

INVESTIGATIONS OF THE MECHANISMS AND APPLICATIONS OF  
PENTATRICOPEPTIDE REPEAT PROTEINS

by

JAMES JOYCE MCDERMOTT

A DISSERTATION

Presented to the Department of Chemistry and Biochemistry  
and the Graduate School of the University of Oregon  
in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

December 2018

## DISSERTATION APPROVAL PAGE

Student: James Joyce McDermott

Title: Investigations of the Mechanisms and Applications of Pentatricopeptide  
Repeat Proteins

This dissertation has been accepted and approved in partial fulfillment of the requirements  
for the Doctor of Philosophy degree in the Department of Chemistry and Biochemistry by:

Diane Hawley	Chairperson
Alice Barkan	Advisor
Mike Harms	Core Member
Vickie DeRose	Core Member
Eric Selker	Institutional Representative

And

Janet Woodruff-Borden Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2018

© 2018 James Joyce McDermott  
This work is licensed under a Creative Commons  
Attribution License



## DISSERTATION ABSTRACT

James Joyce McDermott

Doctor of Philosophy

Department of Chemistry and Biochemistry

December 2018

Title: Investigations of the Mechanisms and Applications of Pentatricopeptide Repeat Proteins

Pentatricopeptide repeat proteins (PPR) proteins are helical-repeat proteins that bind RNA in a modular one-nucleotide:one-repeat fashion. The specificity of a given PPR repeat is dictated by amino acids at two-positions, which recognize a particular nucleotide through hydrogen bonds with the Watson-Crick face. The combinations of amino acids at these positions that give rise to nucleotide specificity is referred to as the PPR-code. The modular and programmable nature of PPR proteins makes them promising candidates for use in applications that require targeting a protein to a specific RNA sequence. One mechanism by which PPR proteins act involves the remodeling of inhibitory RNA hairpins that sequester a ribosome binding site upstream of the gene. However, other evidence suggests that PPR protein-RNA interactions can be inhibited by RNA secondary structure. It is not clear what parameters determine which partner prevails in binding to the RNA. I investigated how the position and strength of an RNA structure impacts PPR:RNA binding and determined that even weak RNA structures are able to inhibit PPR:RNA binding. Additionally, I investigated the driving forces of PPR:RNA binding kinetics. Together, these parameters will benefit the design of synthetic PPR proteins for specific purposes.

Several groups have demonstrated that synthetic PPR proteins can be designed to bind a specified RNA sequence *in vitro*. However, no work has been performed using engineered,

or designer PPR proteins in an *in vivo* setting. I demonstrated the feasibility of using a designer PPR protein to bind a specified RNA *in vivo*, and I applied this capability for a specific application – the purification of an endogenous ribonucleoprotein particle to identify associated proteins.

This dissertation contains unpublished co-authored material.

## CIRRICULUM VITAE

NAME OF AUTHOR: James Joyce McDermott

### GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene  
University of Wisconsin-La Crosse, La Crosse

### DEGREES AWARDED:

Doctor of Philosophy, Chemistry and Biochemistry, 2018, University of Oregon  
Bachelor of Science, Chemistry, 2012, University of Wisconsin-La Crosse

### AREAS OF SPECIAL INTEREST:

Biochemistry  
Molecular Biology  
Synthetic Biology

### PROFESSIONAL EXPERIENCE:

Graduate Research Fellow, Department of Chemistry and Biochemistry,  
University of Oregon, Eugene, 2015 - 2018

Molecular Biology Intern, Digital Assay Feasibility, Flow Diagnostics,  
Intellectual Ventures Laboratory, 2018

Graduate Teaching Fellow, Department of Chemistry and Biochemistry,  
University of Oregon, Eugene, 2013 - 2015, 2018

Research Assistant, Dr. Todd Weaver,  
University of Wisconsin-La Crosse, 2009 - 2012

### GRANTS, AWARDS, AND HONORS:

National Institutes of Health (NIH) Molecular Biology and Biochemistry  
Training Grant (MBBTG) Appointee, University of Oregon, 2015 - 2018

Institute of Molecular Biology (IMB) Graduate Student Representative,  
University of Oregon, 2014 - 2016

Inaugural Biochemistry Representative, Chemistry and Biochemistry  
Graduate Representative Advisory Team (CBGReAT), Department of Chemistry  
and Biochemistry, University of Oregon, 2016 - 2017

Undergraduate Research and Creativity Grant,  
University of Wisconsin-La Crosse, 2011 - 2012

Ray Heath Chemistry Scholarship,  
University of Wisconsin-La Crosse, 2011 - 2012

William Allen Chemistry Scholarship,  
University of Wisconsin-La Crosse, 2010 - 2011

Undergraduate Research Grant,  
American Society for Biochemistry and Molecular Biology, 2010

#### PUBLICATIONS:

McDermott JJ, Williams-Carrier R, Watkins K, Barkan A. Ribonucleoprotein capture via in vivo expression of a designer RNA binding protein. Under review with Nature Biotechnology.

McDermott JJ, Civic B, Barkan A. Factors that influence affinity and binding kinetics of PPR-RNA interactions. Under review with PLOSOne.

Miranda RG, McDermott JJ, Barkan A, RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. Nucleic Acids Research. 2018. 46(5), 2613-2623.

Montoya LA, Shen X, McDermott JJ, Kevil CG, Pluth MD. Mechanistic investigations reveal that dibromobimane extrudes sulfur from biological sulphydryl sources other than hydrogen sulfide. Chemical Science. 2015. 6(1), 294-300.

## ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Alice Barkan, for her mentorship and guidance during my time in her lab and her assistance in preparation of this document. I thank all the current and former Barkan Lab members for their training, support, and feedback. Specifically, Rafael, Kenny, and Margarita for their technical advice and enlightening discussion. I would like to thank the Pete von Hippel Lab for their resources and instruments. I would also like to acknowledge the financial support I received from National Institutes of Health Molecular Biology and Biochemistry Training Grant (NIH MBBTG).

I thank my colleagues in the Department of Chemistry and Biochemistry and the Institute of Molecular Biology (IMB) for their support and friendship, specifically Drs. Matt Hammers and Oggie Golub, and future Drs. Will Storck and Brantly Fulton. I will always remember the laughs I shared with Will Storck and Brantly Fulton throughout the last five years. Without Matt Hammers' friendship during the UO recruiting-weekend, I wouldn't be here. The time with these four gentlemen was too short.

My deepest appreciation goes to Dr. Todd Weaver, my undergraduate research advisor, Dr. Aaron Monte, my undergraduate academic advisor, and the University of Wisconsin-La Crosse Chemistry and Biochemistry Department.

I would like to thank my parents, Jim McDermott and Caron Jagodzinski, for their continued support. I thank them for the sacrifices they made throughout the years to help me get this far. I would not be where I am today without their love and wisdom. I thank my grandparents, Jim and Joyce Jagodzinski, for their interest and support in my scientific education. Finally, I thank my partner, Sophie Sichel, for her continued love, understanding, and companionship during this journey.

To the University of Wisconsin-La Crosse Chemistry and Biochemistry Department and The  
Stone Roses.

## TABLE OF CONTENTS

Chapter	Page
<b>I. INTRODUCTION .....</b>	<b>1</b>
Introduction.....	1
Nuclear encoded RNA binding proteins have co-evolved with RNA metabolism in chloroplasts .....	2
PPR proteins are helical repeat proteins that recognize specific sequences of RNA.....	3
Most functions of P-type PPR proteins can be explained by their ability to sequester long stretches of RNA nucleotides .....	4
The “PPR code”: PPR motifs recognize specific RNA nucleotides through the identity of amino acids at two positions.....	6
Designer PPR proteins: engineered RNA binding proteins for synthetic biology .....	8
Knowledge gap: what is the relationship between RNA structure stability and PPR:RNA binding; and can designer PPR proteins be used for synthetic biology purposes <i>in vivo</i> ?.....	9
Bridge to Chapter II.....	8
<b>II. EFFECTS OF RNA STRUCTURE AND SALT CONCENTRATION ON THE AFFINITY AND KINETICS OF INTERACTIONS BETWEEN PENTATRICOPEPTIDE REPEAT PROTEINS AND THEIR RNA LIGANDS .....</b>	<b>11</b>
Introduction .....	11
Results and discussion .....	13
Impact of RNA secondary structure on PPR10 binding affinity.....	13
Kinetics of PPR-RNA interactions.....	17
Effects of RNA structure on the kinetics of PPR10-RNA interactions.....	20
Contributions of electrostatic interactions to establishing and maintaining the PPR-RNA complex .....	21

Chapter	Page
Summary .....	23
Materials and methods.....	23
Protein expression and purification.....	23
RNA thermal melting analysis .....	24
Gel mobility shift assays .....	24
Surface Plasmon Resonance .....	25
Acknowledgments .....	26
Bridge to Chapter III .....	26
<b>III. RIBONUCLEOPROTEIN CAPTURE VIA <i>IN VIVO</i> EXPRESSION OF A DESIGNER RNA BINDING PROTEIN.....</b>	<b>27</b>
Introduction.....	27
Results and discussion.....	27
Methods.....	35
Development of transgenic lines.....	35
Plant growth .....	35
Chloroplast isolation and fractionation.....	36
Antibodies, SDS-Page, and immunoblot analysis.....	36
Coimmunoprecipitation experiments .....	37
Analysis of immunoprecipitation RNA by RIP-seq and slot blot analysis	38
Acknowledgments .....	39
Bridge to Chapter IV .....	40
<b>IV. SUMMARY AND FUTURE CONSIDERATIONS.....</b>	<b>41</b>
Summary.....	41

Chapter	Page
Implications of the effects of RNA structure on the binding affinity and kinetics of PPR proteins .....	42
Implications and the future applications of designer PPR proteins .....	43
Closing remarks .....	46
APPENDICES.....	47
A. SUPPLEMENTAL MATERIAL FOR CHAPTER II .....	47
B. SUPPLEMENTAL MATERIAL FOR CHAPTER III.....	49
REFERENCES CITED.....	52

## LIST OF FIGURES

Figure	Page
1. PPR protein subfamilies .....	4
2. Mechanisms of PPR proteins.....	6
3. PPR proteins recognize specific nucleotides through the “PPR-code” .....	7
4. Consensus PPR motifs .....	8
5. RNAs used to assess the effects of RNA secondary structure on PPR10 binding .....	15
6. Gel mobility shift assays demonstrating effects of RNA structure on PPR10 binding .....	17
7. Analysis of PPR-RNA interactions by SPR.....	18
8. HCF152- <i>petB</i> RNA binding .....	19
9. Effects of RNA structure on the kinetics of PPR10-RNA interactions .....	21
10. Effect of salt concentration on the kinetics of PPR-atpH RNA .....	22
11. dPPR designs .....	29
12. dPPR chloroplast localization.....	30
13. Analysis of RNAs that coimmunoprecipitate with SCD11 and SCD14.....	31
14. Analysis of the proteins that coimmunoprecipitate with SCD11 and SCD14 .....	33
15. RIP-Seq analysis demonstrating identification of previously-unknown <i>pshA</i> -specific RNA binding proteins in the dPPR-affinity purifications .....	34

# **CHAPTER I**

## **INTRODUCTION**

### **INTRODUCTION**

Ribonucleic acid (RNA) binding proteins mediate many aspects of RNA metabolism and gene expression. These RNA metabolic events include the synthesis, structural rearrangement, post-transcriptional modification, and the processing and degradation of RNAs. From the time of its inception to its degradation, an RNA molecule is covered in a plethora of RNA binding proteins that mediate these events. The proper maintenance of these events drives the life cycle of RNA within a cell and is dependent on the correct functionality of RNA binding proteins. The specificity and affinity of an RNA binding protein is vital to its functionality and the understanding of these binding parameters is leading to the engineering of RNA binding proteins for the development of biotechnological tools surrounding many aspects of gene expression [Lunde et al. 2007, Hall 2016, Chen 2016, Yagi et al. 2013].

Pentatricopeptide repeat (PPR) proteins are alpha helical repeat proteins that recognize specific sequences of RNA. PPR proteins are found in all eukaryotes, but the family has expanded to over 400 members in land plants. PPR proteins are responsible for a multitude of RNA metabolic events in organelles. Each PPR protein is made up of between two and 30 PPR repeats, or motifs, and each PPR motif is made up of about 35 amino acids with conserved sequence and structural features that define it as a PPR motif, but with considerable variability beyond these conserved features [Barkan and Small 2014, Small et al. 2000]. Each repeat recognizes a single RNA nucleotide in a modular one-repeat one-nucleotide fashion. Similar to TALE-repeat and PUF-repeat nucleotide interactions, the

specificity of each PPR repeat is dictated by the identity of amino acids at two positions [Boch et al. 2009, Wang et al. 2002, Barkan et al. 2012]. This modular and predictable binding interface is ideal for engineering customized RNA binding proteins for *in vivo* applications. This dissertation investigates mechanistic details of PPR:RNA binding as well as the applications of engineered PPR proteins for *in vivo* synthetic RNA biology purposes.

### **Nuclear encoded RNA binding proteins have co-evolved with RNA metabolism in chloroplasts**

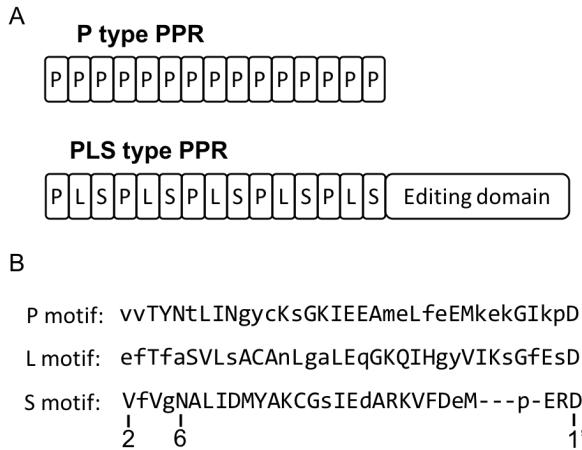
Nuclear encoded proteins, such as PPR proteins, regulate many steps of gene expression and RNA metabolism inside the chloroplast. The chloroplast itself is a product of endosymbiosis and co-evolution between an ancient cyanobacteria and a eukaryotic ancestor cell. Throughout the co-evolution process the chloroplast has transferred many of its original cyanobacterial genes to the nucleus, yet it still retains about 100 of its own genes which include transcriptional and translational components and subunits of the photosynthetic apparatus [Reviewed in Timmis et al. 2004]. Similar to the bacterial ancestor, the chloroplast genome is arranged in polycistronic transcript units. However, the chloroplast has gained many post-transcriptional RNA processing steps that are not found in the bacterial ancestor including cleavage, group II intron splicing, editing, and the processing of full-length transcripts down to shorter isoforms [Barkan 2011]. These events are facilitated by nuclear encoded proteins that harbor RNA binding motifs that are not found in the bacterial ancestors of the DNA comprising organelles, and therefore evolved post-endosymbiosis [Barkan and Small 2014]. The PPR family is a prominent example, and members of the PPR family influence virtually every RNA mediated step in mitochondrial and chloroplast gene expression.

## **PPR proteins are helical-repeat proteins that recognize specific sequences of RNA**

The most widely-recognized RNA binding domains, such as RRM and zinc finger domains, are globular and their RNA recognition modes are idiosyncratic and unpredictable [Lunde et al. 2007]. PPR proteins, on the other hand, are made up of many modular motifs whose nucleotide specificity is, to some extent, predictable and customizable. PPR proteins are made up of many tandem helical motifs that stack to form a super-helix, or alpha solenoid [Small et al. 2000, Shen *et al.* 2016]. Each motif recognizes a single RNA nucleotide and nucleotide specificity is determined by the identity of amino acids at two specific positions. PPR proteins bind to RNA in a “parallel” fashion, that is, the amino-terminus of the protein recognizes the 5’ nucleotides in a binding site and the carboxy-terminus of the protein binds to 3’ nucleotides.

The PPR protein family can be divided into two sub families: P-type, and PLS-type. P-type PPR proteins are comprised entirely of tandem P-type motifs which are all about 35 amino acids in length (Fig 1A). P-type PPR proteins are involved in post-transcriptional RNA isoform processing, as well as cleavage, splicing, and translational activation and repression. Some P-type PPR proteins have additional carboxy-terminal domains such as the MutS-related (SMR) domain, however, the specific function of this domain is unknown [Barkan and Small 2014]. PLS-type PPR proteins are made up of P-type motifs, as well as long (L) and short (S) motifs, which are about 36 amino acids and 31 amino acids respectively (Fig 1B). These motifs are arranged in tandem arrays of PLS triplets (Fig 1A). These proteins almost always contain an additional DYW or E domain which is proposed to be involved in the catalysis of cytidine deamination, (C to U editing) in messenger RNA (mRNA), and also is responsible for recruiting other proteins of the editing complex. There is some evidence that PLS type PPR proteins are involved in cleavage, splicing, and the

processing of transcript isoforms. Both P- and S- type motifs recognize specific RNA nucleotides through the combinatorial action of amino acids at the sixth position in each repeat, and the first position in the following repeat (denoted 6 and 1' in some citations) [Reviewed in Barkan and Small 2014]. In both PLS-type and P-type PPR proteins the PPR motifs define the RNA binding site. The C to U editing function of PLS-type PPR proteins is proposed to be carried out by the carboxy-terminus domain. On the other hand, the diverse functions of P-type PPR proteins are carried out by simply sequestering long segments of RNA – which prevents the RNA from interacting with other proteins or RNA, or exposes a cis-element masked in an RNA hairpin that is then free to interact with other proteins or RNA [Prikryl et al. 2011] [reviewed in Barkan and Small 2014].



