UNIVERSITY OF CALIFORNIA SAN DIEGO

The Fragile X Proteins: Specific or Global Regulators of Translation?

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

in

Chemistry

by

Madison Marie Edwards

Committee in charge:

Professor Simpson Joseph, Chair
Professor Galia Debelouchina
Professor Partho Ghosh
Professor Judy Kim
Professor Eugene Yeo

2021

The dissertation of Madison Marie Edwards is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my family. Especially my mom, dad, and sister. I cannot heal past suffering and injustice, so I look toward the future. I hope this will be a small step toward a brighter and more peaceful life for others.

EPIGRAPH

“I love to see a young girl go out and grab the world by the lapels. Life's a bitch. You've got to go out and kick ass.”

— Maya Angelou

“Of course it is happening inside your head, Harry, but why on earth should that mean that it is not real?”

— J.K. Rowling, Harry Potter and the Deathly Hallows

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ARE	AU-rich element
CDS	Coding sequence
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
FMRP	Fragile X mental retardation protein
FXR1P	Fragile X mental retardation-related protein 1
FXR2P	Fragile X mental retardation-related protein 2
FXS	Fragile X syndrome
FXP	Fragile X proteins or Fragile X protein family
GQ	G-quadruplex
GST	Glutathione S-transferase
IVT	*In vitro* translation
KH	K-homology
MBP	Maltose-binding protein
m^7G	7-Methylguanosine
mRNA	Messenger ribonucleic acid
mRNP	Messenger ribonucleoprotein
NMM	N-methyl mesoporphyrin IX
NTD	N-terminal domain
PAGE	Polyacrylamide gel electrophoresis
RGG	Arginine-glycine-glycine

RLuc	*Renilla* luciferase
RNA	Ribonucleic acid
RRL	Rabbit reticulocyte lysate
ssRNA	Single-stranded ribonucleic acid
TP	Truncated products
tRNA	Transfer ribonucleic acid
UTR	Untranslated region

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Professor Simpson Joseph for welcoming me into his lab; I feel fortunate to have been able to research a topic that is so important to me. Throughout my Ph.D., Simpson not only supported my research goals, but he also supported my teaching, mentoring, and personal goals. Without this flexibility, I would not have grown or developed as much as a scientist, or as a person. For this, I am very thankful. I would like to thank Simpson for treating me like a collaborator and valuing my ideas. During emotionally low points, Simpson asked me how he could support me, and encouraged me to take time for myself, or to be with my family when I needed it. I have learned that I work best with supportive mentors, so I am grateful to have worked for a patient advisor. It was a privilege and an honor to work in the Chemistry and Biochemistry department at UCSD.

I would also like to thank my dissertation committee members for their support over the years. I would especially like to thank professor Partho Ghosh for hosting me for a rotation in his lab, his interest in my research, and his constructive feedback on my research progress.

The Joseph lab has been an enjoyable place to learn and perform research. I enjoyed all the conversations, brainstorming sessions, vent sessions, and laughter I shared with my colleagues. I would like to especially thank Adam Maloney for providing me with RRL, listening to vent sessions, discussing new recipes, and keeping me up to date on the news.

I would like to thank Stacey Brydges, Erilynn Heinrichsen, and my colleagues at the Engaged Teaching Hub. Thank you for sharing your passion for education with me, and for mentoring me throughout my time at UCSD.

I am eternally grateful to my parents for encouraging me to pursue my scientific interests. As a child they encouraged my interest in renting anatomy books at the library, dressing up as a white blood cell for Halloween, and bringing home insects to "investigate." I am thankful to them for patiently enduring all the phone calls during my undergraduate studies where I cried that I had failed my math or physics test, to inevitably find out later I had set the curve. In graduate school my parents called to asked how my science was going (usually not as planned), if I knew how to fix it (no), and how long I thought it would take to figure out how to fix it (exasperated sigh). Yet, they were always right that we would figure out a way to move forward. I am thankful that throughout my time in graduate school we could have visits to the mountains, backpacking trips, and family vacations together.

To my Aunt Maggie, thank you for the many times you listened to problems I wanted to discuss and provided advice. I value your opinion and appreciate that I had someone to talk to about the events in my life. I am also grateful to have someone to chat with who understands my appreciation for reality TV shows, and our trips to Tahoe and to your house.

I would like to thank my family for their love and constant support throughout my Ph.D. To my closest friends, Elizabeth, Jorge, and Krista, thank you for all the laughter, encouragement, and time you spent listening to me talk about my research.

Chapter 2, in full, is a reprint of the material as it appears in A Simple Procedure for Bacterial Expression and Purification of the Fragile X Protein Family, Scientific Reports, 2020. Edwards, Madison; Xu, Mingzhi; Joseph, Simpson, Nature, 2020. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in, The Fragile X Proteins Differentially Regulate Translation of Reporter mRNAs with G-quadruplex Structures, Journal of Molecular Biology, 2021. Edwards, Madison; Joseph, Simpson, Elsevier, 2021. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is currently being prepared for submission for publication as The Fragile X Protein Disordered Regions Bind a Novel RNA Target, Edwards, Madison; Huang, Molly; Joseph, Simpson. The dissertation author was the primary investigator and author of this material.

VITA

2016 Bachelor of Science, Chemistry and Biochemistry
California Polytechnic State University at San Luis Obispo

2018 Master of Science, Chemistry
University of California San Diego

2021 Doctor of Philosophy, Chemistry
University of California San Diego

PUBLICATIONS

M. Edwards, S. Joseph, The fragile X proteins differentially regulate translation of reporter mRNAs with G-quadruplex structures, J. Mol. Biol, (Accepted December 2021).

M. Edwards, M. Xu, S. Joseph, A simple procedure for bacterial expression and purification of the fragile X protein family, Sci. Rep. 10 (2020).

M. Minnes, S.G. Serslev, M. Edwards, Attitude Shifts and Transformation During Computer Science and Engineering Student Internships, (2020).

ABSTRACT OF THE DISSERTATION

The Fragile X Proteins: Specific or Global Regulators of Translation?

by

Madison Marie Edwards

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

Professor Simpson Joseph, Chair

Fragile X Syndrome (FXS) is a neurodevelopmental disorder that is the most prevalent form of inherited mental retardation and the primary monogenetic cause of autism. FXS, as well as some manifestations of autism spectrum disorder, results from improper RNA regulation due to a deficiency of fragile X mental retardation protein (FMRP) in neurons. FMRP and its autosomal paralogs, fragile X related proteins 1 & 2 (FXR1P/2P), have been implicated in many aspects of RNA regulation, from protein synthesis to mRNA stability and decay. The literature on the fragile X related proteins' (FXPs) role in mRNA regulation and their potential mRNA targets is vast, yet there is

little overlap between mRNA targets identified, or the proposed mechanism of mRNA regulation.

There is great interest in studying this family of proteins, yet researchers have faced much difficulty in expressing and purifying the full-length versions of these proteins in sufficient quantities. We developed a simple, rapid, and inexpensive procedure that allows for the recombinant expression and purification of full-length human FMRP, FXR1P, and FXR2P from *Escherichia coli* in high yields, free of protein and nucleic acid contamination. After confirming each protein's identity with mass spectrometry, we assessed the proteins' function after purification, and confirmed their binding to pseudoknot and G-quadruplex forming RNAs as well as their ability to regulate translation *in vitro*.

After developing a method to successfully purify the FXPs, we investigated how each protein regulated the translation of various mRNAs in a simplified *in vitro* translation (IVT) system. We developed an approach to investigate the function of FXPs in translational control using three potential mRNA targets. Briefly, we first selected top mRNA candidates found to be associated with the FXPs and whose translation are influenced by one or more of the FXPs. We then narrowed down the FXPs' binding site(s) within the mRNA and analyzed the strength of this binding *in vitro*. Finally, we determined how each FXP affects the translation of a minimal reporter mRNA containing the FXP-binding site in a minimalistic IVT system we designed. Overall, we observed all FXPs to bind RNAs containing G-quadruplexes with high affinity, for example, cyclin dependent kinase inhibitor p21 and FMRP's own coding region. Interestingly, FMRP inhibited the translation of each mRNA differently, in a

manner that appears to correlate with its binding affinity for each mRNA. In contrast, FXR1P/2P inhibited all mRNAs tested. Finally, although binding of these RNAs was due to the RGG (arginine-glycine-glycine) motif-containing C-terminal region of the FXPs, this region was not sufficient to cause inhibition of translation.

While designing a minimal IVT system, we discovered a novel RNA target of the FXPs. Although the most well-known RNA structural target of this family of proteins is the G-quadruplex (GQ), we discovered a non-GQ forming RNA target that is about 100 nucleotides in length and binds to all three FXPs with nanomolar dissociation constants. Furthermore, we determined that the last 102 amino acids of FMRP, which includes the RGG motif, were necessary and sufficient to bind this novel RNA target. To the best of our knowledge, this is the first example of a non-GQ structure RNA to be bound by the RGG motif/C-termini of FMRP.

Chapter 1:

Introduction

1.1 The Fragile X Proteins' Roles in RNA Regulation are Critical to Human Health

Fragile X Syndrome (FXS) is the predominant type of inherited intellectual disability and autism spectrum disorder, as well as the first genetic disorder to link RNA regulation to human cognitive function [1,2]. Patients with this disorder may experience seizures, hyperactivity, anxiety, and poor language development [1]. On a cellular level, patients with FXS possess a greater density of dendritic spines, and increased numbers of long and immature-shaped spines [3]. It is estimated that 1 in 5,000 males and 1 in 4,000 to 8,000 females possess the full FXS mutation [1]. FXS predominantly results from a CGG trinucleotide repeat expansion in the 5' untranslated region (UTR) of the *FMR1* gene [1,4]. The expanded repeats are hypermethylated causing transcriptional silencing of the *FMR1* gene, leading to a deficiency or absence of fragile X mental retardation protein (FMRP) [1,4–6]. FMRP's role in RNA regulation and translation repression has been studied extensively, particularly as it relates to FXS [4]. Beyond FXS and autism spectrum disorder, genes of FMRP mRNA targets are enriched for those associated with psychiatric disorders: a recent article found that genes with a high probability of being FMRP targets were enriched for association with schizophrenia, bipolar disorder, and major depressive disorder [7]. If we extend our discussion of RNA regulation to the entire fragile X protein family, the impacts of RNA misregulation on human health go beyond neuronal development and cognition.

FMRP is one of three RNA-binding, ribosome-associating proteins involved in translational regulation that comprise the fragile X protein (FXP) family; the other two members are referred to as fragile X-related protein 1 (FXR1P) and fragile X-related

protein 2 (FXR2P) [8–11]. While less studied, FMRP's autosomal paralogs, FXR1P and FXR2P, are noteworthy for their role in translational regulation as their deficiency also leads to developmental abnormalities [8,9,12]. For example, FXR2P-deficient mice have impaired dendritic maturation of new neurons, with new neurons possessing shorter and less complex dendrites compared to wild-type mice [13]. These mice revealed decreased neural connectivity as new neurons with shorter dendrites connected to fewer presynaptic neurons [13]. These mice revealed decreased neural connectivity as new neurons with shorter dendrites connected to fewer presynaptic neurons [13]. Mice deficient in FXR2P displayed atypical gene expression in the brain and altered behavior, such as hyperactivity, reduced sensitivity to heat stimuli, and reduced prepulse inhibition [14,15]. On the other hand, FXR1P is unique among the FXPs as in humans, FXR1P mRNA demonstrates alternative splicing and is abundant in heart and skeletal muscle tissue [17]. Elimination of *Fxr1* leads to neonatal lethality in mice, while reduced levels of FXR1P lead to shortened life spans and reduced limb musculature [19]. Furthermore, FXR1P expression is altered in myoblasts from patients with facioscapulohumeral muscular dystrophy [18].

1.2 Architecture of the Fragile X Proteins

The genes encoding the FXPs are highly homologous through the first 13 exons of FMRP, although FXR1P and FXR2P lack sequences corresponding to exons 11 and 12 of FMRP [20]. After exon 13 of FMRP, the sequences of the three proteins diverge considerably [20]. This suggests that the three proteins likely arose from multiple gene

duplications of a common ancestral gene [20]. The FXPs are multidomain proteins with high amino acid identity over the first 58-70% of their sequences, but lower identity thereafter **(Figure 1.1A)** [21]. All three proteins possess RNA-binding domains of interest: three K homology (KH) domains within the highly conserved N-terminal region, and an arginine-glycine-glycine (RGG) motif with poor conservation located in the divergent C-termini **(Figure 1.1B-D)** [12,22-23]. These domains are noteworthy, since proteins containing KH domains are involved in splicing and regulation of transcription and translation, while RGG motifs are present in greater than 1,000 human proteins, with roles in numerous processes such as translation, transcription, pre-mRNA splicing, DNA damage signaling, and apoptosis regulation [24-25]. In addition, all three proteins possess two Agenet domains which have been shown to bind methylated lysines [26-27]. The high sequence conservation in the N-termini of the FXP family suggests they exhibit some functional redundancy, while their divergent C-termini likely contribute to their unique functions.

A less explored feature of the FXPs is their C-terminal intrinsically disordered region (IDR) which constitutes ~30-43% of the entire protein sequence but has lower sequence conservation [21]. IDRs are enriched in RNA-binding proteins compared to the entire human proteome and can support protein aggregation, phase transitions, and bind to RNA both specifically and non-specifically [28-29]. Within the IDR of FXR1P and FXR2P are arginine-rich motifs that likely impart these paralogs with unique RNA-binding capabilities. The isoform of FXR1P that we study, isoform 2, has one such region, while FXR2P has two. Due to their similarity to the nucleolar-targeting signal (NoS) of the protein Rev of human immunodeficiency virus type 1, these

sequences are referred to as NoS1 and NoS2 respectively [30]. In other proteins, these motifs have been shown to support RNA recognition, where the few non-arginine amino acids mediate specific binding [31].

