

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Dissecting the RNA-Binding Specificity and Role in Regulating mRNA Translation of the Fragile X Mental Retardation Protein

Permalink

<https://escholarship.org/uc/item/3kh837f4>

Author

Athar, Youssi Momen

Publication Date

2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Dissecting the RNA-Binding Specificity and Role in Regulating mRNA Translation of
the Fragile X Mental Retardation Protein

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Youssi Momen Athar

Committee in charge:

Professor Simpson Joseph, Chair
Professor Gourisankar Ghosh
Professor Judy Kim
Professor Ulrich Müller
Professor Nicholas Spitzer

2020

Copyright

Youssi Momen Athar, 2020

All rights reserved.

The dissertation of Youssi Momen Athar is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

In recognition of all my loved ones — those who supported me and made these past few years fun and memorable.

EPIGRAPH

... Dr. Rieux resolved to compile this chronicle, so that he should not be one of those who hold their peace but should bear witness in favor of those plague-stricken people; so that some memorial of the injustice and outrage done them might endure; and to state quite simply what we learn in time of pestilence: that there are more things to admire in men than to despise.

Nonetheless, he knew that the tale he had to tell could not be one of a final victory. It could be only the record of what had had to be done, and what assuredly would have to be done again in the never ending fight... by all who, while unable to be saints but refusing to bow down to pestilences, strive their utmost to be healers.

Albert Camus

TABLE OF CONTENTS

Signature Page.....	iii
Dedication.....	iv
Epigraph.....	v
Table of Contents.....	vi
List of Abbreviations.....	ix
List of Figures.....	xi
List of Tables.....	xii
Acknowledgements.....	xiii
Vita.....	xiv
Abstract of the Dissertation.....	xv
Chapter 1: Introduction.....	1
1.1 Fragile X Syndrome is caused by the FMR1 gene.....	2
1.2 Architecture of FMRP.....	4
1.3 FMRP binds specific sequence or structure motifs in target mRNAs.....	10
1.4 FMRP represses translation of target mRNAs.....	15
1.5 FMRP and the mGluR-LTD pathway.....	17
1.6 Translating FMRP and FXS studies into animal models.....	19
References.....	21
Chapter 2: RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein.....	29
Abstract.....	30
Introduction.....	31

Results and Discussion.....	34
2.1 Isolating the RNA-binding domains of FMRP.....	34
2.2 Design of model RNAs.....	37
2.3 Folding of model RNAs.....	40
2.4 FMRP binds to different GQ RNAs.....	45
2.5 FMRP RGG motif specifically binds to RNA GQ structures.....	46
2.6 The KH domains of FMRP do not bind to the WGGGA and GACR motifs.	51
2.7 FMRP does not bind with high affinity to m ⁶ A RNA.....	53
2.8 The KH domains of FMRP do not bind to short ssRNA motifs.....	55
2.9 Discussion.....	62
Materials and Methods.....	65
2.10 Protein expression and purification.....	65
2.11 Purification of RNAs.....	67
2.12 PAGE analysis of labeled RNAs.....	67
2.13 BG4 antibody assay.....	68
2.14 NMM assay.....	69
2.15 Fluorescence anisotropy RNA binding assay.....	69
2.16 MIDAS DNA template and primers.....	71
2.17 MIDAS: Separation of RNA•protein complex using a native gel.....	72
2.18 MIDAS: Library preparation.....	73
2.19 MIDAS: NGS data analysis.....	74
References.....	76

Chapter 3: The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through its RGG and C-terminal domains.....	84
Abstract.....	85
Introduction.....	86
Materials and Methods.....	89
3.1 Protein expression and purification.....	89
3.2 Preparation of reporter mRNAs.....	91
3.3 In vitro translation assays.....	92
3.4 Human 80S ribosome binding assay.....	93
Results and Discussion.....	94
3.5 FMRP inhibits the translation of various mRNAs.....	95
3.6 FMRP inhibits the elongation step of translation.....	98
3.7 The RGG domain and CTD are sufficient to inhibit translation.....	100
3.8 FMRP binding to the ribosome correlates with inhibition of translation...	103
3.9 Discussion.....	106
References.....	109
Chapter 4: Conclusion and Future Directions.....	115
References.....	120

LIST OF ABBREVIATIONS

CDS	Coding sequence
CrPV	Cricket paralysis virus
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
EMCV	Encephalomyocarditis virus
EMSA	Electrophoretic mobility shift assay
FMRP	Fragile X mental retardation protein
FXS	Fragile X syndrome
GQ	G-quadruplex
GST	Glutathione S-transferase
HCV	Hepatitis C virus
IRES	Internal ribosome entry site
IVT	In vitro translation
KH	K-homology
m ⁷ G	7-Methylguanosine
MIDAS	Motif identification by analysis of simple sequences
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoprotein
NanoLuc	NanoLuc® luciferase
NMM	N-methyl mesoporphyrin IX
Nova-2	Neuro-oncological ventral antigen 2
NTD	N-terminal domain

PAGE	Polyacrylamide gel electrophoresis
PV	Poliovirus
RGG	Arginine-glycine-glycine
RLuc	<i>Renilla</i> luciferase
RNA	Ribonucleic acid
RRL	Rabbit reticulocyte lysate
SGG	Serine-glycine-glycine
ssRNA	Single-stranded ribonucleic acid
tRNA	Transfer ribonucleic acid
UTR	Untranslated region

LIST OF FIGURES

Figure 1.1: Architecture of human FMRP	5
Figure 1.2: 3.0 Å crystal structure of human FMRPΔ depicting the Agenet 2 dimer interface.....	9
Figure 1.3: 2.8 Å crystal structure of cesium-bound human FMRP RGG motif bound to <i>sc1</i> RNA.....	13
Figure 1.4: 0.61 Å crystal structure of (UG ₄ U) ₄	14
Figure 1.5: Models for translational repression by FMRP binding the target mRNA G-quadruplex.....	18
Figure 2.1: Design and purification of human FMRP constructs.....	36
Figure 2.2: Analysis of the structures formed by the model RNAs.....	43
Figure 2.3: (GACG) ₄ may form hairpin or duplex structures.....	44
Figure 2.4: FMRP binding to model RNAs.....	47
Figure 2.5: FMRP RGG domain binding to GQ RNAs is dependent on K ⁺	50
Figure 2.6: KH domains binding to model RNAs (UGGA) ₄ and (GACG) ₄	52
Figure 2.7: NT-hFMRP binding to UGGm ⁶ ACU RNA.....	54
Figure 2.8: Illustration of MIDAS method.....	56
Figure 2.9: KH domains binding to RNAs identified by MIDAS.....	60
Figure 2.10: KH domains binding to MIDAS RNA UCACC.....	61
Figure 3.1: Human FMRP constructs.....	96
Figure 3.2: Human FMRP inhibits translation of different mRNAs in RRL.....	97
Figure 3.3: FMRP inhibits translation of canonical and IRES-driven <i>Renilla</i> luciferase mRNAs.....	99
Figure 3.4: The RGG domain and CTD of FMRP are required for inhibition of translation.....	102
Figure 3.5: NT-hFMRP and GST-hFMRP RGG bind directly to the human 80S ribosome.....	105

LIST OF TABLES

Table 2.1: Nomenclature and sequences of the RNAs.....	39
Table 2.2: Equilibrium dissociation constant, K_D (nM) of FMRP binding to RNAs....	48
Table 2.3: MIDAS enrichment factor (EF) and binding affinity for RNAs.....	59

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Professor Simpson Joseph for his ardent support as my mentor and chair of my committee. Simpson always made it feel like I was working with him, not for him, to develop my career as a scientist and expand our understanding of fragile X syndrome.

I would also like to thank my dissertation committee members for their support over the years, especially professors Gouri Ghosh and Uli Müller for their keen interest in my research and constructive feedback on my manuscripts.

The Joseph lab has been a great place to learn and practice research. I thank my colleagues for their friendship and for making each day enjoyable. I will always remember the hardships we faced together and cherish the fun times we shared.

Finally, I am eternally grateful to my parents and my family for encouraging me on my academic journey. I would not have come this far without their support.

Chapter 2, in full, is a reprint of the material as it appears in RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein, *Journal of Molecular Biology*, 2020. Athar, Youssi M.; Joseph, Simpson, Elsevier, 2020. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through its RGG and C-terminal domains, *Biochemistry*, 2020. Athar, Youssi M.; Joseph, Simpson, American Chemical Society, 2020. The dissertation author was the primary investigator and author of this material.

VITA

- 2011 Bachelor of Arts, Molecular and Cell Biology
University of California, Berkeley
- 2011-2014 Research Assistant
University of California, Los Angeles
- 2016 Master of Science, Chemistry
University of California San Diego
- 2020 Doctor of Philosophy, Chemistry
University of California San Diego

PUBLICATIONS

"RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein"
Journal of Molecular Biology, vol. 432, pp 3851-3868, June 2020.

"The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through its RGG and C-terminal domains" bioRxiv, June 2020.

ABSTRACT OF THE DISSERTATION

Dissecting the RNA-Binding Specificity and Role in Regulating mRNA Translation of the Fragile X Mental Retardation Protein

by

Youssi Momen Athar

Doctor of Philosophy in Chemistry

University of California San Diego, 2020

Professor Simpson Joseph, Chair

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability and is caused by a deficiency of the fragile X mental retardation protein (FMRP) in neurons. FMRP regulates the translation of numerous mRNAs within dendritic synapses, but how FMRP recognizes these target mRNAs remains unknown.

FMRP has KH0, KH1, KH2, and RGG domains, which are thought to bind to specific RNA recognition elements (RREs). Several studies used high-throughput

methods to identify various RREs in mRNAs that FMRP may bind to in vivo. However, there is little overlap in the mRNA targets identified by each study. To determine the specificity of FMRP for the RREs, we performed quantitative in vitro RNA binding studies with various constructs of human FMRP. Our studies show that the KH domains do not bind to the previously identified RREs. To further investigate the RNA-binding specificity of FMRP, we developed a new method called Motif Identification by Analysis of Simple sequences (MIDAS) to identify single-stranded RNA (ssRNA) sequences bound by KH domains. We find that the FMRP KH0, KH1, and KH2 domains bind weakly to the ssRNA sequences suggesting that they may have evolved to bind more complex RNA structures. Additionally, we find that the RGG motif of human FMRP binds with a high affinity to an RNA G-quadruplex (GQ) structure that lacks single-stranded loops, double-stranded stems, or junctions.

FMRP has been proposed to inhibit translation of target mRNAs in neurons. We investigated how human FMRP (hFMRP) regulates the translation of different mRNAs in rabbit reticulocyte lysate. We find that hFMRP inhibits different mRNAs to similar degrees. Furthermore, we observe inhibition is independent of the 5' cap-dependent initiation using . Finally, we dissected the RNA-binding domains of hFMRP and measured their inhibition of translation. We determined the RGG domain and C-terminal domain (CTD) are sufficient to inhibit translation while the KH domains do not inhibit mRNA translation. Interestingly, we see a correlation between ribosome binding and translation inhibition, suggesting the RGG-CTD tail of hFMRP may anchor FMRP to the ribosome during translation inhibition.

Chapter 1:
Introduction

1.1 Fragile X Syndrome is caused by the FMR1 gene

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability (ID). FXS is an inherited X-linked condition and according to the National Fragile X Foundation, FXS affects approximately 1 in 4000 males and 1 in 6000 females worldwide¹. Over 100,000 individuals in the U.S. suffer FXS with costs of treatment exceeding \$200 million per year².

While ID is a common symptom amongst FXS patients, severity of the symptom can vary. Some patients display symptoms observed in patients with autism spectrum disorder (ASD). ASD symptoms may include impaired cognitive function and unusual behavior, but more specifically difficulties with speech, social anxiety, seizures, and attention-deficit/hyperactivity disorder (ADHD)^{3,4}. Patients with FXS may display physical abnormalities as well, such as elongated facial features and ears, flat feet, and macroorchidism⁵.

Postmortem examinations of FXS patients' brains revealed the physiological abnormality behind the disease. The FXS patients' neurons contained abnormally dense and immature dendritic spines⁶. Genetic studies revealed FXS patients had a long expansion of cytosine-guanine-guanine (CGG) trinucleotide repeats at the 5' untranslated region (UTR) of the same gene within the X chromosome, then identified as the Fragile X mental retardation gene (*FMR1*)^{7,8}. The CGG expansion causes FXS by two major mechanisms involving silencing of the *FMR1* gene: (1) hypermethylation of the DNA and (2) hybridization of the encoded mRNA to the *FMR1* gene's complementary CGG array to repress transcription of the gene⁹. While

it is unclear how these two mechanisms control transcription of the *FMR1* gene, it is evident they effectively diminish the expression of the encoded Fragile X mental retardation protein (FMRP). Overall, the studies suggest FXS is caused by FMRP deficiency in the neuron.

Later an isoleucine 304 to asparagine point mutation (I304N) within the K-Homology 2 (KH2) domain was identified in patients suffering a severe form of FXS¹⁰. Investigation of the I304N variant revealed a decrease in FMRP's stability, affinity for target RNAs and association with polyribosomes^{10,11}.

Interestingly, the length of the CGG repeat expansion can cause a different disease in those with shorter repeat expansions. While normal individuals have 6-54 CGG repeats and individuals with FXS have over 200 CGG repeats, individuals with 55-200 CGG repeats can develop pre-mutation diseases such as Fragile X-associated tremor/ataxia syndrome (FXTAS) and Fragile X-related primary ovarian insufficiency (FXPOI)⁵. Unlike FXS symptoms that manifest as ID in childhood, FXTAS symptoms manifest in adulthood and result in neurodegenerative disorders that cause involuntary movements, imbalance, and symptoms generally associated with Parkinson's and Alzheimer's disease¹². FXPOI causes early-onset menopause in women before the age of 40 as well as dysfunctional ovaries which results in infertility¹³. Studies of FXTAS suggest the pre-mutation number of CGG repeats results in elevated levels of the *FMR1* messenger RNA (mRNA) levels but decreased FMRP levels^{14,15}.

1.2 Architecture of FMRP

FMRP contains several highly conserved domains (**Figure 1.1**). Amongst these, four are RNA-binding domains. There are three K-Homology (KH) domains — KH0, KH1 and KH2 — and an arginine-glycine-glycine (RGG) box domain, all four of which are hypothesized to mediate FMRP binding to target mRNAs¹⁶.

The KH domains were first identified as nucleic acid binding domains in heterogeneous nuclear ribonucleoprotein K (hnRNP K) and have since been found in many eukaryotic (Type I) and prokaryotic (Type II) proteins¹⁷. A 1.9 Å resolution crystal structure of the human FMRP KH domains confirms both KH1 and KH2 domains' topologies match the typical Type I found in several eukaryotic nucleic acid binding proteins¹¹. A GXXG loop located in FMRP KH domains are also found in several gene regulatory proteins, such as the KH-type splicing regulatory protein (KSRP), further suggesting FMRP KH domains play a role in specific RNA binding^{18,19}.



Figure 1.1: Architecture of human FMRP. The relative position of each domain is depicted throughout the primary sequence above. The Agenet 1 (cyan) and Agenet 2 (orange) domains are also known as the Tudor 1 and Tudor 2 domains, respectively.

Indeed, patients suffering a severe form of FXS often had a I304N mutation in the KH2 domain. Even with a normal span of CGG repeats at the *FMR1* 5' UTR, the I304N point mutation was sufficient to cause FXS¹¹. Subsequent studies of the I304N FMRP variant revealed the point mutation abolished FMRP's ability to bind target mRNAs and associate with polysomes^{20,21}. Inspection of the KH2 domain's structure shows isoleucine 304 resides within a hydrophobic region of the KH2 domain. Substitution of the hydrophobic isoleucine for a polar asparagine residue could therefore destabilize the protein's structure¹¹. It is still unknown through which mechanism(s) the mutation causes FXS.

Another FXS patient was later found to have a glycine 266 to glutamate mutation within the KH1 domain²¹. The G266E KH1 mutant FMRP was also unable to associate with polysomes and bind target mRNAs. While only two point mutations within the KH domains have been identified, it is possible that other KH1 and KH2 mutations that similarly disrupt FMRP stability or function may cause FXS as well.

In addition to the three KH domains, FMRP also contains an RGG domain composed of two tandem RGG motifs²². An NMR structure of the human FMRP RGG motif bound to the in vitro-selected *sc1* RNA G-quadruplex (GQ) reveals how the disordered RGG motif is stabilized upon binding to the GQ structure²³. A subsequent crystal structure reveals in atomic detail how each RGG motif amino acid interacts with the *sc1* GQ nucleotides²⁴. While the *sc1* RNA was selected as a stable RGG motif binding partner, it is possible that the RGG motif mediates FMRP binding to GQ structures in target mRNAs. And considering the ubiquity of GQ structures

amongst mRNAs, the presence of the RGG motif within FMRP hints at FMRP's potential to bind a vast pool of target mRNAs in vivo.

Besides FMRP's RNA binding domains, FMRP contains a multifunctional domain spanning its N-terminal region known as the N-terminal domain of FMRP (NDF). The NDF is defined as the beginning of FMRP up to the boundary of the KH1 domain, and studies of the NDF suggests it mediates assembly of both FMRP-protein interaction as well as stable FMRP homodimers^{25,26}. The NMR structure of the NDF revealed the presence of two tandem Agenet domains that have been observed to bind trimethylated lysines on histones via hydrophobic patches^{26,27}. There is in fact evidence of FMRP involvement in gametogenesis and the DNA damage response (DDR) via binding of its tandem Tudor domains to chromatin²⁸.

A recent 3.0 Å X-ray crystal structure of a stable region of the NDF spanning residues 1 to 209 of human FMRP (FMRP Δ) revealed a novel KH0 domain upstream of the KH1 and KH2 domains²⁹. The KH0 domain adopts a Type I topology similar to KH1 and KH2 domains, but shows three key differences from other KH domains: (1) the conserved GXXG present in KH domains is replaced with a single lysine residue, (2) the KH0 is missing the conserved hydrophobic residues of the IGXXGXXI motif, and (3) there is no positively charged groove for RNA binding. These three differences suggest the KH0 domain interacts with putative binding partners in a different manner from other KH domains.

In the same study, FMRP Δ was found to exist as stable homodimers. This is in agreement with previous studies showing the NDF mediates FMRP dimerization. However, this study revealed three possible dimer interfaces and attempted to

identify the true dimer interface. Small-angle X-ray scattering (SAXS) analysis of FMRP Δ hinted at an ensemble of dimers. Subsequent minimal ensemble search (MES) to fit the SAXS data suggested all three dimer interfaces are somehow involved in FMRP oligomerization, although the authors propose the Agenet 2 domain contains the primary dimer interface (**Figure 1.2**). It remains unclear which of the proposed interfaces is primarily stabilizing FMRP Δ dimerization.



Figure 1.2: 3.0 Å crystal structure of human FMRP Δ depicting the Agenes 2 dimer interface. The dimerization interface between the the Agenes 2 domains (magenta and orange) of two monomers is proposed to be the primary dimer interface despite the small area of the dimer interface relative to the size of human FMRP. The Agenes 1 and KH0 domains are highlighted in cyan and blue, respectively. Protein Data Bank entry code 4OVA.

FMRP also contains both a nuclear localization signal (NLS) and a nuclear export signal (NES)³⁰. In light of FMRP's RNA-binding domains, FMRP may also shuttle target mRNAs from the nucleus to the cytosol. Interestingly, an arginine 138 mutation to glutamine (R138Q) was identified in a patient suffering developmental delays³¹. More recently, FMRP was found to function in gametogenesis and the DDR within the cell nucleus. Indeed, the R138Q mutation disrupted FMRP's binding to nucleosomes, effectively impairing its role in DDR²⁸.

1.3 FMRP binds specific sequence or structure motifs in target mRNAs

FXS has been traced to the absence of the Fragile X mental retardation protein (FMRP) in neurons. Loss of FMRP in turn has been linked to the deregulation of translation of specific mRNAs by the ribosome^{32,33}. This suggests FMRP is an RNA-binding protein that must be able to recognize its target mRNA and coordinate with the ribosome to regulate their translation. While FMRP has been shown to bind the ribosome directly, it is still unclear how the FMRP, the target mRNA and the ribosome coordinate spatiotemporally to suppress translation of the mRNA^{34,35}. Most endeavors to identify FMRP mRNA targets have employed high-throughput methods such as HITS-CLIP, PAR-CLIP, RNAcompete and TRIBE. Unfortunately, the proposed catalog of target RNAs, as well as RNA sequence and structure motifs, overlap poorly between each study^{34,36–38}. Analysis of the various proposed RNAs and RNA recognition elements yielded two relatively promising RNA recognition elements: the GACR (R is a purine, A or G) and the WGGA (W is an A or

U) sequences³⁹. The GACR sequences were enriched for an FMRP variant composed solely of the KH1 and KH2 RNA binding domains³⁷. The WGGA sequences were proposed to bind specifically to the KH1 domain of FMRP³⁶.

A more recent study has proposed FMRP primarily binds and represses the mRNA encoding diacylglycerol kinase subunit kappa (DGK κ), and that loss of this single interaction is sufficient for causing FXS-like symptoms in mice⁴⁰. The proposition that FMRP's role in FXS pathophysiology is dictated primarily through loss of binding solely to the DGK κ mRNA boldly refutes the field's working hypothesis: FMRP binds a target set of mRNAs by recognizing either a sequence motif or a structural motif, and represses their translation by the ribosome locally.

Despite the controversy, one RNA motif that FMRP has been confirmed to bind experimentally is the RNA G-quadruplex (GQ)^{39,41}. In fact, the SELEX-derived GQ-assembling RNA *sc1* was shown to bind FMRP's arginine-glycine-glycine (RGG) box motif with high affinity (**Figure 1.3**)^{23,24}. However, being selected under optimized conditions in vitro, the *sc1* RNA GQ may not be representative of any RNA GQ structures that FMRP may bind in vivo⁴². It is also unclear if the FMRP RGG motif is sufficient for specifically binding GQ structures in target mRNAs. Altogether it is unknown how FMRP specifically binds the GQ and if the RNA GQ is a general method by which FMRP recognizes target mRNAs.

To answer these questions, we sought to quantify binding affinity between FMRP and various model RNAs. A PolyG₁₇ RNA was originally used to serve as a model RNA GQ. Later we employed a six nucleotide RNA molecule UG₄U that has been shown to assemble a stable intermolecular (UG₄U)₄ GQ structure in potassium,

but remain primarily as a monomer in lithium, as another model RNA GQ (**Figure 1.4**)^{43,44}. Being able to switch a RNA between a GQ and a hexanucleotide would allow FMRP's affinity for a RNA GQ to be determined more definitively.

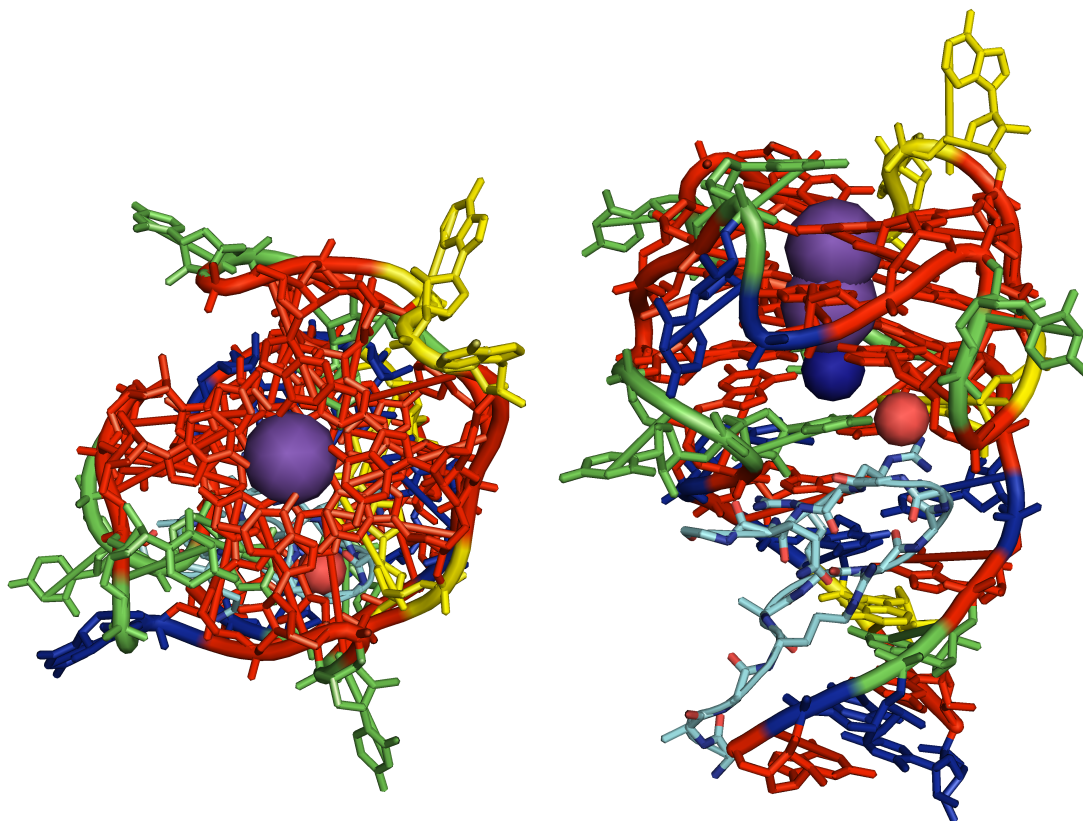


Figure 1.3: 2.8 Å crystal structure of cesium-bound human FMRP RGG motif bound to *sc1* RNA. Top view (left) reveals the guanosines (red) are assembling into G tetrads. Side view (right) shows two K^+ ions (purple spheres) and a Cs^+ ion (indigo sphere) each coordinating to O6 of the tetrad guanines. The Cs^+ coordination suggests K^+ is not specifically required at this site. Uridines, adenosines and cytidines are depicted in green, yellow and blue, respectively. The RGG motif (cyan) binds the *sc1* RNA below the GQ structure. The oxygen atom (red sphere) may mediate binding between the RGG motif and *sc1*. Protein Data Bank entry code 5DEA.

