FUNCTIONAL AND BIOPHYSICAL CHARACTERIZATION OF *TRYPANOSOMA BRUCEI* PENTATRICOPEPTIDE REPEAT PROTEINS

By

Pakoyo Fadhiru Kamba

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Cell and Molecular Biology – Doctor of Philosophy

2013

ABSTRACT

FUNCTIONAL AND BIOPHYSICAL CHARACTERIZATION OF *TRYPANOSOMA BRUCEI* PENTATRICOPEPTIDE REPEAT PROTEINS

By

Pakoyo Fadhiru Kamba

Shuttling of *Trypanosoma brucei* between mammalian hosts and the arthropod vector is aided by a tunable mitochondrion which downregulates its activity in the blood stream and upregulates it in the arthropod gut via differential expression of its genes. Trypanosomes, however, are deficient in transcription factors and regulate gene expression posttranscriptionally via transcript stability, differential processing, and translation. It is hypothesized that a milieu of RNA binding proteins are involved, most of which are still putative. In this thesis, I investigated the function and mode of action of *T. brucei* pentatricopeptide repeat (PPR) proteins, a family of α-helical, organellar RNA binding proteins characterized by tandem repeats of 35 amino acids.

I had three aims: (i) to elucidate the RNA binding specificity of two *T. brucei* PPR proteins, one with a molecular mass of 27 kDa (PPR27) and another with a molecular mass of 41 kDa (PPR41); (ii) to understand the contribution of each PPR motif to PPR27-RNA binding; (iii) to develop suitable methods for production of concentrated, monodisperse, PPR proteins for structural studies. For all aims, recombinant protein was produced in *Escherichia coli* and used for in vitro experiments. First, the RNA binding specificity of PPR27 was determined. We found that PPR27 selectively binds single-stranded polyguanosine RNA quadruplexes over other sequences and nucleic acid forms. PPR27 binding did not disrupt the quadruplex structure commonly adopted by poly-G sequences. There is selective enrichment of G-tracts in maxicircle genes for extensively edited mRNAs, suggesting recognition of G-rich RNA was biologically relevant. A pull-down assay identified uncleaved RNA precursors as biological ligands. An

association between PPR27 and the small mitoribosomal subunit was recently reported. Hence, PPR27 may aid crosstalk between pre-mRNA processing and translation.

Next, the RNA binding activity and solubility of PPR27 deprived of one or more PPR motifs was analyzed. Deletion of PPR motifs modestly reduced the affinity of PPR27 for RNA, and RNA binding still occurred with only two intact PPR motifs. Except for the construct with only two PPR motifs, which suffered a 7-fold drop in RNA binding affinity for G_{12} ssRNA, the RNA binding K_d of truncated variants ranged between 2.5- and 3.5-fold that of wild type. Thus, all the PPR modules in PPR27 are seemingly involved in RNA contacts. Finally, screening of truncated PPR27 variants identified one soluble enough to attain structural biology concentrations.

Thirdly, the RNA binding specificity of PPR41 was studied. Under non-equilibrium conditions, stable binding was only observed with poly(G) ssRNA. Under equilibrium conditions PPR41 showed strong and similar affinities for G_{12}, U_{12}, and $(GGU)_4$, modest affinity for A_{12}, and reduced affinity for C_{12} ssRNAs. PPR41 pulled down precursor RNA transcripts from a mitochondrial RNA extract. RNA editing inserts U's into G-tracts. A role for PPR41 in either edited RNA binding or precursor RNA processing is therefore likely.

Lastly, I probed the application to PPR proteins of some production systems which have previously been successful with some difficult proteins. I found that wild type PPR27 can be purified and refolded from inclusion bodies of its thioredoxin fusion protein, and showed that chimeras of PPR proteins with large solubility tags have promise for NMR spectroscopic and crystallization studies.

DEDICATION

To those who brought me into this world (Mom Ruth, Dad Sowali), all those who gave me a chance and made me smile, particularly my grandparents (Jessica Baluka and Late Silvesti Lodoi), my late senior uncles (John Dumba and Abubaker Mwanga), my late aunties (Sumba Agnes, Hadija Twalo, and Kanifa Katooko), and all those who believe in me.

ACKNOWLEDGEMENTS

A generation has passed since I started this journey as a naive 5 year old unaware of what was at stake, at Kanginima Primary School in January 1983. Such a long adventure cannot be without numerous bumps. Therefore, this dream could not have come true without the love of my family, relatives, friends, teachers, employers, and the Almighty who either cared or taught me or lifted me up or celebrated with me or stood with me or fought with me in both happy and sad times.

The last 5 years have been very challenging. I therefore thank my Graduate Advisor, Prof. Charles Hoogstraten for pulling me through and for giving me a chance in his group even when I seemed so naive. You invested in a very unpredictable startup, and I at least count you as one person I can lean on for a boost to start a high risk enterprise! I am also very thankful to Prof. John Wang, Prof. Honggao Yan, Prof. Donna Koslowsky, Prof. James Geiger, and Prof. Jennifer Ekstrom for their service on my Graduate Guidance Committee, and for counseling me and showing me the right scientific path in order to overcome what was a very daunting task. Prof. Koslowsky's advice, resources, and time were very critical in the success of my RNA work, and I am very thankful. Prof. Ekstrom was fundamental in training me on experimental protein molecular biology and I am also thankful. Because of Prof. Geiger's time, resources, and courtesy, I gained essential experience with protein crystallization. Prof. Wang's sharp thinking was essential in integrating the numerous ideas which evolved out of my Committee and I am thankful. From the outset, Prof. Yan knew how difficult the project I had embarked on was and his counsel shaped our thinking on how a PhD thesis could be produced from such a project.

Whenever Science floors you, your closest ones take the flack. I am therefore not only proud of my wife, Sikola Kamba Nakolantya, my daughters Vanessa Twalo Sumba and Melissa Asio

Ruth, and my son Jason Madingi Kamba for their perseverance, I also apologize to them for the numerous times I turned my back on them in order to bring PPR proteins to order. My mother (Ruth Mary Asio Lodoi), my grandmother (Jessica Baluka), my late grandfather (Silvesti Lodoi), my mother's sisters and brothers, and their loved ones, my special teachers (Rose Apia, Mr. Talima, Mr.Lipoto, Mr. Muzeyi, Keith Obot, Salim Nabiso, Robert Biibi), Uncle Peter Ojakol, and Aunt Florence Kayendeke have been very fundamental to my life and I am thankful. Some special friends in East Lansing have been fundamental to my success on this challenge. The CMB Program Directors (Prof. Susan Conrad and Prof. Kathy Meek) and the Graduate Secretaries have been very nice people and have always made sure the Program supports me financially whenever in need. I am thankful to Prof. James Kelly of MSU's Crop Science Department, whom I met on an airplane, for supporting me financially, socially, and logistically, including welcoming me and my family into his family. I am grateful to Prof. Michael Boivin of MSU's CHM (& his wife, Grace), Dr. Andrew Muganga Kizito (& his wife, Julie), Dr. Joel Lwande, Patrick Ochieng, David Achilla, and Dennis Katuuramu for helping me navigate life in MI. And, I thank the U.S embassy in Kampala and the Fulbright Scholarship Commission for taking a chance on me, and Dorothy Ngalombi of the Public Affairs unit at the Kampala U.S embassy for being such a caring person. And, I thank the following for research support: Dr. Alice Barkan (University of Oregon) for the pMALTEV-E30 plasmid gift; Dr. Lisa Lapidus for the fluorimeter and CD facilities as well as for advice on FRET; Dr. Shelagh Ferguson-Miller for sonicator, thermocycler, and crystallization robot; the MSU DOE-PRL for the fluorescence imager; and Mr Kermit Johnson/Dr. Daniel Holmes for NMR support. Finally, former colleagues in the Hoogstraten Lab who taught me some research techniques (Dr. Minako Sumita, Kristine Julien, and Dr. James Johnson Jr.) or worked with me (David Dickson) are greatly thanked.

TABLE OF CONTENTS

LIST OF TABLES xi

LIST OF FIGURES xiii

KEY TO ABBREVIATIONS xvii

CHAPTER 1 1
REVIEW OF TRYPANOSOME BIOLOGY AND PENTATRICOPEPTIDE REPEAT PROTEINS 1
1.1 BACKGROUND 2
1.1.1 Research goal and outline of thesis 2
1.1.2 Overview of *T. brucei* and trypanosomes 2
1.1.3 *T. brucei* causes severe morbidity and mortality 9
1.1.4 Current anti-HAT chemotherapy is inadequate 10
1.1.5 Novel drug targets include pentatricopeptide repeat (PPR) proteins 13
1.2 *T. BRUCEI* PATHOLOGY AND DEVELOPMENTAL BIOLOGY 15
1.2.1 *T. brucei* pathology 15
1.2.2 *T. brucei* life cycle 16
1.2.3 *T. brucei* metabolic adaptation to contrasting host environments 19
1.3 REGULATION OF *T. BRUCEI* MITOCHONDRIAL GENE EXPRESSION 22
1.3.1 Organization of the *T. brucei* mitochondrial genome 22
1.3.2 Transcription and generation of individual mRNA transcripts 25
1.3.3 RNA editing 29
1.3.4 *T. brucei* mitochondrial RNA 3' end processing 38
1.3.5 *T. brucei* mitochondrial translation 41
1.4 PENTATRICOPEPTIDE REPEAT (PPR) PROTEINS 42
1.4.1 Overview of PPR proteins 42
1.4.2 PPR proteins in mammals 46
1.4.3 PPR proteins in *T. brucei* 47
1.4.4 Mechanism of RNA binding by PPR proteins 48
REFERENCES 55

CHAPTER 2 69
ANALYSIS OF THE RNA BINDING SPECIFICITY OF THE 27 kDa *TRYPANOSOMA BRUCEI* PENTATRICOPEPTIDE REPEAT PROTEIN 69
ABSTRACT 70
2.1 INTRODUCTION 71
2.2 MATERIALS AND METHODS 72
2.2.1 Materials 72
2.2.2 Cloning and site-directed mutagenesis 73
2.2.3 Protein expression 75
2.2.4 Protein purification 76

2.2.5 PPR27 native gel electrophoresis 77
2.2.6 Electrophoretic mobility shift assays (EMSA) 78
2.2.7 Fluorescence polarization spectroscopy 79
2.2.8 Analysis of G-tract content in *T. brucei* mitochondrial transcripts 80
2.2.9 Circular dichroism (CD) spectroscopy of guanine tract RNA 81
2.2.10 RNA native gel electrophoresis 81
2.2.11 Fluorescence resonance energy transfer (FRET) 82
2.2.12 *T. brucei* mitochondrial RNA isolation 83
2.2.13 Pull-down assays of mitochondrial RNA 85
2.3 RESULTS 86
2.3.1 PPR27 construct design, soluble yield and oligomeric state 86
2.3.2 PPR27 preferentially binds guanine-rich single stranded RNA 96
2.3.3 Affinity of PPR27 for G-tract RNA increases with the number of contiguous guanosines 103
2.3.4 Polycistronic maxicircle RNA precursors are among PPR27 targets 111
2.3.5 PPR27 preferentially binds G-quadruplexes over random coil RNA 111
2.3.6 PPR27 binding does not disrupt G-quadruplex structure 116
2.4 DISCUSSION 129
REFERENCES 133

CHAPTER 3 140
STRUCTURE-ACTIVITY STUDIES OF PPR27 140
ABSTRACT 141
3.1 INTRODUCTION 142
3.2 MATERIALS AND METHODS 143
3.2.1 Materials 143
3.2.2 Multiple sequence alignment of *T. brucei brucei* PPR27 and its homologues 143
3.2.3 Cloning and site-directed mutagenesis 144
3.2.4 Protein expression and purification 148
3.2.5 Electrophoretic mobility shift assays (EMSA) 150
3.2.6 Fluorescence polarization spectroscopy 151
3.2.7 PPR27 3-dimensional (3D) structure modeling 152
3.2.8 Circular dichroism (CD) spectroscopy of free PPR27ΔCR3 152
3.2.9 Size exclusion chromatography of ΔCR3 153
3.2.10 Nuclear Magnetic resonance (NMR) spectroscopy of free ^{15}N-PPR27ΔCR3 153
3.3 RESULTS 154
3.3.1 Deletion of PPR motifs reduces but does not abrogate PPR27 RNA binding activity 154
3.3.2 PPR27's putative RNA binding face is contributed by all its PPR motifs 162
3.3.3 A significantly soluble construct was identified through screening of truncated variants 165
3.4 DISCUSSION 171
REFERENCES 174

CHAPTER 4 179

ANALYSIS OF THE RNA BINDING SPECIFICITY OF THE 41 kDa *TRYPANOSOMA BRUCEI* PENTATRICOPEPTIDE REPEAT PROTEIN 179
ABSTRACT 180
4.1 INTRODUCTION 181
4.2 MATERIALS AND METHODS 183
4.2.1 Materials 183
4.2.2 Multiple sequence alignment of *T. brucei brucei* PPR41 and its homologues 183
4.2.3 PPR41 structural modeling and signal peptide prediction 184
4.2.4 Gene design, cloning, and site-directed mutagenesis 184
4.2.5 Protein expression and purification 187
4.2.6 Electrophoretic mobility shift assays (EMSA) 187
4.2.7 Fluorescence anisotropy 188
4.2.8 Pull-down of *T. brucei* total mitochondrial RNA 189
4.3 RESULTS 189
4.3.1 PPR41 structural perspectives and protein production 190
4.3.2 PPR41 has strong affinity for GU-rich RNA 198
4.3.3 Precursor RNA transcripts are among PPR41 targets 203
4.4 DISCUSSION 205
REFERENCES 207

CHAPTER 5 213
PURIFICATION AND REFOLDING OF *TRYPANOSOMA BRUCEI* WILD TYPE PPR27 FROM THIOREDOXIN-PPR27 INCLUSION BODIES 213
ABSTRACT 214
5.1 INTRODUCTION 215
5.2 MATERIALS AND METHODS 216
5.2.1 Materials 216
5.2.2 Cloning 217
5.2.3 TRXPPR27 fusion protein expression 220
5.2.4 Determination of TRXPPR27 yield 222
5.2.5 Free PPR27 purification 222
5.2.6 N-terminal sequencing 225
5.2.7 Mass spectrometry 225
5.2.8 Refolding of PPR27 226
5.2.9 Circular dichroism spectroscopy 226
5.2.10 Size exclusion chromatography (SEC) 226
5.3 RESULTS AND DISCUSSION 227
5.3.1 Tube versus on-column Factor Xa cleavage for PPR27 purification 227
5.3.2 Identification of PPR27 from Factor Xa cleavage 230
5.3.3 PPR27 is monomeric 232
5.3.4 Removal of detergent yields an alpha-helical protein 234
REFERENCES 236

APPENDIX 241
REFERENCES 252

LIST OF TABLES

Table 1.1: Comparison of the sizes in nt (nucleotides) of pre-edited- and edited forms of pan-edited maxicircle transcripts. 31

Table 2.1: Prediction of PPR27 organellar localization..... 87

Table 2.2: Primers used in cloning and site-directed mutagenesis. 88

Table 2.3: Primers used in reverse transcription of PPR27 bound *T. brucei* mitoRNA and PCR of resultant cDNA 105

Table 2.4: PPR27's affinity for various RNA G-tracts..... 110

Table 2.5: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* gRNA genes 122

Table 2.6: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in *T. brucei* pre-edited transcripts of mildly edited mRNA genes..... 124

Table 2.7: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in *T. brucei* edited transcripts of both pan-edited and mildly edited mRNA genes 125

Table 2.8: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* never-edited mitochondrial mRNA genes..... 126

Table 2.9: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* mitochondrial rRNA genes 126

Table 2.10: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* gRNA genes 127

Table 3.1: Primers used for site-directed mutagenesis 146

Table 3.2: Thermodynamics of G_{12} ssRNA interaction with various PPR27 constructs..... 161

Table 3.3: Gross soluble yield of MBP fusion protein of various PPR27 constructs. 167

Table 4.1: Primers used in PPR41 cloning and site-directed mutagenesis 186

Table 4.2: PPR41's affinity for various ssRNA oligonucleotides 202

Table 6.1: Primers used in linker truncation and mutagenesis to dialanine (AA) 250

LIST OF FIGURES

Figure 1.1: Anatomy of *T. brucei* 4

Figure 1.2: Structure of a replicating kDNA disc showing interlocked minicircles still in the disc and released minicircles in the kinetoflagellar zone (KFZ), surrounded by proteins. 8

Figure 1.3: Drugs used in the treatment of HAT. 12

Figure 1.4: Organization of PPR motifs in PPR proteins. 14

Figure 1.5: Life cycle of *T. brucei*. 18

Figure 1.6: Compartmentalization of *T. brucei* metabolism into the glycosome and the mitochondrion. 21

Figure 1.7: *T. brucei* mitochondrial genome. 27

Figure 1.8: Interplay between the RECC (core complex) and specific editing activities. 34

Figure 1.9: WebLogo of 5668 diverse PPR motifs showing the loose consensus sequence of the PPR motif. 44

Figure 1.10: Structure of protein phosphates 5 showing the helix-turn-helix folding of each of its TPR motifs. 45

Figure 1.11: Crystal structures of *Arabidopsis thaliana* protein-only RNase P and human mitochondrial RNA polymerase. 49

Figure 1.12: Structures of human pumilio complexed with cognate RNA (left panel) and noncognate RNA (right panel). 51

Figure 2.1: Organization of PPR motifs in PPR27. 89

Figure 2.2: Predicted secondary structure of PPR27 by PSIPRED. 90

Figure 2.3: LB agar plates showing the viability of toxicity-susceptible BL21 (DE3) and toxicity-resistant C43 (DE3) *E.coli* before and after IPTG induction of PPR27 overexpression. 91

Figure 2.4: Design and expression of PPR27. 93

Figure 2.5: Interaction of PPR27 with 12 nucleotide homopolymeric RNA. 97

Figure 2.6: Nonlinear plots used in the quantitation of the K_d for interaction of PPR27 with G_{12} ssRNA and G_{12} ssDNA. 99

Figure 2.7: EMSA for interaction of PPR27 with $(G.C)_{12}$ and dG_{12} ssDNA. 102

Figure 2.8: G-tract census and pull-down assays. 106

Figure 2.9: 2% agarose gels showing EMSA of PPR27 with G-tracts having fewer than 12 guanines. 108

Figure 2.10: Anisotropy for interaction of PPR27 with G-tracts having fewer than 12 guanines. 109

Figure 2.11: Effect of thermal denaturation of G-tract RNA on PPR27 recognition. 114

Figure 2.12: CD assays for G-tract conformation. 118

Figure 2.13: Effect of PPR27 binding on G-quadruplex conformation. 119

Figure 3.1: Schematic diagram showing the motif composition of various truncated PPR27 constructs relative to the wild type. 145

Figure 3.2: Multiple sequence alignment of *T. brucei* PPR27 and its close homologues. 155

Figure 3.3: SDS-PAGE of MBP fusion proteins of truncated PPR27 constructs with a C-terminal His_6 tag. 157

Figure 3.4: 2 % tris-glycine agarose gels showing how the electrophoretic mobility of G_{12} ssRNA is retarded by increasing concentrations of truncated PPR27 variants. 158

Figure 3.5: Nonlinear plots used in the quantitation of the K_d for interaction of the various PPR27 truncated constructs with G_{12} ssRNA. 159

Figure 3.6: Predicted 3D structure of PPR27. 164

Figure 3.7: Purification and CD spectroscopy of free ΔCR3. 168

Figure 3.8: Size exclusion chromatography and $^{1}H,^{15}N$-TROSY NMR spectroscopy of ΔCR3. 169

Figure 4.1: Multiple sequence alignment of *T. brucei* PPR41 and its close homologues. 191

Figure 4.2: In silico prediction of PPR motifs in PPR41. 193

Figure 4.3: PSIPRED predicted secondary structure of PPR41. 194

Figure 4.4: Predicted 3D structure of PPR41 showing 8 putative antiparallel helical modules organized into a superhelix with a central groove. 195

Figure 4.5: Trace showing the estimation of signal peptide cleavage site by SignalP 4.0. 196

Figure 4.6: Construct design and purification of MBPPPR41. 197

Figure 4.7: 1.5 % tris-glycine agarose gels showing retardation of G_{12} and little effect on A_{12}, U_{12}, or C_{12} with increasing protein concentration. 199

Figure 4.8: Interaction of PPR41 with various ssRNA oligonucleotides. 200

Figure 4.9: 1% agarose gel showing the electrophoretic mobility of PCR amplicons of PPR41-bound *T. brucei* kRNA. 204

Figure 5.1: Design of TRXPPR27. 219

Figure 5.2: Coomassie blue stained 12% polyacrylamide gels showing TRXPPR27 purification and Factor Xa protease test cleavage. 221

Figure 5.3: Coomassie blue stained 12% SDS-PAGE comparing tube- and on-column Factor Xa cleavage of TRXPPR27. 229

Figure 5.4: MALDI-TOF mass spectrum of 0.28 nmol of PPR27 in water. 231

Figure 5.5: Size exclusion chromatography of mature wild type PPR27. 233

Figure 5.6: CD spectra of PPR27. 235

Figure 6.1: 2D NMR spectra of $^{2}H,^{15}N$-double labeled TRXPPR27 obtained from 900 MHz spectrometer. 248

Figure 6.2: Coomassie blue stained 5-12% polyacrylamide gels showing MBPPPR27 with a dialanine linker after tandem Ni^{2+} and amylose affinity chromatography..... 251

KEY TO ABBREVIATIONS

BLAST	Basic local alignment search tool
β-ME	Beta mercaptoethanol
CD	Circular dichroism
CHM	College of Human Medicine
CMB	Cell and Molecular Biology
2D	Two dimensional
3D	Three dimensional
DOE PRL	Department of Energy Plant Research Laboratory
dsRNA	Double stranded RNA
E. coli	*Escherichia coli*
EMSA	Electrophoretic mobility shift assay
ETC	Electron transport chain
FA	Fluorescence anisotropy
FRET	Fluorescence resonance energy transfer
gRNA	Guide RNA
HAT	Human African trypanosomiasis
His_6	Hexahistidine
HSQC	Heteronuclear single quantum coherence
kDa	KiloDalton
kDNA	Kinetoplast DNA
kRNA	Kinetoplast RNA
LB	Luria Bertanni
LIC	Ligation independent cloning

L-Arg	L-Arginine
L-Glu	L-Glutamate
MBP	Maltose binding protein
MBPPPR27	MBP and PPR27 fusion protein
MBPPPR41	MBP and PPR41 fusion protein
mRNA	Messenger RNA
MSU	Michigan State University
NLB	Ni^{2+} lysis buffer
NMR	Nuclear magnetic resonance
PCR	Polymerase chain reaction
PPR	Pentatricopeptide repeat
PUF	Pumilio/*fem-3* mRNA binding factor homology
RBB1	RNA binding buffer 1
rRNA	Ribosomal RNA
SCB	Sarkosyl cleavage buffer
SDS-PAGE	Sodium dodecylsulfate polyacrylamide gel electrophoresis
SLB	Sarkosyl lysis buffer
ssDNA	Single stranded DNA
ssRNA	Single stranded RNA
TALE	Transcription activator-like effector
TEV	Tobacco etch virus
TPR	Tetratricopeptide repeat
tRNA	Transfer RNA
TROSY	Transverse relaxation optimized spectroscopy
TRX	thioredoxin

CHAPTER 1

REVIEW OF TRYPANOSOME BIOLOGY AND PENTATRICOPEPTIDE REPEAT PROTEINS

1.1 BACKGROUND

1.1.1 Research goal and outline of thesis

The goal of my thesis project was to understand the function and mode of action of *Trypanosoma brucei* pentatricopeptide (PPR) proteins, with three specific aims: (i) elucidating the RNA binding specificity of two *T. brucei* PPR proteins, the smallest PPR protein with a molecular mass of 26.9 kDa (PPR27, genbank accession # XM842341) and another with a molecular mass of 40.6 kDa (PPR41, genbank accession # XM_840666). (ii) probing the role of PPR motifs in RNA binding, with a focus on PPR27; and iii) developing suitable protein overexpression, purification, and stabilization/solubilization methods for production of "biophysics grade" PPR27 protein, that is, protein samples amenable to 3-dimensional structural and dynamics studies. To start the thesis, key aspects of trypanosome biology and PPR proteins are reviewed and integrated in chapter 1. In chapter 2, I report on the RNA binding specificity of PPR27 while in chapter 3, I report on how the RNA binding affinity of PPR27 is affected by deprivation of PPR motifs. In chapter 4, the RNA binding specificity of PPR41 is presented. In chapter 5, I describe successful efforts to refold PPR27 from inclusion bodies and demonstrate that PPR27 is an α-helical protein. The thesis is capped by an appendix summarizing preliminary efforts for advanced structural characterization of PPR27.

1.1.2 Overview of *T. brucei* and trypanosomes

T. brucei is an extracellular parasitic protozoan from the family Trypanosomatidae, order Kinetoplastida, a group of protists endowed with a single flagellum and a single large mitochondrion containing a disc-shaped DNA genome in the periflagellar region of its matrix called the kinetoplast (**Figure 1.1**) (*1-4*). Each kinetoplast DNA (kDNA) disc is formed from an intercalated network of 25-50 large circular DNA molecules (20-40 kb) with conserved sequence

called maxicircles and thousands of small DNA circles (0.5- 10 kb) with heterogeneous sequence called minicircles, stabilized by numerous proteins (**Figure 1.2**) (*1-2, 4-7*). Mitochondria are believed to have arisen via the fusion of a bacterial endosymbiont with a primitive nucleated cell early in the evolution of eukaryotes (*1, 8-9*). For having a mitochondrial genome with interlocked circular DNA and no introns, kinetoplastids are similar to prokaryotes and deviant from other eukaryotes, consistent with their early emergence and divergence near the beginning of eukaryotic evolution (*1-2, 10-11*). Kinetoplastids are also different from higher eukaryotes in a few other aspects, notable among which are the execution of glycolysis in unusual peroxisome-like microbodies called glycosomes (there are multiple copies per cell) rather than in the cytosol, and the trans-splicing of a highly conserved spliced leader (SL) RNA sequence on to the 5' end of every nuclear-encoded RNA (*10, 12-14*).

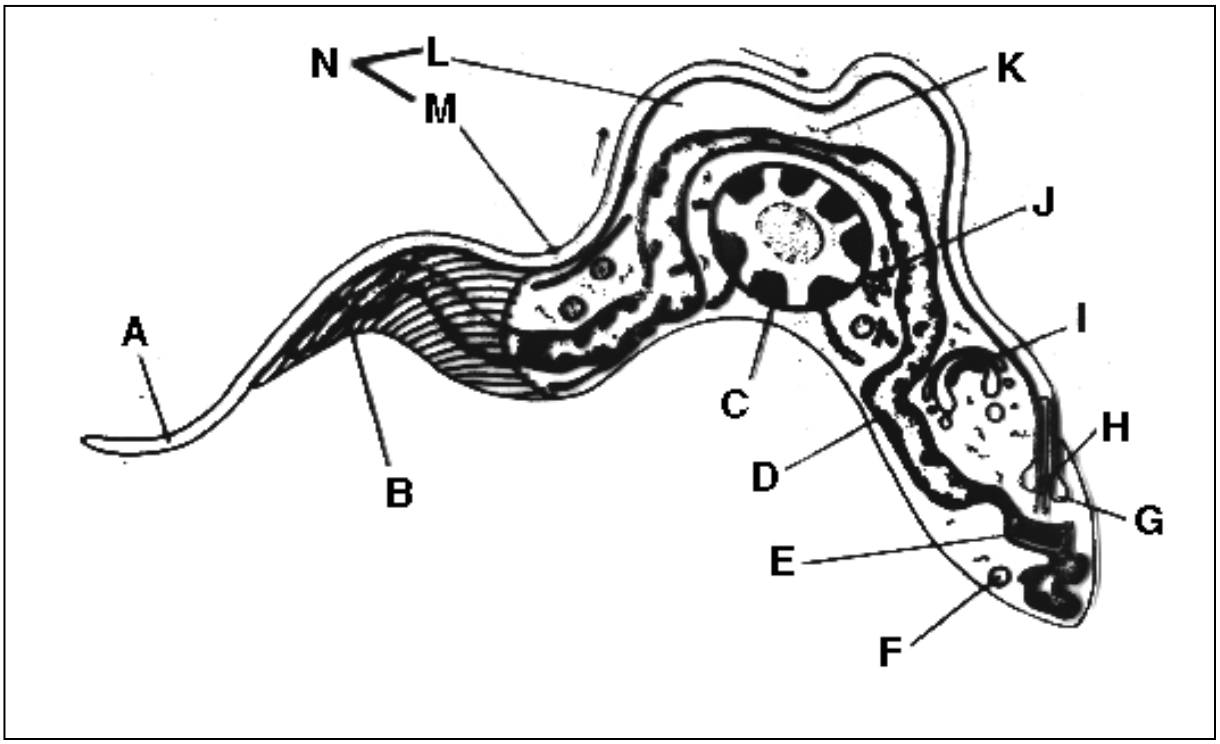


Figure 1.1: Anatomy of *T. brucei*

A, anterior flagellum; B, underlying complex skeleton; C; nucleus; D, mitochondrion; E, kinetoplast; F, glycosome; G, flagellar pocket; H, basal body; I, golgi apparatus; J, endoplasmic reticulum; K, undulating membrane; L, attachment of the flagella to the undulating membrane; M, attachment of the flagella directly to the cell body. Figure taken from http://www.icp.ucl.ac.be/~opperd/parasites/afr_sl_siickness.html.

Within the Trypanosomatidae, there are at least nine genera (*15*) of which two (*Trypanosoma* or trypanosomes and *Leishmania*) have species pathogenic to man. Besides *T. brucei*, the genus *Trypanosoma* includes *T. cruzi*, also a human pathogen, and numerous animal pathogens such as *T. vivax*, *T. congolense*, *T. simiae*, *T. evansi*, and *T. equiperdum* among others (*16*). *T. cruzi* causes human American trypanosomiasis or Chagas disease throughout Latin America and some areas in the USA where about 10 million and 100,000 people are infected respectively (*17-18*). Chagas disease can be acute, indeterminate, or chronic, and is transmitted to humans via the feces of the blood-sucking triatomid (or reduviid) insects (*14, 18-19*). Acute disease presents with symptoms similar to other infections and usually escapes the radar, but it may also affect the heart and brain resulting in meningoencephalitis, myocaditis, and a few deaths (*14, 18*). Indeterminate disease follows the acute phase and is characterized by infected muscles but no symptoms and parasitaemia. In up to 30% of cases, indeterminate disease transforms into chronic form 10-25 years after infection (*14, 18*). Chronic Chagas disease is characterized by severe enlargenment of the heart (cardiomegaly) and gastrointestinal tract (megacolon and megaoesophagus), and is fatal (*14*). On the other hand, up to 21 Leishmania species cause disease in man (*3, 19*). Together, *T. brucei*, *T. cruzi*, and *Leishmania* affect up to 500 million people, and when untreated are invariably fatal (*3*). Infections by the three trypanosomatid species are by and large the most lethal, most debilitating, and most neglected among the neglected tropical diseases (NTDs) (*3, 14, 18, 20*), with over 100,000 associated human fatalities annually (*3*). These parasites are almost exclusively inhabitants of the tropical and subtropical regions of the world, which unfortunately are also home to the world's poorest human populations (*3, 11, 18, 20*).

Trypanosomatids are generally transmitted to their hosts via different arthropod vectors. Being hemoflagellates, *T. brucei*, *T. cruzi*, and *Leishmania* can also be transmitted via blood transfusion, contaminated needles, and in utero from mother to fetus (*3, 14*). Rarely, *T. cruzi* infection can occur via organ transplants, breast feeding, and contaminated food/drinks (*3, 14*). For *T. brucei,* infection primarily occurs through the bite of a tsetse fly (genus *Glossina*), culminating into sleeping sickness or human African trypanosomiasis (HAT) in man and nagana or animal African trypanosomiasis (AAT) in livestock respectively. Vectors for the parasite are only resident in the tropical grasslands, swamps, and forests in Sub-Saharan Africa, hence the connotation of *T. brucei* as the African trypanosome, and the diseases it causes as African trypanosomiases. There are three subspecies of *T. brucei*, namely *T. brucei gambiense, T. brucei rhodesiense*, and *T. brucei brucei*. The *gambiense* subspecies infects only humans, is endemic to West and Central Africa, causes a chronic or West African form of HAT, and is responsible for over 90% of HAT cases. By contrast, the *rhodesiense* subspecies, with a niche in East and Southern Africa, is zoonotic, causes disease in both humans (up to 10% of HAT cases) and animals (over 30% of AAT cases), and causes an acute or East African form of HAT (*14, 18, 21-23*). Finally, *T. b. brucei* is a domestic animal pathogen responsible for a large proportion of AAT in livestock (*1, 18*).

Genetic and immunological differences have been used to type the three subspecies of *T. brucei* [reviewed in (*14*)]. According to the genetic composition, *T. b. brucei* and *T. b. rhodesiense* are quite similar to each other and fairly divergent from *T. b. gambiense*. On the other hand, while the non-human pathogen (*T. b. brucei*) is susceptible to human serum apolipoprotein L-1 induced lysis, the two human pathogenic *T. brucei* subspecies are resistant. Yet again, the genes encoding

serum resistance in each of the two human pathogenic *T. brucei* subspecies are distinct, enabling the discrimination of *gambiense* and *rhodesiense* by molecular diagnostics (*3, 14*).

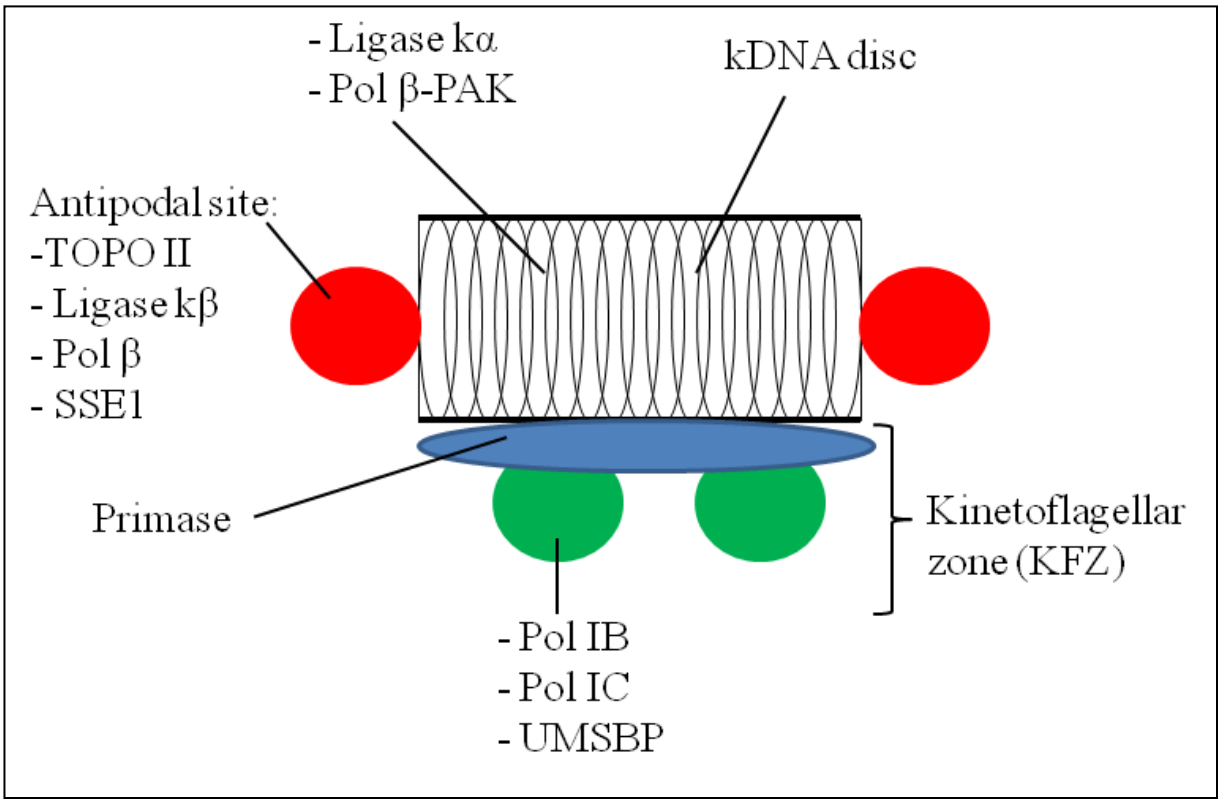


Figure 1.2: **Structure of a replicating kDNA disc showing interlocked minicircles still in the disc and released minicircles in the kinetoflagellar zone (KFZ), surrounded by proteins.**

Figure taken from (*5*) with permission from Elsevier. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

1.1.3 *T. brucei* causes severe morbidity and mortality

T. brucei causes a tandem of two devastating diseases, sleeping sickness or human African trypanosomiasis (HAT) in humans and nagana or animal African trypanosomiasis (AAT) in livestock. The parasite is localized to Sub-Saharan Africa whose swamps, forests, and tropical grasslands are home to the tsetse fly vector that passes the parasite to man, livestock, and wild game (*16*). In humans, *T. brucei* infection causes brain damage resulting in a milieu of neurological symptoms including the archetypical sleep disturbances responsible for the 'sleeping sickness' synonym for HAT; headache and other mental disorders such as irritability, hallucinations, and mania; motor symptoms such as limb tremors, tongue and limb muscle fasciculation, cerebellar ataxia, and polyneuritis; and sensory symptoms such as pruritis and hyperesthesia (*21*). Patients with HAT ultimately die if untreated (*16, 21*). In livestock, *T. brucei* infection usually has a chronic course characterized by intermittent fever, anemia and weight loss (*16*). In severe outbreaks up to 50% of infected animals die if untreated (*16*).

T. brucei is arguably the world's most inhumane parasite, having localized to the rural areas of Sub-Saharan Africa which are home to the world's poorest who have no option but to invade the tsetse fly habitats for land cultivation and livestock production, the only economic activities available to them (*16, 23*). Over 60 million people and 50 million cattle in 36 countries in SSA are in the tsetse fly belt (*3, 14, 16*). In the early 2000's, the number of infected people was about 0.5 million and annual fatalities stood at 50,000 (*14, 24*). Though major investments in prevention over the last decade reduced disease transmission, up to 17,500 new cases and at least 30,000 deaths are still reported annually (*3, 16*). There is no vaccine for HAT and odds for one in the future are slim because the *T. brucei* genome encodes over 800 variants of the glycoprotein lining its surface, the so-called variant surface glycoprotein (VSG), which it continuously

shuffles to escape host immune response by turning off the expression of one gene as it switches on another (*18, 21, 25-26*). Ultimately, chemotherapy is an indispensable remedy against HAT.

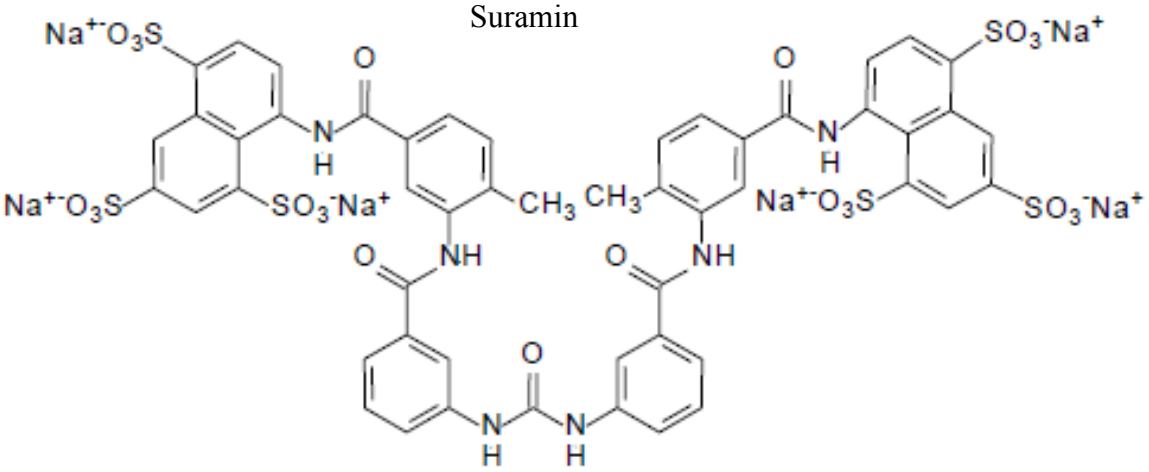
1.1.4 Current anti-HAT chemotherapy is inadequate

There are currently five drugs for treatment of HAT (**Figure 1.3**), namely, suramin (a dye), melarsoprol (an arsenical), pentamidine (a diamidine), eflornithine (a fluorinated amino acid analogue), and nifurtimox (a nitrofuran) (*18, 24, 27*). However, not all of the drugs are efficacious against all of the HAT-causing *T. brucei* subspecies nor are all of them compatible with treatment of central nervous system (CNS) disease (see section 1.2.1 for *T. brucei* pathology). Suramin, a polysulphonated naphthylamine derivative used since 1922, is a considerably safe drug, but is only effective against early stage (hemolymphatic) *T. brucei rhodesiense*- and not CNS disease as its high negative charge precludes it from crossing the blood brain barrier (*18*). On the other hand, pentamidine, a diamidine linked to the inhibition of the parasite's mitochondrial topoisomerase II and used since 1937, is only efficacious against early stage *T. b. gambiense* disease as it is positively charged and unable to accumulate to therapeutic levels in the CNS (*18*). Melarsoprol (MelB), an organic conjugate of arsenic has been used since 1949 and is the only drug with efficacy against CNS disease from both human-infecting subspecies of *T. brucei* (*18, 24*). Unfortunately, melarsoprol is very toxic; about 10% of patients on melarsoprol suffer reactive encephalopathy, half of whom die (*16, 18, 21, 24*). Nevertheless, resistant parasites have been recorded in up to 30% of patients in some localities such as Northern Uganda (*18, 21*). Eflornithine is an anti-hirsutism agent as well as a trypanostatic agent which stops parasite growth by irreversibly inhibiting the enzyme ornithine decarboxylase, thereby aiding parasite clearance by the host's immunity (*18*). Eflornithine is a safe drug but is only effective against late stage *T. b. gambiense* disease where it is currently the

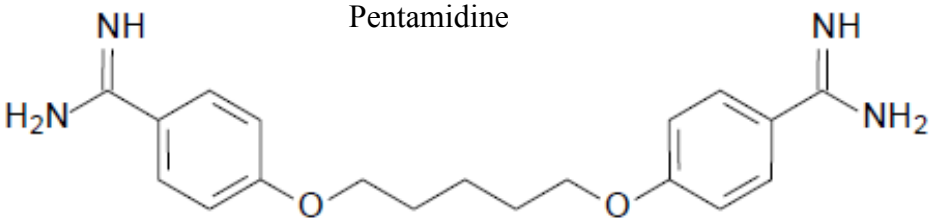
first line treatment; however, unusually large intravenous doses must be administered to attain efficacious concentration in the CNS (*18*). Nifurtimox, originally introduced for the treatment of Chagas disease in 1967 was officially adapted by the World Health Organization (WHO) for treatment of late stage *T. b. gambiense* disease in the form of a combination therapy with eflornithine in 2009 (*28*). When used as monotherapy, nifurtimox is only effective after 1-4 months of therapy (*18*), yet its adverse effects are a hindrance to completion of treatment (*18, 24*). In a nutshell, the available drugs for treatment of HAT either have inadequate efficacy or are very toxic, and new, safer and more efficacious drugs are needed.

A. Drugs used for early stage HAT

Suramin

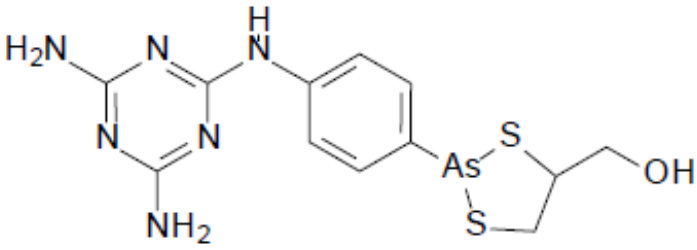


Pentamidine

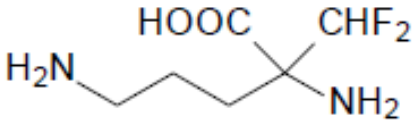


B. Drugs used for early stage HAT

Melarsoprol



Eflornithine



Nifurtimox

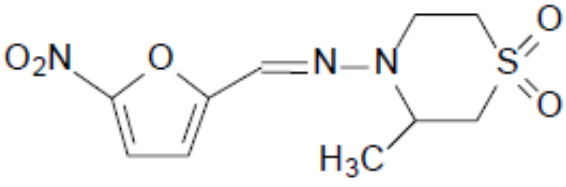


Figure 1.3: Drugs used in the treatment of HAT.

Figure modified from (*24*) with permission of Elsevier.

1.1.5 Novel drug targets include pentatricopeptide repeat (PPR) proteins

Developing new treatments requires elucidation of novel potential drug targets in the parasite which are absent in humans. With the release of complete genome sequences for *T. b. brucei* in 2005 (*25*) and *T. b. gambiense* in 2010 (*29*), identification of putative drug targets, that is genes which are highly conserved in the parasite but absent in humans, has never been any more exciting. Among the putative drug targets in *T. brucei* are the pentatricopeptide repeat (PPR) proteins, a group of organellar RNA binding proteins characterized by tandem repeats of 35 amino acids (**Figure 1.4**). Within the trypanosomes, these proteins are highly conserved (*30*) and RNAi-induced silencing of PPR genes results in adverse phenotypes to *T. brucei*, including death, consistent with an indispensable function (*30-31*). Furthermore, a BLAST (*32*) search generates few if any mammalian homologues of *T. brucei* PPR proteins, meaning they are suitable targets for development of safer and more selective inhibitors.

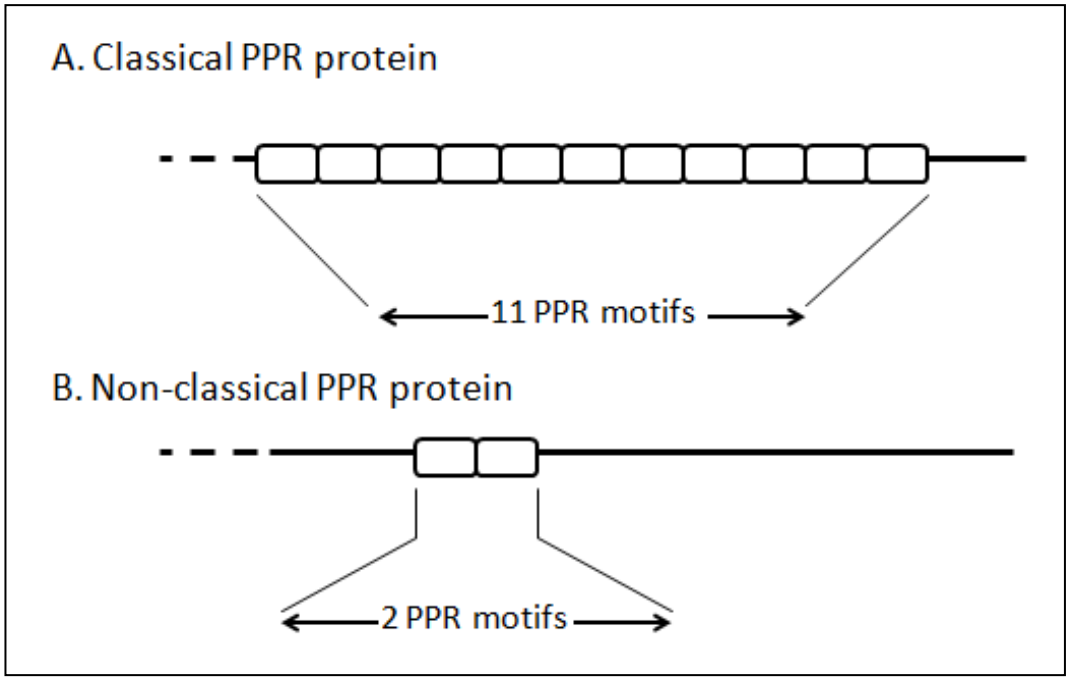


Figure 1.4: Organization of PPR motifs in PPR proteins.

In both A & B, the dashed line is the mitochondrial targeting signal (MTS) at the N-terminus, each open rectangle is a PPR motif, and solid lines represent the non-PPR portion of the protein.

Rational drug design builds on knowledge of how the protein encoded by the gene interacts with its ligand. Unfortunately, the specific RNA ligand(s) for *T. brucei* PPR proteins are unknown, as is the manner in which they contact RNA, presenting obstacles to any prospects of utilizing them in *in silico* drug screening. Structural studies per se require high concentrations of highly pure, monodisperse protein (*33-36*). PPR proteins, however, have very poor aqueous solubility (*37-40*) and are toxic to *Escherichia coli* (*E. coli*), the workhorse of heterologous protein expression. Consequently, structure-function studies of classical PPR proteins have been scanty and not even one structure of a classical PPR protein has been solved yet.

1.2 *T. BRUCEI* PATHOLOGY AND DEVELOPMENTAL BIOLOGY

1.2.1 *T. brucei* pathology

Infection by either subspecies of human pathogenic *T. brucei* ultimately ends in fatality if untreated. Following a bite by an infectious tsetse fly, the parasites multiply locally at the bite site. In about 50% of *rhodesiense* bites and rarely in *gambiense*, a localized inflammatory response to the parasites results in an ulcerative nodule called a chancre within 5-15 days of the bite (*21*). Within 1-3 weeks of the bite the parasites disseminate, first into the lymphatic system and blood where they elicit the first (or hemolymphatic) phase of HAT, and later into the central nervous system where they elicit the late (or encephalitic) phase of the disease (*14, 21*). Meanwhile, the chancre heals within 3-4 weeks to form a scar (*14*). The time course for each phase of the disease and for progression from the blood to the CNS infection differs between *T. b. rhodesiense* and *T. b. gambiense*. For *rhodesiense*, progression from hemolymphatic to CNS disease is acute and may take one to a few weeks (*18*) while the *gambiense* disease is chronic and may take anywhere from several months to a few years to produce CNS symptoms (*18, 21*).

The hemolymphatic disease first manifests with general symptoms like undulating fevers, generalized lymphadenopathy, circinate rash, pruritis, malaise, arthralgia, headache, and weight loss. At this stage and especially for *rhodesiense*, the cardiovascular system may be invaded resulting in pancarditis with severe and often fatal cardiac complications (*14, 21*). On the other hand, onset of CNS disease manifests with a diverse array of neurological and psychiatric symptoms, among which are the synonymous sleep disturbances. If untreated the patient deteriorates into terminal CNS disease characterized by seizures, severe somnolence (sleepiness), severe incontinence, cerebral edema, coma, systemic organ failure, and death (*21*).

1.2.2 *T. brucei* life cycle

Trypanosomes undergo a bimodal life cycle in which half of their development occurs in insect vectors (the tsetse fly for *T. brucei*) and the other half in either humans or other vertebrates (*3, 10, 15*). In each phase of the life cycle, the parasites alternate between proliferative forms which boost the parasite numbers and nondividing forms which are pre-programmed for differentiation in a new and hostile environment after transmission (**Figure 1.5**) (*11, 19, 21, 41*). The salivary glands of infectious tsetse flies contain nondividing short-stumpy (SS) forms of *T. brucei* called metacyclic trypomastigotes which infect the host in the saliva spray during a blood meal. The surface membrane of these insect stage metacyclic parasites is covered in a dense layer of VSG proteins, which enables them to evade antibodies in their new extracellular blood stream habitat. Once in the blood stream, parasites transform into proliferative long-slender (LS) trypomastigotes which divide by binary fission to raise parasite density, after which there is differentiation into non-dividing SS blood stream trypomastigotes pre-conditioned for survival and differentiation in the insect after ingestion in a blood meal. Once in the tsetse fly mid-gut, the parasites replace the VSG coat with a procyclin one to survive proteolysis, after which they

differentiate to proliferative forms called procyclic trypomastigotes (*41-42*). After division by binary fission to raise their density, procyclic trypomastgotes move to the salivary gland where they differentiate into epimastigotes, a form which attaches to the salivary gland with its flagellum (*41, 43*). In the salivary gland, epimastigotes proliferate before differentiating into the infectious, nondividing metacyclic trypomastigotes, a process also accompanied by replacement of the procyclin coat with the VSG (*41*). The insect phase of *T. brucei* life cycle lasts about four weeks (*3, 26*).The parasites are then ready to start a new life cycle once inoculated into a new mammal.

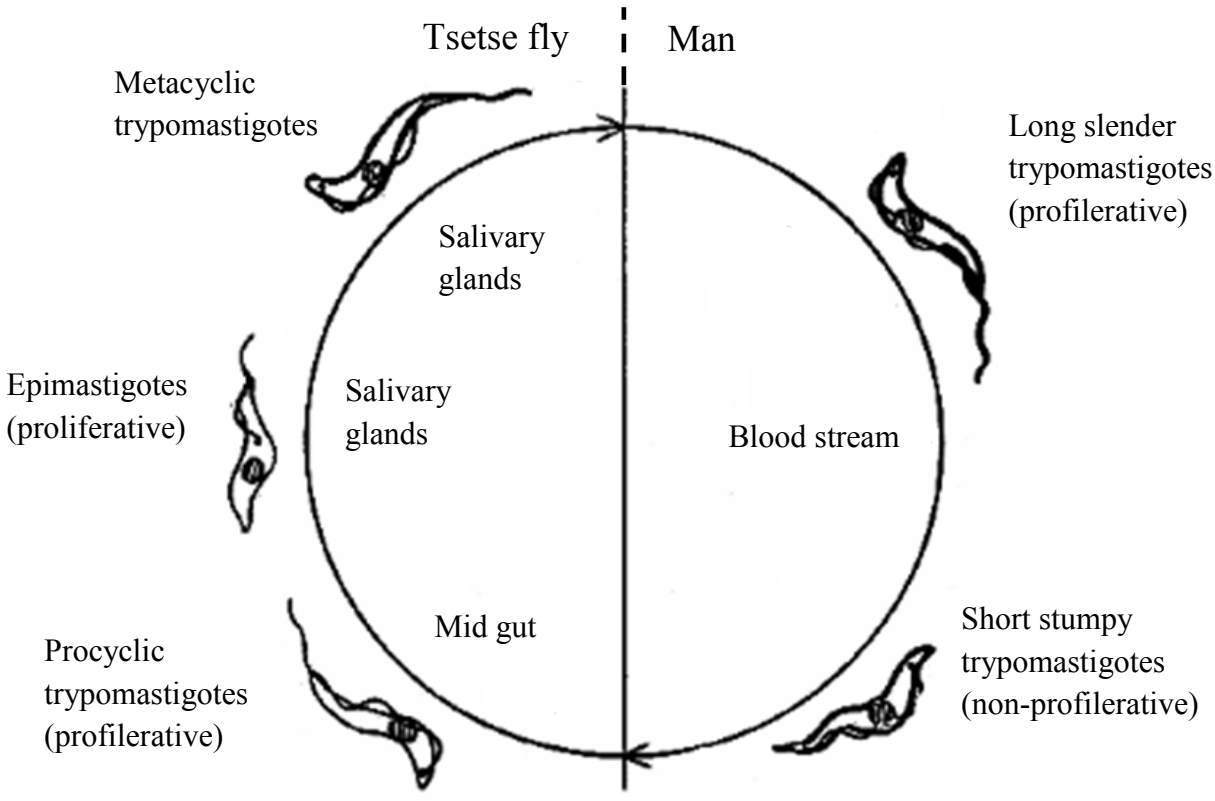


Figure 1.5: Life cycle of *T. brucei*.

Figure taken from http://dna.kdna.ucla.edu/trypanosome/images with some modifications.

1.2.3 *T. brucei* metabolic adaptation to contrasting host environments

Besides escape from hostile defenses in the mammalian blood stream and insect gut, *T.brucei* sustains its lifecycle by navigating through radical changes in its external physicochemical and metabolic environment upon transition from the vector to the mammalian host and vice versa. Not only is there a change in habitat temperature (27 ^{o}C in the tsetse fly gut and 37 ^{o}C in mammalian blood) and decrease in pH, the tsetse fly gut is deficient in sugars whereas the mammalian blood stream is endowed with millimolar concentrations of glucose (*12-13, 44-45*). It has also been argued that the the oxygen potential and osmolality in the insect gut are not the same as those of the mammalian blood (*12*).

To match the drastic environmental changes, *T. brucei* employs contrasting metabolic mechanisms in its insect and mammalian homes by modulating mitochondrial activity (*1, 12-13, 46*). In the sugar-deprived insect gut, mitochondrial activity is upregulated enabling the parasite to derive energy from oxidative phosphorylation and mitochondrial substrate level phosphorylation (*13, 46-47*). The evolution of the mitochondrion from quiescence in the blood stream parasites to metabolic competence in the insect stage procyclics is accompanied by dramatic changes in the mitochondrial transcriptome and proteome. First, while the levels of mature rRNAs are repressed in the LS bloodstream forms of *T. brucei*, they are raised in the procyclic forms (*48*), consistent with recovery to a translationally competent organelle. Secondly, the abundance of mRNA transcripts for some mitochondrial genes (specifically cytochrome oxidase subunits I (COI) and II (COII), and apocytochrome b (CYb)) are differentially elevated in procylic parasites relative to blood stream forms [(*44, 49-50*); reviewed in (*51*)]. Thirdly, the levels of tricarboxylic acid (TCA) cycle enzymes is markedly increased in procyclics over blood stream parasites [(*13*); reviewed in (*46*)]. Strangely, although all enzymes

for TCA cycle are upregulated in procylic *T. brucei*, acetyl CoA (from the catabolism of threonine abundant in the tsetse fly gut and from the catabolism of pyruvate carried over from the blood stream) is oxidized to acetate at the expense of entry into the TCA cycle for complete oxidation to CO_2. Instead, the parasite exhibits an anomalous (incomplete) TCA cycle in which α-ketoglutarate from the breakdown of the abundant proline is the substrate [(*52*); reviewed in (*46*)] (**Figure 1.6**). In contrast, *T. brucei*'s glycosomal activity is upregulated at the expense of the mitochondria in the glucose-endowed mammalian blood, with the effect that the parasite derives energy by aerobic glycolysis using substrate level phosphorylation and oxidation of reducing equivalents with a mitochondrial plant-like alternative oxidase (AOX) (*1, 9, 13, 46-47, 53*). The mitochondrial size decreases from about 25% of cellular volume in procyclics to about 5% in blood stream forms while the proportion of glycolytic enzymes in the glycosome proteome rises to about 90% (*13*). Expression of ETC components and most TCA enzymes is also suppressed in blood steam parasites [reviewed in (*13, 47*)]. Put into perspective, repressing mitochondrial activity in the sugar saturated blood stream parasites could have been evolutionarily driven by a need to circumvent the energetic inefficiency of oxidizing pyruvate to acetate.

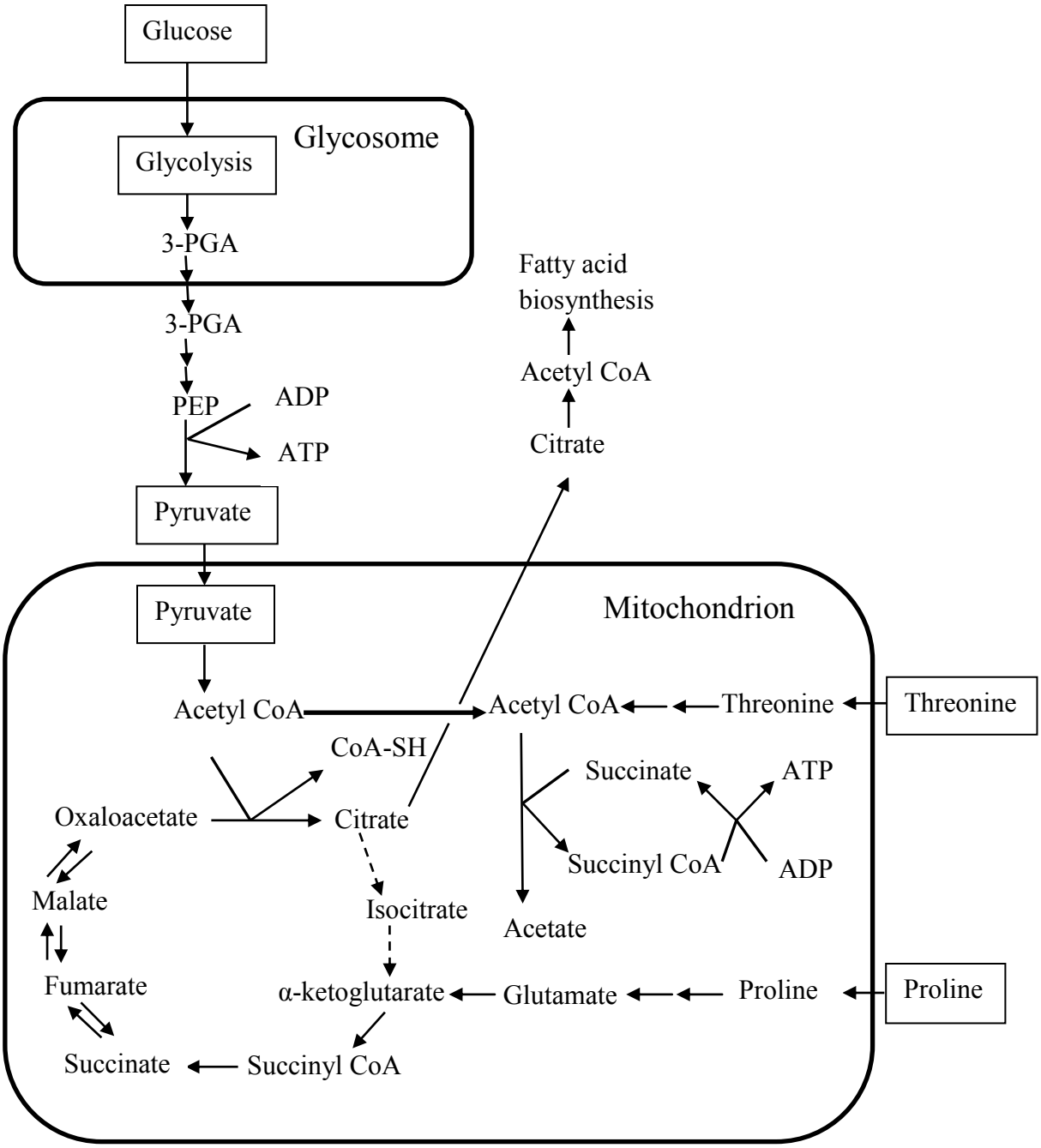


Figure 1.6: Compartmentalization of *T. brucei* metabolism into the glycosome and the mitochondrion.

