6-57	AAGCTTCCATACATCCGCCACTATTTACAGTCGGGAATTC
6-58	AAGCTTCCAAAAACATATACTACACTTACAGTCGGGAATTC
6-59	AAGCTTCCTGCACCCCAACGCTATTTACAGTCGGGAATTC
6-60	AAGCTTCCGCCTACGAGAACTTATATACAGTCGGGAATTC
*blue and green sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	
60 out of 60 sequences (100%) contain the UGU core. The chance that UGU appears in a random 20 nucleotide sequence is 28.1%.	
60 sequences were represented in the logo graph constructed by MEME. 17/60 (28.3%) of the sequences contain the primary 8-mer consensus (UGUAUAUA)	

Raw data from the APUM12 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products

APUM12 SELEX products	**Reverse complementary sequence data**

12-1	AAGCTTCCACACTACACAAATAATATACAGTCGGGAATTC*
12-2	AAGCTTCCTAACGATACAAATACTATACAGTCGGGAATTC
12-3	AAGCTTCCATACTATACATACAACACTCAGTCGGGAATTC
12-4	AAGCTTCCTACGGCCTAATGTACATACAAGTCGGGAATTC
12-5	AAGCTTCCAACATACAGACAATTATACAAGTCGGGAATTC
12-6	AAGCTTCCGACAGCGATAATATACATACAGTCGGGAATTC
12-7	AAGCTTCCATCATACAACTATAATATACAGTCGGGAATTC
12-8	AAGCTTCCGGGCGGCTTAATCTACATACAGTCGGGAATTC
12-9	AAGCTTCCCACTAATATACAACACTGACAGTCGGGAATTC
12-10	AAGCTTCCACATGACACAACTATTATACAGTCGGGAATTC
12-11	AAGCTTCCACTAATATACAACATGTTACAGTCGGGAATTC
12-12	AAGCTTCCAGCCAGATACACTAATATACAGTCGGGAATTC
12-13	AAGCTTCCACGCGAATACTATACATACAAGTCGGGAATTC
12-14	AAGCTTCCAACACATACATATACTATACAGTCGGGAATTC
12-15	AAGCTTCCAACTACAGACACTAATATACAGTCGGGAATTC
12-16	AAGCTTCCACCTCCAACCACTAGTATACAGTCGGGAATTC
12-17	AAGCTTCCGATACAGACAATAATATACAAGTCGGGAATTC
12-18	AAGCTTCCTACAATACACAATAATATACAGTCGGGAATTC
12-19	AAGCTTCCTACGGCTACAACTAATATACAGTCGGGAATTC
12-20	AAGCTTCCAACATACAGACTATTATACAAGTCGGGAATTC
12-21	AAGCTTCCCAAGCCATACTATACATACAAGTCGGGAATTC
12-22	AAGCTTCCTACTATACAACTGACATACAAGTCGGGAATTC
12-23	AAGCTTCCCAGCGATTACTATACATACAAGTCGGGAATTC
12-24	AAGCTTCCTCGGCACTACTATACATACAAGTCGGGAATTC
12-25	AAGCTTCCAACAATATACAATATGATACAGTCGGGAATTC
12-26	AAGCTTCCAACAGACTAATATACAACACAGTCGGGAATTC
12-27	AAGCTTCCAGCAACCTAATATACATACAAGTCGGGAATTC
12-28	AAGCTTCCCTAATATACAACACATCTACAGTCGGGAATTC
12-29	AAGCTTCCAAGATACACACTAATATACAAGTCGGGAATTC
12-30	AAGCTTCCTACCTAATATACATACAACAAGTCGGGAATTC
12-31	AAGCTTCCATTAACTACAACTAATATACAGTCGGGAATTC
12-32	AAGCTTCCGCGCAGGACACTAATATACAAGTCGGGAATTC
12-33	AAGCTTCCTATGATACAGCTAAATTACAAGTCGGGAATTC
12-34	AAGCTTCCGAACCAGATAATATACATACAGTCGGGAATTC
12-35	AAGCTTCCACATAATATACATACATAACAGTCGGGAATTC
12-36	AAGCTTCCATAATATACATGTAATATACAGTCGGGAATTC
12-37	AAGCTTCCAATACACACTACTATACACAAGTCGGGAATTC
12-38	AAGCTTCCACTAGTATACAACAATTTACAGTCGGGAATTC
*blue and green sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	
38 out of 38 sequences (100%) contain the UGU core. The chance that UGU appears in a random 20 nucleotide sequence is 28.1%.	
38 sequences were represented in the logo graph constructed by MEME. 21/38 (55.3%) of the sequences contain the primary 9-mer consensus (UGUAUAUUA)	

Raw data from the APUM13 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products

APUM13 SELEX products	**Reverse complementary sequence data**
13-1	AAGCTTCCCGAATGAGCGCTATATATACAGTCGGGAATTC*
13-2	AAGCTTCCACGCTCCGAATATATGTACAAGTCGGGAATTC