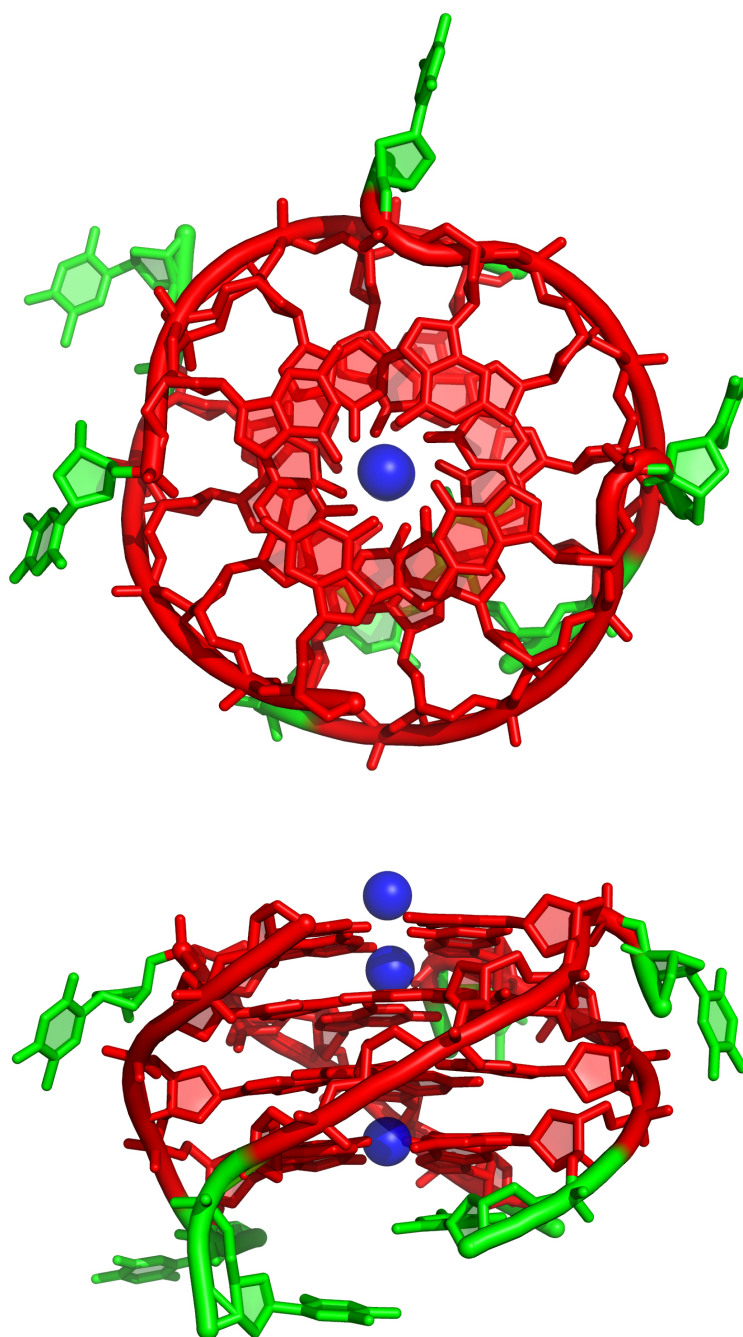


Figure 1.4: 0.61 Å crystal structure of (UG₄U)₄. Top view (top) and side view (bottom) of four UG₄U RNA molecules assembling into the intermolecular G-quadruplex (UG₄U)₄. Uridines are depicted in green and guanoses in red. The blue spheres passing through the axis of the helix represent Sr²⁺ ions which coordinate to the O6 of every other guanine base plane within the quadruplex. Protein Data Bank entry code 1J8G.

1.4 FMRP represses translation of target mRNAs

FMRP was found to be mainly expressed in the mammalian brain's neurons and reproductive organs^{45–47}. In each case, FMRP was found primarily involved within the cytoplasm where it interacted with translating polysomes and messenger ribonucleoprotein (mRNP) complexes^{48,49}.

It was later confirmed that FMRP is indeed repressing translation of target mRNAs locally at the dendritic spines of neurons^{32,50,51}. Early studies of translation repression by FMRP established that FMRP could inhibit target mRNA translation dose-dependently whereas the I304N mutant identified in the FXS patient could not^{32,50}.

While FMRP is understood to repress translation of RNAs, the mechanism(s) by which FMRP interacts with the target mRNA and the translating ribosome is still unknown. One theory is FMRP hinders translation initiation by binding to the target RNA and recruiting the cytoplasmic FMRP-interacting protein (CYFIP1) which then binds the eukaryotic initiation factor 4E (eIF4E) at the mRNA 5' 7-methylguanosine (m⁷G) cap. Binding of CYFIP1 to eIF4E in turn prevents binding of eukaryotic initiation factor 4G (eIF4G) to eIF4E, and the eukaryotic initiation factor 4F (eIF4F) complex is prevented from assembling at the m⁷G cap⁵².

Interestingly however, most of the cytoplasmic FMRP has been found to associate with polysomes that are actively translating mRNAs. Indeed, FMRP has been found to bind target mRNAs associated with stalled ribosomes^{34,53}. Additionally, FMRP has been shown to interact with the ribosome even after

translation initiation is blocked⁵⁴. Altogether, FMRP is suggested to repress translation by binding the target mRNA and stalling the ribosome during elongation.

Recently FMRP was shown to both repress translation of target mRNAs and bind directly to the ribosome in a manner consistent with the model of elongation repression. A cryo-electron microscopy (cryo-EM) model of *Drosophila* FMRP (dFMRP) bound to the *Drosophila* 80S ribosome depicts FMRP binding at a site within the ribosome intersubunit space that would prevent binding of tRNA and translation elongation factors³⁵. Additionally, the FMRP-ribosome structure reveals the KH1 and KH2 domains binding near the ribosome's peptidyl site and the RGG box domain of FMRP residing near the aminoacyl site, presumably available for binding target mRNA motifs (i.e. RNA GQ structures). Together with previous studies showing (1) FMRP primarily binds actively translating ribosomes (polysomes) and (2) FMRP binds RNA GQ structures through its RGG box domain, the cryo-EM structure suggests FMRP may tether itself to the ribosome and the mRNA via its KH and RGG domains, respectively, to stall the ribosome during elongation^{33–35,41,55,56}.

FMRP is also suggested to repress translation via RNA interference (RNAi). FMRP has been shown to associate with the RNA-induced silencing complex (RISC) proteins Dicer and Argonaute 2 (Ago2) as well as specific microRNAs (miRNAs)^{57–59}. In fact, FMRP has been reported to assemble the Ago2 and miRNA-125a inhibitory complex on the postsynaptic density protein 95 (PSD-95) mRNA through phosphorylation of FMRP^{60,61}. The PSD-95 RNAi case may be just one of many RNAi translation repression pathways FMRP employs. It is possible that FMRP can

recruit from a catalog of miRNA complexes to specifically inhibit a variety of mRNAs via RNAi.

Post-translational modification of FMRP can also modulate FMRP activity. One report finds serine 500 of human FMRP must be phosphorylated for FMRP to bind the ribosome and inhibit translation⁵⁴. Additionally, methylation of four specific arginines within the RGG box domain of FMRP has been found to reduce FMRP binding to GQ-assembling target mRNAs^{62,63}. Regarding the aforementioned case of PSD-95 assembly by phosphorylated FMRP, signaling of the metabotropic glutamate receptor (mGluR) has been shown to decrease FMRP phosphorylation and disassemble the miRNA-125a-Ago2-PSD-95 inhibitory complex on the target mRNA⁶¹.

FMRP has been found to repress translation of target mRNAs through different mechanisms and during different stages (**Figure 1.5**)⁶⁴. It is possible FMRP employs all of these methods to specifically repress a wide catalog of target mRNAs under different conditions within the cell.

1.5 FMRP and the mGluR-LTD pathway

Research into FXS therapeutics has implicated FMRP in the Group 1 metabotropic glutamate receptor-long term depression (mGluR-LTD) pathway involved in learning and memory⁶⁵. To activate mGluR-LTD, certain proteins necessary for synaptic function must be synthesized⁶⁶. Studies in *Fmr1* knockout mice revealed a connection between excess activation of mGluR-LTD and increased

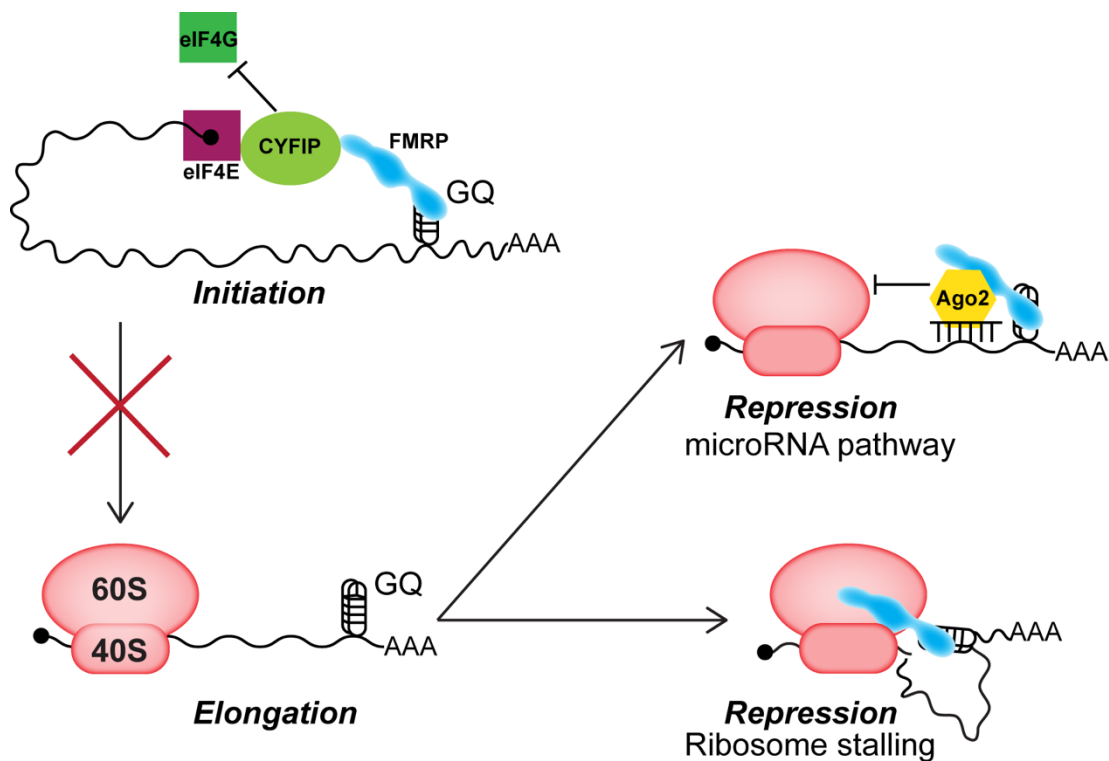


Figure 1.5: Models for translational repression by FMRP binding the target mRNA G-quadruplex. FMRP can inhibit translation initiation by binding mRNA and recruiting CYFIP, which prevents assembly of eIF4E and eIF4G into eIF4F at the 5' m⁷G cap (above). Alternatively, FMRP can inhibit translation elongation via direct binding to the ribosome to stall elongation (right), or via a Ago2/RISC mediator in the microRNA pathway (above).

internalization of the α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor (AMPA) from excessive production of the membrane protein, suggesting FMRP normally represses translation of synaptic proteins required for mGluR-LTD⁶⁷. With loss of regulation by FMRP, these proteins are overly expressed which creates an imbalance between long term depression (LTD) and long term potentiation (LTP). This imbalance in turn is thought to degrade neurological function⁶⁸.

1.6 Translating FMRP and FXS studies into animal models

The FMR1 gene is highly conserved amongst eukaryotes. Several orthologs exist even amongst model organisms such as *Drosophila melanogaster*, zebrafish and mouse. The *Drosophila* dFmr1 gene which encodes *Drosophila* FMRP (dFMRP) shares 56% amino acid sequence identity with human FMRP (hFMRP), including RNA binding domains that are 75% identical. In fact, dFMRP can regulate translation of some of the same mRNAs as hFMRP^{69–71}.

The Gideon Dreyfuss group were the first to characterize dFmr1. They highlighted the RNA binding capacity of dFMRP by mutating the KH1 and KH2 in a manner analogous to the I304N mutant isolated from the patient with the aggravated form of FXS⁷². They isolated two mutations, an I244N in the KH1 domain and an I307N in the KH2 domain. While homozygous expression of dFMRP caused aberrant apoptosis in the fly's eyes, expression of either I244N or I307N caused less apoptosis. However, heterozygous expression of I244N and I307N rescues wild type phenotype. This suggests both mutations are unique loss-of-function mutations

presumably from dFMRP's reduced RNA binding capacities. Further studies on dFMRP found that loss of dFMRP causes abnormal axon morphology within mushroom bodies as well as disrupted courtship behaviors, both similar symptoms found in FXS patients^{69,73,74}.

Like dFmr1, the mouse Fmr1 gene which encodes murine FMRP (mFMRP) shares 97% amino acid sequence identity with hFMRP. Deletion of Fmr1 resulted in several FXS-like symptoms within the knockout mice^{75,76}. Unfortunately, completely replicating the full mutation-causing CGG expansion at the Fmr1 5' UTR has proven futile as transcription of the Fmr1 gene does not seem to be silenced from hypermethylation as FMR1 is in humans¹⁵. However, the large CGG expanse within the 5' UTR did similarly drop mFMRP levels⁷⁷. Conditional knockout mice have been created more recently using Cre-Lox recombination. The Cre-Lox recombination method has empowered researchers to generate specific Fmr1 mutations only to desired cells (such as neurons) and study the effects in the mouse⁷⁸.

Both *Drosophila* and mice should prove invaluable in translating understanding at a molecular level to elucidating the effects of FMRP in FXS.

References

1. Coffee, B., Keith, K., Albizua, I., Malone, T., Mowrey, J., Sherman, S. L., and Warren, S. T. (2009) Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA. *Am J Hum Genet* 85, 503–14.
2. Ouyang, L., Grosse, S., Raspa, M., and Bailey, D. (2010) Employment impact and financial burden for families of children with fragile X syndrome: findings from the National Fragile X Survey. *J. Intellect. Disabil. Res. JIDR* 54, 918–928.
3. Garber, K. B., Visootsak, J., and Warren, S. T. (2008) Fragile X syndrome. *Eur J Hum Genet* 16, 666–72.
4. Maes, B., Fryns, J. P., Ghesquière, P., and Borghgraef, M. (2000) Phenotypic checklist to screen for fragile X syndrome in people with mental retardation. *Ment. Retard.* 38, 207–215.
5. Santoro, M. R., Bray, S. M., and Warren, S. T. (2012) Molecular Mechanisms of Fragile X Syndrome: A Twenty-Year Perspective. *Annu. Rev. Pathol. Mech. Dis.* 7, 219–245.
6. Irwin, S. A., Patel, B., Idupulapati, M., Harris, J. B., Crisostomo, R. A., Larsen, B. P., Kooy, F., Willems, P. J., Cras, P., Kozlowski, P. B., Swain, R. A., Weiler, I. J., and Greenough, W. T. (2001) Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: a quantitative examination. *Am J Med Genet* 98, 161–7.
7. Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P., and et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–14.
8. Fu, Y. H., Kuhl, D. P., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S., Verkerk, A. J., Holden, J. J., Fenwick, R. G., Warren, S. T., and et al. (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 67, 1047–58.
9. Colak, D., Zaninovic, N., Cohen, M. S., Rosenwaks, Z., Yang, W. Y., Gerhardt, J., Disney, M. D., and Jaffrey, S. R. (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* 343, 1002–5.

10. De Boulle, K., Verkerk, A. J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B. A., and Willems, P. J. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet* 3, 31–5.
11. Valverde, R., Pozdnyakova, I., Kajander, T., Venkatraman, J., and Regan, L. (2007) Fragile X mental retardation syndrome: structure of the KH1-KH2 domains of fragile X mental retardation protein. *Structure* 15, 1090–8.
12. Greco, C. M., Berman, R. F., Martin, R. M., Tassone, F., Schwartz, P. H., Chang, A., Trapp, B. D., Iwahashi, C., Brunberg, J., Grigsby, J., Hessel, D., Becker, E. J., Papazian, J., Leehey, M. A., Hagerman, R. J., and Hagerman, P. J. (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain* 129, 243–55.
13. Sherman, S. L. (2000) Premature ovarian failure among fragile X premutation carriers: parent-of-origin effect? *Am J Hum Genet* 67, 11–3.
14. Kenneson, A., Zhang, F., Hagedorn, C. H., and Warren, S. T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum Mol Genet* 10, 1449–54.
15. Brouwer, J., Mientjes, E., Bakker, C., Nieuwenhuizen, I., Severijnen, L., Van der Linde, H., Nelson, D., Oostra, B., and Willemsen, R. (2007) Elevated Fmr1 mRNA levels and reduced protein expression in a mouse model with an unmethylated Fragile X full mutation. *Exp. Cell Res.* 313, 244–253.
16. Ashley, C. T., Wilkinson, K. D., Reines, D., and Warren, S. T. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* 262, 563–6.
17. Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res* 21, 1193–8.
18. Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., and Ramos, A. (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res* 40, 6873–86.
19. Nicastro, G., Taylor, I. A., and Ramos, A. (2015) KH–RNA interactions: back in the groove. *Curr. Opin. Struct. Biol.* 30, 63–70.
20. Zang, J. B., Nosyreva, E. D., Spencer, C. M., Volk, L. J., Musunuru, K., Zhong, R., Stone, E. F., Yuva-Paylor, L. A., Huber, K. M., Paylor, R., Darnell, J. C.,

- and Darnell, R. B. (2009) A mouse model of the human Fragile X syndrome I304N mutation. *PLoS Genet* 5, e1000758.
21. Myrick, L. K., Nakamoto-Kinoshita, M., Lindor, N. M., Kirmani, S., Cheng, X., and Warren, S. T. (2014) Fragile X syndrome due to a missense mutation. *Eur J Hum Genet*.
 22. Thandapani, P., O'Connor, T. R., Bailey, T. L., and Richard, S. (2013) Defining the RGG/RG motif. *Mol Cell* 50, 613–23.
 23. Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B., and Patel, D. J. (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* 18, 796–804.
 24. Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, D. J., and Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci.* 112, E5391–E5400.
 25. Adinolfi, S., Ramos, A., Martin, S. R., Dal Piaz, F., Pucci, P., Bardoni, B., Mandel, J. L., and Pastore, A. (2003) The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. *Biochemistry* 42, 10437–44.
 26. Ramos, A., Hollingworth, D., Adinolfi, S., Castets, M., Kelly, G., Frenkiel, T. A., Bardoni, B., and Pastore, A. (2006) The structure of the N-terminal domain of the fragile X mental retardation protein: a platform for protein-protein interaction. *Structure* 14, 21–31.
 27. Maurer-Stroh, S., Dickens, N. J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C. P. (2003) The Tudor domain “Royal Family”: Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* 28, 69–74.
 28. Alpatov, R., Lesch, B. J., Nakamoto-Kinoshita, M., Blanco, A., Chen, S., Stutzer, A., Armache, K. J., Simon, M. D., Xu, C., Ali, M., Murn, J., Priscic, S., Kutateladze, T. G., Vakoc, C. R., Min, J., Kingston, R. E., Fischle, W., Warren, S. T., Page, D. C., and Shi, Y. (2014) A Chromatin-Dependent Role of the Fragile X Mental Retardation Protein FMRP in the DNA Damage Response. *Cell* 157, 869–81.
 29. Hu, Y., Chen, Z., Fu, Y., He, Q., Jiang, L., Zheng, J., Gao, Y., Mei, P., Chen, Z., and Ren, X. (2015) The amino-terminal structure of human fragile X mental retardation protein obtained using precipitant-immobilized imprinted polymers. *Nat Commun* 6, 6634.

30. Eberhart, D. E., Malter, H. E., Feng, Y., and Warren, S. T. (1996) The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum Mol Genet* 5, 1083–91.
31. Collins, S. C., Bray, S. M., Suhl, J. A., Cutler, D. J., Coffee, B., Zwick, M. E., and Warren, S. T. (2010) Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *Am. J. Med. Genet. A* 152A, 2512–2520.
32. Li, Z., Zhang, Y., Ku, L., Wilkinson, K. D., Warren, S. T., and Feng, Y. (2001) The fragile X mental retardation protein inhibits translation via interacting with mRNA. *Nucleic Acids Res.* 29, 2276–2283.
33. Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., Jin, X., Feng, Y., Wilkinson, K. D., Keene, J. D., Darnell, R. B., and Warren, S. T. (2001) Microarray Identification of FMRP-Associated Brain mRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome. *Cell* 107, 477–487.
34. Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., Stone, E. F., Chen, C., Fak, J. J., Chi, S. W., Licatalosi, D. D., Richter, J. D., and Darnell, R. B. (2011) FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* 146, 247–261.
35. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K., and Joseph, S. (2014) Fragile X Mental Retardation Protein Regulates Translation by Binding Directly to the Ribosome. *Mol. Cell* 54, 407–417.
36. Ascano, M., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., Williams, Z., Ohler, U., and Tuschl, T. (2012) FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382–386.
37. Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. O. F., Lei, E. P., Fraser, A. G., Blencowe, B. J., Morris, Q. D., and Hughes, T. R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–177.
38. McMahon, A. C., Rahman, R., Jin, H., Shen, J. L., Fieldsend, A., Luo, W., and Rosbash, M. (2016) TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 165, 742–753.

39. Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J., and Warren, S. T. (2014) Analysis of FMRP mRNA target datasets reveals highly associated mRNAs mediated by G-quadruplex structures formed via clustered WGGA sequences. *Hum. Mol. Genet.* 23, 5479–5491.
40. Tabet, R., Moutin, E., Becker, J. A. J., Heintz, D., Fouillen, L., Flatter, E., Krężel, W., Alunni, V., Koebel, P., Dembélé, D., Tassone, F., Bardoni, B., Mandel, J.-L., Vitale, N., Muller, D., Le Merrer, J., and Moine, H. (2016) Fragile X Mental Retardation Protein (FMRP) controls diacylglycerol kinase activity in neurons. *Proc. Natl. Acad. Sci.* 113, E3619–E3628.
41. Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. (2001) Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function. *Cell* 107, 489–499.
42. Ramos, A. (2003) G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA* 9, 1198–1207.
43. Deng, J., Xiong, Y., and Sundaralingam, M. (2001) X-ray analysis of an RNA tetraplex (UGGGU)₄ with divalent Sr²⁺ ions at subatomic resolution (0.61 Å). *Proc. Natl. Acad. Sci.* 98, 13665–13670.
44. Kim, J., Cheong, C., and Moore, P. B. (1991) Tetramerization of an RNA oligonucleotide containing a GGGG sequence. *Nature* 351, 331–2.
45. Abitbol, M., Menini, C., Delezoide, A. L., Rhyner, T., Vekemans, M., and Mallet, J. (1993) Nucleus basalis magnocellularis and hippocampus are the major sites of FMR-1 expression in the human fetal brain. *Nat Genet* 4, 147–53.
46. Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P., and Mandel, J. L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet* 4, 335–40.
47. Tamanini, F., Willemsen, R., van Unen, L., Bontekoe, C., Galjaard, H., Oostra, B. A., and Hoogeveen, A. T. (1997) Differential expression of FMR1, FXR1 and FXR2 proteins in human brain and testis. *Hum Mol Genet* 6, 1315–22.
48. Khandjian, E. W., Corbin, F., Woerly, S., and Rousseau, F. (1996) The fragile X mental retardation protein is associated with ribosomes. *Nat Genet* 12, 91–3.
49. Khandjian, E. W., Huot, M.-E., Tremblay, S., Davidovic, L., Mazroui, R., and Bardoni, B. (2004) Biochemical evidence for the association of fragile X mental retardation protein with brain polyribosomal ribonucleoparticles. *Proc. Natl. Acad. Sci.* 101, 13357–13362.

50. Lagerbauer, B., Ostareck, D., Keidel, E. M., Ostareck-Lederer, A., and Fischer, U. (2001) Evidence that fragile X mental retardation protein is a negative regulator of translation. *Hum Mol Genet* 10, 329–38.
51. Mazroui, R., Huot, M. E., Tremblay, S., Fillion, C., Labelle, Y., and Khandjian, E. W. (2002) Trapping of messenger RNA by Fragile X Mental Retardation protein into cytoplasmic granules induces translation repression. *Hum Mol Genet* 11, 3007–17.
52. Napoli, I., Mercaldo, V., Boyl, P. P., Eleuteri, B., Zalfa, F., De Rubeis, S., Di Marino, D., Mohr, E., Massimi, M., Falconi, M., Witke, W., Costa-Mattioli, M., Sonenberg, N., Achsel, T., and Bagni, C. (2008) The Fragile X Syndrome Protein Represses Activity-Dependent Translation through CYFIP1, a New 4E-BP. *Cell* 134, 1042–1054.
53. Stefani, G. (2004) Fragile X Mental Retardation Protein Is Associated with Translating Polyribosomes in Neuronal Cells. *J. Neurosci.* 24, 7272–7276.
54. Ceman, S., O'Donnell, W. T., Reed, M., Patton, S., Pohl, J., and Warren, S. T. (2003) Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum. Mol. Genet.* 12, 3295–3305.
55. Agrawal, R. K., Spahn, C. M., Penczek, P., Grassucci, R. A., Nierhaus, K. H., and Frank, J. (2000) Visualization of tRNA movements on the Escherichia coli 70S ribosome during the elongation cycle. *J Cell Biol* 150, 447–60.
56. Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* 20, 4803–13.
57. Caudy, A. A., Myers, M., Hannon, G. J., and Hammond, S. M. (2002) Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* 16, 2491–6.
58. Ishizuka, A., Siomi, M. C., and Siomi, H. (2002) A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* 16, 2497–508.
59. Plante, I., Davidovic, L., Ouellet, D. L., Gobeil, L. A., Tremblay, S., Khandjian, E. W., and Provost, P. (2006) Dicer-derived microRNAs are utilized by the fragile X mental retardation protein for assembly on target RNAs. *J Biomed Biotechnol* 2006, 64347.
60. Edbauer, D., Neilson, J. R., Foster, K. A., Wang, C. F., Seeburg, D. P., Battersby, M. N., Tada, T., Dolan, B. M., Sharp, P. A., and Sheng, M. (2010)

Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron* 65, 373–84.

61. Muddashetty, R. S., Nalavadi, V. C., Gross, C., Yao, X., Xing, L., Laur, O., Warren, S. T., and Bassell, G. J. (2011) Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. *Mol Cell* 42, 673–88.
62. Stetler, A., Winograd, C., Sayegh, J., Cheever, A., Patton, E., Zhang, X., Clarke, S., and Ceman, S. (2006) Identification and characterization of the methyl arginines in the fragile X mental retardation protein Fmrp. *Hum Mol Genet* 15, 87–96.
63. Blackwell, E., Zhang, X., and Ceman, S. (2010) Arginines of the RGG box regulate FMRP association with polyribosomes and mRNA. *Hum. Mol. Genet.* 19, 1314–1323.
64. Chen, E., and Joseph, S. (2015) Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie* 114, 147–154.
65. Huber, K. M., Kayser, M. S., and Bear, M. F. (2000) Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* 288, 1254–7.
66. Weiler, I. J., Irwin, S. A., Klintsova, A. Y., Spencer, C. M., Brazelton, A. D., Miyashiro, K., Comery, T. A., Patel, B., Eberwine, J., and Greenough, W. T. (1997) Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. *Proc Natl Acad Sci U S A* 94, 5395–400.
67. Snyder, E. M., Philpot, B. D., Huber, K. M., Dong, X., Fallon, J. R., and Bear, M. F. (2001) Internalization of ionotropic glutamate receptors in response to mGluR activation. *Nat. Neurosci.* 4, 1079–1085.
68. Sidorov, M. S., Auerbach, B. D., and Bear, M. F. (2013) Fragile X mental retardation protein and synaptic plasticity. *Mol Brain* 6, 15.
69. Zhang, Y. Q., Bailey, A. M., Matthies, H. J., Renden, R. B., Smith, M. A., Speese, S. D., Rubin, G. M., and Broadie, K. (2001) Drosophila fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell* 107, 591–603.
70. Lee, A., Li, W., Xu, K., Bogert, B. A., Su, K., and Gao, F. B. (2003) Control of dendritic development by the Drosophila fragile X-related gene involves the small GTPase Rac1. *Development* 130, 5543–52.