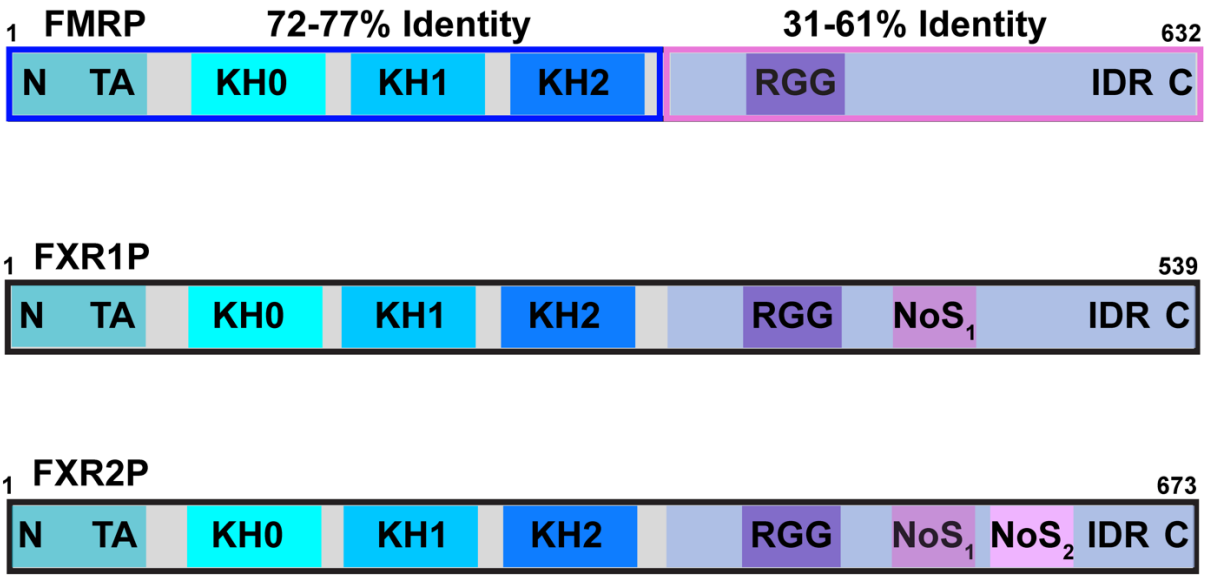


Figure 1.1: Architecture of the fragile X proteins. (A) The fragile X proteins are multidomain proteins that are well-conserved through the sequence RQIG of each protein (outlined in blue), but their conservation diverges after this point (outlined in pink) [21]. (B-D) Schematics of proteins used in this article showing relevant domains. We used human FMRP isoform 1, human FXR1P isoform 2, and human FXR2P.

1.3 The FXPs bind specific sequence or structure motifs in target mRNAs

The RNA-binding capabilities of the FXPs are well-documented, and there appears to be some overlap of mRNA targets amongst the FXPs, as well as evidence that each protein has unique mRNA targets [12-13, 32-35]. Several examples of recognized RNA structural targets of FMRP are G-quadruplexes (GQ) and kissing complexes [12, 36-38].

GQs are an RNA structure that has repeatedly been reported as a target of FMRP and FXR1P, and we have recently shown it is a target of the entire FXP family (Figure 1.2) [12,21,37-39]. An NMR structure of the human FMRP RGG motif bound to the *in vitro*-selected *sc1* RNA G-quadruplex (GQ) was beneficial for modeling how the disordered RGG motif is capable of binding to a GQ [40]. Although *sc1* RNA was selected as a stable RGG motif binding partner, FMRP has also been shown to bind GQ structures in biologically relevant target mRNAs [37-38]. As potential GQs are frequently found in mRNA, within 5' and 3' UTRs as well as the CDS, the RGG motif within FMRP has the potential to bind a great number of target mRNAs *in vivo* [41].

Loop-loop pseudoknots, or "kissing complexes," were identified as an *in vitro* selected target of FMRP [36]. The KH2 domain of FMRP was identified as necessary and sufficient for FMRP binding to an *in vitro* selected kissing complex referred to as KC2 [36]. Furthermore, the KH2 domains of FMRP, FXR1P, and FXR2P were found to bind KC2 RNA with equal affinity [12]. This is not altogether surprising, as the KH domains are highly conserved amongst the FXPs.

For sequence motifs targeted by FMRP, GACR (R is a purine, A or G) and WGGA (W is an A or U) were identified [32]. The GACR sequences were enriched for an FMRP variant composed only of the KH1 and KH2 RNA-binding domains, whereas the WGGA sequences were proposed to bind specifically to the KH1 domain of FMRP [42-43]. For FXR1P, several studies have reported the binding of FXR1P to AU-rich elements (AREs), particularly those located in the 3' UTRs of mRNAs [34, 44-46].

As loss of FMRP has been linked to the deregulation of translation of specific mRNAs by the ribosome, many researchers have focused their attention on expanding the list of known FXP RNA sequence and structural targets [47-48]. In many instances, mRNA targets of the FXPs have been identified by RNA pull-down assays, such as cross-linking immunoprecipitation (CLIP), which report RNA-protein interactions that may not have a strong initial interaction and may not be physiologically relevant [32]. Such errors can be introduced through the crosslinking method utilized, particularly as the specificity of crosslinking is not fully understood [32-33]. This problem is prevalent as many studies have reported results with little overlap and have identified few targets with a validated association with FMRP [32]. In an attempt to disentangle our understanding of FMRP's mRNA targets, Suhl *et al.* performed a comparison of three large FMRP mRNA target studies and found an overlap of only ~3.2%, indicating that the methods used significantly impacted the results [32]. As such, it is imperative that the FXP field focuses on methods to validate such proposed mRNA targets.

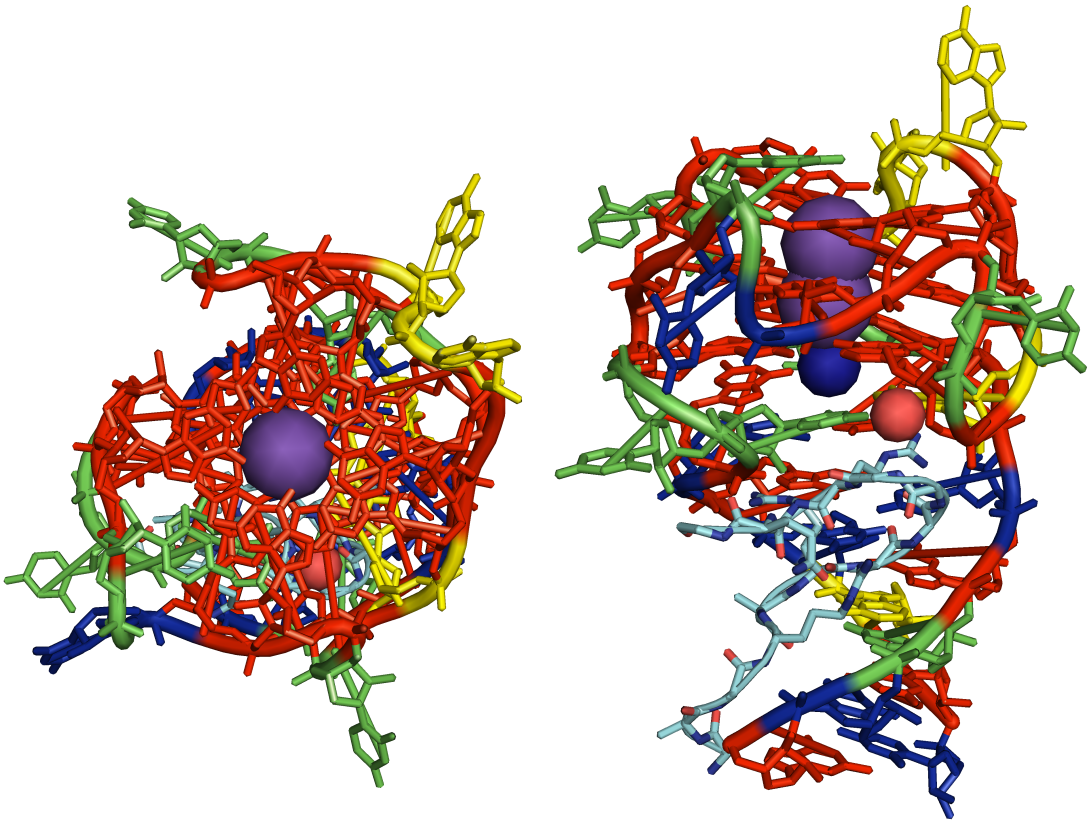


Figure 1.2: A cesium-bound human FMRP RGG motif binds *sc1* RNA. In the top view (left), guanosines (shown in red) form G tetrads. In the side view (right), the RGG motif (cyan) binds the *sc1* RNA below the GQ structure. Two K^+ ions (shown as purple spheres) and a Cs^+ ion (indigo sphere) coordinate to O6 of the tetrad guanines. Uridines, adenosines, and cytidines are depicted in green, yellow and blue, respectively. It is possible that the oxygen atom (shown as a red sphere) may mediate the binding between the RGG motif and *sc1*. 2.8 Å crystal structure from Protein Data Bank entry code 5DEA.

1.4 The FXPs have been reported to regulate translation through diverse mechanisms

Translation regulation is essential for controlling gene expression during development, differentiation, nervous system function, nutrient deprivation, stress, and aging [49]. Compared to transcriptional regulation, it can induce more rapid alterations in the concentrations of proteins in the cell, making it useful for preserving homeostasis as well as regulating lasting changes to the cell [49]. Translation is regulated in multiple ways, such as RNA silencing by microRNAs, mRNA sequestration, or by regulatory proteins [50]. The FXPs are an example of such regulatory proteins. As the FXPs interact strongly with one another, bind RNA, and associate with the ribosome, these capabilities have been explored as mechanisms through which the FXPs regulate translation [10]. Translation regulation by the FXPs, as well as other regulatory proteins, is of great interest as the tightly monitored expression of such mRNAs affects neuron development and integration, which is vital to central nervous system (CNS) function. Identifying a mechanism of mRNA regulation by such regulatory proteins could be used to investigate learning, memory, and speech impairments arising from improper CNS development, which affect more than 3.5 million Americans, costing U.S. citizens more than 262 billion dollars a year [51].

One immense challenge to determining mechanisms of translation regulation by the FXPs has been confounding influences, such as the biological material or cell type used for assays, the assay implemented, the analyses applied, the presence of FXP interaction partners, and even the particular isoform of the FXP, to name just a

few [32]. Differing and contradictory results have been obtained: some suggest the proteins interact primarily with mRNA, while others suggest the proteins also interact with ribosomes [32,52-56]. For example, it has been reported that FXR1P can either repress or activate translation of its target mRNAs in non-neuronal cells depending on the cellular context, while different isoforms of FXR1P appear to inhibit translation through distinct mechanisms such as destabilizing mRNA transcripts or through regulating the process of translation itself [57-58]. Results from our own lab indicate that FMRP recognizes and binds specific sequences within mRNAs, then subsequently influences their translation. Conversely, FXR1P and FXR2P appear to inhibit translation globally, irrespective of their binding affinity to the mRNA in question.

As FMRP has been found to bind target mRNAs associated with stalled ribosomes, the Joseph lab investigated how the ability of FMRP to associate with ribosomes and its presence in polyribosomes is involved in its ability to inhibit translation [2, 59]. Previous work in the Joseph lab revealed that *Drosophila* FMRP (dFMRP) binds to the L5 protein of the 80S ribosome to inhibit translation [52]. A cryo-EM reconstruction of the complex formed between dFMRP and the 80S ribosome indicated that dFMRP likely binds in the inter-subunit space of the ribosome, which could prevent elongation factors and tRNA from binding, thus suggesting a potential mechanism for translational inhibition by dFMRP [52]. As the mechanism of translation inhibition by human and *Drosophila* FMRPs may differ, it is preferable to have a cryo-EM structure of human FMRP bound to human ribosomes. Our lab is currently pursuing this structure.

One area of research has focused on determining protein interaction partners of the FXP family and elucidating the effects these interactions have on the proteins' ability to regulate translation. All three proteins can form homomers, as well as heteromeric complexes with either of the other FXPs [60]. The formation of such heteromeric complexes has been proposed as a mechanism through which the FXP family's functions are regulated. In fact, it was discovered that when certain isoforms of FXR1P form a heterodimer with FMRP, this inhibits the affinity of FMRP for G-quartet RNA, an RNA structure bound by FMRP [61]. Evidence suggests that the N-terminal region of FMRP contains a protein-protein interaction motif involved in its ability to dimerize as well as interact with other proteins, such as nuclear FMRP interacting protein (NUFIP) and cytoplasmic FMRP interacting proteins 1 and 2 (CYFIP1/2) [62-64]. Such interactions appear to be important for modulating the translation regulation activity of the FXP family.

One theory proposes that FMRP hinders translation initiation by binding target RNA and recruiting CYFIP1 which in turn binds eukaryotic initiation factor 4E (eIF4E) at the mRNA 5' 7-methylguanosine (m^7G) cap [65]. Binding of CYFIP1 to eIF4E then precludes binding of eukaryotic initiation factor 4G (eIF4G) to eIF4E, and the eukaryotic initiation factor 4F (eIF4F) complex cannot form [65]. As for FXR1P, a recent study from George et al. determined that FXR1P binds to AREs in the 3' UTR of cMYC mRNA to stabilize its expression [46]. Meanwhile, the RGG domain of FXR1P interacts with eIF4A1 and eIF4E [46]. Together, these interactions result in the circularization of cMYC mRNA to promote recruitment of eukaryotic translation initiation factors, increasing translation of this mRNA in ovarian cancer cells [46].

1.5 Summary of research significance

This work attempts to address several of the previously described obstacles researchers have encountered when studying the FXPs. In Chapter 2, we discuss our method for a simple, rapid purification method that allows for the purification of all three full-length FXPs from *E. coli*. We are unaware of other purification methods that produce such quantities (2-9 mg) of full-length FXPs in a similar amount of time (3 days). For these reasons, we believe our method will allow researchers to advance more rapidly in studying the FXPs.

In Chapter 3, we present a minimalistic *in vitro* translation reporter system we created to observe the FXPs' effect more accurately on specific mRNA target sequences, without conflating the results due to FXP binding sites within the reporter RNAs themselves. Our study allowed us to determine two biologically relevant mRNA sequences to which the FXPs bind through their RGG motifs/C-termini. Moreover, we observed that FMRP shows greater inhibition of mRNAs to which it binds, whereas FXR1/2P appear to inhibit all mRNAs tested regardless of binding.

Finally, in Chapter 4 we identify a novel *in vitro* RNA target that is bound by the RGG motif/C-termini of the FXPs. To our knowledge, this is the first non-GQ RNA target of the FXPs RGG motif/C-termini. As such, we believe researchers may wish to determine if this structure is present in biological mRNAs regulated by the FXPs, and if this new RNA target is bound by RGG motifs in other RNA-binding proteins.

References

[1] R.J. Hagerman, E. Berry-Kravis, H.C. Hazlett, D.B. Bailey, H. Moine, R.F. Kooy, F. Tassone, I. Gantois, N. Sonenberg, J.L. Mandel, P.J. Hagerman, Fragile X syndrome, Nat. Rev. Dis. Prim. 3 (2017) 17065. https://doi.org/10.1038/nrdp.2017.65.

[2] J.C. Darnell, S.J. Van Driesche, C. Zhang, K.Y.S. 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) 247–261. https://doi.org/10.1016/j.cell.2011.06.013.

[3] S.A. Irwin, B. Patel, M. Idupulapati, J.B. Harris, R.A. Crisostomo, B.P. Larsen, F. Kooy, P.J. Willems, P. Cras, P.B. Kozlowski, R.A. Swain, I.J. Weiler, W.T. Greenough, Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: A quantitative examination, Am. J. Med. Genet. 98 (2001) 161–167. https://doi.org/10.1002/1096-8628(20010115)98:2<161::AID-AJMG1025>3.0.CO;2-B.

[4] E. Chen, S. Joseph, Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins, Biochimie. 114 (2015) 147–154. https://doi.org/10.1016/j.biochi.2015.02.005.

[5] M. Pieretti, F. Zhang, Y.-H. Fu, S.T. Warren, B.A. Oostra, C.T. Caskey, D.L. Nelson, Absence of expression of the FMR-1 gene in fragile X syndrome, Cell. 66 (1991) 817–822. https://doi.org/10.1016/0092-8674(91)90125-I.