**Figure 1. PPR protein subfamilies.**

Figure adapted from Cheng *et al.* 2016. (A) PPR proteins can be divided into P-type or PLS-type proteins. PLS-type proteins almost always contain an editing domain. (B) Consensus motifs of P, L, and S PPR motifs. Capital letters represent highly conserved amino acids.

### Most functions of P-type PPR proteins can be explained by their ability to sequester long stretches of RNA nucleotides

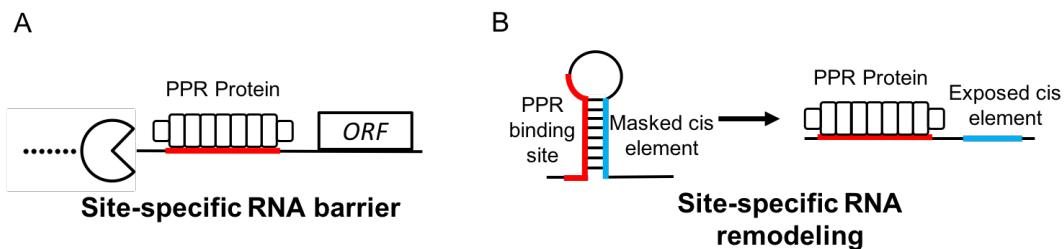
The RNA-metabolic events attributed to P-type PPR proteins can largely be explained by their simple capacity to sequester long stretches of RNA nucleotides. This theme was originally elucidated in studies of a maize protein called PPR10, a protein with 19 PPR repeats. It was observed in *ppr10* mutant plants that processed isoforms of chloroplast *atpH* and *psaJ* transcripts failed to accumulate but their precursors accumulated normally. This

suggested that PPR10 is required to stabilize the processed transcripts that don't accumulate in its absence. It was determined that PPR10 binds intercistronic regions between the *atpI* and *atpH* genes and the *psaJ* and *rpl33* genes and acts as a blockade to exonucleases in the 5' and 3' directions, protecting the *atpH* and *psaJ* genes from exonuclease degradation. By this mechanism PPR10 stabilizes and defines the ends of these processed isoforms. Both of these PPR10 binding sites are similar in sequence and evolutionarily conserved suggesting a sequence specific recognition mechanism by PPR10. Subsequently, it became clear that most RNA termini in chloroplasts are stabilized by PPR proteins, with different members of the family binding different untranslated regions of mRNAs to mediate this effect [Zhelyazkova *et al.* 2012].

Additionally, PPR10 and several other PPR proteins have been shown to activate the translation of genes downstream of their binding sites. *In vitro* data with PPR10 elucidated this mechanism. RNA structure prediction programs predicted PPR10's *atpH* binding site to pair with, and sequester, a ribosome binding site just upstream of the *atpH* start codon. *In vitro* RNase structure probing experiments showed that PPR10 was able to prevent the formation of the RNA hairpin and expose the ribosome binding site [Pfalz *et al.* 2009, Prikryl *et al.* 2011]. Preventing the formation of an RNA hairpin is expected given that PPR proteins bind to the Watson-Crick face of a nucleotide. By sequestering these nucleotides, PPR10 prevents them from basepairing with the ribosome binding site (RBS), exposing the RBS. This suggested an indirect role for PPR10 in translational activation of *atpH* by influencing RNA structure and exposing a masked ribosome binding site.

These findings suggested that many of the RNA metabolic events attributed to PPR proteins could be explained by their ability to sequester long stretches of nucleotides with high affinity. PPR proteins can sequester nucleotides and prevent other proteins, such as

exonucleases, from interacting with that stretch of nucleotides – leading to the stability of processed transcript isoforms (Fig 2A). This sequestration of nucleotides can also prevent the masking of cis-elements by inhibitory RNA hairpins, and by doing so the PPR protein could unmasks cis-elements involved in translational activation, cleavage, and group II intron splicing (Fig 2B) [reviewed in Barkan and Small 2014].

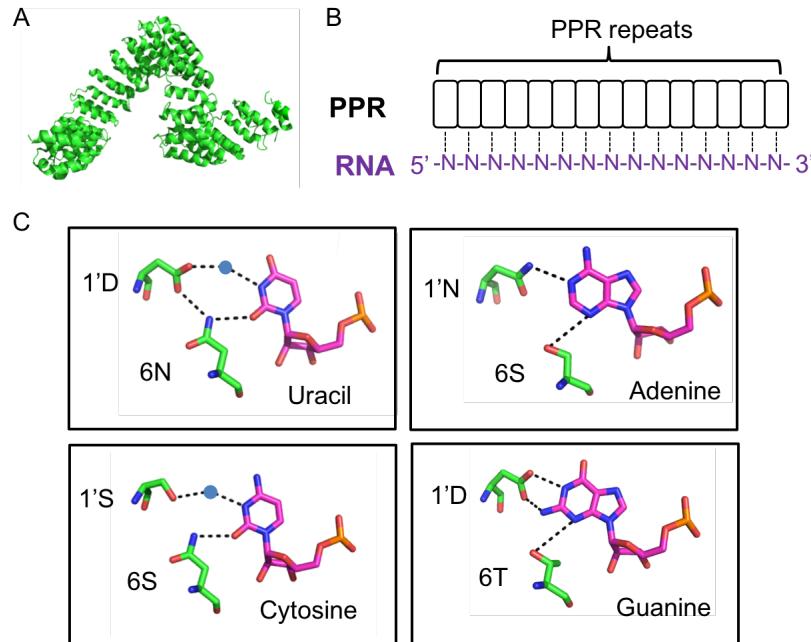


**Figure 2. Mechanisms of PPR proteins.** Figure adapted from Barkan and Small 2014. (A) PPR proteins act as site specific barriers blocking other proteins, such as nucleases, from interacting with the RNA. (B) PPR proteins can influence RNA structure, preventing the formation of RNA hairpins and exposing masked cis-elements like splicing sites, or ribosome binding sites

### The “PPR code”: PPR motifs recognize specific RNA nucleotides through the identity of amino acids at two positions

Before it was recognized that PPR proteins bind RNA, there was only one example of an RNA binding motif consisting of helical repeats – the PUF motif. PUF proteins are helical repeat proteins, independently evolved from PPR proteins, whose repeating motifs recognize RNA through a three-amino-acid code, which was discovered in 2002 [Wang *et al.* 2002]. The sequence specificity of PPR proteins and their repetitive motifs suggested a modular, programmable nucleotide recognition mode similar to PUF-RNA interactions. Through a combination of biochemical and computational approaches that took advantage of the well-defined binding sites of PPR10 and several other PPR proteins, it was determined that the sixth amino acid in each repeat and the first amino acid in the following, carboxy-terminal repeat conferred nucleotide specificity (Fig 2B and C) [Barkan *et al.* 2012].

This “PPR code” was confirmed through a series of biochemical experiments that reprogrammed PPR10 according to the code to bind novel RNA sequences. Subsequently, the structural basis of the code was elucidated through X-ray crystallography of complexes between synthetic PPR proteins (built from consensus PPR motifs) and their RNA ligands (Fig 3C) [Shen *et al.* 2016].



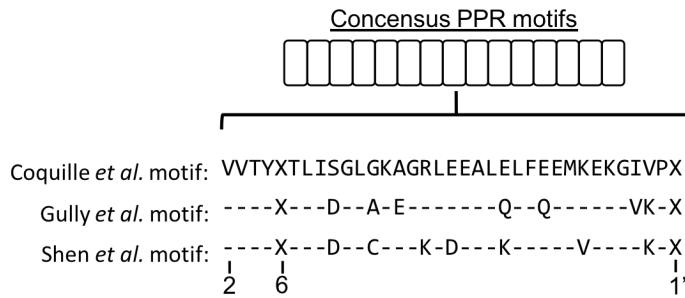
**Figure 3. PPR proteins recognize specific nucleotides through the “PPR-code”.**  
 (A) Crystal structure of PPR10 representing super helical structure (pdb:4M57) [Yin *et al* 2013]. (B) PPR motifs each recognize a single nucleotide. (C) Physical basis for nucleotide specificity of PPR motifs by the 6 and 1' amino acids. Blue circles represent water molecules (pdb: 5I9D, 5I9F, 5I9G, 5I9H) [Shen *et al.* 2016].

Although PPR proteins, to some degree, exhibit a programmable mode of nucleotide recognition, the PPR-code is unable to predict with certainty where native PPR proteins bind *in vivo*. A comprehensive analysis of PPR10’s sequence specificity *in vitro* showed that only a subset of the nucleotides bound with specificity by PPR10 could be explained by the PPR code. Those experiments also showed that the specificity and affinity of the PPR10:*atpH* interaction is not evenly distributed among each individual repeat [Miranda *et al.* 2017]. This

and other data show that native PPR proteins have idiosyncratic features that complicate the prediction of their binding sites and their redesign for new purposes.

### Designer PPR proteins: engineered RNA binding proteins for synthetic biology

To overcome the idiosyncratic features found within native PPR proteins, several groups independently developed consensus PPR motifs to attempt to create more predictable RNA binding proteins based on the PPR architecture (Fig 4) [Shen et al. 2015, Gully et al. 2015, Coquille et al. 2014]. These consensus motifs were created by aligning native P-type PPR motifs found in plants and forming a consensus sequence from the most common amino acids found at each position (aside from the specificity-determining positions). Synthetic, or designer PPR proteins were created using these consensus PPR motifs in combination with the amino acid combinations found in the PPR-code.



**Figure 4. Consensus PPR motifs.** Consensus PPR motifs adapted from Coquille *et al.* 2014, Gully *et al.* 2015, and Shen *et al.* 2015.

The first synthetic PPR proteins were rather simple: only 5-11 repeats long and designed to bind to mononucleotide sequences. Their analysis showed that designer PPR proteins with at least six repeats could bind to their intended target sequences. These proteins showed specificity for their target sequences over other mononucleotide sequences; however the analysis of their sequence specificity was quite superficial [Shen et al. 2015, Gully et al. 2015, Coquille et al. 2014]. In my work, I helped to take this to a more sophisticated level by designing several synthetic PPR proteins (11 and 14 repeats) using previously published

consensus motifs and consensus motifs of my own design. The sequence-specificity of these proteins was then assayed by my collaborator, using an *in vitro* “bind-n-seq” assay [Miranda et al. 2018]. This work showed that the RNA binding specificity landscape of designer PPR proteins could be explained by the PPR-code; that is, PPR tracts made up of consensus motifs do not exhibit the idiosyncratic features that are characteristic of native PPR proteins. The work also suggested that affinity and specificity of designer PPR proteins was greatest when designer PPR proteins are shorter than 12 repeats.

Engineered PPR proteins offer the ability to target specific sequences of RNA through their PPR tracts and to carry out specific functions on RNA through the action of an accessory domain. A *in vivo* PPR protein-based application would not require modification of the RNA sequence itself, and they could be easily targeted to organelles where guide RNAs and oligonucleotides have yet to be successfully targeted. Additionally, the different types of PPR motifs (P, L, and S) offer a flexible template from which the binding kinetics of an engineered PPR protein could be customized to the application. All together these factors offer a promising future for the use of engineered PPR proteins *in vivo*.

**Knowledge gap: What is the relationship between RNA structure stability and PPR:RNA binding; and can designer PPR proteins be used for synthetic biology purposes *in vivo*?**

As described above, PPR proteins, such as PPR10, have the ability to influence local RNA structure around their binding sites. However, other evidence suggests that RNA secondary structure inhibits PPR:RNA binding [Miranda et al. 2017, Kindgren et al. 2015, Zoschke et al 2016], which is to be expected given that PPR motifs make sequence-specific contacts with the Watson-crick face of the aligned nucleotide [Shen *et al.* 2016]. It is unclear

what parameters determine which binding partner the RNA will bind to. In the following chapter I address this question by describing how the stability and position of RNA structure sequestering a PPR binding site influences PPR:RNA binding, a result that has implications for the design of synthetic PPR for *in vivo* use. Additionally, I describe other characteristics of PPR:RNA binding affinity and kinetics in this chapter that aid in our understanding of the mechanisms of PPR:RNA binding and how they compare to globular RNA binding proteins.

My collaboration with Rafael Miranda showed that designer PPR proteins bind to their intended targets with a high degree of specificity *in vitro*. However, there are no reports of exploiting this potential for *in vivo* applications. In Chapter III I develop the use of designer PPR proteins as an *in vivo* technology and describe how they can be used *in vivo* as an “affinity tag” to purify a specified transcript for the purpose of discovering the population of proteins bound to that RNA. Chapter IV is a summary of my work and its future implications in the development of designer PPR proteins as a tool for synthetic RNA biology.

Chapter II of this dissertation contains unpublished co-authored material. I was first author of this work and Alice Barkan was the principle investigator. Bryce Civic was a co-author of this material. Chapter II has been submitted for publication in PLOSONE. Chapter III of this dissertation contains unpublished co-authored material. I was first author of this work and Alice Barkan was the principle investigator. Rosalind Williams-Carrier and Kenneth Watkins co-authored this material. Chapter III has been submitted for publication in Nature Biotechnology.

## BRIDGE TO CHAPTER II

To fully develop the potential of designer PPR proteins for synthetic biology purposes, a better understanding of the details of PPR:RNA interactions are necessary – for example, the kinetics of PPR-RNA interactions and the competition between RNA folding and PPR binding. My research addressed both of these points, as described in the following chapter.

# CHAPTER II

## EFFECTS OF RNA STRUCTURE AND SALT CONCENTRATION ON THE AFFINITY AND KINETICS OF INTERACTIONS BETWEEN PENTATRICOPEPTIDE REPEAT PROTEINS AND THEIR RNA LIGANDS

This chapter contains unpublished co-authored material. Bryce Civic performed some RNA thermal melts. J.J. McDermott performed all other experiments. A. Barkan and J.J. McDermott wrote the manuscript

### INTRODUCTION

Pentatricopeptide repeat (PPR) proteins comprise a large family of RNA binding proteins that function primarily in the context of mitochondrial and chloroplast gene expression [1, 2]. PPR proteins influence every RNA-mediated step in organellar gene expression, including RNA editing, group II intron splicing, RNA stability, and translation. Most PPR proteins act specifically on a handful of RNAs *in vivo*, and this functional specificity is reflected by sequence-specific RNA interactions *in vitro*. PPR proteins consist of tandem degenerate repeats of approximately 35-amino acids, each of which forms a helical hairpin. Consecutive repeats stack to form an elongated superhelix that binds single-stranded RNA. The sequence specificities of PPR proteins are, to some extent, predictable and customizable. Each repeat binds a single nucleotide, with nucleotide specificity dictated by the identities of two amino acids: the sixth amino acid in a given PPR motif and the first amino acid in the next (denoted as the 6 and 1' amino acids according to the nomenclature in ref [3]). These two amino acids form a hydrogen bond network with the Watson-Crick

face of the specified nucleotide [4, 5]. However, this “PPR code” is insufficient to fully explain the sequence specificities of natural PPR proteins, many of which have idiosyncratic features [see, e.g. 6].

PPR proteins are found in all eukaryotes, but the family is particularly large in land plants, where it is made up of more than 400 members containing between two and approximately 30 PPR motifs. These can be divided into two subfamilies, termed P and PLS [7]. PLS proteins consist of variant repeat motifs and function primarily to specify sites of RNA editing [2]. P-type PPR proteins consist primarily of canonical “P-type” motifs, and are involved in group II intron splicing, transcript stabilization, and translational control. It is intriguing that proteins with this simple architecture can elicit such diverse effects on RNA. P-type PPR proteins have only rarely been observed to interact with other proteins [8]; instead, most functions of P-type PPR proteins may result from their capability to form an unusually long protein-RNA interface. For example, many PPR proteins with long repeat tracts stabilize RNA adjacent to their binding sites by blocking exoribonucleases [reviewed in 2]. Furthermore, sequestration of a long RNA segment by a PPR protein can influence local RNA folding [9], which, in turn, may influence RNA stability, processing, or translation. PPR-RNA interactions are inhibited by RNA structures that involve nucleotides in the PPR binding site [6, 10, 11]. This inhibition is to be expected given that PPR motifs bind the Watson-Crick face of nucleobases. However, there is evidence that PPR proteins can bind *in vivo* to RNAs even when a portion of the binding site is complementary to a nearby RNA sequence [9, 11]. In fact, this capability is proposed to underlie the ability of PPR proteins to stimulate translation and group II intron splicing [9]. To develop a better understanding of the parameters that influence the ability of a PPR protein to bind a site capable of pairing with a nearby RNA sequence, we analyzed the effects of RNA structures of varying

stabilities and position on the interaction between the maize protein PPR10 and its *atpH* RNA ligand. Our experiments include the use of Surface Plasmon Resonance (SPR) to examine the kinetics of PPR:RNA interactions, a parameter that is likely to impact the biological functions of PPR proteins and that has, to our knowledge, not been reported previously.