71. Reeve, S. P., Bassetto, L., Genova, G. K., Kleyner, Y., Leyssen, M., Jackson, F. R., and Hassan, B. A. (2005) The *Drosophila* fragile X mental retardation protein controls actin dynamics by directly regulating profilin in the brain. *Curr. Biol. CB* 15, 1156–1163.
72. Wan, L., Dockendorff, T. C., Jongens, T. A., and Dreyfuss, G. (2000) Characterization of dFMR1, a *Drosophila melanogaster* homolog of the fragile X mental retardation protein. *Mol Cell Biol* 20, 8536–47.
73. Morales, J., Hiesinger, P. R., Schroeder, A. J., Kume, K., Verstreken, P., Jackson, F. R., Nelson, D. L., and Hassan, B. A. (2002) *Drosophila* fragile X protein, DFXR, regulates neuronal morphology and function in the brain. *Neuron* 34, 961–72.
74. Tessier, C. R., and Broadie, K. (2008) *Drosophila* fragile X mental retardation protein developmentally regulates activity-dependent axon pruning. *Development* 135, 1547–57.
75. Bakker, C., Verheij, C., Willemsen, R., van der Helm, R., and Oerlemans, F. (1994) *Fmr1* knockout mice: a model to study fragile X mental retardation. The Dutch-Belgian Fragile X Consortium. *Cell* 78, 23–33.
76. Comery, T. A., Harris, J. B., Willems, P. J., Oostra, B. A., Irwin, S. A., Weiler, I. J., and Greenough, W. T. (1997) Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. *Proc Natl Acad Sci U A* 94, 5401–4.
77. Bontekoe, C. J., Bakker, C. E., Nieuwenhuizen, I. M., van der Linde, H., Lans, H., de Lange, D., Hirst, M. C., and Oostra, B. A. (2001) Instability of a (CGG)₉₈ repeat in the *Fmr1* promoter. *Hum. Mol. Genet.* 10, 1693–1699.
78. Mientjes, E. J., Nieuwenhuizen, I., Kirkpatrick, L., Zu, T., Hoogeveen-Westerveld, M., Severijnen, L., Rife, M., Willemsen, R., Nelson, D. L., and Oostra, B. A. (2006) The generation of a conditional *Fmr1* knock out mouse model to study *Fmrp* function in vivo. *Neurobiol Dis* 21, 549–55.

Chapter 2:
RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein

Abstract

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability and is caused by a deficiency of the fragile X mental retardation protein (FMRP) in neurons. FMRP regulates the translation of numerous mRNAs within dendritic synapses, but how FMRP recognizes these target mRNAs remains unknown. FMRP has KH0, KH1, KH2, and RGG domains, which are thought to bind to specific RNA recognition elements (RREs). Several studies used high-throughput methods to identify various RREs in mRNAs that FMRP may bind to in vivo. However, there is little overlap in the mRNA targets identified by each study, suggesting that the RNA-binding specificity of FMRP is still unknown. To determine the specificity of FMRP for the RREs, we performed quantitative in vitro RNA binding studies with various constructs of human FMRP. Unexpectedly, our studies show that the KH domains do not bind to the previously identified RREs. To further investigate the RNA-binding specificity of FMRP, we developed a new method called Motif Identification by Analysis of Simple sequences (MIDAS) to identify single-stranded RNA (ssRNA) sequences bound by KH domains. We find that the FMRP KH0, KH1, and KH2 domains bind weakly to the ssRNA sequences suggesting that they may have evolved to bind more complex RNA structures. Additionally, we find that the RGG motif of human FMRP binds with a high affinity to an RNA G-quadruplex (GQ) structure that lacks single-stranded loops, double-stranded stems, or junctions.

Introduction

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability. FXS is primarily caused by the deficiency of the Fragile X Mental Retardation Protein (FMRP) in neuronal cells¹. The expansion of CGG trinucleotide repeats in the 5'-untranslated region of the *FMR1* gene leads to gene silencing and FMRP deficiency²⁻⁵. FMRP has been proposed to regulate local protein synthesis within dendritic synapses⁶⁻⁸. FMRP deficiency results in the formation of dendrites with long, thin, and immature spines possibly caused by the deregulation of translation of specific neuronal mRNAs⁹⁻¹¹. However, how FMRP binds to specific neuronal mRNAs to regulate their translation is unknown.

FMRP contains several conserved domains. Amongst the conserved domains, four are putative RNA-binding domains. Three K-Homology (KH) domains — KH0, KH1, and KH2 — and an arginine-glycine-glycine (RGG) motif are hypothesized to mediate FMRP binding to mRNAs^{12,13}. The KH domain was first identified as a nucleic acid-binding domain in heterogeneous nuclear ribonucleoprotein K (hnRNP K) and has since been found in many eukaryotic (Type I) and prokaryotic (Type II) proteins^{14,15}. A GXXG loop in the KH1 and KH2 domains of FMRP is conserved in many RNA-binding KH domains, such as the KH-type splicing regulatory protein (KSRP) and neuro-oncological ventral antigen (Nova-1 and 2) proteins, further suggesting FMRP KH domains play a role in binding specific RNAs^{16,17}. A third KH domain was recently discovered upstream of the KH1 domain, termed KH0^{18,19}. Unlike KH1 and KH2, KH0 does not contain the GXXG loop, which

suggests KH0 may not bind RNA or bind a different RNA motif. Mutations in the KH1 and KH2 domains of FMRP also result in FXS, indicating that the RNA-binding function of FMRP is essential for normal health^{20,21}. The KH2 domain was shown to bind to an in vitro-selected kissing complex RNA called KC2²². Additionally, the RGG motif binds to an in vitro-selected G-quadruplex forming RNA called Sc1²³. An NMR structure of the human FMRP RGG peptide bound to Sc1 RNA G-quadruplex (GQ) reveals how the disordered RGG motif is stabilized upon binding to the GQ structure²⁴. A subsequent crystal structure reveals in atomic detail the interaction of each amino acid within the RGG motif with the GQ nucleotides of Sc1²⁵. Interestingly, the RGG peptide interacts not with the GQ region, but binds to the junction between the GQ and a double-stranded stem in SC1. While Sc1 was selected as a high-affinity binding partner for the RGG motif, it suggested that the RGG motif mediates FMRP binding to GQ structures in mRNAs. Indeed, the FMRP RGG peptide has been shown to bind to various GQ structures found in mRNAs with a range of binding affinities²⁶⁻³⁰. Given the ubiquity of potential GQ structures throughout the transcriptome, the presence of the RGG motif within FMRP hints at FMRP's potential to bind a vast pool of target mRNAs in vivo.

The identification of FMRP target mRNAs is crucial for understanding the molecular function of FMRP and has been the focus of numerous studies. Most endeavors to identify FMRP target mRNA motifs have employed high-throughput methods such as coimmunoprecipitation of mRNA-FMRP complexes (Co-IP), cross-linking and immunoprecipitation combined with high-throughput sequencing (HITS-CLIP), photoactivatable ribonucleoside-enhanced cross-linking and

immunoprecipitation (PAR-CLIP), RNAcompete, and targets of RNA-binding proteins identified by editing (TRIBE)³¹⁻³⁵. Unfortunately, the proposed catalog of RNA sequence and structural motifs, or RNA-recognition elements (RREs), overlap poorly between each study³⁶. Analysis of the proposed RREs yielded two promising RREs: clustered WGGA (W = A/U) sequences and the GACR (R = A/G) sequence³⁶. The WGGA clusters were originally proposed to bind specifically to the KH1 domain of FMRP³¹. A subsequent study proposed that WGGA clusters could form GQ structures and therefore bind the RGG motif of FMRP instead³⁶. The GACR sequences were enriched in RNAcompete for an FMRP fragment composed solely of the KH1 and KH2 RNA binding domains, suggesting that the GACR motif binds a KH domain³⁵.

Here, we use fluorescence anisotropy to quantify binding affinities between FMRP and RNAs of a defined length and sequence. We modeled the RNAs after proposed FMRP RREs^{31,35,36}. Unexpectedly, our studies show that the KH0, KH1, and KH2 domains do not bind to the WGGA, ACUK, and GACR motifs. Previous studies showed that many KH domains recognize only four to five nucleotides in ssRNAs^{16,17,37-39}. This inspired us to develop a new method to identify RNA sequences that bind to KH domains, which we named Motif Identification by Analysis of Simple sequences (MIDAS). We synthesized a pool of RNAs having five randomized nucleotides flanked by fixed sequences of eight nucleotides at the 5' and 3' ends. We used the randomized RNA pool to identify RNA molecules that bind to the KH domains of FMRP (the pool will contain $4^5 = 1024$ unique sequences). Surprisingly, the FMRP KH domains bind weakly to the selected RNA sequences

suggesting that the KH domains of FMRP may not interact with simple RREs as the canonical KH domains. Instead, the FMRP KH domains may have evolved to recognize more complex RNA motifs. Additionally, consistent with previous studies, we find that the RGG motif of human FMRP binds to RNA GQ structures with a 100-fold range in affinities showing that FMRP can discriminate between the different GQ structures^{26-30,40,41}.

Results and Discussion

2.1 Isolating the RNA-binding domains of FMRP

To elucidate how FMRP specifically binds its target RNAs, we isolated specific regions of FMRP and quantified their binding affinity to a catalog of RNAs modeled after reported FMRP RREs. We sought to both test the validity of the proposed RREs within RNAs and clarify the roles of the KH0, KH1, KH2, and RGG domains in binding the RREs. We expressed and purified nine human FMRP (hFMRP) variants from *E. coli* for our studies (**Figure 2.1**). The N-terminus truncated human NT-FMRP construct was cloned by deleting the amino acids immediately upstream of the KH1 domain from the full-length sequences (NT-hFMRP encodes Arg218 to Pro632). This truncation eliminates the FMRP N-terminal domain (NTD), which contains the tandem Agenet 1 and Agenet 2 domains as well as the KH0 domain; the NTD is also responsible for FMRP dimerization¹⁸. The NT-hFMRP construct conserves the three proposed RNA-binding domains⁴²⁻⁴⁴. We also generated a NT-hFMRP Δ KH1

construct encoding Phe281 to Pro632 to determine the binding specificity of the KH2 domain. Additionally, we made individual KH0, KH1, KH2, KH1-KH2, and KH0-KH1-KH2 fused downstream of the glutathione S-transferase (GST) tag. The fusion proteins are large enough to generate clear shifts in fluorescence anisotropy should the KH motifs bind the fluorescein-labeled RNAs.

To evaluate the importance of the RGG motif in binding RNA GQ structures, we generated an hFMRP construct encoding Gly531 of the RGG domain to the C-terminal Pro632 and fused it downstream of the GST tag to generate a GST-hFMRP RGG fusion protein. Finally, we mutated all of the arginines to serines in the hRGG-encoding region, generating a GST-hFMRP SGG fusion protein. Together these FMRP constructs should delineate the contributions of the KH domains and the RGG motif towards binding to RNAs.

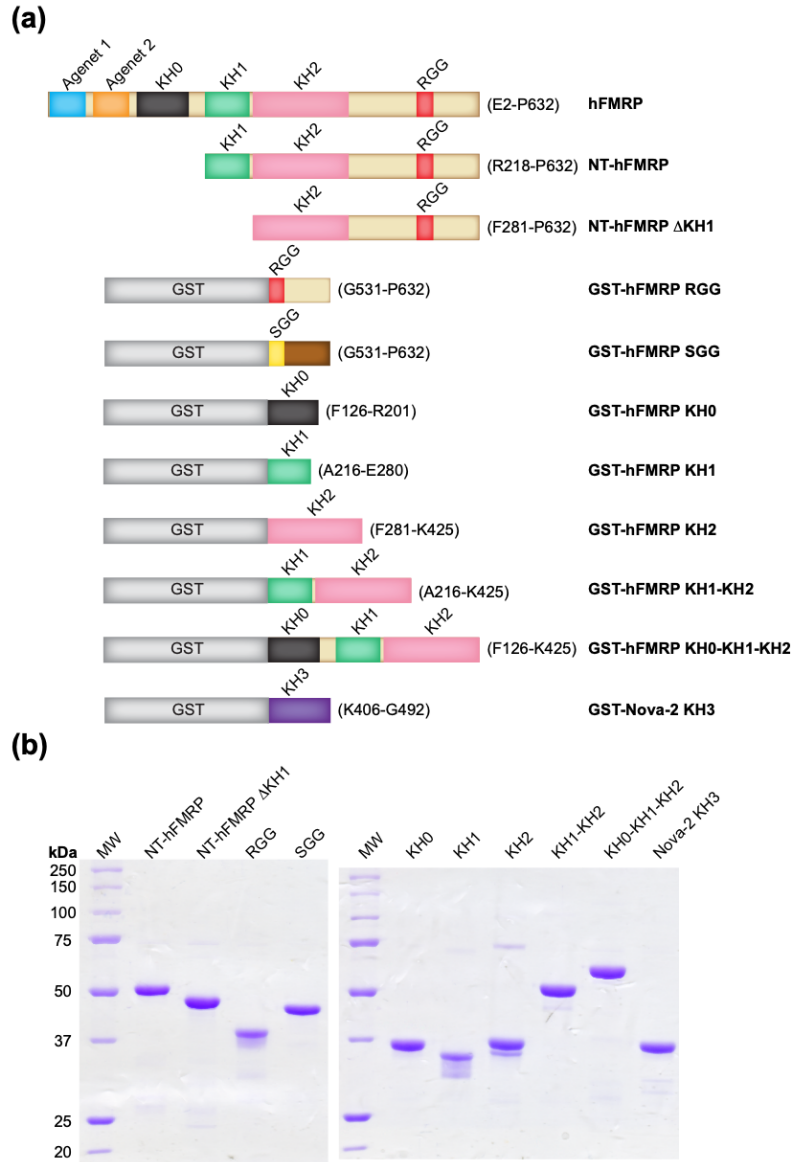


Figure 2.1: Design and purification of human FMRP constructs. (a) hFMRP is the full-length human FMRP isoform 1, spanning E2-P632 as indicated. The different truncation constructs of hFMRP that were made are illustrated. GST-hFMRP SGG is a fusion between the glutathione S-transferase and the RGG motif-containing sequence of hFMRP spanning G531-P632 where the 16 arginines spanning the RGG motif to the C-terminus were mutated to serines, illustrated using a different color for the C-terminus. (b) SDS-polyacrylamide gel showing the purified proteins illustrated above. Lanes are: molecular weight standards (MW), NT-hFMRP (47 kDa), NT-hFMRP ΔKH1 (40 kDa), GST-hFMRP RGG (40 kDa), GST-hFMRP SGG (39 kDa), GST-hFMRP KH0 (38 kDa), GST-hFMRP KH1 (36 kDa), GST-hFMRP KH2 (45 kDa), GST-hFMRP KH1-KH2 (52 kDa), GST-hFMRP KH0-KH1-KH2 (63 kDa), and GST-Nova-2 KH3 (38 kDa).

2.2 Design of model RNAs

We designed a set of RNAs modeled after three promising FMRP RREs — GQ structure, GACR clusters, and WGGGA clusters — to determine if and how FMRP specifically binds these sequences (**Table 2.1**). Apart from the UG₄U 6-mer RNA, we standardized our model RNAs to 18 nucleotides in length to control for length-dependent binding affinities; M1 served as an 8-mer RNA size marker for UG₄U. We used PolyG₁₇ and UG₄U as two different model GQ RNAs to quantify and compare FMRP binding affinities for different RNA GQ structures⁴⁵⁻⁴⁷. UG₄U is noteworthy because the x-ray crystal structure showed that it forms a highly stable, four stacked G-tetrad structure, and we wanted to test whether the RGG domain could bind to a GQ structure that does not contain single-stranded regions, junctions or double-stranded stems⁴⁶.

We generated (GACG)₄ and (UGGA)₄ 18-mer RNAs containing four tandem repeats of GACG and UGGA to test FMRP binding to the GACR cluster and WGGGA cluster, respectively. The GACR array was reported to bind the KH1 and KH2 domains of FMRP³⁵. The WGGGA array was originally reported to bind the KH1 domain but was later hypothesized to assemble an intramolecular GQ structure that may bind the FMRP RGG motif^{31,36}. We also tested FMRP binding to ACUU and UGGA RNAs to determine if a single ACUK (K = G/U) or WGGGA motif is sufficient for binding to FMRP, respectively³¹. Finally, we used two 18-mer RNAs with different base compositions that cannot assemble any secondary structures as control RNAs, denoted PolyC₁₇, and CR1. All RNAs had a 3'-terminal uridine to prevent quenching

of the fluorescein dye attached to the 3' end⁴⁸. The fluorescein dye was used to visualize the RNAs on gels and for performing fluorescence anisotropy binding studies.

Table 2.1. Nomenclature and sequences of the RNAs

RNA	Sequence (5' to 3')
PolyG ₁₇	GGGGGGGGGGGGGGGGGGU
UG ₄ U	UGGGGU
(UGGA) ₄	UUGGAUGGAUGGAUGGAU
(GACG) ₄	UGACGGACGGACGGACGU
UGGm ⁶ ACU	UGGACUUGGm ⁶ ACUUGGACU
PolyC ₁₇	CCCCCCCCCCCCCCCCCU
ACUU	CCCCCCC <u>ACU</u> CCCCCU
UGGA	CCCCCCC <u>UGG</u> CCCCCU
CR1	GCUAUCCAGAUUCUGAUU

2.3 Folding of model RNAs

We first tested whether our model RNAs assemble into folded structures by denaturing and native PAGE analyses (**Figure 2.2**). We tested whether UG₄U assembles the reported 24-mer intermolecular GQ composed of 4 stacked G-tetrads using two methods: (1) a denaturing urea PAGE in the presence of KCl, NaCl or LiCl as described previously, and (2) a native PAGE comparing UG₄U migration to the 8-mer M1, 18-mer PolyC₁₇ and 18-mer CR1 RNAs, which remain unfolded⁴⁹. The denaturing PAGE showed UG₄U migrating as two different species, with the presence of the slow-migrating species being K⁺-dependent (**Figure 2.2a**). The UG₄U GQ assembly was remarkably stable as it endured urea denaturing conditions even with Li⁺. The native PAGE showed UG₄U migrating much slower than the 8-nucleotide M1 RNA and slightly slower than the 18-mer RNAs (**Figure 2.2c**). Together, both PAGE analyses indicate that UG₄U indeed assembles the tetrameric GQ structure. All 18-nucleotide RNAs migrated as expected on a denaturing PAGE (**Figure 2.2b**). However, on a native PAGE, PolyG₁₇ migrated much slower than the control 18-nucleotide RNAs as a heterogeneous array (**Figure 2.2c**). This suggests the PolyG₁₇ array may comprise a mix of 36-nucleotide intermolecular GQ structures in different registers. This is consistent with a recent report that showed that PolyG₄₀ forms stable GQ structures⁴⁷.

The (GACG)₄ and (UGGA)₄ RNAs migrated differently from the other RNAs. The (GACG)₄ RNA migrated as two species, one migrating slowly at a similar rate as PolyG₁₇ and the other migrating faster than the unstructured 18-nucleotide control

RNAs (**Figure 2.2c**). Unlike PolyG₁₇, (GACG)₄ does not contain four repeats of at least two consecutive guanines required to assemble an intramolecular GQ structure^{50,51}. The slower (GACG)₄ species may be a stable dimer structure formed by intermolecular base pairs (**Figure 2.3**). The faster (GACG)₄ species could then be the monomeric 18-mer stem-loop structure, which would be expected to migrate faster than the unfolded control RNAs.

The (UGGA)₄ RNA is predicted to assemble an intramolecular GQ structure composed of 2 stacked G-tetrads^{52,53}. It is indeed migrating at a similar rate as one of the PolyG₁₇ species within the PolyG₁₇ array (**Figure 2.2c**). To verify whether (UGGA)₄ is forming a GQ structure, we used the BG4 antibody that binds specifically to GQ structures⁵⁴. The binding of the BG4 antibody to the fluorescein-labeled RNAs was monitored by fluorescence anisotropy, which should increase because the RNA-BG4 antibody complex will tumble slower than the free RNAs. We observed a substantial increase in anisotropy with UG₄U and PolyG₁₇, which is consistent with the fact that these RNAs form GQ structures (**Figure 2.2d**). The (UGGA)₄ RNA showed a modest increase in anisotropy, suggesting that the BG4 antibody binds poorly to the GQ structure formed by (UGGA)₄. Alternatively, the change in anisotropy is small because the fluorescein attached to the RNA is dynamic in the (UGGA)₄ RNA-BG4 antibody complex and does not correlate with the tumbling rate of the complex. No change in anisotropy was observed with (GACG)₄, indicating that this sequence does not form a GQ structure.

To further validate the formation of GQ structures, we used the N-methyl mesoporphyrin IX (NMM) dye, which shows enhanced fluorescence when it

intercalates into GQ structures^{55,56}. We examined the presence of GQ structures in PolyG₁₇, UG₄U, (UGGA)₄, (GACG)₄, and tRNA reconstituted in buffer containing either KCl or LiCl with the NMM dye. The tRNA served as a negative control. Consistent with the native gel analysis and the BG4 antibody assay, we observed increased fluorescence intensity with PolyG₁₇, UG₄U, and (UGGA)₄ but not with (GACG)₄ and tRNA (**Figure 2.2e**). The increase in the fluorescence intensity was higher when the RNAs were in KCl than in LiCl, demonstrating that PolyG₁₇, UG₄U, and (UGGA)₄ form GQ structures.

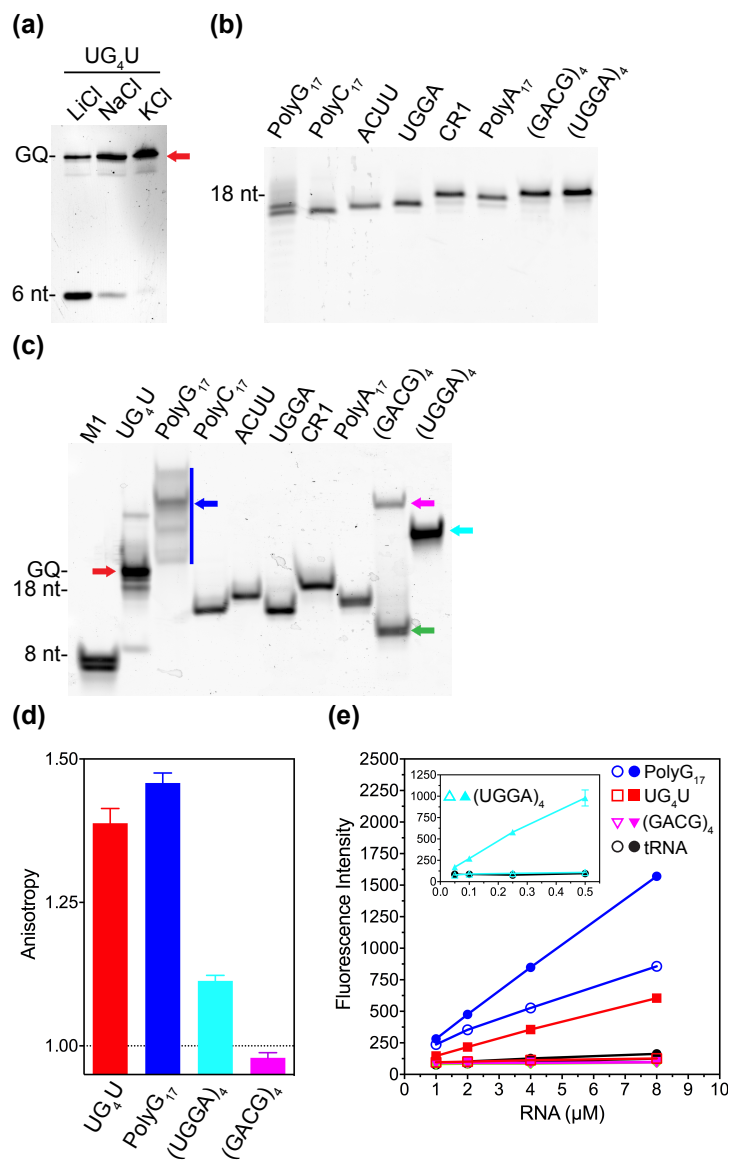


Figure 2.2: Analysis of the structures formed by the model RNAs. (a) Urea PAGE of UG₄U denatured and then refolded in the presence of 100 mM LiCl, 100 mM NaCl or 100 mM KCl. (b) Urea PAGE of the denatured 18-mer model RNAs. (c) Native PAGE of the model RNAs. G-quadruplex species are marked with red arrows and a blue bar; (GACG)₄ hairpin monomer and dimer species are marked with green and magenta arrows, respectively; (UGGA)₄ structure is marked with a cyan arrow. (d) Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled RNAs upon the addition of 430 nM anti-GQ BG4 scFV antibody. All RNAs are labeled at 3' end with fluorescein. The standard deviations from two experiments are shown. (e) Increase in fluorescence intensity of NMM with increasing concentrations of unlabeled RNA. The inset shows the data for the (UGGA)₄ RNA performed at lower concentrations of RNA. The closed and open symbols show RNAs in KCl and LiCl, respectively. The standard deviations from three experiments are shown.



Figure 2.3: (GACG)₄ may form hairpin or duplex structures. (a) Predicted secondary structure of (GACG)₄ is a hairpin stabilized by two GC base pairs and one AU base pair. (b) Two (GACG)₄ RNAs are shown dimerizing through Watson-Crick base pairing to form a duplex structure.

2.4 FMRP binds to different G-quadruplex RNAs

Little is known regarding how FMRP specifically binds its target mRNAs. A previous study suggested that the KH1 and KH2 domains of *Drosophila* FMRP bind to the ribosome while the RGG motif is available to bind the target mRNA⁵⁷. Previous reports also showed that the FMRP RGG motif binds to the in vitro selected Sc1 RNA GQ structure and to a variety of GQ structures from potential mRNA targets of FMRP^{23,25-30,40,41,58}. However, it is unknown whether FMRP can bind with a high affinity to a GQ structure that does not have any single-stranded regions, double-stranded stems, or junctions. We employed fluorescence anisotropy to determine the equilibrium dissociation constants (K_D) of several human FMRP variants binding to RNAs that assemble different GQ structures or contain RREs that have been reported to bind FMRP (**Table 2.1**).