[6] 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. https://doi.org/10.1093/hmg/1.6.397.

[7] N.E. Clifton, E. Rees, P.A. Holmans, A.F. Pardiñas, J.C. Harwood, A. Di Florio, G. Kirov, J.T.R. Walters, M.C. O'Donovan, M.J. Owen, J. Hall, A.J. Pocklington, Genetic association of FMRP targets with psychiatric disorders, Mol. Psychiatry. (2020) 1–14. https://doi.org/10.1038/s41380-020-00912-2.

[8] 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., Eur. Mol. Biol. Organ. J. (1995).

[9] Y. Zhang, J.P. O'Connor, M.C. Siomi, S. Srinivasan, A. Dutra, R.L. Nussbaum, G. Dreyfuss, The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2., EMBO J. 14 (1995) 5358. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394645/ (accessed May 21, 2020).

[10] 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. http://www.ncbi.nlm.nih.gov/pubmed/8668200 (accessed September 8, 2017).

[11] C.E. Bakker, Y. de Diego Otero, C. Bontekoe, P. Raghoe, T. Luteijn, A.T. Hoogeveen, B.A. Oostra, R. Willemsen, Immunocytochemical and Biochemical Characterization of FMRP, FXR1P, and FXR2P in the Mouse, Exp. Cell Res. 258 (2000) 162–170. https://doi.org/10.1006/excr.2000.4932.

[12] J.C. Darnell, C.E. Fraser, O. Mostovetsky, R.B. Darnell, Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs, Hum. Mol. Genet. 18 (2009) 3164–3177. https://doi.org/10.1093/hmg/ddp255.

[13] W. Guo, E.D. Polich, J. Su, Y. Gao, D.M. Christopher, A.M. Allan, M. Wang, F. Wang, G. Wang, X. Zhao, Fragile X Proteins FMRP and FXR2P Control Synaptic GluA1 Expression and Neuronal Maturation via Distinct Mechanisms, Cell Rep. 11 (2015) 1651–1666. https://doi.org/10.1016/j.celrep.2015.05.013.

[14] S. Cavallaro, S. Paratore, F. Fradale, F.M.S. de Vrij, R. Willemsen, B.A. Oostra, Genes and pathways differentially expressed in the brains of Fxr2 knockout mice, Neurobiol. Dis. 32 (2008) 510–520. https://doi.org/10.1016/J.NBD.2008.09.008.

[15] C.J.M. Bontekoe, K.L. McIlwain, I.M. Nieuwenhuizen, L.A. Yuva-Paylor, A. Nellis, R. Willemsen, Z. Fang, L. Kirkpatrick, C.E. Bakker, R. McAninch, N.C. Cheng, M. Merriweather, A.T. Hoogeveen, D. Nelson, R. Paylor, B.A. Oostra, Knockout mouse model for Fxr2: a model for mental retardation, Hum. Mol. Genet. 11 (2002) 487–498. https://doi.org/10.1093/hmg/11.5.487.

[16] L.L. Kirkpatrick, K.A. McIlwain, D.L. Nelson, Alternative Splicing in the Murine and Human FXR1 Genes, Genomics. 59 (1999) 193–202. https://doi.org/10.1006/GENO.1999.5868.

[17] L. Davidovic, S. Sacconi, E.G. Bechara, S. Delplace, M. Allegra, C. Desnuelle, B. Bardoni, Alteration of expression of muscle specific isoforms of the fragile X related protein 1 (FXR1P) in facioscapulohumeral muscular dystrophy patients., J. Med. Genet. 45 (2008) 679–85. https://doi.org/10.1136/jmg.2008.060541.

[18] J.F. Coy, Z. Sedlacek, D. Bächner, H. Hameister, S. Joos, P. Lichter, H. Delius, A. Poustka, Highly conserved 3′ UTR and expression pattern of FXR1 points to a divergent gene regulation of FXR1 and FMR1, Hum. Mol. Genet. 4 (1995) 2209–2218. https://doi.org/10.1093/hmg/4.12.2209.

[19] E.J. Mientjes, R. Willemsen, L.L. Kirkpatrick, I.M. Nieuwenhuizen, M. Hoogeveen-Westerveld, M. Verweij, S. Reis, B. Bardoni, A.T. Hoogeveen, B.A. Oostra, D.L. Nelson, Fxr1 knockout mice show a striated muscle phenotype: implications for Fxr1p function in vivo, Hum. Mol. Genet. 13 (2004) 1291–1302. https://doi.org/10.1093/hmg/ddh150.

[20] L.L. Kirkpatrick, K.A. McIlwain, D.L. Nelson, Comparative Genomic Sequence Analysis of the FXR Gene Family: FMR1, FXR1, and FXR2, Genomics. 78 (2001) 169–177. https://doi.org/10.1006/geno.2001.6667.

[21] M. Edwards, M. Xu, S. Joseph, A simple procedure for bacterial expression and purification of the fragile X protein family, Sci. Rep. 10 (2020) 15858. https://doi.org/10.1038/s41598-020-72984-7.

[22] 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) 1733–1740. https://doi.org/10.1093/hmg/ddu586.

[23] I. Gantois, R.F. Kooy, Targeting fragile X., Genome Biol. 3 (2002) reviews1014. http://www.ncbi.nlm.nih.gov/pubmed/12049671.

[24] R. Valverde, L. Edwards, L. Regan, Structure and function of KH domains, FEBS J. 275 (2008) 2712–2726. https://doi.org/10.1111/j.1742-4658.2008.06411.x.

[25] P. Thandapani, T.R. O'Connor, T.L. Bailey, S. Richard, Defining the RGG/RG Motif, Mol. Cell. 50 (2013) 613–623. https://doi.org/10.1016/j.molcel.2013.05.021.

[26] A. Ramos, D. Hollingworth, S. Adinolfi, M. Castets, G. Kelly, T.A. Frenkiel, B. Bardoni, A. Pastore, The Structure of the N-Terminal Domain of the Fragile X Mental Retardation Protein: A Platform for Protein-Protein Interaction, Structure. 14 (2006) 21–31. https://doi.org/10.1016/J.STR.2005.09.018.

[27] M.A. Adams-Cioaba, Y. Guo, C. Bian, M.F. Amaya, R. Lam, G.A. Wasney, M. Vedadi, C. Xu, J. Min, Structural Studies of the Tandem Tudor Domains of Fragile X Mental Retardation Related Proteins FXR1 and FXR2, PLoS One. 5 (2010) e13559. https://doi.org/10.1371/journal.pone.0013559.

[28] A.I. Järvelin, M. Noerenberg, I. Davis, A. Castello, The new (dis)order in RNA regulation, (2016). https://doi.org/10.1186/s12964-016-0132-3.

[29] S. Calabretta, S. Richard, Emerging Roles of Disordered Sequences in RNA-Binding Proteins, Trends Biochem. Sci. 40 (2015) 662–672. https://doi.org/10.1016/j.tibs.2015.08.012.

[30] F. Tamanini, L.L. Kirkpatrick, J. Schonkeren, L. van Unen, C. Bontekoe, C.

Bakker, D. Nelson, H. Galjaard, B. a Oostra, a T. Hoogeveen, The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins., Hum. Mol. Genet. (2000). https://doi.org/10.1093/hmg/9.10.1487.

[31] R. Tan, A.D. Frankel, A novel glutamine-RNA interaction identified by screening libraries in mammalian cells., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 4247–52. https://doi.org/10.1073/pnas.95.8.4247.

[32] 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 WGGA sequences, Hum. Mol. Genet. 23 (2014) 5479–5491. https://doi.org/10.1093/hmg/ddu272.

[33] S.A. Whitman, C. Cover, L. Yu, D.L. Nelson, D.C. Zarnescu, C.C. Gregorio, Desmoplakin and talin2 are novel mRNA targets of fragile X-related protein-1 in cardiac muscle, Circ. Res. (2011). https://doi.org/10.1161/CIRCRESAHA.111.244244.

[34] J. Garnon, C. Lachance, S. Di Marco, Z. Hel, D. Marion, M.C. Ruiz, M.M. Newkirk, E.W. Khandjian, D. Radzioch, Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level., J. Biol. Chem. 280 (2005) 5750–63. https://doi.org/10.1074/jbc.M401988200.

[35] X.-L. Xu, R. Zong, Z. Li, M.H.U. Biswas, Z. Fang, D.L. Nelson, F.-B. Gao, FXR1P but not FMRP regulates the levels of mammalian brain-specific microRNA-9 and microRNA-124., J. Neurosci. 31 (2011) 13705–9. https://doi.org/10.1523/JNEUROSCI.2827-11.2011.

[36] 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/10.1101/gad.1276805.

[37] 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. https://doi.org/10.1016/s0092-8674(01)00566-9.

[38] C. Schaeffer, B. Bardoni, J.L. Mandel, B. Ehresmann, C. Ehresmann, H. Moine, The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif., EMBO J. 20 (2001) 4803–13. https://doi.org/10.1093/emboj/20.17.4803.

[39] 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 β-turn in the RGG motif of FMRP, Proc. Natl. Acad. Sci. 112 (2015) E5391–E5400. https://doi.org/10.1073/pnas.1515737112.

[40] 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/10.1038/nsmb.2064.

[41] J. Song, J.-P. Perreault, I. Topisirovic, S. Richard, RNA G-quadruplexes and their potential regulatory roles in translation., Transl. (Austin, Tex.). 4 (2016) e1244031. https://doi.org/10.1080/21690731.2016.1244031.

[42] 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. Mecenas, 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.F. 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) 172–177. https://doi.org/10.1038/nature12311.

[43] 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, T. Tuschl, FMRP targets distinct mRNA sequence elements to regulate protein expression., Nature. 492 (2012) 382–6. https://doi.org/10.1038/nature11737.

[44] S. Vasudevan, J.A. Steitz, AU-Rich-Element-Mediated Upregulation of Translation by FXR1 and Argonaute 2, Cell. 128 (2007) 1105–1118. https://doi.org/10.1016/J.CELL.2007.01.038.

[45] A.B. Herman, C.N. Vrakas, M. Ray, S.E. Kelemen, M.J. Sweredoski, A. Moradian, D.S. Haines, M. V. Autieri, FXR1 Is an IL-19-Responsive RNA-Binding Protein that Destabilizes Pro-inflammatory Transcripts in Vascular Smooth Muscle Cells, Cell Rep. (2018). https://doi.org/10.1016/j.celrep.2018.07.002.

[46] J. George, Y. Li, I.P. Kadamberi, D. Parashar, S.-W. Tsaih, P. Gupta, A. Geethadevi, C. Chen, C. Ghosh, Y. Sun, S. Mittal, R. Ramchandran, H. Rui, G. Lopez-Berestein, C. Rodriguez-Aguayo, G. Leone, J.S. Rader, A.K. Sood, M. Dey, S. Pradeep, P. Chaluvally-Raghavan, RNA-binding protein FXR1 drives cMYC translation by recruiting eIF4F complex to the translation start site., Cell Rep. 37 (2021) 109934. https://doi.org/10.1016/j.celrep.2021.109934.

[47] 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. https://doi.org/10.1093/nar/29.11.2276.

[48] B. V, J. P, C. S, D. JC, O. WT, T. SA, J. X, F. Y, W. KD, K. JD, D. RB, W. ST, Microarray Identification of FMRP-associated Brain mRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome, Cell. 107 (2001). https://doi.org/10.1016/S0092-8674(01)00568-2.

[49] N. Sonenberg, A.G. Hinnebusch, J. Pecreaux, M.J. Gilchrist, A. Clark, I. Davis, D.S. Johnston, N. Sonenberg, N.H. Colburn, F.M. White, M.B. Yaffe, Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets, Cell. 136 (2009) 731–745. https://doi.org/10.1016/j.cell.2009.01.042.

[50] J.W.B. Hershey, N. Sonenberg, M.B. Mathews, Principles of translational control: an overview., Cold Spring Harb. Perspect. Biol. 4 (2012) a011528. https://doi.org/10.1101/cshperspect.a011528.

[51] A.V.S. Buescher, Z. Cidav, M. Knapp, D.S. Mandell, Costs of Autism Spectrum Disorders in the United Kingdom and the United States, JAMA Pediatr. 168 (2014) 721. https://doi.org/10.1001/jamapediatrics.2014.210.

[52] 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.

[53] F. Tamanini, L. Van Unen, C. Bakker, N. Sacchi, H. Galjaard, B.A. Oostra, A.T. Hoogeveen, Oligomerization properties of fragile-X mental-retardation protein (FMRP) and the fragile-X-related proteins FXR1P and FXR2P., Biochem. J. 343 Pt 3 (1999) 517–23. https://doi.org/10.1042/BJ3430517.

[54] 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. (1997). https://doi.org/10.1093/hmg/6.9.1465.

[54] E.W. Khandjian, F. Corbin, S. Woerly, F. Rousseau, The fragile X mental retardation protein is associated with ribosomes, Nat. Genet. 12 (1996) 91–93. https://doi.org/10.1038/ng0196-91.

[56] F. Tamanini, FMRP is associated to the ribosomes via RNA, Hum. Mol. Genet. 5 (1996) 809–813. https://doi.org/10.1093/hmg/5.6.809.

[57] L. Davidovic, N. Durand, O. Khalfallah, R. Tabet, P. Barbry, B. Mari, S. Sacconi, H. Moine, B. Bardoni, A Novel Role for the RNA-Binding Protein FXR1P in Myoblasts Cell-Cycle Progression by Modulating

p21/Cdkn1a/Cip1/Waf1 mRNA Stability, PLoS Genet. (2013). https://doi.org/10.1371/journal.pgen.1003367.

[58] D. Cook, M. del Rayo Sanchez-Carbente, C. Lachance, D. Radzioch, S. Tremblay, E.W. Khandjian, L. DesGroseillers, K.K. Murai, Fragile X Related Protein 1 Clusters with Ribosomes and Messenger RNAs at a Subset of Dendritic Spines in the Mouse Hippocampus, PLoS One. 6 (2011) e26120. https://doi.org/10.1371/journal.pone.0026120.

[59] G. Stefani, C.E. Fraser, J.C. Darnell, R.B. Darnell, Fragile X mental retardation protein is associated with translating polyribosomes in neuronal cells., J. Neurosci. 24 (2004) 7272–6. https://doi.org/10.1523/JNEUROSCI.2306-04.2004.

[60] Y. Zhang, J.P. O'Connor, M.C. Siomi, S. Srinivasan, A. Dutra, R.L. Nussbaum, G. Dreyfuss, The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2., EMBO J. 14 (1995) 5358. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394645/ (accessed May 21, 2020).

[61] E. Bechara, L. Davidovic, M. Melko, M. Bensaid, S. Tremblay, J. Grosgeorge, E.W. Khandjian, E. Lalli, B. Bardoni, Fragile X related protein 1 isoforms differentially modulate the affinity of fragile X mental retardation protein for G-quartet RNA structure, Nucleic Acids Res. 35 (2006) 299–306. https://doi.org/10.1093/nar/gkl1021.