3-PGA, 3-phosphoglycerate; PEP, phosphoenolpyruvate. Redrawn in simplified form from (*46, 52*)

1.3 REGULATION OF *T. BRUCEI* MITOCHONDRIAL GENE EXPRESSION

Deprivation of kDNA and deletions in maxicircles block differentiation of parasites to the procyclic form (reviewed in (*4, 50*)). Thus, though over 95% of proteins resident in the *T. brucei* mitochondrion are nuclear-encoded and imported into the mitochondrion from the cytoplasm (*1, 47, 54-55*), developmental adaptation of *T. brucei* is controlled at the level of mitochondrial gene expression. Most eukaryotic genomes encode monocistronic genes under the control of independent promoters, and their expression is predominantly regulated at transcriptional level via a wealth of transcription factors (activators, coactivators, suppresors, and cosuppressors). Trypanosomatids are not only deficient in transcription factors, but both their nuclear and mitochondrial genes are arranged in polycistrons (*12, 43*). Therefore, the preferential enrichment of some mitochondrial transcripts in insect stage *T. brucei* relative to blood stream forms (*44, 48-50*) can only be accounted for by differential posttranscriptional events.

1.3.1 Organization of the *T. brucei* mitochondrial genome

The mitochondrial DNA of kinetoplastids encodes messenger RNA (mRNA), ribosomal RNA (rRNA), and small RNAs (55-70 nucleotides) that specify RNA editing called guide RNAs (gRNAs), but not transfer RNA (tRNA) (*7, 56*). Mitochondrial tRNAs are nuclear encoded, constitutively expressed, and imported into the mitochondria from the cytoplasm (*1, 4, 47, 57*). The mitochondrial DNA of *T. brucei* consists of 50 maxicircles, each 23 kbp in size, and about 10,000 minicircles, each 1.0 kbp long (*4, 58-61*). Canonical mitochondrial genes (mRNA, rRNA) found in other eukaryotes are encoded on the maxicircle, with the minicircle encoding only gRNAs (*2, 4, 7*).

Each *T. brucei* maxicircle consists of a 15 kb gene-dense region fairly conserved across trypanosomatids on which 18 mRNA genes, two rRNA genes, and 3-4 gRNA genes are encoded,

and an 8 kb variable region (VR) devoid of genes but rich in repeated sequences and harboring the maxicircle origin of replication and putative topoisomerase binding sites (*6-7, 56, 58-59, 62-64*) (**Figure 1.7A**). The nucleotide repeats of the VR are characterized by an abundance of adenosines, often occurring in homopolymeric stretches of six (*60, 63*). At 4.5 kb and 3.5 kb upstream of the 5'end of the 12S rRNA, the VR contains a duplicated, highly conserved 11-mer guanine rich sequence (**5' GGGGTTGGTGT 3'**) (*60, 64*) identical to the highly conserved 12-mer universal minicircle sequence (UMS)/conserved sequence block 3 (CSB3) (**5' GGGGTTGGTGTA 3'**) constituting the origin of minicircle replication (*2, 59, 65-67*). However the duplicated sequences are located 2.5-3.0 kb upstream of the origin of maxicircle replication (*59*), and their function remains a puzzle.

Because of their gene composition, the kinetoplastid maxicircles are bona fide equivalents of mitochondrial genomes of other eukaryotes. There are however various features which make maxicircles unique from other mitochondrial DNA. First, not only does a maxicircle exist as one polycistron with no introns, most mRNA genes have overlapping ends, sometimes by as much as 37 nt [reviewed in (*6, 56, 68*)], necessitating machinery to specify gene start and end sites and to create recognizable start and stop codons. Secondly, many mRNA genes are cryptic (or encrypted), that is, they are unreadable by the translation machinery as they are incomplete open reading frames (ORFs) punctuated with frame shifts and often lacking translation initiation codons and stop codons (*69*), and have to be converted into translatable frames by a process of specific insertion and/or deletion of uridines called RNA editing, described in more detail below (**section 1.3.3**) (*2, 4, 6, 47, 56, 70*). Thirdly, both maxicircle strands have coding capacity, with the rRNAs and 11 mRNA genes encoded on one strand and 7 mRNA genes on the complementary strand (*6, 56, 71*) (**Figure 1.7A**), and the transcriptional apparatus works in both

directions from the same promoter (*72*). Fourthly, for still unclear reasons, kinetoplastid mitochondrial ribosomal RNAs (mt-rRNAs) are unusually rich in A and U (over 80% A+U) (*73-75*). To cap it all, the mt-rRNAs of kinetoplastids are severely reduced in size relative to their bacterial homologues (*73-75*). For example, the gene for *T. brucei* mitochondrial small ribosomal RNA (srRNA) is 611 bp (9S) and that for the large ribosomal RNA (lrRNA) is 1150 bp (12S), compared to their *E. coli* homologues which are 1542 bp (16S) and 2904 bp (23S) respectively (*74-75*). Though many critical facets of rRNAs such as the domain responsible for the peptidyl transferase reaction were conserved in the trypanosome mt-rRNAs, most other domains were severely contracted in size and some almost entirely lost (*73*). Essentially, trypanosomatids seem to have retained only the core rRNA function in the course of heavy evolutionary pressure, with the functionally dispensable RNA domains being discarded in favor of accessory proteins or for nothing (*73, 75*).

About 90% of kDNA is minicircles (*1, 7*). Each *T. brucei* kDNA network contains at least 300 unique minicircle sequences (*56, 64, 68, 70, 76*), each of which contains at least one copy of a conserved region (CON) of 100-200 bp (one copy in *T. brucei* and *Leishmania tarentolae*, two direct repeats 180^{o} apart in *Crithidia fasciculata* and *Trypanosoma lewisi*, and four direct repeats 90^{o} apart in *Trypanosoma cruzi*) and at least one variable region (VR) (*58, 67*) (**Figure 1.7B**). Within each CON, there are three short segments of highly conserved sequences called conserved sequence blocks (CSBs) denoted CSB1 (10-mer: 5' **AGGGGCGTTC** 3'), CSB2 (8-mer: 5' **CCCCGTNC** 3'), and CSB3 (12-mer: 5' **GGGGTTGGTGTA** 3') (*58, 67*). CSB3 is the UMS perfectly conserved in all minicircles across kinetoplastids and serves as the origin of minicircle replication (*2, 65-67*). A putative lagging strand replication origin has been reported

within CSB1 in *C. fasciculata* (*67*) though it is yet to be experimentally validated. The function of CSB2 is yet to be elucidated. A single bend of periodically spaced adenosines (aka bent helix) within or close to the conserved region defines each minicircle across kinetoplastid species (*76-77*). The minicircle VR contains the gRNAs (2-5 per *T. brucei* minicircle) and is responsible for the minicircle sequence diversity (*56, 58*). In the minicircle, the 5' and 3' termini of each gRNA gene are flanked by a pair of 18-mer inverted repeats with variable sequence (*58*). These repeats are separated by about 110 bp (*76*) and they are speculated to have a role in gRNA transcription as the 5' repeat is located 31-32 bp upstream of the gRNA transcription start site 5'RYAYA (*70, 76*). Whether the latter projection is true remains to be elucidated.

1.3.2 Transcription and generation of individual mRNA transcripts

Both maxicircles and minicircles employ polycistonic transcription but with some variation on the fate of the 5' end of each transcript. Genes on the maxicircle are decoded as one polycistronic transcript from a single promoter located about 1200 nucleotides upstream of the 5' end of the 12S rRNA (*48*) followed by endonucleolytic cleavage into individual gene transcripts (*2, 6, 68*). Unlike in other eukaryotes where a 7-methyl guanosine (m^7G) cap is added to the 5'end of primary transcripts, the 5'ends of mitochondrial primary transcripts in trypanosomatids are triphosphorylated (*48, 76, 78*). Because individual maxicircle transcripts are not primary transcripts (are products of posttranscriptional cleavage), their 5' termini are monophosphorylated rather than triphosphorylated (*68, 76*). Both minicircle and maxicircle encoded gRNAs are triphosphorylated at the 5' end, consistent with being primary transcripts (*68, 76*). However, despite the association of independent promoters to each gRNA gene on the minicircle, transcription is also polycistronic as the gRNA precursor transcript is about 800 nucleotides (nt), which is much longer than the 55-70 nt mature gRNA [reviewed in (*68*)]. Guide

RNA transcripts are apparently trimmed to the right size by discarding extraneous sequences at the 3'end of the precursor [reviewed in (*68*)], a process in which a RNase III family endoribonuclease mRPN1 (mitochondrial RNA precursor-processing endonuclease 1) has recently been implicated (*79*). A 144 kDa mitochondrial RNA polymerase essential for the replication and transcription of both maxicircles and minicircles in *T. brucei* has been reported (*68, 80-81*). The sequence identity of both maxicircle and minicircle promoters however remains unknown (reviewed in (*78*)).

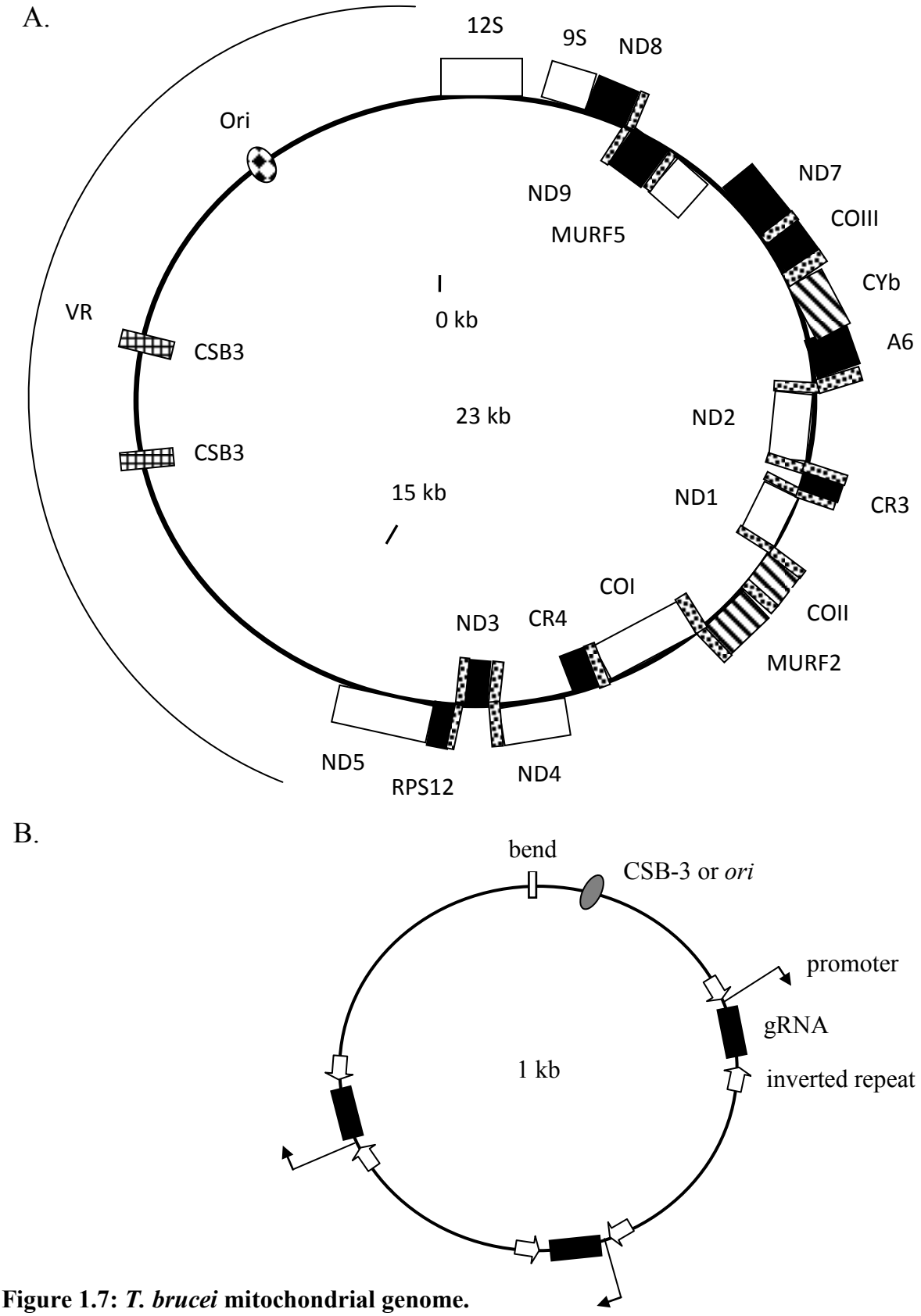


Figure 1.7: *T. brucei* mitochondrial genome.

Figure 1.7 (cont'd).

(A) Structural organization of the *T. brucei* maxicircle: 12S, large rRNA; 9S, small rRNA; ND1-9, NADH dehydrogenase subunits 1-9. Outward-facing genes are those encoded on the leading strand while inward-oriented genes are those encoded on the opposite strand; CR- 3 & 4, C-rich genes 3 & 4; MURF- 2 & 5, maxicircle unidentified reading frames 2 & 5; CO- I, II, & III, cytochrome oxidase subunits I, II, and III; CYb, apocytochrome b; A6, ATP synthase subunit 6; RPS12, ribosomal protein subunit 12. B. Structural organization of the *T. brucei* minicircle [Redrawn from (*6*)]: CSB-3, conserved sequence block 3; *ori*, origin of minicircle replication; gRNA, guide RNA.

1.3.3 RNA editing

The maxicircle DNA sequences for 12 out of the 18 mitochondrial mRNA genes in trypanosomes are not exact templates of encoded proteins and translatable mRNAs. Rather, mature mRNAs for these genes are created posttranscriptionally by a process of guided uridine insertion and/or deletion called RNA editing (*61, 82-85*). The other six *T. brucei* mitochondrial mRNA genes are complete open reading frames (ORFs) and their mRNAs match the encoded proteins. These transcripts are thus referred to as never-edited mRNAs (*2, 6, 86*). Of the 12 mRNAs subject to editing, nine are so heavily edited (pan-edited) that non-encoded nucleotides constitute more than 50% of mature mRNAs for some genes such as cytochrome oxidase subunit III (COIII), ATPase subunit 6 (ATP6), and NADH dehydrogenase subunit 9 (ND9) (*1, 56, 84-85, 87-88*) (**Table 1.1**).

RNA editing is a concerted effort of three main players: gRNA, the 20S editosome or RNA editing core complex (RECC), and various accessory factors with roles extending to as far as gRNA biogenesis, stability, and use (*2, 61, 82*). Both the site and pattern of uridine insertion and deletion are dictated by the gRNAs usually acting *in trans* (if encoded on minicircles) and occasionally *in cis* (if encoded next to the edited gene on the maxicircle) (*4, 82*). The primary structure of each gRNA is divided into three regions with unique roles in editing [reviewed in (*1, 4, 68, 82*)]: (i) a 5' region of 10-15 nucleotides with perfect Watson-Crick complementarity to the sequence 3' of the first editing site (ES) on the pre-edited mRNA that marks the start of the ES called the achor domain; ii) a middle region rich in mismatches to the RNA (starts with a mismatch and ends with a mismatch) which directs uridine insertion or deletion and specifies the pattern and extent of editing called the information or guiding domain; iii) a 3' posttranscriptionally added poly(U) tail of 15-20 nt which loosely anneals to a purine-rich

sequence 5’ of the last ES on the pre-edited RNA to reinforce the gRNA/pre-mRNA interaction and maintain a bridge across the mRNA ES throughout the editing process (*2, 4*). The mRNA stretch containing the editing sites (opposite the gRNA information domain) is called the editing block (*4*).

Table 1.1: Comparison of the sizes in nt (nucleotides) of pre-edited- and edited forms of pan-edited maxicircle transcripts.

Gene	Size of pre-edited mRNA (nt)	Size of edited mRNA (nt)
ND8	361	574
ND9	321	647
ND7	702	1246
COIII	439	970
ATPase A6	369	820
CR3	~165*	300
CR4	270	568
ND3	203	465
RPS12	172	325

Except for the CR3 gene whose boundaries in the *T. brucei* maxicircle are not yet clear, all pre-edited RNA sizes were inferred from the maxicircle genome available in Genbank (accession # M94286). *The size of the *T. brucei* CR3 pre-edited mRNA is estimated from that of its ortholog in *Leishmania amazonensis* (161 nt) (*89*). Edited mRNA sizes are in accordance with the sequences in (*58*).

After binding of the gRNA to the mRNA, one or more unpaired purines in the gRNA information domain signal U insertion at the opposite position in the mRNA editing block via both Watson-Crick and wobble G:U base pairing while a bulge in the mRNA (if mRNA editing block is longer than gRNA information domain) signals deletion of the directly opposed Us on the mRNA (*2, 4, 68, 82*) (**Figure 1.8**). The information domain of each gRNA can specify up to 20 ESs (*2*). Where long stretches of the mRNA are subject to editing, multiple gRNAs are involved in a hierarchical mechanism where the binding site for the anchor domain of the next gRNA is the post-edited domain from the last editing event, thus enforcing a 3' to 5' polarity for mRNA editing (*68, 82, 90*). Editing intermediates in which some editing sites/blocks in the 5' region of the mRNA are yet to be edited are said to be partially edited.

In *T. brucei*, the core editing activity resides in a 0.8 MDa multiprotein complex of at least 20 proteins which sediments at about 20 Svedberg units (20 S) on glycerol gradients called the 20S editosome or RNA editing core complex (RECC) (*61, 91-93*). To date the RECC has been classified into three functional variants each with a unique RNase III type endoribonuclease cleavage specificity: one specialized in trans-gRNA directed U deletion (RECC1), one specific for trans-gRNA directed U insertion (RECC2), and one customized for cis-gRNA directed editing (RECC3) (*4, 92-93*). Each RECC can also be viewed as a chimera of two protein activities: an inner nucleus of six scaffolding/assembly proteins common to all editosomes and an outer sphere of enzymatic activity coordinated by at least one cofactor/assembly protein (*68, 82*) (**Figure 1.8**). Four of the proteins in the inner nucleus, namely MP42, MP24, MP19, and MP18 (MP = mitochondrial protein) possess an oligonucleotide/oligosaccharide binding (OB) fold found in many nucleic acid binding proteins, and RNA binding activity has been reported for MP42, MP24, and MP18, consistent with a dual role in RNA binding and protein-protein

scaffolding (reviewed in (*82*)). Likewise the outer RECC assembly proteins MP81 and MP63 also have an OB fold and C2H2 zinc-finger domains, suggesting they too combine protein scaffolding and RNA binding capabilities (*61*). One hypothesis is that the RNA binding domains of the RECC scaffold proteins could be involved in the coordinated delivery of the RNA substrate to the enzymes (*61*).

When the RECC is recruited to the editing site by gRNA-mRNA duplex, RNA editing is initiated by cleavage of the phosphodiester bond immediately upstream of the anchor duplex by a RNase III type RNA editing endonuclease (REN) specific to the editing type to form a nick with a 3'OH and 5' phosphate (*68, 82, 85*). For U deletion editing, a 3' to 5' U-specific RNA editing exonuclease (REX1) then prunes away the bulging Us after which the mRNA fragments are resealed by an RNA editing ligase (REL1). For U insertion editing, an RNA editing terminal uridylyl transferase (RET2) uses UTP to add Us to the 3' OH of the 5' mRNA fragment till a pyrimidine is encountered in the gRNA information domain (*2, 68, 82, 85*). The RECC enzymatic cascade is summarized in **Figure 1.8**. Finally, two DEAD box RNA editing helicases (REH), one of which is stably associated with the RECC, are candidates for the putative unwinding activity that displaces the gRNA from the mRNA on completion of editing (*61, 82*).

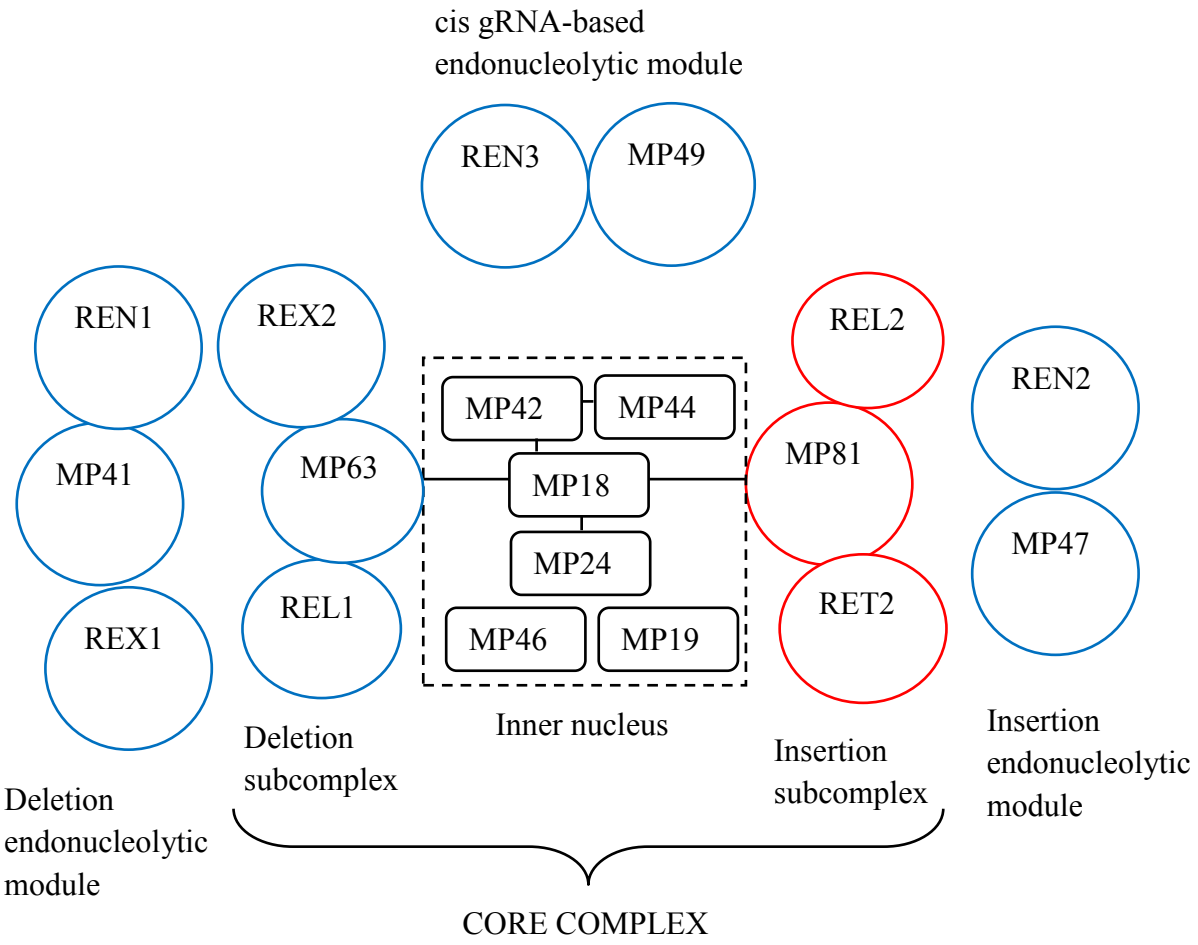


Figure 1.8: Interplay between the RECC (core complex) and specific editing activities.

Shown are the inner nucleus of six mitochondrial proteins (MP) common to both deletion core complex (RECC1) and the insertion core compex (RECC2) surrounded by a layer of effector and co-factor proteins. Figure redrawn from (*82*) with permission of Wiley & Sons, Ltd.

Several other proteins not in the RECC are involved RNA editing. These proteins, called RNA editing accessory factors (REAFs), display biochemical and genetic properties consistent with a role in RNA editing, and typically co-migrate with editing activity and gRNAs on glycerol gradients, even sometimes as stable subunits of complexes bigger than the 20S RECC [(*94*); reviewed in (*2, 4, 82, 95*)]. Based on biochemical and genetics data accumulated to date, REAFs appear to be versatile RNA metabolic players, with knockdowns and/or null mutants sometimes displaying phenotypes beyond those discernible from their known RNA binding specificity.

One set of REAFs is a pair of mitochondrial RNA binding proteins (MRP) denoted 1 (MRP1) and 2 (MRP2). MRP1 and MRP2 form an $\alpha_2\beta_2$ heterotetramer which enhances the recognition of mRNA by gRNA and clamps and stabilizes the gRNA-mRNA duplex (*61, 69, 82*). The MRP1/2 tetramer also makes contact with the RECC and with the RNA editing terminal uridylyl transferase 1 (RET1) (*96*). Strangely, MRP1/2 knockdown causes a pleiotropic phenotype in which it depresses the abundance of one edited mRNA (CYb) and one never-edited mRNA (COI) [reviewed in (*82*)]. In essence, MRP1/2's RNA-RNA match making activity could be of utility to other RNA metabolic functions outside editing.

Another REAF is the the RNA editing associated protein 1 (REAP1), a poly(G) binding protein whose depletion by a monoclonal antibody in vitro inhibits RNA editing (*95*). Strangely, the levels of all mRNA types (pre-edited, partially edited, and never edited) increase in REAP1 null mutants, suggesting it could be a general mt-mRNA instability factor (*97*). Thus, in vitro assays for REAP1 intrinsic nuclease activity and/or its association with global ribonuclease activity are needed for better insights into the activity of this protein.

A third set of REAFs consists of two *T. brucei* Arg-Gly-Gly (RGG) BOX proteins denoted 1 (TbRGG1) and 2 (TbRGG2). Both TbRGG1 and TbRGG2 are oligo(U) RNA binding proteins that associate with editing activity (*98-99*). Silencing of each RGG protein affects the abundance of edited mRNAs with minimal-to-negligible effect on gRNAs and never-edited mRNAs, hence their function could range from a general RNA binding adaptor role to edited mRNA stabilizer [(*99*); reviewed in (*82*)]. Recent reports indicate that the TbRGG proteins operate in the context of a higher molecular ribonucleoprotein complex which sediments at about 30-40 S on glycerol gradients and contains gRNAs plus at least 14 proteins, denoted the mitochondrial mRNA binding complex 1 (MRB1) in *T. brucei* or the gRNA binding complex (GRBC) in *L. tarentolae* (*81, 85, 94, 100-102*). Besides RGG1, the MRB1 complex contains TbRGG2 and a number of other proteins with activities relevant for RNA metabolism, notably the gRNA associated proteins 1 (GAP1), 2 (GAP2) and 3 (GAP3); a DExD/H box RNA helicase; the TUTase RET1; and members of the kineoplast poly(A) polymerase (KPAP1) complex such as PPR proteins (*94, 96*). PPR proteins have recently emerged as sequence specific RNA binding proteins with critical roles at all stages of posttranscriptional RNA processing (*103-104*). Whether the presence of PPR proteins in the MRB1/GRBC is due to a role in mRNA editing is however not clear. The orthologs of GAP1 and GAP2 in *L. tarentolae* are denoted GRBC1 and GRBC2 respectively (*94*).

Through RNAi induced silencing, GAP1, GAP2, and the helicase were found essential for minicircle-encoded gRNA biogenesis and trans-gRNA directed mRNA editing in both *T. brucei* and *L. tarentolae* (*81*). In both *L. tarentolae* and *T. brucei*, knockdown of either GAP1 (GRBC1) or GAP2 (GRBC2) causes accumulation of pre-edited mRNA and decrease in edited mRNA abundance, consistent with a reaction limiting function; additionally, the levels of minicircle-

encoded gRNAs are decreased (*81, 94*). To stabilize the gRNA, GAP1 and GAP2 function as an $\alpha_2\beta_2$ heterotetramer (*94*). GAP3 was first associated with edited mRNA stabilization and denoted the mitochondrial edited mRNA stability factor 1 (MERS1) (*94*). Recent work by Lukes and coworkers however suggests GAP3 may be involved in stabilization of all maxicircle-encoded RNAs (*81*).

In contrast to the RECC, whose sedimentation on glycerol gradients is insensitive to RNase A, the sedimentation rate of the MRB1 complex decreases in the presence of RNase A, an indication that its cohesion involves some RNA contacts (*101*). Most recently, Read and colleagues have reported the existence of a core nucleus of six proteins in MRB1 (GAP1, GAP2, and four uncharacterized ones) on which other proteins land on either directly or via RNA contacts (*100*). In *T. brucei*, MRP1/2 is absent in larger complexes such as the MRB1 complex (*102*) but is stably embedded in the analogous *L. tarentolae* GRBC (*94, 96*).

A fourth REAF is a 16 kDa RNA binding protein (RBP) denoted RBP16, which originally was discovered as an oligo(U) binding protein (reviewed in (*2*)). RBP16 enhances gRNA/pre-mRNA annealing by remodeling RNA secondary structure [(*105*) ; reviewed in (*2, 4, 82*)]. RBP16 knockdown however also causes a pleiotropic effect in which the abundance of one edited mRNA (CYb) and two never-edited mRNAs (COI and ND4) drop [(*106*); reviewed in (*2, 4, 82*)], hence may have other RNA metabolic functions not yet unraveled.

A fifth REAF is a specialized terminal uridylyl transferase (TUTase) that not only adds the oligo(U) to the 3' end of gRNA, rRNA and some mRNAs, but also participates in the synthesis of the long (A/U)-tail on the 3' end of mt-mRNA, denoted RNA editing TUTase 1 (RET1). Silencing of RET1 derails RNA editing, but apparently through decreased gRNA accumulation

(*107*) rather than through oligo(U) tail loss (*108*). The role of the oligo(U) tail in RNA editing is somewhat contentious. Despite being strategically positioned for base pairing to the region 5' of the editing block, in vitro the gRNA U-tail is dispensable for RNA editing (*108*).Given that RNA editing reactions in vivo occur in a more complex biochemical environment than in vitro, there is a possibility that a different phenotype would result from oligo(U) silencing in vivo.

1.3.4 *T. brucei* mitochondrial RNA 3' end processing

All trypanosome mitochondrial RNAs require some form of 3' end processing in order to attain functional competence. For mRNA, 3' end processing involves the synthesis of one of two adenosine (A) rich tails: a short poly(A) tail of 20-30 nts or a long poly(A/U) tail (70% A, 30% U) of 200-300 nts (*2, 68, 86, 109-111*). Short A-tails are added on to mRNA by the kinetoplast poly(A) polymerase 1 (KPAP1) while long (A/U) tails are added by the tandem action of KPAP1 and the TUTase RET1 (*68, 86, 111*). RNAi silencing of KPAP1 reduces the steady state levels of both never-edited and edited mRNA transcripts, and sometimes increases the abundance of pre-edited mRNA, moreover without perturbing the integrity of the 20S editosome and gRNA abundance (*111*). For rRNAs and gRNAs, 3' end processing involves the synthesis of a short oligo(U) tail of 10-20 nts by the TUTase RET1 (*1, 6, 112-113*).

Pre-edited and partially edited mRNAs contain only short A-tails, a small fraction of never-edited mRNA has short A-tails, while all fully edited and most of the never-edited mRNAs exist in long (A/U)-tailed forms (*68, 86, 111*). The interface between RNA editing and the synthesis of the A tails (polyadenylation) is however intriguing. On one hand, the two processes can occur independent of each other, at least in *T. brucei*; notably, RNA editing occurs in uncleaved maxicircle precursor transcripts while polyadenylation occurs unfettered in pre-edited mRNA (*114*). On the other hand, editing and polyadenylation appear to be linked to each other; notably,

polyadenylation is the cis regulatory element for stability of edited mRNA (*68, 109, 111*). Pre-edited mRNAs with short A-tails degrade much faster than their nonadenylated versions, reportedly via the action of a poly(A) specific 3' to 5' hydrolytic exonuclease (*115*). Hence adenylation of pre-edited mRNA would accelerate its degradation and eliminate the substrate for editing. In contrast, partially edited and fully edited mRNAs with A-tails of at least 20 nts degrade much slower than those without any A-tail (*109, 111*). In fact editing of the first editing block by the first gRNA is sufficient to stabilize adenylated mRNA (*109, 111*), an observation that has culminated into the proposal that a protective protein specific for the junction of the poly(A) and the first edited block (or minimal edited element) is the most likely shield against exonuclease action (*109*).

Thus, to have robust metabolism of edited mRNA as is the case in vivo, there must be a mechanism to coordinate editing and polyadenylation, that is, the polyadenylation machinery should be triggered into action immediately editing commences and vice versa. Recent work analyzing the protein partners of the KPAP1 complex suggests such meticulous coordination between editing and polyadenylation is actually feasible in vivo. Using co-immunoprecipitation with anti-KPAP1 antibody, Aphasizhev and colleagues observed RNA independent interactions between *T. brucei* KPAP1 and a small fraction of editing activity in a mitochondrial lysate, both core (REL1 and MP81 present) and accessory (e.g RET 1 present) (*111*). A KPAP1 complex isolated by Aphasizhev and coworkers was found to comprise peptides for at least six components of the *T. brucei* MRB1 RNA editing accessory complex (including TbRGG1), as well as 25 PPR proteins (among them PPR1 (*30*)), ribosomal proteins, and translation factors such as elongation factor Tu and initiation factor 2 (*86, 111*). The presence of the MRB1 complex and some PPR proteins was RNA dependent (*86, 111*).

PPR proteins are sequence-specific RNA binding proteins which can stabilize RNA as well coordinate effector action at the site of RNA processing (*103-104*); therefore their presence was consistent with a putative mechanism of coordination between trypanosome mt-mRNA editing and polyadenylation in vivo. Indeed, a 200 kDa heterodimer of two PPR proteins with RNA dependent connections to KPAP1, denoted kinetoplast polyadenylation/uridylation factors (KPAFs) 1 (CDS Tb927.2.3180 or PPR1 (*30*)) and 2 (CDS Tb11.01.5980) was recently identified as the stimulant for the extension of short A-tails into long A/U tails by KPAP1 and RET1 TUTase in fully edited mRNA and never-edited mRNA (*86*). Though KPA1-2 regulation of long A/U tail synthesis in never-edited mRNA suggests these proteins may act independently of editing factors, transient interactions between KPAF1 and the RECC (evidenced by REL1 and REL2), weak interactions between KPAF1 and RET1 TUTase, and very weak interactions between KPAF1 and GRBC1 have been reported (*86*).

Besides the interplay between polyadenylation and RNA editing, emerging evidence indicates the long A/U tail is what actually signals to the translational machinery that the trypanosome mt-mRNA is ready for translation (*86*) In their recent work, Aphasizheva and Aphasizhev observed that while the majority of the rRNAs sedimented at 60-70S and short A-tailed mRNA sedimented at 25-50S, the minor fraction of rRNAs existed in a heavier particle (>70S) preferentially laced with long (A/U) tailed mRNAs and mature aminoacyl tRNAs; that is, translating ribosomes preferentially accumulated on mRNAs with long A/U tails over those without such tails (*86*). Aphasizheva and Aphasizhev further observed that long (A/U)-tailed mRNAs preferentially interacted with the SSU over the LSU while editing substrates and editing machinery preferentially interacted with the LSU over the SSU, which was consistent with robust coordination and cross-talk between RNA editing and translation (*86*).

1.3.5 *T. brucei* mitochondrial translation

The trypanosome mitochondrion, like those of other eukaryotes, operates its own ribosomal machinery for synthesis of proteins encoded on its mitochondrial DNA. Unlike other eukaryotic mitochondria however, the translational apparatus of trypanosome mitochondria is strange in that: (i) it is not self sustaining as tRNAs are not encoded on kDNA and have to be imported from the cytoplasm (*57*); and (ii) its mitoribosomes contain unusually small rRNAs (9S small rRNA and 12S large rRNA), far smaller than their bacterial ancestors (16S small rRNA, 23S large rRNA) (*73-75*) and mammalian counterparts (12S small rRNA, 16S large rRNA) (*116*). Trypanosome mt-rRNAs lost so much body during evolution from their prokaryotic ancestors that trypanosome mitochondrial translation, unlike bacterial- and mammalian mitochondrial translation, is resistant to chloramphenicol (*117*), their 12S large rRNA having lost one half of a pair of ribonucleotide decamers central to chloramphenicol inhibition of the peptidyl transferase reaction (*73-74*).

Having lost many structural domains during evolution, it is apparent from recent proteomics that more RNA functions may have been delegated to nuclear-encoded proteins in trypanosome mitoribosomes compared to other eukaryotic mitoribosomes (*116*). *T. brucei* mitoribosomes contain at least 133 proteins compared to about 90 for mammalian mitoribosomes, 78 for yeast mitoribosomes, and less than 50 for bacterial ribosomes (*116*). Of the 133 proteins, 77 are from the large ribosomal subunit (LSU) while 56 are from the small ribosomal subunit (SSU) (*116*). Up to 43 of the 133 proteins were found to be homologs of bacterial and/or mammalian mitochondrial ribosomal proteins, though the *T. brucei* proteins were generally much larger than the latter duo (*116*). Of the 90 or so proteins with no bacterial and mammalian homologs, 16 were PPR proteins (*116*). A similar observation was recently made by Aphasizheva and

Aphasizhev and they have revised the number of mitoribosome-associated PPR proteins to at least 20 (*86*). In earlier studies by Schneider and coworkers, not only did conventional localization studies trace PPR proteins to the non-integral membrane fraction typical of mitoribosomes, RNAi silencing also depleted rRNAs rather than mRNA in six out of seven PPR genes assessed (*31*). Finally, despite a high prevalence in the mitoribosomes, RNAi-mediated silencing of most PPR genes so far assessed disrupted mitochondrial metabolism (oxidative phosphorylation) and caused severe phenotypes in the parasites, showing that they are essential and not redundant (*31*).

1.4 PENTATRICOPEPTIDE REPEAT (PPR) PROTEINS

1.4.1 Overview of PPR proteins

The pentatricopeptide repeat (PPR) is a degenerate 35 amino acid RNA binding motif (**Figure 1.9**) that recurs 2-30 times in the polypeptide backbone of certain proteins (the PPR proteins), typically as tandem arrays but sometimes with intervening heterogeneous motifs or sequences (**Figure 1.4**) (*37, 39, 104, 118-119*). The degenerate sequence of different PPR motifs however folds in the same way, each into a pair of antiparallel α-helices similar to that of the 34 amino acid tetratricopeptide repeat (TPR) (*118, 120-121*) (**Figure 1.10**). Multiple PPR motifs, either alone or in concert with other heterogeneous sequences, then are believed to fold into a superhelix with an outer convex face and an inner cavity resembling the concave peptide binding surface of TPR proteins and the RNA binding face of the similarly helical repeat Pumilio/*fem-3* mRNA binding factor homology (PUF) domains (*122*). PPR proteins were first discovered in Arabidopsis by Small and Peeters in 2000 (*118*) and appear to be limited to eukaryotes only (*37*). PPR proteins localize to mitochondria, although in plants 25% reside in the chloroplasts (*37*).

Across different eukaryotes, PPR proteins have a skewed distribution; they are ubiquitous in plant species with *Arabidopsis* and maize having over 450 each, but typically less than 10 in nonplant species (*37*) other than trypanosomes which have at least 37 (*86*). Over the last decade or so, evidence from genetics and limited biochemistry, predominantly of plant PPR proteins, has placed them at the scene of posttranscriptional RNA processing, ranging from demarcation of transcript termini (*123*) and RNA stabilization (*123-126*) through RNA splicing (*38, 127-128*), RNA editing (*129-132*), 3' end processing (*86, 111, 133*), and translation (*40, 125, 134-135*). PPR proteins with catalytic activity have been reported only very rarely (*103, 136-138*), hence the overriding consensus is that PPR proteins are sequence specific single-stranded RNA (ssRNA) sensors which either bind and stabilize pre-RNA or anchor effector proteins to the site of RNA processing (*103*). Knockdown and/or knockout of PPR genes from both plant and nonplant species frequently causes adverse effects to organellar RNA metabolism and the organism at large (*30-31, 139*), an indication that PPR proteins are by and large essential to the organisms, and hence good targets for design of novel inhibitors for herbicides (plants/weeds) and drugs (parasites). Understanding the mechanism by which PPR proteins interact with their RNA ligands is however crucial to any application in medicine and agriculture.

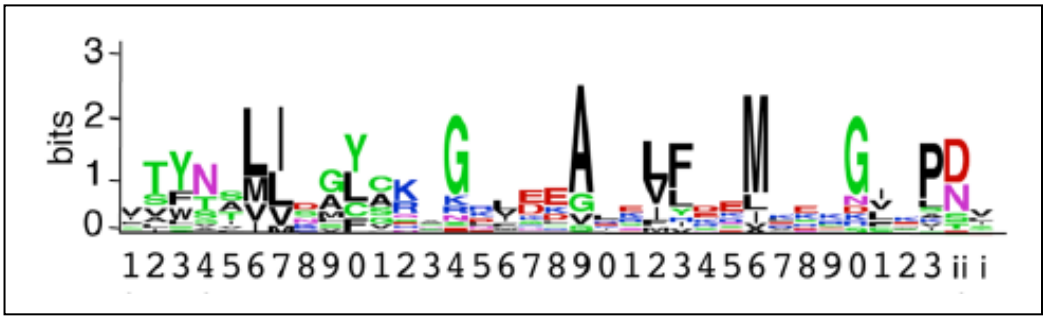


Figure 1.9: **WebLogo of 5668 diverse PPR motifs showing the loose consensus sequence of the PPR motif.**

Colored letters represent amino acid residues commonly found at each position in PPR motifs. The largest letter(s) at each position represent(s) the amino acid(s) most prevalent at that position when diverse PPR motifs are aligned. Figure taken from (*140*).

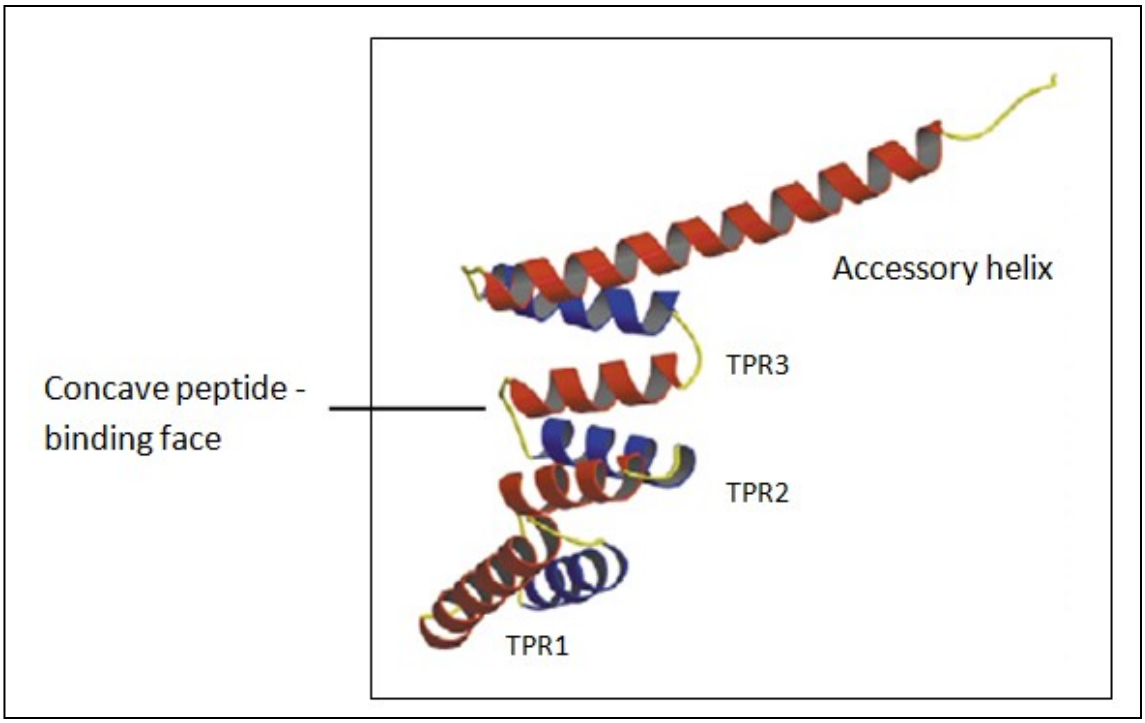


Figure 1.10: Structure of protein phosphates 5 showing the helix-turn-helix folding of each of its TPR motifs.

Figure taken from (*120*) with permission of Elsevier, and labeled.

1.4.2 PPR proteins in mammals

Seven mammalian PPR proteins, all with mitochondrial localization, have been reported to date, including mitochondrial RNA polymerase (POLRMT), mitochondrial RNase P protein 3 (MRPP3), leucine-rich pentatricopeptide repeat cassette (LRPPRC) protein, pentatricopeptide repeat domain proteins 1 (PCTD1), 2 (PCTD2) and 3 (PCTD3), and the mitochondrial ribosomal protein of the small subunit 27 (MRPS27) (*141*). POLRMT is a non-canonical PPR protein with a small domain of 2-4 PPR modules (residues 218-368) flanked by an N-terminal extension (residues 42-218) and a large C-terminal polymerase domain (residues 368-1230) (*141-142*). It has been speculated that the PPR domain may play a role in coupling transcription to RNA processing (*141*), though this requires experimental validation. MRPP3 is one of three subunits of mitochondrial RNase P, an enzyme essential for tRNA processing (*141*). MRPP3 is specifically required for processing the 5' end of tRNAs (*141*). LRPPRC, a 130 kDa protein with 22 PPR modules is an indispensable PPR protein in mammalian mitochondria (*141*). Mutations in LRPPRC disrupt mitochondrial mRNA polyadenylation and translation, and cause the French-Canadian variant of the neurodegenerative disease Leigh syndrome as a result of cytochrome c oxidase deficiency (*133, 141, 143*). PTCD1 is an 8 PPR motif protein which unlike most PPR proteins destaibilizes rather than stabilizes leucine tRNA by a yet to be deciphered mechanism (*136, 141*) while PTCD2 is a five PPR motif protein with a role in the processing of precursor ND5 and CYb mRNA transcripts (*141*). Lastly, PTCD3 (15 PPR motif) and MRPS27 (six PPR motifs) are mitoribosomal small subunit proteins that interact with the small (12S) rRNA; knockdown of each of these proteins represses mitochondrial translation (*141, 144*).

1.4.3 PPR proteins in *T. brucei*

With about 40 PPR genes identified in *T. brucei* to date (*86*), trypanosomes encode an elevated number of PPR proteins compared to other nonplant eukaryotes. Though up to four PPR proteins were predicted to be cytoplasmic in the initial phases of their characterization (*30-31*), proteomics data accumulated over the last five years (*86, 116*) suggests all are mitochondrial. Based on proteomics of mitochondrial RNA processing complexes, these proteins seem to fall into four groups: those unique to mitochondrial polyadenylation (KPAP1) complexes (at least four members), those unique to ribosomal subunit complexes (at least one member), those which sequester into both polyadenylation and mitoribosomal complexes (at least 20 members), and those at large (at least one member) (*86, 126*). The prevalence of peptides for dual sequestering PPR proteins is higher in the mitoribosomal complexes than in the polyadenylation complex (*86*). It is further worth noting that the presence of unique PPR proteins in the polyadenylation complexes is sensitive to both salt and RNase A whereas that of dual segregating members is insensitive to RNase A and seemingly enhanced by higher salt concentration (*86*).

Of the PPR proteins unique to the polyadenylation complex, two (Tb927.2.3180, also known as PPR1or KPAF1; and Tb11.01.5980, also known as KPAF2) operating as a heterodimer are essential to the synthesis of long (A/U) tails in mRNA and hence protein synthesis (*86*). Consistent with its uniqueness to the polyadenylation complex, knockdown of PPR1 selectively depletes mRNAs but not rRNAs, resulting in decreased abundance of long poly(A/U) tailed mRNA forms of at least seven transcripts and decreased steady state mRNA levels for three transcripts (*30*). Of the dual sequestering PPR proteins, five have been genetically characterized and consistent with proteomics, RNAi-mediated knockdown selectively depletes the two mt-rRNAs without affecting the steady state levels of any mRNA (*31*). A role for these proteins in

rRNA biogenesis has been proposed (*31*). The lone at large PPR protein reported to date, Tb11.01.7930 or PPR9, has been genetically characterized and it is essential for stability of COI and COII mRNAs (*126*). Downregulation of the lone PPR protein unique to the small mitoribosomal subunit complex (Tb10.389.0260 (*86*) or PPR4 (*31*)) also depletes the two mt-rRNAs without affecting the steady state levels of any mRNA (*31*). For all *T. brucei* PPR proteins, the specific RNA sequences recognized by the proteins have not yet been elucidated. Thus, not only is there an obstacle to studying the relationship between PPR protein structure and activity, whether the reported phenotypes for the few characterized proteins are due to direct interactions with cognate RNA is not clear.

1.4.4 Mechanism of RNA binding by PPR proteins

Due to poor protein solubility, success in the structural biology of PPR proteins has been minimal and high resolution experimental evidence demonstrating the contacts between PPR motifs and RNA ligands is still lacking (*39*). Only two non-classical PPR proteins, namely human mitochondrial RNAP (POLRMT) and the *Arabidopsis thaliana* dual chloroplast/mitochondrial protein-only RNase P (PRORP1) have their apo structures solved to date (Figure 12) (*142, 145*). The PPR domains of POLRMT (4 PPR motifs) and PRORP1 (4-5 PPR motifs) display the characteristic helix-turn-helix (HTH) fold and the putative RNA binding concave face is visible (**Figure 1.11**) (*142, 145*). The PPR domains in PRORP1 and POLRM T constitute only a small portion of either protein though, being flanked by long sequences of heterogeneous domains on both the N- and C-termini (*142, 145*).

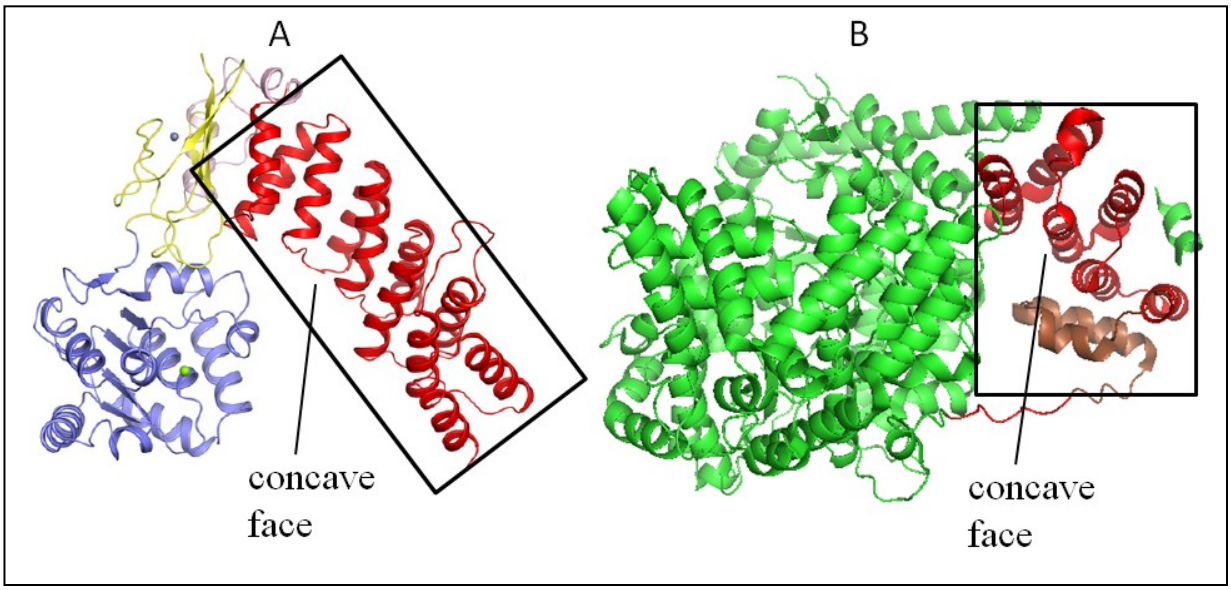


Figure 1.11: Crystal structures of *Arabidopsis thaliana* protein-only RNase P and human mitochondrial RNA polymerase.

(A) 3D structure of *Arabidopsis thaliana* protein-only RNase P (pdb i.d numbers 4G23, 4G24, 4G25, and 4G26). (B) 3D structure of human mitochondrial RNA polymerase (pdb i.d number 3SPA). The characteristic concave face believed to be the RNA binding platform is shown in each. In both (A) and (B), the PPR domains (red ± brown) are enclosed by black rectangles. In (B), red is for PPR modules in which the side chains were discernible from the electron density while brown is for a PPR module which was modeled as polyalanine due to poor resolution of its amino acid side chains on the electron density map (*142*). Figure (A) was adapted from (*145*) while coordinates for Figure (B) (*142*) were downloaded from the protein data bank (pdb) and processed with PyMOL (Schrodinger, LLC).

While the hunt for structures of RNA-bound PPR proteins drags on, five indirect sources have been valuable in drawing some insights on how PPR proteins contact RNA: (i) the similarity of PPR proteins to the well characterized PUF domain RNA binding proteins and the transcription activator-like effector (TALE) dsDNA binding proteins (*39, 146-147*); (ii) comparison of the minimal RNA ligand with the number of PPR motifs in a PPR protein (*125, 148*); (iii) in vitro binding data with truncated PPR proteins (*149*); (iv) PPR motif alignments and phylogenetic analysis; and (v) computational matching of each residue in PPR motifs with each nucleotide in a minimal RNA ligand, in concert with in vitro binding studies (*119, 140*). Though they target chemically different nucleic acids, PUF domains and TALEs bind their cognate ligands in a similar manner in which each repeat module contacts one nucleotide (**Figure 1.12**) (*39, 147*). PUF domains can also display promiscuity in which mutant/longer RNAs are accomodated by flipping out non-cognate bases (**Figure 1.12**) (*146*). A hypothesis of one PPR motif to one nucleotide binding mechanism for PPR proteins is therefore not far-fetched, and unsurprisingly there is consistency with the limited biochemical data available. First, RNA footprinting reports so far indicate that the number of nucleotides protected from RNase digestion by PPR proteins is about the same as the number of PPR motifs in the protein (*125, 148*). Secondly, structure-activity studies have demonstrated retention of RNA binding activity in severely truncated PPR proteins. For example, the high chlorophyll fluorescent mutant 152 (HCF152) comprising12 PPR modules retains activity in all variants with two PPR modules and in one out of four variants with a single PPR module (*149*).

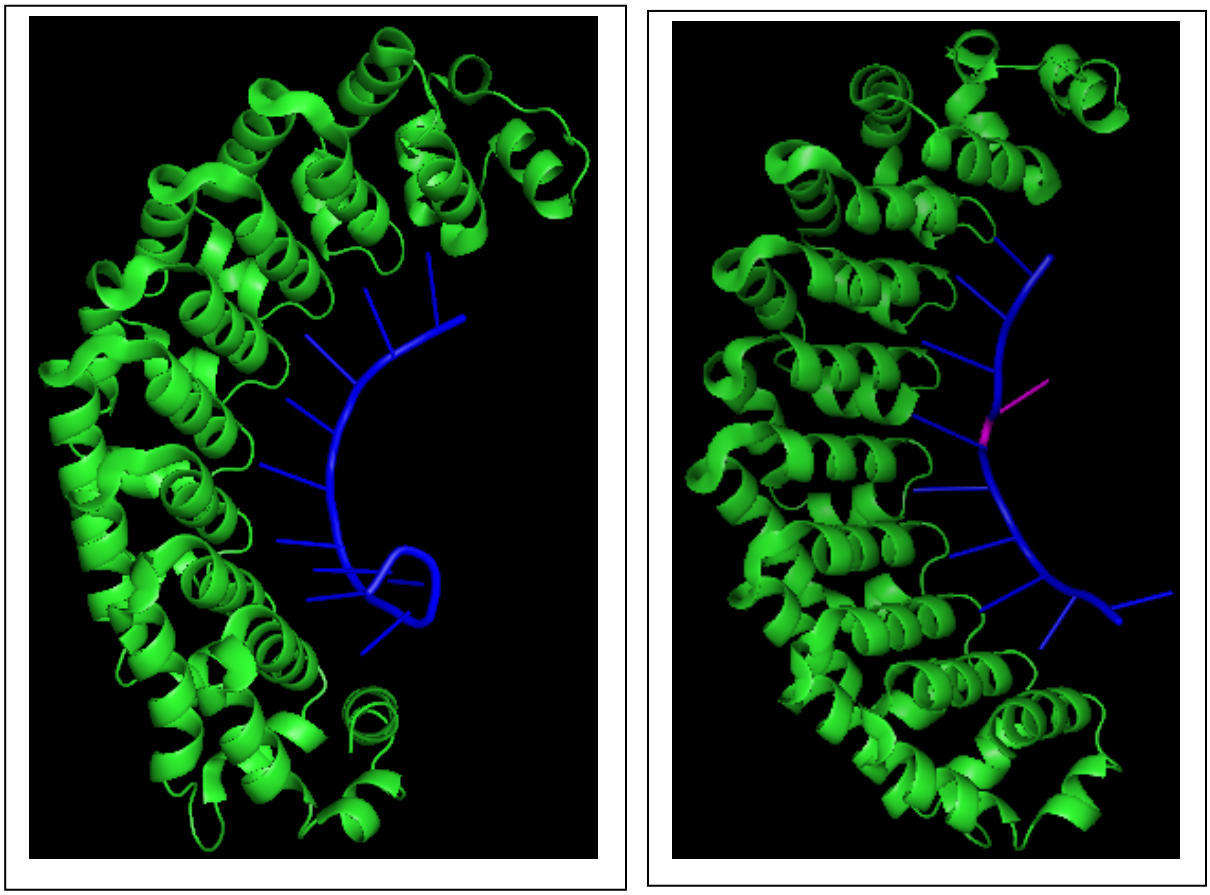


Figure 1.12: Structures of human pumilio complexed with cognate RNA (left panel) and noncognate RNA (right panel).

The PUF domain is colored green while the RNA is colored blue, except the flipped out uridine (U6) in the right panel which is colored magenta. The left panel is the crystal structure of the complex of human Pumilio1 with its cognate NRE2 RNA (pdb number 1M8Y). The right panel is the crystal structure of human Pumilio1 in complex with a noncognate puf5 RNA (pdb number 3BSX). Coordinates for both structures were mined from the pdb and processed in PyMOL. For more insights, see Aggarwal and coworkers (*146*).

In both PUFs and TALEs, the amino acid residues which interface with the nucleic acid ligands and hence determine ligand specificity are localized to specific positions in each repeat motif; for the PUFs it is the amino acid residues at positions 12, 13 , and 16 of each motif while for the TALEs it the residues at positions 12 and 13 only (*39*). For PPR proteins, the emerging consensus based on phylogenetic analysis, computational modeling and limited biochemistry is that the RNA binding residues are localized to positions 1, 3, and 6 of each PPR motif (*119, 140, 150*). Through evolutionary analysis, Small and colleagues found high rates of positive selection for amino acid residues at position 1, 3, and 6 leading them to propose that these could be the RNA binding residues and the RNA ligand specificity determinants (*150*). Then through iterative alignment of amino acids at each position (say first residues) in the PPR motifs with the nucleotides in the RNA footprint of plant PPR10, Small and colleagues again found that amino acids at positions 1 and 6 of each PPR motif correlated best with the base in the RNA ligands (*119*). A putative RNA recognition code in which the amino acid residue at position 6 in the n^{th} PPR motif collaborates with that at position 1 in the $(n+1)^{th}$ motif to bind one nucleotide was thence proposed (*119*). Most recently, Nakamura and coworkers iteratively aligned the most common residues at each position in PPR motifs of 24 plant PPR proteins involved in RNA editing, and found the amino acid residues at positions 1, 3, and 6 to have the best correlation with the bases (*140*). Whether the apparent consensus from computational modeling is universal awaits experimental structure-activity studies with diverse PPR proteins.

In order to extend structure-function studies of PPR proteins to those in trypanosomes, and to further current knowledge of PPR protein function generally, I determined the RNA binding activity of two *T. brucei* PPR proteins (PPR27 and PPR41) for this thesis. In Chapter 2, I

demonstrate using in vitro assays with recombinant protein and synthetic RNA oligonucleotides that PPR27 has preferential affinity for guanosine rich ssRNA in the quadruplex conformation over other ssRNA sequences, dsRNA, and ssDNA, findings which were backed up by in vitro selection (Neil White, unpublished). I further show that PPR27 binding does not perturb the quadruplex structure. Finally, using a pull-down assay, I show that PPR27 is involved in *T. brucei* mitochondrial polycistronic precursor RNA binding. In Chapter 3, I demonstrate using in vitro RNA binding assays with truncated variants of PPR27 that reduction of PPR motifs in PPR27 modestly reduces affinity without abolishing RNA binding, consistent with the hypothesis that all PPR modules are involved in RNA contacts. This work is accompanied by a computationally generated 3D structural model showing the putative RNA binding cavity for PPR27 and predicting the involvement of all PPR motifs in RNA binding. In Chapter 4, I demonstrate using in vitro RNA binding assays with recombinant protein and synthetic RNA that PPR41 preferentially binds ssRNA rich in guanosines and uridines over other sequences. Additionally, I demonstrate using a pull-down assay that PPR41 is also involved in mitochondrial polycistronic precursor RNA binding. All RNA binding assays in chapters 2, 3, and 4 utilized fusion proteins with the *Escherichia coli* maltose binding protein (MBP). Because difficulties in recombinant PPR protein production and poor protein solubility have rendered structural studies of classical PPR proteins unsuccessful to date, I also investigated the application to PPR protein production of some systems which have previously been successful with some difficult proteins. Accordingly, I demonstrate in the last portion of the thesis (Appendix I and II) that free wild type PPR27 can be successfully purified and refolded from inclusion bodies of its thioredoxin fusion protein. I also detail outcomes of preliminary NMR

spectroscopic and crystallization studies of various constructs and chimeric forms of PPR27, and demonstrate the most promising schemes for further optimization.