13-3	AAGCTTCCCCCAAGGGGTGACACCCCACAGTCGGGAATTC
13-4	AAGCTTCCACCACCCCCGACGCTTTCCCAGTCGGGAATTC
13-5	AAGCTTCCCCGACTTCCGCCGGCCGTTGAGTCGGGAATTC
13-6	AAGCTTCCGACACGCCGCTATATACAAAAGTCGGGAATTC
13-7	AAGCTTCCGAATGCGAGCGAATATATACAGTCGGGAATTC
13-8	AAGCTTCCGCCCACACAGCCTATATACAAGTCGGGAATTC
13-9	AAGCTTCCGAACGAGCGAATTATTTACAGTCGGGAATTC
13-10	AAGCTTCCAACGCACCGAAACTATATACAGTCGGGAATTC
13-11	AAGCTTCCACGCTCCGAAATATATTTACAGTCGGGAATTC
13-12	AAGCTTCCACGCCACGAATATTATTTACAGTCGGGAATTC
13-13	AAGCTTCCCGAATACGCGACGCCTACCGAGTCGGGAATTC
13-14	AAGCTTCCACTAGAAAGCGAACGCGAAAAGTCGGGAATTC
13-15	AAGCTTCCGCACACCGAATCATATTTACAGTCGGGAATTC
13-16	AAGCTTCCCCCAGAGGTGACACCTCAACAGTCGGGAATTC
13-17	AAGCTTCCCGTATCTACACTAGGGGACCAGTCGGGAATTC
13-18	AAGCTTCCCCCACGCCCCAACGCGCCCAAGTCGGGAATTC
13-19	AAGCTTCCGACAGTCACCGAGCCTATACAGTCGGGAATTC
13-20	AAGCTTCCAACGCCCGAATAATATTTACAGTCGGGAATTC
13-21	AAGCTTCCTGACTCAAGCAAATATTTACAGTCGGGAATTC
13-22	AAGCTTCCAAACAGCGACAAATATTTACAGTCGGGAATTC
13-23	AAGCTTCCGCCTACCAAGACATATTTACAGTCGGGAATTC
13-24	AAGCTTCCGCGCCCGAAGTTCTATTTACAGTCGGGAATTC
13-25	AAGCTTCCCGGACCAAATCCCAACGCATAGTCGGGAATTC
13-26	AAGCTTCCTTGCCACGTCGACGGCCGCGAGTCGGGAATTC
13-27	AAGCTTCCAGCTAAGCCGAAATATATACAGTCGGGAATTC
13-28	AAGCTTCCGCGCCCACGAATATATACACAGTCGGGAATTC
13-29	AAGCTTCCACACAAAACCAGTGCGCCAAAGTCGGGAATTC
13-30	AAGCTTCCCGCACCCTCACGCATACAACAGTCGGGAATTC
13-31	AAGCTTCCCTTGCCCCTATCGTTCGCGAAGTCGGGAATTC
13-32	AAGCTTCCGTCAAGGGCAACCTACATCGAGTCGGGAATTC
13-33	AAGCTTCCAAGGCCACACAGCTATATACAGTCGGGAATTC
13-34	AAGCTTCCCCACAGACCGGGGCATACAGAGTCGGGAATTC
13-35	AAGCTTCCGCGGCCGAAACATATTTACAAGTCGGGAATTC
13-36	AAGCTTCCAGCGTACCGAATATATATACAGTCGGGAATTC
13-37	AAGCTTCCACGGAGCAGCTTTATGTACAAGTCGGGAATTC
13-38	AAGCTTCCGCGATCCTCTCCGCCGCGCCAGTCGGGAATTC
13-39	AAGCTTCCAGCTCAGCCCGAAAATATATACAGTCGGGAATTC
13-40	AAGCTTCCACGCCCCGAATAATATATACAGTCGGGAATTC
13-41	AAGCTTCCATACACGGGGCGCTATTTACAGTCGGGAATTC
13-42	AAGCTTCCCGGGACAAGATAGACGTACGAGTCGGGAATTC
13-43	AAGCTTCCACGCCCCGAAATGTATTTACAGTCGGGAATTC
13-44	AAGCTTCCGCGCTTCCGAAAATATTTACAGTCGGGAATTC
13-45	AAGCTTCCGAATACGAGCGAATATATACAGTCGGGAATTC
13-46	AAGCTTCCAGCACCCGAATCATATATACAGTCGGGAATTC
13-47	AAGCTTCCGCGCCCGAATACATATATACAGTCGGGAATTC
13-48	AAGCTTCCACTTACACAGCTATATACACAGTCGGGAATTC
13-49	AAGCTTCCGAACACGCGCCAATATTTACAGTCGGGAATTC
13-50	AAGCTTCCGAACCCTACCGAGCCTATACAGTCGGGAATTC
13-51	AAGCTTCCCCCCCGCGCACCGAGACCGAAGTCGGGAATTC
13-52	AAGCTTCCAACGCCCGAAAAATATGTACAGTCGGGAATTC
13-53	AAGCTTCCACGCCTCCGAAAATATATACAGTCGGGAATTC
13-54	AAGCTTCCACGCCCGAAATATATATACAAGTCGGGAATTC
13-55	AAGCTTCCGAATTAAGCACCATATATACAGTCGGGAATTC

13-56	AAGCTTCCGCCTACACAGCTATATACAAAGTCGGGAATTC
13-57	AAGCTTCCCGACATATATACACGGACCGAGTCGGGAATTC
13-58	AAGCTTCCGCGCCCGAAAATATATTTACAGTCGGGAATTC
13-59	AAGCTTCCGGGCGCCGGCCCCCAACCTGAGTCGGGAATTC
13-60	AAGCTTCCATCGACGCAGCAATCTACAAGTCGGGAATTC
13-61	AAGCTTCCGCACGCCGAAGAATATTTACAGTCGGGAATTC
13-62	AAGCTTCCGCAAACTCAGCTATATACACAGTCGGGAATTC
*blue and green sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	
51 out of 62 sequences (82.3%) contain the UGU core. The chance that UGU appears in a random 20 nucleotide sequence is 28.1%.	
51 sequences were represented in the logo graph constructed by MEME. 21/51 (41.2%) of the sequences contain the primary 8-mer consensus (UGUAUAUA)	

Raw data from the APUM18 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products