## RESULTS AND DISCUSSION

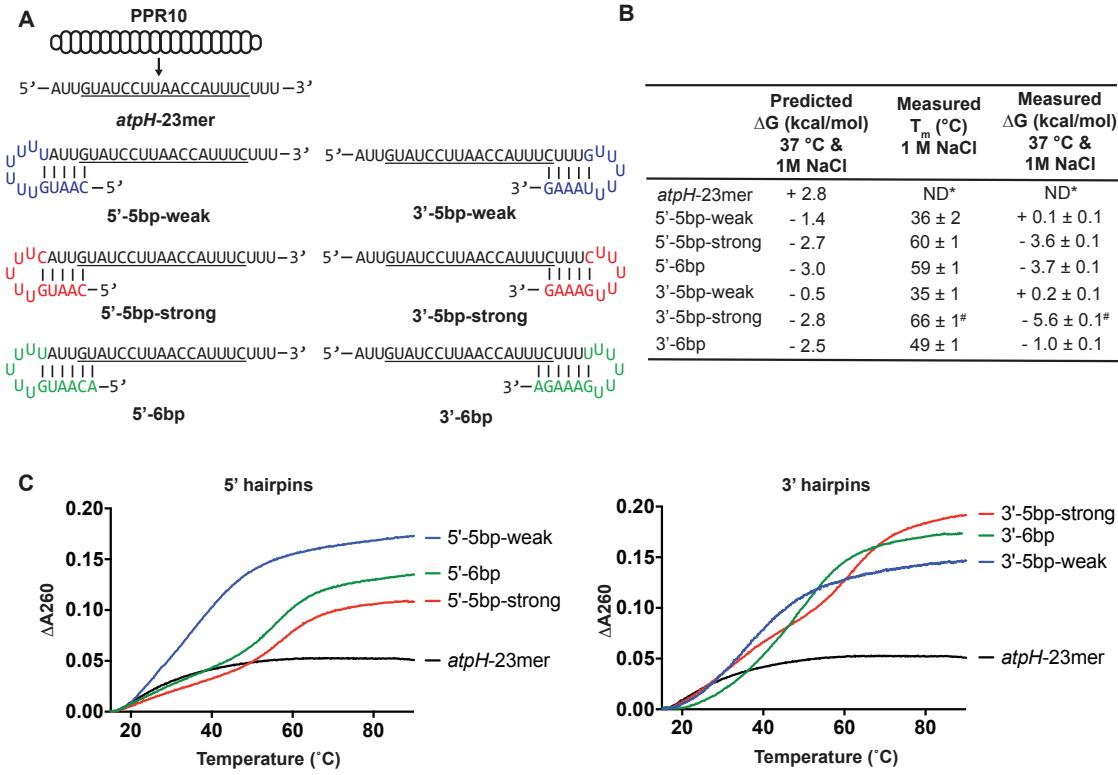
### Impact of RNA secondary structure on PPR10 binding affinity

We selected PPR10 to explore the interplay between RNA folding and PPR binding because PPR10's functions, structure, and sequence specificity have been well characterized [4, 6, 9, 12-15]. PPR10 consists of 19 tandem PPR motifs flanked by capping helices. PPR10 localizes to chloroplasts, where it binds three sites that map in untranslated regions near the *atpH*, *psaJ*, and *psaI* genes. Of these, the site in the *atpH* 5' UTR binds PPR10 with highest affinity, and these interactions have been most thoroughly characterized *in vitro*. PPR10's minimal binding site at *atpH* spans 17-nucleotides, whereas its footprint (the region it protects from exoribonucleases) spans ~23 nucleotides (see Fig 1A). When PPR10 binds this site *in vivo*, it blocks exoribonucleases intruding from both the 5'- and 3'-directions and it also stimulates *atpH* translation. *In vitro* experiments provided evidence that PPR10 activates translation by sequestering RNA that would otherwise form an inhibitory structure with the *atpH* ribosome binding site [9].

To assess the influence of RNA secondary structure on PPR10-RNA interactions, we designed a series of RNAs harboring the PPR10 *atpH* footprint flanked by stem-loops whose stems include nucleotides at either end of the PPR10 binding site (Fig 5A). Constructs were designed such that they had only one predicted structure. Loops were

composed solely of uridines to minimize interactions with other nucleotides in the RNA. We sought to distinguish how the number of binding-site nucleotides sequestered in the stem, the position of those nucleotides in the binding site, and the thermodynamic stability of the RNA structure impact PPR10 binding. Toward that end, we designed RNA hairpins of varying predicted thermodynamic stabilities that intrude on the PPR10 binding site to varying extents (Fig. 5A,B). To distinguish effects of hairpin stability from effects of hairpin position, we designed both “strong” and “weak” RNA structures to sequester the same nucleotides at each end of the binding site.

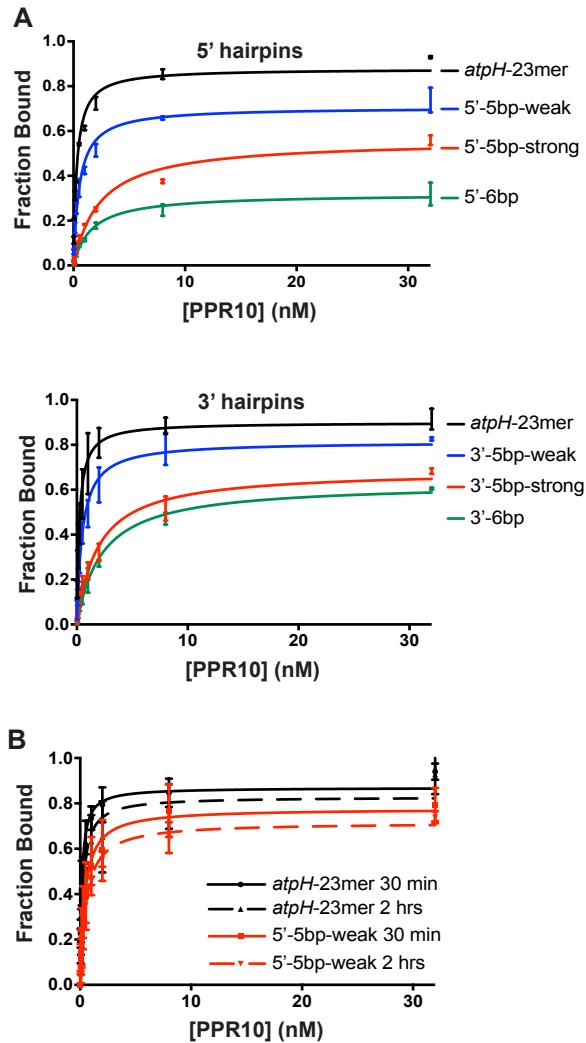
We selected PPR10 to explore the interplay between RNA folding and PPR binding because PPR10’s functions, structure, and sequence specificity have been well characterized [4, 6, 9, 12-15]. PPR10 consists of 19 tandem PPR motifs flanked by capping helices. PPR10 localizes to chloroplasts, where it binds three sites that map in untranslated regions near the *atpH*, *psaJ*, and *psaI* genes. Of these, the site in the *atpH* 5’ UTR binds PPR10 with highest affinity, and these interactions have been most thoroughly characterized *in vitro*. PPR10’s minimal binding site at *atpH* spans 17-nucleotides, whereas its footprint (the region it protects from exoribonucleases) spans ~23 nucleotides (see Fig 5A). When PPR10 binds this site *in vivo*, it blocks exoribonucleases intruding from both the 5'- and 3'-directions and it also stimulates *atpH* translation. *In vitro* experiments provided evidence that PPR10 activates translation by sequestering RNA that would otherwise form an inhibitory structure with the *atpH* ribosome binding site [9].



**Figure 5. RNAs used to assess the effects of RNA secondary structure on PPR10 binding.** (A) Sequences and predicted secondary structures of the RNA ligands. PPR10 is shown aligned to its 23-nt *in vivo* footprint near *atpH* (*atpH*-23mer). PPR10's minimal binding site is underlined [9]. The *atpH*-23mer is not predicted to form any structure. Nucleotides that are appended to the PPR10 footprint to introduce RNA structure are colored. (B) Predicted and measured stabilities of each RNA structure at 1 M NaCl and 2.5  $\mu$ M RNA. Predictions were made with mFold [16], which predicted only one structure for each RNA. The measured  $T_m$  and  $\Delta G$  values were calculated based on thermal melting curves ( $n=3$ , +/- standard error of the mean). Values obtained at 180 mM NaCl, at different RNA concentrations, and from assays performed in reverse (transitions from high to low temperature) are shown in S1A Fig. \*ND- Not determined due to lack of detectable structure. <sup>#</sup>The measured values for the 3'-5bp-strong RNA are based on a single inflection point at 66°C, but the melting curve is biphasic (see panel C). Therefore, these values exaggerate the stability of this structure. (C) Representative melting curves at 1 M NaCl and 2.5  $\mu$ M RNA.

To assess the influence of RNA secondary structure on PPR10-RNA interactions, we designed a series of RNAs harboring the PPR10 *atpH* footprint flanked by stem-loops whose stems include nucleotides at either end of the PPR10 binding site (Fig 5A). Constructs were designed such that they had only one predicted structure. Loops were composed solely of uridines to minimize interactions with other nucleotides in the RNA. We sought to distinguish how the number of binding-site nucleotides sequestered in the stem, the position of those nucleotides in the binding site, and the thermodynamic stability of the RNA structure impact PPR10 binding. Toward that end, we designed RNA hairpins of varying predicted thermodynamic stabilities that intrude on the PPR10 binding site to varying extents (Fig. 5A,B). To distinguish effects of hairpin stability from effects of hairpin position, we designed both “strong” and “weak” RNA structures to sequester the same nucleotides at each end of the binding site.

The maximum fraction of RNA bound was reduced in rough proportion to the stability of the RNA hairpin (Fig 6A), consistent with the anticipated competition between intramolecular RNA interactions and RNA-protein interactions. To determine whether PPR10 can capture additional RNA over time as the RNA transiently unfolds, we determined whether the maximum amount of the 5'-5bp-weak RNA bound to PPR10 increased when the binding reaction was extended from 30 minutes to 2 hours (Fig 6B). The binding curves resulting from the two incubation times were very similar, indicating that the competing binding reactions (PPR10:RNA and intramolecular RNA:RNA) had reached equilibrium by 30 minutes.

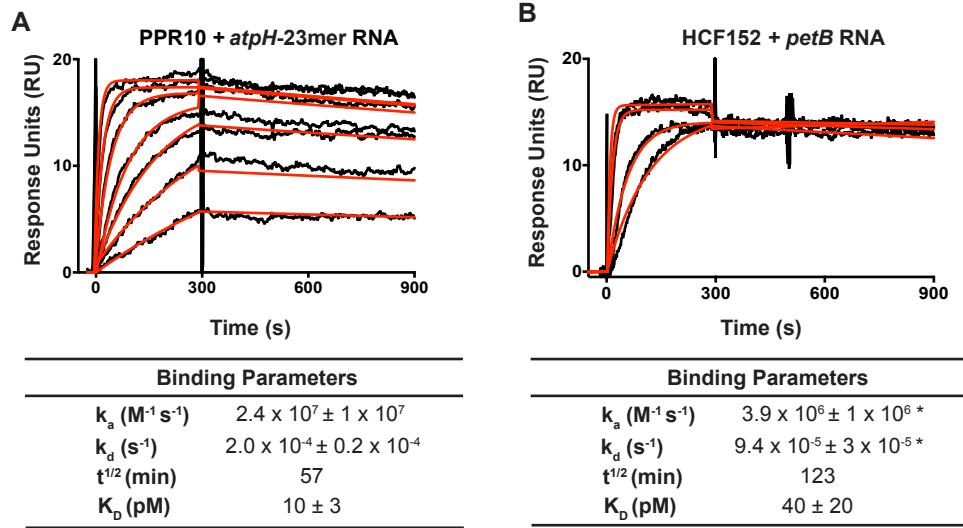


**Fig 6. Gel mobility shift assays demonstrating effects of RNA structure on PPR10 binding.** The RNAs (5 pM) are diagrammed in Fig 1A. PPR10 was used at concentrations of 32 nM, 8 nM, 2 nM six additional 2-fold dilutions. Data for replicate assays ( $n=2$ ) are shown as separate points connected by a vertical line. (A) Summary of binding data for reactions incubated for 30 minutes. Representative gels are shown in S1B Fig. (B) Comparison of results for binding reactions incubated for 30 minutes and 2 hours.

## Kinetics of PPR-RNA interactions

The kinetics of the interactions between PPR proteins and their RNA ligands are likely to impact the outcome of competing protein-RNA and RNA-RNA interactions, and the effects of those interactions on gene expression. To our knowledge, kinetic parameters for PPR-RNA interactions have not been reported. We hypothesized that the long binding interface expected for many PPR-RNA complexes would lead to slower off-rates in comparison with proteins that contact fewer nucleotides. We used SPR to determine on- and off- rates for two PPR-RNA complexes: (i) PPR10 and its *atpH* binding site, and (ii) HCF152 and its binding site in the chloroplast *psbH-petB* intercistronic region (Fig 7).

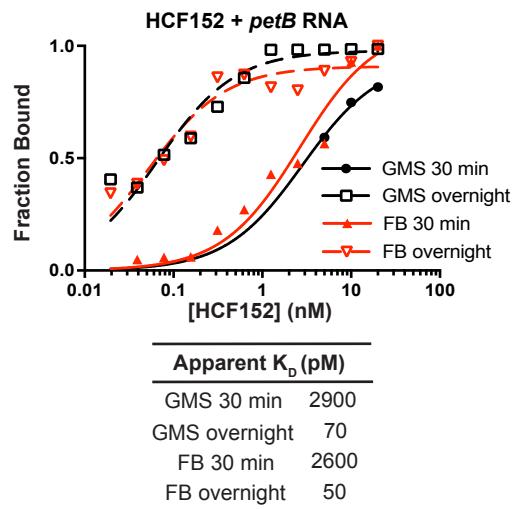
HCF152 leaves a footprint of ~19-nucleotides and blocks 5'- and 3'-exoribonucleolytic degradation *in vivo* [17, 18], similar to PPR10's effect near *atpH*.



**Fig 7. Analysis of PPR-RNA interactions by SPR.** (A) SPR analysis of PPR10-*atpH* RNA interactions. Representative sensorgrams are shown at top. The data (black) were fit with a 1:1 Langmuir binding model (red). The RNAs are diagrammed in Fig 5A. PPR10 was used at a concentration of 5 nM and 2-fold dilutions thereof. Values in the table (+/- standard error of the mean) were calculated from data from three replicate experiments. A negative control demonstrating specificity of PPR10 for *atpH* RNA is shown in S2A Fig. Residuals are shown in S2C Fig. (B) SPR analysis of interactions between HCF152 and *petB* RNA. HCF152 was used at a concentration of 40 nM and 2-fold dilutions thereof. Representative sensorgrams are shown. Values in the table (+/- standard error of the mean) were calculated from four replicate experiments. A negative control demonstrating the specificity of HCF152 for *petB* RNA is shown in S2A Fig. Residuals are shown in S2C Fig. \* Significantly different from data for the PPR10-*atpH* RNA interaction ( $P < 0.05$  according to a students t-test).

The on- and off-rates for the PPR10-*atpH* RNA predict an equilibrium constant ( $K_D$ ) of  $\sim 10$  pM (Fig 7A), similar to that inferred from equilibrium gel-mobility shift assays with the same protein preparation (S2E Fig). Both the on-rate and off-rate for the HCF152-*petB* RNA interaction were several-fold slower than the off-rate for the PPR10-*atpH* interaction (Fig 7B). Notably, the  $K_D$  calculated from the measured kinetic parameters for HCF152-*petB* was roughly 30-fold lower than that we inferred from gel mobility shift assays [17]. Given the long half-life of the complex (approximately 2 hours), we wondered whether the prior

assays had not reached equilibrium and had therefore underestimated binding affinity. To address this possibility, we compared the results of gel mobility shift and filter binding assays incubated for either 30 minutes (as in the prior study) or overnight (Fig 8). The 30-minute incubation resulted in an apparent  $K_D$  similar to that reported previously, whereas the overnight incubation resulted in an apparent  $K_D$  similar to that inferred from kinetic data (Fig 7B). These results highlight the importance of ensuring that binding reactions have reached equilibrium when measuring  $K_D$ 's for PPR-RNA interactions.