We observed that NT-hFMRP binds to PolyG₁₇ with a K_D of 14 ± 2 nM and UG₄U with a K_D of 110 ± 16 nM (**Figure 2.4a, Table 2.2**). Additionally, NT-hFMRP binds to (UGGA)₄ with a K_D of 710 ± 50 nM. Since all three RNAs form GQ structures, our results show that NT-hFMRP binds to the different GQ structures with a 100-fold range in binding affinities. Interestingly, NT-hFMRP failed to bind to the (GACG)₄ RNA, which was modeled after the GACR array reported to bind to the KH domains of FMRP (**Figure 2.4a**). The four tandem GACG repeats may be insufficient for binding to FMRP. The (GACG)₄ RNA could assemble a mixed population of GC stem-loop monomers and double-stranded dimers. FMRP's inability to bind either the GC stem-loop monomer or dimer would suggest FMRP does not bind double-

stranded RNA assemblies even if they are G-rich. NT-hFMRP did not bind to RNAs having a single UGGA or ACUU, which cannot form GQ structures (**Figure 2.4a**). Finally, despite lacking the KH1 domain, NT-hFMRP Δ KH1 showed a binding pattern similar to NT-hFMRP (**Figure 2.4b, Table 2.2**).

2.5 FMRP RGG motif specifically binds to RNA G-quadruplex structures

To determine whether the RGG motif is responsible for binding to the RNA GQ structures, we tested the binding of our model RNAs to GST-hFMRP RGG. Consistent with previous studies, the RGG motif was sufficient for binding PolyG₁₇ and UG₄U GQ structures (**Figure 2.4c, Table 2.2**)^{23,25-30,40,41,58}. In fact, GST-hFMRP RGG bound PolyG₁₇ with a K_D of 8.6 ± 1.2 nM and UG₄U with a K_D of 210 ± 18 nM, similar affinities as NT-hFMRP. The nanomolar affinity of NT-hFMRP for UG₄U shows that the RGG domain can bind to GQ structures that do not contain single-stranded loops, double-stranded stems, or junctions. Interestingly, GST-hFMRP RGG bound (UGGA)₄ with a K_D of > 4000 nM, a significantly lower affinity than NT-hFMRP. This suggests that the region from the KH2 to the RGG domain may be necessary for binding to (UGGA)₄. To confirm that the RGG motif is required for binding to GQ RNAs, we made a GST-hFMRP SGG construct with all sixteen arginines in the RGG motif and C-terminal domain changed to serines. The GST-hFMRP SGG protein did not bind to PolyG₁₇ and UG₄U GQ structures, demonstrating that the RGG motif is essential for binding to GQ RNAs (**Figure 2.4d**).

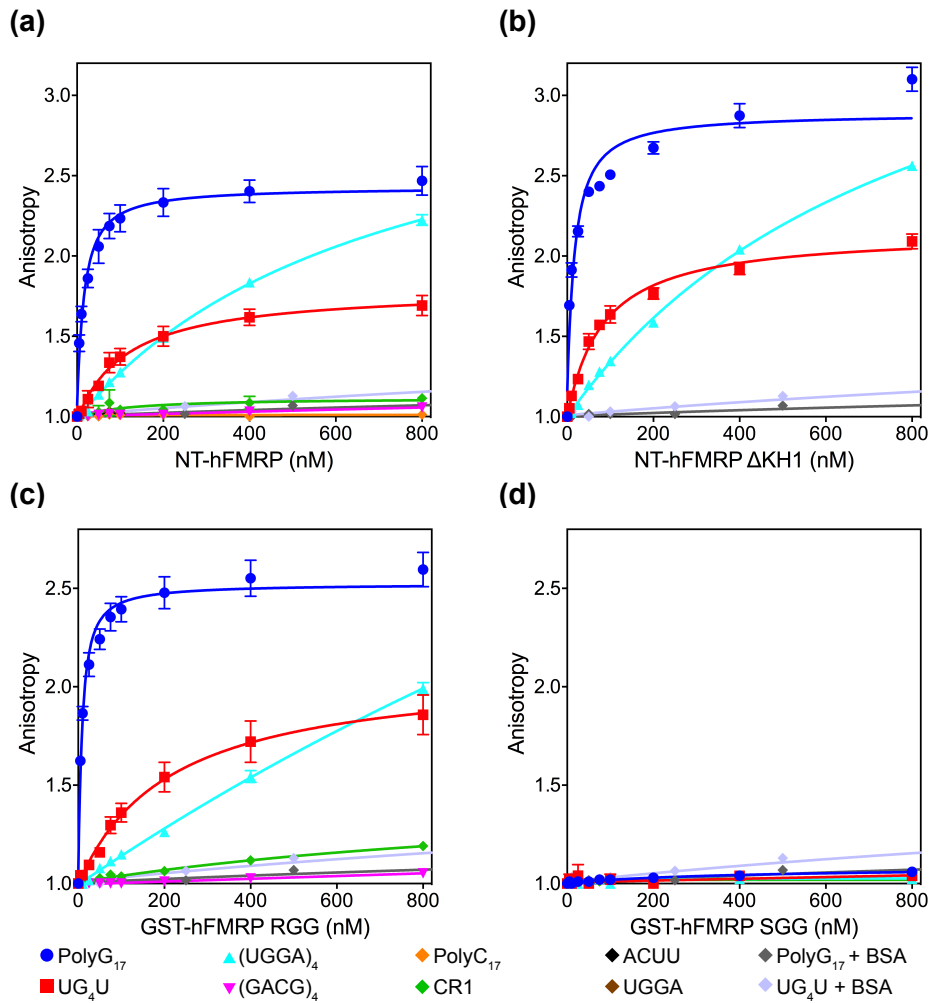


Figure 2.4: FMRP binding to model RNAs. Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled RNAs upon the addition of 0-800 nM of (a) NT-hFMRP, (b) NT-hFMRP Δ KH1, (c) GST-hFMRP RGG, and (d) GST-hFMRP SGG. Bovine serum albumin (BSA) binding to the GQ-assembling RNAs was also measured to test for non-specific protein binding by the two RNAs. Binding trials were performed at least three times, with error bars depicting standard deviation. The symbols indicate the identity of the different RNAs.

Table 2.2. Equilibrium dissociation constant, K_D (nM) of FMRP binding to RNAs

RNA	NT-hFMRP	NT-hFMRP Δ KH1	RGG	SGG
UG ₄ U	110 ± 16	80 ± 7	210 ± 18	No binding
PolyG ₁₇	14 ± 2	17 ± 3	8.6 ± 1.2	No binding
(UGGA) ₄	710 ± 50	840 ± 39	> 4000	No binding
(GACG) ₄	No binding	Not tested	No binding	Not tested
UGGm ⁶ ACU	1250 ± 610	Not tested	Not tested	Not tested

Finally, we tested whether the FMRP RGG motif is binding to the GQ assemblies of PolyG₁₇ and UG₄U, or unfolded forms of the RNAs. To assess specific binding to GQ assemblies, we compared the binding efficiency of GST-hFMRP RGG to PolyG₁₇ and UG₄U in the presence of 75 mM KCl versus 75 mM LiCl. If FMRP could bind unfolded forms of PolyG₁₇ and UG₄U, then we would observe increased binding efficiency in LiCl, where both RNAs favor unfolded states. Conversely, if FMRP specifically binds the folded form, then we would observe decreased binding efficiency due to a drop in the folded population of RNA. Indeed, substituting Li⁺ in place of K⁺ decreased the net anisotropy change for PolyG₁₇, reflecting a smaller population of the RNA binding to FMRP (**Figure 2.5a**). Additionally, GST-hFMRP RGG bound to PolyG₁₇ with a 50-fold lower affinity in the presence of Li⁺ ($K_D = 480 \pm 62$ nM), consistent with a recent report⁴⁰. In the case of UG₄U, we observed only a slight decrease in anisotropy with Li⁺ (**Figure 2.5b**). This may be because Li⁺ does not significantly destabilize the UG₄U GQ assembly. While we see some destabilization by Li⁺ in our urea-PAGE, the urea may be catalyzing it. In both RNA cases, binding was not completely depleted in Li⁺. This suggests the cation substitution did not destabilize the protein. Instead, the protein was tightly binding to the remaining population of stable PolyG₁₇ and UG₄U GQ assemblies. The concomitant drop in the extent of the anisotropy shift with the drop in the GQ population supports the conclusion that the RGG motif specifically binds to RNA GQ structures.

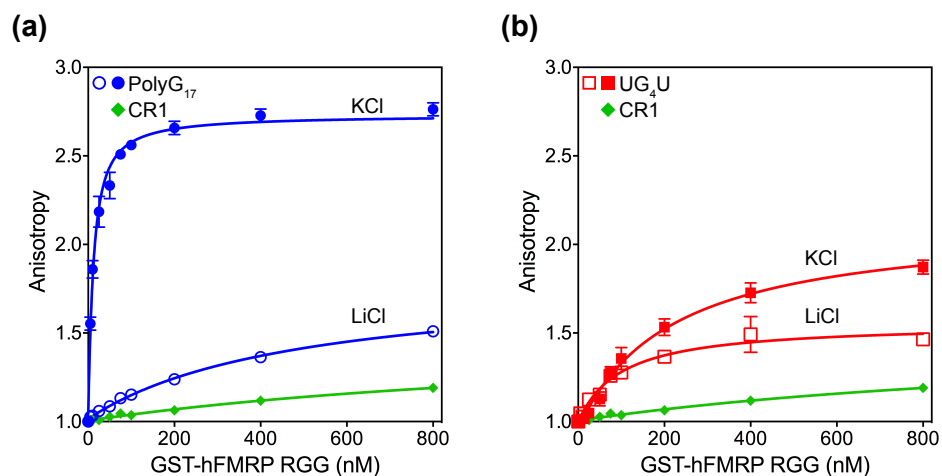


Figure 2.5: FMRP RGG domain binding to GQ RNAs is dependent on K⁺. (a) Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled PolyG₁₇ upon addition of 0-800 nM GST-hFMRP RGG in 75 mM KCl (solid blue circle) and 75 mM LiCl (open blue circle) alongside control RNA CR1. (b) Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled UG₄U upon addition of 0-800 nM GST-hFMRP RGG in 75 mM KCl (solid red square) and 75 mM LiCl (open red square) alongside control RNA CR1. The standard deviations from two experiments are shown.

2.6 The KH domains of FMRP do not bind to the WGGA and GACR motifs

Our studies showed that NT-hFMRP binds with low affinity to (UGGA)₄ RNA and does not bind to (GACG)₄ RNA. Furthermore, we showed that (UGGA)₄ forms a GQ structure suggesting that the RGG motif mediates the interaction of NT-hFMRP with (UGGA)₄. To confirm that the KH domains of FMRP do not bind to (UGGA)₄ and (GACG)₄ RNAs, we performed binding studies with the KH domains of FMRP fused to the GST tag (**Figure 2.1**). GST-hFMRP KH0, GST-hFMRP KH1, GST-hFMRP KH2, GST-hFMRP KH1-KH2, and GST-hFMRP KH0-KH1-KH2 did not bind to (UGGA)₄ and (GACG)₄ RNAs demonstrating that the KH domains of FMRP do not bind to the WGGA and GACR motifs (**Figure 2.6**).

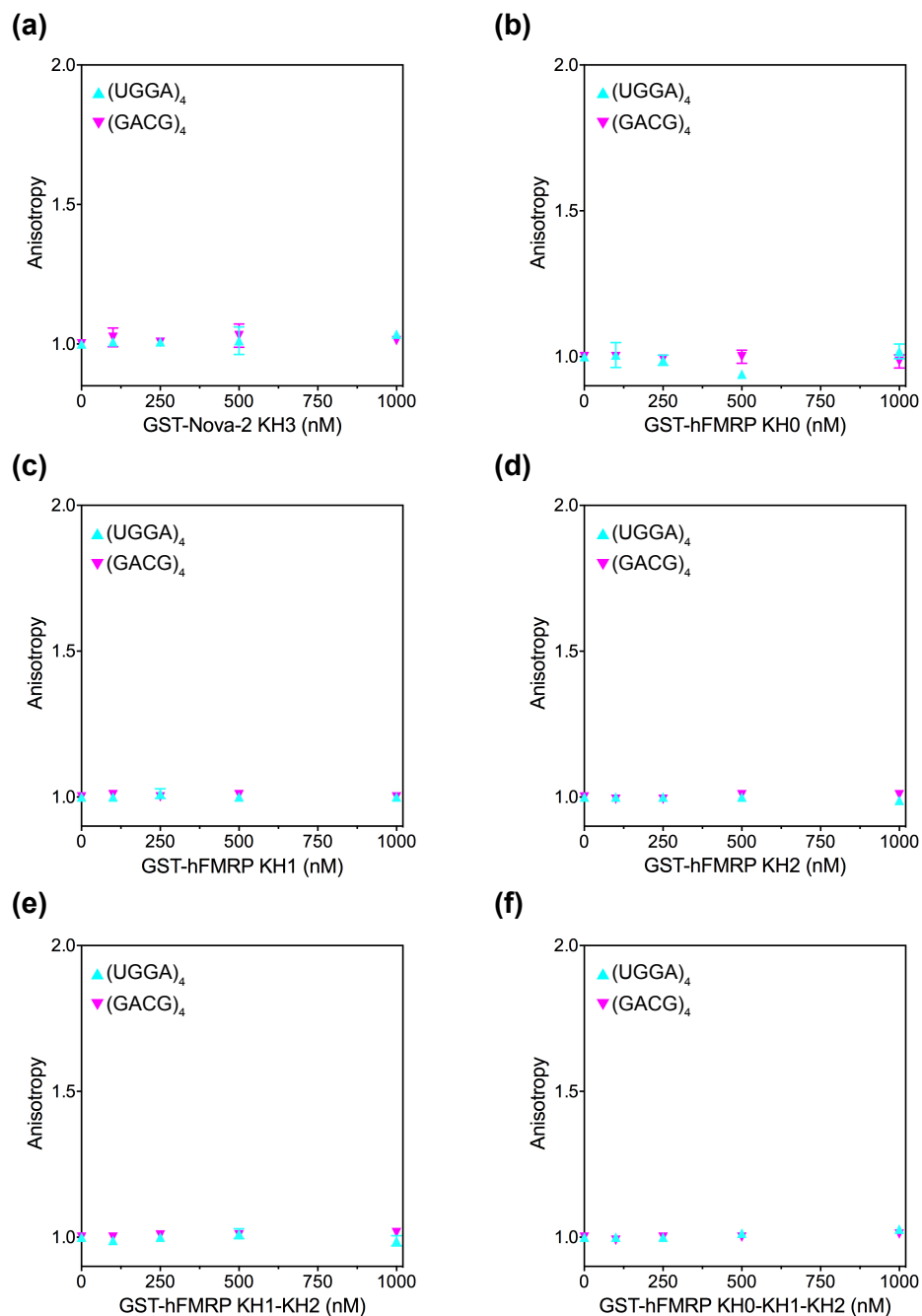


Figure 2.6: KH domains binding to model RNAs (UGGA)₄ and (GACG)₄. Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled (UGGA)₄ and (GACG)₄ upon addition of (a) 0-1000 nM GST-Nova-2 KH3, (b) GST-hFMRP KH0, (c) GST-hFMRP KH1, (d) GST-hFMRP KH2, (e) GST-hFMRP KH1-KH2, and (f) GST-hFMRP KH0-KH1-KH2. The standard deviations from two experiments are shown.

2.7 FMRP does not bind with high affinity to m⁶A RNA

Recent studies indicated that FMRP preferentially binds to mRNAs with the N⁶-methyladenosine (m⁶A) modification⁵⁹⁻⁶¹. We, therefore, tested the binding of NT-hFMRP to the UGGm⁶ACU model RNA having a single m⁶A modification in the middle of the sequence. UGGACU is the consensus sequence for m⁶A modification in mRNAs and is sufficient for binding to YTH-domain proteins⁶²⁻⁶⁴. Surprisingly, our results show that NT-hFMRP binds poorly to the m⁶A modified RNA suggesting that at best, FMRP can only form a low-affinity complex with m⁶A-containing mRNAs **(Figure 2.7)**.

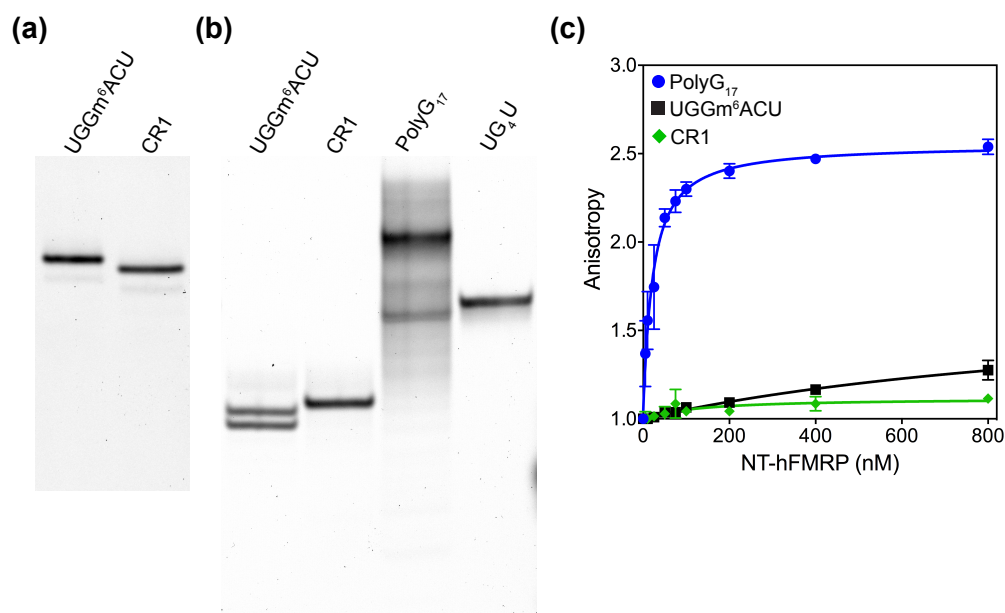


Figure 2.7: NT-hFMRP binding to UGGm⁶ACU RNA. (a) Urea PAGE of UGGm⁶ACU RNA alongside CR1. (b) Native PAGE gel of UGGm⁶ACU RNA alongside linear marker CR1 and GQ RNAs. (c) Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled UGGm⁶ACU, PolyG₁₇, and CR1 upon addition of 0-800 nM NT-hFMRP. The standard deviations from two experiments are shown.

2.8 The KH domains of FMRP do not bind to short ssRNA motifs

The above results indicated that the KH domains of FMRP do not bind to the previously identified RREs. Biochemical and structural studies have shown that KH domains generally recognize 4 to 5 nucleotides in ssRNAs^{16,17,37-39}. For example, the KH3 domain of Nova-2 protein binds to the UCACC sequence^{37,38}. Structural studies showed that the UCACC binds to a cleft formed by the invariant and variable loops of Nova-2 KH3. Similarly, the KH3 and KH4 domains of ZBP1 bind to ACAC and CGGAC sequences, respectively³⁹. Thus, KH domains typically recognize short, ssRNA sequences.

The realization that KH domains recognize 4 to 5 nucleotides motivated us to develop a method called MIDAS (Motif Identification by Analysis of Simple sequences) to determine the RNA-binding specificity of the FMRP KH domains (**Figure 2.8**). We synthesized a pool of RNA molecules having 5 nucleotides randomized in the middle and flanked by 8 nucleotides of fixed sequences in the 5' and 3' ends (the pool will have $4^5 = 1024$ unique sequences). The sequence of the fixed nucleotides was explicitly designed to avoid any of the known KH domain binding RREs. The RNA pool was labeled at the 5'-end with ^{32}P and incubated with 10 μM protein having one or more KH domain. The RNA•protein complex was separated from free RNA using a native gel (**Figure 2.8b**). The band corresponding to the RNA•protein complex was excised, the RNA purified, and the sequences of the purified RNA molecules were identified by next-generation sequencing.

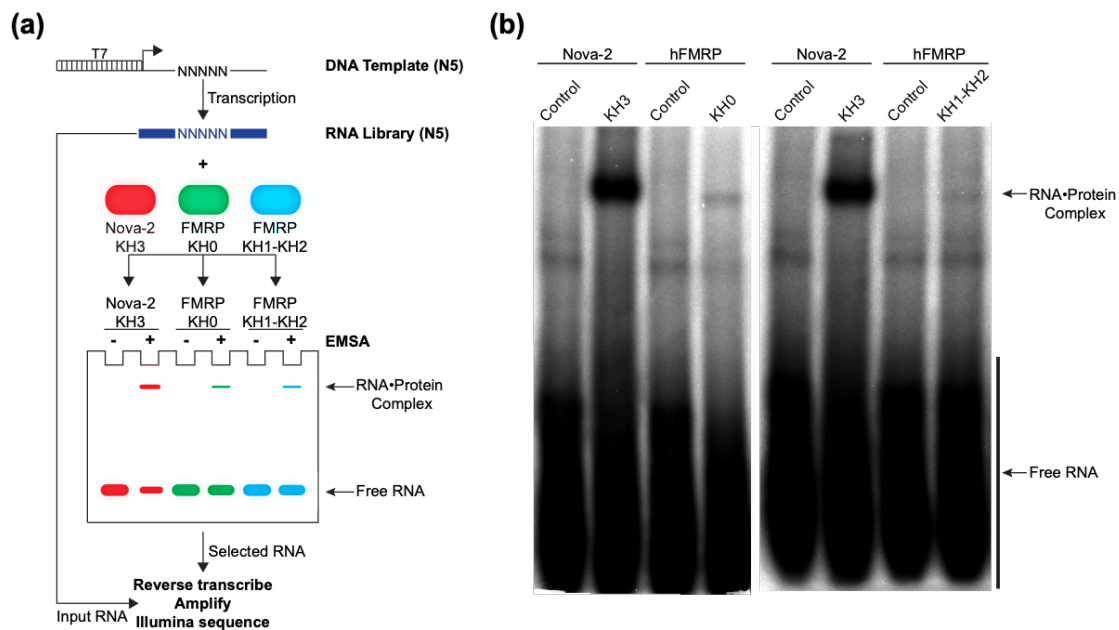


Figure 2.8: Illustration of MIDAS method. (a) Cartoon diagram of the MIDAS process. Purified GST-Nova-2 KH3, GST-hFMRP KH0, and GST-hFMRP KH1-KH2 are individually incubated with the RNA library containing a central N5 sequence. The RNAs bound by each protein are separated by an electrophoretic mobility shift assay (EMSA) and purified from the gel. The input RNA sequences and the selected RNA sequences are then reverse transcribed, amplified, and Illumina sequenced. (b) Autoradiograph of EMSA. Contrast-enhanced image showing the free RNA library migrates as an array (black bar) and the RNAs bound by Nova-2 KH 3, FMRP KH0, and FMRP KH1-KH2 migrate similar distances as distinct bands. Control lanes are the MIDAS N5 RNA library incubated without protein.

To validate the MIDAS method, we used the well-studied Nova-2 KH3 domain^{37,38}. As shown in Figure 4b, Nova-2 KH3 binds to some of the RNA pool molecules resulting in the slower migrating RNA•protein complex band in the native gel. Surprisingly, we observed very faint RNA•protein complex bands with both FMRP KH0 and FMRP KH1-KH2, suggesting very weak binding to the RNA pool molecules. Additionally, we observed no binding to the RNA pool with the individual FMRP KH1 and FMRP KH2 domains (data not shown). We analyzed the RNA sequences that bind to Nova-2 KH3. Interestingly, the highly enriched sequences are UUUCA, UUUA, UUCAC, UUCAU, and UCAAA (**Table 2.3**). The previously studied UCACC sequence was ranked much lower than the highly enriched sequences^{37,38}. To further validate the MIDAS method, we synthesized RNA molecules corresponding to the highly enriched sequences and performed equilibrium binding studies (**Figure 2.9a**). We find excellent agreement between the enrichment factor and the K_D for Nova-2 KH3 binding to the selected RNA molecules, which shows that the MIDAS method is capable of identifying authentic single-stranded RREs (**Table 2.3**).

Analysis of the RNA molecules that bind to the FMRP KH0 and FMRP KH1-KH2 domains indicated that the enriched sequences are UCACC and CCACC. However, the overall enrichment factors for FMRP KH0 and FMRP KH1-KH2 were significantly lower compared to Nova-2 KH3 (**Table 2.3**). Binding studies showed that FMRP KH0 and KH1-KH2 does not bind to UCACC at the standard concentration range used in our anisotropy binding assay (**Figure 2.9b and c**). It was not feasible to test higher concentrations of proteins with this assay. Therefore, we used an electrophoretic mobility shift assay (EMSA) to estimate the binding

affinity of FMRP KH0 and KH1-KH2 for UCACC. The EMSA indeed showed that hFMRP KH0 and KH1-KH2 bind to UCACC, but the affinity is weaker than that of Nova-2 KH3 binding to UCACC ($K_D = 3 \mu\text{M}$) (**Figure 2.10**). This suggests that the MIDAS method can identify even low-affinity RNA ligands.

To test whether the weak binding may be the result of having a single RRE, we synthesized RNAs having two UCACC sequences separated by 5, 10, or 15 nucleotides. The distance between the UCACC sequence was varied because we do not know what the optimal spacing is for the multiple KH domains in FMRP to bind to the RNA. Additionally, we used the GST-hFMRP KH0-KH1-KH2 construct to evaluate potentially synergistic binding when all three KH domains of hFMRP are present. Binding studies showed that hFMRP KH0-KH1-KH2 does not bind to the RNAs having the two UCACC (**Figure 2.9d**). Our results indicate that the KH domains of FMRP do not bind to simple RREs, as proposed by previous studies. Instead, the KH domains of FMRP have likely evolved to bind to a more complex RNA or RNA-protein structure that is not represented in our MIDAS RNA pool. Indeed, FMRP is known to interact with the ribosome, and the KH domains may bind to a unique structure in the ribosome^{33,43,57,65-70}.