APUM18 SELEX products	**Reverse complementary sequence data**
18-1	AAGCTTCCACACAGCCAAGTCAATATACAGTCGGGAATTC*
18-2	AAGCTTCCTGACTGCACCCCACCCCCACAGTCGGGAATTC
18-3	AAGCTTCCCCCCCACCGCCCGCTTACTGAGTCGGGAATTC
18-4	AAGCTTCCTATACAACGTGCATCTAGTGCAGTCGGGAATTC
18-5	AAGCTTCCCCCCCGCTGAACCCTCCCCCAGTCGGGAATTC
18-6	AAGCTTCCCGACAACCACCACCCCTCCCAGTCGGGAATTC
18-7	AAGCTTCCGATCCACGGTTCTTGCCTCAAGTCGGGAATTC
18-8	AAGCTTCCCGACAATACCACCACCTCCTAGTCGGGAATTC
18-9	AAGCTTCCGTATACATATCACGAATCCCAGTCGGGAATTC
18-10	AAGCTTCCCACTCCCACACCCTCACCCCAGTCTCGAGTCGGGAATTC
18-11	AAGCTTCCACCCCTCACCTCCACACTCGAGTCGGGAATTC
18-12	AAGCTTCCGCTCCACGGTTCTTTCCCCAAGTCGGGAATTC
18-13	AAGCTTCCCCAACATTGCCACCACCTCCAGTCGGGAATTC
18-14	AAGCTTCCCCCCCTCCCCTCCCCGCTCGAGTCGGGAATTC
18-15	AAGCTTCCGCTCCACGGTTCTTTCCCCGAGTCGGGAATTC
18-16	AAGCTTCCCAACAACCACCTCCTTCCAAAGTCGGGAATTC
18-17	AAGCTTCCCGACATTACCACCACCTCCCAGTCGGGAATTC
18-18	AAGCTTCCCGACTGCCCCCCCTTCCCTCAGTCGGGAATTC
18-19	AAGCTTCCCACACGCTACAATACTCTACAGTCGGGAATTC
18-20	AAGCTTCCCACACCCTCGCCCACACCCCAGTCGGGAATTC
18-21	AAGCTTCCCCCCGCCACCCTTTACTCGAGTCGGGAATTC
18-22	AAGCTTCCGCCGACCCACACCCGCACCCACACCCGCATTCAGTCGGGAATTC
18-23	AAGCTTCCGCCGACCCACACCCGCACCCACACCCGCATCCAGTCGGGAATTC
18-24	AAGCTTCCCCTACCACGACCCTCCTCCTCCTCCTCCTCCTACAAGTCGGGAATTC
18-25	AAGCTTCCCGCCCACAGATCCTCCTCCTCCTCCTCCTCTATACAGTCGGGAATTC
18-26	AAGCTTCCCCACTCCCTCACCCACTCCCTCACCTTACTGAAGTCGGGAATTC
18-27	AAGCTTCCGCCGACCCACACCCGCACCCACACCTGCATTCAGTCGGGAATTC
18-28	AAGCTTCCCACTCCCTCACCCACTCCCTCACCTTACTGCAGTCGGGAATTC
18-29	AAGCTTCCCCACTCCCTCACCCACTCCCTCACCTTACTGCAGTCGGGAATTC
18-30	AAGCTTCCACCCACGCTCCTCTCTCTCTCTCTTACTCCAGTCGGGAATTC
18-31	AAGCTTCCATATACATATCACAAATTCCGAGTCGGGAATTC

18-32	AAGCTTCCGCCGACCCACACCCGCACCCTCACCCGCATTCAGTCGGGAATTC
18-33	AAGCTTCCCCTACCACGACCCTCCTCCTCCTCCTCCTACAAGTCGGGAATTC
18-34	AAGCTTCCCCACTCCCTCACCCACTCCCTCACCTTACTACAGTCGGGAATTC
18-35	AAGCTTCCGCCGACCCGCTCCCCTCCTCCACCGCCTACAGTCGGGAATTC
*blue and green sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	

Raw data from the APUM23 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products

APUM23 SELEX products	**Reverse complementary sequence data**
23-1	AAGCTTCCGGCACAACCGTCAATTCTCCAGTCGGGAATTC*
23-2	AAGCTTCCAACGCCAAACCGTCAACTCCAGTCGGGAATTC
23-3	AAGCTTCCACGCGTTCCGTCAACTCTACAGTCGGGAATTC
23-4	AAGCTTCCCGTGCAGTCCATCAATTCGCAGTCGGGAATTC
23-5	AAGCTTCCCCCACGCTCCGTCAATTCGTAGTCGGGAATTC
23-6	AAGCTTCCCCAGCAACACCGTCAACTCTAGTCGGGAATTC
23-7	AAGCTTCCCGACACCATCACATCAACTTAGTCGGGAATTC
23-8	AAGCTTCCACCCGTCAATTCGTTAAGCCAGTCGGGAATTC
23-9	AAGCTTCCCAACACCATCACCGTAACTTAGTCGGGAATTC
23-10	AAGCTTCCCGTCAATGCATTCAGTTCCAAGTCGGGAATTC
23-11	AAGCTTCCCGCTATCCGTCAATTCTAAAAGTCGGGAATTC
23-12	AAGCTTCCCCACCTCCATACATAACTTCAGTCGGGAATTC
23-13	AAGCTTCCCCAGCACACCGTCAACTCGTAGTCGGGAATTC
23-14	AAGCTTCCACCGTCAAAACCACAATTCAAGTCGGGAATTC
23-15	AAGCTTCCACCGTCAATTCGTATTCACCAGTCGGGAATTC
23-16	AAGCTTCCCACACGCACCGTCAACTCTCAGTCGGGAATTC
23-17	AAGCTTCCCGGCGGGCCGTCAATTCTTTAGTCGGGAATTC
23-18	AAGCTTCCCCGCACACCACCGTCAAACCATCTTAAGTCGGGAATTC
23-19	AAGCTTCCCACATGCCGTCAATTCGCTGAGTCGGGAATTC
23-20	AAGCTTCCAACCGTCAATTCACACCAAGAGTCGGGAATTC
23-21	AAGCTTCCACGGGCCACCGTCAATTCGTAGTCGGGAATTC
23-22	AAGCTTCCCACATCCATCAACAAACTTCAGTCGGGAATTC
23-23	AAGCTTCCCGTCAATGCCTACAATTCCTAGTCGGGAATTC
23-24	AAGCTTCCCGTCAATACCGTCAACTCCTAGTCGGGAATTC
23-25	AAGCTTCCCACCGTCAATTCATGTTCAGAGTCGGGAATTC
23-26	AAGCTTCCCAACTCCATCAACGAACTTCAGTCGGGAATTC
23-27	AAGCTTCCGTCAATGCACTTCCCACCGTAGTCGGGAATTC
23-28	AAGCTTCCCAACACCATCTCCTCAACTTAGTCGGGAATTC
23-29	AAGCTTCCCCACGCCACGTCAATTCACCAGTCGGGAATTC
23-30	AAGCTTCCCGTCAATGCCTTCAACTTTCAGTCGGGAATTC
23-31	AAGCTTCCCCGCAGCTGCCGTCAACTCAGTCGGGAATTC
23-32	AAGCTTCCATAACACCGTCAATTCACGAAGTCGGGAATTC
23-33	AAGCTTCCCAAGCGATTCCGTCAATTCCAGTCGGGAATTC
23-34	AAGCTTCCCCAGTTTCCGTCAATACACAAGTCGGGAATTC
23-35	AAGCTTCCCAACACCATCACCTCAACTTAGTCGGGAATTC
23-36	AAGCTTCCCAACACCATCACCCTAACTTAGTCGGGAATTC
23-37	AAGCTTCCAGAACTGTCCGTCAATTCACAGTCGGGAATTC
23-38	AAGCTTCCCCACATCCATCAACAACTTCAGTCGGGAATTC
23-39	AAGCTTCCGACGTCCATCAACCAACTTCAGTCGGGAATTC
23-40	AAGCTTCCACCGTCAATACCTCAAATCAAGTCGGGAATTC
23-41	AAGCTTCCCCACGACTCCGTCAATTCACAGTCGGGAATTC
23-42	AAGCTTCCCAACCTCCATACACAACTTCAGTCGGGAATTC