REFERENCES

REFERENCES

1. Schneider, A. (2001) Unique aspects of mitochondrial biogenesis in trypanosomatids, *Int J Parasitol 31*, 1403-1415.

2. Lukes, J., Hashimi, H., and Zikova, A. (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates, *Curr Genet 48*, 277-299.

3. Stuart, K., Brun, R., Croft, S., Fairlamb, A., Gurtler, R. E., McKerrow, J., Reed, S., and Tarleton, R. (2008) Kinetoplastids: related protozoan pathogens, different diseases, *J Clin Invest 118*, 1301-1310.

4. Lukes, J., Hashimi, H., Verner, Z., and Cicova, Z. (2010) The Remarkable Mitochondrion of Trypanosomes and Related Flagellates, *Structures and Organelles in Pathogenic Protists 17*, 227-252.

5. Liu, B., Liu, Y., Motyka, S. A., Agbo, E. E., and Englund, P. T. (2005) Fellowship of the rings: the replication of kinetoplast DNA, *Trends Parasitol 21*, 363-369.

6. Koslowsky, D. J. (2009) Complex interactions in the regulation of trypanosome mitochondrial gene expression, *Trends Parasitol 25*, 252-255.

7. Shapiro, T. A., and Englund, P. T. (1995) The structure and replication of kinetoplast DNA, *Annu Rev Microbiol 49*, 117-143.

8. Gray, M. W., Burger, G., and Lang, B. F. (1999) Mitochondrial evolution, *Science 283*, 1476-1481.

9. Priest, J. W., and Hajduk, S. L. (1994) Developmental regulation of Trypanosoma brucei cytochrome c reductase during bloodstream to procyclic differentiation, *Mol Biochem Parasitol 65*, 291-304.

10. Stevens, J. R., Noyes, H. A., Schofield, C. J., and Gibson, W. (2001) The molecular evolution of Trypanosomatidae, *Adv Parasitol 48*, 1-56.

11. Lee, S. H., Stephens, J. L., and Englund, P. T. (2007) A fatty-acid synthesis mechanism specialized for parasitism, *Nat Rev Microbiol 5*, 287-297.

12. Kramer, S. (2012) Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids, *Mol Biochem Parasitol 181*, 61-72.

13. Vertommen, D., Van Roy, J., Szikora, J. P., Rider, M. H., Michels, P. A., and Opperdoes, F. R. (2008) Differential expression of glycosomal and mitochondrial proteins in the two major life-cycle stages of Trypanosoma brucei, *Mol Biochem Parasitol 158*, 189-201.

14. Barrett, M. P., Burchmore, R. J., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J., and Krishna, S. (2003) The trypanosomiases, *Lancet 362*, 1469-1480.

15. Lopes, A. H., Souto-Padron, T., Dias, F.A., Gomes, M.T., Rodrigues, G.C., Zimmermann, L.T., e Silva, T.L.A., Vermelho, A.B. (2010) Trypanosomatids: odd organisms, devastating diseases, *The Open Parasitology Journal 4*, 30-59.

16. Holmes, P. (2012) Tsetse-transmitted trypanosomes - their biology, disease impact and control, *J. Invertebr. Pathol.*

17. Bern, C., Montgomery, S. P., Herwaldt, B. L., Rassi, A., Jr., Marin-Neto, J. A., Dantas, R. O., Maguire, J. H., Acquatella, H., Morillo, C., Kirchhoff, L. V., Gilman, R. H., Reyes, P. A., Salvatella, R., and Moore, A. C. (2007) Evaluation and treatment of chagas disease in the United States: a systematic review, *JAMA 298*, 2171-2181.

18. Wilkinson, S. R., and Kelly, J. M. (2009) Trypanocidal drugs: mechanisms, resistance and new targets, *Expert Rev Mol Med 11*, e31.

19. Cuervo, P., Domont, G. B., and De Jesus, J. B. (2010) Proteomics of trypanosomatids of human medical importance, *J Proteomics 73*, 845-867.

20. Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., and Savioli, L. (2007) Current concepts - Control of neglected tropical diseases, *New Engl J Med 357*, 1018-1027.

21. Kennedy, P. G. (2004) Human African trypanosomiasis of the CNS: current issues and challenges, *J Clin Invest 113*, 496-504.

22. Simarro, P. P., Jannin, J., and Cattand, P. (2008) Eliminating human African trypanosomiasis: where do we stand and what comes next?, *PLoS Med 5*, e55.

23. Welburn, S. C., Coleman, P. G., Maudlin, I., Fevre, E. M., Odiit, M., and Eisler, M. C. (2006) Crisis, what crisis? Control of Rhodesian sleeping sickness, *Trends Parasitol 22*, 123-128.

24. Fairlamb, A. H. (2003) Chemotherapy of human African trypanosomiasis: current and future prospects, *Trends Parasitol 19*, 488-494.

25. Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M.,

Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E., and El-Sayed, N. M. (2005) The genome of the African trypanosome Trypanosoma brucei, *Science 309*, 416-422.

26. MacGregor, P., Szoor, B., Savill, N. J., and Matthews, K. R. (2012) Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act, *Nat Rev Microbiol 10*, 431-438.

27. Steverding, D. (2010) The development of drugs for treatment of sleeping sickness: a historical review, *Parasit Vectors 3*, 15.

28. Simarro, P. P., Franco, J., Diarra, A., Postigo, J. A. R., and Jannin, J. (2012) Update on field use of the available drugs for the chemotherapy of human African trypanosomiasis, *Parasitology 139*, 842-846.

29. Jackson, A. P., Sanders, M., Berry, A., McQuillan, J., Aslett, M. A., Quail, M. A., Chukualim, B., Capewell, P., MacLeod, A., Melville, S. E., Gibson, W., Barry, J. D., Berriman, M., and Hertz-Fowler, C. (2010) The genome sequence of Trypanosoma brucei gambiense, causative agent of chronic human african trypanosomiasis, *PLoS Negl Trop Dis 4*, e658.

30. Mingler, M. K., Hingst, A. M., Clement, S. L., Yu, L. E., Reifur, L., and Koslowsky, D. J. (2006) Identification of pentatricopeptide repeat proteins in Trypanosoma brucei, *Mol Biochem Parasitol 150*, 37-45.

31. Pusnik, M., Small, I., Read, L. K., Fabbro, T., and Schneider, A. (2007) Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes, *Mol Cell Biol 27*, 6876-6888.

32. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res 25*, 3389-3402.

33. Newby, Z. E., O'Connell, J. D., 3rd, Gruswitz, F., Hays, F. A., Harries, W. E., Harwood, I. M., Ho, J. D., Lee, J. K., Savage, D. F., Miercke, L. J., and Stroud, R. M. (2009) A general protocol for the crystallization of membrane proteins for X-ray structural investigation, *Nat Protoc 4*, 619-637.

34. Eisenstein, E., Gilliland, G. L., Herzberg, O., Moult, J., Orban, J., Poljak, R. J., Banerjei, L., Richardson, D., and Howard, A. J. (2000) Biological function made crystal clear - annotation of hypothetical proteins via structural genomics, *Curr Opin Biotechnol 11*, 25-30.

35. Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C., and Szyperski, T. (2000) Protein NMR spectroscopy in structural genomics, *Nat Struct Biol 7 Suppl*, 982-985.

36. Bagby, S., Tong, K. I., and Ikura, M. (2001) Optimization of protein solubility and stability for protein nuclear magnetic resonance, *Methods Enzymol 339*, 20-41.

37. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, *Plant Cell 16*, 2089-2103.

38. Williams-Carrier, R., Kroeger, T., and Barkan, A. (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand, *RNA 14*, 1930-1941.

39. Filipovska, A., and Rackham, O. (2012) Modular recognition of nucleic acids by PUF, TALE and PPR proteins, *Molecular Biosystems 8*, 699-708.

40. Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giege, P. (2008) PPR336 is associated with polysomes in plant mitochondria, *Journal of Molecular Biology 375*, 626-636.

41. Matthews, K. R. (2005) The developmental cell biology of Trypanosoma brucei, *J Cell Sci 118*, 283-290.

42. Matthews, K. R., Ellis, J. R., and Paterou, A. (2004) Molecular regulation of the life cycle of African trypanosomes, *Trends Parasitol 20*, 40-47.

43. Gull, K. (2001) The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics, *Int J Parasitol 31*, 443-452.

44. Priest, J. W., and Hajduk, S. L. (1994) Developmental regulation of mitochondrial biogenesis in Trypanosoma brucei, *J Bioenerg Biomembr 26*, 179-191.

45. Lam, T. K., Gutierrez-Juarez, R., Pocai, A., and Rossetti, L. (2005) Regulation of blood glucose by hypothalamic pyruvate metabolism, *Science 309*, 943-947.

46. van Hellemond, J. J., Opperdoes, F. R., and Tielens, A. G. (2005) The extraordinary mitochondrion and unusual citric acid cycle in Trypanosoma brucei, *Biochem Soc Trans 33*, 967-971.

47. Cristodero, M., Seebeck, T., and Schneider, A. (2010) Mitochondrial translation is essential in bloodstream forms of Trypanosoma brucei, *Mol Microbiol 78*, 757-769.
48. Michelotti, E. F., Harris, M. E., Adler, B., Torri, A. F., and Hajduk, S. L. (1992) Trypanosoma brucei mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression, *Mol Biochem Parasitol 54*, 31-41.

49. Feagin, J. E., Jasmer, D. P., and Stuart, K. (1985) Apocytochrome b and other mitochondrial DNA sequences are differentially expressed during the life cycle of Trypanosoma brucei, *Nucleic Acids Res 13*, 4577-4596.

50. Feagin, J. E., and Stuart, K. (1985) Differential expression of mitochondrial genes between life cycle stages of Trypanosoma brucei, *Proc Natl Acad Sci U S A 82*, 3380-3384.

51. Adhya, S., Basu, S., Bhattacharyya, S. N., Chatterjee, S., Dhar, G., Goswami, S., Ghosh, S., Home, P., Mahata, B., and Tripathi, G. (2003) Mitochondrial differentiation in kinetoplastid protozoa: a plethora of RNA controls, *Differentiation 71*, 549-556.

52. van Weelden, S. W., Fast, B., Vogt, A., van der Meer, P., Saas, J., van Hellemond, J. J., Tielens, A. G., and Boshart, M. (2003) Procyclic Trypanosoma brucei do not use Krebs cycle activity for energy generation, *J Biol Chem 278*, 12854-12863.

53. Feagin, J. E., Jasmer, D. P., and Stuart, K. (1986) Differential mitochondrial gene expression between slender and stumpy bloodforms of Trypanosoma brucei, *Mol Biochem Parasitol 20*, 207-214.

54. Panigrahi, A. K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R. A., Acestor, N., Myler, P. J., and Stuart, K. D. (2009) A comprehensive analysis of Trypanosoma brucei mitochondrial proteome, *Proteomics 9*, 434-450.

55. Timms, M. W., van Deursen, F. J., Hendriks, E. F., and Matthews, K. R. (2002) Mitochondrial development during life cycle differentiation of African trypanosomes: evidence for a kinetoplast-dependent differentiation control point, *Mol Biol Cell 13*, 3747-3759.

56. Feagin, J. E. (2000) Mitochondrial genome diversity in parasites, *Int J Parasitol 30*, 371-390.

57. Hancock, K., and Hajduk, S. L. (1990) The mitochondrial tRNAs of Trypanosoma brucei are nuclear encoded, *J Biol Chem 265*, 19208-19215.

58. Hong, M., and Simpson, L. (2003) Genomic organization of Trypanosoma brucei kinetoplast DNA minicircles, *Protist 154*, 265-279.

59. Carpenter, L. R., and Englund, P. T. (1995) Kinetoplast Maxicircle DNA-Replication in Crithidia-Fasciculata and Trypanosoma-Brucei, *Mol Cell Biol 15*, 6794-6803.

60. Sloof, P., de Haan, A., Eier, W., van Iersel, M., Boel, E., van Steeg, H., and Benne, R. (1992) The nucleotide sequence of the variable region in Trypanosoma brucei completes the sequence analysis of the maxicircle component of mitochondrial kinetoplast DNA, *Mol Biochem Parasitol 56*, 289-299.

61. Stuart, K. D., Schnaufer, A., Ernst, N. L., and Panigrahi, A. K. (2005) Complex management: RNA editing in trypanosomes, *Trends Biochem Sci 30*, 97-105.

62. Jasmer, D. P., Feagin, J. E., Payne, M., and Stuart, K. (1987) Variation of G-rich mitochondrial transcripts among stocks of Trypanosoma brucei, *Mol Biochem Parasitol 22*, 259-272.

63. de Vries, B. F., Mulder, E., Brakenhoff, J. P., Sloof, P., and Benne, R. (1988) The variable region of the Trypanosoma brucei kinetoplast maxicircle: sequence and transcript analysis of a repetitive and a non-repetitive fragment, *Mol Biochem Parasitol 27*, 71-82.

64. Myler, P. J., Glick, D., Feagin, J. E., Morales, T. H., and Stuart, K. D. (1993) Structural organization of the maxicircle variable region of Trypanosoma brucei: identification of potential replication origins and topoisomerase II binding sites, *Nucleic Acids Res 21*, 687-694.

65. Ntambi, J. M., and Englund, P. T. (1985) A gap at a unique location in newly replicated kinetoplast DNA minicircles from Trypanosoma equiperdum, *J Biol Chem 260*, 5574-5579.

66. Ntambi, J. M., Shapiro, T. A., Ryan, K. A., and Englund, P. T. (1986) Ribonucleotides associated with a gap in newly replicated kinetoplast DNA minicircles from Trypanosoma equiperdum, *J Biol Chem 261*, 11890-11895.

67. Ray, D. S. (1989) Conserved sequence blocks in kinetoplast minicircles from diverse species of trypanosomes, *Mol Cell Biol 9*, 1365-1367.

68. Aphasizhev, R., and Aphasizheva, I. (2011) Mitochondrial RNA processing in trypanosomes, *Res Microbiol 162*, 655-663.

69. Zikova, A., Kopecna, J., Schumacher, M. A., Stuart, K., Trantirek, L., and Lukes, J. (2008) Structure and function of the native and recombinant mitochondrial MRP1/MRP2 complex from Trypanosoma brucei, *Int J Parasitol 38*, 901-912.

70. Grams, J., McManus, M. T., and Hajduk, S. L. (2000) Processing of polycistronic guide RNAs is associated with RNA editing complexes in Trypanosoma brucei, *EMBO J 19*, 5525-5532.

71. Yatawara, L., Le, T. H., Wickramasinghe, S., and Agatsuma, T. (2008) Maxicircle (mitochondrial) genome sequence (partial) of Leishmania major: gene content, arrangement and composition compared with Leishmania tarentolae, *Gene 424*, 80-86.

72. Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K., and Myler, P. J. (2003) Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region, *Mol Cell 11*, 1291-1299.

73. de la Cruz, V. F., Simpson, A. M., Lake, J. A., and Simpson, L. (1985) Primary sequence and partial secondary structure of the 12S kinetoplast (mitochondrial) ribosomal RNA from Leishmania tarentolae: conservation of peptidyl-transferase structural elements, *Nucleic Acids Res 13*, 2337-2356.

74. Eperon, I. C., Janssen, J. W., Hoeijmakers, J. H., and Borst, P. (1983) The major transcripts of the kinetoplast DNA of Trypanosoma brucei are very small ribosomal RNAs, *Nucleic Acids Res 11*, 105-125.

75. Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R., and Noller, H. (1985) Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: a detailed comparison of the 9S and 12S RNAs from Crithidia fasciculata and Trypanosoma brucei with rRNAs from other organisms, *Nucleic Acids Res 13*, 4171-4190.

76. Pollard, V. W., Rohrer, S. P., Michelotti, E. F., Hancock, K., and Hajduk, S. L. (1990) Organization of minicircle genes for guide RNAs in Trypanosoma brucei, *Cell 63*, 783-790.

77. Ntambi, J. M., Marini, J. C., Bangs, J. D., Hajduk, S. L., Jimenez, H. E., Kitchin, P. A., Klein, V. A., Ryan, K. A., and Englund, P. T. (1984) Presence of a bent helix in fragments of kinetoplast DNA minicircles from several trypanosomatid species, *Mol Biochem Parasitol 12*, 273-286.

78. Campbell, D. A., Thomas, S., and Sturm, N. R. (2003) Transcription in kinetoplastid protozoa: why be normal?, *Microbes Infect 5*, 1231-1240.

79. Madina, B. R., Kuppan, G., Vashisht, A. A., Liang, Y. H., Downey, K. M., Wohlschlegel, J. A., Ji, X., Sze, S. H., Sacchettini, J. C., Read, L. K., and Cruz-Reyes, J. (2011) Guide RNA biogenesis involves a novel RNase III family endoribonuclease in Trypanosoma brucei, *RNA 17*, 1821-1830.

80. Grams, J., Morris, J. C., Drew, M. E., Wang, Z., Englund, P. T., and Hajduk, S. L. (2002) A trypanosome mitochondrial RNA polymerase is required for transcription and replication, *J Biol Chem 277*, 16952-16959.

81. Hashimi, H., Cicova, Z., Novotna, L., Wen, Y. Z., and Lukes, J. (2009) Kinetoplastid guide RNA biogenesis is dependent on subunits of the mitochondrial RNA binding complex 1 and mitochondrial RNA polymerase, *RNA 15*, 588-599.

82. Aphasizhev, R., and Aphasizheva, I. (2011) Uridine insertion/deletion editing in trypanosomes: a playground for RNA-guided information transfer, *Wiley Interdiscip Rev RNA 2*, 669-685.

83. Benne, R., Van den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H., and Tromp, M. C. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA, *Cell 46*, 819-826.

84. Brennicke, A., Marchfelder, A., and Binder, S. (1999) RNA editing, *FEMS Microbiol Rev 23*, 297-316.

85. Hashimi, H., Zimmer, S. L., Ammerman, M. L., Read, L. K., and Lukes, J. (2013) Dual core processing: MRB1 is an emerging kinetoplast RNA editing complex, *Trends Parasitol 29*, 91-99.

86. Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes, *Mol Cell 42*, 106-117.

87. Landweber, L. F. (1992) The evolution of RNA editing in kinetoplastid protozoa, *Biosystems 28*, 41-45.

88. Stuart, K. (1989) Trypanosomatids: mitochondrial RNA editing, *Exp Parasitol 68*, 486-490.

89. Maslov, D. A. (2010) Complete set of mitochondrial pan-edited mRNAs in Leishmania mexicana amazonensis LV78, *Mol Biochem Parasitol 173*, 107-114.

90. Maslov, D. A., and Simpson, L. (1992) The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing, *Cell 70*, 459-467.

91. Bohm, C., Katari, V. S., Brecht, M., and Goringer, H. U. (2012) Trypanosoma brucei 20 S editosomes have one RNA substrate-binding site and execute RNA unwinding activity, *J Biol Chem 287*, 26268-26277.

92. Carnes, J., Trotter, J. R., Peltan, A., Fleck, M., and Stuart, K. (2008) RNA editing in Trypanosoma brucei requires three different editosomes, *Mol Cell Biol 28*, 122-130.

93. Panigrahi, A. K., Ernst, N. L., Domingo, G. J., Fleck, M., Salavati, R., and Stuart, K. D. (2006) Compositionally and functionally distinct editosomes in Trypanosoma brucei, *RNA 12*, 1038-1049.

94. Weng, J., Aphasizheva, I., Etheridge, R. D., Huang, L., Wang, X. R., Falick, A. M., and Aphasizhev, R. (2008) Guide RNA-Binding Complex from Mitochondria of Trypanosomatids, *Mol Cell 32*, 198-209.

95. Madison-Antenucci, S., and Hajduk, S. L. (2001) RNA editing-associated protein 1 is an RNA binding protein with specificity for preedited mRNA, *Mol Cell 7*, 879-886.

96. Aphasizhev, R., Aphasizheva, I., Nelson, R. E., and Simpson, L. (2003) A 100-kD complex of two RNA-binding proteins from mitochondria of Leishmania tarentolae catalyzes RNA annealing and interacts with several RNA editing components, *RNA 9*, 62-76.

97. Hans, J., Hajduk, S. L., and Madison-Antenucci, S. (2007) RNA-editing-associated protein 1 null mutant reveals link to mitochondrial RNA stability, *RNA 13*, 881-889.

98. Vanhamme, L., Perez-Morga, D., Marchal, C., Speijer, D., Lambert, L., Geuskens, M., Alexandre, S., Ismaili, N., Goringer, U., Benne, R., and Pays, E. (1998) Trypanosoma brucei TBRGG1, a mitochondrial oligo(U)-binding protein that co-localizes with an in vitro RNA editing activity, *J Biol Chem 273*, 21825-21833.

99. Fisk, J. C., Ammerman, M. L., Presnyak, V., and Read, L. K. (2008) TbRGG2, an essential RNA editing accessory factor in two Trypanosoma brucei life cycle stages, *J Biol Chem 283*, 23016-23025.

100. Ammerman, M. L., Downey, K. M., Hashimi, H., Fisk, J. C., Tomasello, D. L., Faktorova, D., Kafkova, L., King, T., Lukes, J., and Read, L. K. (2012) Architecture of the trypanosome RNA editing accessory complex, MRB1, *Nucleic Acids Res 40*, 5637-5650.

101. Hashimi, H., Zikova, A., Panigrahi, A. K., Stuart, K. D., and Lukes, J. (2008) TbRGG1, an essential protein involved in kinetoplastid RNA metabolism that is associated with a novel multiprotein complex, *RNA 14*, 970-980.

102. Panigrahi, A. K., Zikova, A., Dalley, R. A., Acestor, N., Ogata, Y., Anupama, A., Myler, P. J., and Stuart, K. D. (2008) Mitochondrial complexes in Trypanosoma brucei: a novel complex and a unique oxidoreductase complex, *Mol Cell Proteomics 7*, 534-545.

103. Delannoy, E., Stanley, W.A., Bond, C.S., Small, I.D. (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specific factors in post-transcriptional processes in organelles, *Biochemical Society Transactions*, 1643-1647.

104. Schmitz-Linneweber, C., and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression, *Trends Plant Sci 13*, 663-670.

105. Ammerman, M. L., Fisk, J. C., and Read, L. K. (2008) gRNA/pre-mRNA annealing and RNA chaperone activities of RBP16, *RNA 14*, 1069-1080.

106. Pelletier, M., and Read, L. K. (2003) RBP16 is a multifunctional gene regulatory protein involved in editing and stabilization of specific mitochondrial mRNAs in Trypanosoma brucei, *RNA 9*, 457-468.

107. Aphasizhev, R., Aphasizheva, I., and Simpson, L. (2003) A tale of two TUTases, *Proc Natl Acad Sci U S A 100*, 10617-10622.

108. Cruz-Reyes, J., Zhelonkina, A., Rusche, L., and Sollner-Webb, B. (2001) Trypanosome RNA editing: simple guide RNA features enhance U deletion 100-fold, *Mol Cell Biol 21*, 884-892.

109. Kao, C. Y., and Read, L. K. (2005) Opposing effects of polyadenylation on the stability of edited and unedited mitochondrial RNAs in Trypanosoma brucei, *Mol Cell Biol 25*, 1634-1644.

110. Bhat, G. J., Souza, A. E., Feagin, J. E., and Stuart, K. (1992) Transcript-specific developmental regulation of polyadenylation in Trypanosoma brucei mitochondria, *Mol Biochem Parasitol 52*, 231-240.

111. Etheridge, R. D., Aphasizheva, I., Gershon, P. D., and Aphasizhev, R. (2008) 3' adenylation determines mRNA abundance and monitors completion of RNA editing in T. brucei mitochondria, *EMBO J 27*, 1596-1608.

112. Aphasizheva, I., and Aphasizhev, R. (2010) RET1-catalyzed uridylylation shapes the mitochondrial transcriptome in Trypanosoma brucei, *Mol Cell Biol 30*, 1555-1567.

113. Adler, B. K., Harris, M. E., Bertrand, K. I., and Hajduk, S. L. (1991) Modification of Trypanosoma brucei mitochondrial rRNA by posttranscriptional 3' polyuridine tail formation, *Mol Cell Biol 11*, 5878-5884.

114. Koslowsky, D. J., and Yahampath, G. (1997) Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in Trypanosoma brucei are independent events, *Mol Biochem Parasitol 90*, 81-94.

115. Ryan, C. M., Militello, K. T., and Read, L. K. (2003) Polyadenylation regulates the stability of Trypanosoma brucei mitochondrial RNAs, *J Biol Chem 278*, 32753-32762.

116. Zikova, A., Panigrahi, A. K., Dalley, R. A., Acestor, N., Anupama, A., Ogata, Y., Myler, P. J., and Stuart, K. (2008) Trypanosoma brucei mitochondrial ribosomes: affinity

purification and component identification by mass spectrometry, *Mol Cell Proteomics 7*, 1286-1296.

117. Maslov, A. M., and Agrawal, R. K. (2012) Mitochondrial translation in trypanosomatids, Nucleic Acids and Molecular Biology, In *RNA Metabolism in Trypanosomes* (Bindereif, A., Ed.), Springer-Verlag, Berlin.

118. Small, I. D., and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins, *Trends Biochem Sci 25*, 46-47.

119. Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y. S., Bond, C. S., and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins, *PLoS Genet 8*, e1002910.

120. D'Andrea, L. D., and Regan, L. (2003) TPR proteins: the versatile helix, *Trends Biochem Sci 28*, 655-662.

121. Blatch, G. L., and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions, *Bioessays 21*, 932-939.

122. Edwards, T. A., Pyle, S. E., Wharton, R. P., and Aggarwal, A. K. (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs, *Cell 105*, 281-289.

123. Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts, *EMBO J 28*, 2042-2052.

124. Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts, *Mol Cell Biol 28*, 5337-5347.

125. Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein, *P Natl Acad Sci USA 108*, 415-420.

126. Pusnik, M., and Schneider, A. (2012) A trypanosomal pentatricopeptide repeat protein stabilizes the mitochondrial mRNAs of cytochrome oxidase subunits 1 and 2, *Eukaryot Cell 11*, 79-87.

127. Khrouchtchova, A., Monde, R. A., and Barkan, A. (2012) A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts, *RNA 18*, 1197-1209.

128. Kroeger, T. S., Watkins, K. P., Friso, G., van Wijk, K. J., and Barkan, A. (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing, *Proc Natl Acad Sci U S A 106*, 4537-4542.

129. Kotera, E., Tasaka, M., and Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts, *Nature 433*, 326-330.

130. Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing, *Proc Natl Acad Sci U S A 104*, 8178-8183.

131. Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing, *Journal of Biological Chemistry 281*, 37661-37667.

132. Okuda, K., and Shikanai, T. (2012) A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids, *Nucleic Acids Research 40*, 5052-5064.

133. Ruzzenente, B., Metodiev, M. D., Wredenberg, A., Bratic, A., Park, C. B., Camara, Y., Milenkovic, D., Zickermann, V., Wibom, R., Hultenby, K., Erdjument-Bromage, H., Tempst, P., Brandt, U., Stewart, J. B., Gustafsson, C. M., and Larsson, N. G. (2012) LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs, *EMBO J 31*, 443-456.

134. Tavares-Carreon, F., Camacho-Villasana, Y., Zamudio-Ochoa, A., Shingu-Vazquez, M., Torres-Larios, A., and Perez-Martinez, X. (2008) The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial COX1 mRNA in yeast, *J Biol Chem 283*, 1472-1479.

135. Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R. (2012) An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria, *Plant Cell*.

136. Rackham, O., Davies, S. M. K., Shearwood, A. M. J., Hamilton, K. L., Whelan, J., and Filipovska, A. (2009) Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells, *Nucleic Acids Research 37*, 5859-5867.

137. Nakamura, T., and Sugita, M. (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity, *FEBS Lett 582*, 4163-4168.

138. Okuda, K., Chateigner-Boutin, A. L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts, *Plant Cell 21*, 146-156.

139. Cushing, D. A., Forsthoefel, N. R., Gestaut, D. R., and Vernon, D. M. (2005) Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis, *Planta 221*, 424-436.

140. Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T., and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants, *PLoS One 8*, e57286.

141. Rackham, O., and Filipovska, A. (2012) The role of mammalian PPR domain proteins in the regulation of mitochondrial gene expression, *Biochim Biophys Acta 1819*, 1008-1016.

142. Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P., and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase, *Nature 478*, 269-273.

143. Sasarman, F., Brunel-Guitton, C., Antonicka, H., Wai, T., and Shoubridge, E. A. (2010) LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria, *Mol Biol Cell 21*, 1315-1323.

144. Davies, S. M. K., Rackham, O., Shearwood, A. M. J., Hamilton, K. L., Narsai, R., Whelan, J., and Filipovska, A. (2009) Pentatricopeptide repeat domain protein 3 associates with the mitochondrial small ribosomal subunit and regulates translation, *Febs Letters 583*, 1853-1858.

145. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing, *Proc Natl Acad Sci U S A 109*, 16149-16154.

146. Gupta, Y. K., Nair, D. T., Wharton, R. P., and Aggarwal, A. K. (2008) Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity, *Structure 16*, 549-557.

147. Wang, X. Q., McLachlan, J., Zamore, P. D., and Hall, T. M. T. (2002) Modular recognition of RNA by a human pumilio-homology domain, *Cell 110*, 501-512.

148. Hammani, K., Cook, W. B., and Barkan, A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression, *P Natl Acad Sci USA 109*, 5651-5656.

149. Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., and Nakamura, T. (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein, *Nucleic Acids Research 40*, 2712-2723.

150. Fujii, S., Bond, C. S., and Small, I. D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution, *P Natl Acad Sci USA 108*, 1723-1728.

CHAPTER 2

ANALYSIS OF THE RNA BINDING SPECIFICITY OF THE 27 kDa *TRYPANOSOMA BRUCEI* PENTATRICOPEPTIDE REPEAT PROTEIN

A manuscript [Pakoyo Kamba, David Dickson, Neil White, Jennifer Ekstrom, Donna Koslowsky, Charles Hoogstraten . Mitochondrial RNA precursor transcripts and guanine rich motifs are the targets for the 27 kDa *Trypanosoma brucei* pentatricopeptide repeat protein] based on data in this chapter is being finalized for submission to *Nucleic Acids Research*.

David Dickson, a former Professorial Assistant and undergraduate research aide in our lab collected the fluorescence anisotropy data for G_{12} ssRNA and G_{12} ssDNA under my mentorship using protein prepared by me. Assistance in the growth of *T. brucei* parasites and guidance in mitochondrial preparation and RNA extraction were received from Dr. Donna Koslowsky. Primers used in the pull-down assays were also a gift from Dr. Donna Koslowsky. All other experiments were performed solely by me.

ABSTRACT

Pentatricopeptide repeat (PPR) proteins, a helical repeat family of organellar RNA binding proteins, play essential roles in post-transcriptional RNA processing. In *Trypanosoma brucei* (*T. brucei*), these proteins localize to its single mitochondrion (kinetoplast) where they perform yet to be elucidated roles in RNA processing. We studied the smallest *T. brucei* PPR protein with a molecular mass of 27 kDa (PPR27). Besides a recent report that PPR27 associates via protein-protein interactions to the mitochondrial small ribosomal subunit, neither its RNA binding specificity nor its biochemical function is known. We show here that PPR27 selectively binds nonhelical polyguanosine RNA over other sequences and nucleic acid forms, and that polyguanosine quadruplexes are preferentially recognized over random coil RNA forms. PPR27 binding did not disrupt the quadruplex. A search of the *T. brucei* mitochondrial genome found selective enrichment of guanine tracts in maxicircle genes for extensively edited mRNAs whereas pull-down assays with total kinetoplast RNA identified uncleaved RNA precursors as high affinity ligands. Hence PPR27 may function in crosstalk between pre-mRNA processing and translation. Transition from G_9 to G_6 ssRNA barely increases the K_d but drastically destabilizes the protein-RNA complex under non-equilibrium conditions. Such differential ribonucleoprotein stability could aid PPR27 discrimination among different G-tracts.

2.1 INTRODUCTION

First discovered in *Arabidopsis thaliana* by Small and Peeters in 2000 (*1*), the PPR family is a group of sequence specific RNA binding proteins characterized by tandem repeats of 35 amino acids. Each PPR motif folds into a pair of antiparallel alpha-helices similar to the peptide-binding tetratricopeptide (34 amino acid) repeat (TPR) motif (*1-2*). Over the last decade, these proteins have been linked to all stages of post-transcriptional RNA processing in chloroplasts and mitochondria where they localize [reviewed in (*3-5*)]. Typically, plants contain hundreds of PPR proteins while non-plant eukaryotes contain less than 10 (*6*), with the exception of trypanosomes which have a fairly raised number. At least 36 PPR proteins have been reported in *T. brucei* (*7-8*), the cause of the severely debilitating and fatal human African trypanosomiasis (HAT) or sleeping sickness. PPR proteins are highly conserved across trypanosomatids and RNAi knockdown causes parasite growth retardation, growth arrest, and death (*9-10*), consistent with functional essentiality. Underlying the severe phenotypes is deterioration in oxidative phosphorylation, and hence mitochondrial function (*8-10*).

The mitochondrion is central to the life cycle and pathogenesis of *T. brucei* as modulation in its activity facilitates parasite adaptation from the sugar-deficient tsetse fly gut, where it depends on oxidative phosphorylation, to the sugar-rich environment in human blood, where it depends on glycolysis for energy (*11-16*). Nevertheless, the parasite lacks capacity for transcriptional control, hence most regulation of mitochondrial gene expression occurs at the RNA level (*15, 17-21*). Deciphering the function of RNA binding proteins such as the PPR family is therefore important to understanding critical *T. brucei* biology. Biochemical and biophysical studies are particularly crucial to understanding what these proteins do and how they do it. Poor protein solubility and difficulty in heterologous expression (*6, 22*) have however limited such studies,

and neither the sequence specificity of RNA ligand(s) nor the mode by which these proteins bind RNA is known.

In this report, we demonstrate the specificity of RNA binding by the 27 kDa *T. brucei* PPR protein (CDS: Tb927.8.6040, accession number: XM_842341). Though PPR27 was earlier thought to be a cytoplasmic protein due to negative mitochondrial targeting by classical computational tools (*10*), there is now significant evidence to show that it is mitochondrial: (i) contemporary computational tools support mitochondrial localization (**Table 2.1**); (ii) stable RNA-independent association with mitochondrial small subunit ribosomal complexes has been experimentally demonstrated by others (*7, 23*). We now report that PPR27 has preferential affinity for uncleaved precursor transcripts and RNA motifs selectively enriched in pre-edited mRNA substrates, consistent with involvement in pre-mRNA processing. We also show that PPR27 preferentially binds G-tracts in the quadruplex conformation over those in the random coil state. Finally, we report the thermodynamics of PPR27 interaction with different G-tracts and discuss their implications for PPR27 function.

2.2 MATERIALS AND METHODS

2.2.1 Materials

RNA labeled with either FLUO (fluorescein) at its 5' end or TAMRA (tetramethylrhodamine) at its 3' end was supplied by Thermo Scientific's Dharmacon RNA Technologies (Lafayette, CO) and deprotected according to the manufacturer's instructions. Dry RNA was dissolved in milli-Q water (MQ-H_2O). 5'-FLUO labeled dG_{12} ssDNA was obtained from Integrated DNA Technologies (CoralVille, IA). For cloning and site-directed mutagenesis, desalted, dry DNA primers were obtained from the Michigan State University Macromolecular Structure Facility

(East Lansing, MI), Sigma-Aldrich (St. Louis, MO), and Integrated DNA Technologies and reconstituted in MQ-H_2O. His_6-tagged thioredoxin was expressed from a plasmid created by introduction of a TGA stop codon just before PPR27 in a PPR27pET32Xa/LIC plasmid. The pMalTEV-E30 plasmid vector was a kind donation from Dr. Alice Barkan (University of Oregon). TEV protease was made by IPTG induction and nickel nitrilotriacetic acid (Ni-NTA) (Qiagen GmbH, Hilden, Germany) chromatography from recombinant BL21 (DE3)/RIL *E. coli* kindly donated by Dr.Honggao Yan (Michigan State University). Other reagents and their suppliers in parentheses are: pET32Xa/LIC vector and factor Xa (EMD4Biosciences, Gibbstown, NJ), amylose resin, amylose magnetic beads, BamH1 and SalI restriction enzymes, and T4 DNA ligase (New England BioLabs, Ipswich, MA), 2X PCR Master Mix and recombinant RNasin® (Promega Corporation, Madison, WI), QuikChange® mutagenesis kit (Agilent Technologies, Santa Clara, CA), M-MLV reverse transcriptase and TOPO® TA plasmid (Invitrogen, Grand Island, NY), DNase I and bovine serum albumin (Roche Applied Science, Indianapolis, IN), 2X PCR Master Mix (Syzygy Biotech, Grand Rapids, MI), n-dodecyl-β,D-maltopyranoside, DDM (Anatrace-Affymetrix, Maumee, OH), and Ni-NTA magnetic beads (Qiagen GmbH, Hilden, Germany).

2.2.2 Cloning and site-directed mutagenesis

During heterologous expression, organellar targeting signal peptides can be toxic to *E. coli*. Since targeting signals are cleaved off during shuttling of the polypeptide from the cytoplasm into the organelle, that is, they are not part of mature proteins, the recombinant plasmid for heterologous expression of PPR27 was deprived of the mitochondrial targeting signal (MTS) peptide at its N-terminus. Because PPR27 lacks a classical MTS, the signal peptide was

estimated from predicted repeat organization and secondary structure (details presented below under Results). PPR motif prediction was done by both TPRpred (*24*) (http://toolkit.tuebingen.mpg.de/tprpred/), a classical PPR motif prediction tool, and HHrepID (*25*) (http://toolkit.tuebingen.mpg.de/hhrepid/), a de novo repeat prediction program. Secondary structure prediction was done using PSIPRED (*26*) (http://bioinf.cs.ucl.ac.uk/psipred/).

The DNA for mature PPR27 (lacking the MTS) was initially cloned into a pRSFDuet-1 plasmid (courtesy of Dr. Jennifer Ekstrom, formerly of Michigan State University) to form a construct with a hexahistidine (His_6) tag at the N-terminus. The yield of recombinant PPR27 using this system was however negligible. A colony viability assay upon induction of protein expression by IPTG (isopropyl-β-D-1-thiogalactopyranoside) (*27*) found that mature PPR27 was still toxic to *E. coli* (**Figure 2.3**), which could be, though not exclusively, due to poor solubility. To aid soluble expression, we fused PPR27 to the C-terminus of the solubility- and folding enhancer, maltose binding protein (MBP) from *E. coli*. Briefly, PPR27 was PCR (polymerase chain reaction) amplified from the PPR27pRSFDuet-1 plasmid using primers CGH30 and CGH34 (**Table 2.2**). The PCR amplicon was then subcloned into the pMalTEV-E30 plasmid developed by the Barkan group using BamH1 and SalI restriction enzymes. The ligation mixture of the vector and PPR27 was propagated in DH5α *E. coli* followed by isolation of plasmid DNA using the QIAprep® spin miniprep (Qiagen Sciences, Germantown, MD). The presence of PPR27 was confirmed by DNA sequencing using dideoxynucleotide chain termination at the Michigan State University Genomics Core. The pMalTEV-E30 vector encodes a TEV protease cleavage site just before the BamH1 restriction site, enabling detachment of MBP from PPR27 where necessary.

During expression of MBPPPR27, there was significant leaky expression of MBP, which could not be suppressed by supplementation of growth medium with 0.05% (w/v) glucose (*28*), meaning amylose affinity chromatography alone was insufficient for protein purification. Thus, we inserted a His_6 tag at the C-terminus of PPR27 (**Figure 2.4B)** by site-directed mutagenesis (SDM) to aid tandem divalent metal ion- and amylose affinity chromatography. Assays with $MBPPPR27His_6$ required $MBPHis_6$ as negative control and this was created by inserting His_6 at the C-terminus of MBP just before the TEV protease cleavage site (**Figure 2.4C**) followed by conversion of the first residue of PPR27 in $MBPHis_6PPR27$ to TGA stop codon by SDM. SDM was done by QuikChange® (Agilent Technologies, Santa Clara, CA) using primers designed with the QuikChange® Primer Design program (**Table 2.2**). First, 50 µl PCR reactions were set up according to the QuikChange® protocol (Agilent Technologies). Mutagenesis PCR conditions were one round of initial denaturation (95 ^{o}C for 30 s), 18 cycles of DNA amplification (denaturation at 95 ^{o}C for 30s followed by annealing at 55 ^{o}C for 1 min and extension at 68 ^{o}C for 8 min), and a final round of extension (68 ^{o}C for 8 min). PCR amplicons were digested by DpnI endonuclease, after which 1.0 µl was transformed into either XL1-Blue (Agilent Technologies) or NovaBlue GigaSingles™ (EMD4Biosciences) *E. coli* by heat shock. The presence of desired mutations in the isolated plasmids was verified by DNA sequencing.

2.2.3 Protein expression

The PPR27pMalTEV-E30 plasmid was transformed into BL21 (DE3) *E. coli* by heat shock. Recombinant colonies were selected on LB agar containing 50 µg/ml ampicillin. Thereafter,

single colonies were inoculated into 10 ml LB broth and grown overnight at 37 ^{o}C and 250 rpm to produce seed culture. 20 ml of seed culture was then added to 980 ml of LB broth and grown at 37 ^{o}C, 250 rpm till an OD_{600} of about 0.7 when protein overexpression was induced with 0.1 mM isopropyl-β,D-thiogalactopyranoside (IPTG) at 22 ^{o}C for about 6 hours. Final cultures were harvested by centrifugation at 5000 rpm, 4 ^{o}C, for 15 minutes. Cells were then either immediately lysed for purification or stored at -80 ^{o}C.

2.2.4 Protein purification

MBPPPR27 was purified from *E. coli* under native conditions by tandem affinity chromatography. Briefly, bacteria were resuspended in at least 5 ml of lysis buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-ME) per gram of cells, lysed by a micro-tip sonicator (Heat Systems Ultrasonics, Farmingdale, NJ), and clarified by centrifugation at 14,000 rpm, 4 ^{o}C, for 25 minutes. The supernatant was loaded onto a Ni-NTA column and MBPPPR27His_6 eluted by 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 200 mM imidazole, 5 mM β-ME. Fractions containing MBPPPR27His_6 were immediately further purified by amylose agarose column according to the manufacturer's protocols. Eluates were dialyzed against at least 40 volumes of 10 mM Tris-HCl, pH 7.5, 50 mM NaCl at 4 ^{o}C for at least 12 hours with a 7K MWCO SnakeSkin® tubing (Thermo Scientific, Rockford, IL). Dithiothreitol (DTT) was then added to the dialysates to 1 mM followed by concentration using 10K MWCO amicon® (Millipore, Cork, Ireland) centrifugal filters. Protein concentration was derived from the

absorbance of ultraviolet (U.V) light at 280 nm (A_{280}) and the molar extinction coefficient estimated using Richard's Protein Calculator version 1.1 (http://www.mrc-lmb.cam.ac.uk/ms/methods/proteincalculator.html). Purified proteins were assessed for protein contaminants by 5-12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and nucleic acid contamination by the ratio of U.V light absorbance at 260 nm and 280 nm (A_{260}/A_{280}). Both the A_{280} and A_{260} were from a NanoDrop® (Thermo Scientific) spectrophotometer.

2.2.5 PPR27 native gel electrophoresis

Native gel electrophoresis was done on the TEV protease reaction products of MBPPPR27. MBPPPR27 was purified by tandem Ni-NTA and amylose. Dilute protein was then concentrated to 0.7 mg/ml using 10K MWCO amicon centrifugal filters, followed by TEV protease cleavage in the presence of 50 mM Tris-HCl, 0.5 mM EDTA, 1.0 mM DTT, plus or minus 0.05% (w/v) DDM. The reaction was done at 8 ^{o}C for 13 hours. 1.6 µl of 10x blue native Tris-glycine (BNTG) sample buffer (250 mM Tris, pH 7.0, 30% (w/v) sucrose, 5% (w/v) brilliant blue-G) was then added to 14.4 µl of TEV protease cleavage mixture and run on a 6-12% polyacrylamide gradient resolving gel with a 3.5% polyacrylamide stacking gel at 150 V constant voltage, 4 ^{o}C, for one hour. 25 mM Tris, 250 mM glycine was used as running buffer. Native TEV protease (27 kDa), thioredoxin-His_6 (17 kDa), MBP (45 kDa), MBPPPR27 (69 kDa), and factor Xa protease (43 kDa) were the size markers.

2.2.6 Electrophoretic mobility shift assays (EMSA)

5'-FLUO labeled RNA and DNA oligonucleotides were used. The sequence specificity of RNA ligand was first assessed on 12-mer single stranded RNA (ssRNA) homopolymers. Only G_{12} RNA bound; EMSAs were therefore done on G_{12} single stranded DNA (ssDNA) and $(G.C)_{12}$ double stranded (ds) RNA to assess the chemical specificity of the ligand. $(G.C)_{12}$ was prepared by heating a mixture of G_{12} ssRNA, a three-fold excess of C_{12} ssRNA, and a 5x annealing buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 5 mM EDTA) on a boiling water bath for two minutes followed by slow cooling of the water bath to room temperature for one hour. Finally, EMSAs were done on six shorter G-tracts (G_9, G_6, $G_6C_2U_2A_2$, G_4, $G_4C_2U_3A_3$, and $(GGU)_4$) to establish the minimum size of poly(G) motif required for efficient recognition by PPR27. Agarose, rather polyacrylamide gels were used because the latter were associated with strong well retention of the RNA as the protein concentration rose. Because agarose gels are thick, and due to the propensity of poly(G) RNA/DNA to quench the fluorophore, 150 nM RNA/DNA was required in the reaction mixture for optimal signal in the gel images. The RNA binding reaction consisted of 2.5 µl of RNA/DNA, 7.5 µl of protein (either MBPPPR27His_6 or MBPHis_6), and 2.5 µl of a 5x RNA binding buffer adapted from Stuart and coworkers (100 mM Tris-HCl, pH 7.5, 750 mM KCl, 25 mM $MgCl_2$, 5.0 mM DTT, 0.5 µg/µl BSA, 2 U/µl RNasin) (*29*). 25 mM Tris, 250 mM glycine was used as running buffer. Gels were scanned using a VersaDoc™ MP 4000 fluorescence imager (Bio-Rad Laboratories).

2.2.7 Fluorescence polarization spectroscopy

The anisotropy of 20 nM 5'-FLUO labeled ssRNA and ssDNA was assessed in the presence of increasing PPR27 concentration. A 450 µl reaction mixture for each protein concentration was made from dialysis buffer, MBPPPR27His$_6$, 45 µl of a 10x binding buffer (200 mM Tris-HCl, pH 7.5, 1.5 M KCl, 50 mM $MgCl_2$, 10 mM DTT), and 50 µl of 180 nM RNA/DNA. After at least 15 minutes of equilibration at room temperature, the intensities of vertically- and horizontally polarized fluorescence emission following vertically polarized excitation were obtained at 25 ^{o}C using a microsquare quartz, open top, 5 mm sample cell (Starna Cells, Atascadero, CA) and a QuantaMaster™ spectrofluorometer (Photon Technology International, Birmingham, NJ). Noise from Raman scattering was removed by subtracting the fluorescence emission of PPR27 dialysis buffer from data. We also measured the vertically- and horizontally polarized fluorescence following horizontally polarized excitation and used it to compute the grating factor (G). The slit width of excitation monochromator was 2.0 mm while that of emission was 3.8 mm. Samples were excited by 450 nm wavelength radiation, yielding emission peaks at 520 nm. Fluorescence anisotropy (r) was computed from equation (ii), where I_{vv} and I_{vh} are, respectively, the vertically- and horizontally polarized emissions after vertically polarized excitation.

$$r = \frac{I_{vv} - GI_{vh}}{I_{vv} + 2GI_{vh}} \quad \text{---------- (i)}$$

$$f(c) = \frac{nc}{K_d + c}$$ ---------- (ii)

The mean anisotropies of at least three data sets were then normalized and used to determine the dissociation constant (K_d) by nonlinear fitting with IGOR Pro 6 using equation (ii), where c is the protein concentration and n is the apparent stoichiometry of the interaction. RNA for anisotropy experiments was thawed slowly on ice-cold water as we noticed stronger and more specific interaction of PPR27 with slowly thawed G-tract RNA than with both rapidly thawed- and thermally treated RNA.

2.2.8 Analysis of G-tract content in *T. brucei* mitochondrial transcripts

Sequences of all the kinetoplast pre-mRNA and rRNA genes were decoded from the *T. brucei* maxicircle DNA sequence (GenBank accession # M94286.1). Edited mRNA sequences were mined from literature (*30*). Guide RNA (gRNA) sequences were decoded from available *T. brucei* minicircle sequences in GenBank. A census of nonoverlapping dodecanucleotides containing at least one homopolymeric stretch of four or more Gs in each RNA sequence was then made, whilst ensuring that the concentration of G in the dodecanucleotide was maximized. Next, a count of each G_n homopolymer (n = 4, 5, 6, 7, 8, or 11) in each transcript category (pre-edited form of mildly edited mRNA, pre-edited form of extensively edited mRNA, edited mRNA, never-edited mRNA, rRNA, and gRNA) was made. The percentage of Gs in each dodecanucleotide was computed as well. Finally, the frequency of each G_n homopolymer was divided by the number of genes in each transcript category to get an enrichment index of a given G_n motif.

2.2.9 Circular dichroism (CD) spectroscopy of guanine tract RNA

FA assays showed reduced PPR27 affinity for G-tract RNA relative to native RNA (see Results). Single stranded G-tract nucleic acids tend to exist as advanced structures called quadruplexes in which each guanine base-pairs with two other guanines via Hoogsteen hydrogen bonds (*31-33*)**.** G-quadruplexes have characteristic CD profiles (*31-33*). We therefore collected the CD spectra of our G-tract RNA, using a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK) and a 0.1 mm path length cuvette. To assure native conformation, frozen RNA samples in MQ-H_2O were thawed slowly on ice-cold water. For G_{12}, the concentrations of RNA used were 4.1 μM in MQ-H_2O and 3.7 μM in RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT). For G_4, the concentrations of RNA used were 7.5 μM in MQ-H_2O and 6.8 μM in RNA binding buffer. Each spectrum was baseline corrected with the CD spectrum of MQ-H_2O. Furthermore, an average of three spectra was obtained per sample. Before CD spectroscopy, the G_{12} RNA was desalted by dialysis against two1000-fold volumes of MQ-H_2O at 12 hour intervals for 24 hours at 4 ^{o}C using a 3.5K Slide-A-Lyzer® dialysis cassette (Thermo Scientific). Because G_4 is small relative to the size cutoff of available membranes, we were cautious against RNA loss and did not dialyze it.

2.2.10 RNA native gel electrophoresis

To further understand the basis for better PPR27 binding to native G-tracts than thermally exposed ones, we investigated the gel mobility of native (no heating) and three forms of thermally exposed 5'-FLUO-G_{12} RNA. One set of the thermally exposed RNA was heated on a

heat block at 95 ^{o}C for 10 minutes, then 10X RNA binding buffer was added after which tubes were transferred to 95 ^{o}C hot water and left to slowly cool to room temperature for 4 hours followed by storage at 4 ^{o}C for at least five days prior to assay. The second set of RNA was heated on a heat block at 95 ^{o}C for 10 minutes, snap-cooled on ice for 2 minutes, transferred to room temperature, and 10X RNA binding buffer added. The third set of thermally treated RNA was heated on a heat block at 95 ^{o}C for 5 minutes, snap-cooled on ice for 2 minutes, transferred to room temperature, and 10X RNA binding buffer added. 12 µl of a mixture comprising 10.8 µl of 50 nM RNA and 1.2 µl of a 10X sample buffer (10 mM Tris base, 1.0 mM EDTA, pH 7.5, 50% (v/v) glycerol, 0.01% (w/v) bromophenol blue, 0.01%(w/v) xylene cyanol) were then loaded on to a native 18% polyacrylamide (29:1 acrylamide : N,N'-methylene-bis-acrylamide) gel and run at 150 V constant voltage at 4 ^{o}C using TBE (50 mM Tris-base, 50 mM boric acid, 1.0 mM EDTA) as electrophoresis buffer. A 12 nt poly(A) ssRNA oligo (A_{12}) was loaded as marker for monomeric G_{12} RNA.

2.2.11 Fluorescence resonance energy transfer (FRET)

FRET was used to determine whether PPR27 binding perturbs the native conformation of G-tract RNA. Because G_{12} RNA is short and most likely to form intermolecular quadruplexes, separately labeled 5'-FLUO-G_{12} and 3'-TAMRA labeled G_{12} (G_{12}-TAM-3') ssRNAs were used. First, each RNA was desalted by dialysis against 200-fold volume of MQ-H_2O at 4^{o}C overnight using a 2.0K MWCO Spectra/Por® dialysis membrane (Spectrum Laboratories Inc,

Rancho Dominguez, CA). To anneal the 5'-FLUO-G_{12} and G_{12}-TAM-3' ssRNAs, a mixture of 350 µl of each RNA at 3.0 µM monomeric concentration was heated at 95 ^{o}C on a heat block for 10 minutes, then 245 µl of MQ-H_2O and 105 µl of 10X RNA binding buffer were added, diluting the RNA to 2.0 µM monomeric concentration. Tubes containing the RNA mixture were then transferred to 95 ^{o}C hot water in a plastic dish and allowed to slowly cool to room temperature on the bench for 4 hours. Finally, the RNA was stored at 4 ^{o}C for at least five days to allow further renaturation before FRET assays. The intensities of fluorescence emission of 50 nM monomeric concentrations of 5'-FLUO-G_{12} alone, G_{12}-TAM-3' alone, and of the annealed FLUO-G_{12}/G_{12}-TAM mixture in the absence and presence of various PPR27 concentrations were then collected at 25 ^{o}C following excitation at 490 nm using a microsquare quartz, open top, 5 mm sample cell (Starna Cells, Atascadero, CA) and a QuantaMaster™ spectrofluorometer (Photon Technology International, Birmingham, NJ). The spectrum of each sample was baseline corrected with that of the buffer, and three independent datasets were used to compute an average.

2.2.12 *T. brucei* mitochondrial RNA isolation

Total mitochondrial RNA was extracted from procyclic forms of *T. brucei*. Briefly, cells were grown at 27 ^{o}C in semi-defined maintenance SM medium (*34*) containing 5% (v/v) fetal bovine serum (SDM79) till mid-log phase to a cell density of $3.5x10^7$ per ml (courtesy of Dr. Koslowsky). Cells were pelleted from the culture medium by centrifugation at 6,090g at 4 ^{o}C for

10 min after which they were resuspended in buffer DTE (1.0 mM Tris, 1.0 mM EDTA, pH 7.9) and broken open with a Dounce homogenizer to release organelles. Progress in cell lysis was monitored under a light microscope. Mitochondria were then isolated from other lysate constituents by differential centrifugation. First, organelles were separated from cytoplasm by centrifugation at 15,800g in 7.5% (w/v) sucrose. The pellet was resuspended in STM medium (20 mM Tris pH 7.9, 2.0 mM $MgCl_2$, 8.6% (w/v) sucrose) supplemented with 3.0 mM $MgCl_2$, 0.3 mM $CaCl_2$, 10 U/ml DNase and incubated on ice for one hour. STE (20 mM Tris, pH 7.9, 2.0 mM EDTA, 8.6% (w/v) sucrose) medium was added to inactivate the DNase followed by centrifugation at 15,800g for 10 min, discarding any residual cytoplasmic RNA in the supernatant. The pellet was resuspended in STE medium and spun at 1,000g for 10 min, resolving the nuclei and mitochondria into the pellet and supernatant respectively. A final spin of the supernatant at 15,800g for 15 min pelleted the mitochondria. To extract RNA, mitochondria were lysed by vortexing vigorously while resuspended in 10 ml solution D (6.3 M guanidinium thiocyanate, 39.2 mM sodium citrate, pH 7.0, 0.78% (w/v) N-lauroyl sarcosine (sarkosyl). Lysate was passed through a 26G needle two times to shear any residual DNA. To separate mitoRNA from DNA, 1.0 ml of 2.0 M NaOAc was added to every 10 ml of lysate followed by 10 ml of water saturated phenol and vigorous shaking. Then 1ml of chloroform-isoamyl alcohol (49:1) was added followed by a vigorous shake. After cooling on ice for 15 min, a spin at 10,000g resolved the RNA and DNA into the aqueous (top) phase and phenol-chloroform phase respectively. To precipitate the RNA from the aqueous layer, one volume of isopropanol was added followed by cooling at -20 ^{o}C overnight. A subsequent spin at 10,000g recovered the RNA in the pellet. The pellet was rinsed with 70% (v/v) ethanol and reconstituted in sterile

RNase-free MQ-H_2O. Residual DNA and protein were then mopped up by DNase digestion, standard phenol chloroform-isoamyl alcohol (PCA) extraction (with TE-saturated phenol) and ethanol precipitation.

2.2.13 Pull-down assays of mitochondrial RNA

2.0 μM MBPPPR27His_6 protein was bound on to about 20 μl bed of amylose magnetic beads on a rocker at 4 ^{o}C for one hour after which the beads were washed with RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT). 61 μg of total mitoRNA was then added to the bead-bound protein and gently shaken in hand at ambient temperature for 15 min. After washing off unbound RNA with three 200 μl portions of RNA binding buffer, beads were resuspended in 200 μl of RNase-free MQ-H_2O followed by recovery of bound RNA by standard phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. The RNA pellet was reconstituted in 40 μl of RNase-free MQ-H_2O. The bound RNA was then probed by RT-PCR for the presence of polycistronic RNA precursors, pre-edited mRNA transcripts, and edited mRN**A** using various pairs of specific primers (**Table 2.3**). The Tb12Sc, Tb9S-1, Tb9S-2c, and TbND8 primers are the same as those previously used by Koslowsky and Yahampath (*35*). Reverse transcription of bound RNA utilized M-MLV reverse transcriptase and the primers Tb9S-1, ND8, 3'ND7HR3, and 3'ND7END (**Table 2.3**). PCR reaction mixtures were run on 1% agarose gels in TBE buffer at 100 V to assess the presence of maxicircle RNAs of interest.

2.3 RESULTS

2.3.1 PPR27 construct design, soluble yield and oligomeric state

The MTS of PPR27 is unconventional, hence it is difficult to estimate its length by programs which scan for classical amino acid characteristics at the N-terminus alone. As an alternative, a combination of classical PPR motif prediction (**Figure 2.1A**), de novo repeat prediction (**Figure 2.1B**), and secondary structure prediction (**Figure 2.2**) was used to estimate the MTS as well as gain a better view of the PPR motif organization in PPR27. MTS peptides are generally N-terminal extensions of between 20 and 50 residues in length (*36*) and can be short α-helices or disordered regions (*37-38*). Based on the predicted repeat organization and secondary structure, the first 22 amino acids comprising a short α-helix and a disordered segment were taken as the putative MTS. Our analysis suggests the MTS is followed by six contiguous PPR motifs (**Figure 2.4A**), two more than previously reported (*9-10*).

Table 2.1: Prediction of PPR27 organellar localization

Localization tool*	Targeting prediction
pTARGET	mitochondrial
TargetLoc	mitochondrial
DBSubLoc	mitochondrial
SLPFA	mitochondrial
Phobius	non-cytoplasmic
ESLPred	mitochondrial
LOCtree	mitochondrial
CELLO	
By amino acid composition	mitochondrial
By N-terminal peptide	non-mitochondrial
SHERLOC	
By amino acid composition	mitochondrial
By N-terminal peptide	non-mitochondrial
pLOC	
By amino acid composition	mitochondrial
MultiLoc	
By amino acid composition	mitochondrial
By N-terminal peptide	non-mitochondrial
TargetP 1.1	none
MitoProt II	none

*Multiple online servers with contemporary localization programs, some with ability to predict mitochondrial localization from amino acid composition of the whole polypeptide rather than from the N-terminus alone, were used.

Table 2.2: Primers used in cloning and site-directed mutagenesis.

Activity	Primer code	Primer sequence (5' to 3')
Add BamH1 restriction site on to 5' end of PPR27	CGH30	ACTTCCAGGGATCCGGTCACGTGTACGCCCTTC
Add SalI restriction site on to 3' end of PPR27	CGH34	GCCTGCAGGTCGACTCAACCACGAGGTAAAGT
Insert His_6 tag at C-terminus of PPR27	CGH69_F	CCACTTTACCTCGTGGTCACCACCACCACCACCA-CTGAGTCGACCTGCAG
	CGH70_R	CTGCAGGTCGACTCAGTGGTGGTGGTGGTGGTGA-CCACGAGGTAAAGTGG
Insert His_6 tag between MBP and PPR27ΔNR2[a]	CGH67_F	GGATTTCAGAATTCGGATCTCACCACCACCA-CCACCACGAAAACCTGTACTTCCAGGG
	CGH68_R	CCCTGGAAGTACAGGTTTTCGTGGTGGTGGT-GGTGGTGAGATCCGAATTCTGAAATCC
Convert $MBPHis_6PPR27\Delta NR2$[a] to $MBPHis_6$	CGH77_F	TACTTCCAGGGATCCTGATGCGCTCTTGCAGCC
	CGH78_R	GGCTGCAAGAGCGCATCAGGATCCCTGGAAGTA

[a]A plasmid containing a truncated form of PPR27 lacking the first two N-terminal PPR motifs (hence ΔNR2) was used as template for creation of $MBPHis_6$.