[62] S. Adinolfi, A. Ramos, S.R. Martin, F. Dal Piaz, P. Pucci, B. Bardoni, J.L. Mandel, A. Pastore, The N-Terminus of the Fragile X Mental Retardation Protein Contains a Novel Domain Involved in Dimerization and RNA Binding †, Biochemistry. 42 (2003) 10437–10444. https://doi.org/10.1021/bi034909g.

[63] Bardoni, B., Schenck, A. & Louis Mandel, J. A Novel RNA-binding Nuclear Protein That Interacts With the Fragile X Mental Retardation (FMR1) Protein. *Hum. Mol. Genet.* **8**, 2557–2566 (1999).

[64] Schenck, A., Bardoni, B., Moro, A., Bagni, C. & Mandel, J. L. A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8844–9 (2001).

[65] I. Napoli, V. Mercaldo, P.P. Boyl, B. Eleuteri, F. Zalfa, S. De Rubeis, D. Di Marino, E. Mohr, M. Massimi, M. Falconi, W. Witke, M. Costa-Mattioli, N. Sonenberg, T. Achsel, C. Bagni, The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP., Cell. 134 (2008) 1042–54. https://doi.org/10.1016/j.cell.2008.07.031.

Chapter 2:

A Simple Procedure for Bacterial Expression and Purification of the Fragile X Protein Family

Abstract

The fragile X protein family consists of three RNA-binding proteins involved in translational regulation. Fragile X mental retardation protein (FMRP) is well-studied, as its loss leads to fragile X syndrome, a neurodevelopmental disorder which is the most prevalent form of inherited mental retardation and the primary monogenetic cause of autism. Fragile X related proteins 1 and 2 (FXR1P & FXR2P) are autosomal paralogs of FMRP that are involved in promoting muscle development and neural development, respectively. There is great interest in studying this family of proteins, yet researchers have faced much difficulty in expressing and purifying the full-length versions of these proteins in sufficient quantities. We have developed a simple, rapid, and inexpensive procedure that allows for the recombinant expression and purification of full-length human FMRP, FXR1P, and FXR2P from *Escherichia coli* in high yields, free of protein and nucleic acid contamination. In order to assess the proteins' function after purification, we confirmed their binding to pseudoknot and G-quadruplex forming RNAs as well as their ability to regulate translation *in vitro*.

Introduction

The fragile X protein (FXP) family consists of three RNA-binding, ribosome-associating proteins involved in translational regulation: fragile X-related protein 1 (FXR1P), fragile X-related protein 2 (FXR2P), and the most well-known, fragile X mental retardation protein (FMRP) [1-4]. FMRP's role in translation repression has been studied extensively, as loss of FMRP expression results in a neurodevelopmental disorder called fragile X syndrome (FXS), the most prevalent form of inherited intellectual disability, and the primary monogenic cause of autism spectrum disorders [5-7]. FXS predominantly results from a CGG trinucleotide repeat expansion in the 5' untranslated region (UTR) of the *FMR1* gene [6-7]. The expanded repeats are hypermethylated causing transcriptional silencing of the *FMR1* gene, leading to a deficiency or absence of FMRP [6-9]. Patients with this disorder may experience seizures, hyperactivity, anxiety, and poor language development [7]. On a cellular level, patients with FXS possess a greater density of dendritic spines, and increased numbers of long and immature-shaped spines [7]. It is estimated that 1/5,000 males and 1/4,000-8,000 females possess the full FXS mutation [7].

While perhaps lesser known, FMRP's autosomal paralogs FXR2P and FXR1P are also of interest for their role in translational regulation [1-2,11]. FXR2P-deficient mice have impaired dendritic maturation of new neurons, with new neurons possessing shorter and less complex dendrites compared to wild-type mice [12]. These mice revealed decreased neural connectivity as new neurons with shorter dendrites connected to fewer presynaptic neurons [12]. Mice deficient in FXR2P displayed

atypical gene expression in the brain and altered behavior, such as hyperactivity, reduced sensitivity to heat stimuli, and reduced prepulse inhibition [13-14].

FXR1P is unique among the FXPs in that three (e-g) of the seven isoforms in mice (a-g) show strong expression in cardiac and/or skeletal muscle [4,15-18]. In humans FXR1P mRNA likewise demonstrates alternative splicing and is abundant in heart and skeletal muscle tissue [1,15,17,19]. Elimination of *Fxr1* leads to neonatal lethality in mice, while reduced levels of FXR1P lead to shortened life spans and reduced limb musculature [20]. Furthermore, FXR1P expression is altered in myoblasts from patients with facioscapulohumeral muscular dystrophy [17].

The genes encoding the FXP family are highly homologous through the first 13 exons of FMRP, although FXR1P and FXR2P lack sequences corresponding to exons 11 and 12 of FMRP [21]. After exon 13 of FMRP, the sequences of the three proteins diverge considerably [21]. This suggests that the three proteins likely arose from multiple gene duplications of a common ancestral gene [21]. The amino acid sequences of these proteins display a high similarity over the first 58-70% of their sequences, but a lower similarity thereafter **(Figure 2.1).** Additionally, all three proteins possess RNA-binding domains of interest: three well-conserved K homology (KH) domains, and an arginine-glycine-glycine (RGG) motif with poor conservation [11,22-23]. Another noteworthy feature of the FXPs is their C-terminal intrinsically disordered region (IDR) which constitutes ~30-43% of the entire protein sequence but has lower sequence conservation **(Figure 2.2).** IDRs are enriched in RNA-binding proteins compared to the entire human proteome and can promote protein aggregation and phase transitions, serve as sites for post-translational modifications or protein-protein

interactions, and bind to RNA both specifically and non-specifically [24-25]. The high sequence conservation in the N-termini of the FXP family suggests they exhibit some functional redundancy, while their divergent C-termini likely contribute to their unique functions.

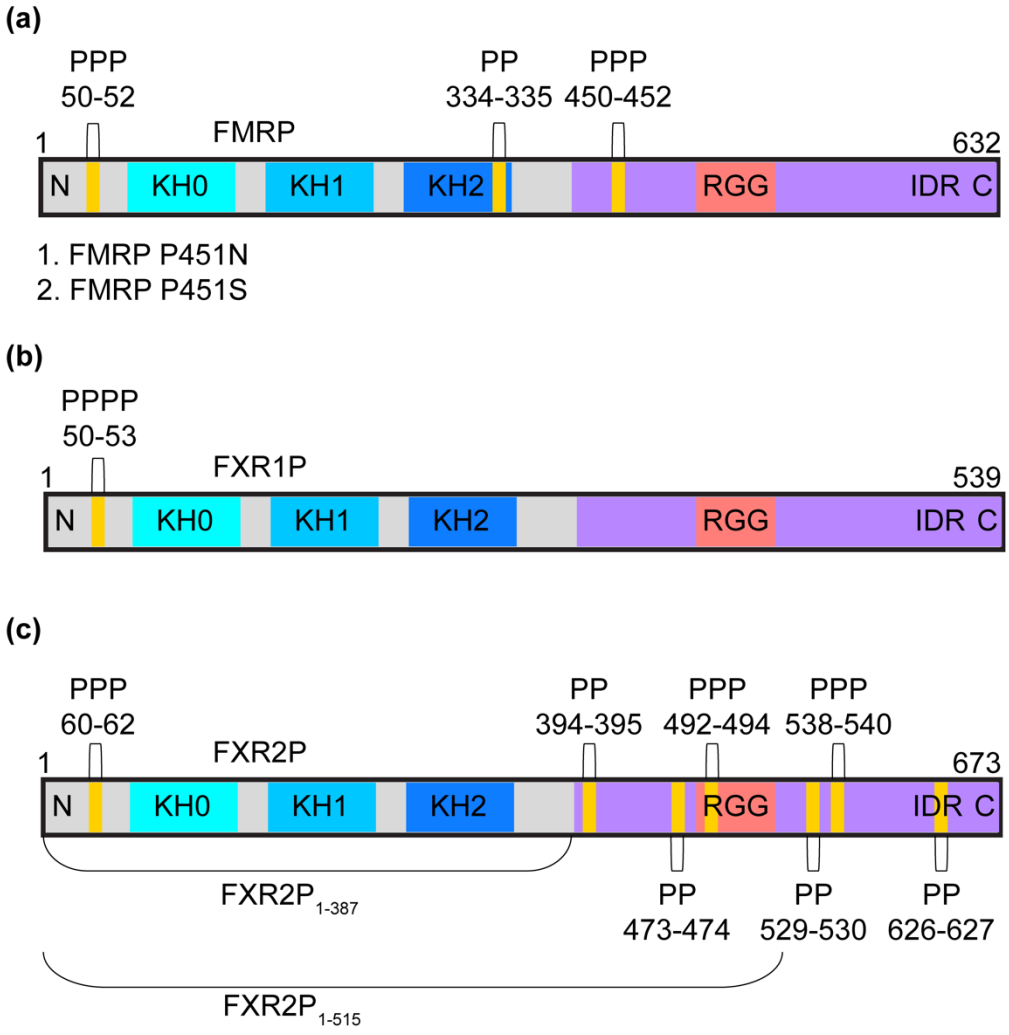


Figure 2.1: **Domains and diprolyl/polyproline stretches of the fragile X protein family.** (**A**) FMRP isoform 1, (**B**) FXR1P isoform 2, and (**C**) FXR2P with relevant protein domains and motifs labeled. Diprolyl/polyproline locations are displayed, and shortened FXR2P constructs and FXR2P/FMRP mutants are listed. The fragile X proteins are well-conserved (72–77% identity) through the sequence RQIG of each protein (located after KH2 domain), but the sequences diverge after this point (31–61% identity). Sequence identities were determined in MUSCLE [63].

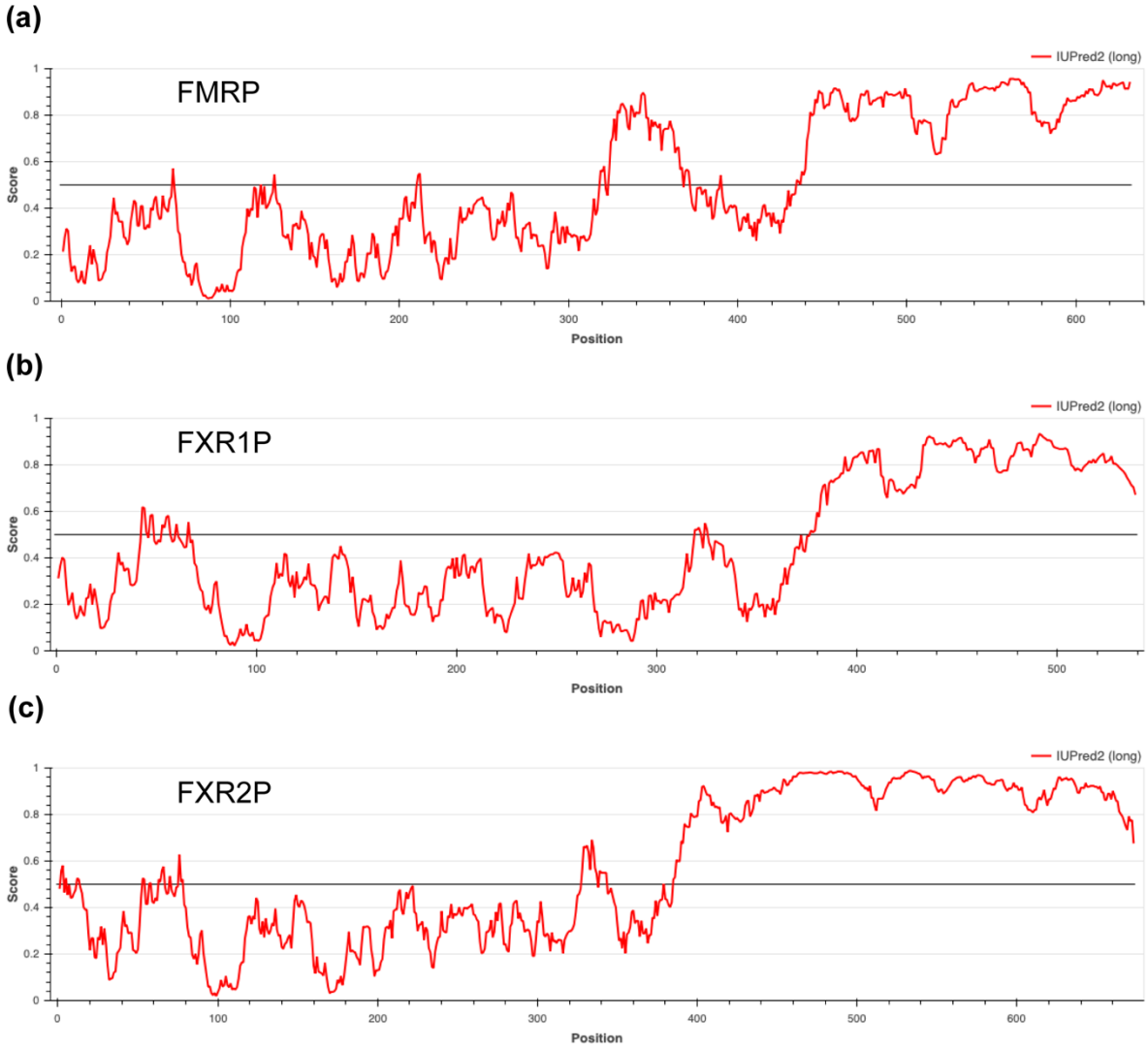


Figure 2.2: The Fragile X Proteins Possess Disordered C-termini. Fragile X protein disordered region predictions from IUPred2A using long disorder settings for (A) human FMRP isoform 1, (B) human FXR1P isoform 2, and (C) human FXR2P [59].

One area of research has focused on determining protein interaction partners of the FXP family, including the effects these interactions have on the proteins' ability to bind RNA and regulate translation. All three proteins are capable of forming homomers, as well as heteromeric complexes with either of the other FXPs [2]. The formation of such heteromeric complexes has been proposed as a mechanism through which the FXP family's functions are regulated. In fact, it was discovered that when certain isoforms of FXR1P form a heterodimer with FMRP, it inhibits the affinity of FMRP for G-quartet RNA, an RNA structure bound by FMRP [26]. Evidence suggests that the N-terminal region of FMRP contains a protein-protein interaction motif involved in its ability to dimerize as well as interact with other proteins, such as nuclear FMRP interacting protein (NUFIP) and cytoplasmic FMRP interacting proteins 1 and 2 (CYFIP1/2) [27-29]. Interestingly, despite the high homology of the FXPs in the N-terminal region, the interactions between FMRP and CYFIP1 or NUF1P appear unique, whereas CYFIP2 can also interact with FXR1P and FXR2P [28-29]. Such interactions appear to be important for modulating the translation regulation activity of the FXP family. For example, recruitment of the CYFIP1-FMRP complex to mRNAs by transactive response DNA-protein 43 kDa (TDP-43) was found to repress translation initiation [30]. While some of these assays are performed *in vivo*, or with protein purified from eukaryotic systems, others have utilized protein purified from *E. coli* for such assays [2,26-27,30].