23-43	AAGCTTCCCACCGTGCCGTCAATTCACCAGTCGGGAATTC
23-44	AAGCTTCCCGTCAACACCTCTCCGACCAAGTCGGGAATTC
23-45	AAGCTTCCCTCAGGCAACCGTCAATTCTAGTCGGGAATTC
23-46	AAGCTTCCCAACATCCATCAATGACTTCAGTCGGGAATTC
23-47	AAGCTTCCCCCATCCATCAACTAACTTCAGTCGGGAATTC
23-48	AAGCTTCCACAAACCGTCAATTCTTCAAAGTCGGGAATTC
23-49	AAGCTTCCAGCACACCCGTCAACTCGTAAGTCGGGAATTC
23-50	AAGCTTCCCAACGTCCATCCATAACTTCAGTCGGGAATTC
23-51	AAGCTTCCACACCGTCAACTCCTTGTCCAGTCGGGAATTC
23-52	AAGCTTCCCAACGTCCATCAATGACTTCAGTCGGGAATTC
23-53	AAGCTTCCGGCACCTCCGTCAATTCACAAGTCGGGAATTC
23-54	AAGCTTCCGGGTCGCTCCGTCAATTCGTAGTCGGGAATTC
23-55	AAGCTTCCACCAACCAACCGTCAATTCCAGTCGGGAATTC
23-56	AAGCTTCCCAACATCCATGCAAAACTTCAGTCGGGAATTC
23-57	AAGCTTCCGTCAATGCCCCGCCGACCAAAGTCGGGAATTC
23-58	AAGCTTCCCAACATCCATCCATAACTTCAGTCGGGAATTC
23-59	AAGCTTCCGTCAATGCAAACAAGACCCGAGTCGGGAATTC
23-60	AAGCTTCCCAACACCATCACTGTAACTTAGTCGGGAATTC
23-61	AAGCTTCCACATAACCGTCAATACCTGAAGTCGGGAATTC
23-62	AAGCTTCCCGACATCCATACATGACTTCAGTCGGGAATTC
23-63	AAGCTTCCCAGCACACCATCAATTCTTCAGTCGGGAATTC
23-64	AAGCTTCCCCGTCAATTCCTATAGCCGAAGTCGGGAATTC
23-65	AAGCTTCCCAACCCCATCACATCAACTTAGTCGGGAATTC
23-66	AAGCTTCCCCTCCGTCAATTCATCTGGTAGTCGGGAATTC
23-67	AAGCTTCCACGCAAACCGTCAATTCATGAGTCGGGAATTC
23-68	AAGCTTCCGTCAATGCATCCAGTCCTACAGTCGGGAATTC
23-69	AAGCTTCCCAGCACGGCCGTCAATTCCAAGTCGGGAATTC
23-70	AAGCTTCCCGTCGCTACCGTCAATTCGCAGTCGGGAATTC
23-71	AAGCTTCCCCGCAGTCCGTCAATTCGTCAGTCGGGAATTC
23-72	AAGCTTCCCAACGTCCATACATGACTTCAGTCGGGAATTC
23-73	AAGCTTCCCACCGTCAATTCATTAGTTCAGTCGGGAATTC
23-74	AAGCTTCCCAAGCCCTCCGTCAATTCGCAGTCGGGAATTC
23-75	AAGCTTCCCGCATTCATCAACAAACTTCAGTCGGGAATTC
23-76	AAGCTTCCCGCATCCATCAACGAACTTCAGTCGGGAATTC
23-77	AAGCTTCCACGACTCGCATGTTGACTTGAGTCGGGAATTC
23-78	AAGCTTCCCGCACATCCATCACAACTTCAGTCGGGAATTC
23-79	AAGCTTCCGGCCAACCGTCAATTCGTTTAGTCGGGAATTC
23-80	AAGCTTCCCGATACACTCCGTCAATACCAGTCGGGAATTC
23-81	AAGCTTCCCACAGGCCGTCAATTCATAGTCGGGAATTC
23-82	AAGCTTCCACGCCAATTCCGTCAAATCCAGTCGGGAATTC
23-83	AAGCTTCCAACACgCTCCGTCAATTCATAGTCGGGAATTC
23-84	AAGCTTCCCAACATCCATGCATGACTTCAGTCGGGAATTC
23-85	AAGCTTCCTACACCGTCAACTCATTGTTAGTCGGGAATTC
23-86	AAGCTTCCCCCATAAACCGTCAATTCGAAGTCGGGAATTC
23-87	AAGCTTCCCGTCCGTCAATTCATAATATAGTCGGGAATTC
23-88	AAGCTTCCCAACTACCATACAAAACTTAAGTCGGGAATTC
23-89	AAGCTTCCCCACATCCATCAATAACTTCAGTCGGGAATTC
23-90	AAGCTTCCCAATCACGCCGTCAATTCTCAGTCGGGAATTC
23-91	AAGCTTCCCCACCATCCATCAAAACTTCAGTCGGGAATTC
23-92	AAGCTTCCCAAACGCACGTCAATTCCTAAGTCGGGAATTC
23-93	AAGCTTCCCAACATCCATCGATAACTTCAGTCGGGAATTC
23-94	AAGCTTCCACAAGCACGTCGTCAACTCCAGTCGGGAATTC
23-95	AAGCTTCCCCAACACCACCGTCAATACCAGTCGGGAATTC
23-96	AAGCTTCCACAACATCCATCACAACTTCAGTCGGGAATTC