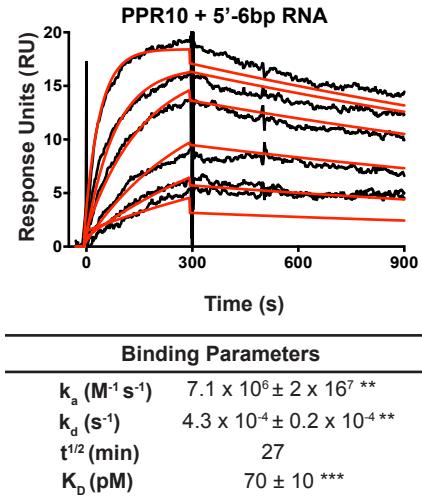


**Figure 8. HCF152-petB RNA binding curves.** HCF152-petB RNA binding curves generated from gel mobility shift (GMS) and filter binding (FB) assays, comparing results from 30 min or overnight (~13 h) binding reactions. Examples of the raw data are shown in S2D Fig.

The kinetic parameters for the PPR10-*atpH* interaction are similar to those reported for the U1A RRM domain and its RNA ligand [19, 20]. However, the on- and off- rates for the HCF152-*petB* interaction were slower than those for PPR10 and U1A with their cognate RNAs. Taken together, these data suggest that some PPR tracts form unusually long-lived complexes with their RNA ligand, but there is not a simple relationship between the length of the RNA-protein interface and the life-time or affinity of the complex.

### Effects of RNA structure on the kinetics of PPR10-RNA interactions

We next addressed the effects of RNA structures sequestering a portion of the PPR10 binding site on the kinetics of the PPR10-*atpH* RNA interaction. An RNA hairpin that includes two nucleotides of PPR10's minimal binding site at the 5'-end (5'-6bp) decreased the on-rate and increased the off-rate several fold (Fig 9). Analogous results were obtained with an RNA harboring a similar structure at the 3'-end (3'-6bp, see S2B Fig); however, this RNA was tethered to the sensor chip at the opposite end from the other RNAs we examined by SPR, and this may impact the binding kinetics. In any case, reduced on-rates are anticipated to result from the competition with adjacent RNA for PPR10 access to its binding site. The accelerated off-rates suggest that the PPR-nucleotide interactions at each end of the PPR-RNA complex can breathe, allowing the intramolecular RNA duplex to intrude on the PPR10 binding site. It is unlikely that substantive interactions can be established without the participation of the two nucleotides at either end of the binding site based on the fact that deletion of one or two nucleotides at either end of the minimal binding site prevents any apparent PPR10 interaction in gel mobility shift assays [9]. Therefore, we favor an interpretation in which the RNA bound to PPR10 can exchange partners to form a competing intramolecular RNA interaction in a manner that is analogous to branch migration at the borders of alternative nucleic acid duplexes, and that once the RNA structure is established the protein dissociates and rarely rebinds in the context of the SPR assay.

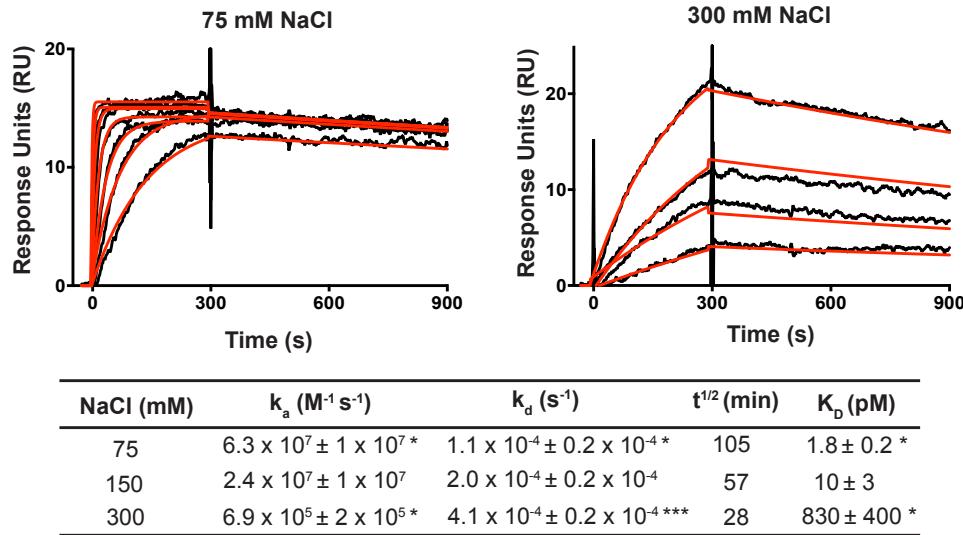


**Figure 9. Effects of RNA structure on the kinetics of PPR10-RNA interactions.** Effects of RNA structure on PPR10-RNA binding kinetics. The data are displayed as in panel (A). Values that are significantly different from those for the *atpH*-23mer are indicated (\*\* =  $P < 0.01$ , \*\*\* =  $P < 0.001$ , according to a ratio paired t-test). Residuals are shown in S2C Fig. The RNAs are diagrammed in Fig 5A. PPR10 was used at a concentration of 5 nM and 2-fold dilutions thereof. Values in the table (+/- standard error of the mean) were calculated from data from three replicate experiments.

### Contribution of electrostatic interactions to establishing and maintaining the PPR-RNA complex

The  $k_a$  for the PPR10-*atpH* RNA interaction ( $\sim 2 \times 10^7 M^{-1}s^{-1}$ , see Fig 3A) is considerably faster than that of diffusion-limited macromolecular interactions ( $\sim 10^6 M^{-1} s^{-1}$ ) [21], suggesting that electrostatic interactions drive encounters between the protein and RNA. Indeed, the consensus PPR motif used in synthetic PPR designs includes a lysine residue that forms a salt-bridge to the phosphate backbone of bound RNA, and replacement of this lysine with alternative amino acids eliminates RNA binding [5, 22]. A basic “stripe” formed by lysines and arginines at this position is apparent also in the PPR10 crystal structure [4]; however, artifactual protein dimerization in the PPR10-RNA crystal structure [4, 13] prevents inferences from that structure about electrostatic contributions to PPR10:RNA binding. To explore how electrostatic forces influence the kinetics of PPR-RNA interactions, we used SPR to monitor the effect of varying salt concentration on the on- and off-rate of PPR10-*atpH* RNA interactions (Fig 10). An increase in salt concentration increases the electrostatic shielding around charged molecules, thereby decreasing their electrostatic interactions with other molecules. Increasing the NaCl concentration from 75

mM to 300 mM caused a dramatic increase in the  $K_D$  of the PPR10-RNA interaction. This was largely due to an effect on the on-rate, which decreased approximately 100-fold. By contrast, the off-rate increased only  $\sim$ 4-fold.



**Fig 10. Effect of salt concentration on the kinetics of PPR10-atpH RNA interactions.** Representative sensorgrams are shown at top. PPR10 was used at 5 nM and two-fold dilutions thereof. The data (black) were fit with a 1:1 Langmuir binding model (red). Residuals are shown in S2C Fig. The table below shows the binding parameters inferred from the data (average of three replicate experiments +/- standard error). Values that show a significant difference from those at 150 mM NaCl (see Fig 7A) are indicated (\* =  $P < 0.05$ , \*\*\* =  $P < 0.001$ , according to a students t-test).

These results suggest that electrostatic forces make a large contribution to establishing interactions between PPR tracts and RNA, but make only a modest contribution to maintaining specific PPR-RNA interactions once established. These trends are similar to those obtained for several proteins with RRM domains [19, 20]. It is intriguing in this context that the “stripe” of positive surface potential adjacent to the RNA binding groove in consensus PPR tracts is flanked by a stripe of negative surface potential [5, 22]. Thus, the electrostatic steering that drives encounters between PPR tracts and RNA likely involves

both attractive and repulsive forces that cooperate to optimize the alignment between the specificity determining amino acids and their cognate nucleobases.

## SUMMARY

Results presented here suggest that non-specific electrostatic interactions drive PPR proteins towards RNA, and that stable binding to specific sequences is established only when most of the binding site is single-stranded. Once established, the complexes between long PPR tracts (e.g. those in HCF152 and PPR10) and their cognate RNAs are generally long-lived, and this likely underlies their effectiveness as barriers to exoribonucleases. The fact that weak secondary structures at the ends of the PPR10 binding site increase PPR10's off-rate suggests that breathing of the RNA-PPR complex provides opportunity for the competing RNA structure to form, which then inhibits reestablishment of protein-RNA contacts. It is notable in this context that several PPR proteins have been shown to occupy RNA sites *in vivo* that are predicted to contribute to RNA hairpins that are substantially more stable than those analyzed here [9, 11, 23]. Thus, it seems likely that RNA helicases and RNA chaperones facilitate PPR action *in vivo* by reducing secondary structures that would otherwise occlude their binding sites. Elucidating the nature of this interplay will be important for the design of synthetic PPR proteins and cognate binding sites, and offers an interesting area for future investigation.

## MATERIALS AND METHODS

### Protein expression and purification

PPR10 and HCF152 were expressed in *E. coli* as fusions with maltose binding protein (MBP), affinity purified on amylose resin, cleaved from the MBP moiety, and further purified by size exclusion chromatography as described previously [12, 17].

### **RNA thermal melting assays**

RNA thermal melting assays were performed as described in ref [24]. Free energies were inferred from the melting curves according to ref [25] using KaleidaGraph. RNAs were purchased from IDT. Assays performed in reverse (from high to low temperatures) and at varying RNA concentrations gave similar values (S1A Fig).

### **Gel mobility shift assays**

Gel mobility shift assays were performed as previously described [9], with minor modifications. In brief, synthetic RNA oligonucleotides (IDT) were 5'-end labeled with T4 polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]-ATP. The binding reactions contained 5 pM RNA, 40 mM Tris-HCl pH 7.5, 140 mM NaCl, 10% glycerol, 4 mM DTT, 10 U RNAsin, 0.1 mg/mL BSA, 0.5 mg/mL heparin, and protein at the indicated concentrations. Unless otherwise noted, binding reactions were incubated for 30 minutes at 25°C. Results were imaged with a Storm phosphorimager and quantified with Image Studio Lite. Curves were fit to the data using a nonlinear regression curve fit using Prism software. The sequences of the *atpH*-related RNAs are shown in Fig 1A. The *petB* RNA used in HCF152 binding assays had the following sequence: 5' GGUAGUUCGACCGUGGA-3'.

## Surface Plasmon Resonance

Biotinylated RNAs with a standard 6-carbon linker were purchased from IDT, with the following sequences: *atpH*: 5'-GAUUGUAUCCUUAACCAUUUCUUUU-3' biotin; 3'-6bp: biotin 5'-AUUGUAUCCUUAACCAUUUCUUUUUUUGAAAGA-3' 5'-6bp: 5'-ACAAUGUUUUUAUUGUAUCCUUAACCAUUUCUUU-3'-biotin; *petB*: 5'-UGGUAGUUCGACGUGGAUUUU-3'-biotin.

SPR streptavidin chips (GE Sensor Chip SA) were labeled with 5 response units (RUs) of biotinylated RNA by injecting RNA (1 pM) in HBS buffer (100 mM HEPES pH 7.5, 150 mM NaCl, 3 mM EDTA, 1 mM β-mercaptoethanol, 0.005% P20 surfactant) at a rate of 10 μL/s for 10 seconds. This yielded a maximum of approximately 25 Response Units (RUs) upon protein injection, the value suggested by Katsamba et al for analysis of high affinity interactions [26]. We used low RNA density on the chip in order to eliminate mass transport effects and ligand rebinding events during the dissociation phase [27, 28]. Prior to each experiment, the instrument was purged three times with fresh HBS buffer and equilibrated for several minutes to establish a flat baseline. For analyses of PPR10 interactions with RNAs harboring secondary structures, lane 1 was left blank for background subtraction, lane 2 was labeled with the *atpH* RNA, lane 3 was labeled with 5'-6bp RNA, and lane 4 was labeled with 3'-6bp RNA. Lanes 1-4 were analyzed in series with the same injections of PPR10 and the resulting data were statistically analyzed using a ratio paired t-test. For experiments that examined the effect of salt concentration on PPR10 binding kinetics, lane 1 was left blank for background subtraction and lane 2 was labeled with the *atpH* RNA; these data were statistically analyzed using a students t-test. For analyses of HCF152 with *petB*, lane 1 was left blank for background subtraction, lane 2 was labeled with *atpH* RNA and lane 4 was labeled with *petB* RNA; these data were statistically analyzed using a students t-test. The

range of protein concentrations in each experiment were injected in a random order with buffer injections every third injection to be used as a second background subtraction. Bound proteins were washed off the chip between protein injections using 0.02% SDS in HBS buffer. Data were analyzed using the Biacore Evaluation software.

## ACKNOWLEDGMENTS

This work was supported by a grant to A.B. from the National Science Foundation (MCB-1243641) and by National Institutes of Health Training Grant T32-GM007759 (J.J.M.). The authors would like to thank Pete Von Hippel for use of his lab instrumentation and for helpful discussions of our data. We are also grateful to Walt Baase for help with the RNA thermal melt experiments, Steve Weitzel for help with the SPR experiments, and Phil Bevilacqua for guidance on the interpretation of RNA melting curves.

## BRIDGE TO CHAPTER III

The use of designer PPR proteins for practical purposes has yet to be shown. In the following chapter I describe how I developed designer PPR proteins into a tool for basic research – purification of a specific RNA in order to identify its bound proteins. These experiments also demonstrated, for the first time, that designer PPR proteins can bind the intended RNA target with considerable specificity *in vivo*, opening up many other possible applications.

## CHAPTER III

### RIBONUCLEOPROTEIN CAPTURE VIA *IN VIVO* EXPRESSION OF A DESIGNER RNA BINDING PROTEIN

This chapter contains unpublished co-authored material. J.J. McDermott, K. Watkins, and R. Williams-Carrier performed experiments. A. Barkan and J.J. McDermott wrote the manuscript

#### **INTRODUCTION**

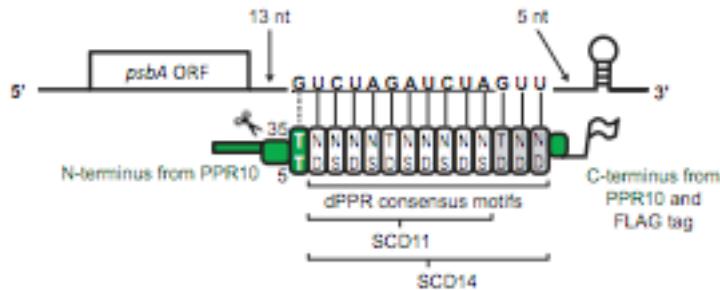
The panoply of proteins that bind an RNA determine many aspects of its function and metabolism. Although excellent approaches are available for identifying RNAs bound to a protein-of-interest, the identification of the proteins bound to particular RNAs remains problematic<sup>1</sup>. We took advantage of the programmable RNA sequence specificity of pentatricopeptide repeat (PPR) proteins to create a customized affinity tag for ribonucleoprotein purification. In a proof-of-concept experiment, we designed PPR proteins to bind the 3'-untranslated region of the chloroplast *psbA* mRNA. RIP-seq analysis of transgenic Arabidopsis expressing these proteins showed that they associate specifically with *psbA* RNA *in vivo*. Analysis of the coimmunoprecipitated proteins by mass spectrometry identified previously-unstudied proteins that we show are bound primarily to *psbA* RNA *in vivo*. Our results demonstrate that PPR proteins can be tailored to bind a specified RNA *in vivo*, and that they can be used as affinity tags to capture specific ribonucleoproteins.

#### **RESULTS AND DISCUSSION**

PPR proteins are helical repeat proteins that influence RNA stability, processing, and translation in mitochondria and chloroplasts<sup>2</sup>. PPR tracts recognize specific RNA sequences via a modular binding mode that is reminiscent of that between Pumilio/Fem3 (PUF) domains and their RNA ligands<sup>3</sup>, such that each PPR motif binds a single nucleotide with a specificity that is strongly influenced by the identities of amino acids at two positions. This amino acid code enables the reprogramming of native PPR proteins to bind novel sequences<sup>4,5</sup>, and the creation of artificial PPR proteins from consensus repeats with predictable sequence specificities<sup>6-8</sup>. The ability to customize the length and sequence specificity of PPR tracts offers promise for applications that require the targeting of a protein to native RNA sequences *in vivo*<sup>3,9</sup>. A comprehensive bind-n-seq analysis confirmed that “designer” PPR proteins built from consensus PPR motifs (dPPRs) can be highly selective for their intended RNA targets *in vitro*<sup>8</sup>, but the degree to which dPPR proteins bind selectively to RNAs *in vivo* has not been reported.