Additionally, much emphasis has been placed on proposing mechanisms of translational regulation for the FXP family, with a particular focus on determining their mRNA targets. Many studies have attempted to identify and validate the mRNA targets

of FMRP, while several papers have identified targets of FXR1P and FXR2P [12,31-37]. Although there appears to be overlap in the mRNA targets of the FXP family, there is evidence that each protein has unique mRNA targets [11-12,31,34,36,38]. In order to validate, analyze, or compare the mRNA targets of the FXP family, researchers often test the direct binding of each protein to its mRNA targets *in vitro*. These studies allow researchers to identify binding sites within a target mRNA or test binding to *in vitro* selected RNAs, leading to the identification of sequence motifs or structural features the proteins may recognize *in vivo* [39-41]. Such studies have identified G-quadruplexes and kissing complexes as RNA features recognized by FMRP [11,39-41]. Thus, it is important to purify these proteins in sufficient quantities, with sufficient purity for *in vitro* assays.

However, researchers have faced difficulty in purifying full-length FXPs due to their poor expression, the production of truncated proteins (TPs), their tendency to aggregate and precipitate, and their instability in solution when not bound to RNA [36,42-44]. In order to overcome such obstacles, researchers have implemented strategies such as plasmids with tRNAs for rare codons to improve expression, extensive and stringent washes to remove contaminant proteins and TPs, purification from inclusion bodies, and purification under denaturing conditions which requires protein refolding and often lengthy dialysis steps [36,43,45]. Many researchers have also purified specific regions or domains instead of purifying the full-length proteins [11,39,42,45-46].

Others have purified the FXPs from mammalian cells or the SF9/baculovirus system, but these systems are more expensive and time consuming than purifying

from *E. coli*, and yields can still be low [26,28,32,36,41,45]. One group noted that when purifying from HEK293 cells, it was challenging to obtain high yields of full-length human FMRP, or to obtain the protein at concentrations above ~1 μM, noting that this could be due to low expression or a tendency of the protein to precipitate at higher concentrations [47]. Purification from *E. coli* produces proteins lacking post-translational modifications unique to mammalian systems, which have been proposed to have a role in the FMRP's function [48-49]. Thus, our method would not be suitable for researchers who desire to characterize the function of the FXPs with the post-translational modifications that occur in mammalian systems. However, our purification from *E. coli* would be advantageous for those who do not require post-translational modifications for their *in vitro* assays, or for those who wish to analyze the effects of specific, individual post-translational modifications.

Our purification protocol improves upon previous methods by allowing the FXP family to be purified using a single, simple protocol. This protocol is fast and inexpensive as the proteins are all recombinantly expressed and purified in *E. coli*, and the materials required are available in most biochemistry laboratories. Briefly, mutations were implemented to disrupt ribosomal stalling proline-rich motifs within the protein sequences. These mutations in tandem with a maltose-binding protein (MBP) tag dramatically boosted the expression of the proteins. The mutations also reduced the production of TPs. An ammonium sulfate precipitation step removed the majority of protein contaminants, while the use of a heparin column removed remaining protein contaminants, TPs, and nucleic acid contamination. The final protein samples were pure and obtained in high yields of 1-9 mg from 2 L of culture. Finally, the purified

proteins bound to G-quadruplex and kissing complex RNAs and inhibited translation *in vitro*, demonstrating that they are functional.

Results

2.1 Expression of Recombinant Fragile X Proteins

We initially attempted to purify FXR2P from *E. coli* using an N-terminal 6X His-tag, and in doing so, encountered extremely poor expression of the full-length protein. In fact, the primary protein obtained was *E. coli* bifunctional polymyxin resistance protein ArnA, which has a similar molecular weight (74 kDa) to His_6-FXR2P (75 kDa). Furthermore, ArnA forms a hexamer with surface-exposed patches of histidine residues that bind to nickel beads [50]. To improve recombinant expression, codon optimized sequences were purchased for human FXR2P and FMRP (isoform 1); FXR1P (isoform 2) expression was sufficient, so a codon optimized sequence was not used. All three genes were cloned into standard protein expression vectors and transformed into *E. coli* Rosetta 2(DE3)pLysS for expression tests.

After codon optimization, the expression of FXR2P was still low, so several fusion tags were tested to boost protein expression. Only an N-terminal SUMO or MBP tag seemed to boost the expression of FXR2P, which is supported by the observation that MBP and thioredoxin (TrX) tags are the best N-terminal tags for promoting protein solubility [51]. There is a decreasing likelihood of soluble expression of mammalian proteins as their molecular weight increases, while the presence of low complexity

regions within a protein correlates with reduced soluble expression in *E. coli,* which could both explain FXR2P's poor expression without the MBP tag [51]. Although the His_6-MBP tag is large (~44 kDa), in some cases it promotes the proper folding of the attached protein into the biologically active conformation [52]. Furthermore, it did not contribute to the RNA-binding or translation regulation capabilities of the FXPs. Moreover, it may be possible to cleave the MBP-tag after purification (data not shown). Thus, we selected the MBP tag for our studies.

The MBP tag boosted expression of FXR2P, but we observed the production of many TPs, which we hypothesized were a result of diprolyl and polyproline stretches within FXR2P **(Figures 2.1C and 2.3C).** As the ring structure of proline makes it a poor peptide bond donor and acceptor, two or more consecutive prolines can cause ribosomes to stall during translation [53-54]. Due to nascent chain-mediated stalling of ribosomes, ribosomal rescue mechanisms may release the ribosomes and unfinished proteins from the mRNA chain, leading to the TPs we observe [55-56].

To enhance expression of the full-length protein and reduce the production of TPs we attempted two approaches (1) co-expression with elongation factor P (EF-P) which alleviates ribosomal stalling in short proline-rich motifs by stimulating peptide bond formation and (2) mutations to disrupt the diprolyl and polyproline motifs [57]. Although co-expression with EF-P enhanced FXR2P expression, we did not see similar results with FMRP **(Figure 2.4).** Furthermore, expression with EF-P did not appear to alleviate the production of TPs. In contrast, mutations dramatically boosted expression of full-length FXR2P and FMRP while reducing the production of TPs **(Figure 2.3 A&C).**

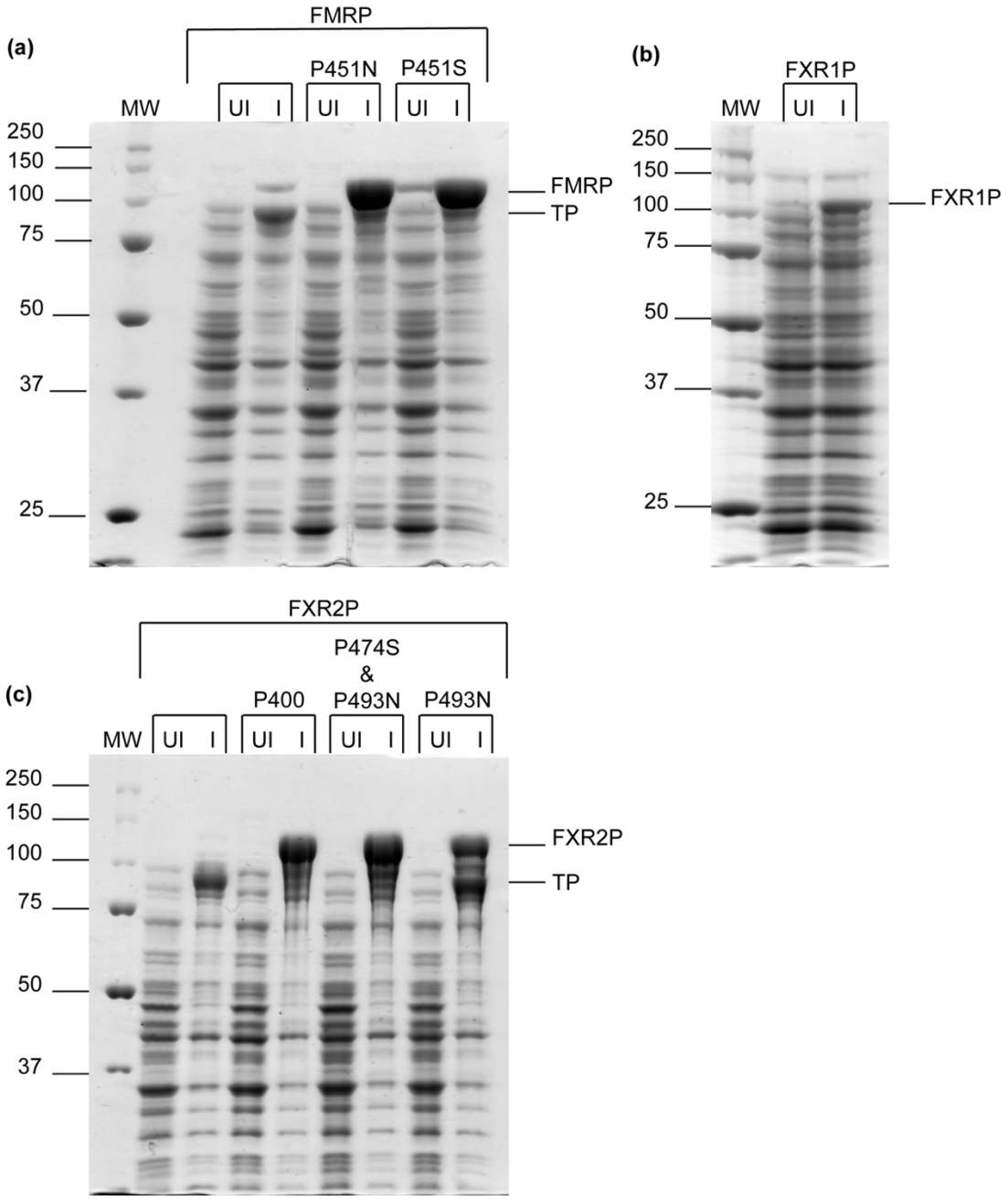


Figure 2.3: Mutations Enhance the Expression of FMRP and FXR2P. (A) His_6-MBP-FMRP (~115 kDa) and mutants, (B) His_6-MBP-FXR1P (~104 kDa), and (C) His_6-MBP-FXR2P (~117 kDa) and mutants. Lanes show the comparison between uninduced samples (UI) and samples induced with IPTG (I). FMRP P451N and FMRP P451S mutants exhibited a 7-fold increase in expression compared to wild type FMRP, and a reduction in truncated proteins (TP). FXR1P was not mutated, and a non-codon optimized sequence was used, which may explain the lower expression compared to FMRP/FXR2P mutants. FXR2P P474S & P493N was selected as it had the highest expression (130-fold greater than wild type FXR2P), less TPs than FXR2P P493N (100-fold greater), and fewer mutations than FXR2P P400 (120-fold greater).

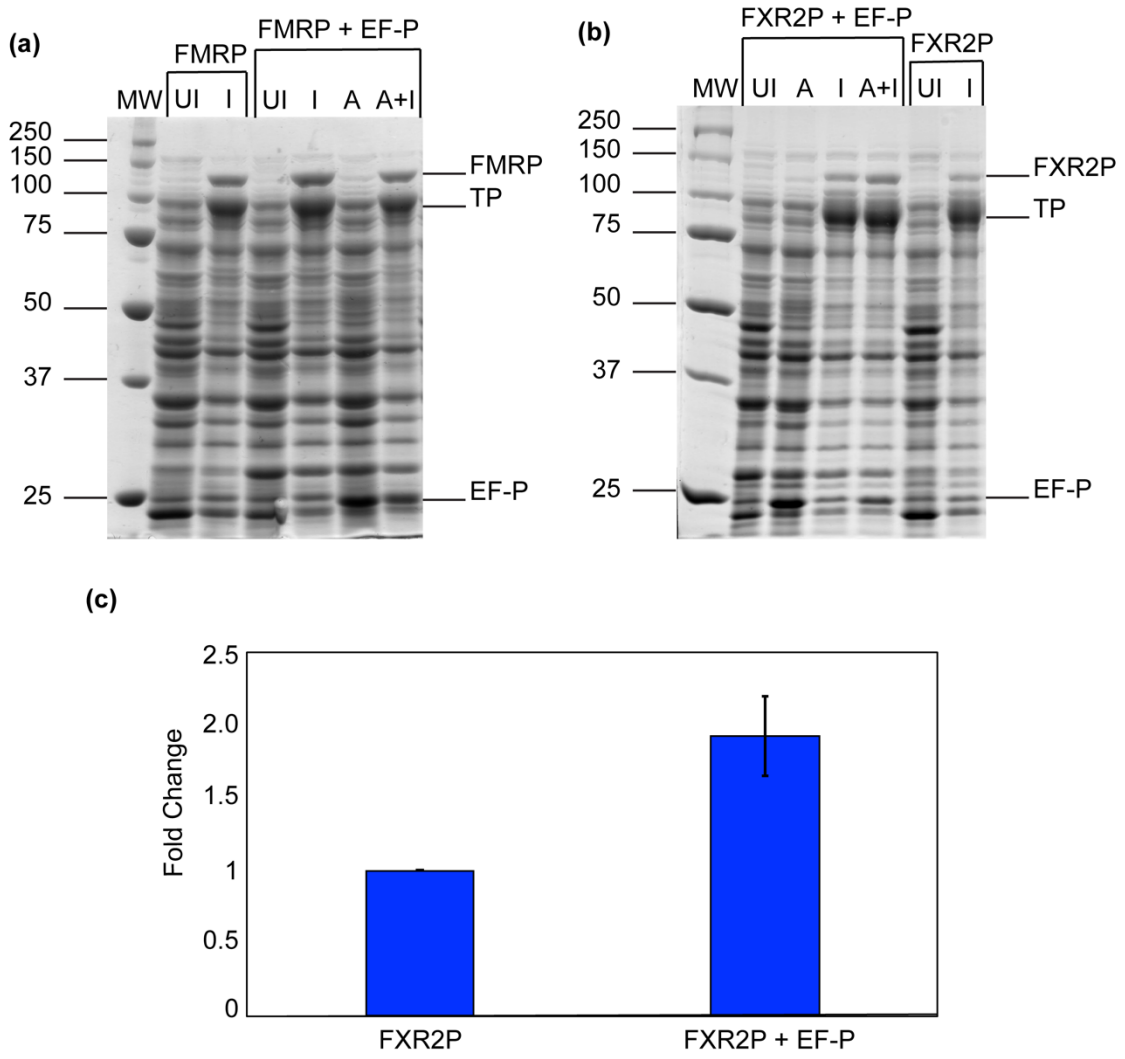


Figure 2.4: Co-expression with EF-P Enhances Expression of Full-length FXR2P. Co-expression with EF-P (~21 kDa) does not appear to enhance (A) FMRP (~115 kDa) expression but appears to enhance (B) FXR2P (~117 kDa) expression. Arabinose (A) was used to induce EF-P expression, and IPTG (I) to induce FMRP or FXR2P expression. Lanes show the comparison between uninduced (UI), arabinose only (A), IPTG only (I), or samples induced with arabinose and IPTG (A + I). Uninduced and IPTG only conditions are shown for cells containing the EF-P plasmid (FMRP/FXR2P + EF-P), and cells without (FMRP/FXR2P). (C) Comparison of full-length FXR2P expression with and without EF-P co-expression. Co-expression with EF-P led to a 1.93 ± 0.27-fold increase in FXR2P expression. Error and error bars represent the standard deviation.