23-97	AAGCTTCCAAGcCgTgCCGTCAACTCCAAGTCGGGAATTC
23-98	AAGCTTCCCAAACAACCGTCAATACACCAGTCGGGAATTC
23-99	AAGCTTCCGTCAATTCACACCGGTACCCAGTCGGGAATTC
23-100	AAGCTTCCAACGCTCCGTCAATTCATGAAGTCGGGAATTC
23-101	AAGCTTCCACGACATCCATCAAAACTTCAGTCGGGAATTC
23-102	AAGCTTCCGGGCGCATACCGTCAATTCAAGTCGGGAATTC
23-103	AAGCTTCCACGCACCGTCAATTCCTCTAAGTCGGGAATTC
23-104	AAGCTTCCGTCAATGCATCAGTTGCAACAGTCGGGAATTC
23-105	AAGCTTCCCCAACGGGCCGTCAATTCTTAGTCGGGAATTC
23-106	AAGCTTCCCAACACTCCATCATAACTTCAGTCGGGAATTC
23-107	AAGCTTCCACAACGTCCGTCAATGCTAAAGTCGGGAATTC
23-108	AAGCTTCCCCAGCATACCGTCAATTCATAGTCGGGAATTC
23-109	AAGCTTCCACACTACCGTCAATTCATACAGTCGGGAATTC
23-110	AAGCTTCCCCGTCAATGCCTGCAAAACCAGTCGGGAATTC
23-111	AAGCTTCCACACCGTCAATTCAACCCAAAGTCGGGAATTC
23-112	AAGCTTCCCACCGTCAATTCATCATAATAGTCGGGAATTC
*blue and red sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	
95 out of 112 sequences (84.8%) contain the UUGA core. The chance that UUGA appears in a random 20 nucleotide sequence is 8.2%.	
105 sequences were represented in the logo graph constructed by MEME. 43/112 (41.0%) of the sequences contain the primary 10-mer consensus (GAAUUGACGG)	

Raw data from the APUM24 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products

APUM24 SELEX products	**Reverse complementary sequence data**
24-1	GCGCTCGAGTAAGCTTCCCGACCGGGTACGCTTAGGACAGTCGGGAATTCGGATC*
24-2	GCGCTCGAGTAAGCTTCCCCCCATCAACAGGTTTTGATAGTCGGGAATTCGGATC
24-3	GCGCTCGAGTAAGCTTCCCGACACGACGATACGAGCTTAGTCGGGAATTCGGATC
24-4	GCGCTCGAGTAAGCTTCCCCATGGCGTACCCCCCCCCATAGTCGGGAATTCGGATC
24-5	GCGCTCGAGTAAGCTTCCCGACCTCCGTGGTCTATCCTAGTCGGGAATTCGGATC
24-6	GCGCTCGAGTAAGCTTCCCGGCAACAGCGATCGACTGGAGTCGGGAATTCGGATC
24-7	GCGCTCGAGTAAGCTTCCCACCCACACCCTTCTGATCGAGTCGGGAATTCGGATC
24-8	GCGCTCGAGTAAGCTTCCTTCCCGCCCCACGTTATATTAGTCGGGAATTCGGATC
24-9	GCGCTCGAGTAAGCTTCCCAAACTGCGGTCAACCCAGCAGTCGGGAATTCGGATC
24-10	GCGCTCGAGTAAGCTTCCCGCAACGATCCCGGTGGCCAAGTCGGGAATTCGGATC
24-11	GCGCTCGAGTAAGCTTCCCAACAAACTTGCCTGCCGCAAGTCGGGAATTCGGATC
24-12	GCGCTCGAGTAAGCTTCCAGAACGTACGCGCTCCCCTCAGTCGGGAATTCGGATC
24-13	GCGCTCGAGTAAGCTTCCATGTGTGCCATTATGCCGGTAGTCGGGAATTCGGATC
24-14	GCGCTCGAGTAAGCTTCCAATTGAGACTGGACGAGGATAGTCGGGAATTCGGATC
24-15	GCGCTCGAGTAAGCTTCCCGTTACAACTGTCCCCCACCAGTCGGGAATTCGGATC
24-16	GCGCTCGAGTAAGCTTCCTCACGCTCCCACCCCCACGGAGTCGGGAATTCGGATC
24-17	GCGCTCGAGTAAGCTTCCAAAGACATCTGGCCTACACTAGTCGGGAATTCGGATC
24-18	GCGCTCGAGTAAGCTTCCGACCAGTCACGTTTAACACCAGTTGGGAATTCGGATC
24-19	GCGCTCGAGTAAGCTTCCCTTAGAGACCTCACACACGAAGTCGGGAATTCGGATC
24-20	GCGCTCGAGTAAGCTTCCCATACCTTCCGCTTACACCTAGTCGGGAATTCGGATC
24-21	GCGCTCGAGTAAGCTTCCAGATCCGGCTTGAGGGTTCTAGTCGGGAATTCGGATC
24-22	GCGCTCGAGTAAGCTTCCCCACCTGGCTAACCCTGCCAAGTCGGGAATTCGGATC
24-23	GCGCTCGAGTAAGCTTCCCAGCAGTCCACTAACTGGACAGTCGGGAATTCGGATC
24-24	GCGCTCGAGTAAGCTTCCACACGTCAACCCACCAAACAAGTCGGGAATTCGGATC