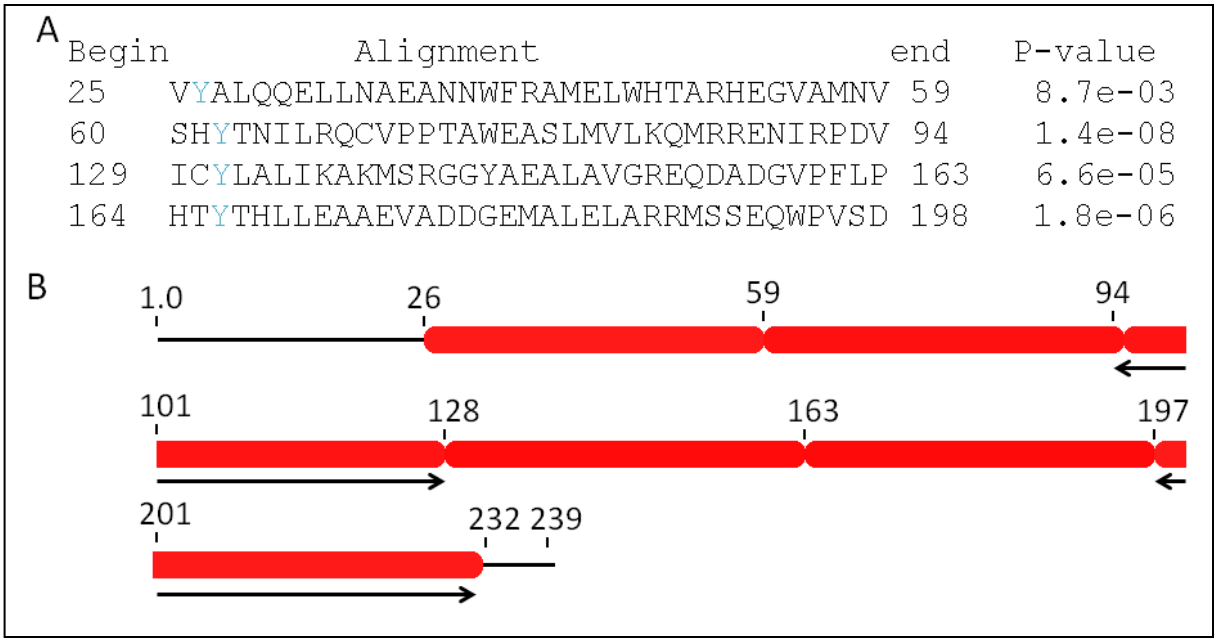


Figure 2.1: Organization of PPR motifs in PPR27.

(A) As predicted by TPRpred program, a classical PPR motif prediction tool. (B) As predicted by HHrepID, a de novo repeat prediction tool; underlying arrows denote the novel PPR motifs 3 (V95-I129) and 6 (D198-L232).

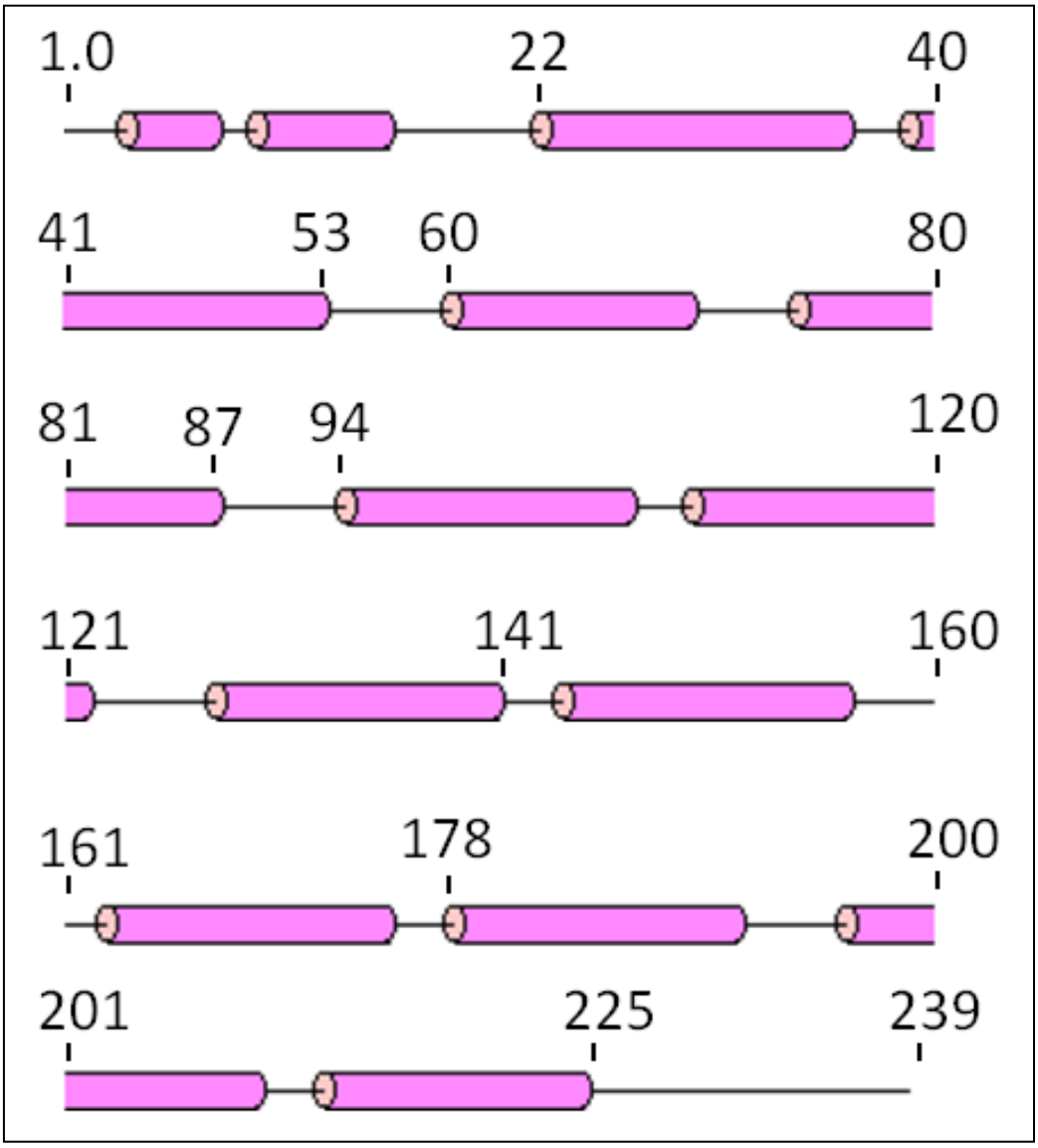


Figure 2.2: Predicted secondary structure of PPR27 by PSIPRED.

Each magenta block is one α-helix. Shown are two short α-helices of the putative mitochondrial targeting signal at the N-terminus and 6 pairs of antiparallel α-helices for the six PPR motifs.

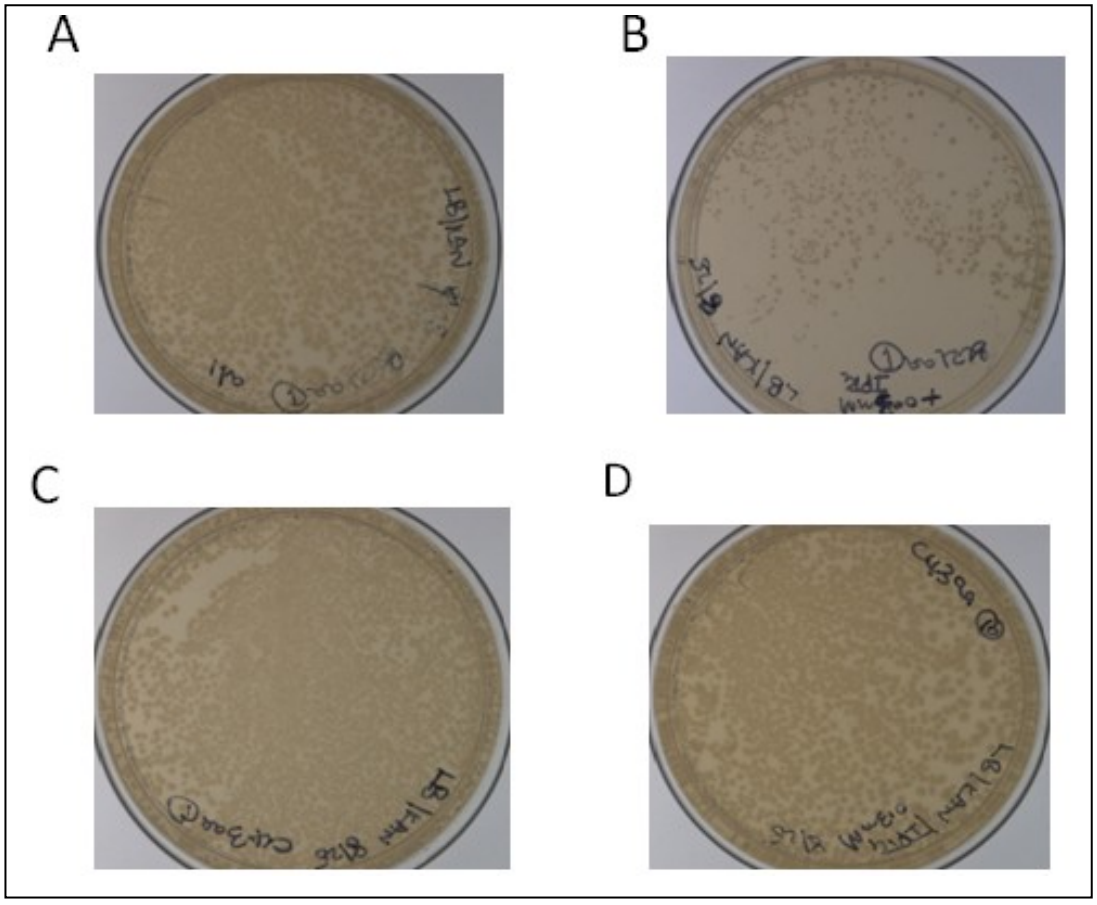


Figure 2.3: LB agar plates showing the viability of toxicity-susceptible BL21 (DE3) and toxicity-resistant C43 (DE3) *E.coli* before and after IPTG induction of PPR27 overexpression.

(A) Lots of recombinant BL21 (DE3) colonies on plate without IPTG. (B) Drastic reduction in population of recombinant BL21 (DE3) colonies on plate with IPTG. (C) Lots of recombinant C43 (DE3) colonies on plate without IPTG. (D) Lots of recombinant C43 (DE3) colonies on plate with IPTG.

Mature PPR27 (lacking the MTS) was expressed as a passenger protein to the C-terminus of MBP to aid soluble expression. A typical soluble yield of 5.0 mg per liter of *E. coli* culture was obtained. After tandem affinity chromatography, we achieved fusion protein of over 90% purity (**Figure 2.4D-E**). After detachment of MBP from PPR27 by TEV protease, no precipitation occurred and free PPR27 was visible on SDS-PAGE (data not shown). However, PPR27 aggregated on purification even in the presence of detergents and glycerol. To establish the oligomeric state of PPR27, we therefore performed blue native PAGE on the TEV protease cleavage mixture of MBPPPR27His_6. On native PAGE, the 24.4 kDa mature PPR27 migrates slightly faster than the 27 kDa TEV protease, indicating it is monomeric (**Figure 2.4F**).

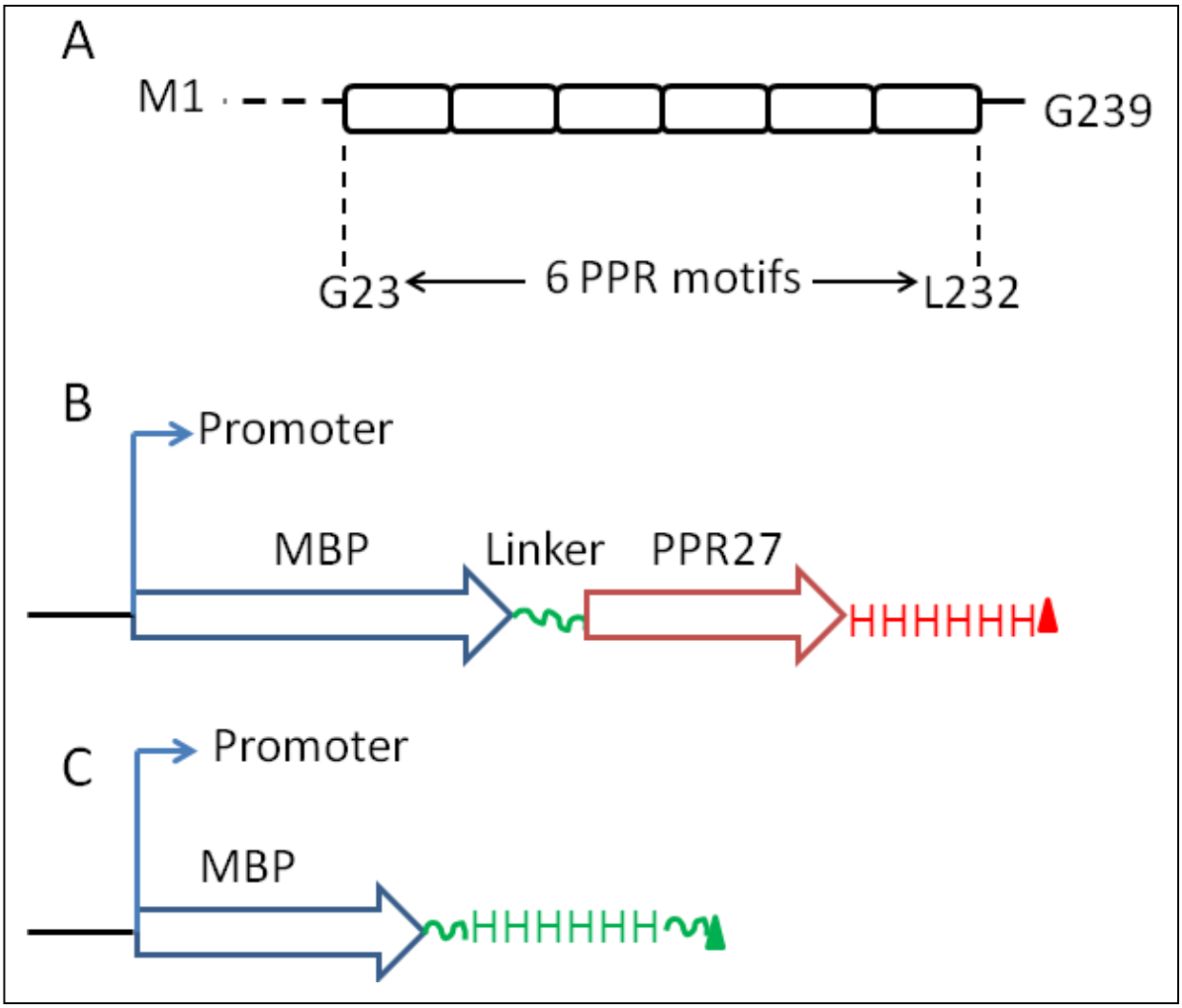
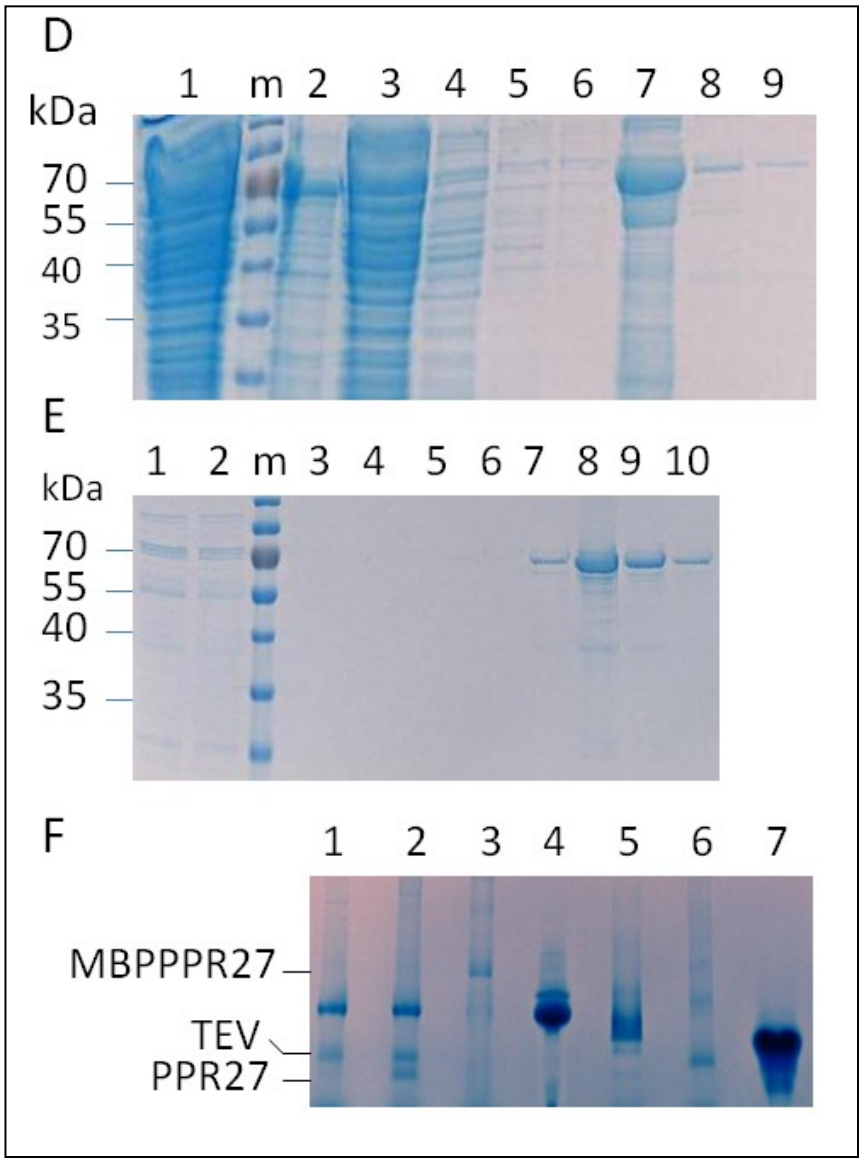


Figure 2.4: Design and expression of PPR27.

(A) Organization of PPR motifs in PPR27. (B) Scheme of the MBPPPR27 fusion protein emphasizing the His_6 tag at the C-terminus of PPR27. (C) Scheme showing free MBP with a His_6 tag at the C-terminus.

Figure 2.4 (cont'd)



(D) SDS-PAGE for MBPPPR27His$_6$ fractions from the Ni-NTA column (lane 1, supernatant; lane m, PageRuler^{TM} size marker; lane 2, inclusion bodies; lane 3, flow through; lane 4, wash 1; lane 5, wash 2; lane 6, wash 5; lane 7, eluate 1; lane 8, eluate 2; lane 9, eluate 3).

Figure 2.4 (cont'd).

(E) SDS-PAGE of MBPPPR27His_6 fractions from the amylose column (lanes 1 and 2, flow through portions 1 and 2 respectively; lane m, PageRuler^{TM} size marker; lane 3, wash 1; lane 4, wash 2; lane 5, wash 3; lane 6, wash 4; lane 7, eluate 1; lane 8, eluate 2; lane 9, eluate 3; lane 10, eluate 4). (F) 6-12% gradient native polyacrylamide gel showing the monomeric state of PPR27 (lane 1, TEV protease cleavage mixture of MBPPPR27 in absence of detergent showing no free PPR27; lane 2, TEV protease cleavage mixture of PPR27 in the presence of DDM detergent showing free PPR27; lane 3, MBPPPR27His_6 alone; lane 4, MBP; lane 5, factor Xa; lane 6, TEV protease alone; lane 7, thioredoxin with a His_6 tag).

2.3.2 PPR27 preferentially binds guanine-rich single stranded RNA

Insights into the sequence- and chemical specificity of PPR27's nucleic acid ligand were obtained by assessing the interaction of its MBP fusion protein with homopolymeric RNAs and DNA. EMSA showed an upward shift in the mobility of 12-nucleotide ssRNA for poly(rG) with no shift visible for poly(rA), poly(rC), or poly(rU) (**Figure 2.5)**, an indication of sequence-specific RNA binding. Similarly, only poly(G) displays an increase in fluorescence anisotropy with PPR27 concentration (**Figure 2.5**). A K_d of about 0.6 μM is obtained from the FA data (**Figure 2.6A**), and the binding isotherms saturate, consistent with specific binding. Using my MBPPPR27His_6 construct, Neil White in our lab also obtained a set of PPR27-binding RNAs dominated by guanine-rich sequences after 11 rounds of in vitro selection from a pool of 2.4 x 10^{18} random sequences with a 20 nucleotide variable region (data not shown), consistent with my findings.

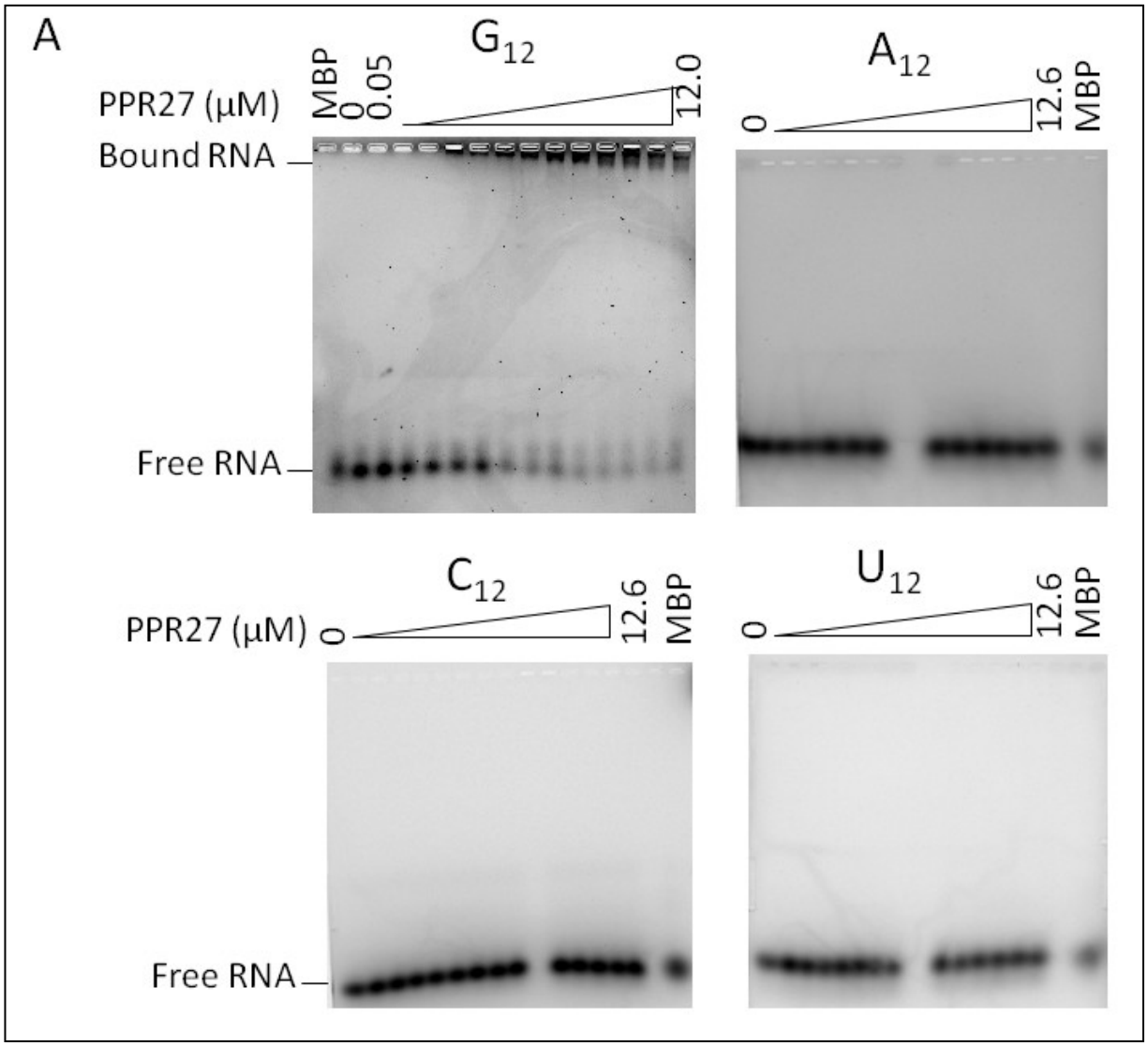
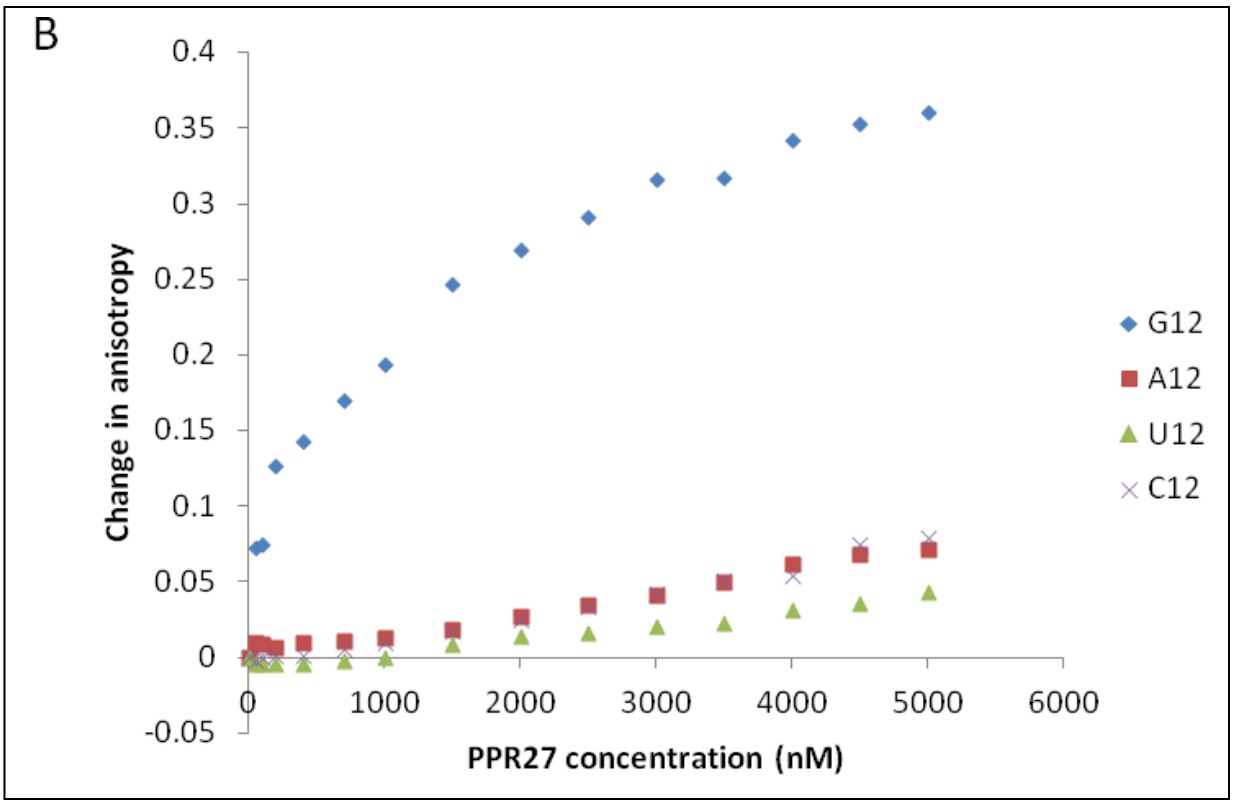


Figure 2.5: Interaction of PPR27 with 12 nucleotide homopolymeric RNA.

(A) 2 % tris-glycine agarose gels showing retardation of G_{12} and not the other RNA homopolymers (A_{12}, U_{12}, and C_{12}) as the protein concentration increased. In all gels, the RNA binding reaction contained 0.15 μM 5'-FLUO labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, 0.1 μg/μl BSA, 0.4 U/μl RNasin), and MBPPPR27. A reaction with 38 μM MBP was also loaded as negative control.

Figure 2.5 (cont'd).



(B) Variation of anisotropy with protein concentration reiterating the selectivity of PPR27 for poly(G) ssRNA. Each reaction for FA contained 20 nM 5'-FLUO labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT), and MBPPPR27.

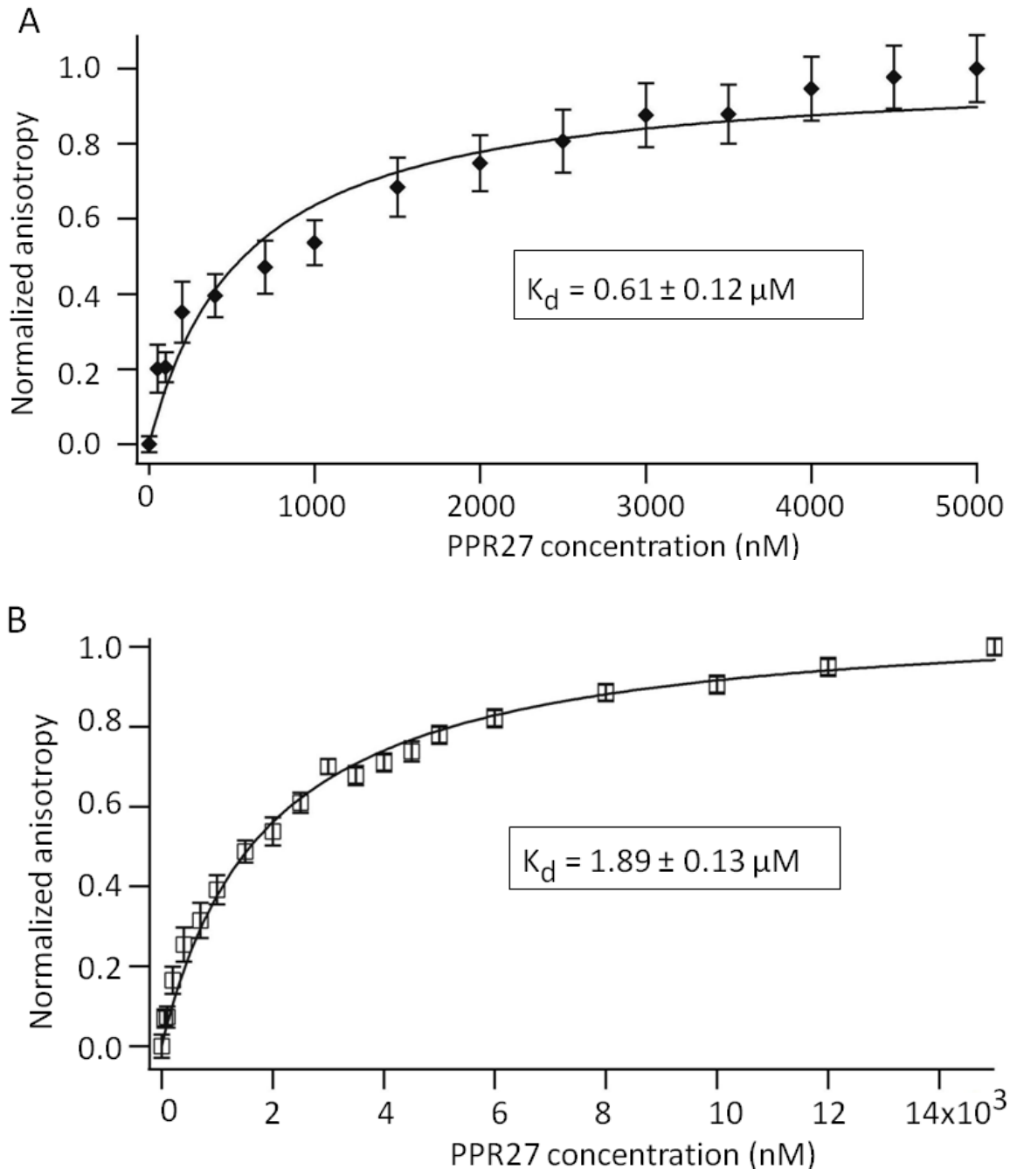


Figure 2.6: Nonlinear plots used in the quantitation of the K_d for interaction of PPR27 with G_{12} ssRNA and G_{12} ssDNA.

Figure (2.6 cont'd).

Inset in the rectangles are the K_d values. (A) Variation of anisotropy of 5'-FLUO-G_{12} ssRNA with increasing PPR27 concentration. (B) Variation of anisotropy of 5'-FLUO-G_{12} ssDNA with increasing PPR27 concentration. The binding reaction contained 20 nM 5'-FLUO-G_{12} ssRNA or ssDNA, RNA binding buffer, and MBPPPR27.

We further investigated the specificity of PPR27 for single-stranded RNA versus double-stranded RNA and single-stranded DNA oligos. Annealing of G_{12}- with C_{12} RNA to form double-stranded $(G.C)_{12}$ abolishes the PPR27-induced upward shift in poly(G) mobility (**Figure 2.7)**. Though PPR27 interacts with G_{12} ssDNA (**Figure 2.7)**, there is a three-fold decrease in affinity (**Figure 2.6B)** relative to ssRNA, an indication of binding selectivity between different chemical forms of the same nucleic acid sequence.

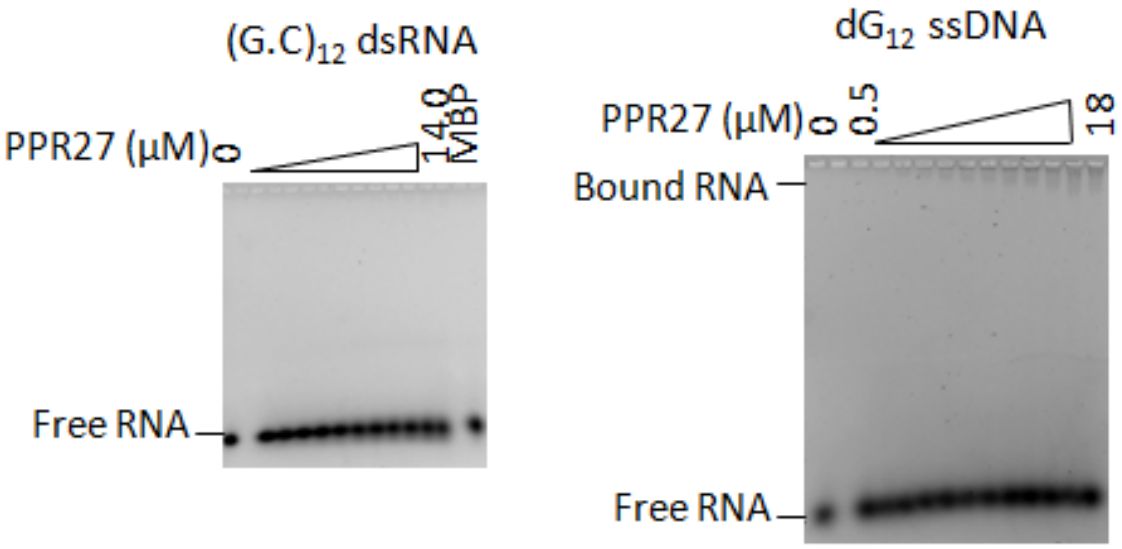


Figure 2.7: EMSA for interaction of PPR27 with $(G.C)_{12}$ and dG_{12} ssDNA.

2% agarose tris-glycine gels showing no retardation in $(G.C)_{12}$ dsRNA electrophoretic mobility (left panel) and modestly increased retardation in dG_{12} mobility with PPR27 concentration (right panel). In the $(G.C)_{12}$ gel, a reaction with 38 μM MBP was loaded as negative control. $(G.C)_{12}$ dsRNA was prepared by annealing 5'-FLUO-C_{12} and 5'-FLUO-G_{12}. The binding reaction contained 0.15 μM 5'-FLUO labeled nucleic acid and RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, 0.1 μg/μl BSA, 0.4 U/μl RNasin).

2.3.3 Affinity of PPR27 for G-tract RNA increases with the number of contiguous guanosines

To understand the biological relevance of PPR27 selectivity for G-tract RNA, the distribution of guanine homopolymers in the *T. brucei* mitochondrial genome was analyzed. We found preferential enrichment of G-tracts in some genes over others. In particular, there is a higher prevalence of G-tracts with at least four consecutive guanines in pan-edited mRNA genes than in never-edited mRNA genes, mildly-edited mRNA genes, gRNA genes, rRNA genes, and edited mRNA forms (see Discussion) (**Figure 2.8A; Tables 2.5-2.10)**. Some pan-edited genes contain G-tracts of up to 11 consecutive guanines. We sought to determine how PPR27 interacts with G-tract ssRNA containing different amounts of consecutive guanines in order to understand how it may impact mRNA processing. We assessed PPR27 interaction with G_9 (GGGGGGGGG), G_6 (GGGGGG), $G_6C_2U_2A_2$ (GGGGGGCCUUAA), G_4 (GGGG), $G_4C_2U_3A_3$ (GGGGCCUUUAAA), and $(GGU)_4$ (GGUGGUGGUGGU) using both EMSA and FA. With EMSA, intense shifted bands on increasing PPR27 concentration only appear with G_9 but not the others (**Figure 2.9)**. FA however shows that the K_d of PPR27 interaction with G_9, G_6, and $G_6C_2U_2A_2$ is about the same at 1.5-2 fold that of G_{12} (**Figure 2.10A; Table 2.4**). Transition from G_{12} to G_4 and $G_4C_2U_3A_3$ increases the K_d about six- and four-fold respectively, while from G_{12} to $(GGU)_4$ increases it by at least one order of magnitude (**Figure 2.10B; Table 2.4**). EMSA is a non-equilibrium assay and is poorly suited to detection of interactions with a fast ligand dissociation rate. On the other hand, FA is an equilibrium assay with a greater dynamic range of analysis for binding affinity than EMSA (*39*). Hence, a potential cause of the

contrasting observations from EMSA and FA as G-tracts decrease in size is the differential off rate of the RNAs from the protein. The running buffer stayed cold during electrophoresis, hence heat-induced dissociation of complexes was a less likely influence on the EMSA results.

Table 2.3: Primers used in reverse transcription of PPR27 bound *T. brucei* mitoRNA and PCR of resultant cDNA

Region amplified	Primer pairs	Primer sequence (5' to 3')
Intergenic region between 12S rRNA and 9S rRNA	Tb12Sc	CGGGTACCGAGCAGGTTAACAAGC
	Tb9S-1[a]	CGAATTCCGCAACGGCTGGCATCC
Intergenic region between 9S rRNA and ND8 mRNA	Tb9S-2c	CGGGTACCGGTATTGTTGCCCACCA
	TbND8[a]	GGAATTCCGCCTCTCTGGTTCTC
5'terminus of ND7	TbND7 5'NED	CATGACTACATGATAAGTAC
	3'ND7HR3[a]	CATTGTTCTACACTTTTATATTCACATAACTT
Full ND7 transcript(s)	TbND7 5'NED	CATGACTACATGATAAGTAC
	3'ND7END[a]	TTTTTATTCAATTTTAAACAATCCT

[a]These primers were used for reverse transcription of the RNA.

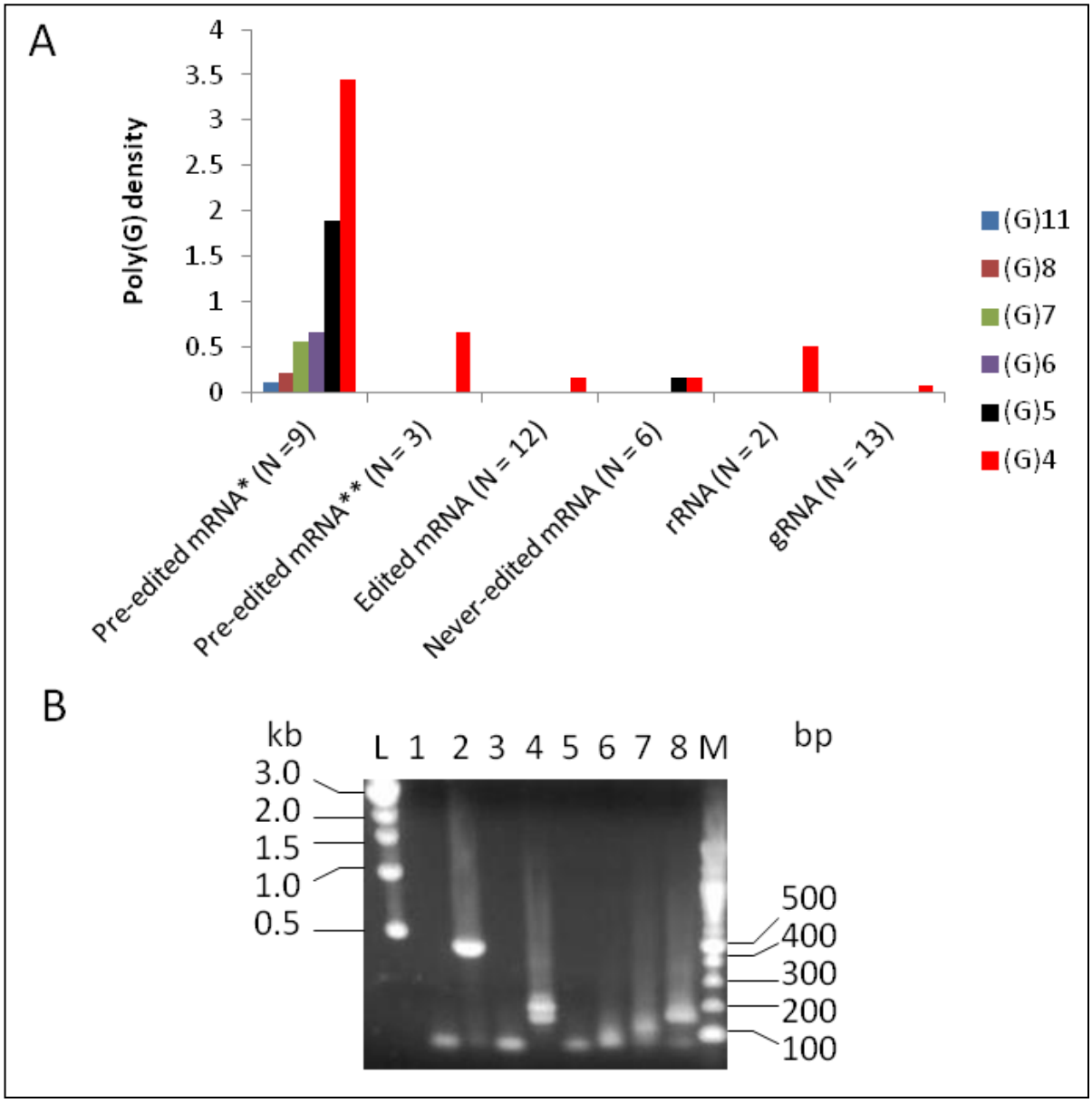


Figure 2.8: G-tract census and pull-down assays.

Figure 2.8 (cont'd).

(A) Prevalence of poly(G) homopolymers of 4, 5, 6, 7, 8, and 11 contiguous guanosine residues. The frequency of homopolymers of a given size was divided by the number of members in each transcript type to get the homopolymer density, which was then plotted against transcript type. The key to the bar colors is indicated on the graph; *pre-edited transcripts of extensively edited mRNA; **pre-edited transcripts of mildly edited mRNA. (B) 1% agarose gel showing the electrophoretic mobility of PCR amplicons of *T. brucei* kRNA bound by PPR27. L, 1 kb DNA marker; M, 100 bp DNA marker; 1, 3, 5, and 7 are negative control reactions (MQ-H_2O) for primer pairs in lanes 2, 4, 6 and 8 respectively; 2, PCR for intergenic region between 12S and 9S rRNAs; 4, PCR for intergenic region between 9S rRNA and ND8 mRNA; 6, PCR for the 5' terminal region of ND7 mRNA; 8, PCR for full ND7.

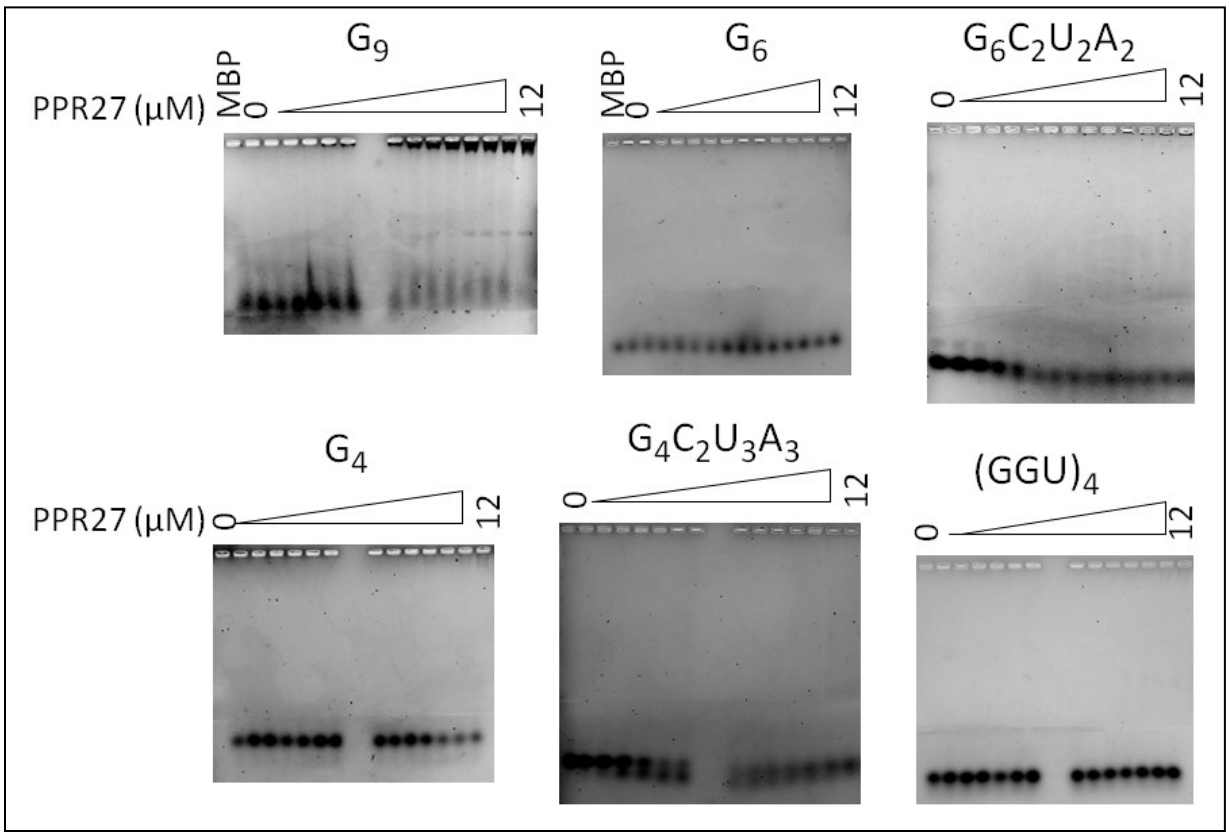


Figure 2.9: 2% agarose gels showing EMSA of PPR27 with G-tracts having fewer than 12 guanines.

Each test reaction contained 0.15 μM 5'-FLUO labeled ssRNA and RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, 0.1 μg/μl BSA, 0.4 U/μl RNAsin), and MBPPPR27. Where applicable, a reaction with 38 μM MBP was loaded as negative control.

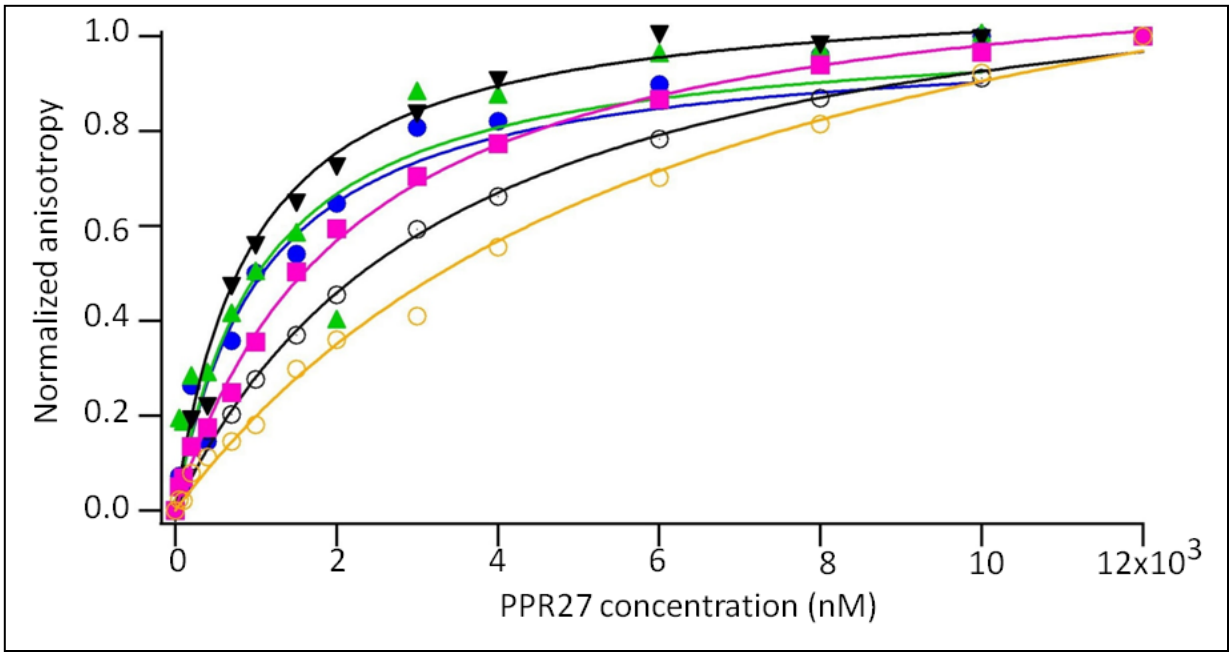


Figure 2.10: Anisotropy for interaction of PPR27 with G-tracts having fewer than 12 guanines.

The G-tracts examined were G_9 (blue closed circles), G_6 (green upward triangles), $G_6C_2U_2A_2$ (inverted, black closed triangles), G_4 (black circles with central dot), $G_4C_2U_3A_3$ (magenta closed squares), and $(GGU)_4$ (orange open circles) ssRNA with increasing PPR27 concentration. Each reaction contained 20 nM 5'-FLUO labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT), and MBPPPR27.

Table 2.4: PPR27's affinity for various RNA G-tracts

RNA	K_d* (μM)
G_{12}	0.61 ± 0.12
G_9	1.08 ± 0.01
G_6	1.07 ± 0.01
$G_6C_2U_2A_2$	0.93 ± 0.01
G_4	3.42 ± 0.19
$G_4C_2U_3A_3$	2.21 ± 0.03
$(GGU)_4$	6.55 ± 0.10

*K_d values were obtained from nonlinear regression of FA data as shown in **Figure 2.10** above.

2.3.4 Polycistronic maxicircle RNA precursors are among PPR27 targets

To gain further insights into the biological RNA ligands for PPR27, the MBP fusion protein immobilized on amylose magnetic beads was used to pull down cognate RNAs from procyclic *T. brucei* total mitoRNA. Procyclic *T. brucei* mitoRNA was used because it is at porcyclic stage when the mitochondrion operates at optimal metabolic capacity. Due to the preferential binding of PPR27 to G-tract RNA, we hypothesized that pull-down of total mitoRNA would capture either pre-edited and partially edited mRNAs or polycistronic precursor transcripts. RT-PCR for NADH dehydrogenase subunit 7 (ND7) was used to test for pre-edited and edited mRNA binding while that of the intergenic regions between 12S rRNA and 9S rRNA, and between 9S rRNA and NADH dehydrogenase subunit 8 (ND8) was used to assess binding of polycistronic RNA precursor transcripts. For both intergenic regions assessed, RT-PCR generated amplicons of the expected size (about 450 bp for the 12S rRNA/9S rRNA shared border and 180 bp for the 9S rRNA/ND8 mRNA intersection) (*35*) (**Figure 2.8B**), consistent with precursor RNA binding. RT-PCR for both the ND7 5' terminus as well as for the full ND7 transcript was negative (**Figure 2.8B**). ND7 transcripts are however downregulated in procyclic parasites (*14*) which were the source of RNA used here, hence the negative pull-down cannot be attributed to lack of PPR27 binding alone.

2.3.5 PPR27 preferentially binds G-quadruplexes over random coil RNA

At the beginning of our RNA binding experiments, the working hypothesis was that PPR27 binds ssRNA in random coil conformation. The traditional approach for ensuring random coil ssRNA conformation is denaturing the sample at 95 ^{o}C for at least two minutes followed by snap-cooling on ice. However, in the course of our FA experiments, specific and pronounced binding of the poly(G) ssRNA ligand by PPR27 was only observed after frozen RNA (-20 ^{o}C)

was slowly thawed on ice-cold water and thereafter kept on ice or used directly in the experiments without high temperature denaturation. When the slowly thawed poly(G) ssRNA was denatured at 95 ^{o}C and snap-cooled before use in the FA assays, the increase in anisotropy with PPR27 concentration was greatly diminished relative to the non-heated (native) RNA, and the binding isotherm did not saturate, consistent with weakened or less specific binding (**Figure 2.11A)**. Similarly, when frozen poly(G) ssRNA (-20 ^{o}C) was quickly thawed on a room temperature water bath and either kept on ice or used directly in RNA binding FA experiments, the increase in anisotropy with PPR27 concentration was much reduced (data not shown) relative to slowly thawed RNA.

Single stranded G-tract nucleic acids are renowned for forming advanced structures with characteristic CD and electrophoretic profiles called quadruplexes in which each guanine base-pairs with two other guanines via Hoogsteen hydrogen bonds (*31-33*)**.** We therefore investigated whether the native conformation of our G-tract RNA ligands is a quadruplex by CD spectroscopy and native gel electrophoresis. We also probed the effect of slow and rapid cooling of the thermally exposed G-tract RNA on its conformation using native gel electrophoresis. The CD spectra of native G_{12} and G_4 ssRNA display minima at 245 nm and maxima at 265 nm (**Figure 2.12A-B)**, consistent with a parallel G-quadruplex conformation (*32, 40*). On native PAGE, native G_{12} ssRNA not only migrates as a ladder-like smear of various RNA sizes larger than that of A_{12} RNA, most of the RNA is also retained in the well (**Figure 2.11B**), consistent with multimeric existence. Following the interpretation of Tluckova and coworkers (*41*), we speculate that that the slowest migrating native G_{12} RNA forms and those retained in the well are

tetrameric (four-stranded) quadruplexes and their aggregates respectively. A denaturing gel of the G-tracts revealed the resistance of some quadruplexes to denaturation (**Figure 2.11C**).

The intensity of native 5'-FLUO-G_{12} bands was severely diminished in the presence of monovalent ions compared to that of the same but thermally treated RNA. Monovalent and divalent ions stabilize quadruplex structure (*42-43*), and it is possible that they drive the G-tract RNA to a state in which the fluorescein stacks against the guanine, thereby reducing its quantum yield. Thermally treated but slowly cooled G_{12} RNA migrates as a single elongated band at the speed of the fastest migrating band of native RNA. It is highly likely that the elongated band results from overlapping migration of a dimeric quadruplex and monomeric coiled coil RNA (*41*) , of which the former severely bleaches fluorescein in the presence of metal ions. On the other hand, when thermally treated RNA is rapidly cooled in the absence of metal ions, up to 20% of G_{12} migrated at a size most likely to be a trimer while the rest migrated largely as monomers. In the presence of either K^+ or Li^+ ions, rapidly cooled G_{12} RNA only migrated as a monomer, and the quantum yield of fluorescein was greatly enhanced. Overall this data is consistent with PPR27 preference for G-quadruplexes over coiled coil RNA forms.

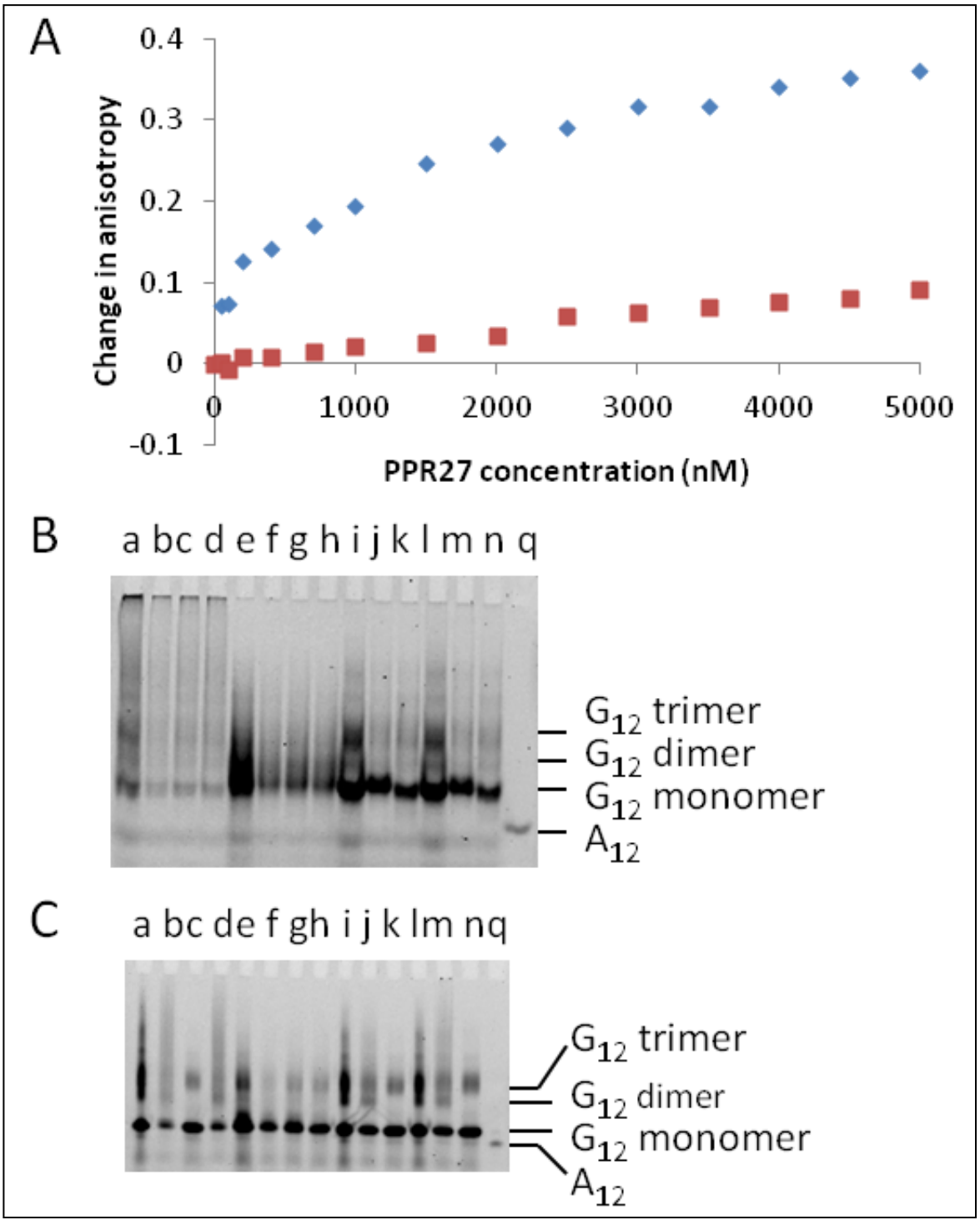


Figure 2.11: Effect of thermal denaturation of G-tract RNA on PPR27 recognition.

(A) Comparison of anisotropy profiles of native (blue) and thermally denatured (brown) G_{12} ssRNA as PPR27 concentration was increased.

Figure 2.11 (cont'd).

(B) Native 18% polyacrylamide gel showing the electrophoretic profile of 5'-FLUO-G_{12} pre-exposed to different thermal and salt conditions: a, native G_{12} with no salt; b, native G_{12} with RNA binding buffer 1 or RBB1 (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT); c, native G_{12} with buffer 2 (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 1.0 mM DTT); d, native G_{12} with buffer 3 (20 mM Tris-HCl, pH 7.5, 150 mM LiCl, 1.0 mM DTT); e, G_{12} heated at 95 ^{o}C for 10 minutes followed by slow annealing, no salt; f, G_{12} heated at 95 ^{o}C for 10 minutes followed by slow annealing, in RBB1; g, G_{12} heated at 95 ^{o}C for 10 minutes followed by slow annealing, in buffer 2; h, G_{12} heated at 95 ^{o}C for 10 minutes followed by slow annealing, in buffer 3; i, G_{12} heated at 95 ^{o}C for 10 minutes followed by snap cool, no salt; j, G_{12} heated at 95 ^{o}C for 10 minutes followed by snap cool, in RBB1; k, G_{12} heated at 95 ^{o}C for 10 minutes followed by snap cool, in buffer 3; l, G_{12} heated at 95 ^{o}C for 5 minutes followed by snap cool, no salt; m, G_{12} heated at 95 ^{o}C for 5 minutes followed by snap cool, in RBB1; n, G_{12} heated at 95 ^{o}C for 5 minutes followed by snap cool, in buffer 3; q, native A_{12} ssRNA in buffer TE (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA). (C) Denaturing 17% polyacrylamide gel (containing 6.8 M urea) showing the electrophoretic profile of 5'-FLUO-G_{12} pre-exposed to different thermal and salt conditions. Lane codes are the same as in (B).