In this study, we show that two dPPR proteins bind specifically to their intended RNA target *in vivo*, and we demonstrate the utility of dPPRs for a particular application – the purification of specific endogenous ribonucleoprotein particles (RNPs) for identification of the associated proteins. We chose the chloroplast *psbA* mRNA as the target for this proof-of-concept experiment because it exhibits dynamic changes in translation in response to light, and it has been studied intensively in that context<sup>10</sup>. We designed dPPR proteins with either 11 or 14 PPR motifs to bind the 3'-untranslated region (UTR) of the *psbA* RNA in *Arabidopsis thaliana* (Arabidopsis) (**Fig. 11, Supplementary Fig. 3a**). We refer to these proteins as SCD11 and SCD14, respectively. Because PPR proteins bind single-stranded RNA, we targeted the proteins to a sequence in the *psbA* 3'UTR that is predicted to be unstructured. To avoid disrupting *psbA* expression, we chose a target sequence that is poorly

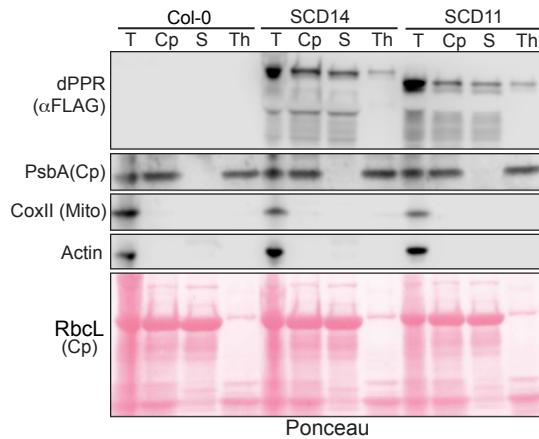
conserved and that begins sufficiently far from the stop codon that the terminating ribosome and dPPR should not occupy the same nucleotides. We designed the proteins according to the scheme described by Shen and coworkers<sup>6</sup>, such that a tract of consensus PPR motifs with the appropriate specificity-determining amino acids is embedded within N- and C-terminal segments of the native chloroplast-localized protein PPR10<sup>11</sup>. We previously reported a comprehensive analysis of the sequence specificity of SCD11 and SCD14 *in vitro*<sup>8</sup>, which confirmed them to be highly selective for their intended target sequence. For the *in vivo* assays described here, the proteins include, in addition, a C-terminal FLAG tag and the N-terminal chloroplast targeting sequence from PPR10, which is cleaved after chloroplast import (**Fig. 11 and Supplementary Fig. 3a**).



**Figure 11. dPPR design.** SCD14 and SCD11 were designed to bind the indicated 14 or 11 (underlined) nucleotide sequence, respectively, in the 3'-UTR of the *psbA* mRNA in Arabidopsis. The targeted sequence begins 13 nucleotides downstream of the stop codon, and ends five (SCD14) or eight (SCD11) nucleotides upstream of the 3'-terminal stem-loop in the *psbA* mRNA. SCD14 and SCD11 contain 13 and 10 consensus PPR motifs, respectively, flanked by sequences from PPR10 (green). The motifs that are found in SCD14 but not in SCD11 are marked in gray. The specificity-determining amino acids (positions 5 and 35 in each motif) are indicated, and each repeat is aligned with its nucleotide ligand. The PPR10-derived sequence at the N-terminus includes a chloroplast targeting sequence and PPR10's first PPR motif, which has a non-canonical specificity code (dotted line). The targeting sequence is cleaved after import into the chloroplast (scissors). Both proteins contain a C-terminal 3x FLAG tag.

Immunoblot analysis of leaf and chloroplast fractions from transgenic Arabidopsis plants expressing SCD11 and SCD14 confirmed that they localize to chloroplasts (**Fig. 12**).

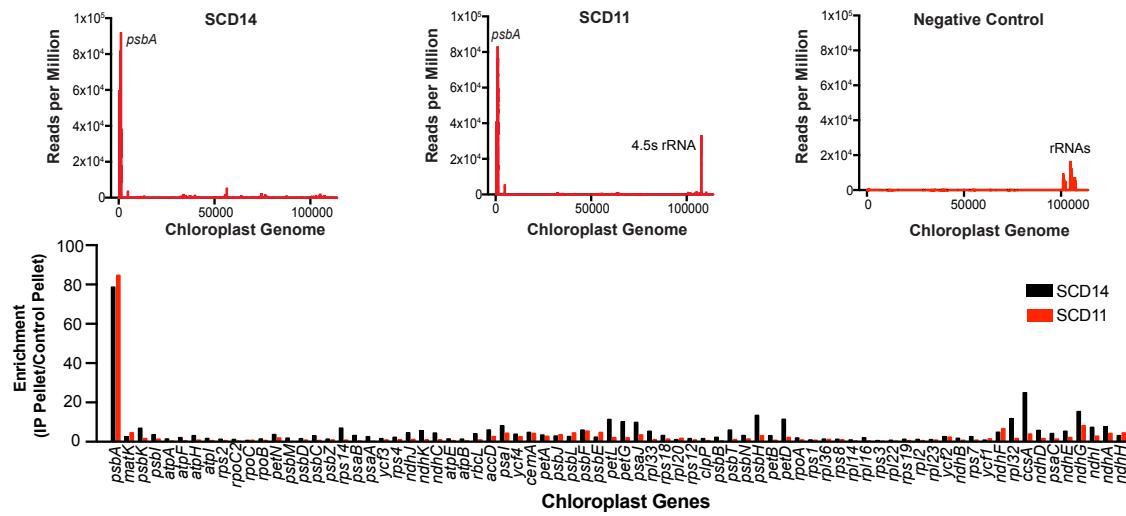
Both proteins were found predominantly in the soluble fraction, which is as expected given that they lack transmembrane segments or thylakoid targeting signals. Laddering beneath the band corresponding to each full-length protein suggests that these artificial proteins are prone to proteolysis *in vivo*. The transgenic plants were phenotypically normal (**Supplementary Fig. 3**) and had normal levels of PsbA protein (**Fig. 12**) indicating that the dPPRs did not disrupt *pshA* expression or have off-target effects that compromised plant growth.



**Figure 12. dPPR chloroplast localization.** Immunoblots demonstrating chloroplast-localization of SCD11 and SCD14. Chloroplasts (Cp) were isolated from total leaf extract (T) of wild-type (Col-0) and transgenic Arabidopsis plants, and fractionated to generate thylakoid membrane (Th) and soluble (S) fractions. Aliquots representing an equivalent amount of starting material were probed to detect markers for cytosol (actin), mitochondria (CoxII), and thylakoid membranes (PsbA). The dPPR proteins were detected with anti-FLAG antibody. The Ponceau S-stained filter is shown below to demonstrate the partitioning of the chloroplast stromal protein RbcL among the fractions.

To identify the RNAs to which SCD11 and SCD14 are bound *in vivo*, we isolated chloroplasts from the transgenic plants and immunoprecipitated each protein from stromal extract by using anti-FLAG antibody (**Supplementary Fig. 4a**). Slot-blot hybridization analysis of RNA from these immunoprecipitates showed that *pshA* RNA was highly enriched from extracts of the transgenic lines in comparison to the wild-type progenitor (Col-0) (**Supplementary Fig. 4b**). Furthermore, RNA from the 3'UTR was more highly enriched than that from the 5'-UTR, consistent with the fact that the two proteins were targeted to a

sequence in the *psbA* 3'-UTR. To gain a comprehensive view of the RNAs bound to each protein, we sequenced the coimmunoprecipitating RNA (RIP-seq) (**Fig. 13**). Comparison of these RNAs to those in a pellet from an immunoprecipitation with an antibody that does not recognize proteins in Arabidopsis showed that the *psbA* RNA strongly and specifically coprecipitated with SCD11 and SCD14 (**Fig. 13**, right and bottom). To our knowledge, this is the first reported evidence that dPPR proteins can bind specifically to intended RNA targets *in vivo*.

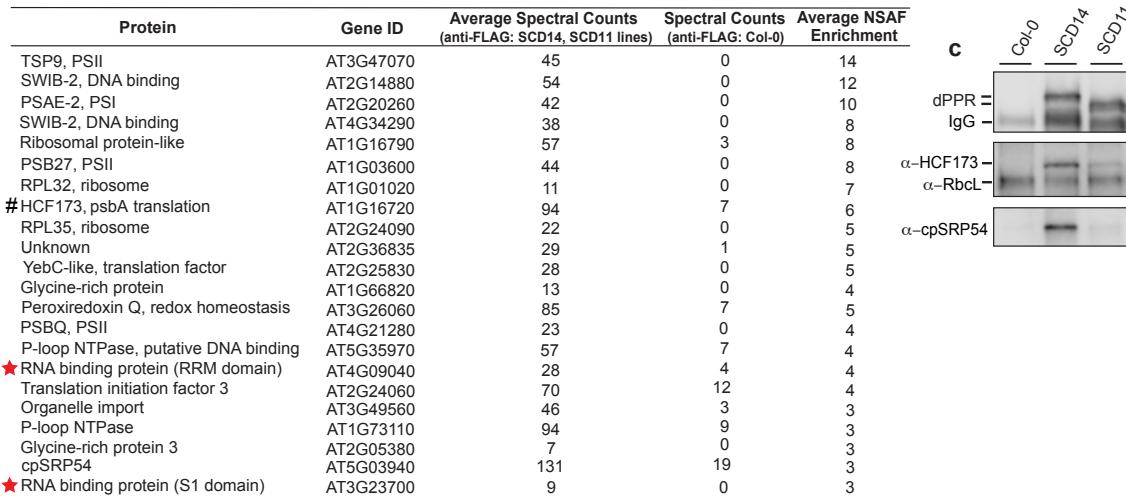


**Figure 13. Analysis of RNAs that coimmunoprecipitate with SCD11 and SCD14.** RIP-seq analysis of RNAs that coimmunoprecipitate with SCD11 and SCD14. Anti-FLAG antibodies were used for immunoprecipitation from extracts of chloroplasts from transgenic plants expressing SCD11 or SCD14. The negative control used chloroplast extract from the Col-0 progenitor and an antibody that does not detect proteins in that extract. Results at top are plotted as the average sequence coverage in consecutive 10-nt windows along the chloroplast genome (Accession NC\_000932.1), per million reads mapped to the chloroplast genome. An alternative view of the data showing the ratio of normalized reads/gene in the experimental and control immunoprecipitations is shown below.

We next used mass spectrometry to identify proteins that coimmunoprecipitate with SCD11 and SCD14 from chloroplast stroma. Approximately 400 different proteins were identified in at least one of the immunoprecipitates. The enrichment of each protein was

calculated with respect to its representation in an anti-FLAG immunoprecipitate from the non-transgenic host line (Col-0). Proteins with an average enrichment of at least 3-fold are listed in **Fig. 14a**. This protein set included several proteins that are known to associate with *pshA* mRNA: HCF173, which activates *pshA* translation<sup>12</sup>, cpSRP54, which binds cotranslationally to PsbA<sup>13</sup>, and various ribosomal proteins and translation factors. Immunoblot analysis of anti-FLAG coimmunoprecipitates confirmed that HCF173 and cpSRP54 coimmunoprecipitate with the dPPR proteins from extracts of the transgenic plants (**Fig. 14b**). The differing efficiency with which these proteins were coprecipitated from the two lines may be due to differing degrees of RNA degradation in the two preparations, as RNA cleavage upstream from the 3'UTR will separate the bound dPPRs from proteins bound elsewhere on the RNA.

Two uncharacterized proteins with predicted RNA binding domains were enriched in the dPPR coimmunoprecipitates: AT3G23700, which has two S1 RNA binding domains, and AT4G09040, which has two RRM RNA binding domains (**Fig. 14a**, stars; sequences shown in **Supplementary Fig. 5a**). We generated antibodies to the maize orthologs of these proteins and confirmed their specificity by immunoblot analysis of corresponding loss-of-function mutants (**Supplementary Fig. 5b**). The antibodies were then used for RIP-seq analysis with maize chloroplast extract; we could not use Arabidopsis extract due to lack of antibody cross-reactivity. The *pshA* RNA was highly enriched in each coimmunoprecipitate, in comparison to its representation in the input RNA (**Fig. 13**). Several other RNA ligands were also identified in each case (**Fig. 14a**). These results validate the utility of the dPPR-affinity tag approach to identify proteins that associate with a specific RNA-of-interest *in vivo*. It will be interesting to explore the roles of these proteins in the expression of their RNA targets in the future.



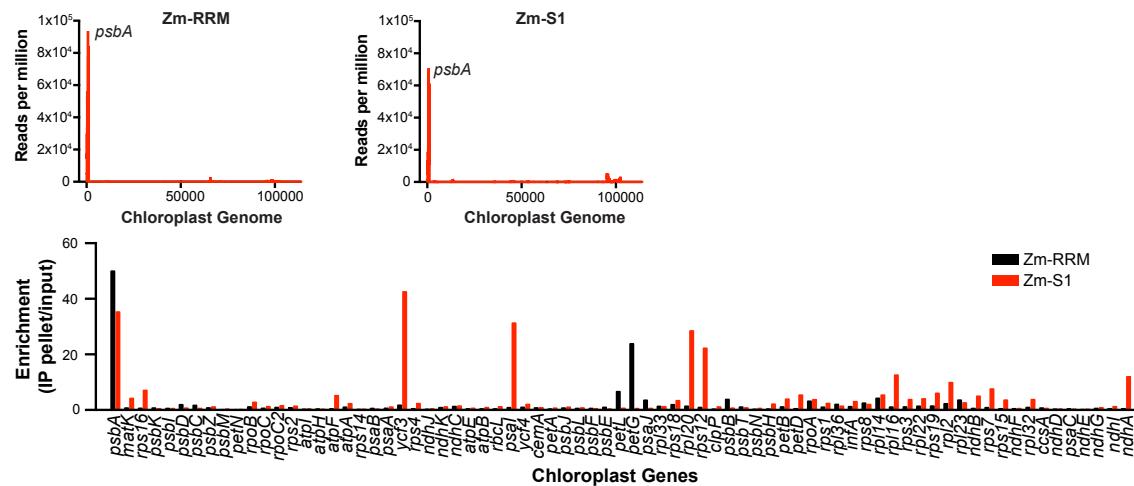
**Figure 14. Analysis of proteins that coimmunoprecipitate with SCD11 and SCD14.**

**(a)** Highly-enriched proteins in SCD11 and SCD14 coimmunoprecipitates, as detected by LC-MS/MS. The table lists proteins whose average enrichment from lines expressing SCD11 or SCD14 in comparison to the Col-0 progenitor was three or greater. Stars mark two uncharacterized RNA binding proteins. HCF173 is known to bind *psbA* RNA<sup>12</sup>.

**(b)** Immunoblot validation of several proteins identified by MS/MS analysis. Chloroplast stroma from plants expressing SCD11 or SCD14, or from the Col-0 progenitor was used for immunoprecipitation with anti-FLAG antibody. Replicate immunoblots were probed to detect SCD11 or SCD14 (anti-FLAG), HCF173, or cpSRP54. The HCF173 blot was initially probed to detect RbcL, an abundant protein that typically contaminates immunoprecipitates, which serves as an internal standard.

Several methods for the RNA-centric purification of RNPs have been reported previously. Some of these rely on the insertion of an RNA affinity tag into the target RNA<sup>14, 15, 16</sup>. However, insertions can alter RNA functionalities, modification of endogenous genes is technically challenging in some experimental systems (such as organelles), and expression of ectopic modified genes can disrupt the balance of *trans*-factors to their *cis*-targets. These limitations are addressed by assays that purify untagged RNPs by coupling *in vivo* crosslinking with post-lysis antisense oligonucleotide purification<sup>17, 18, 19</sup>. However, UV crosslinking is inefficient and is practical only with cultured cells or lysates. Formaldehyde crosslinking provides an alternative, but is prone to capturing both transient and stable interactions. Very recently, a Type VI-related CRISPR-Cas system was engineered to bind and modify the

splicing of endogenous RNAs<sup>20</sup>. Whether this system can achieve the degree of RNA occupancy needed for use in affinity purification approaches remains to be determined.



**Figure 15. RIP-Seq analysis demonstrating identification of previously-unknown *psbA*-specific RNA binding proteins in the dPPR-affinity purifications.** Maize chloroplast stroma was used for immunoprecipitations with antibodies to the maize orthologs of the RRM and S1-domain proteins that were enriched in dPPR-coimmunoprecipitations (AT4G09040 and AT3G23700, respectively). Results at top are plotted as the average sequence coverage in consecutive 10-nt windows along the chloroplast genome, per million reads mapped to the chloroplast genome (NCBI NC\_001666). An alternative view of the data showing the ratio of normalized reads/gene in pellets from the experimental immunoprecipitations versus the input RNA is shown below.