24-25	GCGCTCGAGTAAGCTTCCTGGTCCCGATACATGTAGGCAGTCGGGAATTCGGATC
24-26	GCGCTCGAGTAAGCTTCCGGCACTACCCACCCCATAAAAGTCGGGAATTCGGATC
24-27	GCGCTCGAGTAAGCTTCCCCAGCCAACCCCTACAAGTCAGTCGGGAATTCGGATC
24-28	GCGCTCGAGTAAGCTTCCCTGACCGACAGGCTACACATAGTCGGGAATTCGGATC
24-29	GCGCTCGAGTAAGCTTCCCATTCGACCTGCCATATAACAGTCGGGAATTCGGATC
24-30	GCGCTCGAGTAAGCTTCCTGCACCCCGAAGTCGTACTTAGTCGGGAATTCGGATC
24-31	GCGCTCGAGTAAGCTTCCCACGCAGCACTGATCTCCCCAGTCGGGAATTCGGATC
24-32	GCGCTCGAGTAAGCTTCCTCCGCCATGGGTCCGGAGACAGTCGGGAATTCGGATC
24-33	GCGCTCGAGTAAGCTTCCCCATGTCCCCACACGTCTTAAGTCGGGAATTCGGATC
24-34	GCGCTCGAGTAAGCTTCCCATCCCCCCCCTAAGGAGAAAGTCGGGAATTCGGATC
24-35	GCGCTCGAGTAAGCTTCCCATGGCACCGGCCCGCACGAAGTCGGGAATTCGGATC
24-36	GCGCTCGAGTAAGCTTCCCGTTGCGGAATCGCACCTACAGTCGGGAATTCGGATC
24-37	GCGCTCGAGTAAGCTTCCCCCCCTTTGACTCCCGAGTCAGTCGGGAATTCGGATC
24-38	GCGCTCGAGTAAGCTTCCAGGCATCTGCGGTAGGAGACAGTCGGGAATTCGGATC
24-39	GCGCTCGAGTAAGCTTCCCATTCTCGTGAGGCGGTCAAAGTCGGGAATTCGGATC
24-40	GCGCTCGAGTAAGCTTCCCTCCCCTTCCATAACATGTAAGTCGGGAATTCGGATC
24-41	GCGCTCGAGTAAGCTTCCTGGGATTATCAGCGCGAACTAGTCGGGAATTCGGATC
24-42	GCGCTCGAGTAAGCTTCCGAAAGACTGGCCATCGACGGAGTCGGGAATTCGGATC
24-43	GCGCTCGAGTAAGCTTCCTCACACAGCCCTCCCGCCATAGTCGGGAATTCGGATC
*blue and green sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	

Raw data from the ScNop9 SELEX experiment showing the complementary nucleotide sequences of individually cloned SELEX products.

ScNop9 SELEX products	**Reverse complementary sequence data**
1	GCGCTCGAGTAAGCTTCCCCCAATAACCCTTGCAGACTAGTCGGGAATTCGGATC
2	GCGCTCGAGTAAGCTTCCACAATTCCACCATATGTTGCAGTCGGGAATTCGGATC
3	GCGCTCGAGTAAGCTTCCGTCCGGTGCCCAGGTCGTTCAGTCGGGAATTCGGATC
4	GCGCTCGAGTAAGCTTCCAAGCTGTGACCCCTAAATACAGTCGGGAATTCGGATC
5	GCGCTCGAGTAAGCTTCCGTGATTGCCCCACAATACCAAGTCGGGAATTCGGATC
6	GCGCTCGAGTAAGCTTCCGTCGCCTGTCCATCTTACCAAGTCGGGAATTCGGATC
7	GCGCTCGAGTAAGCTTCCAGACATTCCGTTCATTACCCAGTCGGGAATTCGGATC
8	GCGCTCGAGTAAGCTTCCAACGTCTGGCCATGCCCCATAGTCGGGAATTCGGATC
9	GCGCTCGAGTAAGCTTCCGTACCGCTGCTACCTAATATAGTCGGGAATTCGGATC
10	GCGCTCGAGTAAGCTTCCAGTCCCCCATACAGTCCACAAGTCGGGAATTCGGATC
11	GCGCTCGAGTAAGCTTCCACCACCGCAAATAGACCACAAGTCGGGAATTCGGATC
12	GCGCTCGAGTAAGCTTCCGTCATACACCACAAGAAAGAAGTCGGGAATTCGGATC
13	GCGCTCGAGTAAGCTTCCGTCAGTCCACTATGATAACAAGTCGGGAATTCGGATC
14	GCGCTCGAGTAAGCTTCCAAAACAGTCCTGTCCACCTAAGTCGGGAATTCGGATC
15	GCGCTCGAGTAAGCTTCCCGTCCGTTCCGGCATGTACCAGTCGGGAATTCGGATC
16	GCGCTCGAGTAAGCTTCCAACATGCACCCCCCTTTAGTAGTCGGGAATTCGGATC
17	GCGCTCGAGTAAGCTTCCACGTACCCCAAAGTAGACAAGTCGGGAATTCGGATC
18	GCGCTCGAGTAAGCTTCCTTTCCCATAAACACACCCGCAGTCGGGAATTCGGATC
19	GCGCTCGAGTAAGCTTCCACCTCCCTAAATATCCAGATAGTCGGGAATTCGGATC
20	GCGCTCGAGTAAGCTTCCTGTTCCATGCATGCTACCCCAGTCGGGAATTCGGATC
21	GCGCTCGAGTAAGCTTCCAACACACCCCAGGTCAACTAAGTCGGGAATTCGGATC
22	GCGCTCGAGTAAGCTTCCTGCCGGCTACACCCCTTAGGAGTCGGGAATTCGGATC
23	GCGCTCGAGTAAGCTTCCAACCAGACCCTCGTTGTAACAGTCGGGAATTCGGATC
24	GCGCTCGAGTAAGCTTCCCACTTCAAAAACCCATCAAGAGTCGGGAATTCGGATC