2.3.6 PPR27 binding does not disrupt G-quadruplex structure

The folding of a single G-rich RNA strand to form an intramolecular quadruplex brings distant RNA ends closer in space. If G-tract RNA is double-labeled with fluorescein at one end (say 5') and the opposite end is labeled with a fluorophore such as TAMRA whose excitation spectrum overlaps the emission spectrum of fluorescein, there will be enhanced fluorescence resonance energy transfer (FRET) from the fluorescein (donor) to TAMRA (acceptor) if the RNA is an intramolecular quadruplex. Similarly, the annealing of two or more G-rich RNA strands to form an intermolecular quadruplex brings distant RNA molecules closer. Thus, if one or more G-tract RNA strands monolabeled with fluorescein are annealed to one more strands monolabeled with TAMRA to form an intermolecular quadruplex, excitation of fluorescein will cause some resonance energy transfer (FRET) to the TAMRA. When FRET occurs, the fluorescence emission intensity of the donor (fluorescein) will reduce while that of TAMRA (the acceptor) will increase. Assuming the resonance energy is not quenched by the guanine nucleotides, the FRET efficiency in both intramolecular and intermolecular quadruplexes will be higher the shorter the spatial distance between the donor and acceptor fluorophores. If the RNA is in random coil conformation, the FRET signal will either be low (for double-labeled strands) or nonexistent (for a mixture of monolabeled strands).

Some proteins remodel quadruplex structures (*44-45*). We therefore assessed the impact of PPR27 binding on quadruplex RNA structure using FRET. If PPR27 were to unfold poly$(G)_{12}$ quadruplexes, the FRET signal from the intermolecular quadruplex would reduce with PPR27 concentration. In this work, we found that when annealed to 5'-FLUO-G_{12}, the intensity of G_{12}-TAM-3' emission increases relative to G_{12}-TAM-3' alone (**Figure 2.13A-C**) following

excitation at 490 nm. When the sum of the spectra of free 5'-FLUO-G_{12} and free G_{12}-TAM-3' was deducted from the spectrum of annealed 5'-FLUO-G_{12}/G_{12}-TAM-3', a net positive TAM emission signal (peak intensity of about 8,000 counts/s) and no FLUO emission signal was observed, consistent with a FRET effect. The TAM emission intensity from FRET was however only 5% of the emission intensity of free 5'-FLUO-G_{12}, consistent with low efficiency of resonance energy transfer. Nevertheless, the low FRET signal was not unexpected as G-tracts without intervening heterogeneous nucleotides (A, U, and C) are very stable with melting points above 95 ^{o}C (*43, 46*), and only a small fraction of the RNA is unfolded and available to anneal to new partners. On titration with PPR27, the fluorescence emission of G_{12}-TAM-3' was not reduced but rather modestly increased, suggesting that PPR27 does not remodel G-quadruplexes.

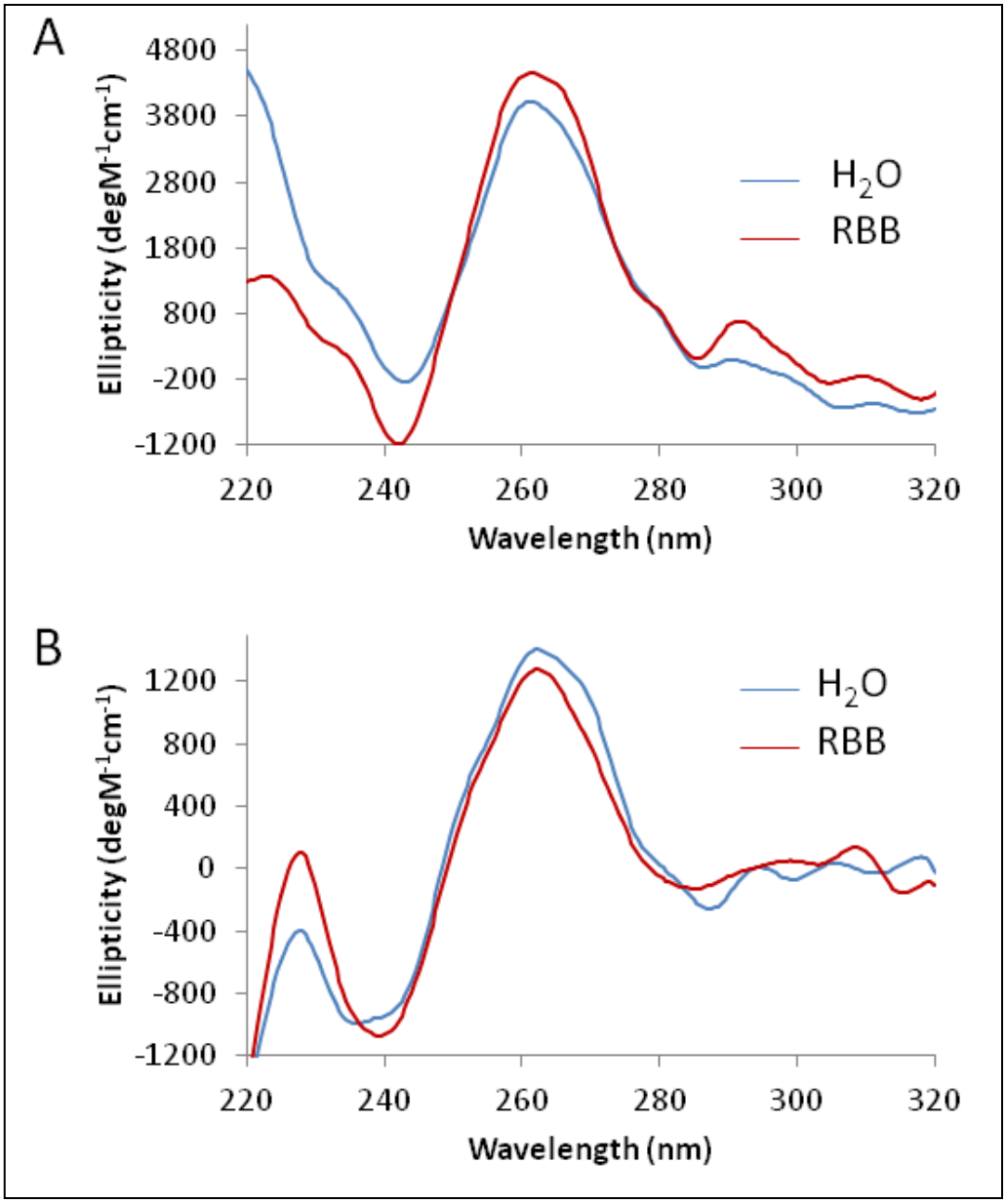


Figure 2.12: CD assays for G-tract conformation.

CD spectra* of G_{12} ssRNA (A) and G_4 ssRNA (B) in MQ-H_2O (blue) and RNA binding buffer or RBB (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT) (brown).

*Minima at 245 nm and maxima at 265 nm are characteristic of parallel G-quadruplexes. The RNA for CD assays was obtained from -20 ^{o}C freezer and slowly thawed on ice-cold water.

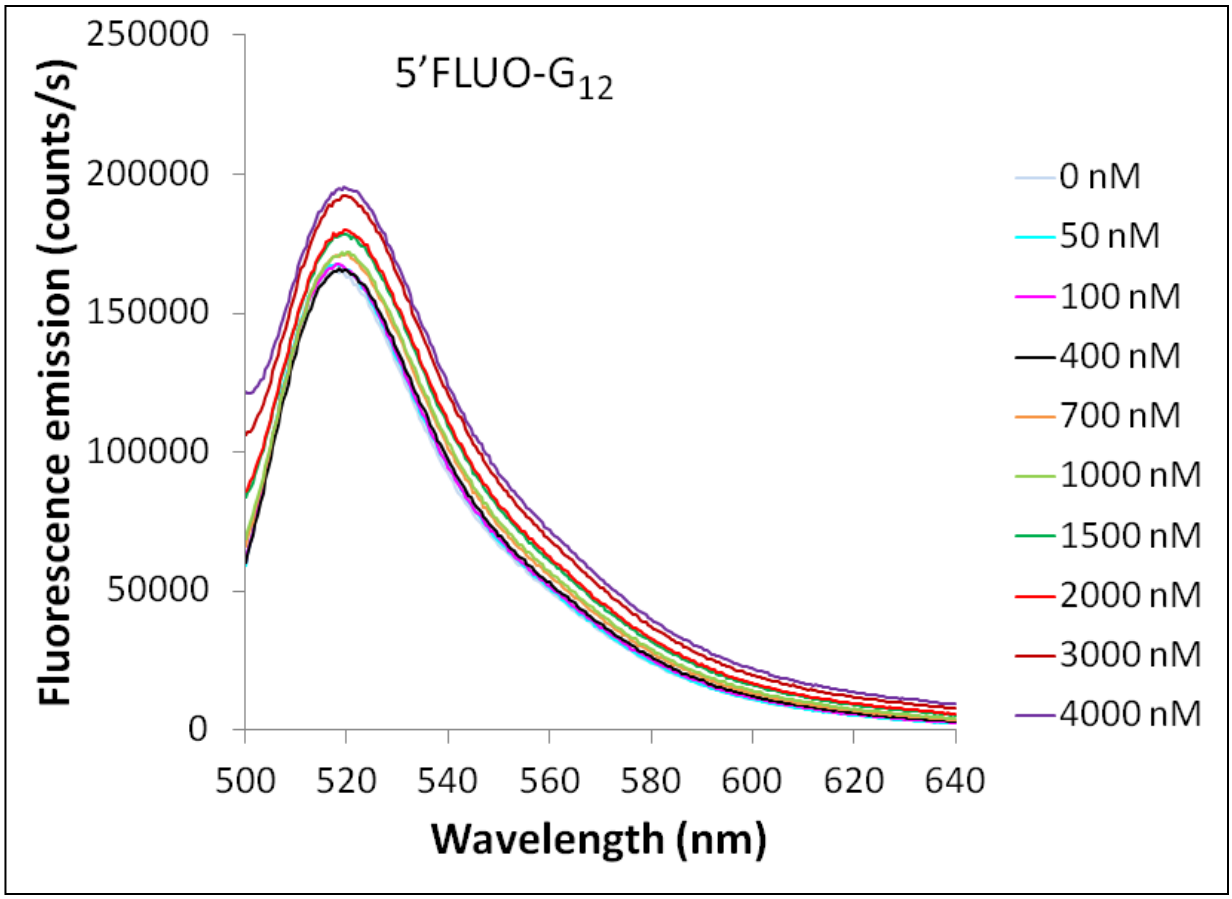
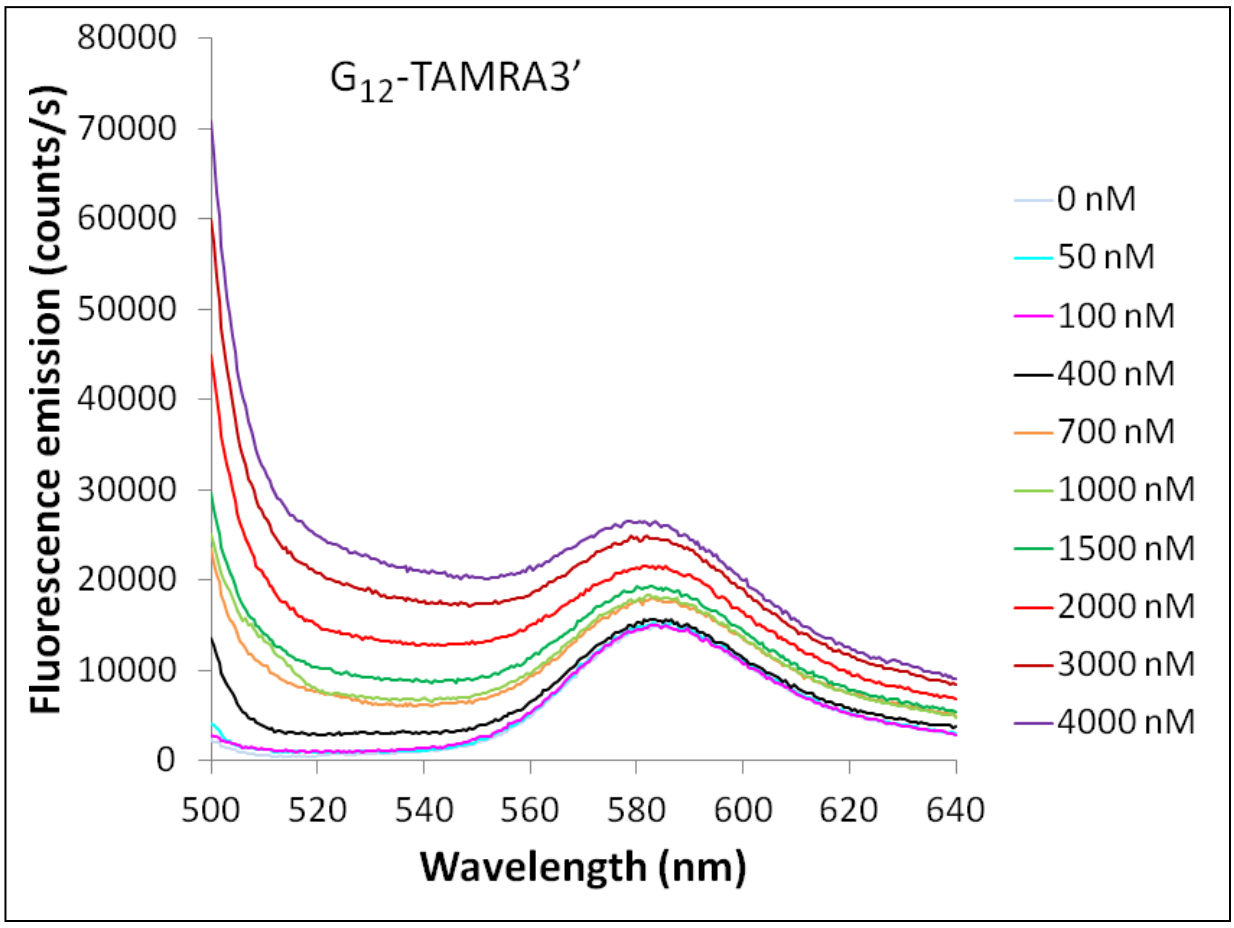


Figure 2.13: Effect of PPR27 binding on G-quadruplex

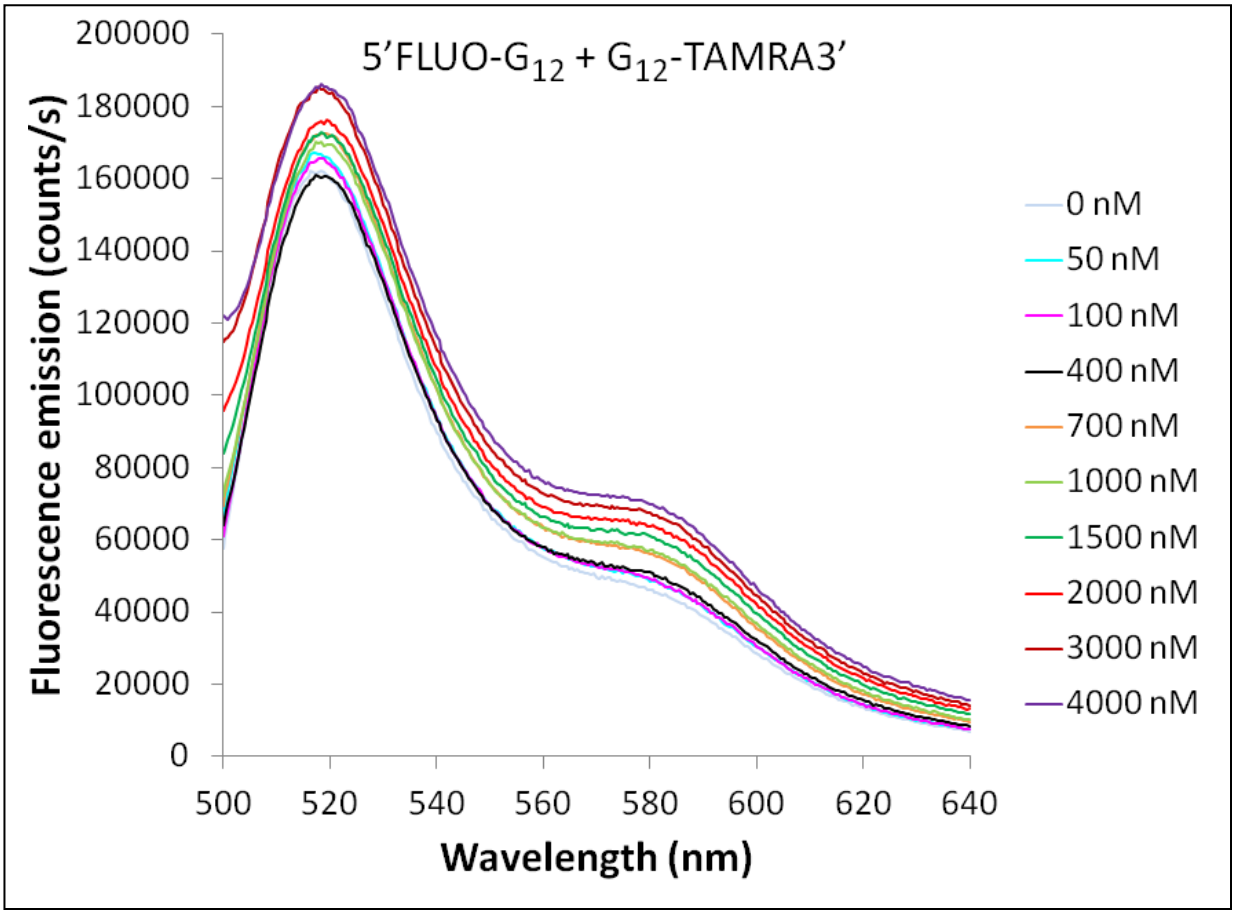
(A) Fluorescence emission intensity of 5'-FLUO-G_{12} in the presence of increasing PPR27 concentration.

Figure 2.13 (cont'd).



(B) Fluorescence emission intensity of G_{12}-TAM-3' in the presence of increasing PPR27 concentration.

Figure 2.13 (cont'd).



(C) Fluorescence emission intensity of an annealed mixture of 5'-FLUO-G_{12} and G_{12}-TAM-3' showing the excitation of G_{12}-TAM-3' by the emitted energy from 5'-FLUO-G_{12} or FRET.

Table 2.5: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* gRNA genes

Gene	G-tract	Percentage of-guanines in G-tract	Number of guanines in longest poly(G)-homopolymer
1) ND7	GGAGGAGAGGGG	75.0%	4
	GAGGGGAAGAGC	58.3%	4
	CCGAGAAGGGGG	58.3%	5
	GAGGGGAAGGGG	75.0%	4
	GGGGCGAGCAGG	66.7%	4
	GAGGGGGAGGGG	83.3%	5
	GAGAGAGAGGGG	66.7%	4
	GCGGCGGGGCAG	66.7%	4
	GGGGGCCGCGAG	66.7%	5
	GAGGGGAGAGUC	58.3%	4
	GGGGGGGGGGGA	91.7%	11
	GGGGGGGGCCGG	83.3%	8
	GAGGAAUGGGGG	66.7%	5
	GAGGGGACCGUA	50.0%	4
2) ND8	GGAAGGUGGGGA	66.7%	4
	GGGGGAGAGCGG	75.0%	5
	GGGGGGGAGGGG	91.7%	7
	GGAAGGGGAGCA	58.3%	4
	GGAGGGGAGCCA	58.3%	4
	GAGGGGGAGAGA	66.7%	5
	GGGGAGAAGGGG	75.0%	4
3) ND9	GGGGAGAGGGUU	66.7%	4
	GGGGAGAGGAGG	75.0%	4
	AGGGGGCGAGGG	75.0%	5
	GGGGCGGGGGGG	91.7%	7
	GCGGGGGAACGC	58.3%	5
	GAGGAGGGGGGG	83.3%	7
	GGAUCCAAGGGG	50.0%	4
	GGGGGGGAGGAG	83.3%	7

Table 2.5 (cont'd).

4) COIII	AAGGGGAGGGGG	75.0%	5
	GGAGGAGGGGGA	75.0%	5
	AGGGGAGGGGAG	75.0%	4
	GGAGAGGGGAGG	75.0%	4
	GGAGGGGUUGGG	75.0%	4
	GAGAGGGGGGGG	83.3%	8
	GGGGUGGGCAAA	58.3%	4
	GAGGGGGGAGAG	75.0%	6
	CGGGGGAAAGGG	66.7%	5
5) ATPase A6	GGGGGGGAGGGG	91.7%	7
	GGGGAAGAGGAG	66.7%	4
	GGGGAGAGGCGG	75.0%	4
	GGAUAAGAGGGG	58.3%	4
	AAGGGGAAAUGG	50.0%	4
	GGGGGAGGAGAG	75.0%	5
6) CR3	AAGGAUUGGGGG	58.3%	5
7) CR4	GGGGCAAGGGUG	66.7%	4
	AGGGGGGUUUGU	58.3%	6
	GGGGGAGAGGAA	66.7%	5
	GGGGGUUUGGGG	75.0%	5
	GAAGGGGAGAAG	58.3%	4
	AAAUUGAAGGGG	41.7%	4
	UUGAUUGGGGGG	58.3%	6
	GGGGAGAAAGUG	58.3%	4
	GGGGUGGGGGAG	83.3%	5
	GGGGAGAGGGGG	83.3%	5
8) ND3	GGGGGGCGGGGU	83.3%	6
	GGGGUGAAGGGA	66.7%	4
	GGGGGGAGAAGG	75.0%	6
	GGGGAGGGAUCA	58.3%	4
9) RPS12	GGGGACGGAGAG	66.7%	4
	GGGAGGCGGGGA	75.0%	4
	GAGGGUGGGGGG	83.3%	6

Table 2.6: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in *T. brucei* pre-edited transcripts of mildly edited mRNA genes

Gene	G-tract	Percentage of-guanines in G-tract	Number of guanines in longest poly(G)-homopolymer
1) Cyb	AUAUGGGGUAGG GGGGAAGUGAAU	50.0% 50.0%	4 4
2) COII	NONE		
3) MURF2	NONE		

Table 2.7: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in *T. brucei* edited transcripts of both pan-edited and mildly edited mRNA genes

Gene	G-tract	Percentage of-guanines in G-tract	Number of guanines in longest poly(G)-homopolymer
1) ND3	NONE		
2) ND7	NONE		
3) ND8	NONE		
4) ND9	NONE		
5) ATPase A6	NONE		
6) COIII	NONE		
7) CR3	NONE		
8) CR4	NONE		
9) RPS12	NONE		
10) Cyb	UAUGGGGUAGGU UUUGGGGAAGUG	50.0% 50.0%	4 4
11) COII	NONE		
12) MURF2	NONE		

Table 2.8: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* never-edited mitochondrial mRNA genes

Gene	G-tract	Percentage of-guanines in G-tract	Number of guanines in longest poly(G)-homopolymer
1) MURF5	NONE		
2) MURF1	NONE		
3) ND1	NONE		
4) COI	UGGUUUUUGGGG GUUGGUUGGGGG	50.0% 66.7%	4 5
5) ND4	NONE		
6) ND5	NONE		

Table 2.9: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* mitochondrial rRNA genes

Gene	G-tract	Percentage of-guanines in G-tract	Number of guanines in longest poly(G)-homopolymer
1) 12S rRNA	GUUUGAUUGGGG	50.0%	4
2) 9S rRNA	NONE		

Table 2.10: Non-overlapping 12-mer guanine RNA tracts (G-tracts) with at least four consecutive guanines in transcripts of *T. brucei* gRNA genes

Accession number of- minicircle encoding- the gRNA	Guide RNA (gRNA)	G-tract	Percentage of- guanines in G-tract	Number of guanines in longest poly(G)- homopolymer
L11652.1	gCOIII gCR4 gND7	NONE NONE NONE		
L25590.1	gCYb gCR3	NONE NONE		
L25589.1	gA6 gCYb gCR3	NONE AUGGGGAGGAUA NONE	50.0%	4
L25588.1	gA6 gCYb gCR3	NONE NONE NONE		
M55992.1	gCOIII	NONE		
U12393.1	gCR3	NONE		
U12394.1	gCR3	NONE		
L16541.1	gCOIII gND7a gND7b gCR5	NONE NONE NONE NONE		

Table 2.10 (cont'd).

L16537.1	gCR5 gND7	NONE NONE		
L16535.1	gCOIIIa gCOIIIb gA6	NONE NONE NONE		
L16540.1	gA6 gCOIII	NONE NONE		
L16538.1	gND8a gCOIII gND8b	NONE NONE NONE		
L16536.1	gA6 gND8a gND8b	NONE NONE NONE		

2.4 DISCUSSION

For the first time, we have demonstrated direct interaction between PPR27 and RNA in which preferential affinity for poly(G) RNA tracts was observed. Between G_{12} and G_6 ssRNA, the decrease in the K_d for PPR27-RNA interaction was modest, suggesting that G_6 could be the minimum RNA motif required for efficient PPR27 recognition.

Williamson and coworkers have demonstrated how knowledge of the minimal RNA motif responsible for protein binding can facilitate the mining of biological RNAs containing the motif, and hence generate leads on the biological relevance of the protein-RNA interaction (*47*). In an adaptation of this approach we compared the enrichment of guanosine tracts containing at least 4 consecutive guanosines in all *T. brucei* mitochondrial RNA, both individually and by type of RNA (**Figure 2.8A; Tables 2.5-2.10**). Based on the ratio of number of G-tracts with n consecutive guanosines to the number of genes, we found selective enrichment of G-tracts with at least 4 consecutive guanosines in pre-edited transcripts of extensively edited mRNA genes, providing potential biological relevance for PPR27 recognition of G-tracts. Pull-down experiments picked up maxicircle polycistronic precursor RNAs, suggesting that PPR27 could alternatively be involved in the processing of the latter to individual transcripts. In the continued absence of genetic data demonstrating the RNA phenotypes of PPR27 knockdown parasites, these are the first findings to provide strong leads as to the function of PPR27 in trypanosome posttranscriptional RNA processing.

In the RNA binding assays, PPR27 preferentially interacted with G-tracts in the quadruplex conformation over those in random coil states. Work on the G-tract rich telomeric DNA, among other systems, has demonstrated that G-quadruplex structures exist in vivo [reviewed in (*48*)]. At

least 30 nucleic acid binding proteins have been reported to preferentially bind G-quadruplex DNA and/or RNA over other structural forms (*48-49*). In these interactions, the biochemical activities of quadruplex nucleic acid binding proteins have varied from quadruplex stabilization to quadruplex destabilization/unwinding, and dissociation constants as low as 3.0 nM have been reported (*48-49*). On the other hand, some guanine quadruplex aptamers have been found to be potent inhibitors of protein function. For example, the dimeric parallel DNA quadruplex formed by the aptamer $G_4TG_3AG_2AG_3T$ potently inhibits HIV-1 integrase with an IC_{50} of 25-50 nM (*50*). Thus, G-quadruplex RNA recognition can modulate biological function.

In our experiments, the quadruplex RNA structure was not disrupted by PPR27 binding (**Figure 2.13**), suggesting that the protein's biological role is likely to anchor other RNA processing proteins to the RNA in a sequence specific manner or to stabilize pre-RNAs, both of which are familiar roles for PPR proteins (*4-5, 51-55*). Comparison of nonequilibrium gel-shift and equilibrium fluorescence anisotropy results suggested that, as the number of consecutive guanines decreased, the off-rate of the RNA increased, effectively destabilizing the RNA-protein complex under nonequilibrium conditions. That is, differential kinetics is at play in the interactions of PPR27 with different sizes of G-tracts, which could mean differential residence times of recruited effectors at sites of action. However, the differential RNA binding kinetics with G-tract size could be offset in vivo as the overall guanosine density in the regions surrounding the G_4, G_5, and G_6 homopolymers in pre-edited transcripts is very high and intervening heterogeneous sequences are very short (**Tables 2.5-2.10**). Hence, complex stabilities similar to those of G_9 and G_{12} are probable for G_4, G_5, and G_6 RNA in vivo.

RNA editing in trypanosomes is dominated by insertion of uridine residues into the mRNA transcripts [reviewed in (*12*)]. Insertion of multiple uridylates within consecutive guanylates eliminates G-tracts from mature mitochondrial mRNA; in fact polyuridylate (poly(U)) tracts populate fully edited mRNA at the expense of poly(G) tracts (*30*). Also, examination of the maxicircle DNA sequence shows that coding strands for never-edited genes are deficient in poly(G) but highly enriched in poly(U) homopolymers, and to a lesser extent adenosines. Both the small (9S) and large (12S) mitochondrial ribosomal RNAs, and the guide RNAs (gRNAs) of *T. brucei* are also preferentially AU-rich and poor in G-tracts (*56-57*). The low affinity of PPR27 for U-rich and A-rich RNA therefore suggests that direct interaction of this protein with translation competent mRNA (fully edited mRNA and never-edited mRNA), gRNA, and rRNA is less likely. For rRNA, our findings are consistent with Aphasizhev and coworkers' recent report of a ribonuclease- and salt- insensitive association of PPR27 with the *T. brucei* mitochondrial small ribosomal subunit (*7*), a sign of non-RNA mediated anchorage. Pre-edited mRNAs and their often overlapping termini (*16, 58-59*) are therefore the potential sites of PPR27 action.

Mitochondrial ribosomes only translate fully edited mRNA with long poly(A/U) tails, and a few protein factors, including two PPR proteins have already been implicated in the synthesis of these tails (*7*). But as elaborated by Maslov and Agrawal in a recent review, extra factors may be necessary to coordinate creation of translatable mRNA (by cleavage of polycistronic precursor RNA, editing and poly(A/U) addition) and recognition of long poly(A/U) tailed mRNA by the ribosome (*60*). Indeed, PPR27's involvement with polycistronic RNA precursors as well as its preferential affinity for motifs only found in pre-edited mRNA transcripts suggests the interplay between RNA maturation and translation could be more complex than is currently understood.

Among the plausible justifications for the existence of precursor/pre-edited RNA binding protein in the small ribosomal subunit (*7, 23*) are: a) it could upregulate the substrates for the ribosome by stabilizing polycistronic precursors and pre-edited transcripts, and hence bridge RNA maturation and translation; b) it could prepare the small ribosomal subunit for translation by binding to premature translation substrates (polycistronic precursors and pre-edited mRNA), and hence drag the small ribosomal subunit closer to mRNA; c) PPR27's differential affinity for G-tracts of different sizes could specify the residency time of some RNA processing enzymes on their sites of action, which in the case of RNA processing nucleases could prevent stray RNA damage; d) it could prevent premature assembly of the complete translational machinery on premature mRNA by masking potential sites of spurious translation initiation. Evidence from plants (*52-55, 61-64*), humans ((*65-67*), yeast [reviewed in (*60*)], and even trypanosomes (*8-10*), among others, shows that PPR proteins are capable of executing any of these functions. Whether PPR27 performs any of these functions will however be settled after evidence from PPR27 knockdowns and in vivo RNA-protein crosslinking.

REFERENCES

REFERENCES

1. Small, I. D., and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins, *Trends Biochem Sci 25*, 46-47.

2. D'Andrea, L. D., and Regan, L. (2003) TPR proteins: the versatile helix, *Trends Biochem Sci 28*, 655-662.

3. Andres, C., Lurin, C., Small, I.D. (2007) The multifarious roles of PPR proteins in plant mitochondrial gene expression, *Physiologica Plantarum 129*, 14-22.

4. Delannoy, E., Stanley, W.A., Bond, C.S., Small, I.D. (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specific factors in post-transcriptional processes in organelles, *Biochemical Society Transactions*, 1643-1647.

5. Schmitz-Linneweber, C., and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression, *Trends Plant Sci 13*, 663-670.

6. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, *Plant Cell 16*, 2089-2103.

7. Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes, *Mol Cell 42*, 106-117.

8. Pusnik, M., and Schneider, A. (2012) A trypanosomal pentatricopeptide repeat protein stabilizes the mitochondrial mRNAs of cytochrome oxidase subunits 1 and 2, *Eukaryot Cell 11*, 79-87.

9. Mingler, M. K., Hingst, A. M., Clement, S. L., Yu, L. E., Reifur, L., and Koslowsky, D. J. (2006) Identification of pentatricopeptide repeat proteins in Trypanosoma brucei, *Mol Biochem Parasitol 150*, 37-45.

10. Pusnik, M., Small, I., Read, L. K., Fabbro, T., and Schneider, A. (2007) Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes, *Mol Cell Biol 27*, 6876-6888.

11. Lukes, J., Hashimi, H., Verner, Z., and Cicova, Z. (2010) The Remarkable Mitochondrion of Trypanosomes and Related Flagellates, *Structures and Organelles in Pathogenic Protists 17*, 227-252.

12. Lukes, J., Hashimi, H., and Zikova, A. (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates, *Curr Genet 48*, 277-299.
13. Michelotti, E. F., Harris, M. E., Adler, B., Torri, A. F., and Hajduk, S. L. (1992) Trypanosoma brucei mitochondrial ribosomal RNA synthesis, processing and developmentally regulated expression, *Mol Biochem Parasitol 54*, 31-41.
14. Schneider, A. (2001) Unique aspects of mitochondrial biogenesis in trypanosomatids, *Int J Parasitol 31*, 1403-1415.
15. Vanhamme, L., and Pays, E. (1995) Control of gene expression in trypanosomes, *Microbiol Rev 59*, 223-240.
16. Koslowsky, D. J. (2009) Complex interactions in the regulation of trypanosome mitochondrial gene expression, *Trends Parasitol 25*, 252-255.
17. Adhya, S., Basu, S., Bhattacharyya, S. N., Chatterjee, S., Dhar, G., Goswami, S., Ghosh, S., Home, P., Mahata, B., and Tripathi, G. (2003) Mitochondrial differentiation in kinetoplastid protozoa: a plethora of RNA controls, *Differentiation 71*, 549-556.
18. Clayton, C., and Shapira, M. (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias, *Mol Biochem Parasitol 156*, 93-101.
19. Cuervo, P., Domont, G. B., and De Jesus, J. B. (2010) Proteomics of trypanosomatids of human medical importance, *J Proteomics 73*, 845-867.
20. Palenchar, J. B., and Bellofatto, V. (2006) Gene transcription in trypanosomes, *Mol Biochem Parasitol 146*, 135-141.
21. Panigrahi, A. K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R. A., Acestor, N., Myler, P. J., and Stuart, K. D. (2009) A comprehensive analysis of Trypanosoma brucei mitochondrial proteome, *Proteomics 9*, 434-450.
22. Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giege, P. (2008) PPR336 is associated with polysomes in plant mitochondria, *Journal of Molecular Biology 375*, 626-636.
23. Zikova, A., Panigrahi, A. K., Dalley, R. A., Acestor, N., Anupama, A., Ogata, Y., Myler, P. J., and Stuart, K. (2008) Trypanosoma brucei mitochondrial ribosomes: affinity purification and component identification by mass spectrometry, *Mol Cell Proteomics 7*, 1286-1296.
24. Karpenahalli, M. R., Lupas, A. N., and Soding, J. (2007) TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences, *BMC Bioinformatics 8*, 2.

25. Biegert, A., and Soding, J. (2008) De novo identification of highly diverged protein repeats by probabilistic consistency, *Bioinformatics 24*, 807-814.

26. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server, *Bioinformatics 16*, 404-405.

27. Dumon-Seignovert, L., Cariot, G., and Vuillard, L. (2004) The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3), *Protein Expr Purif 37*, 203-206.

28. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures, *Protein Expr Purif 41*, 207-234.

29. Tarun, S. Z., Schnaufer, A., Ernst, N. L., Proff, R., Deng, J. P., Hol, W., and Stuart, K. (2008) KREPA6 is an RNA-binding protein essential for editosome integrity and survival of Trypanosoma brucei, *Rna-a Publication of the Rna Society 14*, 347-358.

30. Hong, M., and Simpson, L. (2003) Genomic organization of Trypanosoma brucei kinetoplast DNA minicircles, *Protist 154*, 265-279.

31. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure, *Nucleic Acids Res 34*, 5402-5415.

32. Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA, *Nucleic Acids Res 37*, 1713-1725.

33. Lane, A. N., Chaires, J. B., Gray, R. D., and Trent, J. O. (2008) Stability and kinetics of G-quadruplex structures, *Nucleic Acids Res 36*, 5482-5515.

34. Cunningham, I. (1977) New culture medium for maintenance of tsetse tissues and growth of trypanosomatids, *J Protozool 24*, 325-329.

35. Koslowsky, D. J., and Yahampath, G. (1997) Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in Trypanosoma brucei are independent events, *Mol Biochem Parasitol 90*, 81-94.

36. Pfanner, N. (2000) Protein sorting: recognizing mitochondrial presequences, *Curr Biol 10*, R412-415.

37. Vonheijne, G. (1986) Mitochondrial Targeting Sequences May Form Amphiphilic Helices, *Embo Journal 5*, 1335-1342.

38. Vonheijne, G., Steppuhn, J., and Herrmann, R. G. (1989) Domain-Structure of Mitochondrial and Chloroplast Targeting Peptides, *European Journal of Biochemistry 180*, 535-545.

39. Hill, J. J., and Royer, C. A. (1997) Fluorescence approaches to study of protein-nucleic acid complexation, *Methods Enzymol 278*, 390-416.
40. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A., and Spada, G. P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures, *Org Biomol Chem 8*, 2683-2692.

41. Viglasky, V., Bauer, L., and Tluckova, K. (2010) Structural features of intra- and intermolecular G-quadruplexes derived from telomeric repeats, *Biochemistry 49*, 2110-2120.

42. Juskowiak, B., Galezowska, E., Zawadzka, A., Gluszynska, A., and Takenaka, S. (2006) Fluorescence anisotropy and FRET studies of G-quadruplex formation in presence of different cations, *Spectrochim Acta A Mol Biomol Spectrosc 64*, 835-843.

43. Rachwal, P. A., and Fox, K. R. (2007) Quadruplex melting, *Methods 43*, 291-301.

44. Dominguez, C., Fisette, J. F., Chabot, B., and Allain, F. H. (2010) Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs, *Nat Struct Mol Biol 17*, 853-861.
45. Salas, T. R., Petruseva, I., Lavrik, O., Bourdoncle, A., Mergny, J. L., Favre, A., and Saintome, C. (2006) Human replication protein A unfolds telomeric G-quadruplexes, *Nucleic Acids Res 34*, 4857-4865.

46. Grygoryev, D., and Zimbrick, J. D. (2010) Effect of quadruplex conformation on radiation-induced formation of 8-hydroxyguanine and unaltered base release in polyguanylic acid, *Radiat Res 173*, 110-118.

47. Ryder, S. P., Frater, L. A., Abramovitz, D. L., Goodwin, E. B., and Williamson, J. R. (2004) RNA target specificity of the STAR/GSG domain post-transcriptional regulatory protein GLD-1, *Nat Struct Mol Biol 11*, 20-28.

48. Fry, M. (2007) Tetraplex DNA and its interacting proteins, *Front Biosci 12*, 4336-4351.

49. Shafer, R. H., and Smirnov, I. (2000) Biological aspects of DNA/RNA quadruplexes, *Biopolymers 56*, 209-227.

50. Phan, A. T., Kuryavyi, V., Ma, J. B., Faure, A., Andreola, M. L., and Patel, D. J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase, *Proc Natl Acad Sci U S A 102*, 634-639.

51. Kotera, E., Tasaka, M., and Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts, *Nature 433*, 326-330.

52. Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing, *Journal of Biological Chemistry 281*, 37661-37667.

53. Okuda, K., and Shikanai, T. (2012) A pentatricopeptide repeat protein acts as a site-specificity factor at multiple RNA editing sites with unrelated cis-acting elements in plastids, *Nucleic Acids Research 40*, 5052-5064.

54. Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts, *EMBO J 28*, 2042-2052.

55. Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein, *P Natl Acad Sci USA 108*, 415-420.

56. Eperon, I. C., Janssen, J. W., Hoeijmakers, J. H., and Borst, P. (1983) The major transcripts of the kinetoplast DNA of Trypanosoma brucei are very small ribosomal RNAs, *Nucleic Acids Res 11*, 105-125.

57. Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R., and Noller, H. (1985) Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: a detailed comparison of the 9S and 12S RNAs from Crithidia fasciculata and Trypanosoma brucei with rRNAs from other organisms, *Nucleic Acids Res 13*, 4171-4190.

58. Aphasizhev, R., and Aphasizheva, I. (2011) Mitochondrial RNA processing in trypanosomes, *Res Microbiol 162*, 655-663.

59. Feagin, J. E. (2000) Mitochondrial genome diversity in parasites, *Int J Parasitol 30*, 371-390.

60. Maslov, A. M., and Agrawal, R. K. (2012) Mitochondrial translation in trypanosomatids, Nucleic Acids and Molecular Biology, In *RNA Metabolism in Trypanosomes* (Bindereif, A., Ed.), Springer-Verlag, Berlin.

61. Fujii, S., Bond, C. S., and Small, I. D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution, *P Natl Acad Sci USA 108*, 1723-1728.

62. Hammani, K., Cook, W. B., and Barkan, A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression, *P Natl Acad Sci USA 109*, 5651-5656.

63. Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R. (2012) An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria, *Plant Cell*.

64. Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K., and Shikanai, T. (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing, *Proc Natl Acad Sci U S A 104*, 8178-8183.

65. Davies, S. M. K., Rackham, O., Shearwood, A. M. J., Hamilton, K. L., Narsai, R., Whelan, J., and Filipovska, A. (2009) Pentatricopeptide repeat domain protein 3 associates with the mitochondrial small ribosomal subunit and regulates translation, *Febs Letters 583*, 1853-1858.

66. Rackham, O., Davies, S. M. K., Shearwood, A. M. J., Hamilton, K. L., Whelan, J., and Filipovska, A. (2009) Pentatricopeptide repeat domain protein 1 lowers the levels of mitochondrial leucine tRNAs in cells, *Nucleic Acids Research 37*, 5859-5867.

67. Sasarman, F., Brunel-Guitton, C., Antonicka, H., Wai, T., and Shoubridge, E. A. (2010) LRPPRC and SLIRP interact in a ribonucleoprotein complex that regulates posttranscriptional gene expression in mitochondria, *Mol Biol Cell 21*, 1315-1323.

CHAPTER 3

STRUCTURE-ACTIVITY STUDIES OF PPR27

All protein used in RNA binding experiments was designed and prepared by me. For six of the eight truncated PPR27 variants examined here, the raw fluorescence anisotropy data used in the computation of averages was contributed by both me and David Dickson, a former undergraduate researcher in the lab. Anisotropy data for wild type PPR27 and one truncated variant (ΔCR1) were collected by David Dickson. All other experiments were performed solely by me and I merged and analyzed all the data. Finally, Mr. Kermit Johnson of the campus Max T. Rogers NMR Facility provided assistance and training in NMR data acquisition on the 900 MHz spectrometer.

ABSTRACT

Pentatricopeptide repeat (PPR) proteins have recently emerged as key players in posttranscriptional RNA processing in mitochondria. These proteins are therefore critical to organisms such as trypanosomes where the mitochondrion drives adaptation to different host environments, and have been touted as potential drug targets. Understanding how these proteins contact their RNA is therefore important. Unfortunately, high resolution studies remain unsuccessful due to poor protein solubility and other factors. Hence, low resolution structure-activity studies such as by mutagenesis remain important, as are efforts to improve protein solubility. Here, a comprehensive analysis of the RNA binding activity and solubility of PPR27 constructs deprived of one or more PPR motifs was done. Deletion of PPR motifs modestly reduced the affinity of PPR27 for RNA, and measurable RNA binding still occurred with as few as two intact PPR motifs. Except for the construct with only two PPR motifs which suffered a 7-fold drop in RNA binding affinity for G_{12} ssRNA, the RNA binding K_d of truncated variants ranged between 2.5- and 3.5-fold that of wild type. Thus, all the PPR modules in PPR27 are seemingly involved in RNA contacts. Finally, screening of truncated PPR27 variants identified one soluble enough to attain structural biology concentrations in the presence of stabilizers.

3.1 INTRODUCTION

Thirteen years after their discovery, pentatricopeptide repeat (PPR) proteins have become household names in the fields of mitochondrial and chloroplast RNA processing. In plants where these proteins have been most studied, there are hundreds of them per species, anchoring diverse functions involving sequence specific RNA recognition, including RNA stabilization, splicing, editing, polyadenylation, and translation (*1*). With the exception of trypanosomes which have at least three dozen (*2*), non-plant species contain few PPR proteins (*3*). Gene knockdown studies in *Trypanosoma brucei* have shown that its PPR proteins are not functionally redundant despite their abundance, causing adverse phenotypes ranging from growth retardation to growth arrest and death (*4-5*). Thus, trypanosome PPR proteins are putative drug targets, offering a glimmer of hope for the development of safer and more potent anti-trypanosomal drugs to replace the highly toxic and sub-efficacious treatments currently in use (*6-7*).

To be of utility in rational drug design, the mechanism of ligand (RNA) binding by PPR proteins ought to be understood, both at individual and family level. Experimentally derived atomic structures are the gold standard for elucidating the details of biomacromolecular ligand binding modes. However, for classical PPR proteins (*8*), poor protein solubility (*3, 9-10*) has hindered structural biology. Thus efforts to understand the molecular mechanism of RNA binding by PPR proteins have centered on biochemical characterization of mutant proteins. A major tendency has been to examine the RNA binding activity of truncated proteins lacking one or more PPR motifs. Outcomes of these studies have however not been consistent among members of the protein family. While the RNA binding activity is spread across all PPR motifs in most PPR proteins examined so far (*11-13*), for some such as the mammalian LRP130 (LRPPRC) it is localized to a C-terminal domain possessing only two of its 22 PPR motifs (*14-15*). Even where all PPR motifs

contribute to RNA binding, different motifs bind RNA with different strengths (*11*). Nevertheless, the evolutionary plasticity of the PPR motif and the variable composition and organization of PPR motifs in different PPR proteins means mechanistic biochemistry of their RNA binding is best done on case-by-case basis.

In this Chapter, we report the RNA binding activity of truncated constructs of the 27 kDa *T. brucei* PPR protein (PPR27) and discuss mechanistic implications in context of its computationally modeled 3-dimensional structure. Secondly, in an effort towards high resolution structural biology, we explore the effect of various PPR motif deletions on PPR27 protein solubility. Through this approach, we identify one construct with promising properties including high solubility.

3.2 MATERIALS AND METHODS

3.2.1 Materials

L-arginine (L-Arg), L-glutamate (L-Glu), deuterium oxide (D_2O) and DEAE sephadex® were obtained from Sigma-Aldrich (St. Louis, MO). ^{15}N-labeled ammonium chloride was obtained from Cambridge Isotope Laboratories (Andover, MA). TEV protease was initially obtained from Invitrogen but was later produced in-house from a plasmid donated to us by Dr. Honggao Yan, Michigan State University.The rest of the materials and their sources were described in chapter 2.

3.2.2 Multiple sequence alignment of *T. brucei brucei* PPR27 and its homologues

PPR27 homologues were searched in NCBI's nucleotide collection using the basic local alignment search tool (BLAST) (*16*). The amino acid sequences of PPR27 and its homologues in

FASTA format were then saved as text file and uploaded into the TCOFFEE sequence alignment server (http://igs-server.cnrs-mrs.fr/Tcoffee). The resultant alignment was formatted in ESPript 2.2 (http://espript.ibcp.fr/ESPript).

3.2.3 Cloning and site-directed mutagenesis

Based on the positions of PPR motifs along the PPR27 backbone as inferred from repeat prediction and secondary structure modeling (chapter 2), ten truncated PPR27 variants lacking one, two, three, or four PPR modules were designed (**Figure 3.1**). Mature wild type PPR27 was cloned into the pMalTEV-E30 plasmid to create an MBPPPR27 fusion as described in chapter 2. PPR motifs were deleted and a His_6 tag inserted either at the C-terminus of PPR27 or between MBP and PPR27 to create His_6-tagged MBP fusions using QuikChange® site-directed mutagenesis (SDM) as described in chapter 2. Primers used for SDM were designed by the QuikChange® Primer Design tool and are in **Table 3.1**. For ΔNR1, ΔNR2, NR3, ΔCR2, ΔCR3, and ΔCR4, a His_6 tag was inserted at the PPR27 C-terminus of already truncated constructs. For ΔCR1, ΔNCR11, ΔNCR12, and ΔNCR21 C-terminal truncation and insertion of His_6 tag were done concomitantly in a single SDM reaction with one set of primers for each.

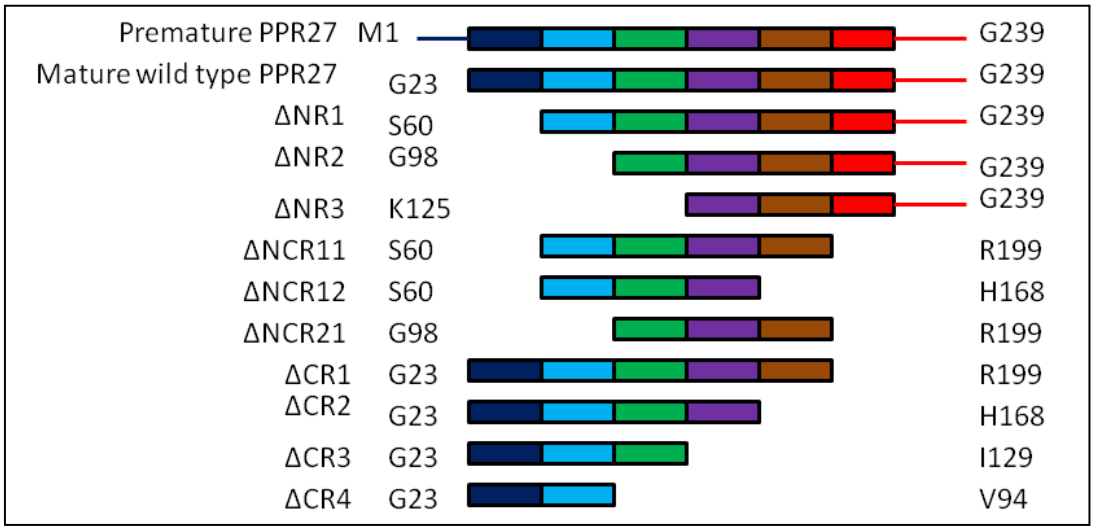


Figure 3.1: **Schematic diagram showing the motif composition of various truncated PPR27 constructs relative to the wild type.**

Each colored solid block is a PPR motif.

Table 3.1: Primers used for site-directed mutagenesis

Activity	Primer code*	Primer sequence (5' to 3')
Truncate WT to ΔNR1	CGH53_F	CTGTACTTCCAGGGATCCTCCCACTACACAA-ACATC
	CGH54_R	GATGTTTGTGTAGTGGGAGGATCCCTGGAAG-TACAG
Truncate WT to ΔNR2	CGH55_F	GTACTTCCAGGGATCCGGTTGCGCTCTTGCAG
	CGH56_R	CTGCAAGAGCGCAACCGGATCCCTGGAAGTAC
Truncate WT to ΔNR3	CGH57_F	AACCTGTACTTCCAGGGATCCAAGTTGGATAG-TATTTGTTAC
	CGH58_R	GTAACAAATACTATCCAACTTGGATCCCTGGA-AGTACAGGTT
Truncate WT to ΔCR2	CGH63_F	TTCCACACACGTATACGCACTGACTCGAAGCGG
	CGH64_R	CCGCTTCGAGTCAGTGCGTATACGTGTGTGGAA
Truncate WT to ΔCR3	CGH61_F	GGCAAGCTAAAGTTGGATAGTATTTGATACCT-TGCGCTCATTAA
	CGH62_R	TTAATGAGCGCAAGGTATCAAATACTATCCAA-CTTTAGCTTGCC
Truncate WT to ΔCR4	CGH59_F	GAACATCCGACCAGATGTCTGAGGTGTCGGTT-GCGCTCTTG
	CGH60_R	CAAGAGCGCAACCGACACCTCAGACATCTGGT-CGGATGTTC
Insert His_6 tag at C-terminus ΔNR1, ΔNR2, and ΔNR3	CGH69_F	CCACTTTACCTCGTGGTCACCACCACCACCAC-CACTGAGTCGACCTGCAG
	CGH70_R	CTGCAGGTCGACTCAGTGGTGGTGGTGGTGGT-GACCACGAGGTAAAGTGG
Create ΔCR1, ΔNCR11, and ΔNCR21 with His_6 at C-terminus from WT, ΔNR1, and ΔNR2 respectively	CGH105_F	GGTGTCCGATCGGCACCACCACCACCACCACT-GAGGACGGGAGGCAC
	CGH106_R	GTGCCTCCCGTCCTCAGTGGTGGTGGTGGTGG-TGCCGATCGGACACC

Table 3.1 (cont'd).

Create ΔCR2 and ΔNCR12 with His_6 tag at C-terminus from ΔCR2 and ΔNR1 respectively	CGH75_F CGH76_R	ACACACGTATACGCACCACCACCACCACCACA-CTGATTACTCGAAGCGGCGG CCGCCGCTTCGAGTAATCAGTGGTGGTGGTGGT-GGTGGTGCGTATACGTGTGT
Insert His_6 tag at ΔCR3 C-terminus	CGH73_F CGH74_R	GGCAAGCTAAAGTTGGATAGTATTCACCACCACC-ACCACCACTGATGTTACCTTGCGCTCAT ATGAGCGCAAGGTAACATCAGTGGTGGTGGTGGT-GGTGAATACTATCCAACTTTAGCTTGCC
Insert His_6 tag at ΔCR4 C-terminus	CGH71_F CGH72_R	CATCCGACCAGATGTCCACCACCACCACCACC-ACTGAGTTGGTGTCGGTTGCG CGCAACCGACACCAACTCAGTGGTGGTGGTGGT-GGTGGACATCTGGTCGGATG
Insert His_6 tag between MBP and ΔCR3	CGH67_F CGH68_R	GGATTTCAGAATTCGGATCTCACCACCACCA-CCACCACGAAAACCTGTACTTCCAGGG CCCTGGAAGTACAGGTTTTCGTGGTGGTGGT-GGTGGTGAGATCCGAATTCTGAAATCC
Convert $MBPHis_6$ΔNR2 to $MBPHis_6$	CGH77_F CGH78_R	TACTTCCAGGGATCCTGATGCGCTCTTGCAGCC GGCTGCAAGAGCGCATCAGGATCCCTGGAAGTA

3.2.4 Protein expression and purification

Overexpression of the MBP fusion proteins of the truncated variants was assessed mostly in BL21 (DE3) and occasionally in Rosetta (DE3) *E. coli*. The PPR27pMalTEV-E30 plasmid was transformed into *E. coli* by heat shock. LB agar containing 50 μg/ml ampicillin for BL21 (DE3) and both ampicillin (50 μg/ml) and chloramphenicol (34 μg/ml) for Rosetta (DE3) was used to select for transformants. Single colonies were used to generate 10 ml overnight cultures in LB medium with appropriate antibiotics, 20 ml of which were used to seed 980 ml of LB broth containing appropriate antibiotics. After incubation in a shaker at 37 ^{o}C, 250 rpm to an OD_{600} of about 0.7 protein overexpression was induced with 0.1 mM IPTG. For ΔNR2, ΔCR1, ΔCR2, ΔCR3, ΔNCR4, ΔNCR12, and ΔNCR21, post-induction cultures were incubated at 30 ^{o}C whereas for ΔNR1 they were incubated at a lower temperature (22 ^{o}C) because its soluble yield was low at 30 ^{o}C. After six hours, post-induction cultures were harvested by centrifugation in a SORVALL® RC5 PLUS centrifuge (Kendro Laboratory Products, Newtown, CT) at 5000 rpm, 4 ^{o}C, for 15 minutes. Cells were either immediately lyzed for purification or stored at -80 ^{o}C till required. ^{15}N-labeled MBP fusion protein of ΔCR3 for NMR spectroscopy was overexpressed in Rosetta(DE3) following the method of Bracken and coworkers (*17*). Two 10 ml LB media with the antibiotics were inoculated with single colonies and incubated overnight at 250 rpm, 37 ^{o}C after which the two were pooled and used to seed 980 ml of LB broth containing the antibiotics. After incubation in a shaker at 37 ^{o}C, 250 rpm to mid-log phase (OD_{600} of about 0.7), bulk LB cultures were harvested by centrifugation at 5,000 rpm, 15 min, 4 ^{o}C. The pellet was then

resuspended in 1X minimal salts devoid of glucose, Mg^{2+}, Ca^{2+}, and NH_4^+ to wash off residual LB. New pellet was resuspended in 1.0 L of ^{15}N-labeled minimal medium (1X M9 minimal salts, $^{15}NH_4Cl$, pH 7.2, 0.4% (w/v) glucose, 2.0 mM $MgSO_4$, 0.1 mM $CaCl_2$) and incubated at 225 rpm, 37 ^oC for one hour followed by induction of protein expression with 0.2 mM IPTG at 30 ^oC for 4 hours. Cells were harvested as above and kept at -80 ^oC.

All protein purification was by gravity at room temperature. For RNA binding assays, fusion proteins with a His_6 tag at the C-terminus were purified by tandem Ni-NTA and amylose columns as described in chapter 2. On the other hand, fusion proteins for TEV protease reaction were purified by amyolse only by adaptation of the manufacturer's instructions. Briefly, cells were lyzed by microsonication in at least 5 ml of buffer MWB1 (20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 1.0 mM EDTA, 1.0 mM NaN_3) per gram of cells after which lysate was clarified by centrifugation at 14,000 rpm for 25 min at 4 ^oC. Supernatants were then loaded onto an amylose column equilibrated with MWB1, and purified under gravity at room temperature. After drainage of unbound impurities in the flow through and washes with lysis buffer, MBPPPR27 fusion proteins were eluted from the column in maltose containing TEV protease cleavage buffer or MEB4 (20 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, 10 mM maltose). TEV protease cleavage was then done in the presence of 1.0 mM dithiothreitol (DTT) with or without 0.05% (w/v) n-dodecyl-β,D-maltopyranoside (or dodecyl maltoside, DDM) (Anatrace-Affymetrix, Maume, OH) at 8 ^oC for at least 12 hours. At least 0.1 units (or 0.01 μg) of TEV protease were required for every 1.0 μg of fusion protein in the cleavage reaction for optimal yield of free PPR27. Free

PPR27ΔCR3 was subsequently purified from the TEV protease reaction by DEAE sephadex anion exchange chromatography in the presence of 0.05% (w/v) DDM to prevent on-column aggregation. Briefly, the reaction mixture was dialyzed against at least 50-fold volumes of buffer IEXB (20 mM Tris-HCl, pH 8.0) for at least 12 hours at 4 ^{o}C to remove EDTA and salt, using a 7K MWCO SnakeSkin dialysis tubing (Thermo Scientific, Rockford, IL). The dialysate was then supplemented with more DDM equivalent to 0.05% (w/v) after which it was loaded onto the column equilibrated with buffer IEXB containing 0.05% DDM. Free PPR27 comes out in the flow through and very low salt washes while MBP, uncleaved MBPPPR27, and TEV protease emerge later during the salt gradient. To attain high concentrations of free ΔCR3 prior to size exclusion chromatography (SEC) and nuclear magnetic resonance (NMR) spectroscopy, sample was diluted with an equal volume of 1.0 M L-Arg as a stabilizer (*18*). Protein in 500 mM L-Arg was then concentrated, first to a volume below 20 ml by the fast filtering, conical membrane amicon centrifugal filters (stirred cell pressure concentrator explored with no advantage) at 5,000 rpm, 4 ^{o}C. Then the residue was dialyzed against over 100 volumes of buffer TSBRE50 (50 mM NaH_2PO_4, 50 mM L-Arg, 50 mM L-Glu, pH 6.5, 100 mM NaCl) for at least 12 hours. Then DTT was added to 3.0 mM followed by final concentration to 0.5-1.0 ml by the slow filtering, flat membrane centricon centrifugal filters, each of 3K MWCO (Millipore, Cork, Ireland) at 6,000 rpm, 4 ^{o}C.

3.2.5 Electrophoretic mobility shift assays (EMSA)

5'-FLUO labeled G_{12} ssRNA was used as the ligand. Agarose, rather than polyacrylamide gels were used for electrophoresis because the latter were associated with strong well retention of the

RNA as the protein concentration rose. For fluorescent RNA bands to be detected, 150 nM G_{12} ssRNA was used in the reaction mixture as explained in chapter 2. The composition of the RNA binding reaction was the same as described in chapter 2. 25 mM Tris, 250 mM glycine was used as running buffer. A VersaDoc™ MP 4000 fluorescence imager (Bio-Rad Laboratories) was used to scan the gels.

3.2.6 Fluorescence polarization spectroscopy

The anisotropy of 20 nM 5'-FLUO labeled G_{12} ssRNA was assessed in the presence of increasing concentration of PPR27 truncated variants. Following the observation in chapter 2 that thermal exposure of G_{12} RNA reduced its affinity for PPR27, the RNA used here was accordingly thawed slowly on ice-cold water. 450 µl RNA binding reactions were set up from dialysis buffer, protein, RNA binding buffer, and RNA for each protein concentration as described in chapter 2. After at least 15 minutes of equilibration at room temperature, polarized fluorescence data was collected as described in chapter 2.

Fluorescence anisotropy (r) was computed from raw data using equation (i), where G is the grating factor, I_{vv} and I_{vh} are vertically polarized and horizontally polarized emission intensities respectively, following vertically polarized excitation.

$$r = \frac{I_{vv} - GI_{vh}}{I_{vv} + 2GI_{vh}} \quad \text{---------- (i)}$$

$$f(c) = \frac{nc}{K_d + c}$$ ---------- (ii)

For each protein, average anisotropies of at least three datasets were normalized and used to determine the K_d by nonlinear fitting with IGOR Pro 6 using equation (ii), where c is the protein concentration and n is the apparent stoichiometry of the interaction.