In creating mutations, we aimed to preserve the original protein sequence as much as possible **(Figure 2.5).** We therefore created and compared the expression of two shortened constructs of FXR2P and seven mutants, which enabled us to determine that the polyproline stretch from residues 492-494 has the greatest contribution towards ribosomal stalling and the production of TPs **(Figure 2.6).** However, we found we could further enhance expression and reduce TPs by also disrupting the nearby diprolyl motif from residues 473-474. Interestingly, the diprolyl and polyproline stretches at residues 529-530 and 538-540, which are situated closer to one another, do not seem to cause much ribosomal stalling or TPs **(see FXR2P P500 mutant, Figure 2.6).** This ultimately led us to select a FXR2P double mutant (FXR2P P474S & P493N) for future purification attempts **(Figure 2.2C).**

Based on the results for FXR2P, we decided to mutate the polyproline motif in FMRP that was also located within its disordered region, residues 450-452. Mass spectrometry results of the major TP of FMRP indicated that truncation was occurring after P451, providing further evidence that the TPs are a result of ribosomal stalling at proline-rich motifs (data not shown). Of the two mutants, we selected FMRP P451S for purification attempts **(Figure 2.2A).** FXR1P did not need to be mutated as it contains only one polyproline motif, however we did observe a truncation for FXR1P **(Figures 2.1B and 2.3B).** This TP does not appear to be due to ribosomal stalling as mass spectrometry results suggested the truncation occurs within the KH1 domain (data not shown). This TP is lacking the KH2 domain and C-terminal disordered region and appeared to be more stable than full-length FXR1P.

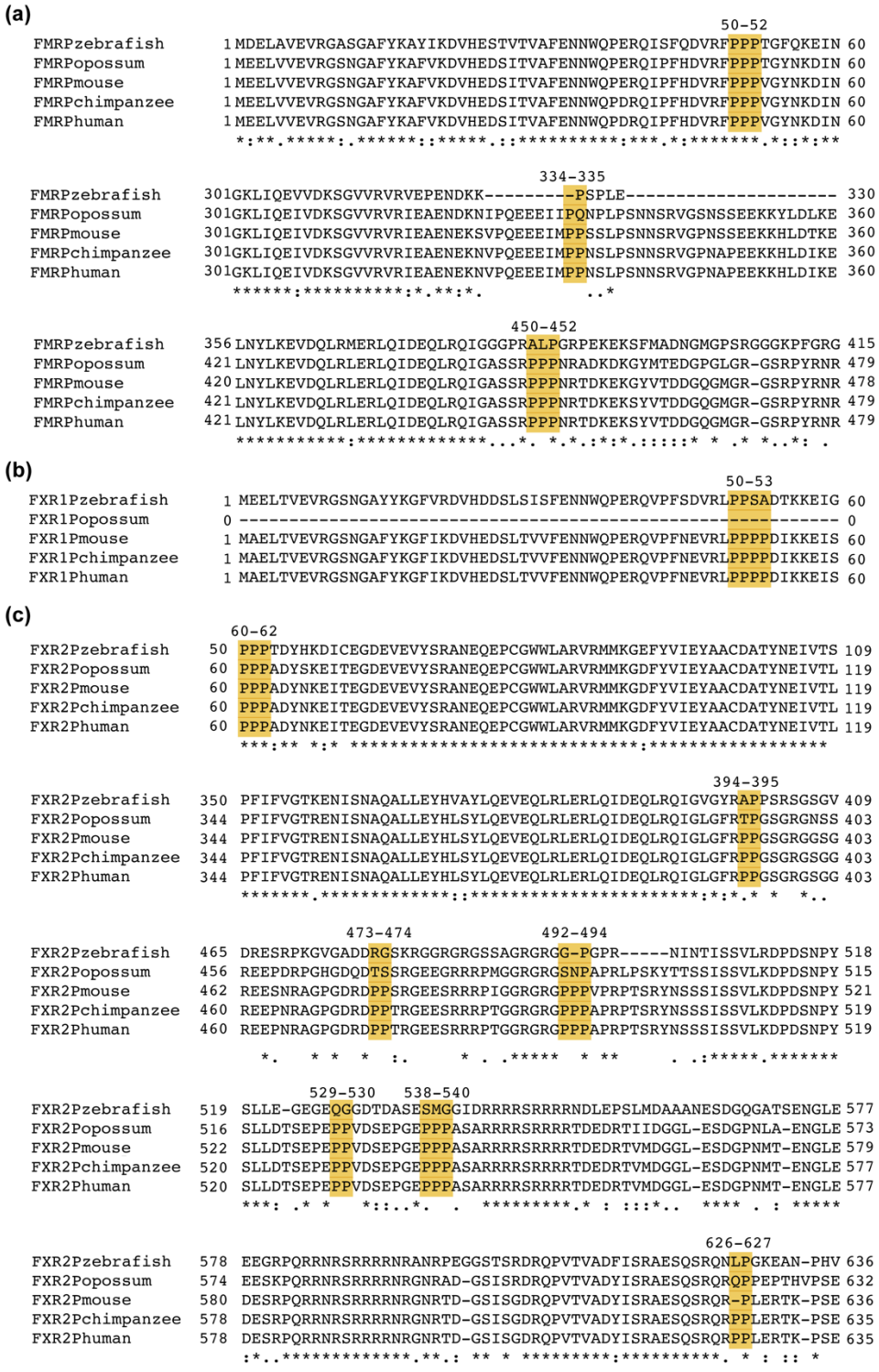


Figure 2.5: **Sequence alignments of FXPs.** Multiple sequence alignments of (A) FMRP, (B) FXR1P, and (C) FXR2P from multiple organisms, performed in MUSCLE [63]. Alignments of regions containing polyproline motifs are displayed, with the numbering of polyproline motifs referring to the position within the human sequence.

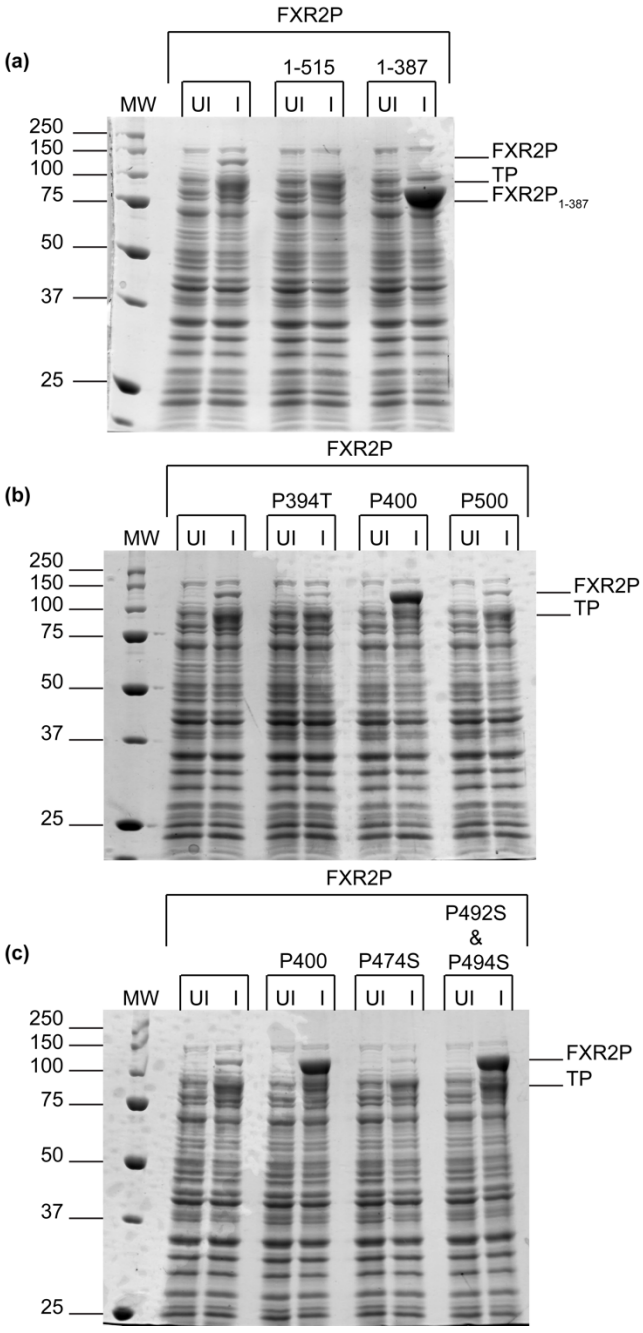


Figure 2.6: Mutating Proline-rich Regions Enhances Full-length FXR2P Expression. Expression tests of (A) shortened FXR2P constructs and (B-C) FXR2P mutants used to determine which consecutive prolines cause ribosomal stalling. Lanes show the comparison between uninduced samples (UI) and samples induced with IPTG (I). Full-length His_6-MBP-FXR2P and mutants are ~117 kDa. $FXR2P_{1-515}$ (not visibly expressed) is ~100 kDa and $FXR2P_{1-387}$ is ~87 kDa. Removing amino acids 388-673 of FXR2P ($FXR2P_{1-387}$) or mutating the polyproline stretch from 492-494 (FXR2P P492S & P494S) increases protein expression.

2.2 Purification of the Fragile X Proteins

After selecting a FXR2P and FMRP mutant for purification, we set out to identify a single expression and purification scheme for all three proteins. We grew our cells at 37°C until the OD_{600} was ~0.4, then allowed the cells to grow an additional 14-16 hours at ~14°C after induction with Isopropyl β- d-1-thiogalactopyranoside (IPTG). Initially we attempted amylose resin for protein affinity chromatography, however, we encountered poor binding to the amylose resin unless we implemented a purification step prior to batch binding, and we were unable to remove all the TPs. We therefore switched to an ammonium sulfate precipitation followed by a heparin column. After harvesting and lysing the cells and clarifying the lysate, we were pleased to discover that all three proteins could be precipitated at relatively low percentages of ammonium sulfate, while the majority of *E. coli* proteins remained in the supernatant **(Figures 2.7-2.9).** After allowing the FXP to precipitate, the sample was centrifuged, the supernatant removed, and the pellet resuspended. The resulting solution was dialyzed overnight to remove ammonium sulfate as salt must be removed prior to the heparin column. After dialysis, soluble protein was further purified through fast protein liquid chromatography (FPLC) with a heparin column. Using an increasing salt gradient from 0-1 M NaCl (or KCl) we were able to remove nucleic acid contamination, residual protein contaminants, and TPs from the FXPs **(Figure 2.10).** Interestingly, we observed that the C-terminally truncated His_6-MBP-FXR1P that is predicted to be missing the KH2 domain and disordered region (~75 kDa) did not appear to bind to the heparin column and was present predominantly in the flow-through **(Figure 2.8).** Additionally, the FXR2P TPs

began eluting prior to the full-length protein, while the full-length was predominantly eluted at high salt concentrations **(Figure 2.9).** In fact, in subsequent purifications, a step gradient was implemented to remove the TPs prior to eluting full-length FXR2P. Thus, it appears the FXPs interact with the heparin column through their C-terminal disordered regions. Through the use of a single ammonium sulfate precipitation step and a heparin column, we were able to purify all three FXPs **(Figure 2.11).**

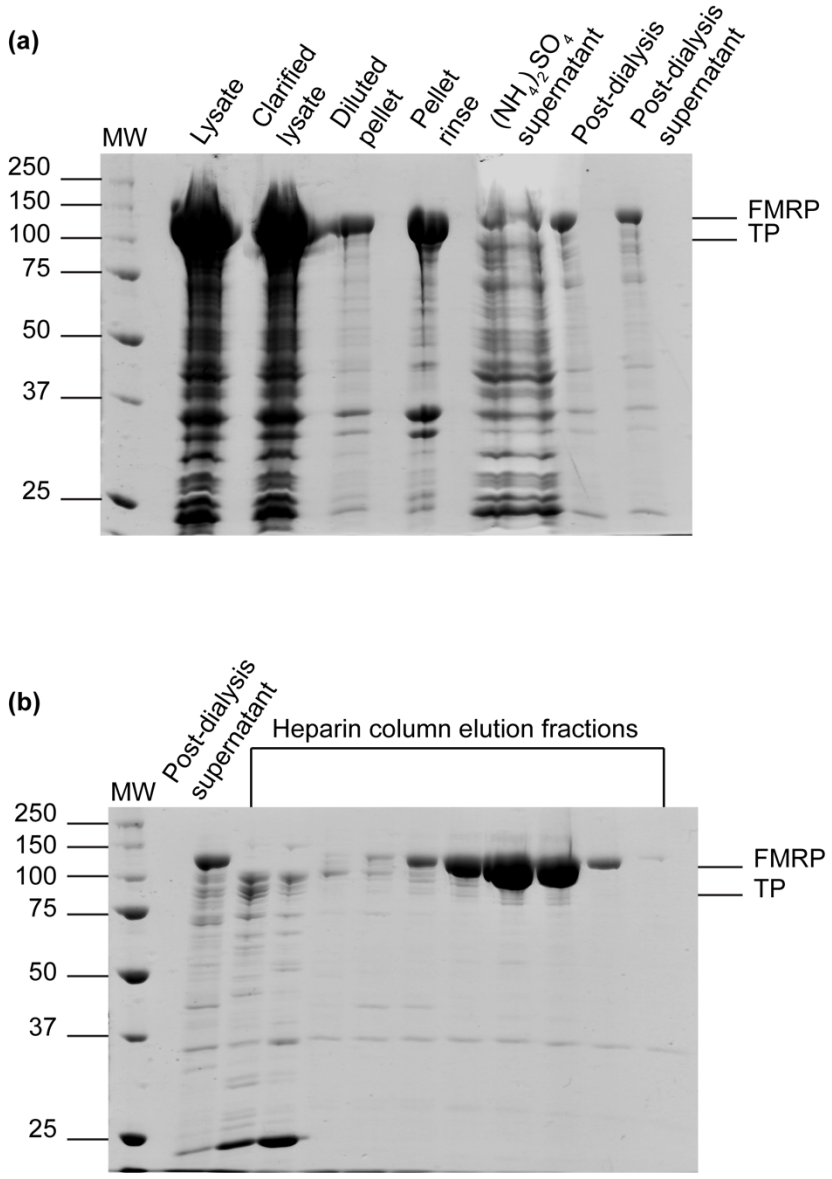


Figure 2.7: FMRP Purification. (A) His_6-MBP-FMRP P451S (~115 kDa) is obtained in the lysate and pelleted from *E. coli* proteins at 25% $(NH_4)_2SO_4$. (B) A heparin column removes nucleic acid contamination and the majority of residual *E. coli* and truncated proteins.