25	GCGCTCGAGTAAGCTTCCTAATACCACCAAACCATCAAAGTCGGGAATTCGGATC
26	GCGCTCGAGTAAGCTTCCACCTAATGACCATGCTACCCAGTCGGGAATTCGGATC
27	GCGCTCGAGTAAGCTTCCCGCACCCCCCCAAATATGCAAGTCGGGAATTCGGATC
28	GCGCTCGAGTAAGCTTCCAGCTCCCACATAATCCATCCAGTCGGGAATTCGGATC
29	GCGCTCGAGTAAGCTTCCAACGGCGTGCAGCTCATCATAGTCGGGAATTCGGATC
30	GCGCTCGAGTAAGCTTCCCTTAGTCGTCCAGTCCAACAAGTCGGGAATTCGGATC
31	GCGCTCGAGTAAGCTTCCACCACCCAGAATGTACAGCAAGTCGGGAATTCGGATC
32	GCGCTCGAGTAAGCTTCCGACGACGCACCATCCCTCGTAGTCGGGAATTCGGATC
33	GCGCTCGAGTAAGCTTCCGTCAGACACGCCCATCTTTGAGTCGGGAATTCGGATC
34	GCGCTCGAGTAAGCTTCCAAACATCCATGAGGCCAGTAAGTCGGGAATTCGGATC
35	GCGCTCGAGTAAGCTTCCCCACATGTCCTCTGTAAGCCAGTCGGGAATTCGGATC
36	GCGCTCGAGTAAGCTTCCCTACTTACCACCACTCCCCCAGTCGGGAATTCGGATC
37	GCGCTCGAGTAAGCTTCCCTGTCCCCCCAGCATCGCCCAGTCGGGAATTCGGATC
38	GCGCTCGAGTAAGCTTCCATACACCTTCCAATACTACTAGTCGGGAATTCGGATC
39	GCGCTCGAGTAAGCTTCCAACCGTACTACTTCACTCCTAGTCGGGAATTCGGATC
40	GCGCTCGAGTAAGCTTCCCTTGCATCCACCCATGTCATAGTCGGGAATTCGGATC
41	GCGCTCGAGTAAGCTTCCCTACCCAACATCCTTCAACAAGTCGGGAATTCGGATC
42	GCGCTCGAGTAAGCTTCCTGTTCCCTCTACATCACCAGAGTCGGGAATTCGGATC
43	GCGCTCGAGTAAGCTTCCCCGCCAGTTCCTTTATCGTAAGTCGGGAATTCGGATC
44	GCGCTCGAGTAAGCTTCCCCCCCTACTATTATTACCCTAGTCGGGAATTCGGATC
45	GCGCTCGAGTAAGCTTCCCCTAAATCCGCAAAAGACACAGTCGGGAATTCGGATC
46	GCGCTCGAGTAAGCTTCCTCATGTTCCACTGTCGTCCTAGTCGGGAATTCGGATC
47	GCGCTCGAGTAAGCTTCCAGTTCCTTTTAGCCGCACCCAGTCGGGAATTCGGATC
48	GCGCTCGAGTAAGCTTCCAATTCTACCCCTCACAAACAAGTCGGGAATTCGGATC
49	GCGCTCGAGTAAGCTTCCAAATCCGGGGGTGCCACGTAGTCGGGAATTCGGATC
*blue and red sequences are portions of the forward and reverse SELEX primers. Black sequence is the 20 nucleotide region enriched by SELEX	

APPENDIX 3 - SUMMARY OF SOME OF THE TRANSGENIC APPROACHES ATTEMPTED IN ORDER TO IDENTIFY *IN VIVO* TARGETS OF APUM23

A3.1 Subcellular localization of APUM23 fusions to epitope tagged proteins

Numerous constructs were produced in an attempt to express epitope tagged fusions of APUM23 as a strategy to generate transgenic plants that contain epitope-tagged APUM23 or to be used in transient expression. These were: 6xHis-FLAG-APUM23 in the vector pH2GW7, APUM23-FLAG-6xHis in the vector pH2GW7, APUM23-GFP in the vector pK7FWG2 and GFP-FLAG-6xHis in the vector pH2GW7 (control). The construct containing APUM23-GFP was tested for transient expression in fava bean leaf epidermal cells using biolistic particle bombardment, and showed expression in nucleoli in pavement cells (Figure A3.1A). These constructs were transformed into both wild-type background and T-DNA insertion knock-out *apum23* line (CS833857, homozygote). The stable transformation line containing APUM23-EGFP construct showed green fluorescence stained nuclei in multiple cell types (Figure A3.1B-E).

The T-DNA insertion knock-out *apum23* line CS833857 has been reported to show various traits of developmental defects, including late germination, short root at cotyledon stage and small and serrated rosette leaves (Abbasi et al. 2010). Complementation by expression of a 6xHis-FLAG-APUM23 T-DNA construct into an insertion knock-out mutant effectively restored the wild-type phenotype (Figure A3.2).