3.2.7 PPR27 3-dimensional (3D) structure modeling

To gain further insights into the molecular mechanism of PPR27 RNA binding, the 3-dimensional (3D) structure of PPR27 was modeled computationally. Fold recognition rather than sequence homology was used to generate templates for modeling PPR27 3D-structure due to the paucity of structures with high sequence identity to PPR27 in the protein data bank (pdb). The Fold & Function Assignment System (FFAS03) (*19*) (http://ffas.ljcrf.edu/ffas-cgi/cgi/ffas.pl) was used. Templates with significant similarity to PPR27 according to the FFAS statistical scores (*19*) were then used to model PPR27 3-dimensional structure with SCWRL (*20*). Finally, the best model was selected by assessment of the stereochemical quality using WHAT-IF (*21*) and PROCHECK (*22*), and the extent of target sequence coverage by the model.

3.2.8 Circular dichroism (CD) spectroscopy of free PPR27ΔCR3

Free PPR27ΔCR3 was concentrated from to 15 ml to 3 ml to reduce volume followed by dialysis against over 300 volumes of buffer CD2 (10 mM NaH_2PO_4, pH 7.5, 50 mM NaCl) at 4 ^{o}C for 4 hours using Slyde-A-Lyzer cassettes (Thermo Scientific). CD spectroscopy was performed on 0.43 mg/ml protein at 25 ^{o}C with a 0.5 nm step size using a Chirascan CD spectrometer (Applied

Photophysics, Surrey, UK) and 0.1 mm path length cuvette. Each sample was scanned three times and the traces averaged. The secondary structure composition of ΔCR3 was then estimated by spectral deconvolution using the DICHROWEB program (*23*).

3.2.9 Size exclusion chromatography of ΔCR3

Concentrated protein at about 5 mg/ml was passed through a 0.45 m membrane syringe filter (Millipore) to remove particulates after which 500 µl was injected into an AKTA fast performance liquid chromatography (FPLC) system (GE Life Sciences, Pittsburgh, PA) for SEC (gel filtration chromatography) through a TSK-GEL® G2000SW column (TOSOH Bioscience, Montgomeryville, PA) at a flow rate of 2.0 ml/min at 8 ^{o}C. The SEC running buffer was TSBR500 (50 mM NaH_2PO_4, 500 mM L-Arg, pH 8.5, 100 mM NaCl), employing the L-Arg as stabilizer (*24*).

3.2.10 Nuclear Magnetic resonance (NMR) spectroscopy of free ^{15}N-PPR27ΔCR3

To further assure that free ΔCR3 remained unaggregated in 50 mM L-Arg and 50 mM L-Glu at NMR scale concentrations (sub-mM to mM), a 2D ^{1}H,^{15}N-TROSY (transverse relaxation optimized spectroscopy)-HSQC (heteronuclear single quantum coherence) spectrum of concentrated protein was acquired on a 900 MHz Bruker Avance NMR spectrometer (Bruker BioSpin Corporation, Billerica, MA) equipped with a cryoprobe. The ^{1}H spectrometer frequency and sweep width were 899.7992 MHz and 15,290.5199 Hz while those for ^{15}N were 91.1861 MHz and 3191.5130 Hz. Free ΔCR3 was purified and concentrated to 0.52 mM as earlier described. A sample of 0.46 mM protein, 10% (v/v) D_2O, and 0.2 mM DSS (4,4-dimethyl-4-

silapentane-1-sulfonic acid) reference standard was loaded into a 5.0 mm shigemi reduced-volume NMR tube (Shigemi, Inc., Allison Park, PA). 64 transients, each with 1024 and 512 complex points in the 1^{st} and 2^{nd} dimensions respectively were acquired for 8 hours at 298 K. NMR data was processed with a line broadening of 5 Hz using Felix 2002 software (Felix NMR Inc., San Diego, CA).

3.3 RESULTS

3.3.1 Deletion of PPR motifs reduces but does not abrogate PPR27 RNA binding activity

Within the trypanosomes, PPR27 is highly conserved (**Figure 3.2**) and besides the N-terminal signal peptide and a small C-terminal disordered peptide, its core is exclusively PPR motifs (**Figure 3.1**). To gain insights into the mechanism of RNA binding by PPR27, the contribution of its PPR motifs in RNA binding was analyzed by both EMSA and fluorescence anisotropy. For each construct, recombinant MBP fusion protein of at least 95% purity (**Figure 3.3**) and negligible nucleic acid contamination (U.V A_{260}:A_{280} was typically < 0.8) was used. All PPR27 truncated variants, including the one with only two PPR motifs (out of six) retard the electrophoretic mobility of G_{12} ssRNA (**Figure 3.4**). Fluorescence anisotropy revealed that with the exception of the construct lacking 4 PPR motifs whose RNA binding affinity is significantly reduced, deletion of 2-3 motifs only modestly reduces functional activity , with a slightly larger effect when N-terminal PPR motifs are deleted (**Figure 3.5 and Table 3.2**). As the number of PPR motifs increased, the reductions in standard free energy of binding were less than additive (**Table 3.2**). Thus, though all PPR motifs appear to be involved in RNA binding, the energetic contribution of each PPR motif to the interaction is not uniform.

```
Tbbrucei  MHPRLSGQGVFPRYGGGWMHQFGHVYALQQELLNAEANNWFRAMELWHTARH
Lbrazilie MHPRTSGQGLFPRYGGGWMHQFGHVYALQQELLDPSASNWFRAVELWHTARQ
Linfantum MHPPTSGQGLFPRYGGGWMHQFGHVYALQQELLDPSANNWFRAVELWHTARQ
Tcmarinke MHPRVSGQGVFPRYGGGWMHQFGHVYALQQELLNPQANNWFRAVELWHTARH
Tcbrener  MHPRASGQGVFPRYGGGWMHQFGHVYALQQELLNPEASNWFRAVELWHTARH
Tvivax    MHPRMSGQGIFPRYGGGWMHQFGHVYALQQELLSTQANNWFRAVELWHTARH
          ***  ****:**********************. .*.*****:*******:

Tbbrucei  EGVAMNVSHYTNILRQCVPPTAWEASLMVLKQMRRENIRPDVVGVGCALAAC
Lbrazilie EGVAMNVSHFTSILRQCIQPGAWVQSLQVLEQMKRENIRPDVVGVGCALAAC
Linfantum EGVAMNVSHFTSILRQCVHPGAWAQSLQVLKQMKRENVRPDVVGVGCALAAC
Tcmarinke EGVALNSAHYTSILRQCVQPKAWEASLRVLRQMQRDGIRPDVVGVGCVLATC
Tcbrener  EGVALNSAHYTSILRQCVQPKAWEASLRVLRQMQREGIRPDVVGVGCVLATC
Tvivax    EGVALNSSHYTNILRQCVPAKAWEASMLVLRQMQREGIRPDVVGVGCALASC
          ****:* :*:*.*****:   **  *: **.**:*: :*********.**:*

Tbbrucei  ADANRSEEVEKVFSCFSGKLKLDSICYLALIKAKMSRGGYAEALAVGREQDA
Lbrazilie VEGMHYGEVEDIFAAFHKTMLLDSICYLALIRARQESGNAKAAIAAGRQQQA
Linfantum VEGKHYDEVENIFATFHKTMLLDSVCHLALIRARQESGNAKGAIAAGRQQQA
Tcmarinke ADADRIREVEEVFDEFSGKMALDSVCYLALIKAKMAGGKWKEAIAVGKRQEA
Tcbrener  ADADRISEVEEVFDEFSGKMALDSVCYLALIKAKMAGGRWKEAIAVGKRQEA
Tvivax    ADAGRISEVEKTFSKFSEKMKLDSVCYLALIKARMSQGRFAEALVAGRQQEA
          .:. :  ***. *  *.: ***:*:****:*:   *    *:..*:.*:*.

Tbbrucei  DGVPFLPHTYTHLLEAAEVADDGEMALELARRMSSEQWPVSDRGREALKKLS
Lbrazilie EQVPFLPYTYTHLLEACNAEDDSAYAMELVRNMHSEQWAPSQRAMLAIEDLG
Linfantu  DQVPFLPYTYTHLLEACNAEDDAAYAMELVHNMHSEQWAPSQCAMLAIEDLG
Tcmarinke DGVHFLPYTYTHLLEAAKEADDGEFGLELVRRMKREQWELSERGRVAFKKLC
Tcbrener  DGLHFLPYTYTHLLEAAKEADDGEFGLELVRRMKREQWELSERGRVAFKKLC
Tvivax    SGTPLLPYTYSHLLEAAEKADDDVYALELVRRMSAEQWPLSESGRKAVKKLS
          .   :**:**:*****.:  **   .:**.:.*  ***  *: .  *.:.*

Tbbrucei  TRHHWEEVAEYKQFVAVEDHKLSLPPTLPRG-------
Lbrazilie RRHKGLATVSRPLLSTSDKTDGALRSLDASKCLPRLQ-
Linfantum RRHERLAAFSRPLLSASDKAHSALPSLDASKRLFPPEW
Tcmarinke VLHNWEA--EFDRLAGFSEEDASRPQLPETSL------
Tcbrener  VLHNWEA--EFDRLAGFSEEGVSRPQLPERSL------
Tvivax    ARHGWDV-AEYEQLVGNAHLAGSTRQLTDK--------
            *      .   :    .   :
```

Figure 3.2: Multiple sequence alignment of *T. brucei* PPR27 and its close homologues.

Star (*), absolutely conserved residues; colon (**:**), highly conserved residues; fullstop (**.**), moderately conserved residues; all others, less conserved residues. Amino acids 1-239 were aligned by Clustal Omega followed by manual adjustments.

Figure 3.2 (cont'd).

Tbbrucei, *Trypanosoma brucei brucei*; Lbrazilie, *Leishmania braziliensis*; Linfantum, *Leishmania infantum*; Tcm, *Trypanosoma cruzi marinkellei*; Tcbrener, *Trypanosoma cruzi* strain Brener; Tvivax, *Trypanosoma vivax*.

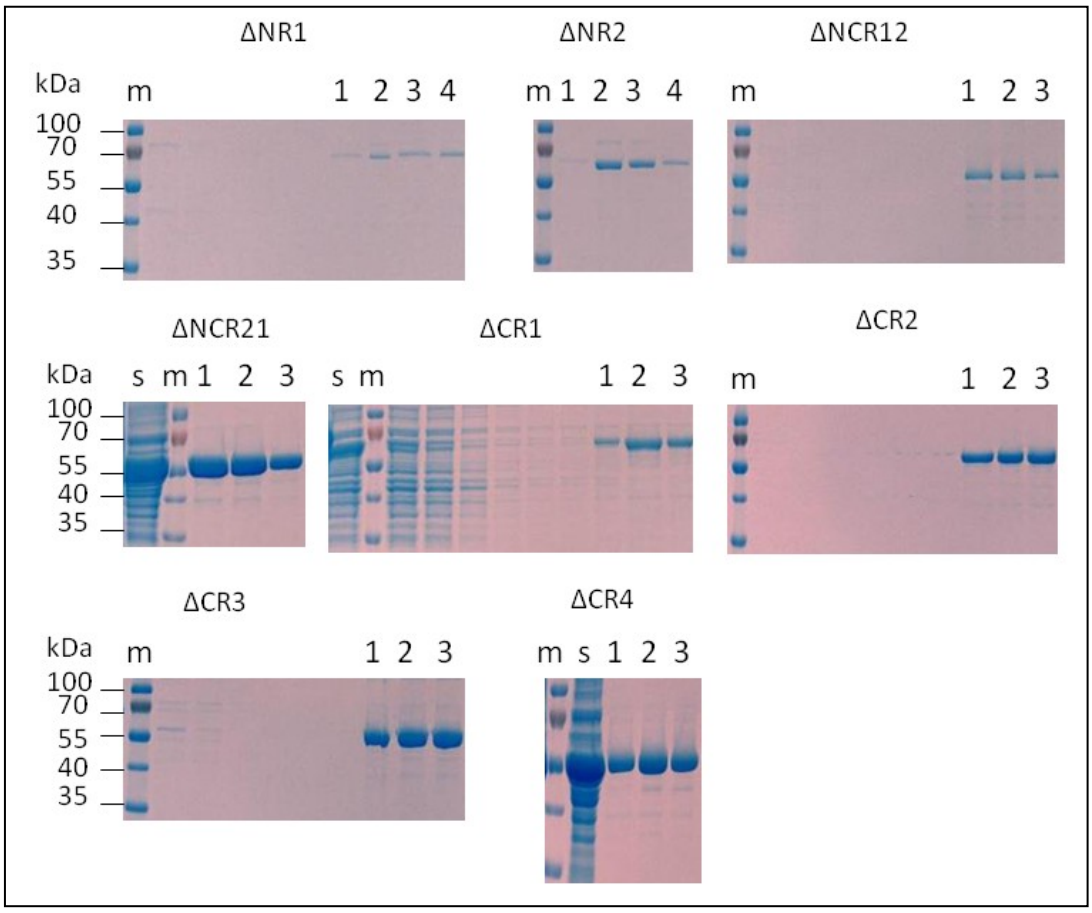


Figure 3.3: SDS-PAGE of MBP fusion proteins of truncated PPR27 constructs with a C-terminal His_6 tag.

Shown are the pure proteins in the eluate fractions 1, 2, 3, (and 4 where applicable) following Ni-NTA purification alone (ΔCR1) or tandem Ni-NTA and amylose affinity purification (the rest of the gels); m, size marker; s, supernatant. Unlabeled lanes between the marker and eluates are for wash fractions.

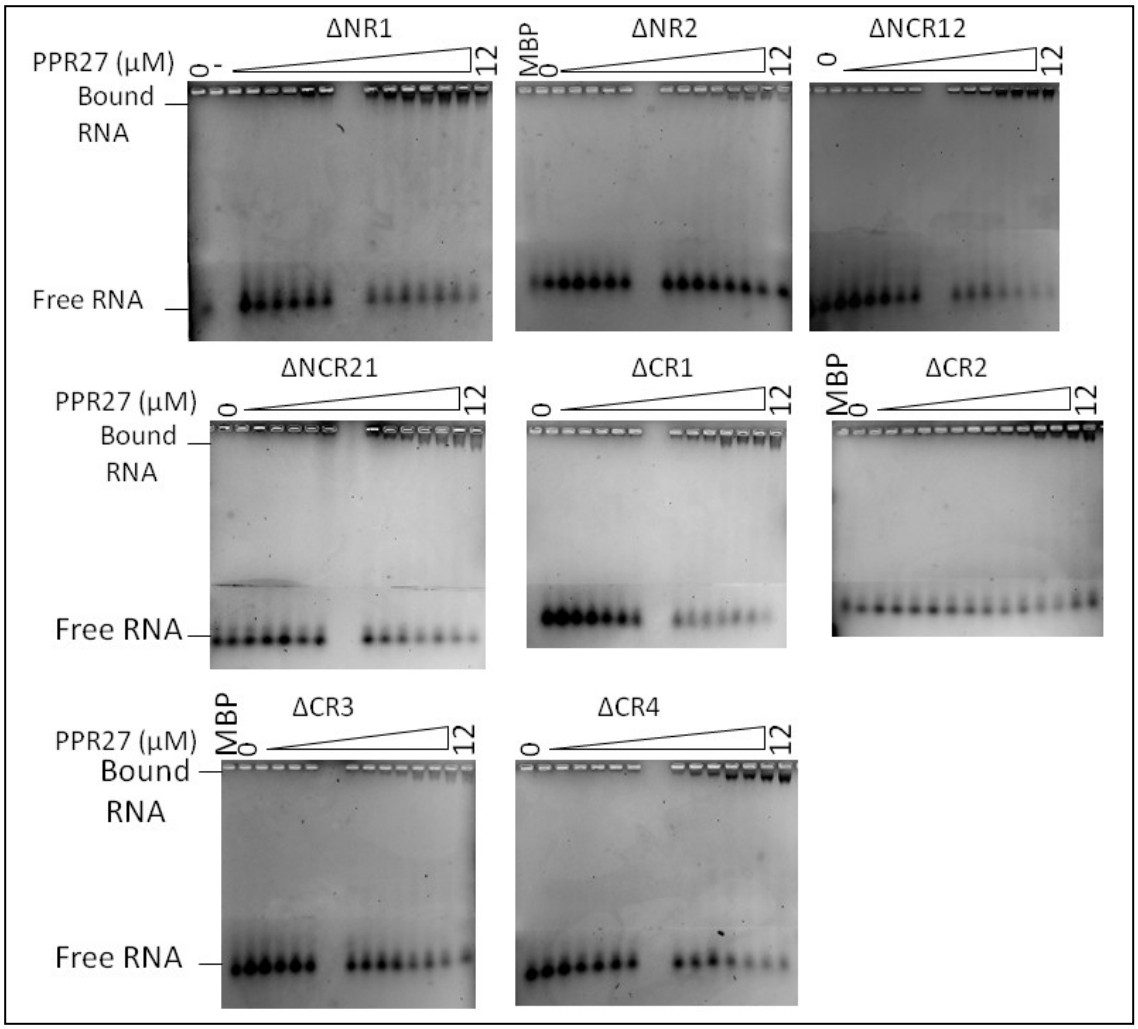


Figure 3.4: 2 % tris-glycine agarose gels showing how the electrophoretic mobility of G_{12} ssRNA is retarded by increasing concentrations of truncated PPR27 variants.

Each test reaction contained 0.15 μM 5'-Fluorescein labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, 0.1 μg/μl BSA, 0.4 U/μl RNasin), and MBPPPR27. All eight PPR27 constructs caused increased retardation of RNA mobility with concentration. In the ΔNR2, ΔCR2, and ΔCR3 gels, a reaction mixture of RNA and MBP was loaded as negative control.

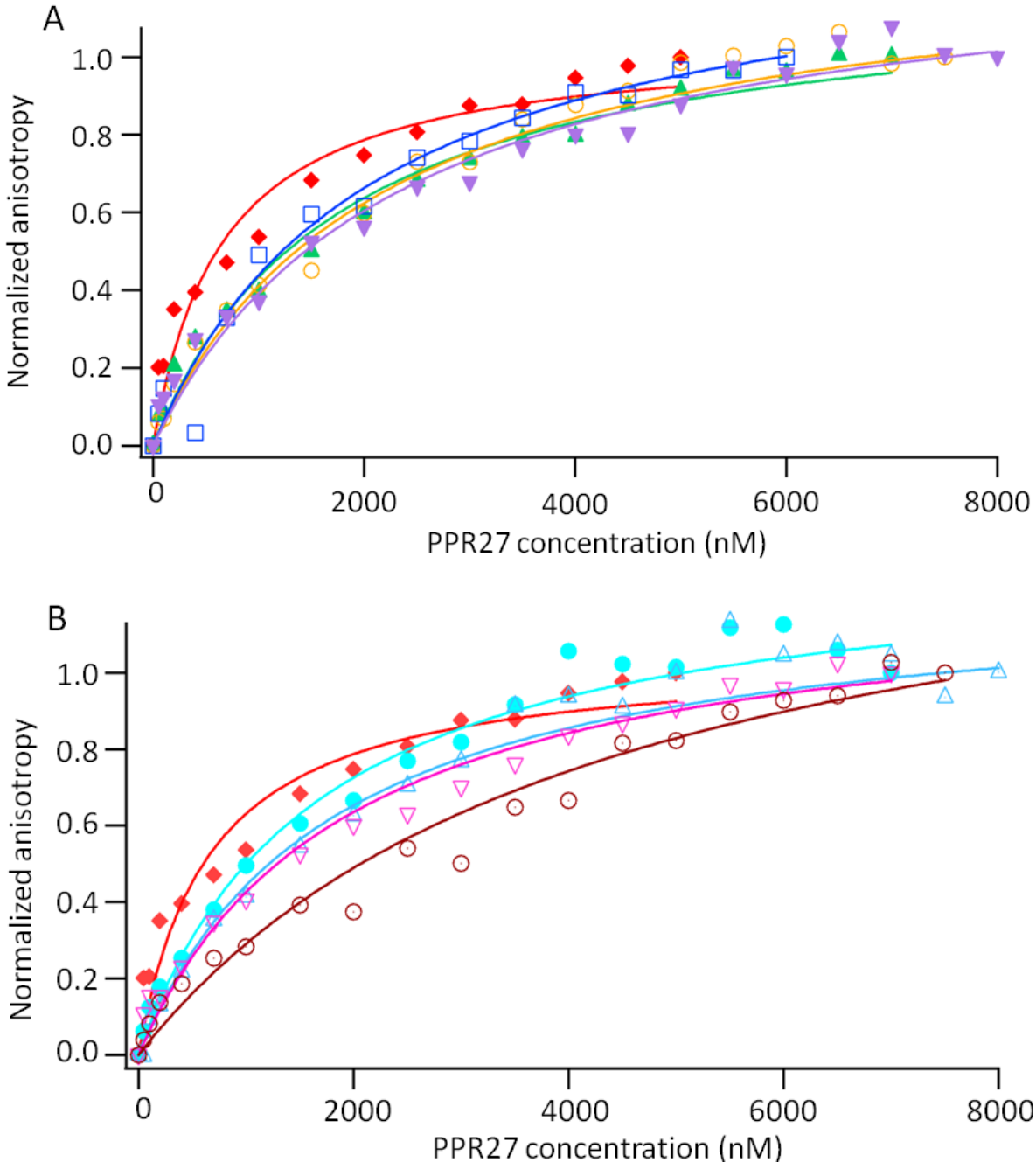


Figure 3.5: Nonlinear plots used in the quantitation of the K_d for interaction of the various PPR27 truncated constructs with G_{12} ssRNA.

Figure 3.5 (cont'd).

(A) Plots for ΔNR1 (upward closed triangles), ΔNR2 (open circles), ΔNCR12 (open squares), and ΔNCR21 (inverted closed triangles) relative to wild type (closed squares). (B) Plots for ΔCR1 (closed circles), ΔCR2 (triangles with central dot), ΔCR3 (inverted open triangles), and ΔCR4 (circles with central dot) relative to wild type (closed squares). Reactions used for FA experiments contained 20 nM 5'-FLUO labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT), and MBP fusions of truncated PPR27.

Table 3.2: Thermodynamics of G_{12} ssRNA interaction with various PPR27 constructs

PPR27 construct	Dissociation constant K_d* (μM)	Standard free energy-change, ΔG° (kJ/mol)**	$\Delta\Delta G^o$ (kJ/mol) relative to wild type PPR27
Wild type	0.61 ± 0.12	-35.5	0
ΔNR1	1.77 ± 0.10	-32.8	2.7
ΔNR2	2.15 ± 0.09	-32.3	3.2
ΔCR1	1.66 ± 0.09	-33.0	2.5
ΔCR2	1.80 ± 0.11	-32.8	2.7
ΔCR3	1.93 ± 0.12	-32.6	2.9
ΔCR4	4.32 ± 0.30	-30.6	4.9
ΔNCR12	2.06 ± 0.13	-32.4	3.1
ΔNCR21	2.33 ± 0.16	-32.1	3.4

* K_d values were obtained from nonlinear regression of FA data as shown in **Figure 3.5** above.

* Values of ΔG° were computed from:

$$\Delta G^o = -RT \ln\left(\frac{1}{K_d}\right).$$

3.3.2 PPR27's putative RNA binding face is contributed by all its PPR motifs

PPR27 3D structure modeling predicts a super-helical conformation in which six anti-parallel α-helical pairs form an inner concave face and an outer convex face (**Figure 3.6**) similar to predicted structures of other classical PPR proteins (*13, 25-26*). The concave face is also morphologically similar to the peptide binding surface of the structurally homologous TPR proteins (*13, 25-26*) and the RNA binding face of the helical repeat Pumilio/*fem-3* mRNA binding factor homology (PUF) domains (*27*). Relative to TPR proteins and PUF domains however, the PPR27 model and those of other classical PPR proteins (*13, 25-26*) show more curvature with the concave face approximating a cavity or groove (**Figure 3.6**). All PPR motifs in PPR27 contribute to the concave groove, an arrangement which is visible in the PPR domains of the recently solved crystal structures of human mitochondrial RNA polymerase (POLRMT) (*28*) and *Arabidopsis thaliana* protein only RNase P (PRORP1) (*29*) (Chapter 1 for details). Thus, like the RNA binding data, PPR27's 3D model also suggests that all its PPR modules are involved in RNA contacts. The best 3D model was generated using the crystal structure of TTC0263 (*30*), a *Thermus thermophilus* tetratricopeptide repeat (TPR) protein with 14% sequence identity to PPR27 as template. Major methodological advancements have improved the reliability of 3D computational predictions, and models good enough to guide experimental design can be generated from templates with as low as 6% identity to the target (*31*). The deviation of the PPR27 model's backbone residue conformations from those of well refined structures as determined by the WHAT-IF Ramachandran plot z-score was -2.429 while PROCHECK determined that 91.0% of amino acid residues in the model fall in the core region of the Ramachandran plot, both of which indicate good stereochemical quality. Thus, despite the

modest target-template identity the quality of the 3D model is sufficient to provide a global structural perspective of PPR27 and aid the interpretation of functional data.

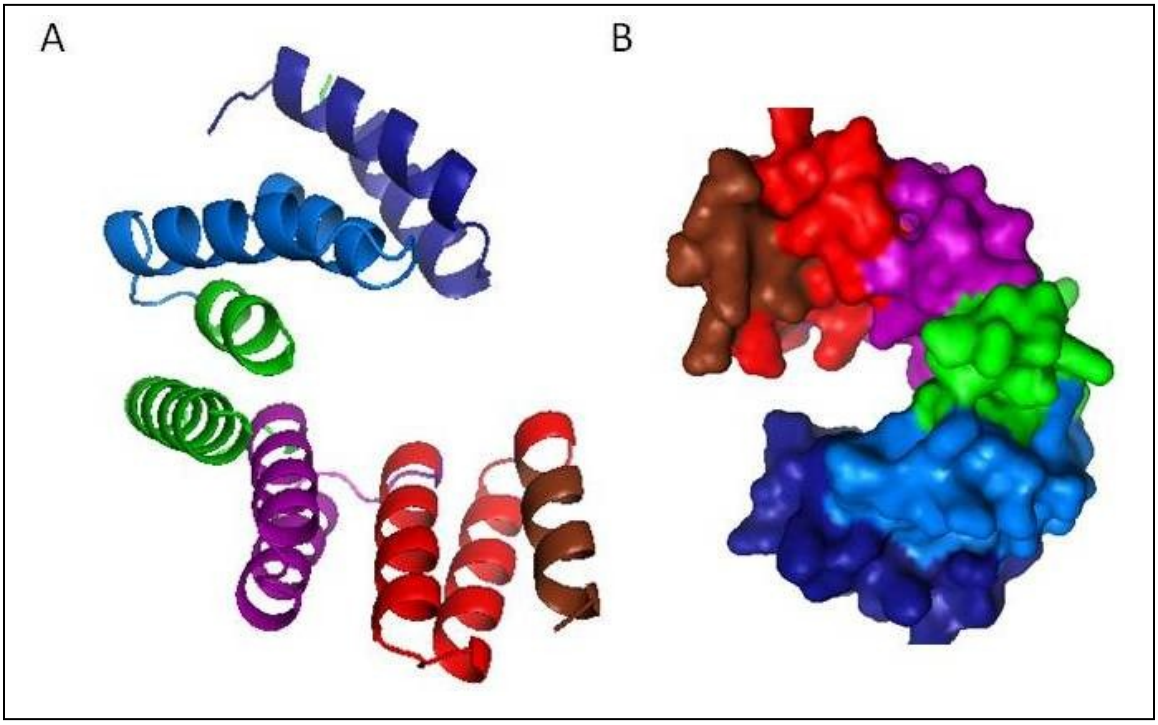


Figure 3.6: Predicted 3D structure of PPR27.

(A) Ribbon view of the predicted 3D structure of PPR27 showing a superhelical organization of the six PPR motifs (dark blue, first or most N-terminal PPR motif; blue, second PPR motif; green, third PPR motif; purple, fourth PPR motif; red, fifth PPR motif; brown, sixth or most C-terminal PPR motif). PPR27 regions with very low structural homology to the template are not represented (hence the break in the loop of the third motif and the missing helix of the sixth motif). The model was generated using the crystal structure of the tetratricopeptide repeat (TPR) protein TTC0263, a protein with 14% sequence identity to PPR27, as template. (B) Surface view of the predicted 3D structure of PPR27 showing a putative ligand binding cavity formed from a wall contributed by all six PPR motifs. The color code in (B) parallels that in (A).

3.3.3 A significantly soluble construct was identified through screening of truncated variants

Given the losses which occur during purification of enzymatic cleavage mixtures of fusion proteins, a soluble yield of 5.0 mg per liter of *E. coli* culture for the MBP fusion protein of wild type PPR27 is too low to generate sufficient free PPR27 for structural biology (chapter 2). Additionally, free wild type PPR27 aggregates on purification columns after removal of the MBP tag by TEV protease, even in the presence of stabilizers like glycerol and nonionic detergents (data not shown). Since construct engineering through truncations, single amino acid changes or both improves the solubility of some proteins (*32-33*), we screened for free PPR27 solubility in six truncated variants, including five whose MBP fusion protein yield was elevated at least 4-fold relative to wild type (**Figure 3.3 and Table 3.3**). Free PPR27 solubility was assessed following removal of MBP by TEV protease, both in the reaction mixture and on the column. All the six constructs yielded free PPR27 of the expected size (data not shown) and precipitation was not observed in the TEV cleavage reactions even after at least 24 hours of reaction. However, only ΔCR3 was soluble or stable enough not to aggregate on purification columns under native conditions, and only in the presence of the nonionic detergent dodecyl maltoside (DDM). Free ΔCR3 has been purified to over 95% purity using DEAE sephadex® anion exchange chromatography in the presence of 0.05% (w/v) DDM (**Figure 3.7A**). ΔCR3 displays a CD spectrum of a highly α-helical protein (**Figure 3.7B**) consistent with PPR27 being a helical repeat protein. Deconvolution of the CD spectrum with DICHROWEB (*23*) put the proportion of alpha-helical secondary structure at between 55% and 80% . In the presence of DDM alone, ΔCR3 could only be concentrated up to 0.1 mM, beyond which it formed oily globules which run as a smear on SDS-PAGE (data not shown). To improve solubility, stabilization with the

charged amino acids L-Arg and L-Glu was explored as described elsewhere (*18*). Under these conditions, no aggregation was observed at 0.52 mM protein and SDS-PAGE showed normal electrophoretic profile (**Figure 3.8A**). SDS-PAGE of SEC fractions showed ΔCR3 migrating as a monomeric protein (**Figure 3.8C**) and a size-exclusion chromatogram was sharp (**Figure 3.8B**), implying that the preparation was monodisperse and free of amorphous aggregates, even after concentration to the 0.40 mM used in SEC. The chemical shift dispersion of the $^{1}H,^{15}N$-TROSY spectrum of ΔCR3 was sufficiently wide for an α-helical repeat protein (**Figure 3.8D**). Moderate peak broadening was however visible, probably due to the persistence of DDM detergent post-dialysis and hence the reformation of protein-detergent complexes on concentration.

Table 3.3: Gross soluble yield of MBP fusion protein of various PPR27

Construct	Soluble yield (mg/L of culture)	
	At 30 ^{o}C	At 22 ^{o}C
Wild type[a]	≤ 3.0	10
ΔNR1[a]	≤ 3.0	24
ΔNR2[a]	39	
ΔNR3[a]	0	0
ΔNCR11	0	0
ΔNCR12[c]	10	
ΔNCR21[c]	60	
ΔCR1[b]	14	
ΔCR2[a]	48	
ΔCR3[a]	48	
ΔCR4[b]	62	

[a]Mean of protein yield from at least three different batches of cultures

[b]Mean of protein yield from at least two different batches of cultures

[c]Yield from only one batch of culture

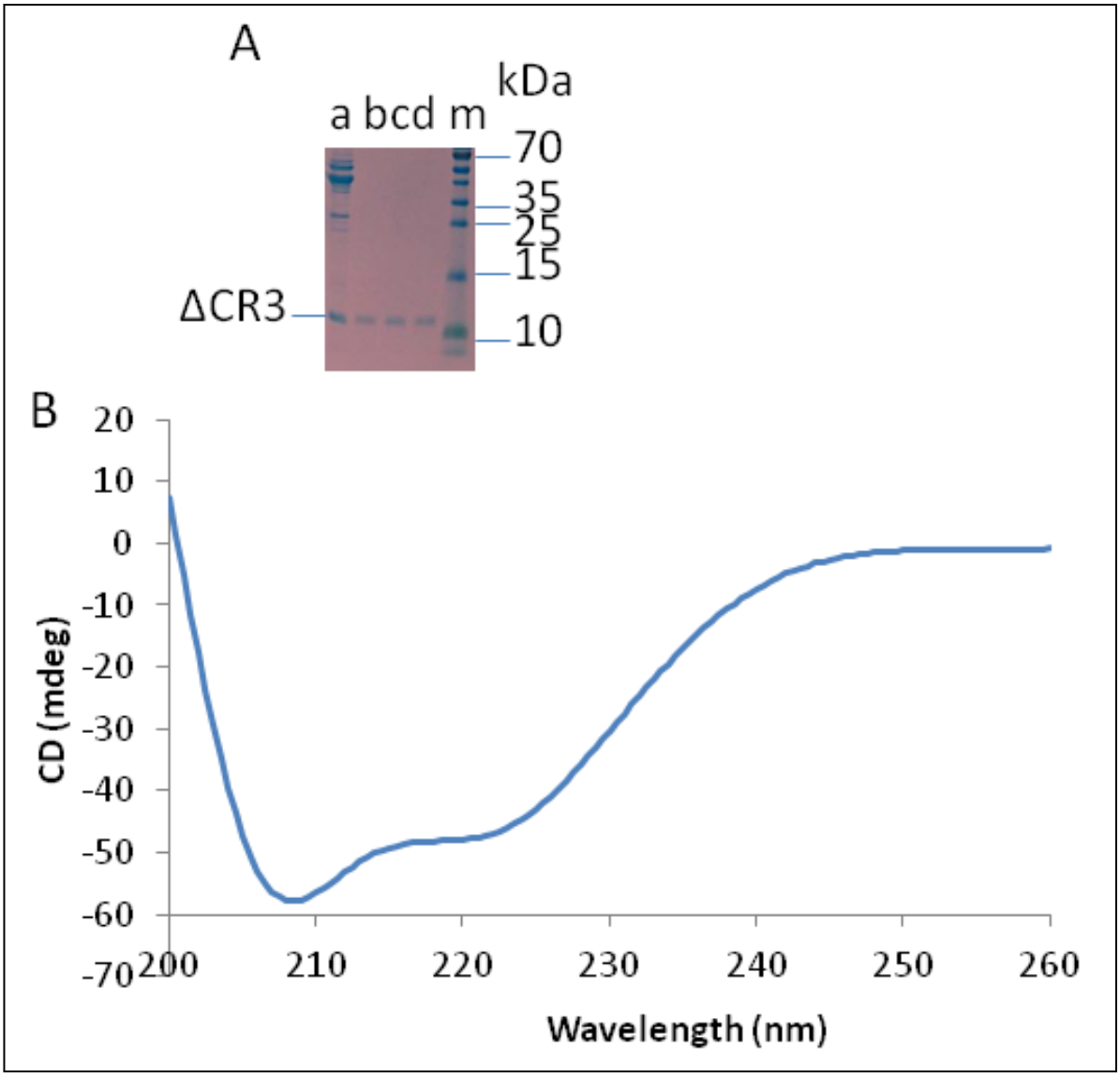


Figure 3.7: Purification and CD spectroscopy of free ΔCR3.

(A) SDS-PAGE of pure free ΔCR3 after anion exchange chromatography of the TEV protease cleavage reaction of MBPPPR27ΔCR3 in the presence of 0.05% (w/v) DDM detergent; a, TEV cleavage reaction of MBPPPR27ΔCR3; pure free ΔCR3 is shown in the flow through (b) and salt-free washes (c and d); m, protein size marker. (B) CD spectrum of free ΔCR3 at 0.43 mg/ml concentration in low salt buffer (10 mM NaH_2PO_4, 100 mM NaCl, 1.0 mM DTT) at 25 ^{o}C.

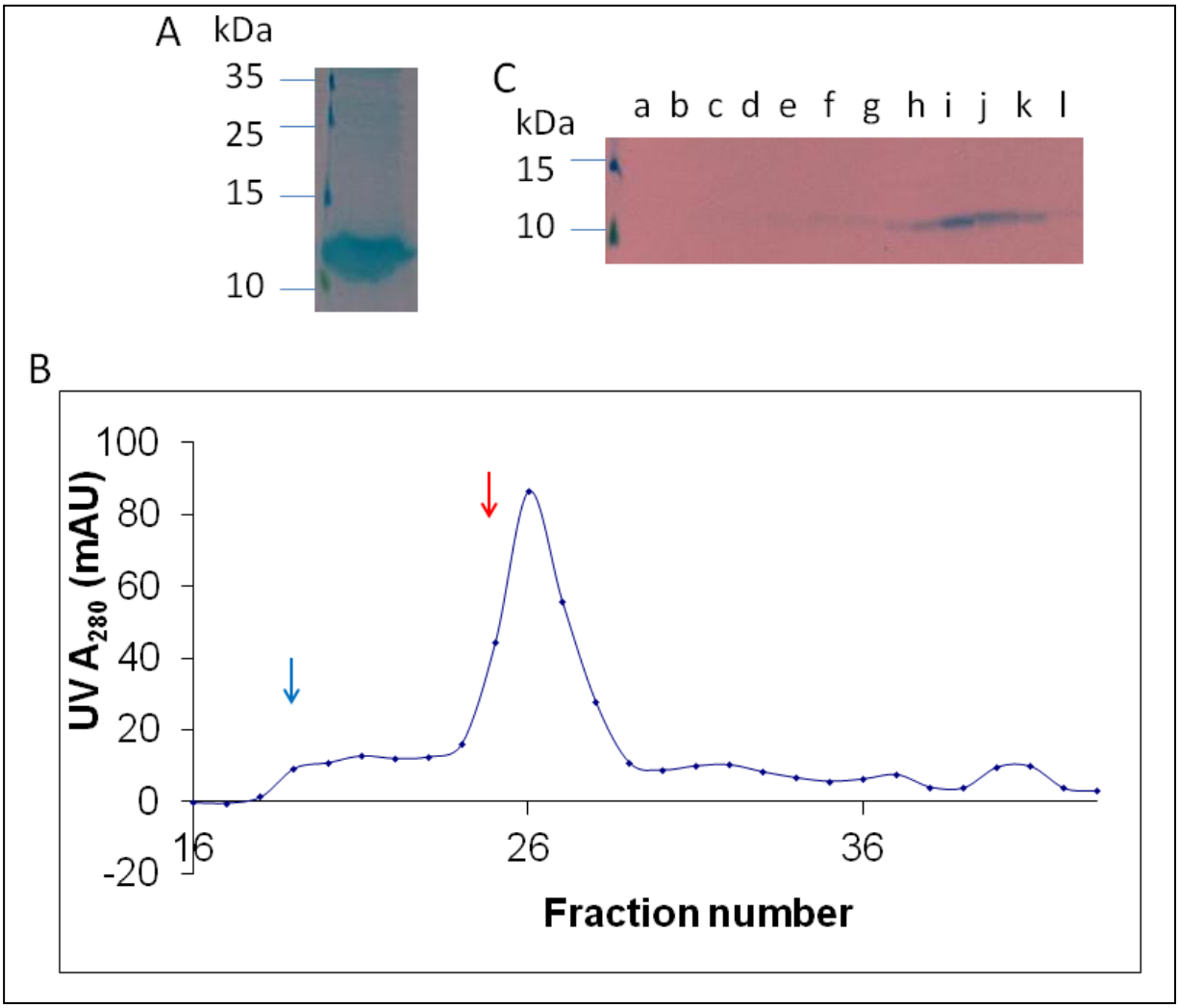
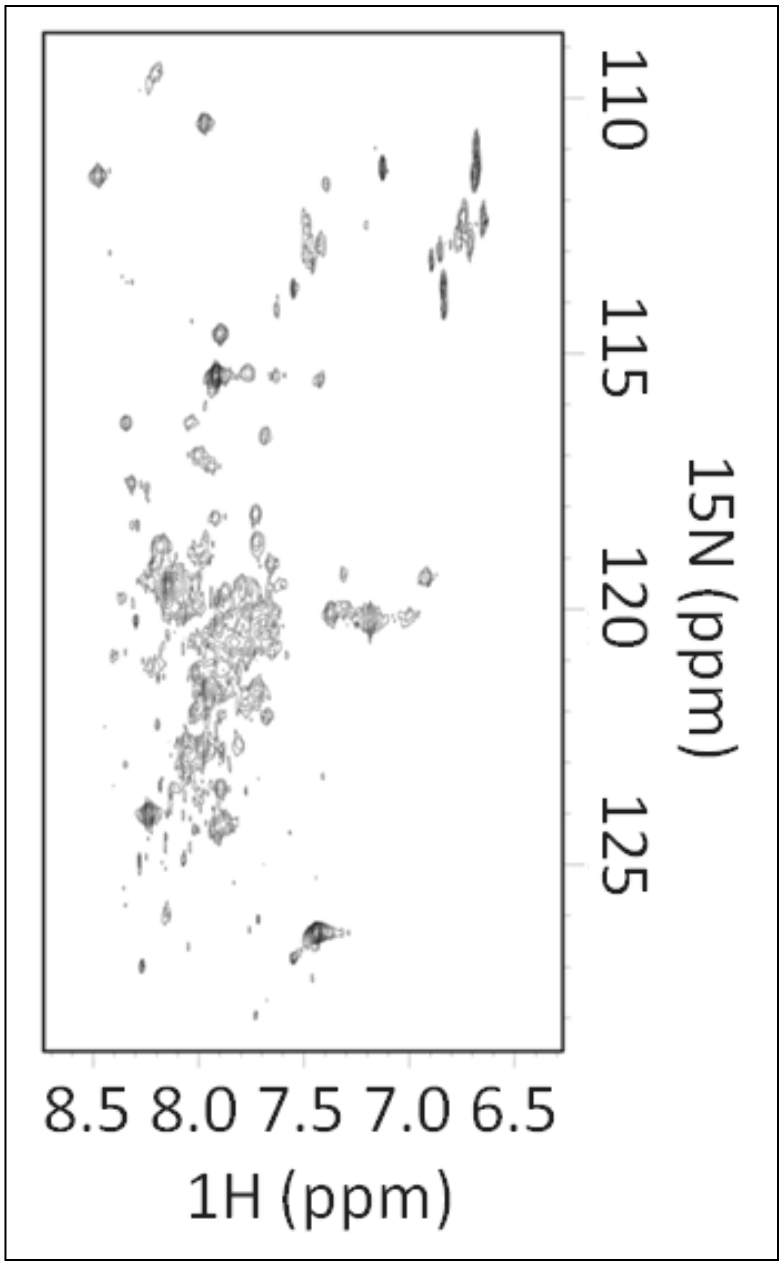


Figure 3.8: Size exclusion chromatography and $^{1}H,^{15}N$-TROSY NMR spectroscopy of ΔCR3

(A) SDS-PAGE of pure ^{15}N-labeled ΔCR3 concentrated to 0.5 mM in the presence of L-Arg and L-Glu. (B) Gel filtration chromatogram of concentrated ΔCR3 with a running buffer containing 500 mM L-Arg showing a sharp peak between fractions 24 and 29 where ΔCR3 elutes; using the same column, thyroglobulin (66 kDa) peaks at position of blue arrow while ribonuclease A (14.7 kDa) peaks at red arrow. (C) SDS-PAGE of fractions surrounding the peak in the gel filtration chromatogram in (B); a-l are consecutive fractions 18-29.

Figure 3.8 (cont'd).



(D) $^{1}H,^{15}N$-TROSY NMR spectrum of 0.46 mM ^{15}N-labeled ΔCR3 in 10 mM NaH_2PO_4, 100 mM NaCl, 1.0 mM DTT, 50 mM L-Arg, 50 mM L-Glu obtained using a 900 MHz spectrometer.

3.4 DISCUSSION

Crystallographic structures of RNA-bound PUF domains portray an intricate relationship between RNA ligands and protein repeats in which 1:1 binding of cognate RNA ligands occurs by each PUF domain contacting one base (*34*), and promiscuous binding of mutant/longer RNA ligands occurs via flipping out of non-cognate bases (*35*). In both mechanisms of RNA binding by PUF domains, the protein residues which contact RNA are distributed throughout all the protein repeats, and deletion or swapping of domains is not expected to abolish RNA binding. Similarly, if the mode of PPR27-RNA interface is consistent with the predicted 3D structure, then deletion or rearrangement of domains should only weaken but not abolish RNA binding activity, and this is what was observed. Similar studies on other classical PPR proteins though still few predominantly project a similar picture. For example, in the high chlorophyll fluorescent mutant 152 (HCF152) protein, all eight constructs with two PPR motifs and one out of four constructs with a single PPR motif retained RNA binding capacity (*11*). Literature on mapping experiments of the PPR protein binding sites on target RNAs also lends further credence to the involvement of all PPR motifs in RNA contacts. A case in point is the work by Barkan and coworkers on maize PPR10 (*36*) and *Arabidopsis* HCF107 (*37*) which found the number of nucleotides constituting the protein binding site to be approximately equal to the number of PPR and PPR-like modules in each protein (*36-37*). It is therefore increasingly apparent that one classical PPR module is sufficient to bind RNA.

The *Drosophila melanogaster* Pumilio homology domain (DmPUM-HD) interacts with its Nanos Response Elements (NREs) cognate RNA with a K_d of about 0.5 nM (*38*) while the *Homo sapiens* Pumilio homology domain (HsPUM-HD) binds to the same RNA with a K_d of

less or equal to 0.06 nM (*34*). In the crystal structures of HsPUM-HD complexed to NREs, each nucleotide of the minimal bound RNA element stacks between two amino acid side chains provided by residues at position 13 of consecutive helical repeats (*34*). Each of the stacked nucleotides is also specifically bound by 2-3 hydrogen bonds from amino acid residues at position 12 and 16 in each helical repeat (*34*). Altogether, there are about 20 hydrogen bonds between the 8 nucleotides of the bound RNA element and the HsPUM-HD protein (*34*). In contrast with the PUF domain proteins, PPR27 interacts with the putative G_6 RNA element with a K_d of 1.0 μM (chapter 2), which is 2000-fold less than the RNA binding affinity of PUF domains. Thus, intermolecular stacking and weaker forces (Van der Waals) interactions likely dominate the energetics of PPR27-RNA interaction than hydrogen bonding. The role of electrostatic interactions in PPR27-RNA contacts is likely to be minimal since binding is not affected by the high salt (150 mM KCl) used in our assays.

The predicted 3D structure of PPR27 presents a groove or channel as the putative RNA binding cavity (**Figure 3.6B**), implying that the protein's G-quadruplex cognate RNA (chapter 2) should anchor into the channel. Insights from docking of the dimeric G-quadruplex DNA aptamer $G_4TG_3AG_2AG_3T$ into the crystal structure of its HIV-1 integrase protein target is consistent with the latter proposition (*39*).

The hypothetical RNA binding channel of PPR27 is formed by a wall contributed by all its PPR motifs, hence truncating the protein is likely to reduce steric restrictions to the RNA binding cavity and alter the RNA binding specificity. Such an effect has been reported in Arabidopsis HCF152 (*12*) and may be a salient feature of classical PPR proteins. In the latter case, some shortened proteins concomitantly displayed stronger binding capacity for RNA sequences poorly

bound by the wild type protein and weak interaction with RNA sequences preferred by the wild type (*12*). Thus, concerted action by multiple PPR modules probably serves not only to strengthen the RNA-protein interaction but also to enhance RNA binding specificity.

Enhancing PPR protein solubility and stability is critical to high resolution structural and dynamics studies. Using PPR27 as a case study, we demonstrate that classical construct engineering and solvent tinkering can be effective in efforts to get PPR proteins into the structural biology pipeline. Using these approaches, we obtained a promising preliminary NMR spectrum for a three repeat PPR27 construct, which though a work in progress, constitutes a significant milestone since it is the first NMR spectrum of a classical PPR protein to our knowledge. This is also the first example of a classical PPR protein concentrated to NMR scale levels. Finally, though their record in protein crystallography is not clear, L-Arg and L-Glu hold promise for crystallization of classical PPR proteins since they ensure homogeneity in concentrated protein.

REFERENCES

REFERENCES

1. Schmitz-Linneweber, C., and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression, *Trends Plant Sci 13*, 663-670.

2. Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes, *Mol Cell 42*, 106-117.

3. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, *Plant Cell 16*, 2089-2103.

4. Mingler, M. K., Hingst, A. M., Clement, S. L., Yu, L. E., Reifur, L., and Koslowsky, D. J. (2006) Identification of pentatricopeptide repeat proteins in Trypanosoma brucei, *Mol Biochem Parasitol 150*, 37-45.

5. Pusnik, M., Small, I., Read, L. K., Fabbro, T., and Schneider, A. (2007) Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes, *Mol Cell Biol 27*, 6876-6888.

6. Fairlamb, A. H. (2003) Chemotherapy of human African trypanosomiasis: current and future prospects, *Trends Parasitol 19*, 488-494.

7. Wilkinson, S. R., and Kelly, J. M. (2009) Trypanocidal drugs: mechanisms, resistance and new targets, *Expert Rev Mol Med 11*, e31.

8. Andres, C., Lurin, C., Small, I.D. (2007) The multifarious roles of PPR proteins in plant mitochondrial gene expression, *Physiologica Plantarum 129*, 14-22.

9. Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giege, P. (2008) PPR336 is associated with polysomes in plant mitochondria, *Journal of Molecular Biology 375*, 626-636.

10. Filipovska, A., and Rackham, O. (2012) Modular recognition of nucleic acids by PUF, TALE and PPR proteins, *Molecular Biosystems 8*, 699-708.

11. Kobayashi, K., Kawabata, M., Hisano, K., Kazama, T., Matsuoka, K., Sugita, M., and Nakamura, T. (2012) Identification and characterization of the RNA binding surface of the pentatricopeptide repeat protein, *Nucleic Acids Research 40*, 2712-2723.

12. Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G. (2003) RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA, *European Journal of Biochemistry 270*, 4070-4081.

13. Tavares-Carreon, F., Camacho-Villasana, Y., Zamudio-Ochoa, A., Shingu-Vazquez, M., Torres-Larios, A., and Perez-Martinez, X. (2008) The pentatricopeptide repeats present in Pet309 are necessary for translation but not for stability of the mitochondrial COX1 mRNA in yeast, *J Biol Chem 283*, 1472-1479.

14. Mili, S., and Pinol-Roma, S. (2003) LRP130, a pentatricopeptide motif protein with a noncanonical RNA-binding domain, is bound in vivo to mitochondrial and nuclear RNAs, *Mol Cell Biol 23*, 4972-4982.

15. Rackham, O., and Filipovska, A. (2012) The role of mammalian PPR domain proteins in the regulation of mitochondrial gene expression, *Biochim Biophys Acta 1819*, 1008-1016.

16. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res 25*, 3389-3402.

17. Marley, J., Lu, M., and Bracken, C. (2001) A method for efficient isotopic labeling of recombinant proteins, *J Biomol NMR 20*, 71-75.

18. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A., and Lian, L. Y. (2004) A simple method for improving protein solubility and long-term stability, *J Am Chem Soc 126*, 8933-8939.

19. Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005) FFAS03: a server for profile--profile sequence alignments, *Nucleic Acids Res 33*, W284-288.

20. Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction, *Protein Sci 12*, 2001-2014.

21. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program, *J Mol Graph 8*, 52-56, 29.

22. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures, *Journal of Applied Crystallography 26*, 283-291.

23. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res 32*, W668-673.

24. Tischer, A., Lilie, H., Rudolph, R., and Lange, C. (2010) L-arginine hydrochloride increases the solubility of folded and unfolded recombinant plasminogen activator rPA, *Protein Sci 19*, 1783-1795.

25. Fujii, S., Bond, C. S., and Small, I. D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution, *P Natl Acad Sci USA 108*, 1723-1728.

26. Small, I. D., and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins, *Trends Biochem Sci 25*, 46-47.

27. Edwards, T. A., Pyle, S. E., Wharton, R. P., and Aggarwal, A. K. (2001) Structure of Pumilio reveals similarity between RNA and peptide binding motifs, *Cell 105*, 281-289.

28. Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P., and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase, *Nature 478*, 269-273.

29. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing, *Proc Natl Acad Sci U S A 109*, 16149-16154.

30. Lim, H., Kim, K., Han, D., Oh, J., and Kim, Y. (2007) Crystal structure of TTC0263, a thermophilic TPR protein from Thermus thermophilus HB27, *Mol Cells 24*, 27-36.

31. Tramontano, A., and Morea, V. (2003) Assessment of homology-based predictions in CASP5, *Proteins-Structure Function and Genetics 53*, 352-368.

32. Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., dhe-Paganon, S., Park, H. W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S. H., Rao, Z., Shi, Y., Terwilliger, T. C., Kim, C. Y., Hung, L. W., Waldo, G. S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J. L., Stevens, R. C., Lesley, S. A., Wilson, I. A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M. I., Eschenfeldt, W. H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S. K., Emtage, J. S., Sauder, J. M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S. C., Bonanno, J. B., Fiser, A., Swaminathan, S., Studier, F. W., Chance, M. R., Sali, A., Acton, T. B., Xiao, R., Zhao, L., Ma, L. C., Hunt, J. F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C. K., Wang, D., Wang, H., Jiang, M., Montelione, G. T., Stuart, D. I., Owens, R. J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K., and Gunsalus, K. C. (2008) Protein production and purification, *Nat Methods 5*, 135-146.

33. Graslund, S., Sagemark, J., Berglund, H., Dahlgren, L. G., Flores, A., Hammarstrom, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P., and Weigelt, J. (2008) The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins, *Protein Expr Purif 58*, 210-221.

34. Wang, X. Q., McLachlan, J., Zamore, P. D., and Hall, T. M. T. (2002) Modular recognition of RNA by a human pumilio-homology domain, *Cell 110*, 501-512.
35. Gupta, Y. K., Nair, D. T., Wharton, R. P., and Aggarwal, A. K. (2008) Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity, *Structure 16*, 549-557.

36. Prikryl, J., Rojas, M., Schuster, G., and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein, *P Natl Acad Sci USA 108*, 415-420.

37. Hammani, K., Cook, W. B., and Barkan, A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression, *P Natl Acad Sci USA 109*, 5651-5656.

38. Zamore, P. D., Bartel, D. P., Lehmann, R., and Williamson, J. R. (1999) The PUMILIO-RNA interaction: a single RNA-binding domain monomer recognizes a bipartite target sequence, *Biochemistry 38*, 596-604.

39. Phan, A. T., Kuryavyi, V., Ma, J. B., Faure, A., Andreola, M. L., and Patel, D. J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase, *Proc Natl Acad Sci U S A 102*, 634-639.

CHAPTER 4

ANALYSIS OF THE RNA BINDING SPECIFICITY OF THE 41 kDa *TRYPANOSOMA BRUCEI* PENTATRICOPEPTIDE REPEAT PROTEIN

Assistance in the growth of *T. brucei* parasites and guidance in mitochondrial preparation and RNA extraction was received from Dr. Donna Koslowsky. Primers used in the pull-down assays were also a gift from Dr. Donna Koslowsky. I designed and performed all the other experiments herein.

ABSTRACT

Shuttling of *Trypanosome brucei* (*T. brucei*) between the contrasting environments in the animal blood stream and the arthropod vector is aided by a tunable mitochondrion which downregulates its activity in the animal blood stream and ups it in the arthropod gut via differential expression of its genes. Trypanosomes, however, are deficient in transcription factors and regulate gene expression posttranscriptionally via transcript stability, differential processing, and translation. It is hypothesized that a milieu of RNA binding proteins are involved, most of which are still putative. Here we heterologously expressed the 41 kDa *T. brucei* pentatricopeptide repeat (PPR) protein in *E. coli* and probed its RNA binding activity using both RNA oligonucleotides and a mitochondrial RNA (mitoRNA) extract from procyclic parasites. Under non-equilibrium conditions, stable binding is only observed with poly(G) ssRNA. However, under equilibrium conditions PPR41 showed strong and similar affinities for G_{12}, U_{12}, and $(GGU)_4$ ssRNA, modest affinity for A_{12}, and reduced affinity for C_{12} ssRNA. From a procyclic mitoRNA extract, PPR41 pulled down precursor RNA transcripts. RNA editing converts G-tracts in pre-edited mRNAs to GU patterns. A role for PPR41 in either edited RNA binding or precursor RNA processing is therefore likely.

4.1 INTRODUCTION

To adapt from the sugar-rich environment in the human blood stream to the sugar deficient environment in the tsetse fly gut and vice versa, *Trypanosoma brucei* (*T. brucei*) modulates the metabolic activity of its mitochondrion. In the blood stream forms (BFs) of the parasite, mitochondrial activity is downsized and expression of most mitochondrial genes is suppressed, allowing the parasite to generate energy from glucose by glycolysis, while in the procyclic forms (PFs) mitochondrial activity is reactivated, enabling the parasite to derive energy from proline and threonine by oxidative phosphorylation (*1*). Controlling gene expression is the most robust mechanism for metabolic regulation. While such regulation generally occurs during transcription via a milieu of transcription factors (TFs), trypanosomes are an exception as they are deficient in TFs and predominantly regulate gene expression at posttranscriptional level (*2-5*). For the trypanosome mitochondrion, posttranscriptional RNA processing even takes on extra significance as many of its genes are encoded as incomplete open reading frames (ORFs) on a polycistron and must undergo extensive processing to form translatable ORFs (*6-7*).

Due to their essential nature in the parasite life cycle, the unique biochemical processes involved in trypanosome mitochondrial RNA (mitoRNA) metabolism are attractive targets for development of novel anti-trypanosome drugs. At each of the main stages of trypanosome mitoRNA processing, namely nucleolytic cleavage of polycistronic primary transcripts, polyadenylation and polyuridylation, RNA editing, and translation, there are opportunities for inhibition of RNA binding proteins and enzymes. But while trypanosome genome sequencing (*8-10*) and proteomics (*11*) have ushered in many putative mitoRNA processing proteins, our knowledge of the actual role of these proteins and the specific RNA sequences/motifs they recognize remains limited.

We investigated the biochemical activity of the 41 kDa pentatricopeptide repeat (PPR) protein (aka PPR41, genbank accession # XM_840666, CDS Tb927.7.1350), one member of a family of at least 37 nuclear-encoded, sequence specific RNA binding proteins targeted to the trypanosome mitochondrion (*12*). PPR proteins are defined by two or more degenerate 35 amino acid motifs, each of which folds into an antiparallel helical pair analogous to that of the 34 amino acid tetratricopeptide repeat (TPR) motif (*13-14*). Across different trypanosome species, PPR proteins are highly conserved and essential to the parasites (*15-16*). In plants, where they are better characterized, they have been implicated in diverse roles in posttranscriptional RNA metabolism in the mitochondria and chloroplasts where they localize, including RNA stabilization, RNA splicing, RNA editing, RNA polyadenylation, and translation (*17-24*). Recent reports have linked two *T. brucei* PPR proteins to the synthesis of long poly(A/U) tails and translational control (*12*). Another *T. brucei* PPR protein, the so-called TbPPR9 was recently found to be required for the stability of cytochrome oxidase subunits 1 (COX1) and 2 (COX2) mRNAs (*25*). Many trypanosome PPR proteins are also found in mitoribosomal subunit complexes, both in RNA-mediated and RNA-independent manner (*12, 26*), though PPR41 is not one of them. However, the cognate RNA sequences for all *T. brucei* PPR proteins remain unknown. In this work, PPR41 was overexpressed in *E. coli*. Using the recombinant protein, we observed that PPR41 binds with highest affinity to GU-rich sequences and pulls down precursor RNA transcripts from total mitoRNA, consistent with a role in either edited RNA binding or precursor RNA processing.

4.2 MATERIALS AND METHODS

4.2.1 Materials

RNA labeled with FLUO (fluorescein) at its 5' end was supplied by Dharmacon RNA Technologies (Lafayette, CO), treated as described in chapter 2, and reconstituted in MQ-H_2O. For cloning and site-directed mutagenesis, dry primers were obtained from Sigma-Aldrich (St. Louis, MO) and reconstituted in MQ-H_2O. The pMalTEV-E30 plasmid vector was a kind donation from Dr. Alice Barkan (University of Oregon). Amylose resin, amylose magnetic beads, BamH1 and SalI restriction enzymes, and T4 DNA ligase were obtained from New England BioLabs (Ipswich, MA); recombinant RNasin® was from Promega Corporation (Madison, WI); QuikChange® mutagenesis kit was from Agilent Technologies (Santa Clara, CA); M-MLV reverse transcriptase was from Invitrogen (Grand Island, NY); DNase I and bovine serum albumin were from Roche Applied Science (Indianapolis, IN); and 2X PCR Master Mix was from Syzygy Biotech (Grand Rapids, MI).

4.2.2 Multiple sequence alignment of *T. brucei brucei* PPR41 and its homologues

I searched for PPR41homologues in NCBI's nucleotide collection using the BLAST program (*27*). A text file of the amino acid sequences of PPR41 and its homologues in FASTA format was then uploaded into the TCOFFEE sequence alignment server (http://igs-server.cnrs-mrs.fr/Tcoffee). Finally, the alignment was formatted with ESPript 2.2 (http://espript.ibcp.fr/ESPript).