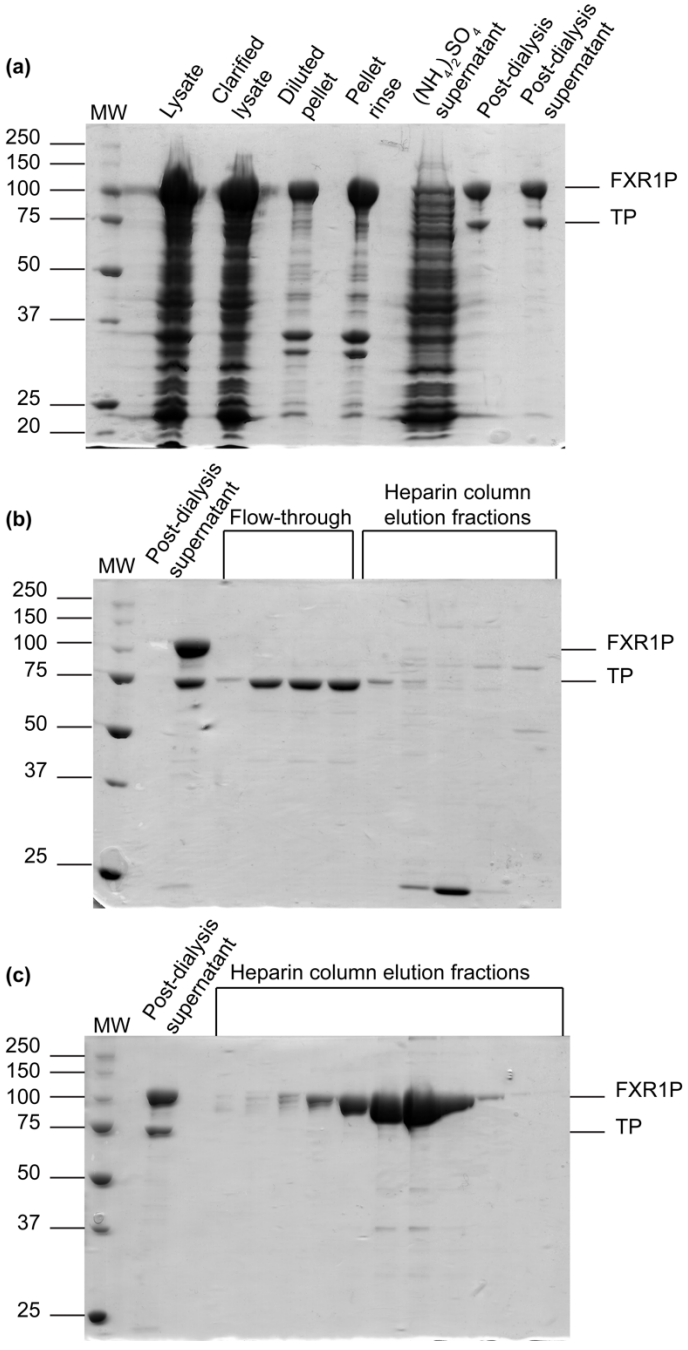


Figure 2.8: FXR1P Purification. (A) His_6-MBP-FXR1P (~104 kDa) is obtained in the lysate and pelleted from *E. coli* proteins at 20% $(NH_4)_2SO_4$. (B-C) A heparin column removes nucleic acid contamination and the majority of residual *E. coli* and truncated proteins. (B) The FXR1P truncated protein (~75 kDa) does not appear to bind to the column and is found predominantly in the flow-through.

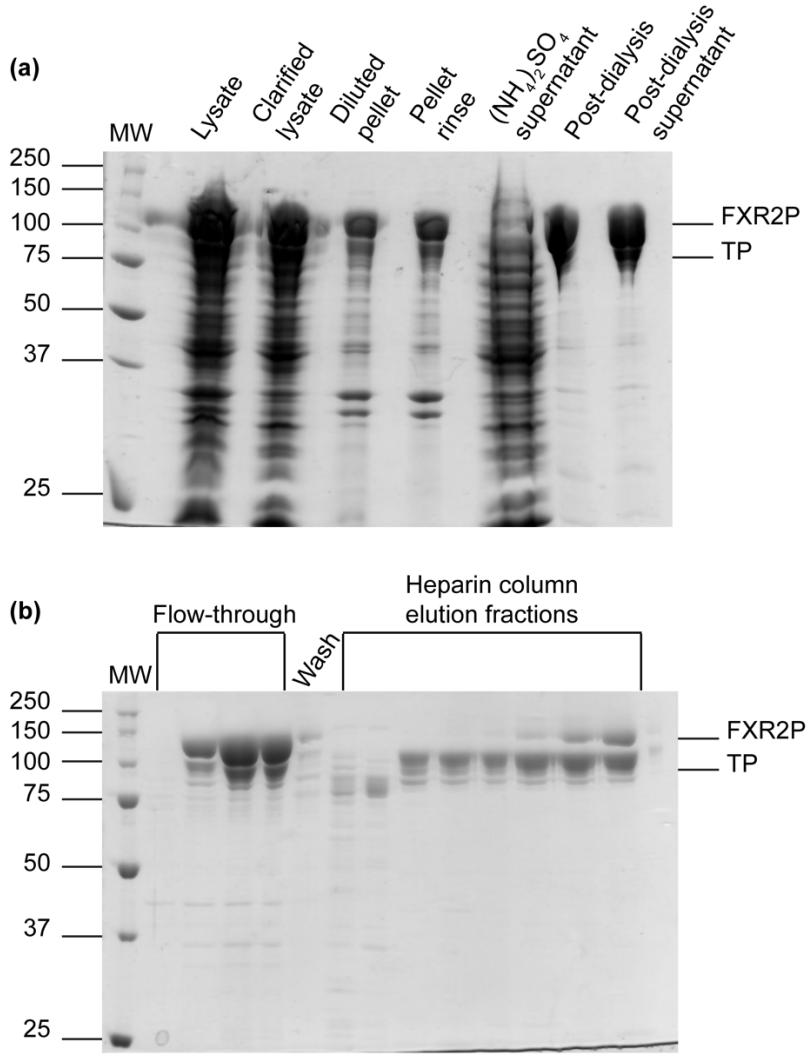


Figure 2.9: (A) His_6-MBP-FXR2P P474S & P493N (~117 kDa) is obtained in the lysate and pelleted from *E. coli* proteins at 20% $(NH_4)_2SO_4$. (B-C) A heparin column removes nucleic acid contamination and the majority of residual *E. coli* and truncated proteins. Truncated proteins elute before full-length FXR2P. (D) Diluted FXR2P elution fractions reveal that fractions eluted at higher salt concentrations contain less truncated proteins relative to full-length FXR2P.

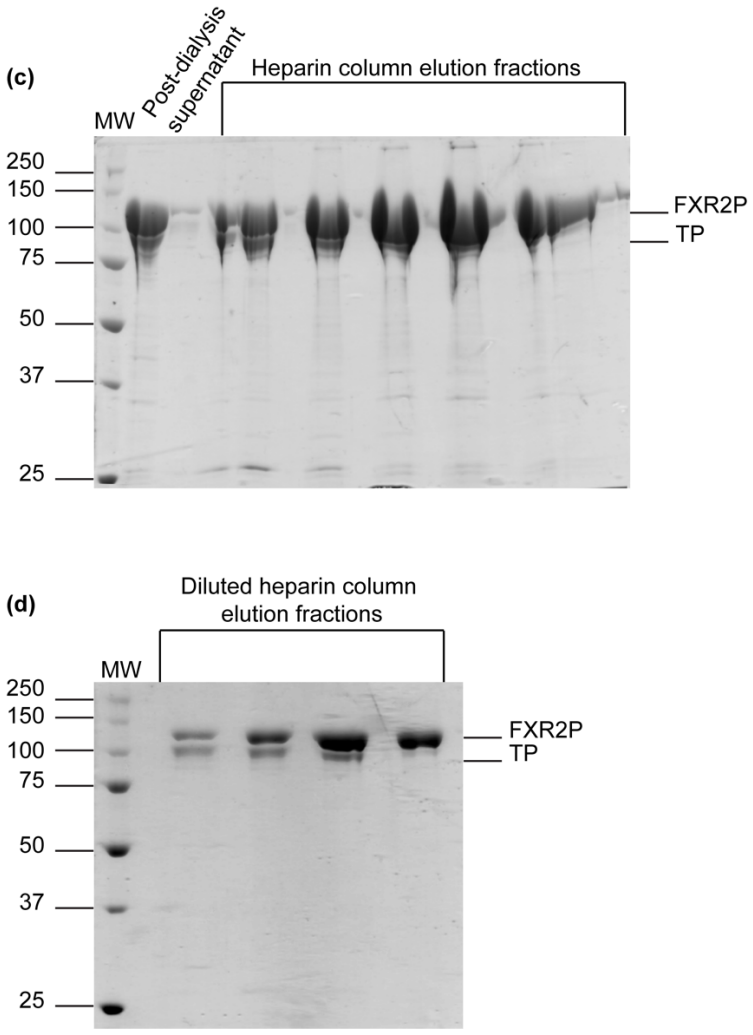


Figure 2.9 Continued.

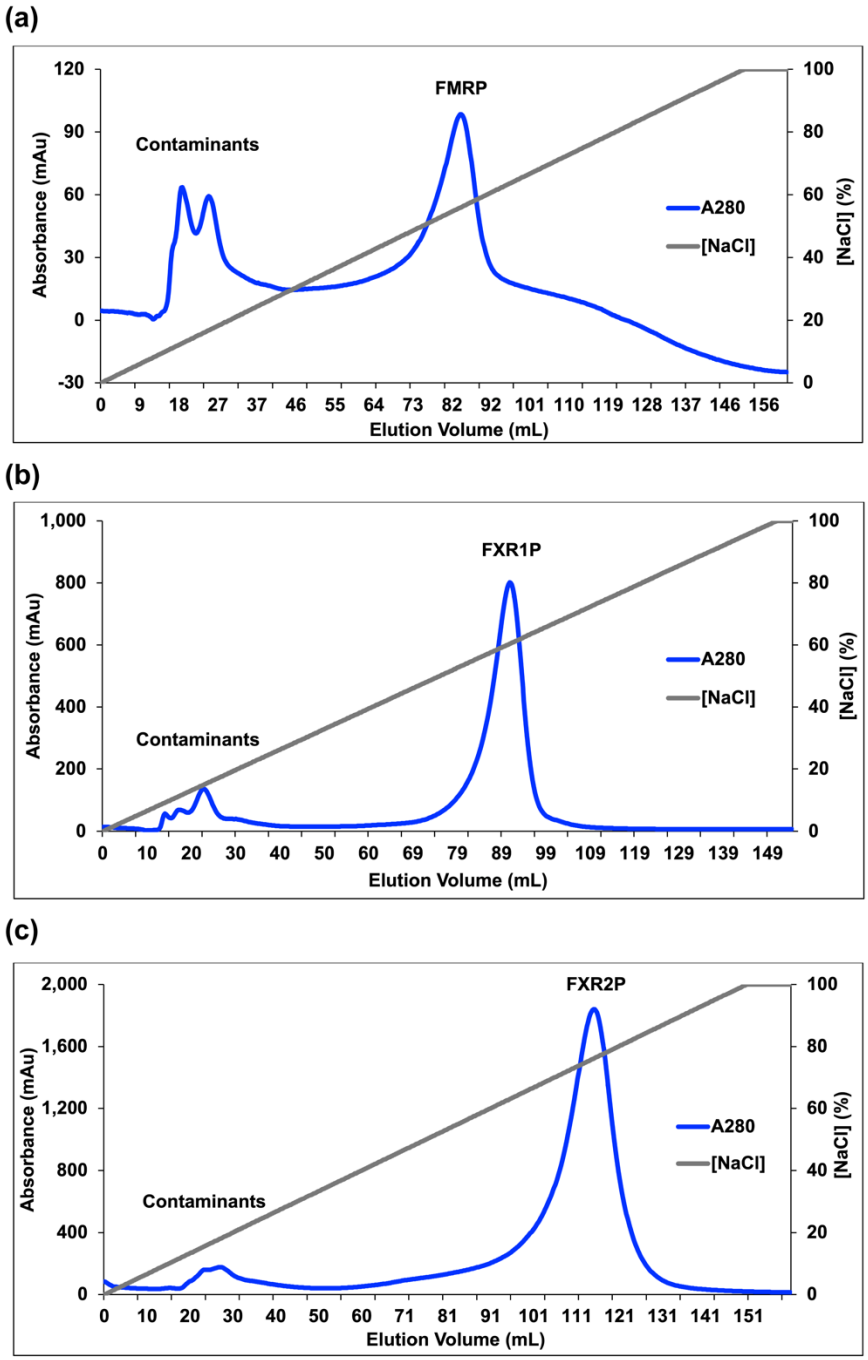


Figure 2.10: High Salt Concentrations Elute the Fragile X Proteins. Heparin column elution fractions for the fragile X proteins displaying A280 absorbance and NaCl concentration. *E. coli* protein contaminants elute at lower salt concentrations than the fragile X proteins. (A) FMRP elutes at ~500-600 mM NaCl, with the peak at ~560 mM NaCl, (B) FXR1P elutes at ~560-700 mM NaCl, with a peak at ~640 mM NaCl, and (C) FXR2P elutes over a large range, however the most full-length with the least truncated proteins elutes from ~700-830 mM NaCl with the peak at ~730 mM.

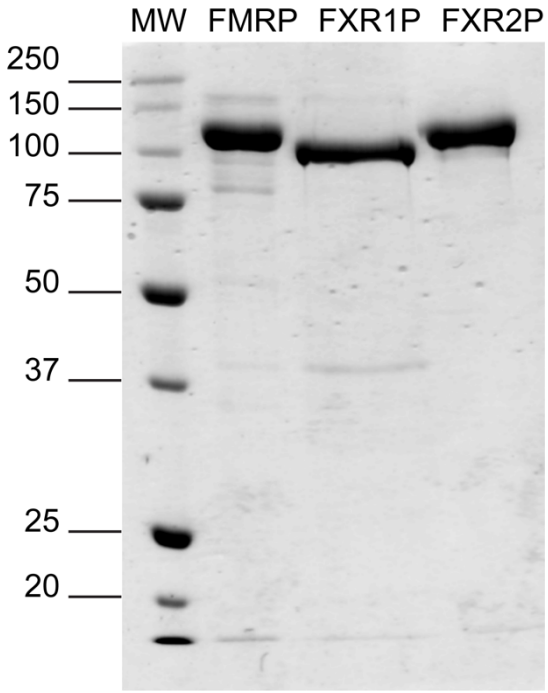


Figure 2.11: Purified Fragile X Proteins. Approximately 2 μg of each fusion protein was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). His_6-MBP-FMRP P451S (~115 kDa), His_6-MBP-FXR1P (~104 kDa), and His_6-MBP-FXR2P P474S & P493N (~117 kDa). The identity of each protein was confirmed by mass spectrometry.