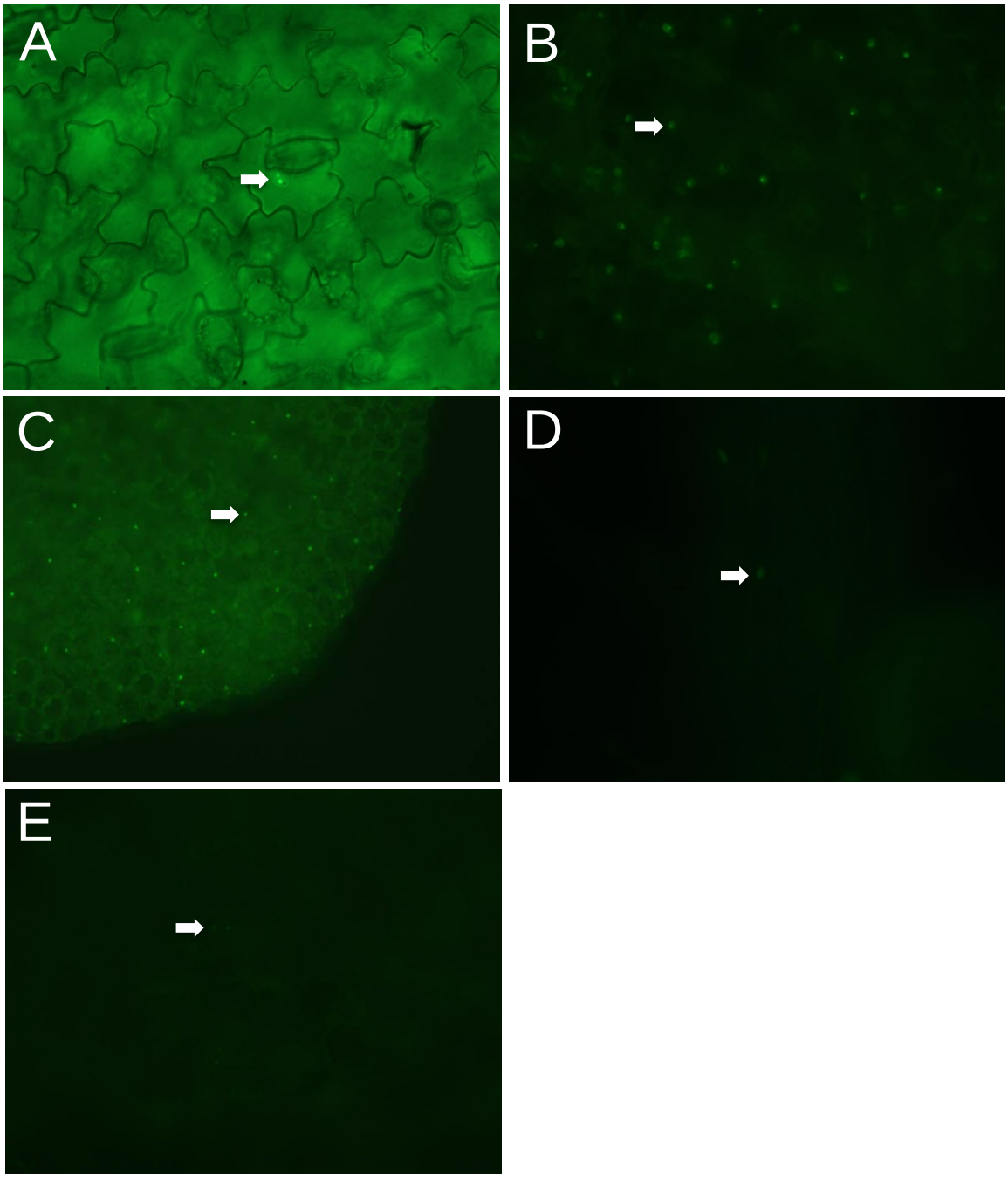


Figure A3.1 Images showing the nucleolar localization of the APUM23-GFP fusion protein. (A) Transient expression in fava bean leaf epidermal cells using biolistic particle bombardment. (B) Stable transformation in Arabidopsis wild type (Col-0) plants showing nucleolar APUM23-GFP expression in epidermal cells from mature rosette leaves. (C) Expression of nucleolar APUM23-GFP in 5 day-old Arabidopsis seedling cotyledons (T1 transformant). (D) Expression in 5 day-old Arabidopsis seedling roots (T1 transformant). (E) Expression in leaf guard cell nucleoli.

Figure A3.2 Phenotypic comparison between T-DNA insertion *apum23* knock-out line (CS833857)(left) and a plant in the same background with the introduction of the construct 6xHis-FLAG-APUM23 in pH2GW7 driven by the 35S CaMV promoter (right). Arabidopsis plants in both pots were cultured in normal condition for 14 days.

A3.2 Western blotting and attempted crosslinking immunoprecipitation using APUM23 transgenic lines

Transgenic Arabidopsis lines expressing epitope-tagged APUM23 would be an ideal resource used to perform *in vivo* experiments to identify APUM23's endogenous RNA targets and associated cofactors. In this study, all constructs made for epitope-tagged APUM23 were expressed with the constitutive 35S CaMV promoter. However, the expression level of these fusion proteins was surprisingly low based on the observation gained from the microscopic studies. The expression was not temporally consistent. For example, 5 or 6 day-old seedlings were the only period at the seedling stage that expression can be effectively traced. Western blotting to test the protein abundance in seedlings or rosette leaves at different stages of development was performed over 30 times, with no consistent results obtained. Although the expected result was obtained occasionally (Figure A3.3), protein degradation or completely loss was seen more frequently. Crosslinking combined with immunoprecipitation as described in Chapter 2 was attempted as well. Unfortunately, neither recombinant APUM23 protein nor 18S rRNA was detected in the IP product.

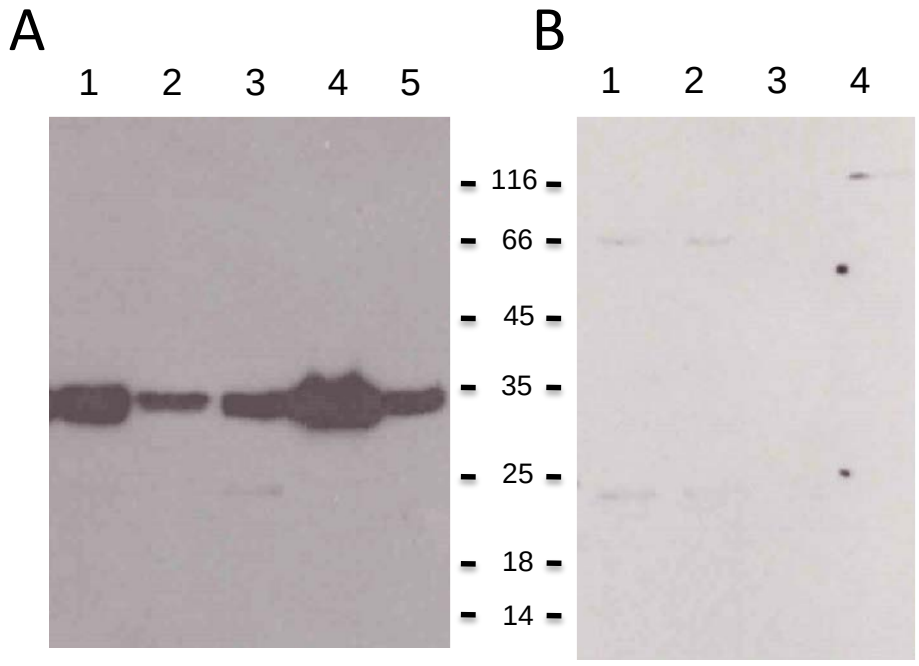


Figure A3.3 Protein gel blot analysis of immunoprecipitated protein from transgenic Arabidopsis seedlings expressing APUM23-FLAG tag fusion protein. FLAG antibody immunoprecipitation experiments of UV-crosslinked Arabidopsis seedling extract was performed. A) 35S::EGFP-cFH control transgenic seedlings expressing EGFP with C-terminal FLAG-6xHis tag fusion. B) 35S::nHF-APUM23 transgenic seedlings expressing APUM23 fused with the N-terminal fusion of 6xHis-FLAG. Anti-FLAG antibody and antibody-conjugated agarose beads were used in this experiment. 1, lysate. 2, pellet from centrifugation of lysate. 3, supernatant after incubation with anti-FLAG antibody conjugated agarose beads. 4, elution from the agarose using 3xFLAG peptide. 5, agarose beads after elution. The protein molecular weights are shown in kD.