4.2.3 PPR41 structural modeling and signal peptide prediction

To assist in protein construct design during cloning and site-directed mutagenesis (SDM), and to gain insights into PPR41's RNA binding mechanism, the size of the mitochondrial targeting signal (MTS), the repeat composition, the secondary structure, and the 3-dimensional (3D) structure of PPR41 were modeled from amino acid sequence. First, the signal peptide at the N-terminus of full PPR41 was predicted online from amino acid sequence using SignalP 4.0 (*28*) (http://www.cbs.dtu.dk/services/SignalP/). Secondly, PPR41's repeat portfolio was re-evaluated using the classical PPR motif prediction tool TPRpred (*29*) (http://toolkit.tuebingen.mpg.de/tprpred/) and the de novo repeat prediction tool HHrepID (*30*) (http://toolkit.tuebingen.mpg.de/hhrepid/). Then PPR41secondary structure was modeled from sequence by the PSIPRED program (*31*) (http://bioinf.cs.ucl.ac.uk/psipred/). Next, the 3D structure of PPR41 was modeled by fold recognition using I-TASSER (*32*), a server which has consistently performed as the most reliable in Critical Assessment of Protein Structure Prediction (CASP) experiments (http://zhanglab.ccmb.med.umich.edu/I-TASSER). We then reconciled the secondary structure, tertiary structure, and predicted repeat organization to generate a harmonized PPR motif map of PPR41.

4.2.4 Gene design, cloning, and site-directed mutagenesis

DNA coding for mature PPR41 (lacking the MTS) was optimized to codons of highly expressed *E. coli* genes by GeneDesign (*33*). The codon-optimized DNA was cloned to the 3' end of MBP in a pMalTEV-E30 plasmid vector as described in chapter 2, using BamH1 and SalI restriction enzymes and the primers CGH115 and CGH116 (**Table 4.1**). To aid soluble expression and tandem affinity purification of the MBP fusion protein, a variant lacking the last three C-terminal

PPR motifs (ΔCR3) and containing a His_6 tag at its C-terminus was created by SDM as described in chapter 2 using the primers CGH141 and CGH142 (**Table 4.1**).

Table 4.1: Primers used in PPR41 cloning and site-directed mutagenesis

Activity	Primer code	Primer sequence (5' to 3')
Add BamH1 restriction site on to 5' end of PPR41	CGH115	GTACTTCCAGGGATCCGGTAACGACGAATCTCT-GCC
Add SalI restriction site on to 3' end of PPR41	CGH116	CTTGCCTGCAGGTCGACTCACAGGATAACAGA-TTCA
Insert His_6 tag after T229 alongside deletion of last two C-terminal PPR motifs	CGH141	CGGTAAACACCTGACCCACCACCACCACCACC-ACTGAGCTGCTACCTACGAAT
	CGH142	ATTCGTAGGTAGCAGCTCAGTGGTGGTGGTGG-TGGTGGGTCAGGTGTTTACCG

4.2.5 Protein expression and purification

MBPPPR41ΔCR3 was overexpressed in BL21 (DE3) *E. coli* by induction with 0.1 mM IPTG (isopropyl-β-D-1-thiogalactopyranoside) at 30 ^{o}C for 6 hours as described in chapter 2. Cells were harvested by centrifugation at 5,000 rpm and stored at -80 ^{o}C until protein purification. For pull-down assays, protein was purified under native conditions by nickel nitrilotriacetic aid (Ni-NTA) affinity chromatography alone. For other assays, protein was purified by tandem Ni-NTA and amylose affinity chromatography as noted previously (chapter 2). Buffers and methods for protein extraction and purification are the same as in chapter 2. All protein was dialyzed against - buffer DCB (20 mM Tris-HCl, pH 7.5, 50 mM NaCl) at 4 ^{o}C for at least 12 hours, using 7K MWCO SnakeSkin tubing. Dialysate was then concentrated and concentration determined from U.V A_{280} as in chapter 2.

4.2.6 Electrophoretic mobility shift assays (EMSA)

5'-FLUO labeled G_{12}, A_{12}, U_{12}, and C_{12} ssRNAs were used as model ligands. Electrophoresis with Tris-glycine agarose gels as established previously for PPR27-RNA EMSAs (chapter 2) was done. The RNA binding reactions were set up as in chapter 2 and contained 150 nM RNA, 20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, 0.1 μg/μl BSA, and 0.4 U/μl RNasin for each protein concentration. Gels were scanned for fluorescent bands using a VersaDoc^{TM} MP 4000 scanner (Bio-Rad Laboratories).

4.2.7 Fluorescence anisotropy

The anisotropy of 5'-FLUO labeled G_{12}, A_{12}, U_{12}, C_{12}, and $(GGU)_4$ ssRNA was assessed in the presence of increasing MBPPPR41ΔCR3 concentration. Following the precedent with PPR27 and G-rich ssRNA (chapter 2), frozen RNA in MQ-H_2O was slowly thawed on ice-cold water to maintain native conformation. RNA binding reactions were set up as in chapter 2 and included 20 nM RNA, 20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, and protein. Binding was allowed to proceed at ambient temperature for at least 15 minutes before polarized fluorescence data was collected as described previously (chapter 2). Using equation (i), the fluorescence anisotropy of the RNA was computed, normalized, and used to determine the dissociation constant (K_d) of PPR41-RNA interaction by nonlinear regression with equation (ii) where G is the grating factor; I_{vv} and I_{vh} are vertically polarized and horizontally polarized emission intensities respectively following vertically polarized excitation; c is the protein concentration; and n is the apparent stoichiometry of the interaction. For each RNA, analysis was done on an average of two data sets.

$$r = \frac{I_{vv} - GI_{vh}}{I_{vv} + 2GI_{vh}} \quad \text{---------- (i)}$$

$$f(c) = \frac{nc}{K_d + c} \quad \text{---------- (ii)}$$

4.2.8 Pull-down of *T. brucei* total mitochondrial RNA

MitoRNA was extracted from PF *T. brucei* and reconstituted in MQ-H_2O as described previously (chapter 2). 2.0 µM of Ni-NTA purified MBPPPR41ΔCR3 in buffer DCB was immobilized onto about 20 µl of amylose magnetic beads as previously (chapter 2). 54 µg of total mitoRNA (in 40 µl) was added to the immobilized protein and left to bind with gentle hand shaking at ambient temperature for 15 min. Unbound RNA was washed off followed by phenol-chloroform-isoamyl alcohol extraction of bound RNA and ethanol precipitation as done previously (chapter 2). The dry RNA pellet was reconstituted in 40 µl of RNase-free MQ-H_2O followed by analysis for the major mitoRNA forms (polycistronic RNA precursors, pre-edited mRNA transcripts, and edited mRNA) by reverse transcription PCR (RT-PCR) using the same set of primers as outlined in chapter 2. PCR reactions were analyzed on 1% agarose gels in TBE buffer at 100 V.

4.3 RESULTS

Like other *T. brucei* PPR proteins, neither the specific RNA sequence it recognizes nor the mechanism of RNA binding is known for PPR41. We thus sought to decipher the RNA binding specificity of this protein using in vitro binding assays with both synthetic RNA oligonucleotides and procyclic mitoRNA extract. A recombinant fusion protein of MBP and a truncated construct of PPR41 lacking the last three C-terminal PPR motifs (ΔCR3) was used in these assays. Both the cloning and truncation of PPR41 to form ΔCR3 were informed by computational structure prediction.

4.3.1 PPR41 structural perspectives and protein production

PPR41 is a highly conserved gene across the trypanosomatids (**Figure 4.1**). A BLAST search did not return any homologues with high sequence similarity from non-trypanosome species. Previously, *T. brucei* PPR41 was reported to contain two PPR motifs (*15-16*). The difficulties encountered in heterologous expression of wild type PPR41 are however inconsistent with the presence of only two PPR modules within the 350 amino acids long polypeptide. Due to the limitations of computational structure prediction in general, in order to get a better structural outline of PPR41, I re-examined its PPR motif repertoire using a combination of contemporary motif prediction programs, secondary structure modeling, and 3D structural modeling. TPRpred (*29*), a profile-based PPR motif recognition program predicted 4 PPR motifs (**Figure 4.2A**) while the de novo repeat prediction tool HHrepID (*30*) identified 5 repeats with high probability (**Figure 4.2B**). According to the predicted secondary structure, PPR41 has at least 7 antiparallel α-helical pairs, which could translate into at least 7 PPR modules (**Figure 4.3**). Finally, the 3D structural model suggests mature PPR41 has eight antiparallel α-helical pairs (**Figure 4.4A**). Putting everything together, it seems that PPR41 has up to 8 PPR modules (**Figure 4.6A**). The signal peptide was predicted to comprise the first 25 amino acids (**Figure 4.5**) and this was consistent with the outlook from secondary- and 3D structure modeling. With this broad structural perspective in hand, DNA for mature wild type PPR41 (amino acids 26-350) was cloned downstream of the *E. coli* MBP. However, production of the MBP fusion protein of mature wild type PPR41 has so far been unsuccessful. After deleting the last three C-terminal PPR motifs, about 4.0 mg of over 90% pure fusion protein (MBPPPR41ΔCR3) (**Figure 4.6B**) was obtained from 1.0 L of culture when expressed at 30 ^{o}C. This protein was then used in RNA binding assays.

```
Lbr RSGLPLYTRRLVAAVPLHCRKGVRRC--------NGGPFWPVHRGFTSGTTPCSSPPV
Ldo RCGLPLSTPRHVPAALAHSGKGVVRR--------NGGATWTARRSFTSSTTLRSYPPV
Tbb ----------------------MFRYRCRNA---------KNLLLMRNIRRWYTSAG-
Tbg ----------------------MFRYRCRNA---------KNLLLMRNIRRWYTSAG-
Tcb -----------MTAVFFHSPSHFAHCVCGGMRVFSDAYCWPRVSRFVHAVRWFSTSG-
Tcm -----------MTTPFFHSPSHVVHWVGGGMRGFSGAYCLPHVSRLVHTARWFSTSE-
                          . :                    :      :

Lbr SERL-STPVDASPLPFSCATGASDADKTAAVPSAGTSSS-LGTGTAAAVTSTVNDLDG
Ldo SELPSSTPAAPSPLPFISAFTASDAAETAAVPSADTSSP-LGTGTTAAATATVNDLHG
Tbb --NDESLPSAQGP------------------PLAGVSSPTLWRN---GGVLGEGMLGN
Tbg --NDENLPSAQEP------------------PLAGVSSPTLWRN---GGVLGEGMLGN
Tcb --VGGATPASRAEMSH---------------VKPASSSSELWRS---GEDSRTVTIGR
Tcm --VGSATPASPAEAPH---------------VKPASSSSELWRS---GEDSRTVTIGK
           *                           **  *      .       :

Lbr GEVPLNTPAEVEEAFACSVQARTVSNVQLRRYIDHLPPKDYSLALAAVKGARAAGLRI
Ldo DEVPLNTPAEVEEAFACSVQARTVSNVQLRRYIDHLPPKDYSLALAAVKGAKAAGLRI
Tbb PRVAATTPQEVEETFRCSVLARTVSNLKLRRYLQKLEPKDHTLVLAALRGAQAGGLRL
Tbg PRVAATTPQEVEETFRCSVLARTVSNLKLRRYLQKLEPKDHTLVLAALRGAQAGGLRL
Tcb PHVEATTPEQVEKTFRLSVLARTASNVQLRRYLEQLQPKDHALALAAVRGAEAGGLRV
Tcm LHLEATTPEQVEETFRLSVLARTASNVQLRRYLEQLQPKDHALALAAVRGAEAGGLRV
     .:  .** :**::*  ** ***.**::*****:::* ***::*.***::**.*.***:

Lbr SALTYESLLASLMSGGQLRASMELYQLMIKQRMTPTPNTYAALMDMCLQRDMPKACQS
Ldo SASTYEALLSSLMSGGQLRASMELYQVMIKQHMAPTPNMYAALMEMCLQRGMPKACQS
Tbb DTKTNEVALSKLMDGGQLQASMEIYRKMIQNRMNPTANTYATLMHMCIERDMPEACQK
Tbg DTKTNEVALSKLMDGGQLQASMEIYRKMIQNRMNPTANTYATLMHMCIERDMPEACQK
Tcb DAKTNEVHLMKLMDGGQLRASMELYRRMIDRRMTPTANTYATLMHMCIQREMPEACQK
Tcm DAKTNEVHLMKLMDGGQLRASMELYRRMIGGRMTPTANTYATLMHMCLQREMPEACQK
    .: * *  * .**.****:****:*: **  :* ** * **:**.**::* **:***.

Lbr LFNDLQKRGVRPSAQNYELMITSLSTEVPPQWKRAIDIFDKISRERKS-RITSKTYNA
Ldo LFDDMQQRGVRPAARNYELMITSLSAEVPPQWARAIEVFDRISRERRS-RITAKTYNA
Tbb LFEEMVKRGQSPNTRNYEMYIGSLAMENPPKWKNAIEVFDRMSRERHGKHLTAATYES
Tbg LFEEMVKRGQSPNTRNYEMYIGSLAMDNPPKWKKAIEVFDRMSRERHGKHLTAATYES
Tcb MFEEMVKRGQSPNTKNYELYITSLAMENPPKWEKAVEVFDHISRDRRGKHLTAMTYNS
Tcm MFEEMLKRGQSPNTRNYELYITSLAMENPPKWEKAVEVFDHISRDRRGKHLTATTYNS
    :*::: :**  * ::***: * **: : **:* .*:::**::**:*:. ::*: **::
```

Figure 4.1: Multiple sequence alignment of *T. brucei* PPR41 and its close homologues.

Star (*), absolutely conserved residues; colon (**:**), highly conserved residues; fullstop (**.**), moderately conserved residues; all others, less conserved residues. Amino acids 3-350 were aligned by Clustal Omega followed by manual adjustments.

Figure 4.1 (cont'd).

```
Lbr LMRVYMNMDPFDWRVVYNCYSEMRNRRPRVPLEWESYLILAEALRMGRAGYVRRGMAY
Ldo LMRVYMNMDPFDWRVVYNCYSEMRNRRPRVPLEWESYLILCEALRKGNAGYVRRGMAY
Tbb LMRVYLNMTPFDWRVVYNCYYEMRSRQPQIQLRWETYHLVAEALRRGQAGYTRRLITY
Tbg LMRVYLNMTPFDWRVVYNCYYEMRSRRPQIQLRWETYHLVAEALRRGQAGYTRRLITY
Tcb LMRVYLNMTPFDWRVVYNCYHELRNRKPPILLQWESYLLVAEALRRGRVGYIRRFITY
Tcm LMRVYLNMTPFDWRVVYNCYHELRNRKPPILLQWESYLLVAEALRRGRVGYIRRFITY
    *****:** ********** *:*.*:* : *.**:* ::.**** *..** ** ::*

Lbr IDAWIAVTPLRSWNFLMGAMMYLTFMMLLKSLAGYLVVWYYEMSVPSTS---DSVLSR
Ldo IDAWIAVTPLRSWNFLIGAMAYLAFMMVLKSIVGYFVVWYYESSVPSTS---DSVLPR
Tbb LDAWFCITYFRSWEFFMGFMFYICLMFLIKGLISWVVVWYYGRAVNASRGASESVIL-
Tbg LDAWFCITYFRSWEFFMGFMFYICLMFLIKGLISWVVVWYYGRAVNASRGASESVIL-
Tcb LDAWFCVTPIRSVEFLFGMALYVGAMLVLKAIISWLVVIYYERVIAAKKGVSESIL--
Tcm LDAWFCVTPIRSVEFLFGMALYVGAMLVLKAIISWLVVIYYERVIAAKKGASEPVF--
    :***:.:* :** :*::*   *:  *:::*.: .:.** **   : :.    : ::
```

Lbr, *Leishmania braziliensis*; Ldo, *Leishmania donovani*; Tbb, *Trypanosoma brucei brucei*; Tbg, *Trypanosoma brucei gambiense*; Tcb, *Trypanosoma cruzi brener*; Tcm, *Trypanosoma cruzi marinkellei*.

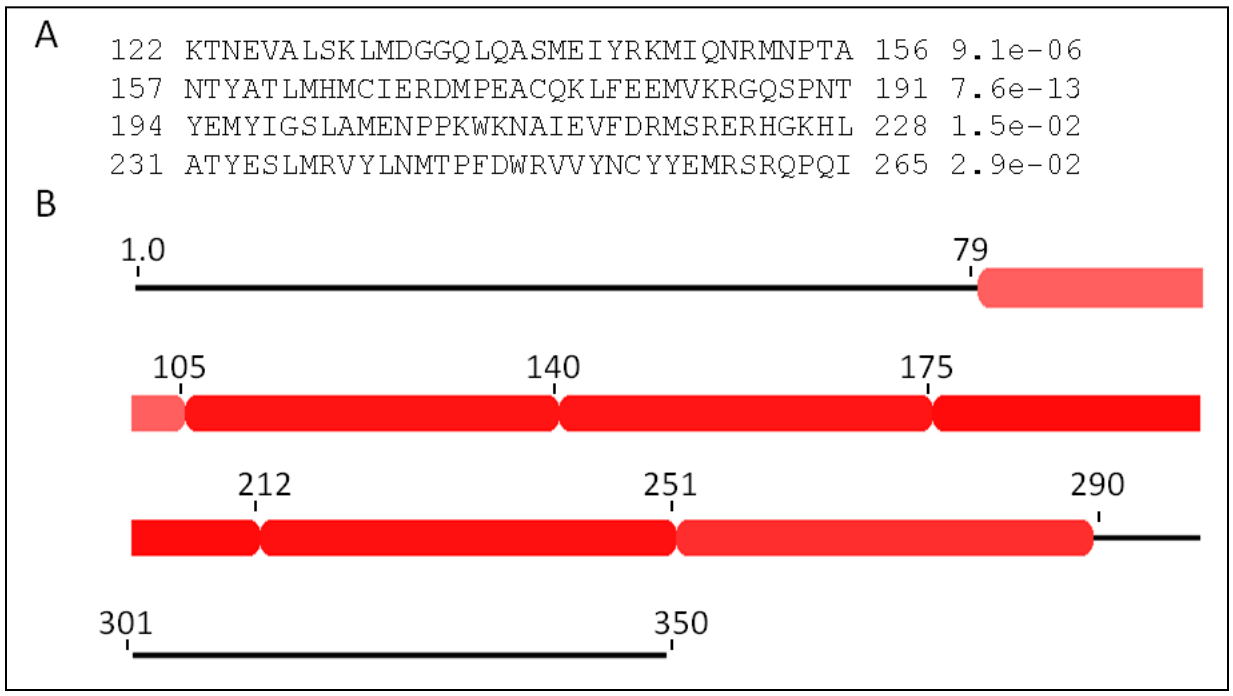


Figure 4.2: In silico prediction of PPR motifs in PPR41.

(A) 4 motifs predicted by TPRpred, a classical PPR motif prediction tool. (B) Five motifs predicted by HHrepID, a de novo repeat prediction tool, with high probability (deep red).

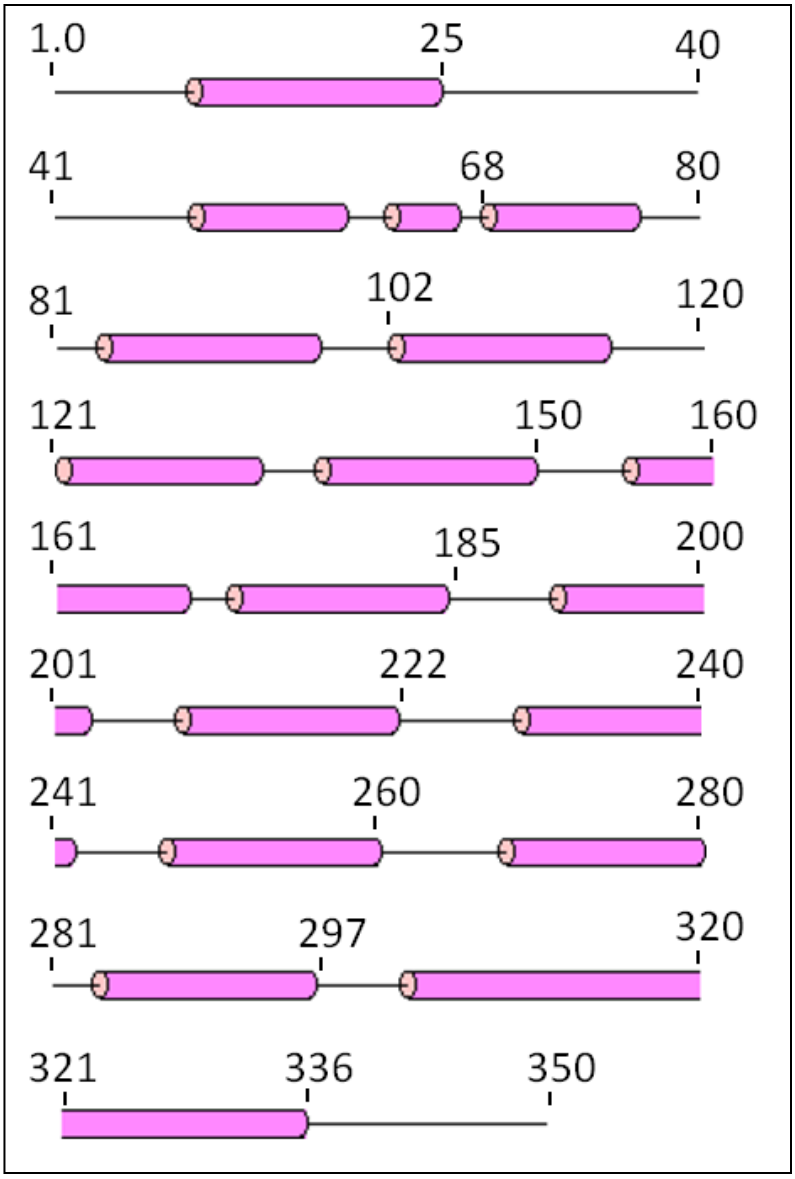


Figure 4.3: PSIPRED predicted secondary structure of PPR41.

Shown are: (i) the N-terminal organellar targeting signal encompassing a putative disordered end followed by an isolated α-helix and disordered end; (ii) the 7-8 putative α-helical pairs of the main polypeptide.

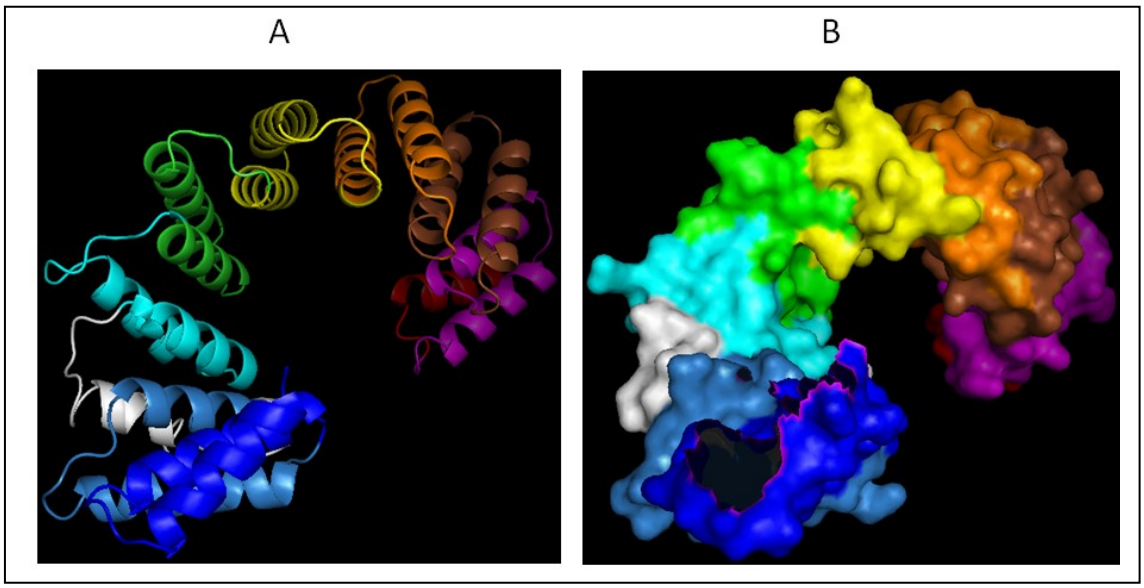


Figure 4.4: Predicted 3D structure of PPR41 showing 8 putative antiparallel helical modules organized into a superhelix with a central groove.

(A) Ribbon view. (B) Surface view. In both (A) and (B), the putative PPR modules are residues 26-60 (deep blue), 62-96 (sky blue), 122-156 (cyan), 157-191 (green), 194-228 (yellow), 231-265 (orange), 266-300 (brown), and 301-335 (magenta). Represented in grey is a putative intervening linker helix (residues 97-121) between the sky blue (2^{nd}) and cyan (3^{rd}) PPR modules. The 3D model was generated by I-TASSER (*32*) using the crystal structure of the tetratricopeptide repeat (TPR) domain of O-linked GlcNAc transferase (*34*) as template. Sequence identity between PPR41 and the aligned region of the template was 12.8%.

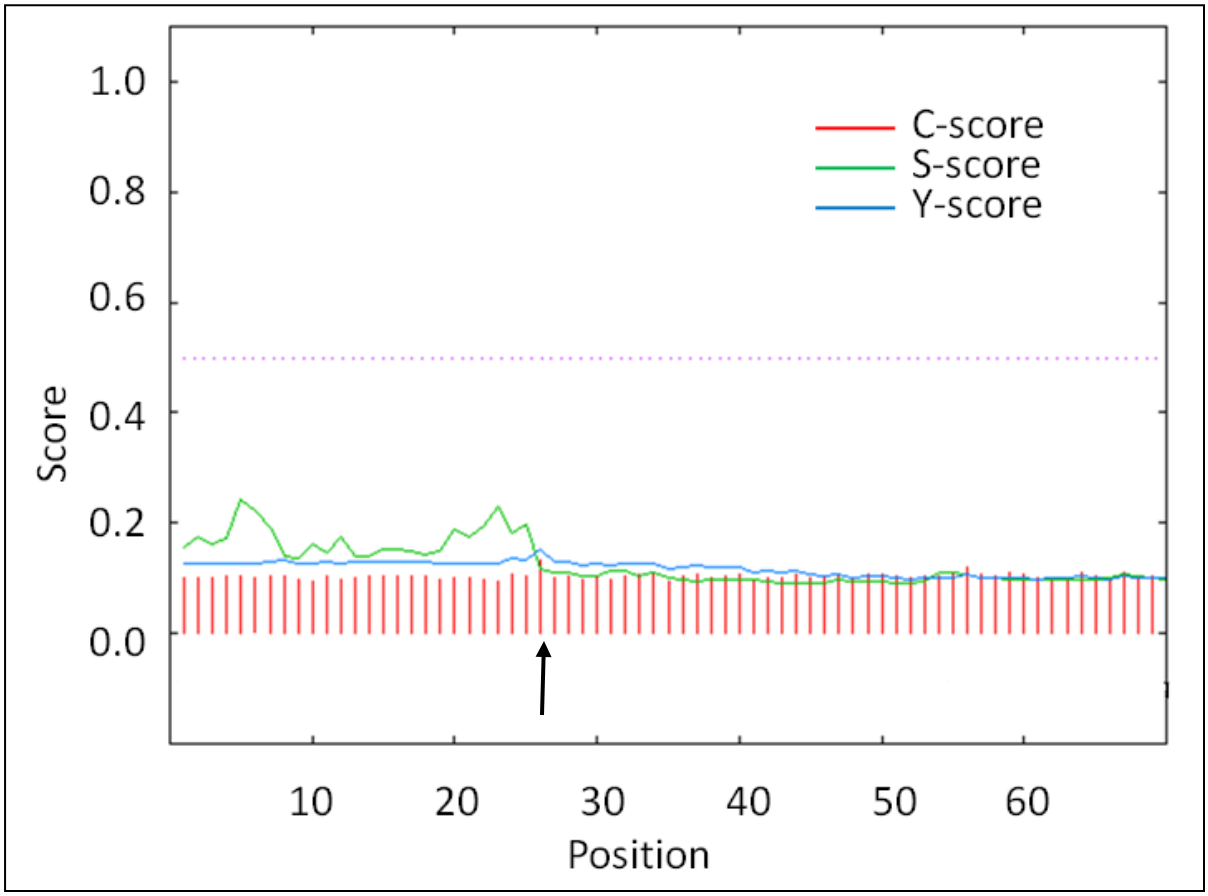


Figure 4.5: Trace showing the estimation of signal peptide cleavage site by SignalP 4.0.

A spike in the cleavage score (C-score) at G26 (black arrow) is an indicator of the signal peptidase cleavage site.

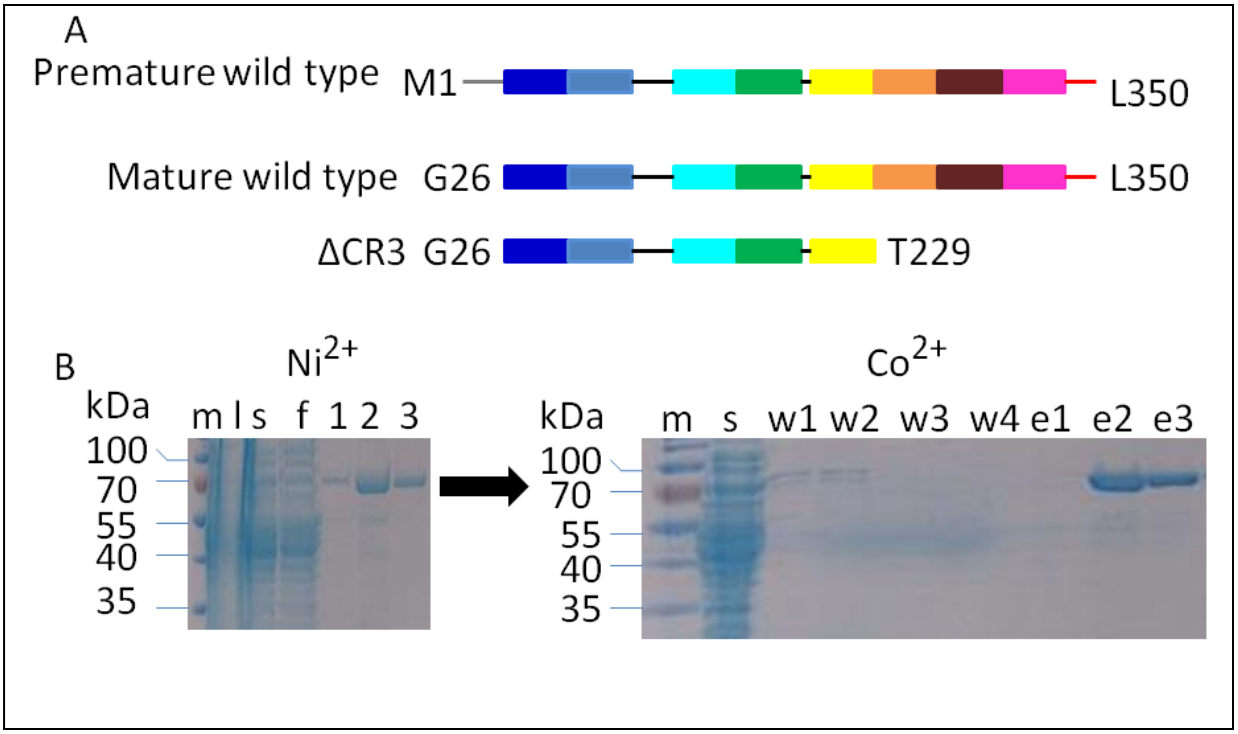


Figure 4.6: Construct design and purification of MBPPPR41.

(A) Schematic diagram showing the design of mature wild type- and ΔCR3 forms of PPR41 from the gene encoded polypeptide; each colored rectangular block is a PPR module. (B) 5-12% discontinuous SDS-PAG (sodium dodecylsulfate polyacrylamide gels) showing the tandem purification of the MBP fusion protein of ΔCR3 (67.6 kDa). In both panels, m is the size marker and s is the supernatant after clarification of bacterial cell lysate. In the left panel (Ni^{2+}), f is unbound flow through; 1-3 are consecutive eluates 1-3. In the right panel (Co^{2+}), w1-w4 are amylose column washes; e1-e3 are amylose column eluates.

4.3.2 PPR41 has strong affinity for GU-rich RNA

The RNA binding specificity of PPR41 was investigated under both nonequilibrium and equilibrium conditions. Despite the PPR41 protein used in the assays lacking its last three C-terminal PPR modules, robust RNA binding activity was still observed. Under non-equilibrium conditions (EMSA), stable binding was most prominent with poly(G) among the ssRNA homopolymers (**Figure 4.7**). However equilibrium assays (fluorescence anisotropy) found PPR41 to have strong (sub-micromolar) affinity for G_{12}, U_{12}, and $(GGU)_4$ ssRNA, modest affinity for A_{12}, and reduced affinity for C_{12} ssRNA (**Figure 4.8A-B and Table 4.2**). These results are consistent with the biological RNA target for PPR41 being GU-rich. It was demonstrated previously that MBP lacks RNA binding activity ((*19*); Chapter 2), hence the observed RNA binding activity is due to PPR41.

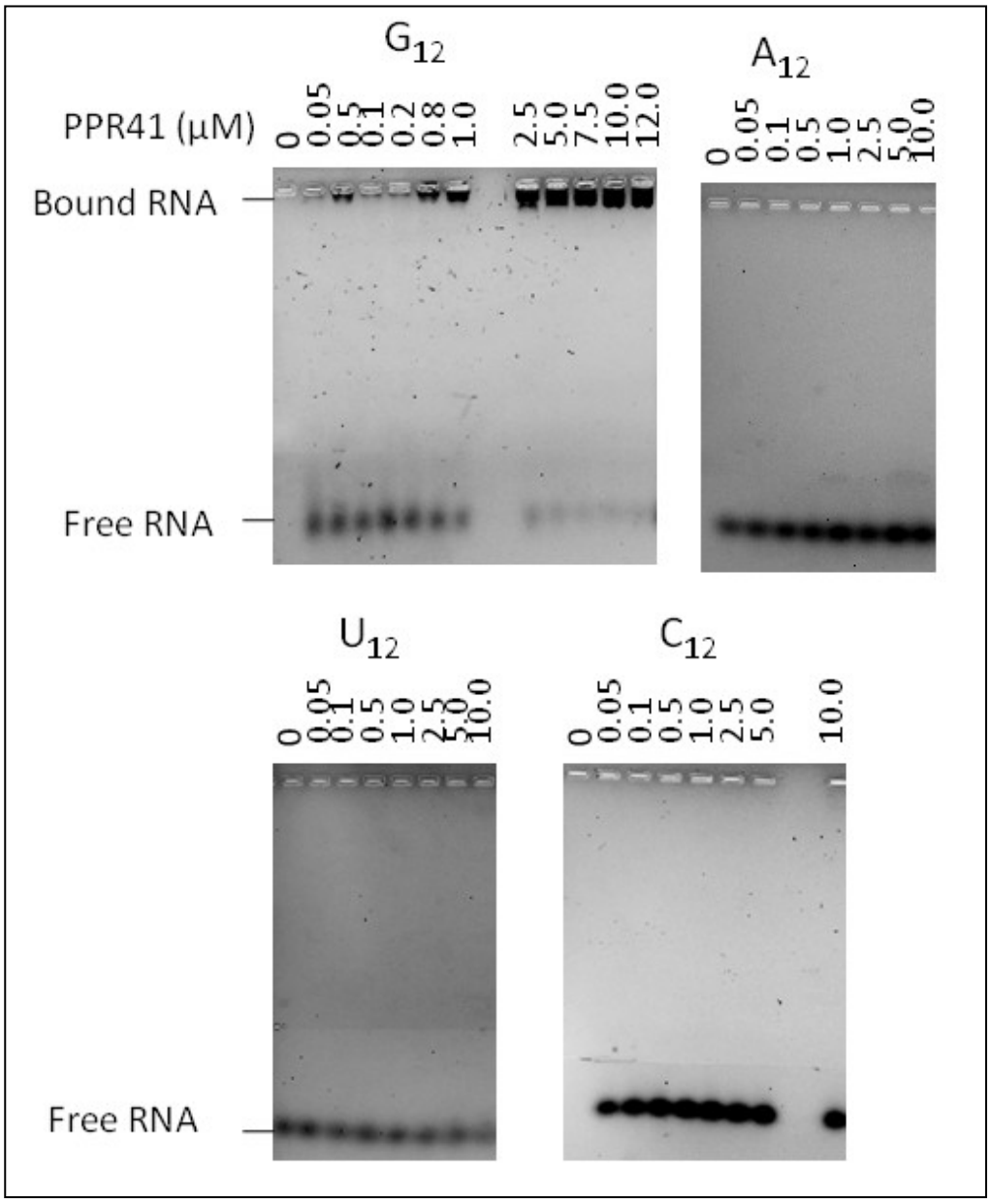


Figure 4.7: 1.5 % tris-glycine agarose gels showing retardation of G_{12} and little effect on A_{12}, U_{12}, or C_{12} with increasing protein concentration.

The RNA binding reaction contained 0.15 μM 5'-FLUO labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT, 0.1 μg/μl BSA, 0.4 U/μl RNasin), and MBPPPR41.

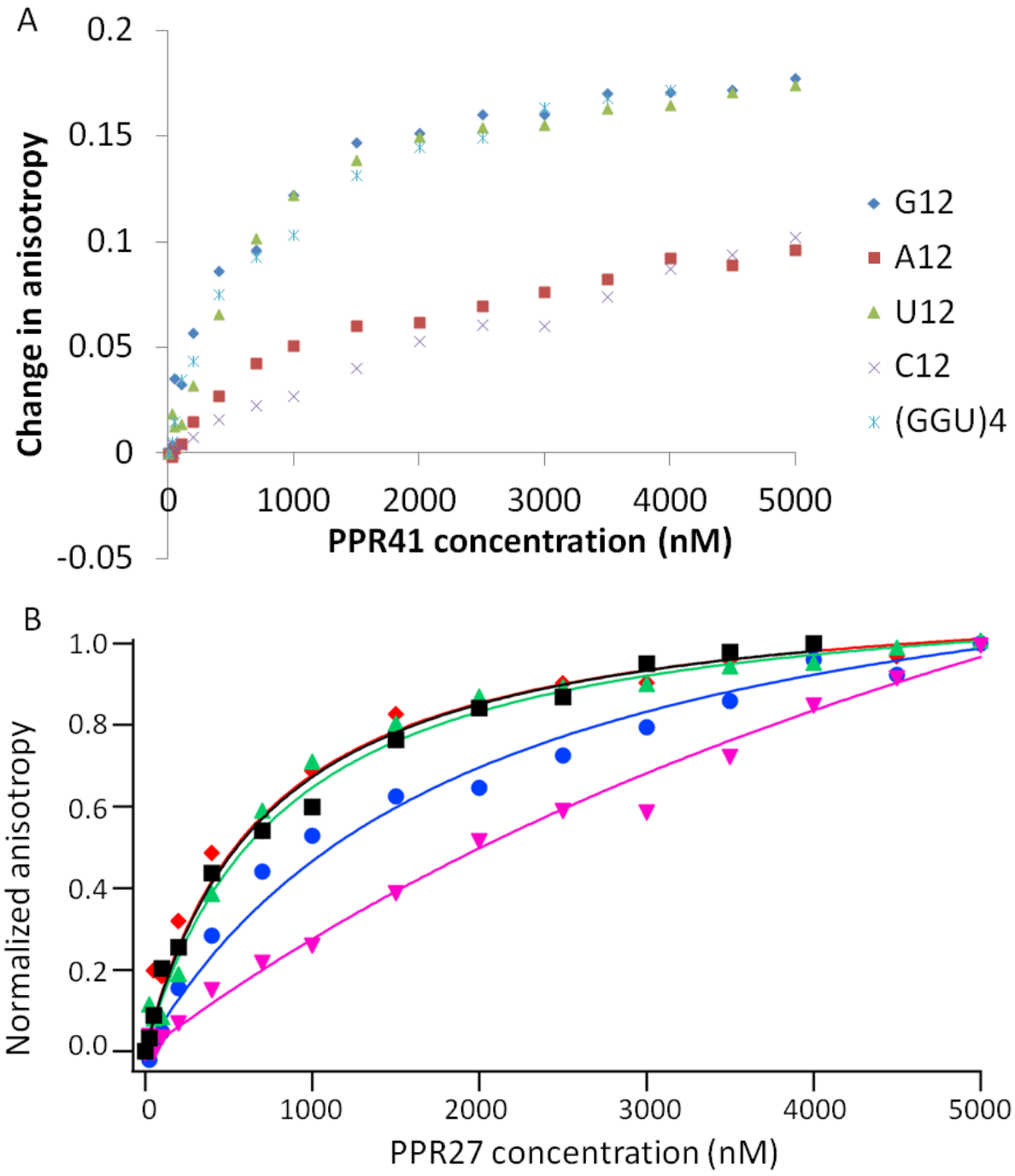


Figure 4.8: Interaction of PPR41 with various ssRNA oligonucleotides.

(A) Plot of the increase in RNA anisotropy (relative to free RNA) with protein concentration demonstrating the high affinity of PPR41 for G_{12}, U_{12}, and $(GGU)_4$ ssRNA compared to A_{12} and C_{12}.

Fig 4.8 (cont'd).

(B) Nonlinear fits of the normalized RNA anisotropy versus protein concentration for each of the RNAs in (A); G_{12} (red closed squares), U_{12} (green upward triangles), $(GGU)_4$ (black closed squares), A_{12} (blue closed circles), and C_{12} (magenta inverted closed triangles). The RNA binding reactions contained 20 nM 5'-FLUO labeled ssRNA, RNA binding buffer (20 mM Tris-HCl, pH 7.5, 150 mM KCl, 5.0 mM $MgCl_2$, 1.0 mM DTT), and MBPPPR41.

Table 4.2: PPR41's affinity for various ssRNA oligonucleotides

RNA	K_d (μM) from FA
G_{12}	0.70 ± 0.00
U_{12}	0.81 ± 0.00
$(GGU)_4$	0.72 ± 0.06
A_{12}	1.94 ± 0.06
C_{12}	*

*Reliable K_d could not be determined as binding isotherm was far from saturation within the 5.0 μM limits of the performed titration.

4.3.3 Precursor RNA transcripts are among PPR41 targets

Efforts were made to isolate the biological RNA target(s) for PPR41 from total *T. brucei* mitoRNA using a pull-down experiment. Pulled down RNA was analyzed by RT-PCR and agarose gel electrophoresis (**Figure 4.9**). For the intergenic region between the 12S and 9S rRNA, three bands were detected: one band was of the expected size (about 450 bp) (*35*), one was about 700 bp which was larger than expected, and one was about 180 bp which was smaller than expected. For the intergenic stretch between 9S rRNA and ND8 mRNA, one band of the expected size (about 180 bp) was observed. RT-PCR for both the 5' end of ND7 mRNA and its full transcript were negative, though this could have also been due to absence of ND7 mRNA in the procyclic mitoRNA extract since ND7 expression is downregulated in procyclic parasites (*36*).

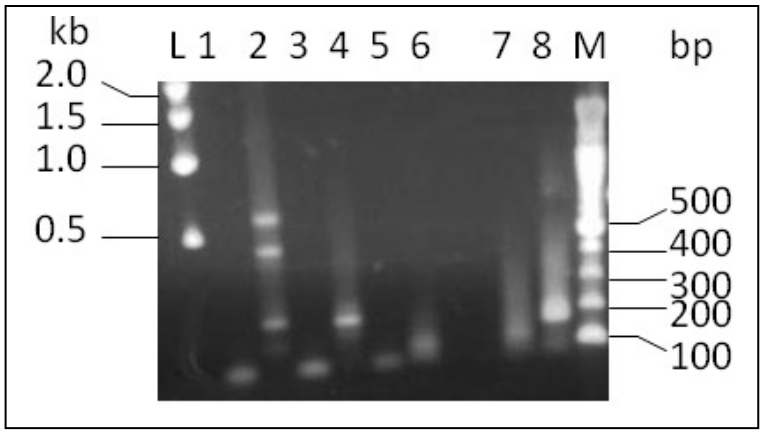


Figure 4.9: 1% agarose gel showing the electrophoretic mobility of PCR amplicons of PPR41-bound *T. brucei* kRNA.

L, 1 kb DNA marker; M, 100 bp DNA marker; 1, 3, 5, and 7 are negative control reactions (MQ-H_2O) for primer pairs in lanes 2, 4, 6 and 8 respectively; 2, RT-PCR for intergenic region between 12S and 9S rRNAs; 4, RT-PCR for intergenic region between 9S rRNA and ND8 mRNA; 6, RT-PCR for the 5' terminal region of ND7 mRNA; 8, RT-PCR for full ND7.

4.4 DISCUSSION

The cognate RNA ligands for almost all *T. brucei* PPR proteins remain unclear. Earlier, we demonstrated for the first time the sequence specificity of RNA ligands for *T. brucei* PPR27 (chapter 2). Using a similar approach, we now demonstrate for the first time that PPR41 binds RNA and that it prefers RNA rich in guanosines and uridines over other sequences. Nevertheless, despite a similar K_d under equilibrium conditions, the complex of poly(U) ssRNA with PPR41 does not withstand the stringency of electrophoresis (nonequilibrium assay) suggesting that RNA off rates may have a role in the biological function of PPR41. Sub-micromolar dissociation constants were observed with the protein construct lacking 3 C-terminal PPR motifs. Difficulties in production of mature wild type PPR41 have so far prevented us from determining the RNA binding affinity of the biological form of the protein. However, insights from the *Arabidopsis* PPR protein HCF152 (*37*) as well as from our own *T. brucei* PPR27 (chapter 2) show that wild type proteins typically bind RNA the tightest. In the biological context in which it exists with all its PPR motifs, PPR41 is therefore likely to bind RNA even tighter than observed with its ΔCR3 variant.

In the *T. brucei* mitochondria, maxicircle pre-edited mRNA transcripts have a high density of guanosine tracts (G-tracts) (*38*) which are broken up to form GU-rich combinations in partially edited and fully edited mRNA (*39-40*). On the other hand, rRNAs, gRNAs, and never-edited mRNA transcripts have high density of uridines and adenosines (*41-42*). The high GU-rich RNA binding affinity and the modest interaction with A-rich RNA therefore suggests that in the presence of competing sequences as is the case in vivo, PPR41 preferentially recognizes edited RNA over other RNA types. A search for both pre-edited and edited forms of ND7 mRNA within the pulled down material bordered from negative to indeterminate. On the other hand RT-

PCR analysis of pulled down material was positive for precursor RNA transcripts, though it also came up with two extra bands whose basis is unclear. Sequencing of the RTPCR products in the gel bands is anticipated to resolve the conundrum. Furthermore, extension of RT-PCR analysis to cover more RNA species should also narrow down the scope of RNA recognized by PPR41 in vivo. Ultimately, the biochemical rationale for recognizing sequence elements within these RNAs awaits genetic studies.

REFERENCES

REFERENCES

1. van Hellemond, J. J., Opperdoes, F. R., and Tielens, A. G. (2005) The extraordinary mitochondrion and unusual citric acid cycle in Trypanosoma brucei, *Biochem Soc Trans 33*, 967-971.

2. Campbell, D. A., Thomas, S., and Sturm, N. R. (2003) Transcription in kinetoplastid protozoa: why be normal?, *Microbes Infect 5*, 1231-1240.

3. Palenchar, J. B., and Bellofatto, V. (2006) Gene transcription in trypanosomes, *Mol Biochem Parasitol 146*, 135-141.

4. Clayton, C., and Shapira, M. (2007) Post-transcriptional regulation of gene expression in trypanosomes and leishmanias, *Mol Biochem Parasitol 156*, 93-101.

5. Clayton, C. E. (2002) Life without transcriptional control? From fly to man and back again, *Embo Journal 21*, 1881-1888.

6. Feagin, J. E. (2000) Mitochondrial genome diversity in parasites, *Int J Parasitol 30*, 371-390.

7. Koslowsky, D. J. (2009) Complex interactions in the regulation of trypanosome mitochondrial gene expression, *Trends Parasitol 25*, 252-255.

8. Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L., Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., Melville, S. E., and El-Sayed, N. M. (2005) The genome of the African trypanosome Trypanosoma brucei, *Science 309*, 416-422.

9. El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler,

E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., and Andersson, B. (2005) The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease, *Science 309*, 409-415.

10. Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchettin, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabbinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., and Myler, P. J. (2005) The genome of the kinetoplastid parasite, Leishmania major, *Science 309*, 436-442.

11. Panigrahi, A. K., Ogata, Y., Zikova, A., Anupama, A., Dalley, R. A., Acestor, N., Myler, P. J., and Stuart, K. D. (2009) A comprehensive analysis of Trypanosoma brucei mitochondrial proteome, *Proteomics 9*, 434-450.

12. Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes, *Mol Cell 42*, 106-117.

13. D'Andrea, L. D., and Regan, L. (2003) TPR proteins: the versatile helix, *Trends Biochem Sci 28*, 655-662.

14. Small, I. D., and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins, *Trends Biochem Sci 25*, 46-47.

15. Mingler, M. K., Hingst, A. M., Clement, S. L., Yu, L. E., Reifur, L., and Koslowsky, D. J. (2006) Identification of pentatricopeptide repeat proteins in Trypanosoma brucei, *Mol Biochem Parasitol 150*, 37-45.
16. Pusnik, M., Small, I., Read, L. K., Fabbro, T., and Schneider, A. (2007) Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes, *Mol Cell Biol 27*, 6876-6888.

17. Schmitz-Linneweber, C., and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression, *Trends Plant Sci 13*, 663-670.

18. Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts, *Mol Cell Biol 28*, 5337-5347.

19. Kroeger, T. S., Watkins, K. P., Friso, G., van Wijk, K. J., and Barkan, A. (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing, *Proc Natl Acad Sci U S A 106*, 4537-4542.

20. Okuda, K., Hammani, K., Tanz, S. K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid ndhB and ndhG transcripts, *Plant J 61*, 339-349.

21. Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing, *Journal of Biological Chemistry 281*, 37661-37667.

22. Pfalz, J., Bayraktar, O. A., Prikryl, J., and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts, *EMBO J 28*, 2042-2052.

23. Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giege, P. (2008) PPR336 is associated with polysomes in plant mitochondria, *Journal of Molecular Biology 375*, 626-636.

24. Williams-Carrier, R., Kroeger, T., and Barkan, A. (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand, *RNA 14*, 1930-1941.

25. Pusnik, M., and Schneider, A. (2012) A trypanosomal pentatricopeptide repeat protein stabilizes the mitochondrial mRNAs of cytochrome oxidase subunits 1 and 2, *Eukaryot Cell 11*, 79-87.

26. Zikova, A., Panigrahi, A. K., Dalley, R. A., Acestor, N., Anupama, A., Ogata, Y., Myler, P. J., and Stuart, K. (2008) Trypanosoma brucei mitochondrial ribosomes: affinity

purification and component identification by mass spectrometry, *Mol Cell Proteomics 7*, 1286-1296.

27. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res 25*, 3389-3402.

28. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nat Methods 8*, 785-786.

29. Karpenahalli, M. R., Lupas, A. N., and Soding, J. (2007) TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences, *BMC Bioinformatics 8*, 2.

30. Biegert, A., and Soding, J. (2008) De novo identification of highly diverged protein repeats by probabilistic consistency, *Bioinformatics 24*, 807-814.

31. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server, *Bioinformatics 16*, 404-405.

32. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics 9*, 40.

33. Richardson, S. M., Wheelan, S. J., Yarrington, R. M., and Boeke, J. D. (2006) GeneDesign: rapid, automated design of multikilobase synthetic genes, *Genome Res 16*, 550-556.

34. Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A., and Conti, E. (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha, *Nat Struct Mol Biol 11*, 1001-1007.

35. Koslowsky, D. J., and Yahampath, G. (1997) Mitochondrial mRNA 3' cleavage/polyadenylation and RNA editing in Trypanosoma brucei are independent events, *Mol Biochem Parasitol 90*, 81-94.

36. Schneider, A. (2001) Unique aspects of mitochondrial biogenesis in trypanosomatids, *Int J Parasitol 31*, 1403-1415.

37. Nakamura, T., Meierhoff, K., Westhoff, P., and Schuster, G. (2003) RNA-binding properties of HCF152, an Arabidopsis PPR protein involved in the processing of chloroplast RNA, *European Journal of Biochemistry 270*, 4070-4081.

38. Jasmer, D. P., Feagin, J. E., Payne, M., and Stuart, K. (1987) Variation of G-rich mitochondrial transcripts among stocks of Trypanosoma brucei, *Mol Biochem Parasitol 22*, 259-272.

39. Lukes, J., Hashimi, H., and Zikova, A. (2005) Unexplained complexity of the mitochondrial genome and transcriptome in kinetoplastid flagellates, *Curr Genet 48*, 277-299.
40. Hong, M., and Simpson, L. (2003) Genomic organization of Trypanosoma brucei kinetoplast DNA minicircles, *Protist 154*, 265-279.

41. Eperon, I. C., Janssen, J. W., Hoeijmakers, J. H., and Borst, P. (1983) The major transcripts of the kinetoplast DNA of Trypanosoma brucei are very small ribosomal RNAs, *Nucleic Acids Res 11*, 105-125.

42. Sloof, P., Van den Burg, J., Voogd, A., Benne, R., Agostinelli, M., Borst, P., Gutell, R., and Noller, H. (1985) Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: a detailed comparison of the 9S and 12S RNAs from Crithidia fasciculata and Trypanosoma brucei with rRNAs from other organisms, *Nucleic Acids Res 13*, 4171-4190.

CHAPTER 5

PURIFICATION AND REFOLDING OF *TRYPANOSOMA BRUCEI* WILD TYPE PPR27 FROM THIOREDOXIN-PPR27 INCLUSION BODIES

This portion of the thesis reports successful efforts to produce free PPR27 by protein refolding.

ABSTRACT

The flagellate protozoan *Trypanosoma brucei* (*T. brucei*) causes an invariably fatal disease, human African trypanosomiasis (HAT) or sleeping sickness. There is no anti-HAT vaccine yet available drugs are few, toxic, and sub-efficacious. New drugs are thus needed. Recently, RNA-binding pentatricopeptide repeat (PPR) proteins were identified as potential drug targets as they are highly conserved among trypanosomatids. Understanding the structure and mechanism of RNA binding by PPR proteins is however important to drug design. Unfortunately, structural studies of PPR proteins are hindered by difficulties in heterologous protein production, owing to their poor aqueous solubility. We sought to produce the 27 kDa PPR protein (PPR27), the smallest among at least 37 *T. brucei* PPR proteins, in *Escherichia coli*. PPR27 overexpression was only successful after fusion to a large soluble tag. As reported earlier, with the maltose binding protein as soluble carrier, low amounts of soluble fusion protein are obtained with the rest segregating into inclusion bodies. With thioredoxin as the soluble tag, there is huge accumulation of protein, but all partitions into inclusion bodies. Using N-lauroyl sarcosine as stabilizer and Factor Xa protease to remove thioredoxin, we have obtained free PPR27 by metal affinity chromatography and successfully refolded it by dialysis.

5.1 INTRODUCTION

Safe and potent treatments for the neglected tropical diseases (NTDs) (*1*) such as human African Trypanosomiasis (HAT) or "sleeping sickness", the disease caused by the flagellate protozoan *Trypanosoma brucei* (*T. brucei*), are lacking. In the last two decades however, numerous putative drug targets have been uncovered with the advent of genomics. In *T. brucei*, the pentatricopeptide repeat (PPR) family of RNA binding proteins is among those which have elicited great interest for drug targeting as they are essential to the parasite (*2-3*) and lack close homologues in humans. The PPR motif is a eukaryote-only loosely conserved repeat of 35 amino acids each of which folds into a pair of antiparallel α-helices (*4*). *T. brucei* has at least 37 PPR proteins (*5*). Typically, PPR motifs dominate the polypeptide backbone of PPR proteins (*6*) but there are also nonclassical cases such as human mitochondrial RNA polymerase (mtRNAP) (*7*) and the *Arabidopsis thaliana* protein-only mitochondrial RNase P (PRORP1) (*8*) in which PPR modules occupy a minor portion of the polypeptide.

For in silico drug design, the 3-dimensional structure of the target/template is required. Unfortunately, poor aqueous solubility (*9-11*) and toxicity to *E. coli* expression hosts have hindered production of PPR proteins in sufficient amounts and quality for structural biology, and no 3D structure of a classical PPR protein has been solved to date. Solubility enhancing protein tags such as the maltose binding protein (MBP), thioredoxin (TRX), and glutathione S-transferase (GST) have aided expression of PPR proteins in *E. coli* lately (*12-17*). But while fusion proteins are compatible with some in vitro applications such as biochemical assays, they are not suitable for high resolution structural biology. For X-ray crystallography, large tags introduce conformational flexibility and/or heterogeneity which hinder crystallization (*18-20*) while for nuclear magnetic resonance (NMR) spectroscopy, large tags increase signal overlap

and line broadening which undermine chemical shift assignments (*21-22*). PPR proteins have been successfully resolved from large solubility tags only rarely, and even then yields have often been low for structural biology. For example, Barkan and coworkers have reported yields of 0.1 mg and 0.5 mg per liter of culture for maize CRP1 and PPR5 respectively following enzymatic cleavage and column resolution from MBP (*23*). Yet still, some PPR fusion proteins segregate into inclusion bodies (IBs) (*24*).

IBs can be an invaluable source for recombinant protein as they enable the bacterial host to accommodate huge amounts of poorly soluble and toxic proteins (sometimes up to over 50% of cellular protein) without being harmed (*25-26*). But IBs comprise unfolded and misfolded protein (*25-28*) and pose the challenge of solubilization and refolding into native state. Here, we present a scheme for the purification of mature wild type *T. brucei* PPR27 in native form from IBs of its fusion with TRX. PPR27 is the smallest of *T. brucei* PPR proteins with a molecular mass of 27 kDa. We have previously used the MBP fusion of PPR27 in RNA binding assays because purification of free PPR27 after detachment of MBP was unsuccessful (chapter 2). Prior attempts at expressing PPR27 with only a hexahistidine (His_6) tag were also futile (data not shown). This scheme is therefore a major step in recombinant production of free PPR27 and has potential utility to other recalcitrant proteins.

5.2 MATERIALS AND METHODS

5.2.1 Materials

BL21 (DE3), Rosetta (DE3), C41 (DE3), C43 (DE3), and C43 (DE3)pLysS *E. coli* expression strains and PPR27pRSFDuet-1 recombinant plasmid were a kind gift from Dr. Jennifer Ekstrom (formerly of Michigan State University). Dry, desalted PCR primers were ordered from the

Macromolecular Structure Facility of Michigan State University (East Lansing, MI) and reconstituted in nuclease free H_2O. Other materials with their sources in parentheses: 2X PCR master mix (Promega Corporation, Madison, WI); pET-32 Xa/LIC vector, NovaBlue GigaSingles *E. coli*, dGTP, T4 DNA polymerase, 10X T4 DNA polymerase, Factor Xa protease and Xarrest™ agarose(Novagen®/EMD4Biosciences, Gibbstown, NJ); Talon CellThru resin (Clontech Laboratories, Mountain View, CA); Ni-NTA agarose (Qiagen GmbH, Hilden, Germany); QIAprep spin miniprep- and QIAquick gel extraction kits (Qiagen Sciences, Germantown, MD); polyvinylidene fluoride (PVDF) membrane and amicon centrifugal filters (EMD Millipore, Billerica, MA) Slide-A-Lyzer® cassettes (Thermo Scientific, Rockford, IL); 6-Cyclohexyl-1-Hexyl-β-D-Maltoside (CYMAL-6), 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS), N-lauroyl sacrcosine (sarkosyl), n-Octyl-β-D-Glucopyranoside (BOG) (Anatrace-Affymetrix, Maume, OH); n-Dodecyl-β-D-Maltopyranoside (DDM) (Calbiochem®/EMD4Biosciences, Gibbstown, NJ); and Triton X-100 (Sigma-Aldich, St. Louis, MO).

5.2.2 Cloning

DNA for mature PPR27 DNA (lacking the first 22 amino acids of the mitochondrial targeting signal) was subcloned into a pET-32 Xa/LIC vector by ligation independent cloning (LIC) (*29-30*) according to the manufacturer's instructions (Novagen®/EMD4Biosciences). The vector encodes a 53 residue spacer (S) separating thioredoxin (TRX) from PPR27 (**Figure 5.1**). Within the spacer are an N-terminal His_6 tag and a C-terminal Factor Xa protease cleavage site which facilitate resolution of TRX from PPR27 during protein purification. PPR27 DNA was amplified

from a PPR27pRSFDuet-1 plasmid using a pair of primers containing at least 15 vector-encoded nucleotides or LIC sites, namely 5'GGTATTGAGGGTCGCGGTCACGTGTACGCCCTTCA3' (forward primer) and 5'AGAGGAGAGTTAGAGCCTCAACCACGAGGTAAAGTGG3' (reverse primer). PCR conditions were one round of initial denaturation at 95 ^{o}C for 2.0 min; 35 rounds of 95 ^{o}C denaturation for 30 s, 55 ^{o}C annealing for 45 s, and 72 ^{o}C extension for 2.0 min; a final round of 72 ^{o}C extension for 5 min. The PCR reaction mixture was resolved on a 1% agarose gel in Tris acetate EDTA (TAE) buffer (40 mM Tris, 1.0 mM EDTA, pH 8.0, 20 mM acetic acid glacial). Amplicons were recovered from the gel using the QIAquik gel extraction kit. They were then digested with T4 DNA polymerase to create long sticky ends (LIC overhangs), and subsequently annealed into the pET32 Xa/LIC vector following the manufacturer's instructions. The annealed plasmids were propagated in NovaBlue GigaSingles *E. coli* followed by isolation of plasmid DNA using the QIAprep® spin miniprep. The presence of PPR27 was confirmed by both PCR (using T7 promoter- and terminator primers) and DNA sequencing, the latter using dideoxynucleotide dye terminators at the Genomics Core facility of Michigan State University. Sequencing PCR was done using the T7 terminator primer.