Table 2.3. MIDAS enrichment factor (EF) and binding affinity for RNAs

	RNA	EF ^a	<i>K</i> _D (nM)
Nova-2 KH3	UUUCA	17.3	179 ± 23
	UUCAC	14.7	739 ± 82
	UUUAA	17.5	1038 ± 120
	UCACC	2.8	2991 ± 450
	CCACC	1.1	6991 ± 1848
FMRP KH0	UCACC	0.9	> 10,000 ^b
	CCACC	0.6	> 10,000 ^b
	GACUC	0.2	> 10,000 ^b
	UUCAC	0.2	> 10,000 ^b
	UUUCA	0.1	> 10,000 ^b
FMRP KH1-KH2	UCACC	0.5	> 10,000 ^b
	CCACC	0.3	> 10,000 ^b
	GACUC	0.3	> 10,000 ^b
	UUCAC	0.3	> 10,000 ^b
	UUUCA	0.1	> 10,000 ^b

^a Enrichment factors were normalized to the relative abundance of selected RNAs from the EMSA gel

^b Estimated affinity based on EMSA selection using 10 μM FMRP KH domains

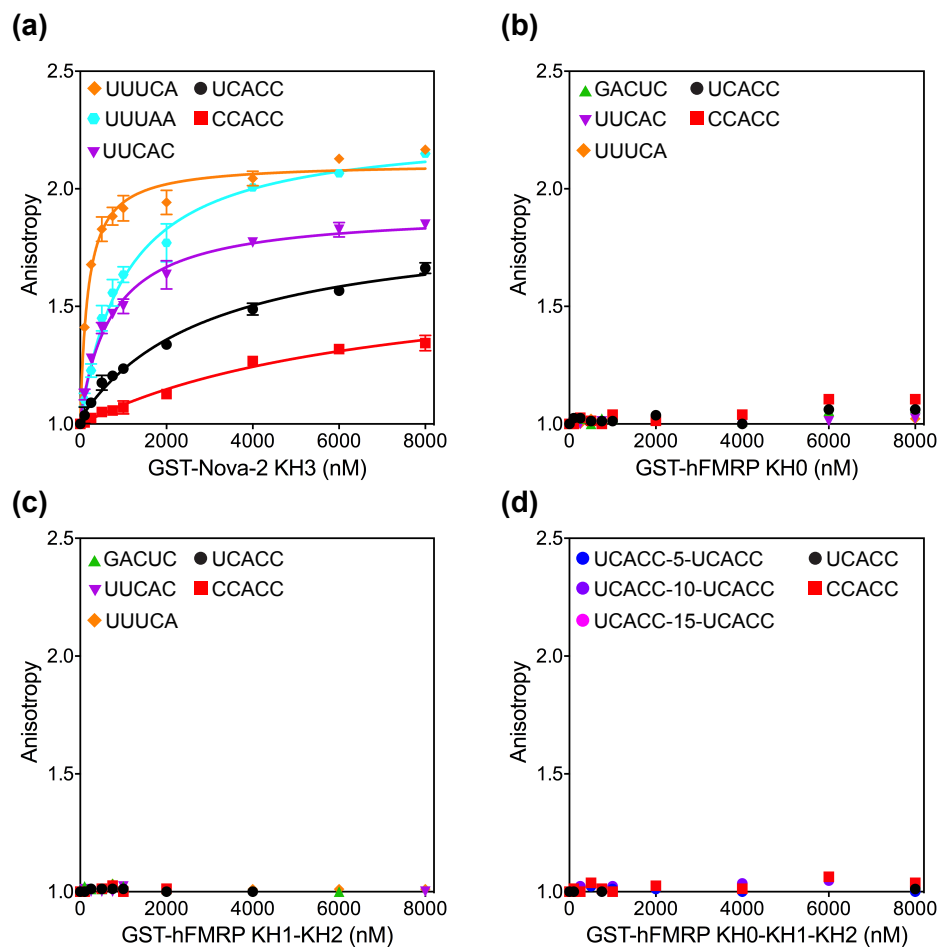


Figure 2.9: KH domains binding to RNAs identified by MIDAS. Fold increase in fluorescence anisotropy of 5 nM fluorescein-labeled RNAs upon the addition of 0-8000 nM of (a) GST-Nova-2 KH3, (b) GST-hFMRP KH0, (c) GST-hFMRP KH1-KH2, and (d) GST-hFMRP KH0-KH1-KH2. Binding trials were performed at least three times, with error bars depicting standard deviation. The symbols indicate the identity of the different RNAs tested for each protein construct.

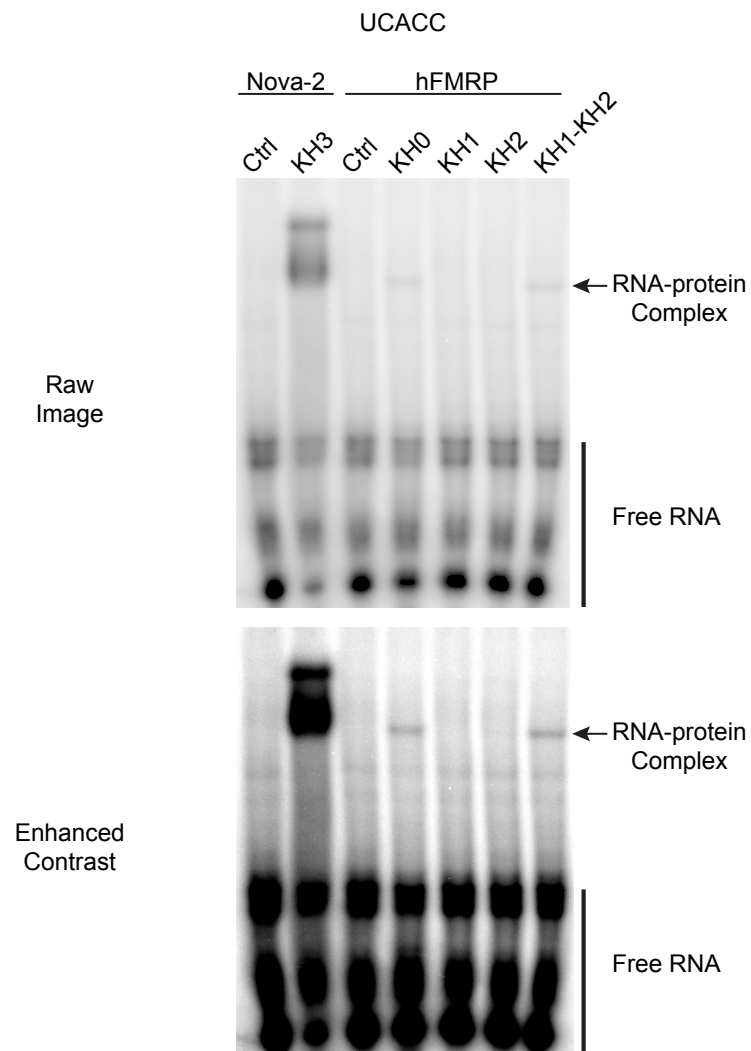


Figure 2.10: KH domains binding to MIDAS RNA UCACC. EMSA of radiolabeled MIDAS RNA UCACC incubated with 10 μ M Nova-2 KH3, hFMRP KH0, hFMRP KH1, hFMRP KH2, and hFMRP KH1-KH2 domains.

2.9 Discussion

FMRP has been estimated to regulate the translation of nearly 4% of the mRNAs in the human fetal brain¹². However, identifying the RRE of FMRP in these mRNAs has remained challenging. Early RNA in vitro selection studies showed that the KH2 motif binds to kissing complex RNAs, whereas the RGG motif binds to a GQ-forming RNA^{22,23}. In addition, high-throughput approaches have been used to determine the RREs of FMRP. For example, the WGGA and the ACUK motifs were shown to be enriched in FMRP mRNA targets by the PAR-CLIP method³¹. The WGGA and ACUK motifs were proposed to bind to the KH1 and KH2 domains of FMRP, respectively. In contrast to the above study, the GACR motif was identified using the KH1 and KH2 domains of FMRP by the in vitro technique called RNAcompete³⁵. Further in-depth analysis of the FMRP target datasets indicated that the GACR motif is highly enriched, the WGGA motif is modestly enriched, and the ACUK motif is not enriched³⁶. Moreover, the WGGA motifs are clustered together in the mRNA targets suggesting that they are enriched in the FMRP dataset because they form GQ structures³⁶.

In the PAR-CLIP method, 4-thiouridine is incorporated into nascent mRNA transcripts to improve the efficiency of cross-linking proteins that bind to mRNAs. However, as noted previously, this may bias the selection of mRNA sequences having more uridine residues, which may explain the enrichment of the ACUK motif in the FMRP target dataset³⁶. Consistent with this interpretation, our studies show that FMRP does not bind to an RNA having a single ACUU sequence. Furthermore, the

PAR-CLIP method used full-length FMRP to identify mRNA targets, which may bias the selection of mRNAs having GQ structures because of the presence of the RGG domain. Consistent with this explanation, our results show that an RNA with four repeats of the WGGA motif forms a GQ structure and binds to the RGG domain but not to the KH domains of hFMRP.

Surprisingly, our studies show that FMRP does not bind to (GACG)₄, which was identified by the RNAcompete method and appears to be highly enriched in the FMRP target datasets^{35,36}. The discrepancy between our data and the previous RNAcompete data may be explained by the size of the FMRP protein or the length of the RNA used in both studies. We used NT-hFMRP and several deletion constructs to determine K_D . In contrast, the RNAcompete method used a KH1-KH2 fragment, which may not have the same structure and the RNA-binding specificity as the larger protein fragment. Alternatively, we used 18 nucleotide RNA for our binding studies, whereas the RNAcompete used 30 - 41 nucleotides RNAs, which may fold into more complex structures that could bind to the KH1-KH2 fragment.

Interestingly, we show that FMRP has the highest binding affinity for PolyG₁₇ and UG₄U RNAs, which form stable GQ structures. Furthermore, the RGG domain of FMRP is sufficient for binding to PolyG₁₇ and UG₄U. These results are consistent with previous studies that showed the RGG domain of FMRP binds specifically to GQ structures^{23-30,40,58}. Taken together, our results show that the KH0, KH1, and KH2 domains do not bind to the WGGA, ACUK, and GACR motifs, whereas the RGG domain binds with high affinity to RNA GQ structures. We developed the MIDAS method to identify in an unbiased manner single-stranded RREs that could bind to

the KH0, KH1, and KH2 domains of hFMRP. Although the MIDAS method was able to identify RREs that bind with high affinity to the Nova-2 KH3 domain, we were only able to identify RREs that bind with weak affinity to the KH0 and KH1-KH2 domains of FMRP. The inability to identify any RREs that bind to the KH domains of hFMRP using the MIDAS method is consistent with our data showing that they do not bind to model RNAs having the previously identified RREs. One possibility for our unexpected results is that the phosphorylation of hFMRP at Ser500 is essential for the KH domains to bind to single-stranded RNAs with high affinity^{71,72}. Although Ser500 is not located within the KH domains, FMRP phosphorylation may induce some structural rearrangement in the protein that increases the binding affinity of the KH domains for single-stranded RNAs⁷¹.

A second possibility is that the KH domains of hFMRP recognize more complex RNA motifs than represented in our randomized RNA pool. Indeed, Darnell and co-workers showed that an in vitro selection experiment performed with an RNA pool having 25 randomized nucleotides did not yield any RNA binders for the hFMRP KH1-KH2 domains²². They obtained RNA binders only when they used an RNA pool with 52 randomized nucleotides. The in vitro selected RNA binders formed a kissing complex RNA motif, which binds to the KH2 domain of FMRP, most likely in a different mode than simple RREs binding to the canonical KH domains. These results suggest that the KH domains of FMRP are more specialized than the canonical KH domains that recognize 4-5 nucleotides in single-stranded RNA. The evolutionary changes in the sequence of the KH domains of FMRP may inhibit their

binding to simple single-stranded RNA sequences so that FMRP may bind to a unique structure such as the ribosome^{33,43,57,65-70}.

Materials and Methods

2.10 Protein expression and purification

The human FMR1 isoform 1 gene was assembled from *E. coli* codon-optimized gene blocks (IDT). The gene encoding the full-length isoform 1 human FMRP (hFMRP) spanning residues E2-P632 was subcloned into the LIC expression vector pMCSG7 (DNASU plasmid repository) with the 5' TEV cleavage site deleted, conferring a N-terminal hexahistidine tag. The N-terminus truncated human FMRP (NT-hFMRP) spanning residues R218-P632 was subcloned into the LIC expression vector pMCSG7, conferring a N-terminal hexahistidine tag with the 5' TEV cleavage site deleted. The NT-hFMRP Δ KH1 construct spanning residues F281-P362 was subsequently generated from NT-hFMRP via PCR deletion of the KH1 domain. The human FMRP RGG domain (GST-hFMRP RGG) spanning residues G531-P632 was subcloned into the LIC expression vector pMCSG10 (DNASU) which confers a N-terminal hexahistidine-GST fusion tag that is cleavable using TEV protease. The mutant SGG domain (GST-hFMRP SGG) was assembled from an *E. coli* codon-optimized gene block where all 16 arginine residues of the RGG region were mutated to serines. The resulting SGG fragment was subcloned into the LIC expression vector pMCSG10. The human Nova-2 KH3 domain (GST-Nova-2 KH3) spanning

residues K406-G492 was subcloned into the LIC expression vector pMCSG10. The human FMRP KH0 domain (GST-hFMRP KH0) spanning residues F126-R201, human FMRP KH1 domain (GST-hFMRP KH1) spanning residues A216-E280, human FMRP KH2 domain (GST-hFMRP KH2) spanning residues F281-K425, human FMRP tandem KH1-KH2 domains (GST-hFMRP KH1-KH2) spanning residues A216-K425, and human FMRP tandem KH0-KH1-KH2 domains (GST-hFMRP KH0-KH1-KH2) spanning residues F126-K425 were also subcloned into the LIC expression vector pMCSG10.

The NT-hFMRP, NT-hFMRP Δ KH1, GST-hFMRP RGG, GST-hFMRP SGG, GST-Nova-2 KH3, and the five GST-hFMRP KH construct expression plasmids were all transformed in *E. coli* BL21(DE3) cells (Novagen). NT-hFMRP and NT-hFMRP Δ KH1 were purified using Ni-NTA affinity chromatography (Qiagen). GST-hFMRP RGG, GST-hFMRP SGG, and the five GST-hFMRP KH constructs were purified using glutathione affinity chromatography (GE Healthcare). All proteins were further purified using Superdex 75 16/60 or Superdex 200 16/60 (GE Healthcare) gel filtration chromatography and stored in their respective running buffers. Gel filtration buffer composition varied between FMRP constructs. NT-hFMRP and NT-hFMRP Δ KH1 gel filtration running buffer contained 25 mM Tris pH 7.4, 150 mM KCl and 1 mM DTT. GST-hFMRP RGG and GST-hFMRP SGG gel filtration running buffer contained 50 mM Tris pH 7.5 and 1 mM DTT. GST-Nova-2 KH3 gel filtration running buffer contained 10 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl₂ and 1 mM DTT. All five GST-hFMRP KH domain constructs were purified in gel filtration running buffer containing 50 mM Tris pH 7.4, 150 mM NaCl and 1 mM DTT.

2.11 Purification of RNAs

The PolyG₁₇, UG₄U, CR1, PolyC₁₇, PolyA₁₇, ACUU, UGGA, (GACG)₄, (UGGA)₄ and UGGm⁶ACU RNAs were purchased from Thermo Scientific/Dharmacon with a fluorescein tag at the 3' end. For the NMM assay, PolyG₁₇ and UG₄U were also purchased without a fluorescein tag, and (UGGA)₄ was in vitro transcribed from annealed oligonucleotide template using T7 RNA polymerase. Each RNA was purified under standard denaturing conditions by urea polyacrylamide gel electrophoresis (PAGE) and then extracted from the gel.

The RNA in solution was purified by chloroform extractions followed by ethanol precipitation. The purified RNAs were dissolved in RNase-free water. The RNA concentration was measured by absorbance at 260 nm using a spectrophotometer.

2.12 PAGE analysis of labeled RNAs

Denaturing urea PAGE analysis of the UG₄U RNA was performed as described previously⁴⁹. 20 pmol UG₄U was reconstituted in 50 mM Tris pH 7.8 and either 100 mM KCl, 100 mM NaCl or 100 mM LiCl. The UG₄U samples were heated at 37°C for 90 minutes, then incubated at 25°C for 5 minutes in a thermal cycler. Then formamide loading solution was added to each RNA sample to a final 8 µL volume. The samples were heated at 60°C for 5 minutes and then incubated at 25°C for 5 minutes in a thermal cycler. Finally, the samples were run on a 16% urea polyacrylamide gel at room temperature for 4 hours at 600 V.

For denaturing urea PAGE analysis of the 18-mer model RNAs, 10 pmol PolyG₁₇, 3 pmol PolyC₁₇, 2 pmol ACUU, 2 pmol UGGA, 6 pmol CR1, 4 pmol PolyA₁₇, 6 pmol (GACG)₄, and 10 pmol (UGGA)₄ in 20 mM Tris-HCl pH 7.45 and 100 mM KCl were run on a 15% urea polyacrylamide gel at room temperature for 2 hours at 25 W.

For native PAGE analysis of all the model RNAs, 10 pmol M1 5' UTR, 50 pmol UG₄U, 10 pmol PolyG₁₇, 3 pmol PolyC₁₇, 2 pmol ACUU, 2 pmol UGGA, 6 pmol CR1, 4 pmol PolyA₁₇, 6 pmol (GACG)₄, and 10 pmol (UGGA)₄ in 20 mM Tris-HCl pH 7.45, 100 mM KCl and 5% (v/v) glycerol were run on a 15% native polyacrylamide gel containing 1X TBE. The gel was run at 4°C for 4 hours at 11 Watts.

2.13 BG4 antibody assay

Antibody BG4 scFv was purchased from Millipore Sigma. 100 nM BG4 was mixed with 5 nM fluorescein-labeled RNA by diluting BG4 into fluorescence anisotropy buffer (20 mM Tris pH 7.7, 75 mM KCl, 5 mM MgCl₂, and 100 ng/μL total tRNA from *E. coli*) followed by mixing with fluorescence anisotropy buffer supplemented with 5 nM fluorescein-labeled RNA directly in the well of a non-binding 96-well black flat bottom plate (Greiner) in a final 200 μL volume. The plate was incubated at room temperature for 30 minutes. Samples were excited at 470 nm and emission was measured at 520 nm with a 20 nm bandwidth; optimal signal gain was determined per read. Fluorescence anisotropy was measured on a Tecan Safire 2 plate reader in polarization mode. All experiments were performed a minimum of 2 or 3 times to determine the standard deviations.

2.14 NMM assay

N-methyl mesoporphyrin IX (NMM) was purchased from Frontier Scientific. NMM stock solution was prepared by dissolving the stock in 0.2 N HCl to 5 mg/mL (equal to 8.61 μ M). Fresh working solution was prepared by diluting the NMM stock solution to 400 μ M in 10% (v/v) DMSO. Various concentrations of the RNAs were reconstituted in 158 μ L solution containing 20 mM Tris pH 7.7, 5 mM $MgCl_2$, 1 mM DTT, and either 100 mM KCl or 100 mM LiCl. The samples were then heated at 90°C for 2 minutes and slowly cooled to room temperature over 1 hour. Samples were briefly spun down and collected before adding 2 μ L of 400 μ M NMM to each 158 μ L sample, to a final 5 μ M NMM per sample. The samples were incubated at room temperature for 10 minutes before taking fluorescence measurements (JASCO FP-8500). Samples were excited at 400 nm and emission was scanned from 560 nm to 650 nm with a 5 nm excitation and emission bandwidths; automatic signal gain was used. The peak NMM fluorescence (610 nm) was plotted as a function of RNA concentration. All experiments were performed a minimum of 2 or 3 times to determine the standard deviations.

2.15 Fluorescence anisotropy RNA binding assay

The various FMRP constructs were titrated against 5 nM fluorescein-labeled RNAs. Fluorescence anisotropy was measured on a non-binding 96-well black flat bottom plate (Greiner) using a Tecan Safire 2 plate reader in polarization mode.

Some of the RNA fluorescence anisotropy measurements were also performed using a Tecan Spark plate reader.

FMRP was diluted into fluorescence anisotropy buffer supplemented with DTT (20 mM Tris pH 7.7, 75 mM KCl, 5 mM MgCl₂, 100 ng/μL total tRNA from *E. coli*, and 1 mM DTT) and mixed with fluorescence anisotropy buffer supplemented with fluorescein-labeled RNA within the plate wells in a final 200 μL volume. The plate was incubated at room temperature for 30 minutes. Samples were excited at 470 nm and emission was measured at 520 nm with a 20 nm bandwidth; optimal signal gain was determined per read.

To quantify the binding affinity between FMRP and RNA, the anisotropy data from each binding assay was normalized to the initial value without protein, plotted and fit to the quadratic equation below to calculate the equilibrium dissociation constant (K_D) as described previously⁷³:

$$\frac{[P + RNA]}{[RNA]} = \frac{([P] + [RNA] + K_D) - \sqrt{([P] + [RNA] + K_D)^2 - 4[P][RNA]}}{2[P]}$$

Where $[P+RNA]/[RNA]$ is the anisotropy value, $[RNA]$ is the RNA concentration, and $[P]$ is the protein concentration. GraphPad Prism software (Graphpad Software Inc.) was used to perform the curve fits. All experiments were performed a minimum of 2 or 3 times to determine the standard deviations.

2.16 MIDAS DNA template and primers

The MIDAS N5 DNA template sequence: 5'-mAmATGCGTANNNNTGGATCCCTATAGTGAGTCGTATTA-3' was purchased from IDT and has two 2'-O-methyl substitutions (mA) at the 5' terminus to reduce the amount of n+1 product during in vitro transcription. The five randomized positions were produced by hand mixing the nucleotides to give equal ratio of all four nucleotides.

18T7T sequence:

5'-TAATACGACTCACTATAG-3'.

First-strand primer:

5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG AATGCGTA-3'. Second-strand primer:

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGATCCA-3'. Index 1

Read primer:

5'-CAAGCAGAAGACGGCATACGAGAT [i7] GTCTCGTGGGCTCGG-3'. Index 2

Read primer:

5'-AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGTC-3'.

2.17 MIDAS: Separation of RNA•protein complex using a native gel

The MIDAS RNAs were generated by in vitro transcription of annealed 18T7T and MIDAS template DNAs using T7 RNA polymerase. The sequence of the 21-mer MIDAS N5 RNA is 5'-GGGAUCCANNNNNUACGCAUU-3'. The transcription products were purified under standard denaturing conditions by urea polyacrylamide gel electrophoresis (PAGE) and then extracted from the gel. Each RNA was purified by chloroform extractions followed by ethanol precipitation. The purified RNAs were dissolved in RNase-free water. The RNA concentration was measured by absorbance at 260 nm using a spectrophotometer.

The MIDAS N5 RNA pool was 5' dephosphorylated with alkaline phosphatase (NEB) and labeled with [γ -³²P]-ATP and T4 polynucleotide kinase (NEB). An EMSA was used to detect and isolate protein-bound MIDAS RNAs from the pool. Briefly, in a 10 μ L sample, 100,000 cpm of MIDAS N5 RNA was incubated with 10 μ M protein (or protein storage buffer in control lane) in RNA binding buffer (50 mM Tris acetate pH 7.7, 50 mM potassium acetate and 5 mM magnesium acetate, 1 mM DTT, 0.1 mg/mL BSA, and 100 ng/ μ L total tRNA from *E. coli*) for 45 minutes at room temperature. Glycerol was added to 5% final concentration and the samples were run on an 8% (79:1 acrylamide:bisacrylamide) 0.5X TBE native polyacrylamide gel at 4°C for 2 hours and 30 minutes at a constant 300 V. The gel was exposed to a phosphor screen at -80°C and imaged on a Typhoon FLA 9500 phosphorimager (GE) after 18 hours. The gel was thawed at room temperature and the shifted band for each protein was extracted from the gel. Each selected RNA was purified by

chloroform extractions followed by ethanol precipitation. The purified RNAs were dissolved in 35 μ L RNase-free water.

2.18 MIDAS: Library preparation

The selected MIDAS RNAs were used to generate a cDNA library for next generation sequencing. This was accomplished in three simple steps: (1) synthesize the first-strand cDNA by reverse transcription, (2) synthesize the second-strand by primer extension, and (3) amplify the dsDNA by PCR. First, 8 μ L of the selected RNA and 2 μ L of 10 μ M MIDAS reverse transcription primer were combined and annealed by heating at 65°C for 5 minutes and cooling to room temperature for 2 minutes before placing on ice. The 10 μ L annealed RNA-primer was mixed with 10 μ L 2X cDNA synthesis mix containing 2X FS buffer (100 mM Tris pH 8.3, 150 mM KCl, and 6 mM MgCl₂), 1 mM dNTPs, 10 mM DTT, and 10 U Superscript III reverse transcriptase (ThermoFisher). The reverse transcription was carried out in a thermal cycler: (1) 10 minutes at 25°C, (2) 1 hour and 30 minutes at 30°C, (3) 15 minutes at 70°C, (4) 37°C hold for RNase H digestion. On step 4, the 20 μ L first-strand samples were placed on ice for 1 minute and spun down. Then 1 μ L of 5 U/ μ L RNase H (NEB) was added and the sample was incubated at 37°C for 20 minutes.

To the 21 μ L first-strand cDNA sample, we added: 3.5 μ L of 100 mM NaCl, 6.4 μ L of 100 mM MgCl₂, 10 μ L of 10 mM dNTPs, 10 μ L of 10 μ M MIDAS cDNA Klenow primer, 2 μ L of 5 U/ μ L Klenow fragment 3' $\rightarrow$ 5' exo⁻ (NEB) and water to a final 100 μ L volume. This made an ideal 100 μ L primer extension reaction containing 10 mM

Tris pH 8.2, 15 mM KCl, 35 mM NaCl, 7 mM MgCl₂, and 1 mM DTT. The reaction proceeded for 30 minutes at room temperature and was then deactivated for 20 minutes at 75°C. The second-strand product was then amplified by PCR with sample-specific Index 1 and Index 2 primers containing their unique i7 and i5 index sequences.

PCR was carried out using a KOD HS polymerase system (MilliporeSigma). 50 µL PCRs were carried according to manufacturer protocol using 5 µL of the 100 µL second-strand product. The DNA was amplified over 25 cycles (annealing temperature of 60°C and a 3 second extension time). The DNA library was finally cleaned up using an Amico Ultra-0.5 mL centrifugal filter with a 30,000 Da molecular weight cut-off. The 50 µL PCR product was diluted in 450 µL water and concentrated down to approximately 50 µL a total of three times. The DNA concentration was measured by absorbance at 260 nm using a spectrophotometer. The Nova-2 KH3, hFMRP KH0, hFMRP KH1-KH2, and total input DNA libraries were mixed in equimolar ratios and then sequenced with Illumina HiSeq 4000 at the UCSD IGM Genomics Center.

2.19 MIDAS: NGS data analysis

We obtained more than four million reads for each Illumina sequencing run. To identify and rank the favourite motifs of each protein we used a custom Python script, which first filtered the data by sequence length to show only the N5 region sequences. Doing this revealed exactly 1,024 unique 5-mer sequences. Counts

from each of the 1,024 sequences were compared between the protein-selected libraries and the total input library to generate a list of enrichment factors. To determine if any of the top enriched sequences had a significant number of counts coming from mutated sequences, the data were also filtered by sequence length to show the N5 region plus 3 bases upstream and 3 bases downstream. This revealed how many counts for each N5 motif came from the wild type sequence and how many came from mutated sequences. More than 99% of the reads were the correct wild type sequence and only counts coming from the wild type sequences were used to calculate the enrichment factor. Identical sequencing data were obtained when the Illumina sequencing was repeated with another set of independent library preparations showing that the process is reproducible.

Chapter 2, in full, is a reprint of the material as it appears in RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein, Journal of Molecular Biology, 2020. Athar, Youssi M.; Joseph, Simpson, Elsevier, 2020. The dissertation author was the primary investigator and author of this paper.