Designer PPR affinity tags add to this toolkit by binding unmodified RNPs within intact tissues, and allowing their purification under non-denaturing conditions. Given the successes in using designer PUF proteins to modify the expression of specific mRNAs<sup>21,22</sup> and to track untagged RNAs *in vivo*<sup>23</sup>, they may also be useful as affinity tags for the purification of specific RNPs. However, the greater flexibility in repeat tract length with the PPR scaffold<sup>3</sup> may facilitate customization of RNA binding affinity, kinetics, and specificity for this application. A caveat relevant to the general applicability of our approach is that PPR proteins in nature function almost exclusively inside mitochondria and chloroplasts, and it is

as yet unclear how they will perform in the nuclear-cytosolic compartment. Thus, an important next step is to test this approach on a cytosolic target RNA.

## METHODS

### Development of transgenic lines

Genes for SCD14 and SCD11 were codon-optimized for *Arabidopsis* and assembled by PCR from several overlapping synthetic DNA fragments (IDT). The protein sequences are provided in **Supplemental Fig. 3a**. They were designed with the PPR nucleotide specificity codes described previously<sup>4,8</sup> and summarized in **Fig. 11**. The PPR-encoding genes were inserted into a modified form of pCambia1300<sup>24</sup> that encodes a 3X-FLAG tag at the C-terminus of the inserted ORF (a gift from Jie Shen and Zhizhong Gong, China Agricultural University). The plasmids were used to transform *Arabidopsis thaliana* (ecotype Col-0) (*Arabidopsis*) using the floral dip method<sup>25</sup>. Lines were screened by immunoblotting for dPPR expression, and those with the highest expression were used for further experiments. An additional transgenic line was developed using the MCD14 protein design we reported previously<sup>8</sup>; however, MCD14 transgenic lines failed to express the protein.

### Plant growth

*Arabidopsis* seeds were sterilized by incubation for 10 min in a solution containing 1% bleach and 0.1% SDS, followed by a 70% ethanol wash. The seeds were then washed three times with sterile water. Seeds were plated and grown in tissue culture dishes containing MS agar medium: 4.33 g/L Murashige and Skoog basal salt medium (Sigma), 2% sucrose, 0.3% Phytagel, pH 5.7. Transgenic plants were selected by the addition of 50 µg/mL hygromycin to the growth medium. Plants used for chloroplast isolation and immunoprecipitation assays

were grown in a growth chamber in diurnal cycles (10 h light at 120 µE light intensity, 14 h dark, 22°C) for 14 days.

### **Chloroplast isolation and fractionation**

Maize chloroplast stroma for use in RIP-seq assays was prepared as described previously<sup>26</sup>. Arabidopsis chloroplast stroma was prepared from chloroplasts isolated from the aerial portion of 2-week old seedlings (40 g tissue) as described<sup>27</sup>, with the following modifications: seedlings were not placed in ice water before homogenization, sorbitol concentration in the homogenization buffer was reduced to 0.33 M, and plants were homogenized in a blender using three 5-second bursts. Purified chloroplasts were resuspended and lysed in Hypotonic Lysis Buffer (30 mM HEPES-KOH pH 8.0, 10 mM MgOAc-4H<sub>2</sub>O, 60 mM KOAc, 2 mM DTT, 2 µg/mL aprotinin, 2 µg/mL leupeptin, 1 µg/mL pepstatin A, 0.8 mM PMSF), using a minimal buffer volume so as to maximize protein concentration in the extract. Lysed chloroplasts were centrifuged for 40 min at 18,000 x g at 4°C in a tabletop microcentrifuge to pellet membranes and particulates. The supernatant was removed, and the pellet was resuspended in Hypotonic Lysis Buffer and centrifuged again under the same conditions. Supernatants were combined, aliquoted, and frozen at -80 °C. The thylakoid membranes (pellet fraction) were aliquoted and frozen at -80 °C.

### **Antibodies, SDS-Page, and immunoblot analysis**

SDS-PAGE and immunoblot analyses were performed as described previously<sup>28</sup>. A mouse monoclonal anti-FLAG M2 antibody was purchased from Sigma. Polyclonal

antibodies were raised in rabbits to recombinant fragments of the maize orthologs of HCF173, AT4G09040 and AT3G23700 (the uncharacterized RRM and S1-domain proteins, respectively, that coimmunoprecipitated with the dPPRs); these correspond to maize genes GRMZM2G397247, GRMZM2G023591, and GRMZM2G016084, respectively (see <http://cas-pogs.uoregon.edu/#/> for evidence of orthology). The amino acids used for the S1 and RRM protein antigens and evidence for the specificity of the resulting antisera are shown in **Supplementary Fig. 3**. Amino acids 364-633 were used for Zm-HCF173. The maize CRP1 antibody was described previously<sup>29</sup>. Antibodies were affinity purified against their antigen prior to use.

### Coimmunoprecipitation experiments

Immunoprecipitation for analysis of proteins by mass spectrometry was performed as described previously<sup>30</sup> with minor modifications. In brief, experiments used Arabidopsis stromal extract, anti-FLAG antibodies were crosslinked to magnetic Protein A/G beads (Pierce), the beads were pre-washed in CoIP Buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% NP-40, 5 µg/mL aprotinin), and the antibody cross-linked beads were titrated to determine the amount required to deplete the dPPR from the stromal extract. Stromal extract (400 µL at 6 mg protein/ml) was supplemented with RNAsin (Promega) to a concentration of 1 unit/µL and precleared by centrifugation for 10 min at 18,000 x g at 4 °C. The supernatant was removed to a new tube, antibody-bound beads were added and the mixture was incubated at 4 °C for one hour while rotating. Beads were captured with a magnet (Invitrogen) and the supernatant was removed. The beads (pellets) were washed three times with CoIP Buffer and then twice with 50 mM ammonium bicarbonate (pH 7.5). Proteins were digested on the beads with trypsin (Promega mass

spectrometry grade at 25 ng/ $\mu$ L in 50 mM ammonium bicarbonate pH 7.5) overnight at 25 °C while shaking. Beads were captured and the supernatant was transferred to a new tube. This step was repeated five times to ensure the removal of all beads. LC-MS/MS was performed by the UC Davis Proteomics Core Facility, where the data were analyzed using Scaffold2 (Proteome Software Inc). Protein enrichment was calculated by dividing the average Normalized Spectral Abundance Factor (NSAF) values<sup>31</sup> from the two dPPR lines by NSAF values from the control immunoprecipitation using extract of Col-0 plants. To avoid division by zero, a correction term of 0.001 was added to each NSAF value in the control; therefore, the actual enrichment of proteins that were not detected in the control is under-estimated.

Immunoprecipitations for RIP-seq analysis were performed similarly, except that antibodies were not crosslinked to the beads, the Arabidopsis experiments used 200  $\mu$ L of extract, and the maize experiments used 70  $\mu$ L of stromal extract at ~10 mg protein/ml and did not include RNAsin. Antibody to maize CRP1 was used as a negative control for the dPPR RIP-seq assays; this antibody does not recognize proteins in Arabidopsis chloroplasts.

### **Analysis of immunoprecipitation RNA by RIP-seq and slot blot analysis**

An equal volume of TESS Buffer (10 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 0.2% SDS) supplemented with Proteinase K (0.2  $\mu$ g/ $\mu$ L) was added to the supernatant and pellet fractions and incubated for 30 min at 37 °C. RNA was then purified by phenol-chloroform extraction and ethanol precipitation, resuspended in H<sub>2</sub>O, and quantified by Qubit. The RNA was used directly for slot-blot hybridizations as previously described<sup>32</sup>, or processed for sequencing using the BIOO Scientific NEXTflex Small RNA-Seq Kit v3 (Cat# NOVA-5132-06). For Arabidopsis, 50 ng of pellet RNA was used as the input for

sequencing libraries. The maize experiments used 20 ng of pellet or input RNA for library preparation, and included an RNA fragmentation step: RNA was fragmented by incubation at 95 °C for 4 min in 40 mM Tris-Acetate pH 8, 100 mM KOc, 30 mM MgOAc2. The reaction was stopped by the addition of EDTA to a final concentration of 50 mM, and the RNA was ethanol precipitated in the presence of 1.5 µg GlycoBlue (Thermo Fisher). Libraries were gel purified to enrich for inserts between 15 and 100 nucleotides. Libraries were sequenced at the University of Oregon Genomics and Cell Characterization Core Facility, with read lengths of 75 or 100 nucleotides. Sequencing data were processed as in<sup>33</sup> except that all read lengths were included and reads were aligned only to the chloroplast genome.

## ACKNOWLEDGMENTS

We are grateful to Jie Shen (Chinese Academy of Sciences) and Zhizhong Gong (China Agricultural University) for their gift of the pCAMBIA1300 vector modified to encode a FLAG tag and for helpful advice. We are also grateful to Masato Nakai (Osaka University) for the generous gift of cpSRP54 antibody, Carolyn Brewster for immunoblot quantification of dPPRs in transformed lines, the UniformMu project (University of Florida) for maize insertion lines, and the UC-Davis Proteomics Core for LC-MS/MS proteomic analyses. This work was supported by National Science Foundation grant MCB-1616016 (to A.B.) and National Institutes of Health Training Grant T32-GM007759 (to J.J.M.).

## BRIDGE TO CHAPTER IV

In this chapter, I described the target specificity of designer PPR proteins *in vivo* and their utility as an *in vivo* biotechnological tool. The results showed that dPPR proteins bind to their

targets *in vivo* and can be used as an “affinity-tag” to purify endogenous transcripts and identify bound RNPs. In the following chapter I summarize the findings of this dissertation and describe the implications and possible future directions for the findings, with a specific focus on the possible applications of designer PPR proteins.

## CHAPTER IV

### SUMMARY AND FUTURE CONSIDERATIONS

#### SUMMARY

PPR proteins function in organellar RNA metabolism and are critical for the development of all eukaryotes. In the first chapter of this dissertation I introduce PPR proteins and the cellular compartments where they function – the DNA containing organelles – with a focus on the PPR proteins found in plants. I discuss the *in vivo* functions and the binding mechanisms of PPR proteins, and I discuss knowledge gaps in the literature of PPR proteins that I answer in the proceeding chapters.

In the second chapter of this dissertation I focused on developing a deeper understanding of the mechanisms of PPR:RNA interactions, specifically what parameters influence binding affinity and kinetics. PPR proteins are crucial for the proper expression of genes in organelles and yet we don't fully understand how the dynamic nature of RNA structure impacts PPR:RNA binding and function. I demonstrate how the stability and position of RNA secondary structure inhibits PPR10:*atpH* interactions and show that even weak RNA structures inhibit PPR10:*atpH* binding. Additionally, I show how RNA secondary structure affects the binding kinetics of PPR10:*atpH* interactions. RNA secondary structure slowed down the PPR10:*atpH* on-rate – which was an expected outcome based on the fact that the secondary structure would sterically inhibit PPR10 from interacting with the Watson-Crick face of its binding site nucleotides. A more perplexing result was the fact that RNA secondary structure increased the off-rate of PPR10:*atpH* interactions. We hypothesized that the PPR10:*atpH* complex can “breathe”, similar to DNA duplex breathing, and during this breathing, the RNA structure can reform which causes the protein

to prematurely dissociate thus increasing the off-rate. I also demonstrated how electrostatic forces influence the binding kinetics, and how there is not a simple relationship between PPR protein length and the binding kinetics when comparing two native PPR proteins.

The customizable binding specificity of PPR proteins makes them attractive for a variety of applications. In Chapter III of this dissertation I demonstrate the feasibility of using synthetic PPR proteins as tools for basic science research *in vivo*. I showed that synthetic PPR proteins can be engineered to bind to a specific chloroplast transcript *in vivo* and can be used to co-purify the transcript and bound ribonucleoproteins (RNPs). These designer PPR proteins bound quite specifically to their intended RNA target (chloroplast *psbA* RNA), but this finding is likely influenced by the fact that chloroplasts have a limited sequence space (~110,000 bp). Nonetheless, this was the first demonstration, to my knowledge, that a designer PPR protein behaves predictably *in vivo*. In addition, this provides a new tool for the purification of endogenous RNPs for proteomic analysis. Our success using it for this purpose was likely enhanced by the fact that *psbA* mRNA is the most abundant mRNA in a plant cell. I highlight other potential designer uses of synthetic PPR proteins in the subsequent sections.

### **Implications of the effects of RNA structure on the binding affinity and kinetics of PPR proteins**

Results presented in Chapter II revealed that RNA structure stability and position impact PPR:RNA interactions, and that even weak RNA structures of  $\Delta G^\circ \sim 0$  kcal/mol inhibit PPR:RNA interactions. It is likely that many PPR protein binding sites could form RNA structures *in vivo* thus it is unclear how PPR proteins are able to bind to these sequestered binding sites. It is likely they are assisted by RNA helicases or chaperones, so a paramount

next step would be to examine the impact of these proteins on PPR binding *in vivo*. The influence of RNA secondary structure on PPR binding will impact the design of future synthetic PPR proteins and will be an important factor to consider. I also showed how electrostatic forces influence the binding kinetics, specifically the on-rate of PPR10:*atpH* interactions. With two native PPR proteins, I demonstrated that there is not a simple relationship between the length of the PPR protein and the binding kinetics. Because native proteins like the ones I examined are known to have idiosyncratic features, it would be interesting to do analogous experiments with artificial proteins built from consensus PPR motifs in the future. Investigations in this area would also be fruitful in the design of synthetic PPR proteins. Understanding the structure-function relationship of PPR proteins and their binding kinetics and affinity could lead to the development of designer PPR proteins with customized binding kinetics and affinities for specialized purposes that I highlight in the section below.

### **Implications and the future applications of designer PPR proteins**

Chapter III of this dissertation developed the use of synthetic PPR proteins for designer purposes. I successfully showed that designer PPR proteins bind specifically to their target sequences *in vivo* and can be used as a tool to purify an endogenous transcript and identify proteins bound to that transcript.

The use of engineered PPR proteins as a biotechnology has been almost 25 years in the making. This journey began as an observation concerning the basic science of RNA metabolism in chloroplasts in 1994 [Barkan *et al.* 1994] and has come full circle to the development of engineered PPR proteins for biotechnology applications in 2018.

I have demonstrated one potential future use of PPR proteins as a biotechnology. There are many other potential uses of PPR proteins that have yet to be realized [Yagi et al. 2013]. Many of these future applications could stem from the functions of native PPR proteins. Native PPR proteins function by i.) sequestering long segments of RNA which prevents the RNA from interacting with other proteins or RNA or exposes a cis-element masked in an RNA hairpin that is then free to interact with other proteins or RNA, or ii.) the action of an accessory domain attached to the PPR protein [reviewed in Barkan and Small 2014]. By the first mechanism, designer PPR proteins could be used as sequence specific binders to block cis elements (such as RNA splice sites, or protein binding sites) from interacting with RNA or proteins and affect downstream RNA actions; this could have applications in basic research and biotechnology – similar to how small RNA oligonucleotides are used as tools for synthetic RNA biology [Reviewed in Isaacs et al. 2006 and Sazani and Kole 2003].

Through the second mechanism, an accessory domain could be attached to the designer PPR protein to provide a secondary action after the PPR motifs have guided the protein to the site of interest. These accessory domains could be i.) nucleases or RNA editing domains which would disrupt or change the fate of the RNA transcript, ii.) fluorescent domains, such as GFP, which could be used to track the localization of a transcript of interest [Bertrand 1998], iii.) a localization factor which could localize the RNA to specific cellular compartments, or iv.) as I showed in Chapter III of this dissertation, the designer PPR protein could be linked with an epitope like a FLAG tag which could be used to purify the PPR protein, target transcripts, and RNPs for the identification of RNA binding proteins associating with a particular transcript. Other methods are capable of purifying endogenous transcripts and RNPs, but many require a change in the RNA sequence [Panchapakesan et al. 2017, Butter et al. 2009, Ramanathan et al 2018]. Genetic engineering can be difficult in

organelles and can also create unexpected complications for the RNA. Other methods use oligonucleotides to recognize and purify the transcript, but these procedures have yet to be adopted into organelles [Rogell et al. 2017, McHugh et al. 2015, Chu et al. 2015]. PPR proteins offer the advantage of using a protein-based system which can easily be transformed or transfected into the nuclear genomes of model organisms and customized to bind to an existing RNA sequence. They are also easily targeted to organelles as I demonstrated in Chapter III.

Other customizable RNA binding proteins, such as CRISPR-CAS or engineered PUF proteins, could be used in these same applications [Nelles et al. 2016, Konermann et al 2018, Campbell et al. 2014], however the diversity of PPR motifs (P, L, and S motifs) provides a more malleable template for which to customize the binding affinity and kinetics to achieve the variety of designer functions mentioned above – a parameter that these other systems may be unable to achieve. PPR proteins also offer an advantage in organelles, where PUF and CRISPR-CAS proteins have yet to be successfully targeted. Additionally, CRISPR-CAS systems require a guide RNA and currently it is technically impossible to introduce guide RNAs into mitochondria and chloroplasts.