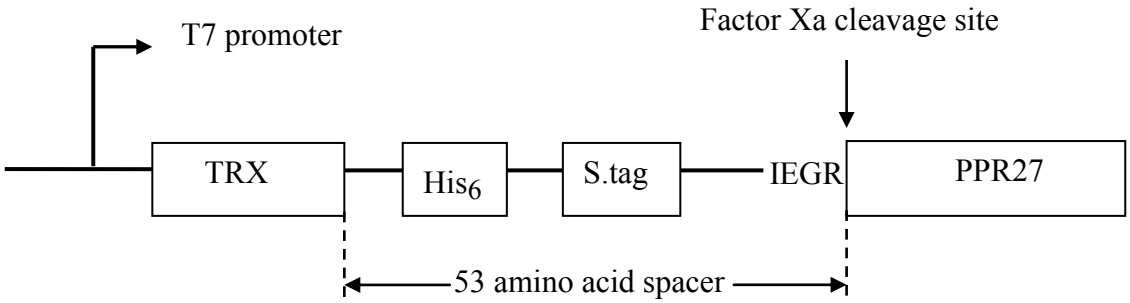


Figure 5.1: Design of TRXPPR27.

Scheme showing the peptide constituents of TRXPPR27; note the location of His_6, the S-peptide (S.tag), and the Factor Xa protease cleavage site in the 53 amino acid spacer between TRX and PPR27.

5.2.3 TRXPPR27 fusion protein expression

We first established the suitable *E. coli* expression strains, IPTG (isopropyl-β,D-thiogalactopyranoside) concentration, and expression temperature through induction analysis in Luria Bertani (LB) medium. Between 0.1 mM and 1.0 mM IPTG, there wasn't any difference in TRXPPR27 yield and the protein segregated into inclusion bodies at all temperatures and IPTG concentrations (data not shown). On the other hand, assessment of five *E. coli* expression strains showed TRXPPR27 overexpression in BL21 (DE3), Rosetta (DE3), and C41 (DE3) but not C43 (DE3) and C43 (DE3)pLysS (data not shown). Further exploration showed some truncated protein contamination in TRXPPR27 purified eluates following expression in BL21 (DE3). Thus, Rosetta (DE3) became the expression strain of choice going forward. The PPR27pET-32 Xa plasmid was transformed into the bacteria by heat shock followed by selection of transformants on LB agar containing 50 μg/ml ampicillin and 34 μg/ml chloramphenicol at 37 ^{o}C for 16 hours. Single colonies were then inoculated into 10 ml LB broth and grown overnight in an innova4200 incubator shaker (New Brunswick Scientific, Edison, NJ) at 250 rpm and 37 ^{o}C overnight. 20 ml of the latter cultures were used to seed 980 ml of LB broth to generate the bulk cultures for protein expression. Bulk cultures were grown at 37 ^{o}C, 250 rpm till an OD_{600} of about 0.8 when heterologous expression was induced with 0.1 mM IPTG. Post-IPTG cultures were incubated at 30 ^{o}C, 150 rpm for 6 hours, after which they were harvested by centrifugation in a SORVALL® RC5 PLUS centrifuge (Kendro Laboratory Products, Newtown, CT) at 5000 rpm, 4 ^{o}C, for 15 minutes. Pelleted cells were either resuspended in lysis buffer for purification or frozen at -80 ^{o}C.

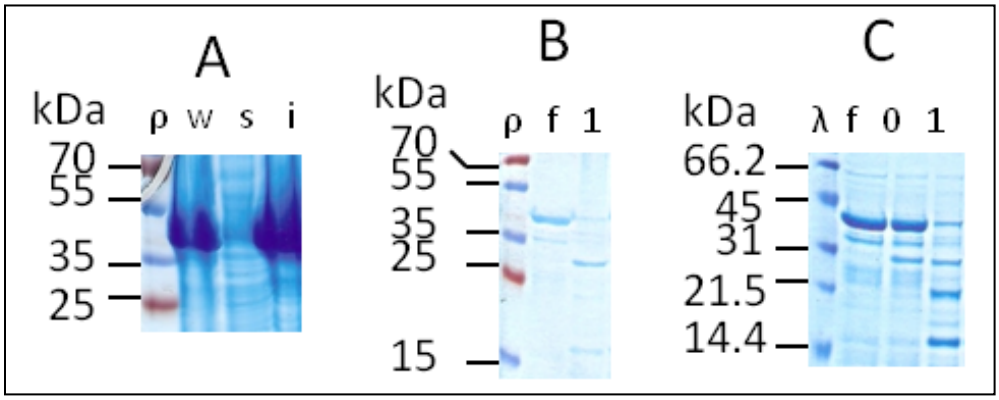


Figure 5.2: Coomassie blue stained 12% polyacrylamide gels showing TRXPPR27 purification and Factor Xa protease test cleavage.

(A) Gel of native lysates of BL21(DE3) cells showing that TRXPPR27 sequesters into inclusion bodies; ρ, size marker; w, whole cell lysates; s, supernatant; and i, inclusion bodies. (B) Gel of the Factor Xa protease cleavage products after reaction in Novagen's standard cleavage buffer; ρ, size marker; f, uncleaved TRXPPR27; 1, cleavage reaction mixture. (C) Gel of the Factor Xa protease cleavage products after reaction in buffer SCB, free PPR27 being the band near 21.5 kDa; λ, size marker; f, uncleaved TRXPPR27; 0, cleavage reaction mixture at start of reaction (or t = 0); 1, reaction mixture after over 12 hours of reaction. The band between 25 and 35 kDa in (B) and slightly below 31 kDa in (C) is for Factor Xa protease. The band near 15 kDa in (B) and 14.4 kDa in (C) is for TRX.

5.2.4 Determination of TRXPPR27 yield

E. coli cells were lyzed by five 20 s sonication cycles in native lysis buffer, NLB (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol (β-ME) using a microtip (Heat Systems Ultrasonics, Farmingdale, NJ). After centrifugation of lysate at 14,000 rpm in a SORVALL® RC5 PLUS centrifuge, all TRXPPR27 segregated into the insoluble fraction (IBs) as soluble host contaminants were lost in the supernatant (**Figure 5.2A**). To quantitate the gross yield of TRXPPR27 per liter of culture, complete solubilization of IBs was done using10 ml of the iFOLD™ protein refolding system 2 (Novagen®/EMD4Biosciences) guanidine denaturation buffer (50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2.0 mM EDTA, 7.0 M GuHCl) per gram of paste by stirring at room temperature. The protein concentration in the solubilizate was derived from the absorbance of U.V radiation at 280 nm (A_{280}) using Richard's protein calculator (http://www.mrc-lmb.cam.ac.uk/ms/methods/proteincalculator.html), a tool which auto-computes the protein's molar extinction coefficient (ε) and molecular mass from the amino acid sequence.

5.2.5 Free PPR27 purification

IBs were isolated from *E. coli* cells as in the last section. For optimization of Factor Xa reaction conditions, IBs were solubilized in 8 ml of Qiagen's buffer B (100 mM NaH_2PO_4, 10 mM Tris.Cl, pH 8.0, 8.0 M urea) per gram of paste. Denatured supernatant was loaded onto a divalent metal affinity column equilibrated with lysis buffer and purified under gravity at room temperature. Column was then washed with buffer NLB followed by eluation with native elution buffer, NEB (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 200 mM imidazole, 5 mM β-ME).

Protein was then exchanged into Factor Xa cleavage buffer by repeated dilution and centrifugal filtration with 10K MWCO amicon filters. After trials, though specific cleavage was observed in Novagen's Factor Xa cleavage buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $CaCl_2$) all free PPR27 precipitated (**Figure 5.2B**) causing us to explore the use of protein stabilizers during the cleavage reaction. After examining six detergents (CYMAL-6, CHAPS, sarkosyl, BOG, DDM, and Triton X-100), the L-Arg/L-Glu combination and some commercial refolding conditions, specific activity was only observed with sarkosyl (**Figure 5.2C**). Thus, going forward, fusion protein for Factor Xa cleavage was prepared by solubilizing IBs in 8 ml of sarkosyl lysis buffer or SLB (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 5 mM β-ME, 0.4 % (w/v) sarkosyl) per g of solid paste by stirring at room temperature for 1-2 hours followed by centrifugation in a SORVALL® RC5 PLUS centrifuge at 14, 000 rpm for 30 minutes to remove insoluble debris. The supernatant was loaded onto either a Ni^{2+} (Ni-NTA) or Co^{2+} (Talon CellThru) column equilibrated with buffer NLB. The Ni^{2+} column showed superior protein binding capacity than the Co^{2+} column though fusion protein purity was better with the latter. Non-specifically bound contaminants from the column were washed off with at least 4 column volumes of buffer SLB followed by either elution of TRXPPR27 with sarkosyl elution buffer, SEB (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 200 mM imidazole, 5 mM β-ME, 0.4% (w/v) sarkosyl) for subsequent use in tube Factor Xa reaction or by addition of Factor Xa protease to the column for on-column detachment of TRX from PPR27. In the on-column Factor Xa reaction, immobilized TRXPPR27 was equilibrated with buffer sarkosyl cleavage buffer, SCB (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $CaCl_2$, 0.5% (w/v) sarkosyl) after which

50 units of Factor Xa protease diluted in 3 ml of SCB were added to the column with resin bed disruption. After settling of the resin bed, flow through was collected and poured back. The latter was repeated every 30 minutes for the first two hours of Factor Xa reaction after which incubation of the column continued at 8 ^{o}C for at least 10 more hours. Note that sarkosyl precipitates in the presence of $CaCl_2$, hence the actual concentration of sarkosyl available to stabilize free PPR27 in solution is lower than the 0.5 % (w/v) added to the cleavage buffer. For tube Factor Xa protease reaction, pure TRXPPR27 was exchanged into buffer SCB by repeated dilution and centrifugation using 10 K MWCO amicon centrifugal filters at 5000 rpm, 4 ^{o}C in an Eppendorf 5804 R centrifuge (Eppendorf, Hauppage, NY) after which Factor Xa reactions were set up and incubated at 8 ^{o}C for at least 12 hours. The reaction mixture was then loaded onto a Ni^{2+} column equilibrated with 3 column volumes of buffer NLB to fish out TRXHis_6S and uncleaved TRXPPR27, leaving free PPR27 in the flow through and column washes. For the on-column reaction, the flow through and 3-5 column volume washes with buffer SLB were pooled and purified one more time with a Ni^{2+} column. Ni^{2+} columns rather than Co^{2+} were used to mop up TRX and TRXPPR27 because binding of His_6 to Co^{2+} after a prior exposure was poor. The specificity of the Factor Xa reaction in sarksoyl was confirmed by N-terminal sequencing and mass spectrometry. The yield and purity of fractions from both the Co^{2+} and the mop-up Ni^{2+} column were evaluated by SDS-PAGE.

5.2.6 N-terminal sequencing

The N-terminus of PPR27 was sequenced by Edman degradation using an ABIPROCISE-HT sequencer (Applied Biosystems, Carlsbad, CA) at the Macromolecular Structure Facility of Michigan State University. Briefly, 950 picomoles of PPR27 were electroblotted from a 12% SDS polyacrylamide gel onto a PVDF membrane using the XCell II™ blot module (Invitrogen Life Technologies, Grand Island, NY) with 1X CAPS transfer buffer (10 mM CAPS minimum, 10 % (v/v) methanol) at 30 V voltage for one hour. After blotting, the membrane was rinsed with MQ-H_2O followed by a 30 s soak in 100% methanol and a 30 s stain with 0.1 % (w/v) Coomassie R250. Next, the membrane was destained with fresh 100% methanol every hour for two hours to visualize PPR27 bands which were then cut out and submitted for N-terminal sequencing.

5.2.7 Mass spectrometry

The amount of salt and detergent (sarkosyl) in the sample were rendered as low as possible by three cycles of dialysis, each against 70 volumes of detergent-free solvent, using a 20K Slide-A-Lyzer® cassette at 4 ^{o}C. Dialysis involved a first round of low salt buffer (10 mM NaH_2PO_4, pH 7.5, 100 mM, 0.5 mM DTT) overnight followed by two rounds of fresh MQ-H_2O 12 hours apart. The dialysate was then concentrated to 4.5 mg/ml using 10K MWCO amicon centrifugal filters. 1 µl of this sample was then mixed with an ammonium citrate matrix and subjected to matrix assisted laser desorption time-of-flight (MALDI-TOF) mass spectrometry in an Axima CFRplus™ mass spectrometer (Shimadzu Biotech, Kyoto, Japan).

5.2.8 Refolding of PPR27

PPR27 was refolded by 3 rounds of dialysis against 100 volumes each of low salt, detergent-less buffer (10 mM NaH_2PO_4, pH 7.5, 100 mM NaCl, 0.5 mM DTT). Dialysis was done at 4 ^{o}C using 10K MWCO Slide-A-Lyzer® cassettes or 7K MWCO SnakeSkin dialysis tubing, with changes into fresh buffer occurring every 8-12 hours. Successful refolding was confirmed by circular dichroism (CD) spectroscopy.

5.2.9 Circular dichroism spectroscopy

The secondary structure of PPR27 was investigated by CD spectroscopy of 0.22 mg/ml PPR27 in low salt buffer (10 mM NaH_2PO_4, pH 7.5, 100 mM NaCl, 0.5 mM DTT) at 25 ^{o}C using a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK), 0.1 mm path length cuvette, and 0.5 nm step size. Each sample was scanned three times and the traces averaged. The CD spectrum was then deconvoluted by the CDSSTR and CONTIN spectral analysis programs in DICHROWEB (*31*) to determine the contribution of different secondary structure elements to the overall structure of PPR27. We also examined the thermal stability of PPR27 by monitoring its CD signal at 222 nm as temperature was increased from 4 ^{o}C through 92 ^{o}C at 4 ^{o}C increments.

5.2.10 Size exclusion chromatography (SEC)

1.0 ml of a mixture of PPR27, TRXHis$_6$S tag, and TRXPPR27 in 50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, 5.0 mM imidazole, 0.4% sarkosyl, and 5 mM β-ME was injected into an AKTA fast performance liquid chromatography (FPLC) system (GE Life Sciences, Pittsburgh, PA) for SEC through a HiLoad Superdex 200, 26/60 column (GE Life Sciences) at a flow rate of 1.5 ml/min.

The running buffer comprised 10 mM NaH_2PO_4, pH 7.5, 100 mM NaCl, 0.5 mM dithiothreitol (DTT), 0.2% (w/v) sarkosyl. 4.0 ml fractions were collected.

5.3 RESULTS AND DISCUSSION

Inclusion bodies can be a very rich reservoir of heterologous protein even though the protein is in a structurally and functionally compromised state (*25-28*). In this work, the fusion protein of thioredoxin and PPR27 (TRXPPR27), a protein recalcitrant to heterologous expression in the absence of a solubility enhancer at its N-terminus was overexpressed in inclusion bodies. After removal of soluble host contaminants by native lysis of *E. coli* cells, the recovered pellet comprised almost entirely of TRXPPR27 (**Figure 5.2A**). The yield of TRXPPR27 was approximated at 480 ± 98 (n = 3) mg per liter of culture and about 15% of harvested cell mass. PPR27 (24.4 kDa) constitutes about 59% of the molecular mass of TRXPPR27, hence an equivalent of about 280 mg free PPR27 is produced per liter of culture. With NMR spectroscopy, about 250 µl of PPR27 at 0.5 - 1.0 mM concentration (*32-33*) or an absolute amount of 3.0 - 6.0 mg is ideal. For crystallization, protein at 5.0-50 mg/ml concentration is required (*34-35*). With sub-microliter amounts used for coarse screening nowadays (*36*), 500 µl of protein would screen at least 500 crystallization conditions, hence at least 2.5 mg absolute amount of pure PPR27 is desirable. Therefore, given a system to purify native PPR27 from the TRXPPR27 IBs, structural biology yields of wild type PPR27 are attainable.

5.3.1 Tube versus on-column Factor Xa cleavage for PPR27 purification

Initially, the protocol for the Factor Xa reaction was done only in tubes. The Factor Xa : TRXPPR27 reaction ratio was varied between 1 unit : 20 µg and 1 unit : 90 µg with no change in

yield of free PPR27 relative to the amount of fusion protein substrate (data not shown). However, removal of the TRX tag and uncleaved fusion protein required at least 3 rounds of Ni^{2+} affinity chromatography, and significant loss of PPR27 was observed (**Figure 5.3A**). In contrast, we observed that on-column Factor Xa cleavage of TRXPPR27 produced higher yield of free PPR27 per unit enzyme and shortened the protein purification cycle to one extra round of Ni^{2+} column compared to cleavage in the tube (**Figure 5.3B**). It appears that on-column enzyme cleavage maximizes the amount of available substrate and raises the surface area for the enzyme reaction. Overall, loss of PPR27 along the purification chain was minimized.

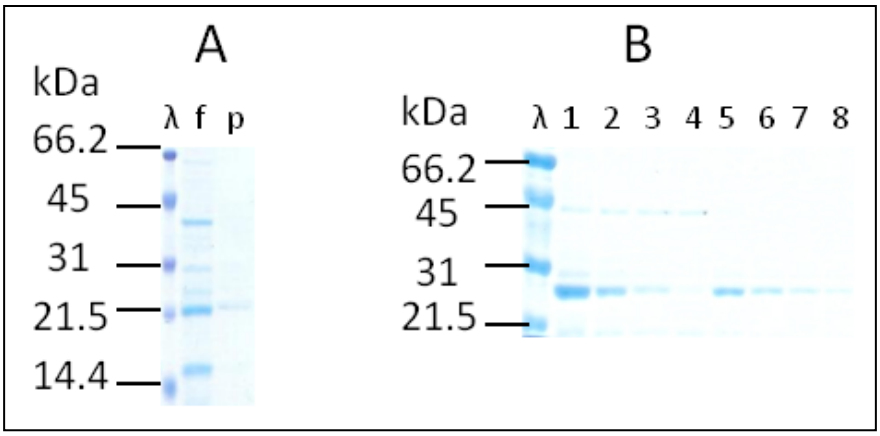


Figure 5.3: Coomassie blue stained 12% SDS-PAGE comparing tube- and on-column Factor Xa cleavage of TRXPPR27.

(A) PPR27 purified by tube cleavage of TRXPPR27 and three rounds of Ni^{2+} affinity chromatography; λ, size marker; f, TRXPPR27; p, pure PPR27. (B) PPR27 purified by Factor Xa cleavage of Co^{2+} column bound TRXPPR27 (lanes 1-4) followed by one round of Ni^{2+} purification (lanes 5-8); λ, size marker; 1, Co^{2+} column flow through; 2-4, Co^{2+} column washes 1-3; 5, Ni^{2+} column flow through; 6-8, Ni^{2+} column washes 1-3.

Besides the observed higher enzyme efficiency during the on-column reaction, it is worth noting that enzymatic cleavage of fusion proteins in tubes involves buffer exchange either by dialysis or repeated dilution and centrifugal ultrafiltration. When using detergents for protein stabilization, dialysis can be costly as protein volumes rise. And though buffer exchange by dilution and ultrafiltration is economical on detergent, precipitation from localized concentration and leakage through the filter membranes cause reasonable loss of fusion protein substrate for the Factor Xa reaction. Yet both methods of buffer exchange consume significant time resource. In contrast, on-column Factor Xa cleavage follows on-column buffer exchange, which is not only economical on detergent and time but also circumvents protein loss to concentration-induced precipitation and leakage.

5.3.2 Identification of PPR27 from Factor Xa cleavage

N-terminal sequencing and mass spectrometry provide confidence that you are not purifying a host adulterant, and where applicable that the enzymatic resolution of fusion protein components is specific. MALDI-TOF mass spectrometry of pure PPR27 showed an abundant species of 24.39 kDa and a minor species of 24.45 kDa (**Figure 5.4**), consistent with the theoretical molecular mass of PPR27. Detergents suppress ionization and must be removed prior to mass spectrometry (*37-38*). Because we used sarkosyl during protein purification, uneven ionization due to interference from residual detergent could be responsible for the polymorphism in the molecular masses obtained.

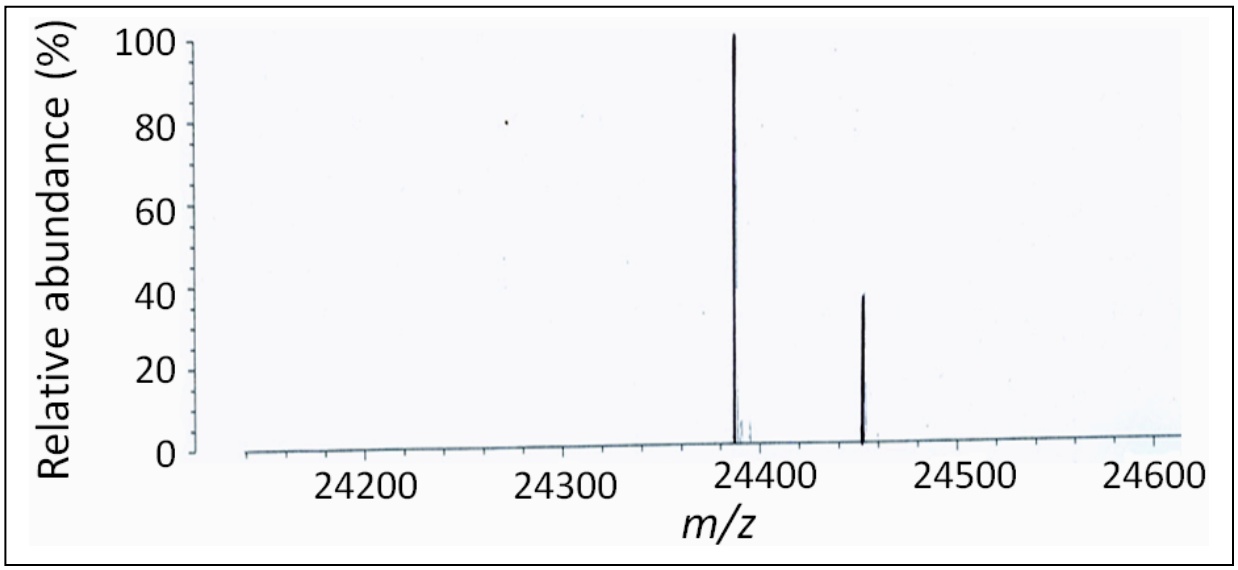


Figure 5.4: MALDI-TOF mass spectrum of 0.28 nmol of PPR27 in water.

A highly abundant peak of 24.39 kDa and a minor peak of 24.45 kDa are shown.

Using Edman degradation, we obtained G(W)P(H)VYAL as the first six amino acid residues at the N-terminus (chromatograms not shown, but are available). Barring the noise in positions 1 and 2, the sequence obtained was that expected of PPR27. Thus, the Factor Xa enzymatic reaction in sarkosyl was specific.

5.3.3 PPR27 is monomeric

The aggregation state and monodispersity of a protein are important in structural studies. During SEC, aggregated and heterogeneous proteins will elute in the void volume and as broadened amorphous peaks respectively. On the other hand, monomeric and monodisperse proteins elute at volumes corresponding to their size and as sharp peaks respectively. We thus investigated the hydrodynamic characteristics of PPR27 using SEC. PPR27 elutes after the 41.4 kDa TRXPPR27 but together with the 17 kDa TRXHis$_6$S (**Figure 5.5**). Assuming the oligomeric state of the proteins is not affected by low sarkosyl concentrations, tthe PPR27 size exclusion chromatogram is consistent with monomeric molecular mass. A similar observation was obtained earlier by native gel electrophoresis (chapter 2).

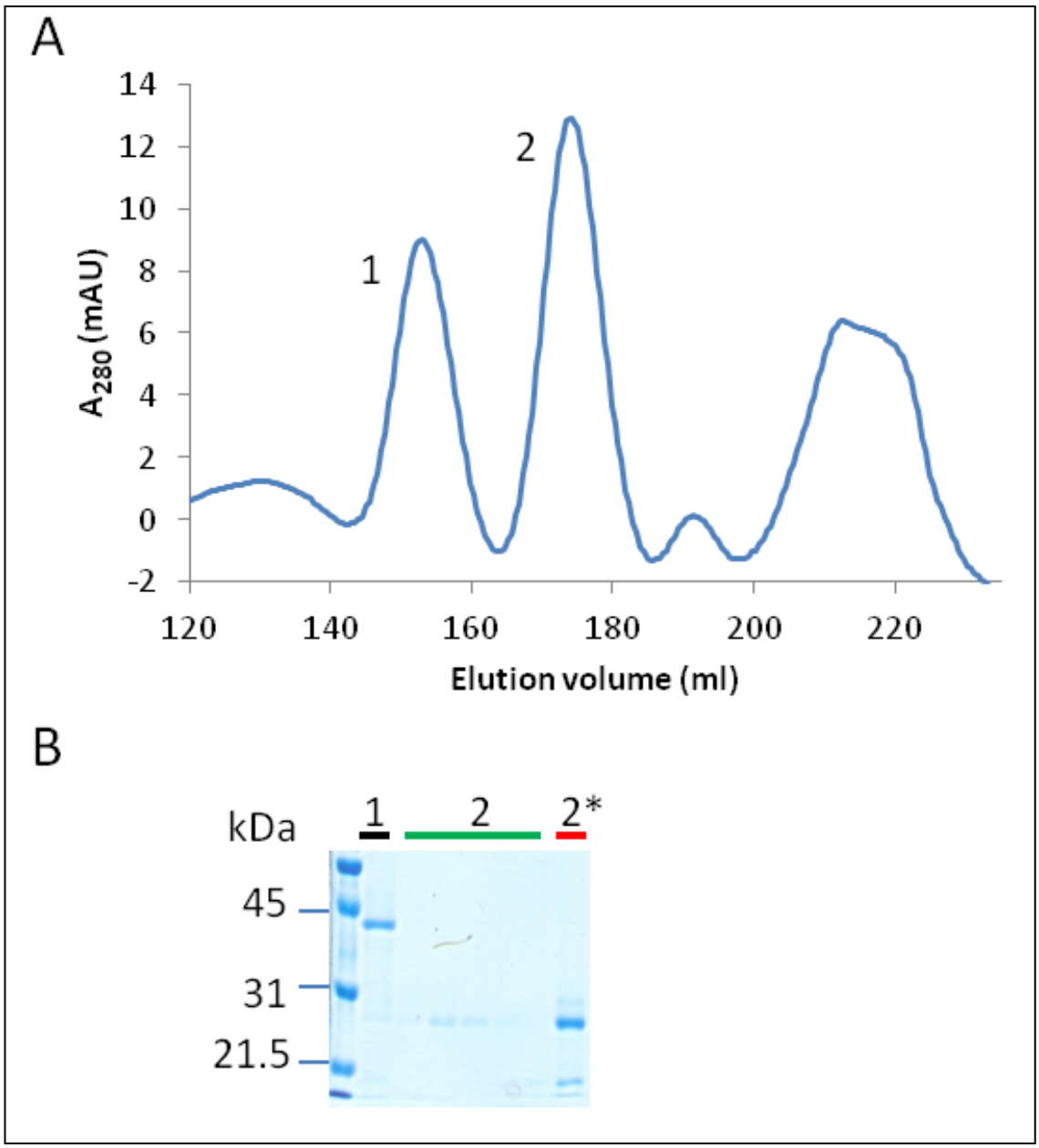


Figure 5.5: Size exclusion chromatography of mature wild type

(A) Gel filtration chromatogram of the Factor Xa cleavage mixture of TRXPPR27 showing peak 1 which is TRXPPR27 in gel (B) and peak 2 which contains free PPR27 and TRX according to gel (B). For (B), 1 is protein under peak1 after concentration; 2, dilute fractions under peak 2; 2* is the concentrated residue of fractions under peak 2.

5.3.4 Removal of detergent yields an alpha-helical protein

A protein's CD profile in the near U.V region is a good indicator of its structural integrity (*31, 39-40*). After removal of detergent by dialysis, the CD spectrum of PPR27 was characteristic of an α-helical protein (**Figure 5.6A**) with an helical content of between 55% and 63%, consistent with theoretical prediction and experimentally reported values of structurally similar plant PPR proteins (*23*).

Thermal melting experiments suggest PPR27 undergoes multi-state unfolding with early onset of unfolding (about 15 ^{o}C), a gradual transition region, a less distinct upper baseline by 90 ^{o}C, and an approximate transition temperature of at least 50 ^{o}C (**Figure 5.6B**). Because different modules in a multi-repeat protein may fold independently (*41*), multistate unfolding is possible. For example the mitochondrial tetratricopeptide repeat (TPR) protein Tom70 undergoes three state unfolding (*42*). Hence the unfolding profile of PPR27 is not unusual. But despite the apparent transition state being above 50 ^{o}C, the slope of the transition region is small which means reduced difference in free energy between native and denatured PPR27 and translates into less thermodynamic stability. So far the thermal melting experiment has been done on PPR27 dissolved in salt-free MQ-H_2O and it is also possible that instability from lack of pH control could have induced soluble aggregates and distorted the melting profile. The sample used for thermal melting had been stored at -20 ^{o}C for about two months before analysis, hence the effect of storage on the protein's structural integrity cannot be ruled out at this time too.

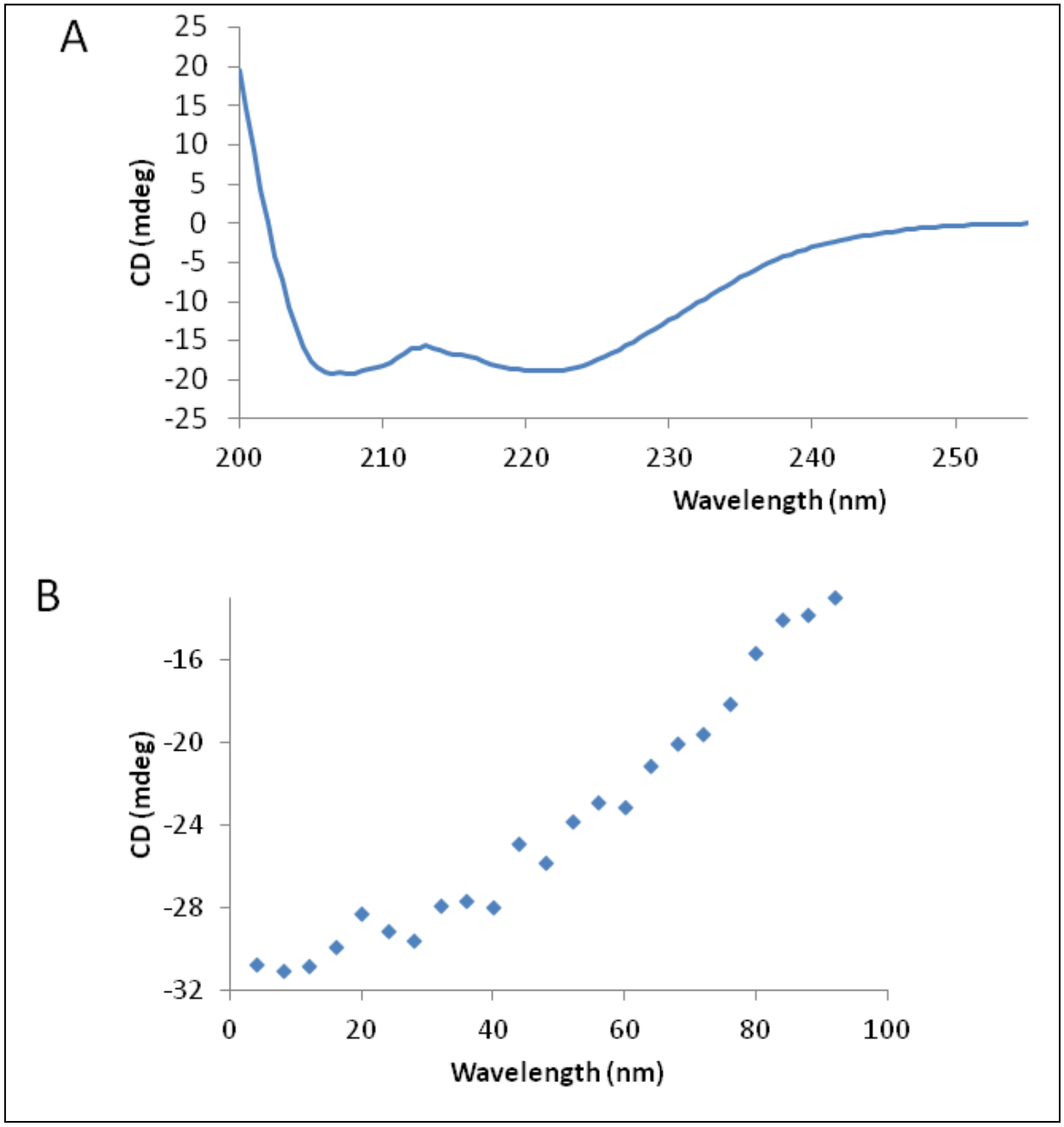


Figure 5.6: CD spectra of PPR27.

(A) CD spectrum of 0.22 mg/ml PPR27 with the two characteristic minima of α-helical proteins. (B) Variation of the CD signal at 222 nm with temperature showing a transition temperature of about 50 ^{o}C.

REFERENCES

REFERENCES

1. Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., and Savioli, L. (2007) Current concepts - Control of neglected tropical diseases, *New Engl J Med 357*, 1018-1027.

2. Mingler, M. K., Hingst, A. M., Clement, S. L., Yu, L. E., Reifur, L., and Koslowsky, D. J. (2006) Identification of pentatricopeptide repeat proteins in Trypanosoma brucei, *Mol Biochem Parasitol 150*, 37-45.

3. Pusnik, M., Small, I., Read, L. K., Fabbro, T., and Schneider, A. (2007) Pentatricopeptide repeat proteins in Trypanosoma brucei function in mitochondrial ribosomes, *Mol Cell Biol 27*, 6876-6888.

4. Small, I. D., and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins, *Trends Biochem Sci 25*, 46-47.

5. Aphasizheva, I., Maslov, D., Wang, X., Huang, L., and Aphasizhev, R. (2011) Pentatricopeptide repeat proteins stimulate mRNA adenylation/uridylation to activate mitochondrial translation in trypanosomes, *Mol Cell 42*, 106-117.

6. Saha, D., Prasad, A. M., and Srinivasan, R. (2007) Pentatricopeptide repeat proteins and their emerging roles in plants, *Plant Physiol Biochem 45*, 521-534.

7. Ringel, R., Sologub, M., Morozov, Y. I., Litonin, D., Cramer, P., and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase, *Nature 478*, 269-273.

8. Howard, M. J., Lim, W. H., Fierke, C. A., and Koutmos, M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing, *Proc Natl Acad Sci U S A 109*, 16149-16154.

9. Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis, *Plant Cell 16*, 2089-2103.

10. Filipovska, A., and Rackham, O. (2012) Modular recognition of nucleic acids by PUF, TALE and PPR proteins, *Molecular Biosystems 8*, 699-708.

11. Uyttewaal, M., Mireau, H., Rurek, M., Hammani, K., Arnal, N., Quadrado, M., and Giege, P. (2008) PPR336 is associated with polysomes in plant mitochondria, *Journal of Molecular Biology 375*, 626-636.

12. Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B., and Barkan, A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts, *Mol Cell Biol 28*, 5337-5347.

13. Hammani, K., Cook, W. B., and Barkan, A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression, *P Natl Acad Sci USA 109*, 5651-5656.

14. Khrouchtchova, A., Monde, R. A., and Barkan, A. (2012) A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts, *RNA 18*, 1197-1209.

15. Kroeger, T. S., Watkins, K. P., Friso, G., van Wijk, K. J., and Barkan, A. (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing, *Proc Natl Acad Sci U S A 106*, 4537-4542.

16. Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast psbB-psbT-psbH-petB-petD RNAs, *Plant Cell 15*, 1480-1495.

17. Okuda, K., Nakamura, T., Sugita, M., Shimizu, T., and Shikanai, T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing, *Journal of Biological Chemistry 281*, 37661-37667.

18. Derewenda, Z. S. (2004) The use of recombinant methods and molecular engineering in protein crystallization, *Methods 34*, 354-363.

19. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags, *Protein Sci 12*, 1313-1322.

20. Cura, V., Gangloff, M., Eiler, S., Moras, D., and Ruff, M. (2008) Cleaved thioredoxin fusion protein enables the crystallization of poorly soluble ERalpha in complex with synthetic ligands, *Acta Crystallogr Sect F Struct Biol Cryst Commun 64*, 54-57.

21. Foster, M. P., McElroy, C. A., and Amero, C. D. (2007) Solution NMR of large molecules and assemblies, *Biochemistry 46*, 331-340.

22. Wider, G. (2005) NMR techniques used with very large biological macromolecules in solution, *Methods Enzymol 394*, 382-398.

23. Williams-Carrier, R., Kroeger, T., and Barkan, A. (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand, *RNA 14*, 1930-1941.

24. Manavski, N., Guyon, V., Meurer, J., Wienand, U., and Brettschneider, R. (2012) An Essential Pentatricopeptide Repeat Protein Facilitates 5' Maturation and Translation Initiation of rps3 mRNA in Maize Mitochondria, *Plant Cell*.

25. Baneyx, F., and Mujacic, M. (2004) Recombinant protein folding and misfolding in Escherichia coli, *Nat Biotechnol 22*, 1399-1408.

26. Singh, S. M., and Panda, A. K. (2005) Solubilization and refolding of bacterial inclusion body proteins, *J Biosci Bioeng 99*, 303-310.

27. Burgess, R. R. (2009) Refolding solubilized inclusion body proteins, *Methods Enzymol 463*, 259-282.

28. Cabrita, L. D., and Bottomley, S. P. (2004) Protein expression and refolding--a practical guide to getting the most out of inclusion bodies, *Biotechnol Annu Rev 10*, 31-50.

29. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR), *Nucleic Acids Res 18*, 6069-6074.

30. Aslanidis, C., de Jong, P. J., and Schmitz, G. (1994) Minimal length requirement of the single-stranded tails for ligation-independent cloning (LIC) of PCR products, *PCR Methods Appl 4*, 172-177.

31. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res 32*, W668-673.

32. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A., and Lian, L. Y. (2004) A simple method for improving protein solubility and long-term stability, *J Am Chem Soc 126*, 8933-8939.

33. Montelione, G. T., Zheng, D., Huang, Y. J., Gunsalus, K. C., and Szyperski, T. (2000) Protein NMR spectroscopy in structural genomics, *Nat Struct Biol 7 Suppl*, 982-985.

34. Souza, D. H., Selistre-de-Araujo, H. S., and Garratt, R. C. (2000) Determination of the three-dimensional structure of toxins by protein crystallography, *Toxicon 38*, 1307-1353.

35. Goulding, C. W., and Perry, L. J. (2003) Protein production in Escherichia coli for structural studies by X-ray crystallography, *J Struct Biol 142*, 133-143.

36. DeLucas, L. J., Bray, T. L., Nagy, L., McCombs, D., Chernov, N., Hamrick, D., Cosenza, L., Belgovskiy, A., Stoops, B., and Chait, A. (2003) Efficient protein crystallization, *J Struct Biol 142*, 188-206.

37. Norris, J. L., Porter, N. A., and Caprioli, R. M. (2005) Combination detergent/MALDI matrix: functional cleavable detergents for mass spectrometry, *Anal Chem 77*, 5036-5040.

38. Yeung, Y. G., Nieves, E., Angeletti, R. H., and Stanley, E. R. (2008) Removal of detergents from protein digests for mass spectrometry analysis, *Anal Biochem 382*, 135-137.

39. Kelly, S. M., and Price, N. C. (1997) The application of circular dichroism to studies of protein folding and unfolding, *Biochim Biophys Acta 1338*, 161-185.

40. Sreerama, N., and Woody, R. W. (2004) Computation and analysis of protein circular dichroism spectra, *Methods Enzymol 383*, 318-351.

41. Itzhaki, L. S., and Lowe, A. R. (2012) From artificial antibodies to nanosprings: the biophysical properties of repeat proteins, *Adv Exp Med Biol 747*, 153-166.

42. Bushell, S. R., Bottomley, S. P., Rossjohn, J., and Beddoe, T. (2006) Tracking the unfolding pathway of a multirepeat protein via tryptophan scanning: evidence of localized instability in the mitochondrial import receptor Tom70, *J Biol Chem 281*, 24345-24350.

APPENDIX

PERSPECTIVES ON ADVANCED STRUCTURAL CHARACTERIZATION OF *T. BRUCEI* PPR27

This portion of the thesis provides a status report on exploratory nuclear magnetic resonance (NMR) spectroscopy and crystallization studies of various recombinant forms of PPR27.

I received assistance, supervision, and training on setting up crystallization screens and examining plates for crystals from Dr. Stacy Hovde (Dr. Geiger and Dr. Henry labs, Michigan State University). Besides the proteins which I designed and produced, all other crystallization materials were a kind gift from Dr. James Geiger. Finally, Mr. Kermit Johnson of the campus Max T. Rogers NMR Facility provided assistance and training in NMR data acquisition on the 900 MHz spectrometer.

Summary

Because of the solubility limitations of free PPR27, the feasibility of using chimeras of PPR27 with the TRX and MBP solubility enhancers in NMR spectroscopy was examined. We also did crystallization screening of the MBP fusion proteins of wild type PPR27 and its truncated variant lacking the last two C-terminal PPR motifs after shortening the linker between MBP and PPR27 to dialanine. Thirdly, the use of cognate RNA to facilitate PPR27 crystallization was attempted. Additionally, the free form of a truncated variant of PPR27 lacking the last three C-terminal PPR motifs was screened for crystallization.

NMR spectroscopy of TRXPPR27 and MBPPPR27

Because of the relatively small size of TRX (17 kDa), the TRXPPR27 fusion protein (41.4 kDa) isn't too big by modern NMR standards, hence is potentially amenable to less complex NMR investigations such as chemical shift perturbation assays and chemical shift assignments. On the other hand the presence of a 34 amino acid flexible linker between MBP and PPR27 left a possibility for PPR27 to hydrodynamically behave as an autonomus unit, and being smaller than MBP (44.0 kDa) had potential to yield most of the narrow peaks in the NMR spectrum. We thus assessed the chemical shift dispersion of $^{2}H,^{15}N$-TRXPPR27, ^{15}N-labeled free MBP, and ^{15}N-labeled MBP fusions of wild type PPR27 and a variant lacking two C-terminal PPR motifs (ΔCR2) using $^{1}H,^{15}N$-HSQC (heteronuclear single quantum coherence) and $^{1}H,^{15}N$-TROSY (transverse relaxation optimized spectroscopy) NMR experiments. A His_6 tag was present at the C-terminus of free MBP and that of PPR27 in MBPPPR27 to facilitate tandem Ni^{2+} and amylose affinity chromatography.

$^{2}H,^{15}N$-TRXPPR27 was overexpressed in Rosetta (DE3) *E. coli* while the ^{15}N-labeled MBPPPR27 fusion proteins and free MBP were overexpressed in BL21(DE3) *E. coli*. Briefly, bacteria containing recombinant plasmids were grown in 1.0 L of Luria Bertani (LB) medium at 37 ^{o}C, 250 rpm till an OD_{600} of 0.7-1.0 after which it was pelleted and pellet washed with 1X M9 minimal salts as described in chapter 3. For TRXPPR27, the washed pellet was resuspended in ^{2}H (93% (v/v D_2O) and ^{15}N dual labeled minimal medium while for free MBP and MBPPPR27 proteins, the pellets were resuspended in ^{15}N monolabeled minimal medium. Cultures were then incubated at 225 rpm, 37 ^{o}C for one hour followed by induction of protein expression with 0.1-0.2 mM IPTG at 30 ^{o}C for 4-6 hours. Cells were harvested as described in chapters 2-5 and stored at -80 ^{o}C.

$^{2}H,^{15}N$-TRXPPR27 was solubilized from inclusion bodies by 0.4% (w/v) sarksoyl, bound to a Talon CellThru resin under gravity and eluted by 200 mM imidazole in 0.1% (w/v) sarkosyl. Eluates were dialyzed against 75 volumes of low salt buffer (10 mM NaH_2PO_4, pH 7.5, 100 mM NaCl, 0.7 mM DTT) at 4 ^{o}C for at least 12 hours using 10K MWCO SnakeSkin dialysis tubing. Dialysates were concentrated to 0.18 mM using 10K MWCO amicon centrifugal filters. On the other hand, free ^{15}N-MBP and ^{15}N-MBPPPR27 fusion proteins were purified from bacteria under native conditions by tandem Ni-NTA and amylose affinity chromatography as described in chapter 2. Eluates were dialyzed against at least 50 volumes of the above low salt buffer (less DTT) for at least 12 hours as described above. Then DTT was added to each dialysate to 1.0 mM

followed by centrifugal concentration using amicon filters. For wild type PPR27, five liters of culture were required to produce 250 µl of 0.16 mM ^{15}N-MBPPPR27 while for ^{15}N-MBPPPR27ΔCR2 and free ^{15}N-MBP, 1.0 L of culture was sufficient to produce at least 250 µl of millimolar protein.

The suitability of the concentrated proteins for NMR investigations was examined by determining the chemical shift dispersion of ^{1}H/^{15}N cross peaks using 2D HSQC and TROSY NMR assays. For ^{2}H,^{15}N-TRXPPR27, the NMR sample contained 0.15 mM protein, 5.0% (w/v) D_2O, and 1.0 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) reference standard. For wild type ^{15}N-MBPPPR27, the NMR sample contained 0.14 mM protein, 10% (w/v) D_2O, and 0.2 mM DSS. For ^{15}N-MBPPPR27ΔCR2, the NMR sample contained 0.34 mM protein, 10% (w/v) D_2O, and 0.2 mM DSS. For ^{15}N-MBP, the NMR sample contained 0.7 mM protein, 10% (w/v) D_2O, and 0.2 mM DSS. Samples were loaded into a 5.0 mm shigemi reduced-volume NMR tube (Shigemi, Inc., Allison Park, PA) for NMR spectroscopy in a 900 MHz NMR spectrometer equipped with a cryoprobe.

Free ^{15}N-MBP (negative control) yielded well dispersed peaks (data not shown) as expected to due to high solubility and stability on concentration. For PPR27, only ^{2}H,^{15}N-TRXPPR27 yielded well dispersed 2D NMR spectra. There was too much peak overlap in the NMR spectra of ^{15}N-MBPPPR27 and ^{15}N-MBPPPR27ΔCR2 (data not shown), rendering them unsuitable for any NMR investigations. On the other hand, both the ^{1}H,^{15}N-HSQC- and ^{1}H,^{15}N-TROSY NMR

spectra of TRXPPR27 show a wide chemical shift dispersion characteristic of native proteins (**Figure 6.1**), though the number of peaks is smaller than expected. TRX is a very soluble protein whose NMR spectra in both reduced and oxidized states have wide chemical shift dispersions (*1*). Due to the poor solubility of PPR27, there is a possibility that the PPR portion of one TRXPPR27 molecule may stick to that of other TRXPPR27 molecules resulting in slowly tumbling PPR27 aggregates but freely tumbling TRX ends. If this happened, the observed NMR spectra of TRXPPR27 could be dominated by TRX cross-peaks at the expense of PPR27 as the latter's spectra would suffer from severe line broadening. Furthermore, PPR27 is a multi-repeat α-helical protein; hence, loss of cross-peaks in TRXPPR27 could also arise from the inherent problem of signal overlap in alpha-helical proteins. As such, there is potential for improvement of these spectra with further optimization of protein solubility. For example, supplementation with 50 mM Arg/50mM Glu could raise the sample concentration above 0. 15 mM and improve the signal. If the reduced number of peaks were due to some aggregation, Glu/Arg could still improve sample quality, and raise the number of discernible peaks.

Coarse crystallization screens of PPR27

MBP fusions of a few poorly soluble proteins have been crystallized after reducing the flexible linker between MBP and the passenger to a 2-6 amino acid peptide, typically rich in alanine (*2-3*). Additionally, crystallization of some proteins has only been possible after complexation with cognate ligands (*4*). Here, I did crystallization screening for the MBP fusion proteins of wild type PPR27 and its truncated variant lacking the last two C-terminal PPR motifs (ΔCR2) after shortening and mutating the linker between MBP and PPR27 to dialanine. Truncation of the linker from 34 amino acids to two (Asp (N) and Ser (S)) and subsequent mutation to AA were done by QuikChange site-directed mutagenesis as described previously (chapter 2) using the

primers in **Table 6.1**. These shortened-linker fusion proteins were then overexpressed in BL21 (DE3) *E. coli* by induction with 0.1 mM IPTG, followed by tandem affinity purification with Ni^{2+} followed by amylose (**Figure 6.2**). Pure protein was dialyzed against at least 50-fold volumes of buffer DCB (20 mM Tris-HCl, pH 7.5, 50 mM NaCl) at 4 ^{o}C for at least 12 hours, using 7K MWCO SnakeSkin tubing. Dilute dialysate was concentrated by 10K MWCO amicon centrifugal filters and concentration derived from U.V A_{280} as in chapter 2.Then coarse crystallization screens were set up for the proteins in the presence and absence of of G_6 ssRNA ligands. The apo form of MBP-AA-PPR27 (wild type) yielded microcrystals in at least one well. Well B8 [10% (w/v) PEG 6000, 0.1 M HEPES/sodium hydroxide, pH 7.0] from Wizard III/IV screen (Emerald Bio) appears to have the most promising conditions for optimization.

The untagged form of a truncated variant of PPR27 lacking the last three C-terminal PPR motifs (ΔCR3) was also screened for crystallization. The protein was purified by DEAE sephadex anion exchange chromatography and concentrated in the presence of 50 mM L-Arg anf 50 mM L-Glu as described in chapter 3. Microscrystals were observed in at least one well. Well F8 [15% (v/v) ethanol, 0.1 M MES/sodium hydroxide, pH 6.0, 0.2 M zinc acetate] from Wizard I/II screen (Emerald Bio) appears to have the most promising conditions for optimization.

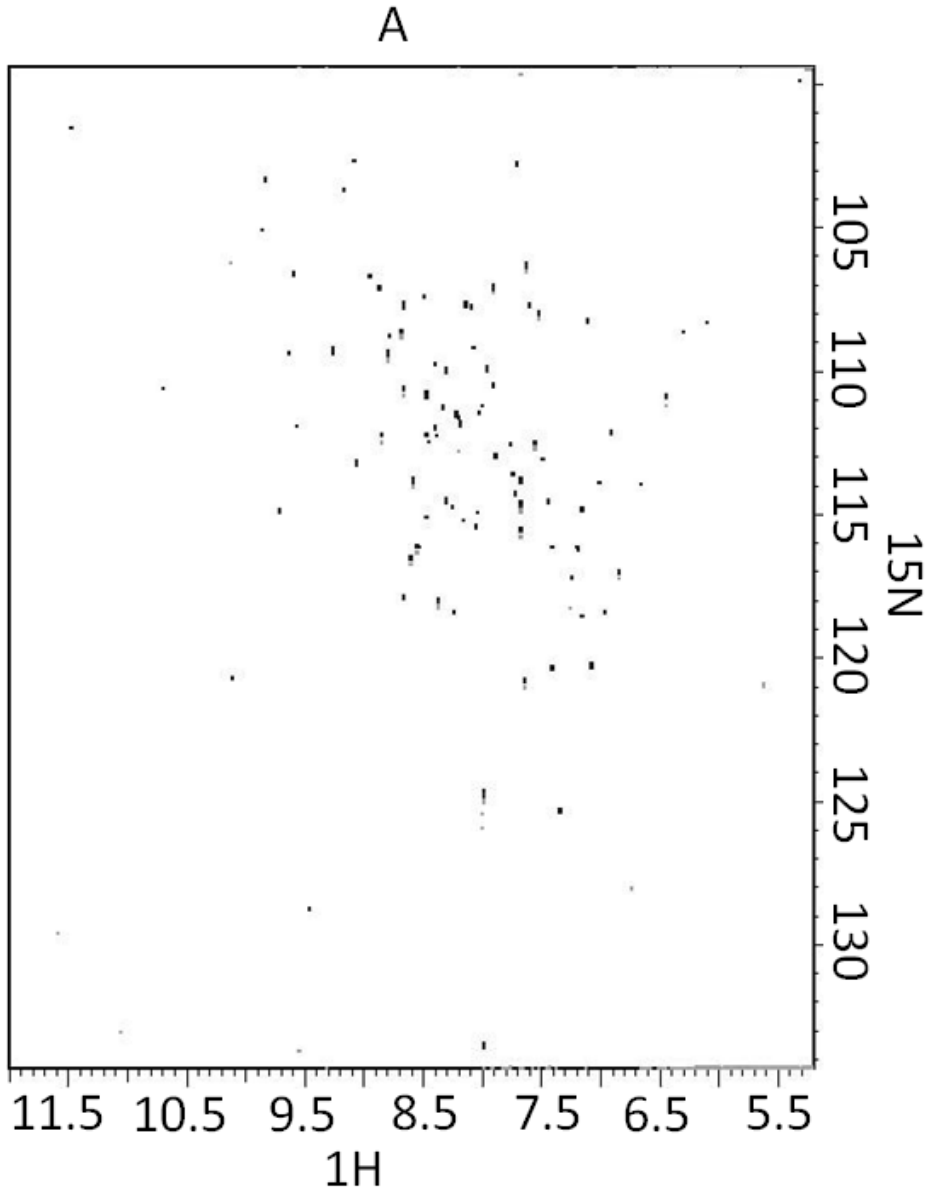
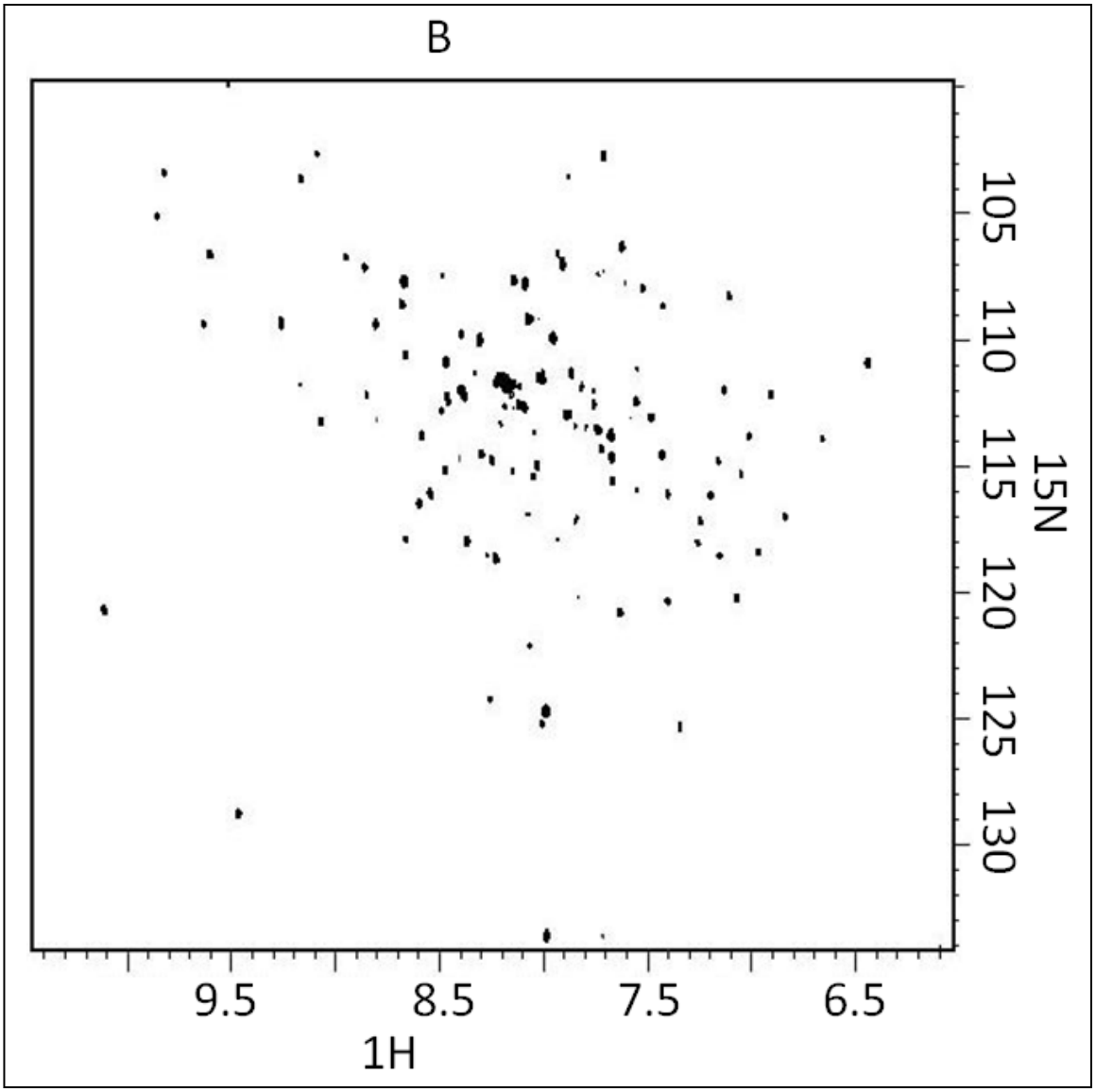


Figure 6.1: 2D NMR spectra of ^{2}H,^{15}N-double labeled TRXPPR27 obtained from 900 MHz spectrometer.

(A) ^{1}H,^{15}N-HSQC spectrum obtained using 32 transients at 25 ^{o}C, each with 1024 and 256 complex points in the 1st and 2nd dimensions respectively. The spectral width for ^{1}H was 14,367.8161 Hz while that for ^{15}N was 2735.5098 Hz.

Figure 6.1 (cont'd).



(B) $^{1}H,^{15}N$-TROSY spectrum obtained using 200 transients at 25 ^{o}C, each with 1024 and 256 complex points in the 1^{st} and 2^{nd} dimensions respectively. The spectral width for ^{1}H was 15,290.5199 Hz while that for ^{15}N was 3191.5130 Hz. NMR data was processed with a line broadening of 5 Hz using Felix 2002 software (Felix NMR Inc., San Diego, CA).

Table 6.1: Primers used in linker truncation and mutagenesis to dialanine (AA)

Activity	Primer-code	Primer sequence (5' to 3')
Truncate the 34 amino acid linker to two amino acids (NS)	CGH85	GCGCAGACTAATTCGGGTCACGTGTACGCC
	CGH86	GGCGTACACGTGACCCGAATTAGTCTGCGC
Mutate NS to AA	CGH87	GCCCTGAAAGACGCGCAGACTGCTGCGGGTCACGTG
	CGH88	CACGTGACCCGCAGCAGTCTGCGCGTCTTTCAGGGC

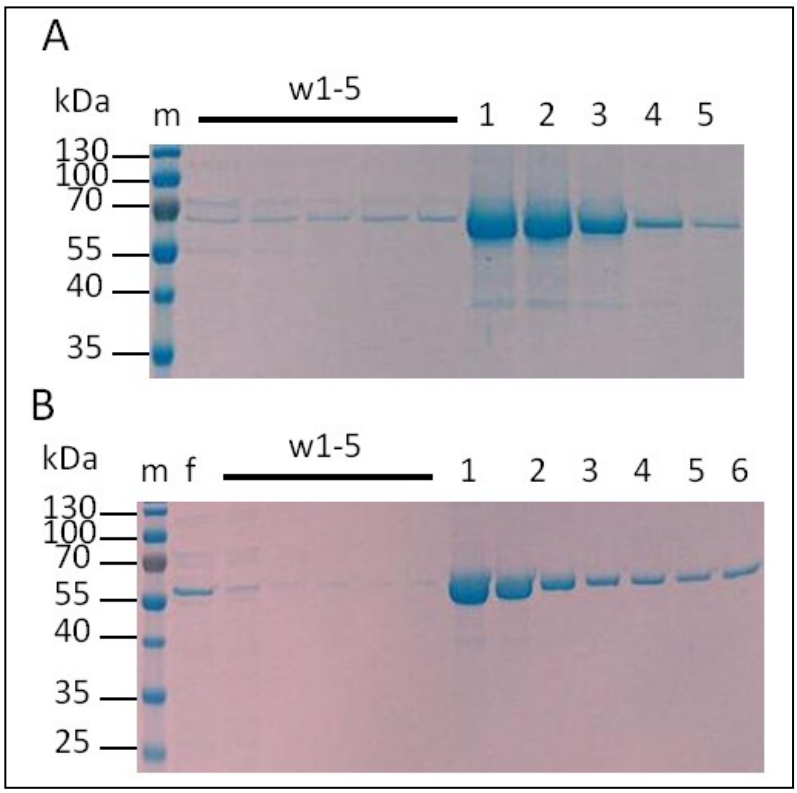


Figure 6.2: Coomassie blue stained 5-12% polyacrylamide gels showing MBPPPR27 with a dialanine linker after tandem Ni^{2+} and amylose affinity chromatography.

In both gels, m is the protein size marker, w1-5 are consecutive amylose column washes, and 1-5/6 are the eluates containing pure MBPPPR27. In (B), f is the amylose column flow through.

REFERENCES

REFERENCES

1. Chandrasekhar, K., Krause, G., Holmgren, A., and Dyson, H. J. (1991) Assignment of the 15N NMR spectra of reduced and oxidized Escherichia coli thioredoxin, *FEBS Lett 284*, 178-183.

2. Derewenda, Z. S. (2004) The use of recombinant methods and molecular engineering in protein crystallization, *Methods 34*, 354-363.

3. Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags, *Protein Sci 12*, 1313-1322.

4. Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J., Jr., Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H., Hui, R., Marsden, B. D., Nordlund, P., Sundstrom, M., Weigelt, J., and Edwards, A. M. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination, *Proc Natl Acad Sci U S A 103*, 15835-15840.