References

1. M.R. Santoro, S.M. Bray, S.T. Warren, Molecular mechanisms of fragile X syndrome: a twenty-year perspective, *Annu Rev Pathol.* 7 (2012) 219–45. <https://doi.org/10.1146/annurev-pathol-011811-132457>.
2. B. Coffee, F. Zhang, S.T. Warren, D. Reines, Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells, *Nat Genet.* 22 (1999) 98–101. <https://doi.org/10.1038/8807>.
3. D.L. Nelson, H.T. Orr, S.T. Warren, The unstable repeats--three evolving faces of neurological disease, *Neuron.* 77 (2013) 825–43. <https://doi.org/10.1016/j.neuron.2013.02.022>.
4. J.S. Sutcliffe, D.L. Nelson, F. Zhang, M. Pieretti, C.T. Caskey, D. Saxe, S.T. Warren, DNA methylation represses FMR-1 transcription in fragile X syndrome, *Hum Mol Genet.* 1 (1992) 397–400.
5. A.J. Verkerk, M. Pieretti, J.S. Sutcliffe, Y.H. Fu, D.P. Kuhl, A. Pizzuti, O. Reiner, S. Richards, M.F. Victoria, F.P. Zhang, et al, Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome, *Cell.* 65 (1991) 905–14. [https://doi.org/0092-8674\(91\)90397-H](https://doi.org/0092-8674(91)90397-H) [pii].
6. G.J. Bassell, S.T. Warren, Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function, *Neuron.* 60 (2008) 201–14. [https://doi.org/S0896-6273\(08\)00847-7](https://doi.org/S0896-6273(08)00847-7) [pii] 10.1016/j.neuron.2008.10.004.
7. B. Lagerbauer, D. Ostareck, E.M. Keidel, A. Ostareck-Lederer, U. Fischer, Evidence that fragile X mental retardation protein is a negative regulator of translation, *Hum Mol Genet.* 10 (2001) 329–38.
8. Z. Li, Y. Zhang, L. Ku, K.D. Wilkinson, S.T. Warren, Y. Feng, The fragile X mental retardation protein inhibits translation via interacting with mRNA, *Nucleic Acids Res.* 29 (2001) 2276–83.
9. T.A. Comery, J.B. Harris, P.J. Willems, B.A. Oostra, S.A. Irwin, I.J. Weiler, W.T. Greenough, Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits, *Proc. Natl. Acad. Sci. U. S. A.* 94 (1997) 5401–4.
10. A.W. Grossman, G.M. Aldridge, I.J. Weiler, W.T. Greenough, Local protein synthesis and spine morphogenesis: Fragile X syndrome and beyond, *J Neurosci.* 26 (2006) 7151–5. <https://doi.org/26/27/7151> [pii] 10.1523/JNEUROSCI.1790-06.2006.

11. S.A. Irwin, R. Galvez, W.T. Greenough, Dendritic spine structural anomalies in fragile-X mental retardation syndrome, *Cereb Cortex*. 10 (2000) 1038–44.
12. C.T. Ashley, K.D. Wilkinson, D. Reines, S.T. Warren, FMR1 protein: conserved RNP family domains and selective RNA binding, *Science*. 262 (1993) 563–6.
13. H. Siomi, M.C. Siomi, R.L. Nussbaum, G. Dreyfuss, The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein, *Cell*. 74 (1993) 291–8.
14. H. Siomi, M.J. Matunis, W.M. Michael, G. Dreyfuss, The pre-mRNA binding K protein contains a novel evolutionarily conserved motif, *Nucleic Acids Res*. 21 (1993) 1193–8.
15. R. Valverde, I. Pozdnyakova, T. Kajander, J. Venkatraman, L. Regan, Fragile X mental retardation syndrome: structure of the KH1-KH2 domains of fragile X mental retardation protein, *Structure*. 15 (2007) 1090–8. [https://doi.org/S0969-2126\(07\)00280-8](https://doi.org/S0969-2126(07)00280-8) [pii] 10.1016/j.str.2007.06.022.
16. D. Hollingworth, A.M. Candel, G. Nicastro, S.R. Martin, P. Briata, R. Gherzi, A. Ramos, KH domains with impaired nucleic acid binding as a tool for functional analysis, *Nucleic Acids Res*. 40 (2012) 6873–86. <https://doi.org/gks368> [pii] 10.1093/nar/gks368.
17. G. Nicastro, I.A. Taylor, A. Ramos, KH-RNA interactions: back in the groove, *Curr. Opin. Struct. Biol*. 30 (2015) 63–70. <https://doi.org/10.1016/j.sbi.2015.01.002>.
18. Y. Hu, Z. Chen, Y. Fu, Q. He, L. Jiang, J. Zheng, Y. Gao, P. Mei, Z. Chen, X. Ren, The amino-terminal structure of human fragile X mental retardation protein obtained using precipitant-immobilized imprinted polymers, *Nat Commun*. 6 (2015) 6634. <https://doi.org/10.1038/ncomms7634>.
19. L.K. Myrick, H. Hashimoto, X. Cheng, S.T. Warren, Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain, *Hum Mol Genet*. 24 (2015) 1733–40. <https://doi.org/10.1093/hmg/ddu586>.
20. K. De Boulle, A.J. Verkerk, E. Reyniers, L. Vits, J. Hendrickx, B. Van Roy, F. Van den Bos, E. de Graaff, B.A. Oostra, P.J. Willems, A point mutation in the FMR-1 gene associated with fragile X mental retardation, *Nat Genet*. 3 (1993) 31–5. <https://doi.org/10.1038/ng0193-31>.
21. L.K. Myrick, M. Nakamoto-Kinoshita, N.M. Lindor, S. Kirmani, X. Cheng, S.T. Warren, Fragile X syndrome due to a missense mutation, *Eur J Hum Genet*. (2014). <https://doi.org/10.1038/ejhg.2013.311>.

22. J.C. Darnell, C.E. Fraser, O. Mostovetsky, G. Stefani, T.A. Jones, S.R. Eddy, R.B. Darnell, Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes, *Genes Dev.* 19 (2005) 903–18. <https://doi.org/gad.1276805> [pii] 10.1101/gad.1276805.
23. J.C. Darnell, K.B. Jensen, P. Jin, V. Brown, S.T. Warren, R.B. Darnell, Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function, *Cell.* 107 (2001) 489–99.
24. A.T. Phan, V. Kuryavyi, J.C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, R.B. Darnell, D.J. Patel, Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction, *Nat Struct Mol Biol.* 18 (2011) 796–804. <https://doi.org/nsmb.2064> [pii] 10.1038/nsmb.2064.
25. N. Vasilyev, A. Polonskaia, J.C. Darnell, R.B. Darnell, D.J. Patel, A. Serganov, Crystal structure reveals specific recognition of a G-quadruplex RNA by a beta-turn in the RGG motif of FMRP, *Proc. Natl. Acad. Sci. U. S. A.* 112 (2015) E5391–400. <https://doi.org/10.1073/pnas.1515737112>.
26. A.C. Blice-Baum, M.R. Mihailescu, Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms, *RNA.* 20 (2014) 103–14. <https://doi.org/10.1261/rna.041442.113>.
27. L. Menon, M.R. Mihailescu, Interactions of the G quartet forming semaphorin 3F RNA with the RGG box domain of the fragile X protein family, *Nucleic Acids Res.* 35 (2007) 5379–92. <https://doi.org/10.1093/nar/gkm581>.
28. L. Menon, S.A. Mader, M.R. Mihailescu, Fragile X mental retardation protein interactions with the microtubule associated protein 1B RNA, *RNA.* 14 (2008) 1644–55. <https://doi.org/10.1261/rna.1100708>.
29. S. Stefanovic, B.A. DeMarco, A. Underwood, K.R. Williams, G.J. Bassell, M.R. Mihailescu, Fragile X mental retardation protein interactions with a G quadruplex structure in the 3'-untranslated region of NR2B mRNA, *Mol. Biosyst.* 11 (2015) 3222–3230. <https://doi.org/10.1039/c5mb00423c>.
30. Y. Zhang, C.M. Gaetano, K.R. Williams, G.J. Bassell, M.R. Mihailescu, FMRP interacts with G-quadruplex structures in the 3'-UTR of its dendritic target Shank1 mRNA, *RNA Biol.* 11 (2014) 1364–74. <https://doi.org/10.1080/15476286.2014.996464>.
31. M. Ascano, N. Mukherjee, P. Bandaru, J.B. Miller, J.D. Nusbaum, D.L. Corcoran, C. Langlois, M. Munschauer, S. Dewell, M. Hafner, Z. Williams, U. Ohler, T.

- Tuschl, FMRP targets distinct mRNA sequence elements to regulate protein expression, *Nature*. 492 (2012) 382–6. <https://doi.org/10.1038/nature11737>.
32. V. Brown, P. Jin, S. Ceman, J.C. Darnell, W.T. O'Donnell, S.A. Tenenbaum, X. Jin, Y. Feng, K.D. Wilkinson, J.D. Keene, R.B. Darnell, S.T. Warren, Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome, *Cell*. 107 (2001) 477–87.
 33. J.C. Darnell, S.J. Van Driesche, C. Zhang, K.Y. Hung, A. Mele, C.E. Fraser, E.F. Stone, C. Chen, J.J. Fak, S.W. Chi, D.D. Licatalosi, J.D. Richter, R.B. Darnell, FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism, *Cell*. 146 (2011) 247–61. [https://doi.org/S0092-8674\(11\)00655-6](https://doi.org/S0092-8674(11)00655-6) [pii] 10.1016/j.cell.2011.06.013.
 34. A.C. McMahon, R. Rahman, H. Jin, J.L. Shen, A. Fieldsend, W. Luo, M. Rosbash, TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins, *Cell*. 165 (2016) 742–53. <https://doi.org/10.1016/j.cell.2016.03.007>.
 35. D. Ray, H. Kazan, K.B. Cook, M.T. Weirauch, H.S. Najafabadi, X. Li, S. Gueroussov, M. Albu, H. Zheng, A. Yang, H. Na, M. Irimia, L.H. Matzat, R.K. Dale, S.A. Smith, C.A. Yarosh, S.M. Kelly, B. Nabet, D. Mecnas, W. Li, R.S. Laishram, M. Qiao, H.D. Lipshitz, F. Piano, A.H. Corbett, R.P. Carstens, B.J. Frey, R.A. Anderson, K.W. Lynch, L.O. Penalva, E.P. Lei, A.G. Fraser, B.J. Blencowe, Q.D. Morris, T.R. Hughes, A compendium of RNA-binding motifs for decoding gene regulation, *Nature*. 499 (2013) 172–7. <https://doi.org/10.1038/nature12311>.
 36. J.A. Suhl, P. Chopra, B.R. Anderson, G.J. Bassell, S.T. Warren, Analysis of FMRP mRNA target datasets reveals highly associated mRNAs mediated by G-quadruplex structures formed via clustered WGA sequences, *Hum Mol Genet*. 23 (2014) 5479–91. <https://doi.org/10.1093/hmg/ddu272>.
 37. K.B. Jensen, K. Musunuru, H.A. Lewis, S.K. Burley, R.B. Darnell, The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 5740–5745. <https://doi.org/10.1073/pnas.090553997>.
 38. H.A. Lewis, K. Musunuru, K.B. Jensen, C. Edo, H. Chen, R.B. Darnell, S.K. Burley, Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome, *Cell*. 100 (2000) 323–32. [https://doi.org/S0092-8674\(00\)80668-6](https://doi.org/S0092-8674(00)80668-6) [pii].
 39. G. Nicastro, A.M. Candel, M. Uhl, A. Oregioni, D. Hollingworth, R. Backofen, S.R. Martin, A. Ramos, Mechanism of β -actin mRNA Recognition by ZBP1, *Cell Rep*. 18 (2017) 1187–1199. <https://doi.org/10.1016/j.celrep.2016.12.091>.

40. B.A. Ozdilek, V.F. Thompson, N.S. Ahmed, C.I. White, R.T. Batey, J.C. Schwartz, Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding, *Nucleic Acids Res.* 45 (2017) 7984–7996. <https://doi.org/10.1093/nar/gkx460>.
41. K.J. Zanotti, P.E. Lackey, G.L. Evans, M.R. Mihailescu, Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA, *Biochemistry.* 45 (2006) 8319–30. <https://doi.org/10.1021/bi060209a>.
42. M.C. Siomi, H. Siomi, W.H. Sauer, S. Srinivasan, R.L. Nussbaum, G. Dreyfuss, FXR1, an autosomal homolog of the fragile X mental retardation gene, *EMBO J.* 14 (1995) 2401–8.
43. M.C. Siomi, Y. Zhang, H. Siomi, G. Dreyfuss, Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them, *Mol Cell Biol.* 16 (1996) 3825–32.
44. L. Wan, T.C. Dockendorff, T.A. Jongens, G. Dreyfuss, Characterization of dFMR1, a *Drosophila melanogaster* homolog of the fragile X mental retardation protein, *Mol Cell Biol.* 20 (2000) 8536–47.
45. C. Cheong, P.B. Moore, Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures, *Biochemistry.* 31 (1992) 8406–14.
46. J. Deng, Y. Xiong, M. Sundaralingam, X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr(2+) ions at subatomic resolution (0.61 Å), *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 13665–70. <https://doi.org/10.1073/pnas.241374798>.
47. X. Wang, K.J. Goodrich, A.R. Gooding, H. Naeem, S. Archer, R.D. Paucek, D.T. Youmans, T.R. Cech, C. Davidovich, Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines, *Mol Cell.* 65 (2017) 1056–1067 e5. <https://doi.org/10.1016/j.molcel.2017.02.003>.
48. N.G. Walter, J.M. Burke, Fluorescence assays to study structure, dynamics, and function of RNA and RNA-ligand complexes, *Methods Enzym.* 317 (2000) 409–40.
49. J. Kim, C. Cheong, P.B. Moore, Tetramerization of an RNA oligonucleotide containing a GGGG sequence, *Nature.* 351 (1991) 331–2. <https://doi.org/10.1038/351331a0>.
50. A. Bugaut, S. Balasubramanian, 5'-UTR RNA G-quadruplexes: translation regulation and targeting, *Nucleic Acids Res.* 40 (2012) 4727–41. <https://doi.org/10.1093/nar/gks068>.

51. G.W. Collie, S.M. Haider, S. Neidle, G.N. Parkinson, A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex, *Nucleic Acids Res.* 38 (2010) 5569–80. <https://doi.org/10.1093/nar/gkq259>.
52. O. Kikin, L. D'Antonio, P.S. Bagga, QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, *Nucleic Acids Res.* 34 (2006) W676–82. <https://doi.org/10.1093/nar/gkl253>.
53. V.T. Mukundan, A.T. Phan, Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences, *J Am Chem Soc.* 135 (2013) 5017–28. <https://doi.org/10.1021/ja310251r>.
54. G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells, *Nat Chem.* 6 (2014) 75–80. <https://doi.org/10.1038/nchem.1805>.
55. T. Endoh, A.B. Rode, S. Takahashi, Y. Kataoka, M. Kuwahara, N. Sugimoto, Real-Time Monitoring of G-Quadruplex Formation during Transcription, *Anal. Chem.* 88 (2016) 1984–1989. <https://doi.org/10.1021/acs.analchem.5b04396>.
56. P. Ivanov, E. O'Day, M.M. Emara, G. Wagner, J. Lieberman, P. Anderson, G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 18201–18206. <https://doi.org/10.1073/pnas.1407361111>.
57. E. Chen, M.R. Sharma, X. Shi, R.K. Agrawal, S. Joseph, Fragile X mental retardation protein regulates translation by binding directly to the ribosome, *Mol Cell.* 54 (2014) 407–417. <https://doi.org/10.1016/j.molcel.2014.03.023>.
58. A. Ramos, D. Hollingworth, A. Pastore, G-quartet-dependent recognition between the FMRP RGG box and RNA, *RNA.* 9 (2003) 1198–207.
59. A.E. Arguello, A.N. DeLiberto, R.E. Kleiner, RNA Chemical Proteomics Reveals the N(6)-Methyladenosine (m(6)A)-Regulated Protein-RNA Interactome, *J Am Chem Soc.* 139 (2017) 17249–17252. <https://doi.org/10.1021/jacs.7b09213>.
60. R.R. Edupuganti, S. Geiger, R.G.H. Lindeboom, H. Shi, P.J. Hsu, Z. Lu, S.Y. Wang, M.P.A. Baltissen, P. Jansen, M. Rossa, M. Muller, H.G. Stunnenberg, C. He, T. Carell, M. Vermeulen, N(6)-methyladenosine (m(6)A) recruits and repels proteins to regulate mRNA homeostasis, *Nat Struct Mol Biol.* 24 (2017) 870–878. <https://doi.org/10.1038/nsmb.3462>.
61. P.J. Hsu, H. Shi, A.C. Zhu, Z. Lu, N. Miller, B.M. Edens, Y.C. Ma, C. He, The RNA-binding protein FMRP facilitates the nuclear export of N6-methyladenosine-containing mRNAs, *J. Biol. Chem.* 294 (2019) 19889–19895. <https://doi.org/10.1074/jbc.AC119.010078>.

62. K.D. Meyer, S.R. Jaffrey, Rethinking m(6)A Readers, Writers, and Erasers, *Annu Rev Cell Dev Biol.* 33 (2017) 319–342. <https://doi.org/10.1146/annurev-cellbio-100616-060758>.
63. C. Xu, X. Wang, K. Liu, I.A. Roundtree, W. Tempel, Y. Li, Z. Lu, C. He, J. Min, Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain, *Nat Chem Biol.* 10 (2014) 927–9. <https://doi.org/10.1038/nchembio.1654>.
64. T. Zhu, I.A. Roundtree, P. Wang, X. Wang, L. Wang, C. Sun, Y. Tian, J. Li, C. He, Y. Xu, Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of N6-methyladenosine, *Cell Res.* 24 (2014) 1493–6. <https://doi.org/10.1038/cr.2014.152>.
65. F. Corbin, M. Bouillon, A. Fortin, S. Morin, F. Rousseau, E.W. Khandjian, The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes, *Hum Mol Genet.* 6 (1997) 1465–72. <https://doi.org/dda192> [pii].
66. Y. Feng, D. Absher, D.E. Eberhart, V. Brown, H.E. Malter, S.T. Warren, FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association, *Mol Cell.* 1 (1997) 109–18.
67. A. Ishizuka, M.C. Siomi, H. Siomi, A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins, *Genes Dev.* 16 (2002) 2497–508. <https://doi.org/10.1101/gad.1022002>.
68. E.W. Khandjian, F. Corbin, S. Woerly, F. Rousseau, The fragile X mental retardation protein is associated with ribosomes, *Nat Genet.* 12 (1996) 91–3. <https://doi.org/10.1038/ng0196-91>.
69. R. Mazroui, M.E. Huot, S. Tremblay, N. Boilard, Y. Labelle, E.W. Khandjian, Fragile X Mental Retardation protein determinants required for its association with polyribosomal mRNPs, *Hum Mol Genet.* 12 (2003) 3087–96. <https://doi.org/10.1093/hmg/ddg335> ddg335 [pii].
70. D. Simsek, G.C. Tiu, R.A. Flynn, G.W. Byeon, K. Leppek, A.F. Xu, H.Y. Chang, M. Barna, The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity, *Cell.* 169 (2017) 1051-1065.e18. <https://doi.org/10.1016/j.cell.2017.05.022>.
71. B. DeMarco, S. Stefanovic, A. Williams, K.R. Moss, B.R. Anderson, G.J. Bassell, M.R. Mihailescu, FMRP - G-quadruplex mRNA - miR-125a interactions: Implications for miR-125a mediated translation regulation of PSD-95 mRNA, *PLoS One.* 14 (2019) e0217275. <https://doi.org/10.1371/journal.pone.0217275>.

72. R.S. Muddashetty, V.C. Nalavadi, C. Gross, X. Yao, L. Xing, O. Laur, S.T. Warren, G.J. Bassell, Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling, *Mol Cell*. 42 (2011) 673–88. [https://doi.org/S1097-2765\(11\)00337-6](https://doi.org/S1097-2765(11)00337-6) [pii] 10.1016/j.molcel.2011.05.006.
73. T.D. Pollard, A guide to simple and informative binding assays, *Mol Biol Cell*. 21 (2010) 4061–7. <https://doi.org/10.1091/mbc.E10-08-0683>.

Chapter 3:

The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through its RGG and C-terminal domains

Abstract

Fragile X mental retardation protein (FMRP) is an RNA-binding protein that regulates the translation of numerous mRNAs in neurons. The precise mechanism of translational regulation by FMRP is unknown. Some studies have indicated that FMRP inhibits the initiation step of translation, whereas other studies have indicated that the elongation step of translation is inhibited by FMRP. To determine whether FMRP inhibits the initiation or the elongation step of protein synthesis, we investigated m⁷G-cap-dependent and IRES-driven, cap-independent translation of several reporter mRNAs in vitro. Our results show that FMRP inhibits both m⁷G-cap-dependent and cap-independent translation to similar degrees, indicating that the elongation step of translation is inhibited by FMRP. Additionally, we dissected the RNA-binding domains of hFMRP to determine the essential domains for inhibiting translation. We show that the RGG domain, together with the C-terminal domain (CTD), is sufficient to inhibit translation while the KH domains do not inhibit mRNA translation. However, the region between the RGG domain and the KH2 domain may contribute as NT-hFMRP shows more potent inhibition than the RGG-CTD tail alone. Interestingly, we see a correlation between ribosome binding and translation inhibition, suggesting the RGG-CTD tail of hFMRP may anchor FMRP to the ribosome during translation inhibition.

Introduction

Fragile X syndrome (FXS) is the most common form of inherited intellectual disability and is caused by the reduced expression of the fragile X mental retardation 1 (*Fmr1*) gene in neurons¹. The expression of the *Fmr1* gene is reduced because of the expansion of the CGG trinucleotide repeats in the 5'-untranslated region of the gene, which results in abnormal methylation of the gene and the repression of transcription²⁻⁵. *Fmr1* encodes an RNA binding protein, fragile X mental retardation protein (FMRP), that is highly expressed in the brain⁶⁻¹⁰. FMRP has four RNA-binding domains: one RGG domain that is rich in arginines and glycines and three hnRNP K-homology domains (KH0, KH1, and KH2) (**Figure 1**)^{6,7,11}. FMRP is thought to regulate the translation of about 4% of the human fetal brain mRNAs by directly binding to specific sequences or structures in the target mRNAs¹². Additionally, the majority of FMRP in the cell is associated with polyribosomes, consistent with its proposed role in regulating protein synthesis¹³. Interestingly, missense mutations in the KH1 and KH2 domains (Gly266Glu and Ile304Asn, respectively, of human FMRP) abolishes the binding of FMRP to polyribosomes and cause FXS in humans^{14,15}. These results suggest that the RNA-binding activity of FMRP is critical for the healthy development and function of the brain.

Several studies have focused their efforts on trying to identify the mRNA targets of FMRP. Although thousands of mRNA targets have been identified over the years, there is minimal overlap among the mRNA targets from individual studies^{12,13,16-19}. These results suggest that it is difficult to discern the authentic

mRNA targets of FMRP. Interestingly, in vitro selection experiments identified a G-quadruplex (GQ) structure and a pseudoknot structure as the potential RNA ligands for the RGG and KH2 domains, respectively^{12,20–22}. Based on these results, it was proposed that FMRP may bind to mRNAs that possess GQ or pseudoknot forming sequences to repress their translation^{12,20,22,23}. We previously showed that the KH domains of FMRP do not bind to simple four or five nucleotide RNA sequence motifs compared to the canonical KH domains found in other RNA-binding proteins^{24–29}. It is possible that the KH domains of FMRP bind to pseudoknot structures in the mRNAs; however, such structures are not common in mRNAs and do not agree with the fact that FMRP appears to interact with thousands of mRNAs. In contrast, the RGG domain of FMRP binds to GQ structures with high affinity, and potential GQ structures are prevalent in mRNAs^{24,30–36}. Interestingly, GQ structures are even present in the 28S and 18S rRNA expansion segments^{37,38}. Therefore, possibly the RGG domain of FMRP interacts with GQ structures in the mRNA target or in the rRNAs to regulate translation.

The mechanism(s) by which FMRP regulates translation, however, is still unknown. One theory is that FMRP hinders translation initiation by binding to the target mRNA and recruiting the cytoplasmic FMRP-interacting protein (CYFIP1), which then binds the eukaryotic initiation factor 4E (eIF4E) at the mRNA 5' 7-methylguanosine (m⁷G) cap. Binding of CYFIP1 to eIF4E, in turn, prevents binding of eukaryotic initiation factor 4G (eIF4G) to eIF4E, and the eukaryotic initiation factor 4F (eIF4F) complex is prevented from assembling at the m⁷G cap³⁹. However, most of the cytoplasmic FMRP has been found associated with stalled polyribosomes^{13,40}.

Additionally, FMRP has been shown to interact with the ribosome even after translation initiation is blocked⁴¹. These results suggest that FMRP inhibits translation by binding to the target mRNA and stalling the ribosome during the translational elongation step.

Consistent with the idea that FMRP inhibits the elongation step, a cryo-electron microscopy (cryo-EM) structure of the *Drosophila* FMRP (dFMRP) bound to the *Drosophila* 80S ribosome showed that FMRP binds to a site within the ribosome that would prevent binding of tRNA and translation factors⁴². Additionally, the FMRP-ribosome structure revealed that the KH1 and KH2 domains bind near the ribosome's peptidyl site, and the RGG domain of FMRP residing near the aminoacyl site, presumably available for binding target mRNA motifs. Thus, FMRP may tether itself to the ribosome and the mRNA via its KH and RGG domains, respectively, to stall the ribosome during the elongation step of translation^{12,13,20,21,42,43}.

FMRP is also suggested to repress translation via RNA interference (RNAi). FMRP has been shown to associate with the RNA-induced silencing complex (RISC) proteins Dicer and Argonaute 2 (Ago2) as well as specific microRNAs (miRNAs)^{44–46}. In fact, FMRP has been reported to assemble the Ago2 and miRNA-125a inhibitory complex on the postsynaptic density protein 95 (PSD-95) mRNA through phosphorylation of FMRP^{47,48}. The PSD-95 RNAi case may be just one of many RNAi translation repression pathways FMRP employs. It is possible that FMRP can recruit from a catalog of miRNA complexes to inhibit a variety of mRNAs via RNAi.

FMRP may use multiple mechanisms to specifically repress an extensive catalog of target mRNAs under different conditions within the cell. To study the

mechanism of translational inhibition by FMRP, we analyzed the translation of 5'-m⁷G-capped mRNAs and uncapped mRNAs initiated using different IRES structures. Our studies show that FMRP reduces both m⁷G-cap-dependent and the cricket paralysis virus (CrPV) IRES-driven translation to a similar extent demonstrating that FMRP inhibits the elongation step of translation. Furthermore, we show that the RGG domain, together with the C-terminal domain (CTD), is sufficient to inhibit translation, whereas the KH domains do not inhibit mRNA translation. Interestingly, the RGG-CTD tail of FMRP binds to the ribosome suggesting that translation inhibition may depend on the interaction of FMRP with potential GQ structures in the rRNAs.