The binding kinetics and affinity will likely impact the efficiency of the designed function of the synthetic PPR. Synthetic PPR proteins with an additional nuclease or editing domain will function best with fast on-rates and off-rates. Ideally, these proteins would quickly find their targets, perform their functions, and then dissociate to interact with a new transcript. On the other hand, synthetic proteins designed to activate translation, block a cis-element, or be used as an “affinity tag” would function optimally with fast on-rates and slow off-rates. Similarly, these proteins would quickly find their targets, but would seldom dissociate – an ideal function for their purpose.

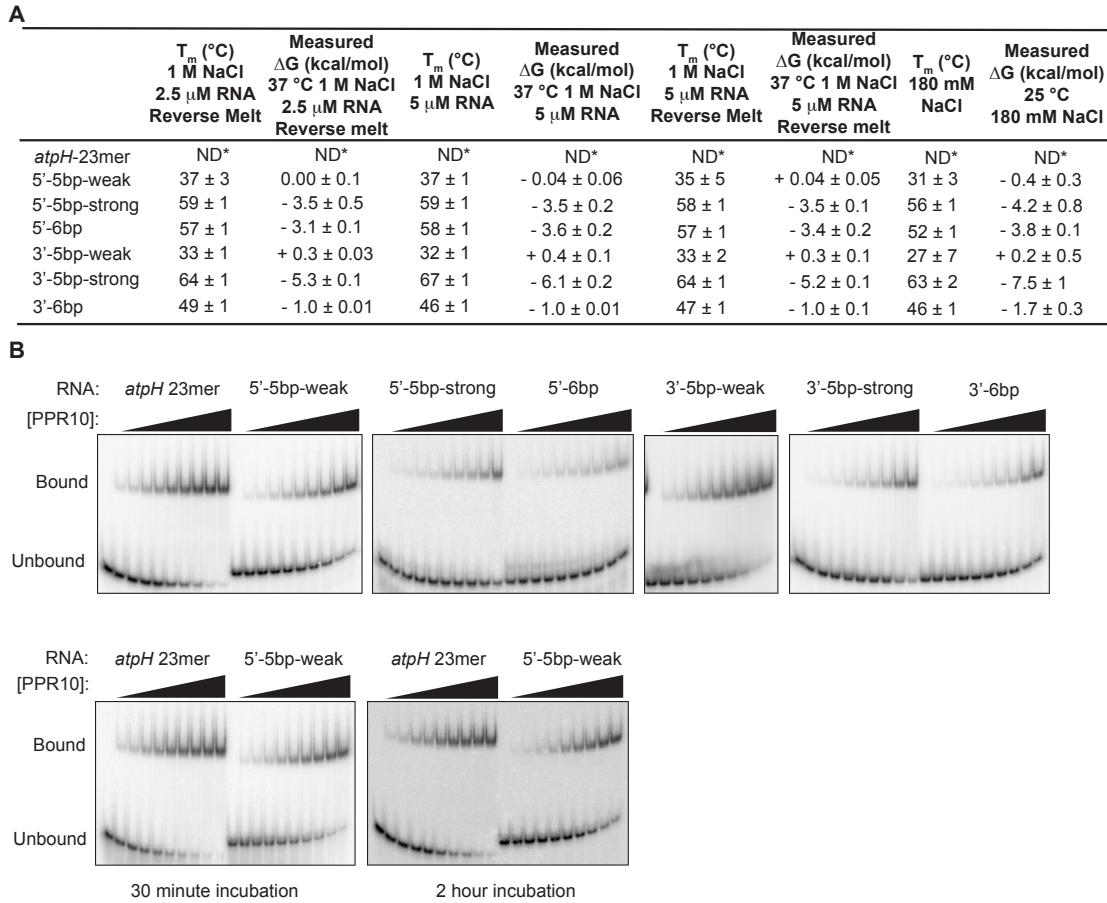
One of the key challenges for the future development of designer PPR proteins is to determine if they will function outside the context of an organelle. Very few PPR proteins are predicted to localize to the cytosol [Hammani *et al* 2016], and so many potential applications will hinge on the ability of synthetic PPR proteins to function in the cytosol. It will be necessary to perform future experiments and optimizations in this area.

## CLOSING REMARKS

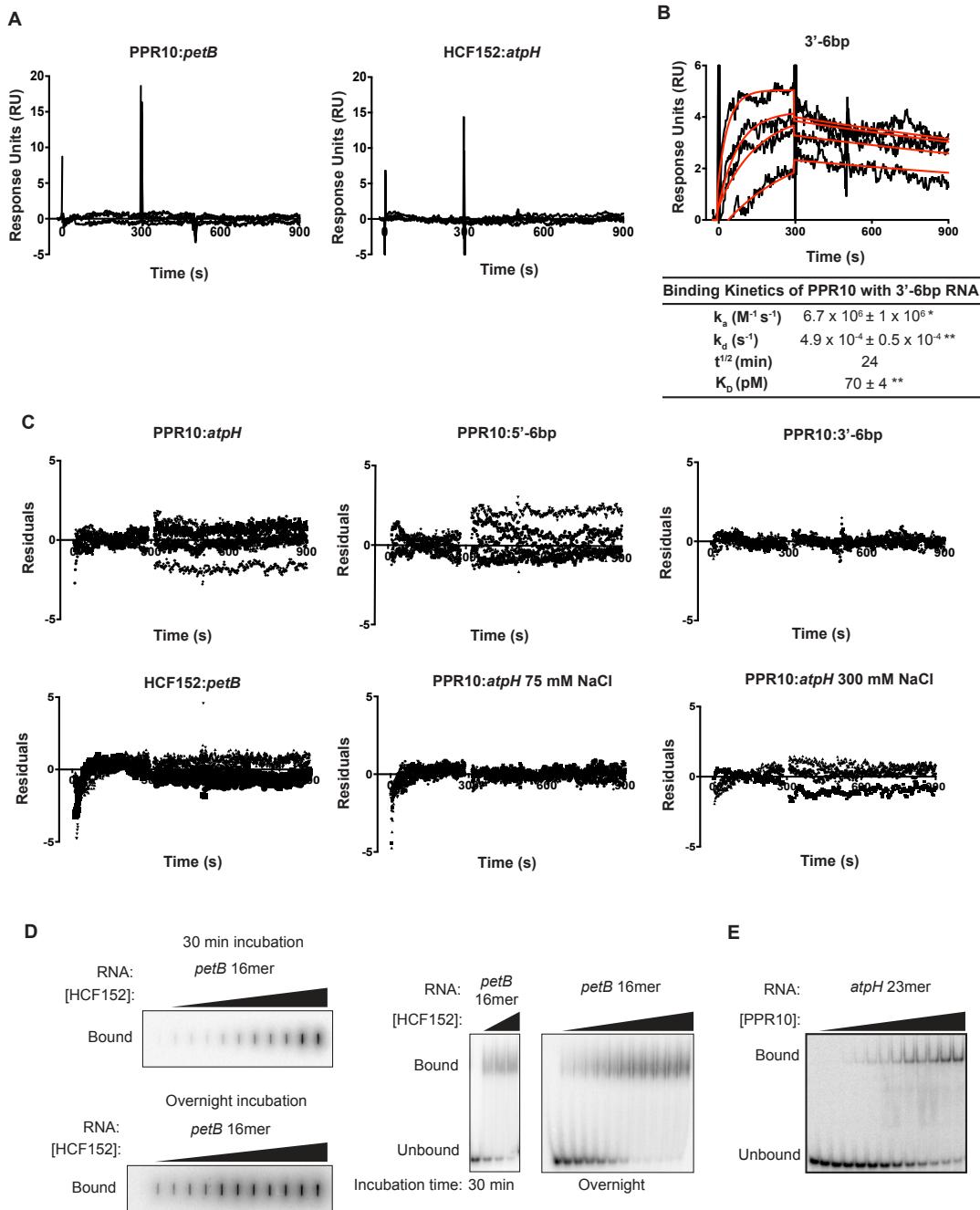
In this dissertation I have made significant contributions towards understanding the mechanisms of PPR:RNA interactions, specifically how they interact with RNA secondary structure, a parameter that will likely impact the *in vivo* function of PPR proteins. Additionally, the modular nature of PPR proteins make them excellent targets for the design of bespoke proteins. This dissertation investigated one potential avenue for designer PPR proteins and shows their viability as a method to purify endogenous transcripts and associated RNPs. Given these results and the malleable, modular, and programmable nature of PPR proteins there is a bright future for designer PPR proteins in translational control, the purification of transcripts and RNPs, RNA editing, and RNA tracking in organelles.

## APPENDIX A

### SUPPLEMENTAL MATERIAL FOR CHAPTER II



**S1 Fig.** (A) Predicted and measured stabilities of the *atpH*-related RNAs diagrammed in Fig 1A. (B) Representative gel mobility shift assays underlying the curves presented in Fig 2. See Fig 2 for details.



**S2 Fig** (A) Specificity controls for PPR10 and HCF152 SPR assays. Sensorgrams are shown for the analysis of PPR10 interaction with HCF152's *petB* RNA ligand (left) and HCF152's interaction with PPR10's *atpH* RNA ligand (right). (B) SPR analysis of PPR10 interaction with the 3'-6bp RNA ligand (see Fig 1A). The experiment was performed as in Fig 3A except that the RNA was tethered to the SPR chip via biotin at its 5'-end. PPR10 was used at a concentration of 5 nM and 2-fold dilutions thereof. (C) Residuals for SPR assays. (D) Examples of gel mobility shift and filter binding data supporting the curves shown in Fig 3C. (E) Gel mobility shift assay of PPR10-*atpH* 23 mer interactions, using the same PPR10 protein preparation as used in the SPR assays. PPR10 was used at a concentration of 2.5 nM and 2-fold dilutions thereof.

## APPENDIX B

### SUPPLEMENTAL MATERIAL FOR CHAPTER III

**a**

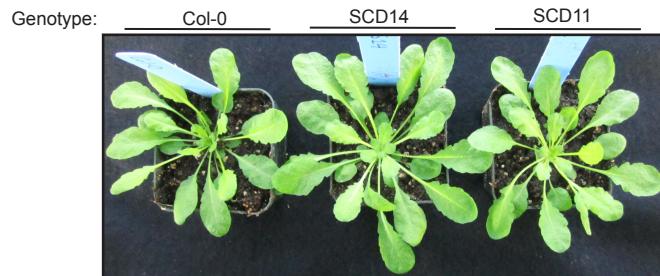
**Amino acid sequence of SCD14:**

MEATGRGLFPNKPTLPAGPRKRGPLPAAPPPPSPSSLPLDSLLLHTAPAPAPAPRRSHQTPTPPHSFLSP  
 DAQVLVLAISHPLPTLAFASRRDELLRADITSLKALELSGHWEWALLLRWAGKEADASALEMVRALR  
 GREGQHDAVCALLDETPLPPGSRLDVRAYTVLHALSRAGRYERALELFEALRRQGVAPTVVTYNTLIDGLCK  
 AGKLDEALKLFEEMVEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPSVVTYNTLIDGLCKAGKLKD  
 EALKLFEEMVEKGIKPDVVTYSTLIDGLCKAGKLDEALKLFEEMVEKGIKPNVVTTIDGLCKAGKLDEALKL  
 FEEMVEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPSVVTYNTLIDGLCKAGKLDEALKLFEEM  
 VEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPNVVTTIDGLCKAGKLDEALKLFEEMVEKGI  
 KPDVVTYSTLIDGLCKAGKLDEALKLFEEMVEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDVV  
 TYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPDVV  
 VESYCRAKRFEEARGFLSEVSETDLDFDKKALEAYIEDAQFGRGDYKDHDGDYKDHDIDYKDDDDDK

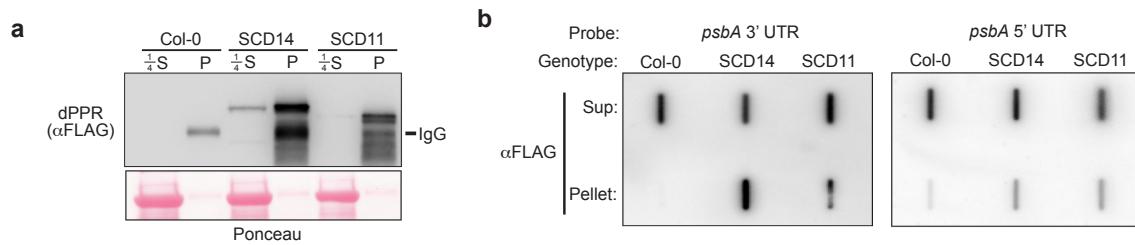
**Amino acid sequence of SCD11:**

MEATGRGLFPNKPTLPAGPRKRGPLPAAPPPPSPSSLPLDSLLLHTAPAPAPAPRRSHQTPTPPHSFLSP  
 DAQVLVLAISHPLPTLAFASRRDELLRADITSLKALELSGHWEWALLLRWAGKEADASALEMVRALR  
 GREGQHDAVCALLDETPLPPGSRLDVRAYTVLHALSRAGRYERALELFEALRRQGVAPTVVTYNTLIDGLCK  
 AGKLDEALKLFEEMVEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPSVVTYNTLIDGLCKAGKLKD  
 EALKLFEEMVEKGIKPDVVTYSTLIDGLCKAGKLDEALKLFEEMVEKGIKPNVVTTIDGLCKAGKLDEALKL  
 FEEMVEKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPSVVTYNTLIDGLCKAGKLDEALKLFEEMV  
 EKGIKPDVVTYNTLIDGLCKAGKLDEALKLFEEMVEKGIKPNVELTYRRVESYCRAKRFEEARGFLSEVSETDLDFDKK  
 ALEAYIEDAQFGRGDYKDHDIDYKDDDDDK

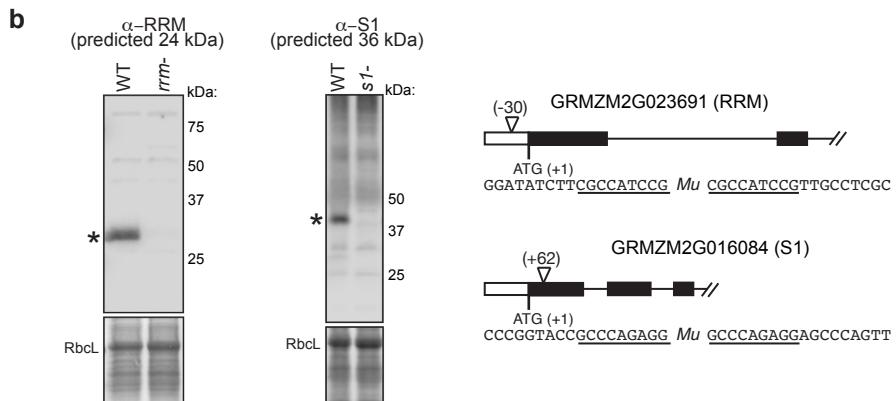
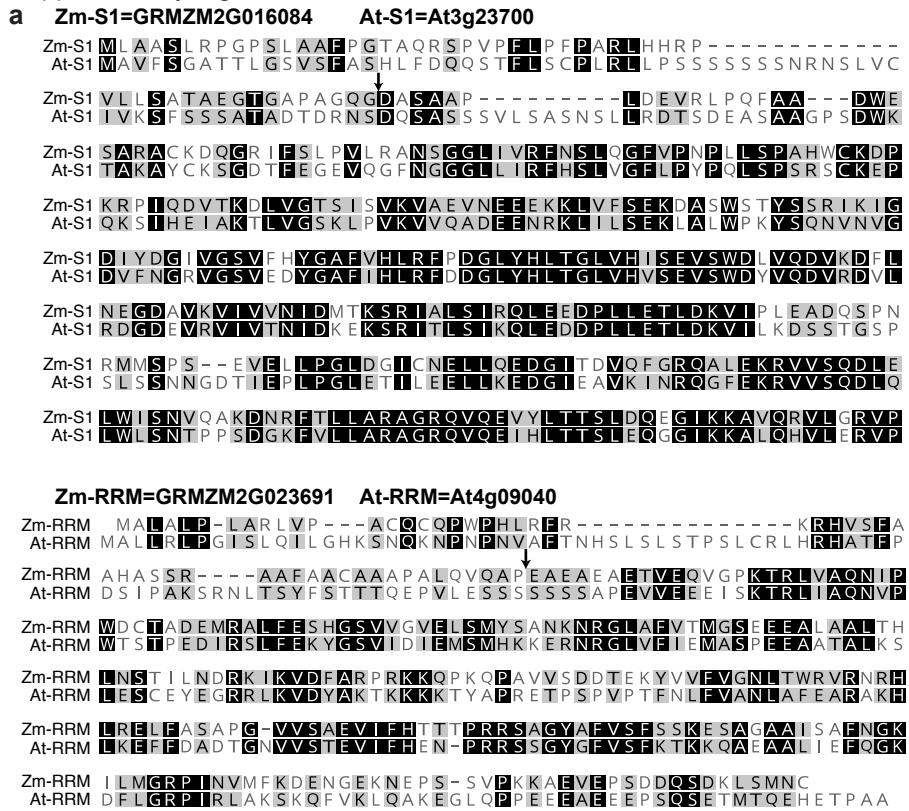
**b**



**Supplementary Figure 3. Sequences of SCD14 and SCD11 and their lack of effect on plant phenotype.** (a) Amino acid sequences of SCD14 and SCD11. Amino acids derived from PPR10 are in green. The C-terminal FLAG tag is on bold black font. (b) Visible phenotype of transgenic Arabidopsis plants expressing SCD14 and SCD11. Col-0 is the wild-type progenitor of the transgenic lines.