Materials and Methods

3.1 Protein expression and purification

The human FMR1 isoform 1 gene was assembled from *E. coli* codon-optimized gene blocks (IDT). The gene encoding the full-length isoform 1 human FMRP (hFMRP) spanning residues E2-P632 was subcloned into the LIC expression vector pMCSG7 (DNASU plasmid repository) with the 5' TEV cleavage site deleted, conferring a N-terminal hexahistidine tag. The N-terminus truncated human FMRP (NT-hFMRP) spanning residues R218-P632 was subcloned into the LIC expression vector pMCSG7, conferring a N-terminal hexahistidine tag with the 5' TEV cleavage site deleted. The human FMRP RGG domain (GST-hFMRP RGG) spanning

residues G531-P632 was subcloned into the LIC expression vector pMCSG10 (DNASU) which confers a N-terminal hexahistidine-GST fusion tag that is cleavable using TEV protease. The mutant SGG domain (GST-hFMRP SGG) was assembled from an *E. coli* codon-optimized gene block where all 16 arginine residues of the RGG region were mutated to serines. The resulting SGG fragment was subcloned into the LIC expression vector pMCSG10. The human FMRP tandem KH0-KH1-KH2 domains (GST-hFMRP KH0-KH1-KH2) spanning residues F126-K425 was also subcloned into the LIC expression vector pMCSG10.

The NT-hFMRP, GST-hFMRP RGG, GST-hFMRP SGG, and GST-hFMRP KH0-KH1-KH2 expression plasmids were all transformed in *E. coli* BL21(DE3) cells (Novagen). NT-hFMRP was purified using Ni-NTA affinity chromatography (Qiagen). GST-hFMRP RGG, GST-hFMRP SGG, and GST-hFMRP KH0-KH1-KH2 were purified using glutathione affinity chromatography (GE Healthcare). All proteins were further purified using Superdex 75 16/60 or Superdex 200 16/60 (GE Healthcare) gel filtration chromatography and stored in their respective running buffers. Gel filtration buffer composition varied between FMRP constructs. NT-hFMRP gel filtration running buffer contained 25 mM Tris pH 7.4, 150 mM KCl and 1 mM DTT. GST-hFMRP RGG and GST-hFMRP SGG were purified in gel filtration running buffer containing 50 mM Tris pH 7.5 and 1 mM DTT. GST-hFMRP KH0-KH1-KH2 was purified in gel filtration running buffer containing 50 mM Tris pH 7.4, 150 mM NaCl and 1 mM DTT.

3.2 Preparation of reporter mRNAs

The gene encoding the *Renilla* luciferase (RLuc) protein was modified with an poly(A)₂₅ insertion at the 3' end of the gene within the pRL-null vector (Promega). The gene encoding the NanoLuc luciferase was transferred from the pNL1.1 vector (Promega) into the pRL-null vector using 5' NheI and 3' XbaI restriction enzymes, then a poly(A)₂₅ was inserted at the 3' end of the NanoLuc gene. This generated RLuc and NanoLuc mRNAs with uniform 3' poly(A)₂₅ tails to bypass the need for post-transcriptional 3' polyadenylation. The PV IRES-, EMCV IRES-, HCV IRES-, and CrPV IRES-driven RLuc genes were subcloned into the EMCV-TurboGFP parental vector pT7CFE1 (ThermoFisher Scientific) in two steps. In the first step, the TurboGFP gene was replaced with the RLuc gene using 5' NdeI and 3' XhoI restriction enzymes to generate the EMCV-RLuc plasmid. In the second step, the EMCV IRES was replaced with either the PV, HCV, or CrPV IRES using 5' ApaI and 3' NdeI restriction enzymes to generate the PV-RLuc, HCV-RLuc, and CrPV-RLuc plasmids, respectively. The four IRES-driven RLuc genes contain a 3' poly(A)₃₀ sequence from the pT7CFE1 vector, which resulted in the synthesis of mRNAs with uniform 3' poly(A)₃₀ tails.

All 6 mRNAs were synthesized by run-off in vitro transcription using T7 RNA polymerase at 37°C for 4-6 hours. The RLuc template DNA was prepared by linearizing the vector using XbaI, while the NanoLuc and IRES-driven template DNAs were prepared by PCR amplification. The transcription reactions were treated with RQ1 DNase (Promega) at 37°C for 30 minutes to remove the DNA templates. Each

transcription reaction was cleaned up in two steps, first by chloroform extraction and ethanol precipitation followed by Monarch RNA cleanup kit (NEB).

The RLuc and NanoLuc mRNAs were post-transcriptionally 5' capped with 7-methylguanosine (m⁷G) using a homemade Vaccinia capping enzyme (REF). The mRNAs were first heated for 5 minutes at 65°C and then put on ice for 5 minutes. Capping was carried out for 2 hours at 37°C in 50 µL reactions with 1X capping buffer (50 mM Tris pH 8, 5 mM KCl, 1 mM MgCl₂, and 1 mM DTT), 0.5 mM GTP, 0.1 mM SAM, and 4 µL homemade Vaccinia capping enzyme; the amount of mRNA substrate to be modified was limited to maintain a large stoichiometric excess of both GTP and SAM to ensure all mRNA molecules were capped. The 5' m⁷G-capped mRNAs were then cleaned up using the Monarch RNA cleanup kit.

3.3 In vitro translation assays

The rabbit reticulocyte lysate was prepared following the method described by (1) Feng and Shao, and (2) Sharma *et al.* ^{49,50}. In vitro translation (IVT) reactions in rabbit reticulocyte lysate (RRL) were carried out in a final 20 µL volume at 30°C for 90 minutes. First, in a 10 µL binding reaction, 24 mM HEPES pH 7.5, 100 mM potassium acetate, 20 nM mRNA, and 500-2000 nM protein were combined and incubated for 15 minutes at room temperature; the control sample contained the protein storage buffer equivalent to the highest volume of protein used in the titration (maximum 2 µL). Then 10 µL RRL was added to make the final 20 µL IVT reaction. After 90 minutes at 30°C, the samples were brought to room temperature for 5-10

minutes while 2 μ L of 30 mM coelenterazine (Promega), or 15 μ L of 1-to-50 diluted NanoLuc substrate (Promega), was pipetted into a 384-well white flat bottom plate (Greiner). In the case of IVT reactions containing RLuc mRNAs, 18 μ L of the 20 μ L sample was mixed with 2 μ L coelenterazine, and luminescence readings were taken immediately on a Tecan Spark plate reader. In the case of IVT reactions containing NanoLuc mRNAs, 15 μ L of the 20 μ L sample was mixed with 15 μ L of 1-to-50 diluted NanoLuc substrate, and luminescence readings were taken immediately on a Tecan Spark plate reader. Luminescence readings of samples in each titration were normalized to the signal of their respective control sample.

For the GST-hFMRP RGG cleavage reactions with TEV protease, GST-hFMRP RGG fusion protein was combined with either TEV protease or storage buffer (50 mM Tris pH 7.4 and 1 mM DTT) at a 4:1 ratio for a final 20 μ M GST-hFMRP RGG. The cleavage reaction was carried out for 2 hours and 30 minutes at room temperature with gentle agitation, after which the protein samples were serially diluted in storage buffer for the IVT assay.

3.4 Human 80S ribosome binding assay

In a 20 μ L binding reaction, 0.8 μ M human 80S ribosomes and 7.2 μ M protein were incubated with 50 mM Tris acetate pH 7.7, additional potassium acetate (final 50 mM K⁺), additional magnesium acetate (final 3.8 mM Mg²⁺), 10 mM DTT, and 30 μ g/mL total tRNA from *E. coli*. The 20 μ L reactions were incubated for 15 minutes at room temperature, placed on ice for 2 minutes, and then carefully pipetted onto a

pre-cooled 200 μ L 40% (w/w) sucrose cushions reconstituted in RBB (50 mM Tris acetate pH 7.7, 50 mM potassium acetate, 5 mM magnesium acetate, and 10 mM DTT). The sample was centrifuged at 4°C for 1 hour and 40 minutes at 182,753 x g (38,000 rpm in a Beckman Coulter Type 42.2 Ti rotor).

The supernatant was carefully removed then 16 μ L RBB was gently pipetted onto the ribosome pellet and incubated covered at 4°C for 30 minutes. The 16 μ L sample was carefully mixed by gentle pipetting in place and transferred to a clean tube containing 4 μ L 5X SDS sample loading solution. 10 μ L of the 20 μ L gel sample was analyzed by 10% SDS-PAGE and stained with Coomassie brilliant blue.

Results and Discussion

3.5 FMRP inhibits the translation of various mRNAs

We sought to identify how the RNA-binding domains of human FMRP (hFMRP) contributed to the regulation of mRNA translation. We dissected hFMRP into two halves, the KH0-KH2 domains and the RGG domain with the C-terminal domain (CTD) (**Figure 3.1**). We used a nuclease-treated rabbit reticulocyte lysate (RRL) in vitro translation (IVT) system to test FMRP effects on the *Renilla* luciferase (RLuc) reporter mRNA translation. First, we measured the importance of two canonical post-transcriptional modifications, the 5'-m⁷G cap and 3' polyA tail, in increasing translation efficiency in the RRL system (**Figure 3.2**). We observed that the individual modifications are not critical for efficient translation, but removing the

5'-m⁷G cap and 3' polyA tail, significantly decreases translation efficiency in RRL (**Figure 3.2b**). FMRP has been proposed to regulate the translation of some mRNAs during the initiation step through interactions involving the 5' cap³⁹. Therefore, we tested how NT-hFMRP regulates the translation of a polyadenylated RLuc with and without the 5' cap. We observed FMRP inhibits both mRNAs similarly, suggesting that the inhibition is independent of the 5' cap machinery.

However, it was unclear whether FMRP specifically inhibits translation of RLuc mRNA because it contains a feature that makes it a target mRNA, or if FMRP would have a similar effect on the translation of different mRNAs. Therefore, we compared the effect of NT-hFMRP on the translation of *Renilla* luciferase (RLuc) and NanoLuc luciferase (NLuc) reporter mRNAs. Surprisingly, we observed nearly identical effects on both mRNAs, suggesting FMRP indiscriminately inhibits translation through a common mechanism (**Figure 3.2c**).

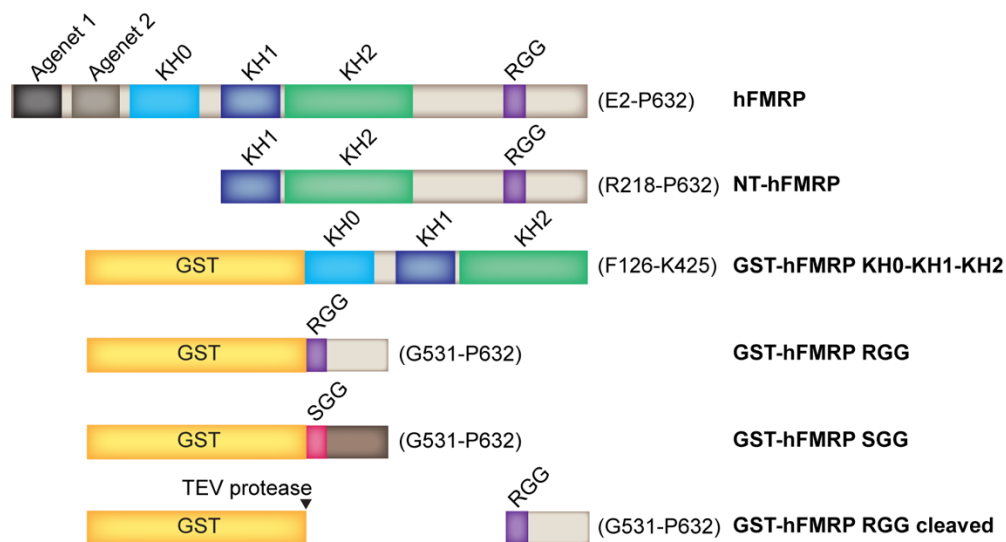


Figure 3.1: Human FMRP constructs. hFMRP is the full-length human FMRP isoform 1, spanning E2-P632, alongside the truncation constructs that were used. GST-hFMRP SGG is a fusion between the glutathione S-transferase and the RGG motif-containing sequence of hFMRP spanning G531-P632 where the 16 arginines spanning the RGG motif to the C-terminus were mutated to serines, illustrated using a different color for the C-terminus.

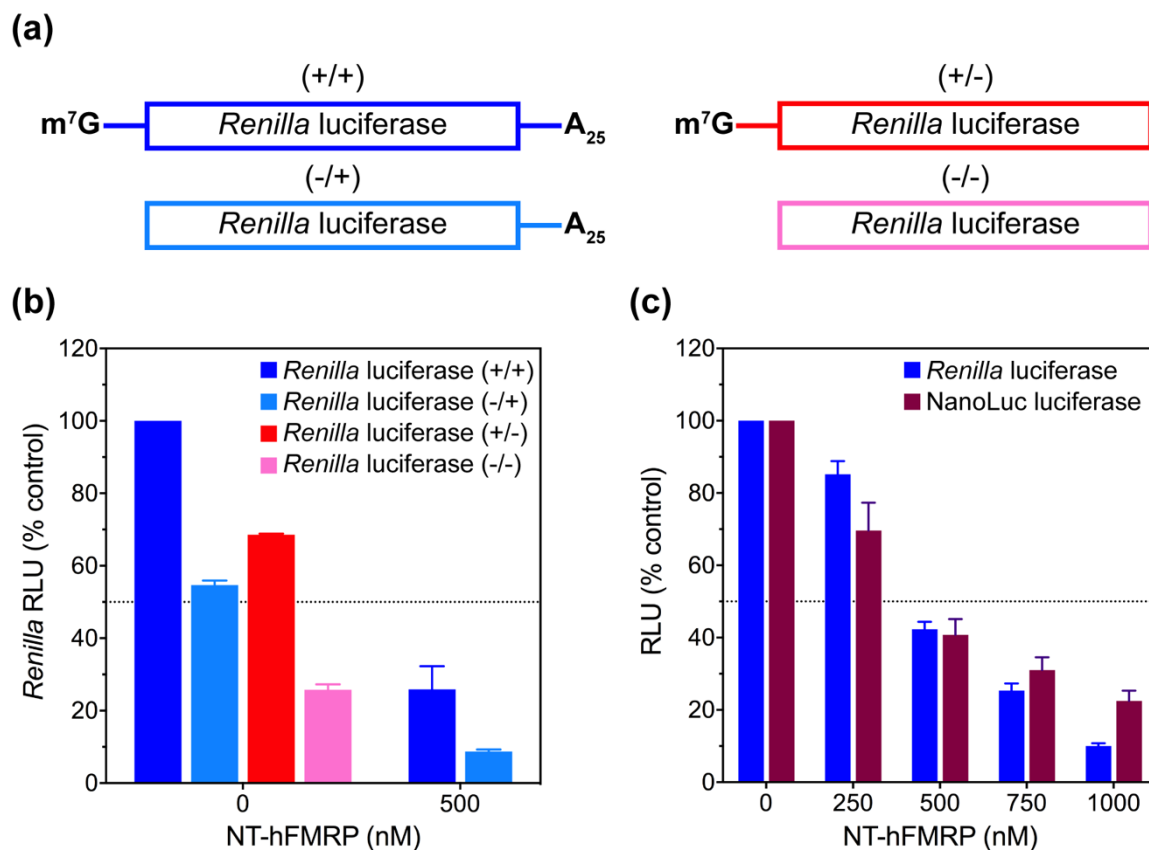


Figure 3.2: Human FMRP inhibits translation of different mRNAs in RRL. (a) Diagram of *Renilla* luciferase (RLuc) mRNAs containing: both 5'-m⁷G cap and 3'-polyA₂₅ tail (+/+), only 3'-polyA₂₅ tail (-/+), only 5'-m⁷G cap (+/-), or neither 5'-m⁷G cap nor 3'-polyA₂₅ tail (-/-). (b) Relative luminescence units (RLU) from 10 nM of the four aforementioned *Renilla* luciferase mRNAs, and upon the addition of 500 nM NT-hFMRP, all as a percentage of the (+/+) RLuc mRNA. (c) RLU from 10 nM *Renilla* luciferase (blue) and 10 nM NanoLuc luciferase (maroon) upon the addition of 0-1000 nM NT-hFMRP as a percentage of the control containing no NT-hFMRP. The standard deviations from three experiments are shown.

3.6 FMRP inhibits the elongation step of translation

We observed that hFMRP equally inhibited the translation of 5'-capped and uncapped RLuc mRNAs (**Figure 3.2b**). Therefore, we sought to test hFMRP inhibition of canonical capped-and-tailed RLuc mRNA against internal ribosome entry site (IRES)-driven mRNAs. We generated RLuc mRNAs driven by a representative IRES from the four different classes of IRESes: poliovirus (PV-RLuc), encephalomyocarditis virus (EMCV-RLuc), hepatitis C virus (HCV-RLuc), and cricket paralysis virus (CrPV-RLuc)⁵¹. Unfortunately, the PV-RLuc and HCV-RLuc mRNAs did not translate efficiently enough for IVT studies, so we proceeded with the EMCV-RLuc and CrPV-RLuc mRNAs at a fixed 10 nanomolar concentration for comparison against 10 nanomolar canonical RLuc (**Figure 3.3a**). While one may expect more protein synthesis from having more mRNAs in the IVT reaction, too much of the EMCV IRES-driven RLuc mRNA actually decreases translation efficiency⁵². Consistent with our previous observation, we saw FMRP inhibits the translation of CrPV-RLuc as efficiently as the canonical RLuc, suggesting that FMRP indeed inhibits translation of RLuc exclusively at the elongation step because CrPV IRES mRNA does not require any of the initiation factors to bind to the ribosome and initiate protein synthesis^{53–56}. Inhibition of EMCV-RLuc was slightly milder compared to canonical RLuc and CrPV-RLuc (**Figure 3.3b**).

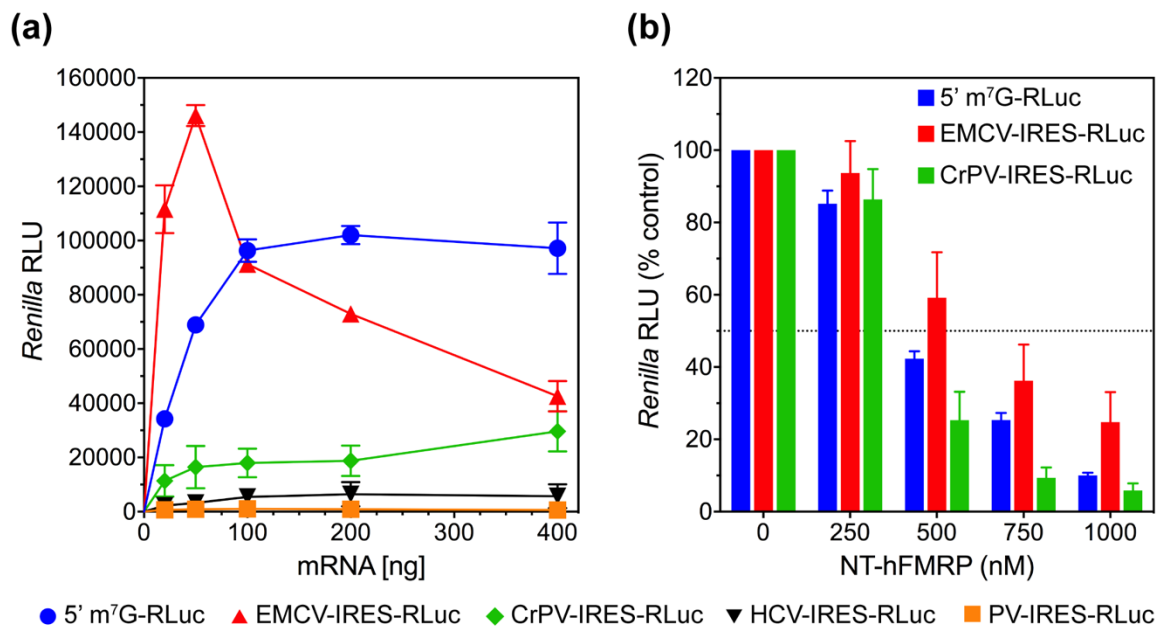


Figure 3.3: FMRP inhibits translation of canonical and IRES-driven *Renilla* luciferase mRNAs. (a) Relative luminescence units (RLU) from *Renilla* luciferase upon the addition of 0-400 ng of the indicated mRNAs in RRL. (b) RLU from canonical *Renilla* luciferase (blue), EMCV IRES-driven *Renilla* luciferase (red), and CrPV IRES-driven *Renilla* luciferase (green) upon the addition of 0-1000 nM NT-hFMRP as a percentage of the control containing no NT-hFMRP. The standard deviations from three experiments are shown.

3.7 The RGG domain and CTD are sufficient to inhibit translation

To identify which regions and RNA-binding domains of hFMRP contribute to translation inhibition, we measured the effects of the GST-tagged KH0-KH1-KH2 region (GST-KH0-KH1-KH2), GST-tagged RGG-CTD region (GST-hFMRP RGG), and a GST-tagged mutant SGG-CTD region (GST-hFMRP SGG) on RLuc mRNA translation. We found that GST-hFMRP RGG was sufficient to inhibit the translation of RLuc (**Figure 3.4a**). In contrast, GST-KH0-KH1-KH2 did not inhibit translation and had a similar effect as the mutant GST-hFMRP SGG in which the RNA G-quadruplex (GQ)-binding arginines spanning the RGG domain and CTD of FMRP were all mutated to serines (**Figure 3.4a**). Furthermore, we tested whether the RNA GQ-binding RGG domain was sufficient to inhibit translation. We used the 18 amino acid RGG peptide previously found to bind the sc1 GQ-forming RNA and observed it was insufficient to inhibit translation, suggesting the CTD of hFMRP plays an essential role in translation inhibition (**Figure 3.4a**)^{23,57}. Finally, the GST-hFMRP RGG is not as efficient at inhibiting RLuc translation compared to NT-hFMRP. This suggests there may indeed be a cooperative effect between the C-terminus of hFMRP, containing the RGG and CTD, the KH1 and KH2 domains, and the unstructured region between the KH2 and RGG domains. Altogether, it is possible the RGG and CTD are critical for the initial binding interaction as an anchor while the other regions of FMRP further stabilize that interaction.

To determine if the GST tag is involved in translation inhibition by the RGG-CTD region, we cleaved the GST-hFMRP RGG with TEV protease to separate the

GST tag and RGG-CTD fragments before IVT (**Figure 3.4b**). We observed no significant difference in translation inhibition between the intact GST-hFMRP RGG fusion protein and the cleaved protein, showing that the GST protein is not involved in translation inhibition by the RGG-CTD fragment (**Figure 3.4c**).

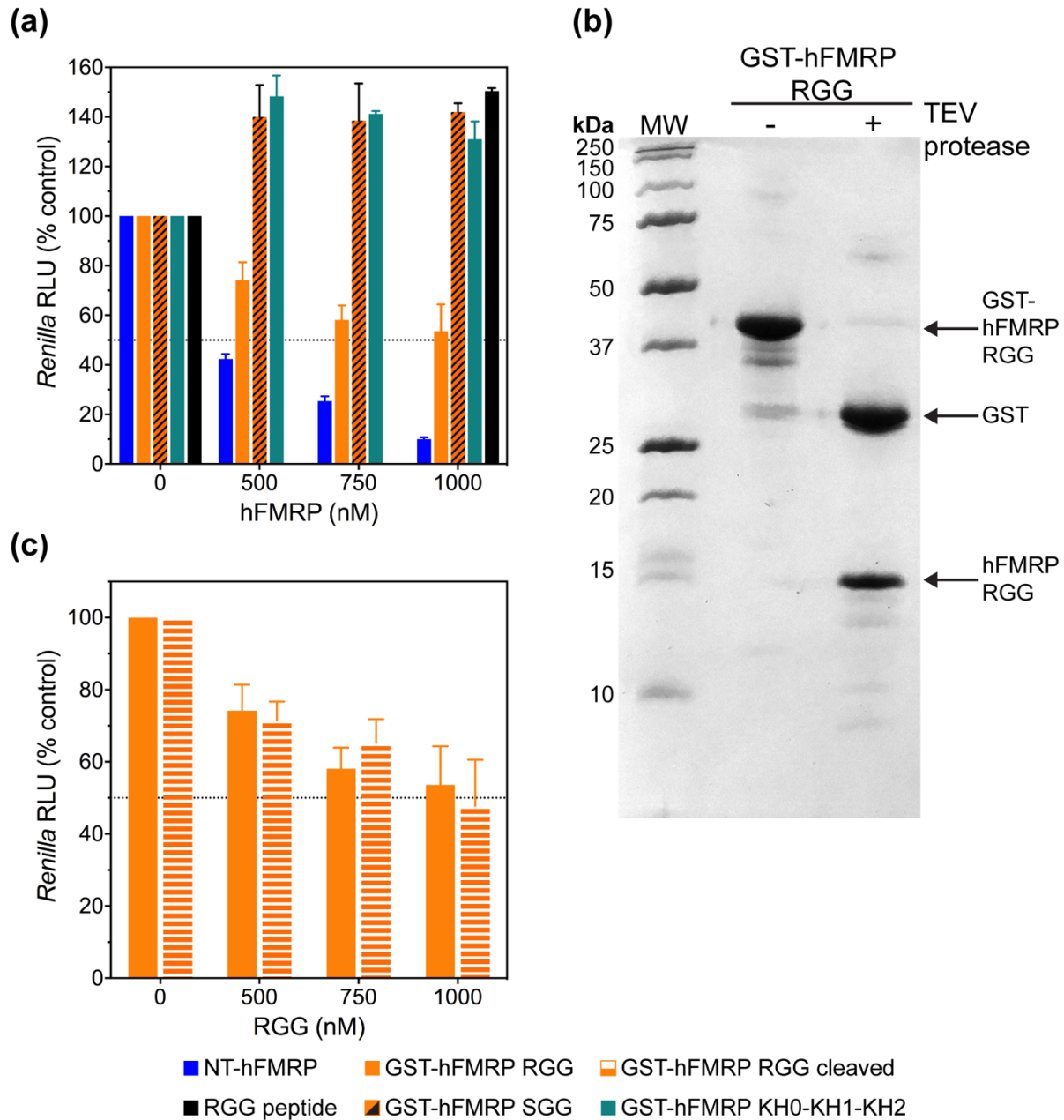


Figure 3.4: The RGG domain and CTD of FMRP are required for inhibition of translation. (a) Relative luminescence units (RLU) from *Renilla* luciferase upon the addition of 0-1000 nM NT-hFMRP (blue), GST-hFMRP RGG (orange), GST-hFMRP SGG (orange with black diagonal stripes), GST-hFMRP KH0-KH1-KH2 domains (teal), and RGG peptide (black) in RRL. (b) 12% SDS-PAGE of GST-hFMRP RGG before (- TEV protease) and after (+ TEV protease) cleaving with TEV protease. (c) RLU from *Renilla* luciferase upon the addition of 0-1000 nM GST-hFMRP RGG (orange) and GST-hFMRP RGG pre-cleaved using TEV protease (orange with white horizontal stripes) in RRL. The standard deviations from three experiments are shown.

3.8 FMRP binding to the ribosome correlates with inhibition of translation

Based on past findings and our observations that FMRP inhibits translation at the elongation step and the RGG-CTD region is sufficient for inhibition, we hypothesized the RGG-CTD might be serving as the anchor during direct binding of FMRP to the ribosome. Two recent studies found specific rRNA expansion segments in the ribosomes of higher eukaryotes contain G-quadruplex structures^{37,38}. Together with our recent study showing the RGG domain of hFMRP can bind even a minimal G-quadruplex structure, we investigated whether the RGG-CTD region of hFMRP could be binding directly to a G-quadruplex on the ribosome to inhibit translation²⁴. To determine whether binding to the ribosome is correlated with translation inhibition, we tested the binding of the human 80S ribosome to NT-hFMRP, GST-hFMRP RGG, and GST-hFMRP KH0-KH1-KH2 by sucrose cushion co-sedimentation. We observed both NT-hFMRP and GST-hFMRP RGG bind directly to the ribosome, while GST-hFMRP KH0-KH1-KH2 does not bind (**Figure 3.5**). Our results suggest that the direct binding of FMRP to the ribosome is crucial for inhibition of translation.

Figure 3.5: NT-hFMRP and GST-hFMRP RGG bind directly to the human 80S ribosome. (a) A representative 10% SDS-PAGE of 40% sucrose co-sedimentation assay showing NT-hFMRP (blue dot) co-sediments with human 80S ribosomes (h80S). Samples in lanes 4 to 6 were sedimented through the sucrose cushion. Lanes: 1, molecular weight ladder; 2, input h80S; 3, input NT-hFMRP; 4, NT-hFMRP in pellet; 5, h80S in pellet; 6, h80S+NT-hFMRP complex in pellet. NT-hFMRP migrated close to a ribosomal protein; however, quantification showed that the band intensity of the NT-hFMRP and the ribosomal protein in lane 6 increased by ~30% compared to the same ribosomal protein band in lane 5. (b) A representative 10% SDS-PAGE of 40% sucrose co-sedimentation assay showing GST-hFMRP RGG (orange dot) co-sediments with human 80S ribosomes, but GST-hFMRP KH0-KH1-KH2 (teal dot) does not. Samples in lanes 4 to 8 were sedimented through the sucrose cushion. Lanes: 1, molecular weight ladder; 2, input h80S; 3, input KH0-KH1-KH2; 4, KH0-KH1-KH2 in pellet; 5, h80S+KH0-KH1-KH2 in pellet; 6, h80S in pellet; 7, h80S+GST-RGG in pellet; 8, GST-RGG in pellet; 9, input GST-RGG; 10, molecular weight ladder. GST-RGG migrated close to a ribosomal protein; however, quantification showed that the band intensity of the GST-RGG and the ribosomal protein in lane 7 increased by ~40% compared to the same ribosomal protein band in lane 6. (c) Model for the inhibition of translation by FMRP. Top: The RGG-CTD tail of FMRP binds to a GQ sequence in a rRNA expansion segment and the KH domains interact with a functional site on the ribosome to inhibit translation. Bottom: The RGG-CTD tail of FMRP binds to a GQ sequence in the mRNA and the KH domains interact with a functional site on the ribosome to inhibit translation. Labels: FMRP (red), 80S ribosome (blue), rRNA expansion segments (dark blue lines), and mRNA (black).

3.9 Discussion

FMRP has been proposed to regulate the translation of thousands of mRNA targets suggesting that it either has relaxed RNA-binding specificity or the RNA recognition element is present abundantly in mRNAs^{12,13,16–19}. Alternatively, FMRP may globally inhibit translation by repressing the function of a specific translation factor or the ribosome. Our IVT studies in RRL lead us to conclude that human FMRP inhibits mRNAs indiscriminately and to similar degrees when concentrations of two different mRNAs of different lengths are standardized. This suggests FMRP does not necessarily regulate the translation of only a subset of target mRNAs, but can also regulate mRNAs globally through a common upstream translation mechanism independent of sequence or structure features in the mRNA.

FMRP also potently inhibits the translation of RLuc mRNA independent of the 5' cap, showing that potent inhibition can occur without FMRP interacting with the initiation machinery. To confirm our findings, we used CrPV IRES, which binds directly to the ribosomal P site to initiate translation without the need for any initiation factors^{53–56}. FMRP inhibits CrPV-RLuc just as potently as the canonical RLuc, supporting the hypothesis that FMRP inhibits at the elongation step regardless of the initiation rate and initiation machinery. We observe milder inhibition of EMCV-RLuc by FMRP, which may suggest FMRP's ability to inhibit translation is linked to factors beyond translation elongation. A recent study looked into the differences between canonical and EMCV-driven translation and found EMCV IRESes are indistinguishable from canonical 5' UTRs in terms of initiation and elongation rates.

However, the EMCV IRES is less efficient at recruiting ribosomes than the 5' cap machinery⁵⁸. Ribosome recruitment and availability, together with elongation, seem to be major factors for FMRP inhibition of translation.

Upon dissecting FMRP, we determined the arginine-rich RGG domain and CTD are the primary drivers of translation inhibition, while the region spanning the putative RNA-binding KH0 to KH2 domains alone cannot inhibit translation. Although the RGG-CTD region may be sufficient for mild translation inhibition, potent inhibition requires additional features found in NT-hFMRP, such as the unstructured region between the KH2 and RGG domains. Based on our ribosome-binding study, it is possible that the KH domains cannot inhibit translation alone because they cannot form a stable interaction with the ribosome. Perhaps the KH domains cooperate with the RGG-CTD to stabilize the initial interaction between the RGG-CTD and the ribosome to enhance translation inhibition.

Overall our studies suggest hFMRP can globally inhibit mRNA translation by binding directly to ribosomes through its RGG domain and CTD tail to (1) effectively decrease the population of active ribosomes available to translate mRNAs, and (2) stall actively translating ribosomes at the elongation step (**Figure 3.5c**). It remains to be seen whether RREs, such as G-quadruplexes, within the mRNAs can alter translation regulation by FMRP.

ACCESSION CODES

UniProtKB: Q06787

Chapter 3, in full, has been submitted for publication of the material as it may appear in The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through its RGG and C-terminal domains, Biochemistry, 2020. Athar, Youssi M.; Joseph, Simpson, American Chemical Society, 2020. The dissertation author was the primary investigator and author of this material.

References

1. Santoro, M. R., Bray, S. M., and Warren, S. T. (2012) Molecular Mechanisms of Fragile X Syndrome: A Twenty-Year Perspective. *Annual Review of Pathology: Mechanisms of Disease* 7, 219–245.
2. Coffee, B., Zhang, F., Warren, S. T., and Reines, D. (1999) Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat Genet* 22, 98–101.
3. Nelson, D. L., Orr, H. T., and Warren, S. T. (2013) The unstable repeats--three evolving faces of neurological disease. *Neuron* 77, 825–43.
4. Sutcliffe, J. S., Nelson, D. L., Zhang, F., Pieretti, M., Caskey, C. T., Saxe, D., and Warren, S. T. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet* 1, 397–400.
5. Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P., and et al. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905–14.
6. Siomi, H., Siomi, M. C., Nussbaum, R. L., and Dreyfuss, G. (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* 74, 291–8.
7. Ashley, C. T., Wilkinson, K. D., Reines, D., and Warren, S. T. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* 262, 563–6.
8. Devys, D., Lutz, Y., Rouyer, N., Bellocq, J. P., and Mandel, J. L. (1993) The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet* 4, 335–40.
9. Siomi, H., Choi, M., Siomi, M. C., Nussbaum, R. L., and Dreyfuss, G. (1994) Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* 77, 33–9.
10. Hinds, H. L., Ashley, C. T., Sutcliffe, J. S., Nelson, D. L., Warren, S. T., Housman, D. E., and Schalling, M. (1993) Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. *Nat Genet* 3, 36–43.

11. Myrick, L. K., Hashimoto, H., Cheng, X., and Warren, S. T. (2015) Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Hum Mol Genet* 24, 1733–40.
12. Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., Jin, X., Feng, Y., Wilkinson, K. D., Keene, J. D., Darnell, R. B., and Warren, S. T. (2001) Microarray Identification of FMRP-Associated Brain mRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome. *Cell* 107, 477–487.
13. Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y. S., Mele, A., Fraser, C. E., Stone, E. F., Chen, C., Fak, J. J., Chi, S. W., Licatalosi, D. D., Richter, J. D., and Darnell, R. B. (2011) FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* 146, 247–261.
14. Myrick, L. K., Nakamoto-Kinoshita, M., Lindor, N. M., Kirmani, S., Cheng, X., and Warren, S. T. (2014) Fragile X syndrome due to a missense mutation. *Eur J Hum Genet*.
15. De Boulle, K., Verkerk, A. J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B. A., and Willems, P. J. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet* 3, 31–5.
16. Ascano, M., Mukherjee, N., Bandaru, P., Miller, J. B., Nusbaum, J. D., Corcoran, D. L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., Williams, Z., Ohler, U., and Tuschl, T. (2012) FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* 492, 382–386.
17. McMahon, A. C., Rahman, R., Jin, H., Shen, J. L., Fieldsend, A., Luo, W., and Rosbash, M. (2016) TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 165, 742–753.
18. Ray, D., Kazan, H., Cook, K. B., Weirauch, M. T., Najafabadi, H. S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L. H., Dale, R. K., Smith, S. A., Yarosh, C. A., Kelly, S. M., Nabet, B., Mecnas, D., Li, W., Laishram, R. S., Qiao, M., Lipshitz, H. D., Piano, F., Corbett, A. H., Carstens, R. P., Frey, B. J., Anderson, R. A., Lynch, K. W., Penalva, L. O. F., Lei, E. P., Fraser, A. G., Blencowe, B. J., Morris, Q. D., and Hughes, T. R. (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature* 499, 172–177.
19. Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J., and Warren, S. T. (2014) Analysis of FMRP mRNA target datasets reveals highly associated mRNAs mediated by G-quadruplex structures formed via clustered WGGA sequences. *Human Molecular Genetics* 23, 5479–5491.

20. Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T., and Darnell, R. B. (2001) Fragile X Mental Retardation Protein Targets G Quartet mRNAs Important for Neuronal Function. *Cell* 107, 489–499.
21. Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* 20, 4803–13.
22. Darnell, J. C., Fraser, C. E., Mostovetsky, O., Stefani, G., Jones, T. A., Eddy, S. R., and Darnell, R. B. (2005) Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev* 19, 903–18.
23. Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B., and Patel, D. J. (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nature Structural & Molecular Biology* 18, 796–804.
24. Athar, Y. M., and Joseph, S. (2020) RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein. *J. Mol. Biol.* 432, 3851–3868.
25. Hollingworth, D., Candel, A. M., Nicastro, G., Martin, S. R., Briata, P., Gherzi, R., and Ramos, A. (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. *Nucleic Acids Res* 40, 6873–86.
26. Nicastro, G., Taylor, I. A., and Ramos, A. (2015) KH–RNA interactions: back in the groove. *Current Opinion in Structural Biology* 30, 63–70.
27. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K., and Darnell, R. B. (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5740–5745.
28. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B., and Burley, S. K. (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* 100, 323–32.
29. Nicastro, G., Candel, A. M., Uhl, M., Oregioni, A., Hollingworth, D., Backofen, R., Martin, S. R., and Ramos, A. (2017) Mechanism of β -actin mRNA Recognition by ZBP1. *Cell Rep* 18, 1187–1199.
30. Blice-Baum, A. C., and Mihailescu, M. R. (2014) Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms. *RNA* 20, 103–14.

31. Menon, L., and Mihailescu, M. R. (2007) Interactions of the G quartet forming semaphorin 3F RNA with the RGG box domain of the fragile X protein family. *Nucleic Acids Res* 35, 5379–92.
32. Menon, L., Mader, S. A., and Mihailescu, M. R. (2008) Fragile X mental retardation protein interactions with the microtubule associated protein 1B RNA. *RNA* 14, 1644–55.
33. Stefanovic, S., DeMarco, B. A., Underwood, A., Williams, K. R., Bassell, G. J., and Mihailescu, M. R. (2015) Fragile X mental retardation protein interactions with a G quadruplex structure in the 3'-untranslated region of NR2B mRNA. *Molecular BioSystems* 11, 3222–3230.
34. Zhang, Y., Gaetano, C. M., Williams, K. R., Bassell, G. J., and Mihailescu, M. R. (2014) FMRP interacts with G-quadruplex structures in the 3'-UTR of its dendritic target Shank1 mRNA. *RNA Biology* 11, 1364–1374.
35. Zanoliti, K. J., Lackey, P. E., Evans, G. L., and Mihailescu, M. R. (2006) Thermodynamics of the fragile X mental retardation protein RGG box interactions with G quartet forming RNA. *Biochemistry* 45, 8319–30.
36. Ozdilek, B. A., Thompson, V. F., Ahmed, N. S., White, C. I., Batey, R. T., and Schwartz, J. C. (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Research* 45, 7984–7996.
37. Mestre-Fos, S., Penev, P. I., Suttapitugsakul, S., Hu, M., Ito, C., Petrov, A. S., Wartell, R. M., Wu, R., and Williams, L. D. (2019) G-Quadruplexes in Human Ribosomal RNA. *J. Mol. Biol.* 431, 1940–1955.
38. Mestre-Fos, S., Penev, P. I., Richards, J. C., Dean, W. L., Gray, R. D., Chaires, J. B., and Williams, L. D. (2019) Profusion of G-quadruplexes on both subunits of metazoan ribosomes. *PLoS ONE* 14, e0226177.
39. Napoli, I., Mercaldo, V., Boyl, P. P., Eleuteri, B., Zalfa, F., De Rubeis, S., Di Marino, D., Mohr, E., Massimi, M., Falconi, M., Witke, W., Costa-Mattioli, M., Sonenberg, N., Achsel, T., and Bagni, C. (2008) The Fragile X Syndrome Protein Represses Activity-Dependent Translation through CYFIP1, a New 4E-BP. *Cell* 134, 1042–1054.
40. Stefani, G. (2004) Fragile X Mental Retardation Protein Is Associated with Translating Polyribosomes in Neuronal Cells. *Journal of Neuroscience* 24, 7272–7276.
41. Ceman, S., O'Donnell, W. T., Reed, M., Patton, S., Pohl, J., and Warren, S. T. (2003) Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Human Molecular Genetics* 12, 3295–3305.

42. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K., and Joseph, S. (2014) Fragile X Mental Retardation Protein Regulates Translation by Binding Directly to the Ribosome. *Molecular Cell* 54, 407–417.
43. Agrawal, R. K., Spahn, C. M., Penczek, P., Grassucci, R. A., Nierhaus, K. H., and Frank, J. (2000) Visualization of tRNA movements on the Escherichia coli 70S ribosome during the elongation cycle. *J Cell Biol* 150, 447–60.
44. Caudy, A. A., Myers, M., Hannon, G. J., and Hammond, S. M. (2002) Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* 16, 2491–6.
45. Ishizuka, A., Siomi, M. C., and Siomi, H. (2002) A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* 16, 2497–508.
46. Plante, I., Davidovic, L., Ouellet, D. L., Gobeil, L. A., Tremblay, S., Khandjian, E. W., and Provost, P. (2006) Dicer-derived microRNAs are utilized by the fragile X mental retardation protein for assembly on target RNAs. *J Biomed Biotechnol* 2006, 64347.
47. Edbauer, D., Neilson, J. R., Foster, K. A., Wang, C. F., Seeburg, D. P., Battersby, M. N., Tada, T., Dolan, B. M., Sharp, P. A., and Sheng, M. (2010) Regulation of synaptic structure and function by FMRP-associated microRNAs miR-125b and miR-132. *Neuron* 65, 373–84.
48. Muddashetty, R. S., Nalavadi, V. C., Gross, C., Yao, X., Xing, L., Laur, O., Warren, S. T., and Bassell, G. J. (2011) Reversible inhibition of PSD-95 mRNA translation by miR-125a, FMRP phosphorylation, and mGluR signaling. *Mol Cell* 42, 673–88.
49. Feng, Q., and Shao, S. (2018) In vitro reconstitution of translational arrest pathways. *Methods* 137, 20–36.
50. Sharma, A., Mariappan, M., Appathurai, S., and Hegde, R. S. (2010) In vitro dissection of protein translocation into the mammalian endoplasmic reticulum. *Methods Mol. Biol.* 619, 339–363.
51. Mailliot, J., and Martin, F. (2018) Viral internal ribosomal entry sites: four classes for one goal. *Wiley Interdiscip Rev RNA* 9.
52. Zeenko, V. V., Wang, C., Majumder, M., Komar, A. A., Snider, M. D., Merrick, W. C., Kaufman, R. J., and Hatzoglou, M. (2008) An efficient in vitro translation system from mammalian cells lacking the translational inhibition caused by eIF2 phosphorylation. *RNA* 14, 593–602.

53. Jan, E., and Sarnow, P. (2002) Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J Mol Biol* 324, 889–902.
54. Pestova, T. V., Lomakin, I. B., and Hellen, C. U. (2004) Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO reports* 5, 906–13.
55. Spahn, C. M., Jan, E., Mulder, A., Grassucci, R. A., Sarnow, P., and Frank, J. (2004) Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* 118, 465–75.
56. Pestova, T. V., and Hellen, C. U. (2005) Reconstitution of eukaryotic translation elongation in vitro following initiation by internal ribosomal entry. *Methods* 36, 261–9.
57. Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, D. J., and Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proceedings of the National Academy of Sciences* 112, E5391–E5400.
58. Koch, A., Aguilera, L., Morisaki, T., Munsky, B., and Stasevich, T. J. (2020) Quantifying the spatiotemporal dynamics of IRES versus Cap translation with single-molecule resolution in living cells. *bioRxiv* 2020.01.09.900829.

Chapter 4:
Conclusion and Future Directions

Dissecting the RNA-binding specificity of human FMRP (hFMRP) and the role of hFMRP in regulating mRNA translation has shed some light on (1) how the different RNA-binding domains of hFMRP contribute to binding specific RREs and (2) how these specific interactions between hFMRP and RREs may be triggering regulation of mRNA translation^{1,2}.

Our investigation into the RNA-binding specificity of hFMRP provided much needed clarity on what RREs the KH domains of hFMRP actually bind¹. During our study we developed the method Motif Identification by Analysis of Simple sequences (MIDAS) to isolate and identify the short ssRNA target sequences of an RNA-binding domain of interest. Using the Nova-2 KH3 domain as the benchmark protein in MIDAS, we identified new, higher affinity target ssRNA sequences of the Nova-2 KH3 domain. By comparison, the hFMRP KH0, KH1, and KH2 domains could not bind to any short ssRNA sequence with biologically relevant affinity, either in isolation or in tandem.

The MIDAS library preparation is simple enough to be accomplished quickly without expensive kits, and the next-generation sequencing data are easy to sort and interpret without the need for complex computational workflows. In our study, MIDAS was employed such that the 5 nucleotide variable region was flanked by two linear sequences. However, MIDAS is a versatile technique. It can also be used, for example, to display a short 4-6 nucleotide variable region as a loop within a stable hairpin backbone. This could identify new target sequences if the RNA-binding domain of interest binds to ssRNA sequences within loops of RNA hairpin structures³. Furthermore, the identified sequences can be introduced in tandem RNA

hairpin structures to determine whether the protein of interest simultaneously binds to tandem loop sequences.

We also found that the hFMRP RGG domain and CTD can not only interact with a variety of RNA GQ structures but also bind with high affinity to a minimal RNA GQ core element¹. We determined the interaction is indeed mediated by arginine residues spanning the RGG-CTD tail region, but it remains unknown how this region binds each of the different GQ structures and which specific arginine residues are important for binding these various RNA GQ structures. Considering the disordered nature of the RGG-CTD tail region, it is likely there are multiple GQ-binding conformations that involve different sets of arginines within the RGG-CTD that confer these observed differences in binding affinity between the various RNA GQ structures⁴⁻⁶. We know hFMRP can bind with high affinity to GQ structures in various mRNAs, and now we have shown that hFMRP can also bind to a minimal RNA GQ core element^{1,7-11}. As such, hFMRP can probably bind to many RNA GQ structure with a range of affinities.

From our in vitro translation studies, we see that the RGG-CTD tail region of hFMRP is indeed both necessary and sufficient for inhibiting mRNA translation, and for mediating binding to the human 80S ribosome². While we did not test whether the RGG domain alone could bind the GQ structures, we determined the CTD is required for translation inhibition. If we consider hFMRP RNA GQ-binding affinity together with translation inhibition potency, we can make a few connections: (1) the KH0-KH1-KH2 domains alone are insufficient to mediate binding to RNA or inhibit translation, (2) the RGG-CTD shows weaker binding to RNA GQ structures and a milder

inhibition of mRNA translation, and (3) NT-hFMRP shows strongest binding to RNA GQ structures and the most potent inhibition of mRNA translation. Altogether this suggests there is likely a connection between FMRP binding to RNA GQ structures and FMRP inhibiting mRNA translation. In consideration of the recent finding that ribosomes of higher eukaryotes, including humans and rabbits, contain GQ structures in their RNA extension segments, we tested if our hFMRP constructs could bind directly to the purified human 80S ribosome^{12,13}. To our surprise, we observed both RNA GQ-binding proteins, RGG-CTD and NT-hFMRP, bind to the ribosome; the KH0-KH1-KH2 domains, which could neither bind RNA nor inhibit mRNA translation, did not bind to the ribosome. Altogether these findings suggest FMRP can globally regulate mRNA translation perhaps by binding directly to a GQ structure on an extension segment of the 80S ribosome and preventing the ribosome from translating mRNAs. However, this does not exclude the possibility that other mechanisms could contribute to how FMRP regulates a given mRNA. For example, if an mRNA has GQ structures of its own, then it is possible for that mRNA to be inhibited by FMRP binding to the GQ structures in the mRNA and stalling actively translating polyribosomes. This could be tested by introducing stable GQ structures within a reporter mRNA (e.g. *Renilla* luciferase) and measuring how its translation is regulated compared to the original reporter mRNA.

Elucidating the atomic interactions between the 80S ribosome and FMRP would clarify many of the questions regarding how FMRP regulates mRNA translation. An earlier study by Chen *et al.* from our lab showed that *Drosophila* FMRP (dFMRP) can bind directly to purified *Drosophila* 80S (d80S) ribosomes, and we have since found that human FMRP

can also bind directly to purified human 80S ribosomes^{2,14}. Interestingly, unlike in higher eukaryotic ribosomes, there is no evidence of d80S ribosomes having extension segments with GQ structures. Indeed the NT-dFMRP•d80S cryo-EM structure model suggests (1) KH1 and KH2 domains are occupying the ribosomal intersubunit space peptidyl site (P site) that would normally be occupied by a tRNA during translation, and (2) the C-terminus containing the RGG domain resides near the A site and is therefore possibly interacting with the mRNA. Overall, the model put forth by Chen *et al.* is that *Drosophila* FMRP inhibits translation by binding to the d80S ribosome intersubunit space to block binding of the eEF1A•GTP•aminoacyl-tRNA ternary complex and eEF2 components and therefore stall elongation¹⁴. Human FMRP on the other hand could be inhibiting translation by binding to a GQ structure in an extension segment far from the intersubunit space where the elongation factors assemble. The manner in which FMRP interacts with ribosomes in humans is likely different from how FMRP interacts with ribosomes in *Drosophila*. Therefore, the mechanisms by which FMRP regulates translation of mRNAs in eukaryotes could be significantly different between species. Obtaining high-resolution structures of both human FMRP•human ribosome and *Drosophila* FMRP•*Drosophila* ribosome complexes would help tremendously towards understanding how FMRP interacts with the ribosome in different manners to regulate translation of mRNAs in different organisms.

References

1. Athar, Y. M., and Joseph, S. (2020) RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein. *J. Mol. Biol.* 432, 3851–3868.
2. Athar, Y. M., and Joseph, S. (2020) The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through its RGG and C-terminal domains. *bioRxiv* 2020.06.25.171967.
3. Jensen, K. B., Musunuru, K., Lewis, H. A., Burley, S. K., and Darnell, R. B. (2000) The tetranucleotide UCAY directs the specific recognition of RNA by the Nova K-homology 3 domain. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5740–5745.
4. Phan, A. T., Kuryavyi, V., Darnell, J. C., Serganov, A., Majumdar, A., Ilin, S., Raslin, T., Polonskaia, A., Chen, C., Clain, D., Darnell, R. B., and Patel, D. J. (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat. Struct. Mol. Biol.* 18, 796–804.
5. Vasilyev, N., Polonskaia, A., Darnell, J. C., Darnell, R. B., Patel, D. J., and Serganov, A. (2015) Crystal structure reveals specific recognition of a G-quadruplex RNA by a β -turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci.* 112, E5391–E5400.
6. Ozdilek, B. A., Thompson, V. F., Ahmed, N. S., White, C. I., Batey, R. T., and Schwartz, J. C. (2017) Intrinsically disordered RGG/RG domains mediate degenerate specificity in RNA binding. *Nucleic Acids Res.* 45, 7984–7996.
7. Blice-Baum, A. C., and Mihailescu, M. R. (2014) Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms. *RNA* 20, 103–14.
8. Menon, L., and Mihailescu, M. R. (2007) Interactions of the G quartet forming semaphorin 3F RNA with the RGG box domain of the fragile X protein family. *Nucleic Acids Res* 35, 5379–92.
9. Menon, L., Mader, S. A., and Mihailescu, M. R. (2008) Fragile X mental retardation protein interactions with the microtubule associated protein 1B RNA. *RNA* 14, 1644–55.
10. Stefanovic, S., DeMarco, B. A., Underwood, A., Williams, K. R., Bassell, G. J., and Mihailescu, M. R. (2015) Fragile X mental retardation protein interactions with a G quadruplex structure in the 3'-untranslated region of NR2B mRNA. *Mol. Biosyst.* 11, 3222–3230.

11. Zhang, Y., Gaetano, C. M., Williams, K. R., Bassell, G. J., and Mihailescu, M. R. (2014) FMRP interacts with G-quadruplex structures in the 3'-UTR of its dendritic target Shank1 mRNA. *RNA Biol.* 11, 1364–1374.
12. Mestre-Fos, S., Penev, P. I., Suttapitugsakul, S., Hu, M., Ito, C., Petrov, A. S., Wartell, R. M., Wu, R., and Williams, L. D. (2019) G-Quadruplexes in Human Ribosomal RNA. *J. Mol. Biol.* 431, 1940–1955.
13. Mestre-Fos, S., Penev, P. I., Richards, J. C., Dean, W. L., Gray, R. D., Chaires, J. B., and Williams, L. D. (2019) Profusion of G-quadruplexes on both subunits of metazoan ribosomes. *PloS One* 14, e0226177.
14. Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K., and Joseph, S. (2014) Fragile X Mental Retardation Protein Regulates Translation by Binding Directly to the Ribosome. *Mol. Cell* 54, 407–417.