**Supplementary Figure 4. Controls for coimmunoprecipitation experiments.** (a) Immunoprecipitation of SCD14 and SCD11 from transgenic lines. Stromal extracts isolated from transgenic Arabidopsis expressing the indicated protein or from the Col-0 progenitor were used for immunoprecipitation with anti-FLAG antibody. The pellet (P) and supernatant (S) fractions were analyzed by immunoblot analysis, using anti-FLAG antibody. An excerpt of the Ponceau S-stained filter is shown to illustrate the abundance of the large subunit of Rubisco (RbcL), which serves as a loading control. An equal proportion of each pellet fraction was analyzed; 1/4<sup>th</sup> that proportion of each supernatant was analyzed to avoid overloading the lane. (b) Slot blot hybridization analysis of RNAs that coimmunoprecipitate with SCD14 and SCD11 from chloroplast stroma. RNA was extracted from the same immunoprecipitations analyzed in panel (a) and applied to nylon membrane via a slot-blot manifold. The same proportion of all RNA samples was analyzed to illustrate the partitioning of *psbA* RNA between the pellet and supernatant fractions. Replicate blots were hybridized with synthetic oligonucleotide probes specific for the *psbA* 5'-UTR or 3'-UTR, as indicated.



**Supplementary Figure 5. Validation of antibodies used for RIP-seq assays.**

(a) Multiple sequence alignments of the maize and Arabidopsis orthologs of the RRM-domain and S1-domain proteins enriched in the SCD11/14 coimmunoprecipitations. The RRM protein is encoded by gene AT4G09040 and GRMZM2G023591 in Arabidopsis and maize, respectively. The S1 domain protein is encoded by genes AT3G23700 and GRMZM2G016084 in Arabidopsis and maize, respectively. Arrows mark the first amino acid of the recombinant proteins used to raise polyclonal antibodies. (b) Specificity of antibodies demonstrated by immunoblot analysis of maize insertion alleles. The transposon insertion alleles are diagrammed to the right, and were obtained from the UniformMu project: the GRMZM2G023591/RRM mutant corresponds to line (mu1032521, UFMu-02565); the GRMZM2G016084/S1 mutant corresponds to line (mu1076060, UFMu-09028). Total leaf extracts from wild-type or mutant seedlings were probed with the indicated antiserum. Excerpts of the Ponceau-stained blots are shown below to illustrate equal sample loading.

## REFERENCES CITED

### CHAPTER I

- Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **8**, 479–90 (2007).
- Hall, T.M. De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr Opin Struct Biol* **36**, 116-121 (2016)
- Chen, Y. *et al.* Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat. Chem. Biol.* **12**, 717–723 (2016).
- Barkan, A., & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annual Review of Plant Biology*, 65.1: 415–442 (2014).
- Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326: 1509–1513 (2009).
- Wang, X., McLachlan, J., Zamore, PD., Hall, TM. Modular Recognition of RNA by a Human Pumilio-Homology Domain. *Cell*, 110.4: 501–512 (2002).
- Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8**, e1002910 (2012).
- Timmis, J. N., Ayliff, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
- Barkan, A. Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* **155**, 1520–32 (2011).
- Prikryl, J., Rojas, M., Schuster, G. & Barkan, A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 415–20 (2011).
- Pfalz, J., Bayraktar, O. A., Prikryl, J. & Barkan, A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J.* **28**, 2042–52 (2009).
- Miranda, R. G., McDermott, J. J. & Barkan, A. RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR–RNA interactions. *Nucleic Acids Res.* **46**, 2613–2623 (2018).
- Yagi, Y., Nakamura, T. & Small, I. The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.* 772–782 (2013).

Shen, C. *et al.* Specific RNA Recognition by Designer Pentatricopeptide Repeat Protein. *Mol. Plant* **8**, 667–670 (2015).

Gully, B. S. *et al.* The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 196–208 (2015).

Coquille, S. *et al.* An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.* **5**, 5729 (2014).

Miranda RG, Rojas M, Montgomery MP, Gribbin KP, Barkan A. RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA*; **23**, 586-99 (2017)

Kindgren P, Yap A, Bond CS, Small I. Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell*, **27**, 403-16 (2015)

Zoschke R, Watkins KP, Miranda RG, Barkan A. The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J.* **85**, 594-606 (2016)

Zhelyazkova, P. *et al.* Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.* **40**, 3092–105 (2012).

## CHAPTER II

1. Small I, Peeters N. The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 2000; 25: 46-7.
2. Barkan A, Small I. Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol* 2014; 65: 415-42.
3. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 2012; 8: e1002910. doi: 10.1371/journal.pgen.1002910.
4. Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, et al. Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 2013; 504: 168-71. doi: 10.1038/nature12651.
5. Shen C, Zhang D, Guan Z, Liu Y, Yang Z, Yang Y, et al. Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat Commun* 2016; 7: 11285. doi: 10.1038/ncomms11285.
6. Miranda RG, Rojas M, Montgomery MP, Gribbin KP, Barkan A. RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA* 2017; 23: 586-99. doi: 10.1261/rna.059568.116.
7. Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, et al. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J* 2016; 85: 532-47. doi: 10.1111/tpj.13121.
8. Guillaumot D, Lopez-Obando M, Baudry K, Avon A, Rigaill G, Falcon de Longevialle A, et al. Two interacting PPR proteins are major *Arabidopsis* editing factors in plastid and mitochondria. *Proc Natl Acad Sci U S A* 2017; 114: 8877-82. doi: 10.1073/pnas.1705780114.
9. Prikryl J, Rojas M, Schuster G, Barkan A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci USA* 2011; 108: 415-20.
10. Kindgren P, Yap A, Bond CS, Small I. Predictable alteration of sequence recognition by RNA editing factors from *Arabidopsis*. *Plant Cell* 2015; 27: 403-16. doi: 10.1105/tpc.114.134189.
11. Zoschke R, Watkins KP, Miranda RG, Barkan A. The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J* 2016; 85: 594-606. doi: 10.1111/tpj.13093.
12. Pfalz J, Bayraktar O, Prikryl J, Barkan A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J* 2009; 28: 2042-52.

13. Gully BS, Cowieson N, Stanley WA, Shearston K, Small ID, Barkan A, et al. The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Res* 2015; 43: 1918-26. doi: 10.1093/nar/gkv027.
14. Rojas M, Ruwe H, Miranda RG, Zoschke R, Hase N, Schmitz-Linneweber C, et al. Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5 and PPR10. *Nucleic Acids Res* 2018. doi: 10.1093/nar/gky737.
15. Zoschke R, Watkins K, Barkan A. A rapid microarray-based ribosome profiling method elucidates chloroplast ribosome behavior in vivo. *Plant Cell* 2013; 25: 2265-75.
16. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31: 3406-15.
17. Zhelyazkova P, Hammami K, Rojas M, Voelker R, Vargas-Suarez M, Borner T, et al. Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res* 2012; 40: 3092-105. doi: gkr1137 [pii]
18. Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* 2003; 15: 1480-95.
19. Katsamba PS, Myszka DG, Laird-Offringa IA. Two functionally distinct steps mediate high affinity binding of U1A protein to U1 hairpin II RNA. *J Biol Chem* 2001; 276: 21476-81. doi: 10.1074/jbc.M101624200.
20. Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* 2006; 25: 163-73. doi: 10.1038/sj.emboj.7600918.
21. Northrup SH, Erickson HP. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A* 1992; 89: 3338-42.
22. Coquille S, Filipovska A, Chia T, Rajappa L, Lingford JP, Razif MF, et al. An artificial PPR scaffold for programmable RNA recognition. *Nat Commun* 2014; 5: 5729. doi: 10.1038/ncomms6729.
23. Williams-Carrier R, Kroeger T, Barkan A. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA* 2008; 14: 1930-41.
24. deLorimier E, Coonrod LA, Copperman J, Taber A, Reister EE, Sharma K, et al. Modifications to toxic CUG RNAs induce structural stability, rescue mis-splicing in a myotonic dystrophy cell model and reduce toxicity in a myotonic dystrophy zebrafish model. *Nucleic Acids Res* 2014; 42: 12768-78. doi: 10.1093/nar/gku941.

25. Siegfried NA, Bevilacqua PC. Thinking inside the box: designing, implementing, and interpreting thermodynamic cycles to dissect cooperativity in RNA and DNA folding. *Methods Enzymol* 2009; 455: 365-93. doi: 10.1016/S0076-6879(08)04213-4.
26. Katsamba PS, Park S, Laird-Offringa IA. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* 2002; 26: 95-104. doi: 10.1016/S1046-2023(02)00012-9.
27. Majka J, Speck C. Analysis of protein-DNA interactions using surface plasmon resonance. *Adv Biochem Eng Biotechnol* 2007; 104: 13-36.
28. Rich RL, Myszka DG. Survey of the year 2007 commercial optical biosensor literature. *J Mol Recognit* 2008; 21: 355-400. doi: 10.1002/jmr.928.

### CHAPTER III

1. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. & Krzyzosiak, W.J. Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. *Nucleic Acids Res* **44**, 9050-9070 (2016).
2. Barkan, A. & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol* **65**, 415-442 (2014).
3. Hall, T.M. De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr Opin Struct Biol* **36**, 116-121 (2016).
4. Barkan, A. et al. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS genetics* **8**, e1002910 (2012).
5. Kindgren, P., Yap, A., Bond, C.S. & Small, I. Predictable alteration of sequence recognition by RNA editing factors from Arabidopsis. *Plant Cell* **27**, 403-416 (2015).
6. Shen, C. et al. Specific RNA recognition by designer pentatricopeptide repeat protein. *Molecular plant* **8**, 667-670 (2015).
7. Coquille, S. et al. An artificial PPR scaffold for programmable RNA recognition. *Nature communications* **5**, 5729 (2014).
8. Miranda, R.G., McDermott, J.J. & Barkan, A. RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR-RNA interactions. *Nucleic Acids Res* **46**, 2613-2623 (2018).
9. Yagi, Y., Nakamura, T. & Small, I. The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J* **78**, 772-782 (2014).
10. Sun, Y. & Zerges, W. Translational regulation in chloroplasts for development and homeostasis. *Biochimica et biophysica acta* **1847**, 809-820 (2015).
11. Pfalz, J., Bayraktar, O., Prikryl, J. & Barkan, A. Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J* **28**, 2042-2052 (2009).
12. Schult, K. et al. The nuclear-encoded factor HCF173 is involved in the initiation of translation of the psbA mRNA in *Arabidopsis thaliana*. *Plant Cell* **19**, 1329-1346 (2007).
13. Nilsson, R. & van Wijk, K.J. Transient interaction of cpSRP54 with elongating nascent chains of the chloroplast-encoded D1 protein; 'cpSRP54 caught in the act'. *FEBS Lett* **524**, 127-133 (2002).
14. Ramanathan, M. et al. RNA-protein interaction detection in living cells. *Nat Methods* **15**, 207-212 (2018).

15. Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631 (2009).
16. Panchapakesan, S.S.S. et al. Ribonucleoprotein purification and characterization using RNA Mango. *RNA* **23**, 1592-1599 (2017).
17. Rogell, B. et al. Specific RNP capture with antisense LNA/DNA mixmers. *RNA* **23**, 1290-1302 (2017).
18. McHugh, C.A. et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).
19. Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).
20. Konermann, S. et al. Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676 (2018).
21. Campbell, Z.T., Valley, C.T. & Wickens, M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* **21**, 732-738 (2014).
22. Wang, Y., Cheong, C.G., Hall, T.M. & Wang, Z. Engineering splicing factors with designed specificities. *Nat Methods* **6**, 825-830 (2009).
23. Yoshimura, H. & Ozawa, T. Monitoring of RNA Dynamics in Living Cells Using PUM-HD and Fluorescent Protein Reconstitution Technique. *Methods Enzymol* **572**, 65-85 (2016).
24. Lee, L.Y. et al. Novel plant transformation vectors containing the superpromoter. *Plant Physiol* **145**, 1294-1300 (2007).
25. Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W. & Chua, N.H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc* **1**, 641-646 (2006).
26. Ostheimer, G. et al. Group II intron splicing factors derived by diversification of an ancient RNA binding module. *EMBO J* **22**, 3919-3929 (2003).
27. Kunst, L. Preparation of physiologically active chloroplasts from *Arabidopsis*. *Methods Mol Biol* **82**, 43-48 (1998).
28. Barkan, A. Approaches to investigating nuclear genes that function in chloroplast biogenesis in land plants. *Methods Enzymol.* **297**, 38-57 (1998).
29. Fisk, D.G., Walker, M.B. & Barkan, A. Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.* **18**, 2621-2630 (1999).

30. Watkins, K. et al. A ribonuclease III domain protein functions in group II intron splicing in maize chloroplasts. *Plant Cell* **19**, 2606-2623 (2007).
31. Zhang, Y., Wen, Z., Washburn, M.P. & Florens, L. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal Chem* **82**, 2272-2281 (2010).
32. Schmitz-Linneweber, C., Williams-Carrier, R. & Barkan, A. RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5'-region of mRNAs whose translation it activates. *Plant Cell* **17**, 2791-2804 (2005).
33. Chotewutmontri, P. & Barkan, A. Multilevel effects of light on ribosome dynamics in chloroplasts program genome-wide and psbA-specific changes in translation. *PLoS genetics* **14**, e1007555 (2018).

## CHAPTER IV

- Barkan, A., Walker, M., Nolasco, M., Johnson, D. (1994). A Nuclear Mutation in Maize Blocks the Processing and Translation of Several Chloroplast mRNAs and Provides Evidence for the Differential Translation of Alternative mRNA Forms. *The EMBO Journal*, 13.13: 3170–3181.
- Small, ID., & Peeters, N. The PPR Motif – a TPR-Related Motif Prevalent in Plant Organellar Proteins. *Trends in Biochemical Sciences*, **25.2**, 45–47 (2000).
- Barkan, A. *et al.* A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8**, e1002910 (2012).
- Miranda RG, Rojas M, Montgomery MP, Gribbin KP, Barkan A. RNA-binding specificity landscape of the pentatricopeptide repeat protein PPR10. *RNA*, **23**, 586-99 (2017)
- Shen, C. *et al.* Specific RNA Recognition by Designer Pentatricopeptide Repeat Protein. *Mol. Plant* **8**, 667–670 (2015).
- Gully, B. S. *et al.* The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 196–208 (2015).
- Coquille, S. *et al.* An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.* **5**, 5729 (2014).
- Yagi, Y., Nakamura, T. & Small, I. The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.* 772–782 (2013). doi:10.1111/tpj.12377
- Barkan, A., & Small, I. Pentatricopeptide Repeat Proteins in Plants. *Annual Review of Plant Biology*, **65.1**, 415–442 (2014)
- Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. *Nat. Biotechnol.* **24**, 545–554 (2006).
- Sazani, P. & Kole, R. Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing. *J. Clin. Invest.* **112**, 481–486 (2003).
- Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
- Nelles, D. A. *et al.* Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* **165**, 488–496 (2016).
- Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676 (2018).
- Campbell, Z.T., Valley, C.T. & Wickens, M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* **21**, 732-738 (2014).

Ramanathan, M. et al. RNA-protein interaction detection in living cells. *Nat Methods* **15**, 207-212 (2018).

Butter, F., Scheibe, M., Morl, M. & Mann, M. Unbiased RNA-protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci U S A* **106**, 10626-10631 (2009).

Panchapakesan, S.S.S. et al. Ribonucleoprotein purification and characterization using RNA Mango. *RNA* **23**, 1592-1599 (2017).

Rogell, B. et al. Specific RNP capture with antisense LNA/DNA mixmers. *RNA* **23**, 1290-1302 (2017).

McHugh, C.A. et al. The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232-236 (2015).

Chu, C. et al. Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404-416 (2015).

Hammani, K. et al. An *Arabidopsis* Dual-Localized Pentatricopeptide Repeat Protein Interacts with Nuclear Proteins Involved in Gene Expression Regulation. *Plant Cell* **23**, 730-740 (2011).