2.3 Concentration and Storage of the Fragile X Proteins

After purification we concentrated FMRP and FXR1P by centrifugation. However, we observed a loss in yield, perhaps due to the protein sticking to the concentrator and due to protein precipitation. In our hands, even with a His_6-MBP tag which promotes solubility, the proteins will precipitate if concentrated too much after removal of nucleic acid contamination [51-52]. FMRP has been concentrated to ~16 μM but is not stable for long at this concentration and drops to ~3-6 μM over time. FXR1P and FXR2P precipitate more readily. We were able to concentrate FXR1P to 8 μM but precipitation began occurring at ~3 μM. FXR1P and FXR2P fractions from the heparin column at ~5-7 μM seem stable whereas fractions at ~9-13 μM had visible precipitation as they eluted. Once purified from nucleic acids, FMRP seems stable at 16 μM and FXR1P/FXR2P at ~6 μM initially, but the proteins appear to precipitate slowly over time if not stored at -80°C.

To maximize yield, we suggest researchers avoid concentrating, concentrate minimally, or concentrate right before use or storage at -80°C. When we centrifuged our concentrated protein samples after storage at 4°C or -15°C to remove precipitated protein prior to use, we noticed a decrease in concentration of the samples over time. This is likely due to precipitation. Storage at -80°C appears prevent this, although thawing may induce precipitation. The tendencies of the FXPs to form aggregates, precipitate during concentration, or not concentrate past a certain concentration have been noted by other researchers [42-44,47]. Concentrating appears to induce

aggregation and precipitation, which may have a function *in vivo*, namely in the proteins' presence in ribonucleoprotein granules [44]. In rat brain, FXR1P predominantly forms oligomers or insoluble aggregates, while monomers are nearly undetectable [45]. We therefore recommend the MBP tagged FXPs be stored at -80°C for long-term storage. Prior to use in assays, we suggest researchers centrifuge stored samples to remove insoluble aggregates then remeasure protein concentration.

2.4 Confirmation of Fragile X Protein Identities

After purifying the FXPs, the final samples were analyzed by mass spectrometry. The correct identity of each protein was confirmed with 94% sequence coverage obtained for FXR1P, and 96% for FMRP and FXR2P.

2.5 Analysis of the RNA-binding Activity of the Fragile X Proteins

The functionality of the purified proteins was verified by testing the FXP family's binding to a G-quadruplex forming RNA, a well-known target of FMRP, which is bound by the RGG motif, and a target of FXR1P [11,26,35,40-41,58]. We chose to test the proteins' binding to poly-$G_{17}U$ as our lab has identified poly-$G_{17}U$ as a G-quadruplex forming RNA [46]. For a negative control we used CR1, an RNA with no predicted G-quadruplex forming capability [46].

As predicted based on our previous observations for an N-terminally truncated FMRP and the FMRP RGG motif, all proteins showed high affinity binding to poly-$G_{17}U$,

and no binding to CR1 in the concentration range tested **(Figure 2.12A-C)** [46]. Our fluorescence anisotropy results reinforce the hypothesis that the FXPs have different affinities for mRNA targets: FXR2P showed the greatest affinity for poly- $G_{17}U$, and FXR1P the least (FXR2P $K_D = 3.1 \pm 0.4$ nM, FMRP $K_D = 5.6 \pm 0.6$ nM, and FXR1P $K_D = 11.7 \pm 1$ nM). To ensure that the His_6-MBP tag did not contribute to the observed RNA-binding, we tested its binding to poly-$G_{17}U$ and CR1 and observed no binding **(Figure 2.12D).** Furthermore, it appears that the mutation implemented to boost FMRP expression does not impair the protein's ability to bind poly-$G_{17}U$ RNA, as we obtained higher affinity binding than previously determined for an N-terminally truncated FMRP construct (R218-P632 of FMRP) or a glutathione S-transferase tagged FMRP RGG fusion protein (G531-P632 of FMRP): K_D of 14 ± 2 nM and 8.6 ± 1.2 nM, respectively [46].

After assessing the binding of the FXPs to a G-quadruplex forming RNA, we tested for binding to a loop-loop pseudoknot, or "kissing complex" RNA, ΔKC2, a shortened version of an *in vitro* selected target of FMRP called KC2 [39]. The KH2 domain of FMRP was found to be necessary and sufficient for FMRP binding to KC2 [39]. Furthermore, the KH2 domains of FMRP, FXR1P, and FXR2P bind KC2 RNA with equal affinity [11]. As binding of FMRP to KC2 is dependent on the integrity of the KH2 domain, we felt it valuable to assess the FXP family's ability to bind ΔKC2 [11,39].

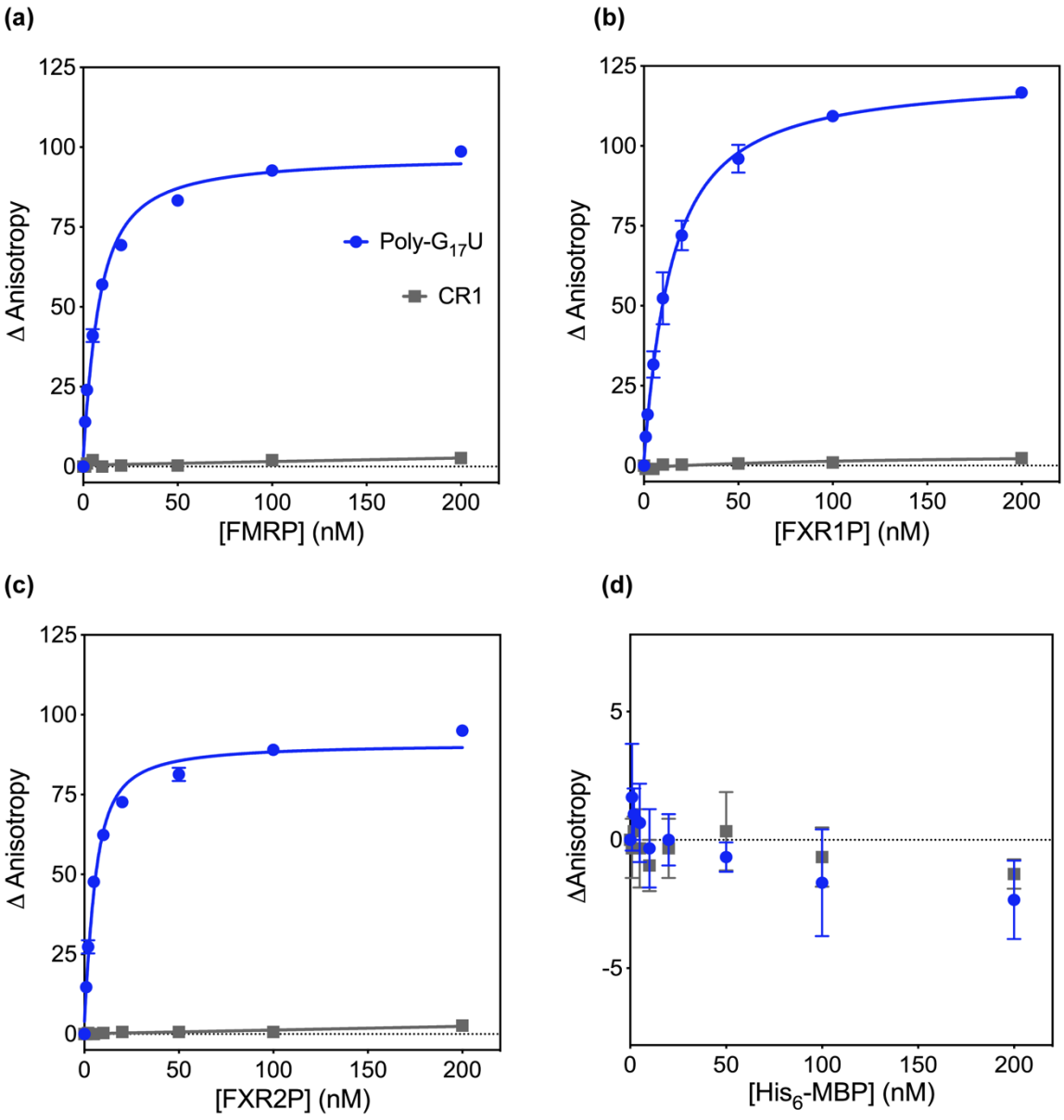


Figure 2.12: The Fragile X Proteins Bind Poly-$G_{17}U$. The RNA-binding capabilities of (A) FMRP, (B) FXR1P, and (C) FXR2P were assessed by fluorescence anisotropy. All three proteins bound to poly-$G_{17}U$ with high affinity: FMRP $K_D = 5.6 \pm 0.6$, FXR1P $K_D = 11.7 \pm 1$, and FXR2P $K_D = 3.1 \pm 0.4$. No binding was observed for CR1 in the concentration range tested. Data are from three individual trials with error bars for the standard deviation displayed (with the exception of FMRP 1 and 2 nM points for which there was only one trial). (D) His_6-MBP was tested for binding to poly-$G_{17}U$ and CR1 through fluorescence anisotropy and shows no RNA-binding capabilities.

Due to the size of our ΔKC2 (72 nucleotides), we tested for binding by electrophoretic mobility shift analysis (EMSA) using CR1 and poly-$G_{17}U$ as negative and positive controls, respectively. As predicted, all proteins showed binding to poly-$G_{17}U$ and ΔKC2, indicated by the reduction in free RNA upon the addition of protein, but not to CR1 **(Figure 2.13A).** It is worth mentioning that there are two free RNA species visible in the control lane for ΔKC2. As this is a native gel, it is likely that the faster migrating species is more compact and folded in the correct confirmation. Darnell et al. also observed two free RNA species for KC2 RNA, and only the faster migrating species was observed to shift with added protein [39].

2.6 *In vitro* Translation Regulation by the Fragile X Proteins

We further analyzed the functionality of the FXPs by testing their ability to regulate translation in an *in vitro* translation system (IVTS) comprised of rabbit reticulocyte lysate treated to reduce endogenous mRNAs. We chose *Renilla* luciferase mRNA as the reporter for protein synthesis as other researchers have used it previously with the FXPs, it has 3 G-rich sequences that may form G-quadruplex structures, and we have previously observed inhibition of this mRNA by an N-terminally truncated *Drosophila* FMRP in *Drosophila* embryo extract [35,59]. We monitored the translation of a 5' capped *Renilla* luciferase mRNA with a 3' poly (A) tail through bioluminescence and observed that all the FXPs inhibited translation, but to different extents **(Figure 2.13B).** Similar to our anisotropy results for poly-$G_{17}U$, we observed the greatest translation inhibition by FXR2P, and the least by FXR1P (FXR2P 2.06%

± 0.54 percent luciferase activity, FMRP 11.7% ± 2.1, and FXR1P 20.2% ± 4.3). As expected, the His-MBP tag did not inhibit the translation of *Renilla* luciferase mRNA (103.8% ± 7.4 percent activity). Our results suggest that our purified FXPs maintain their ability to regulate translation.

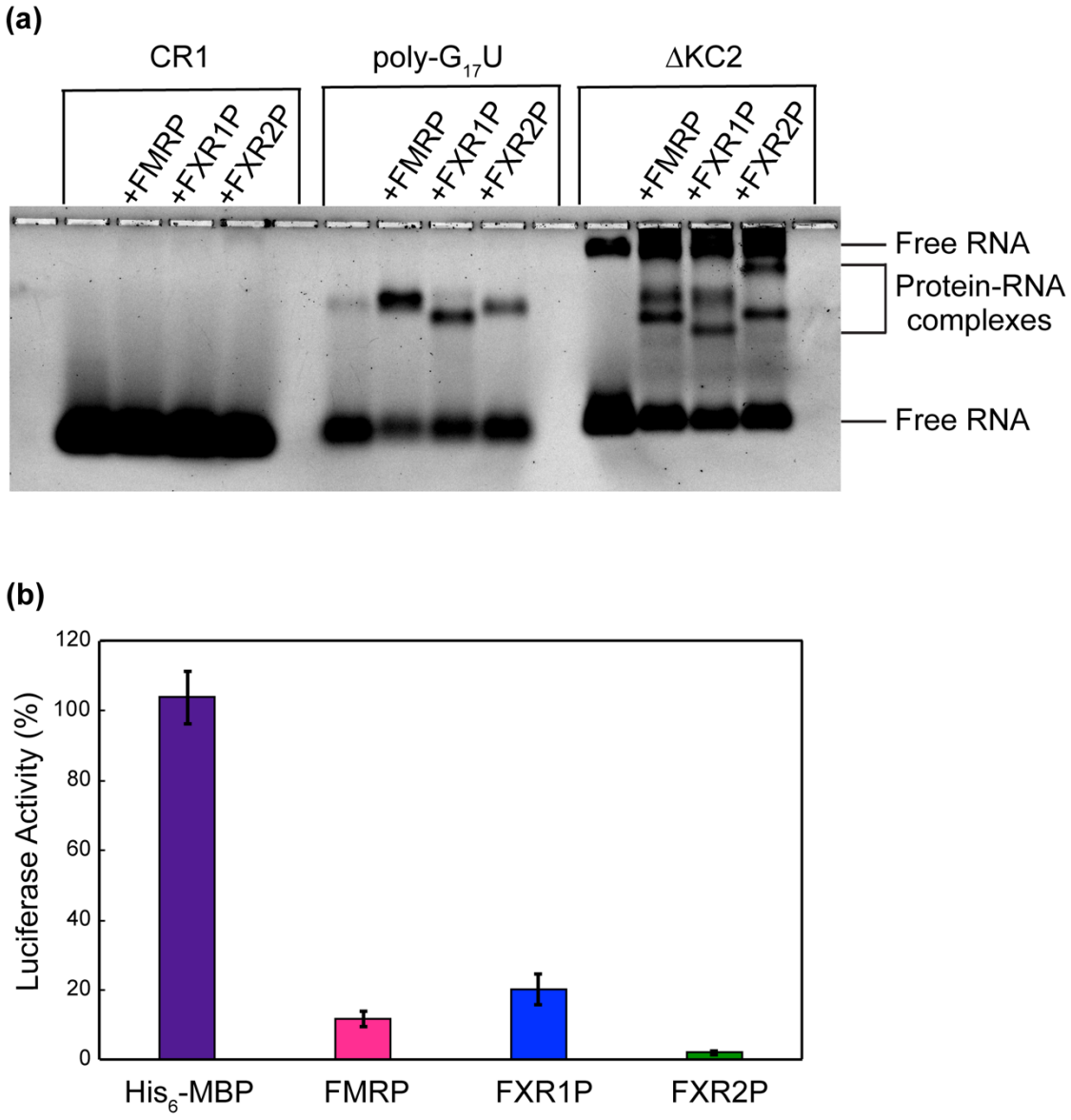


Figure 2.13: The Fragile X Proteins Bind ΔKC2 and Inhibit Translation of Renilla Luciferase mRNA. (A) Binding to ΔKC2 was assessed using an agarose EMSA with CR1 and poly-$G_{17}U$ included as negative and positive controls, respectively. (B) The addition of fragile X proteins led to a reduction in the luciferase percent activity, indicating a reduction in the translation of *Renilla* luciferase mRNA (FMRP 11.7% ± 2.1, FXR1P 20.2% ± 4.3, and FXR2P 2.06% ± 0.54). The His_6-MBP tag did not inhibit luciferase activity (103.8% ± 7.4 percent activity). For all reactions, 500 nM of protein was combined with 10 nM *Renilla* luciferase mRNA.

2.7 Discussion

Our RNA-binding and *in vitro* translation studies suggest the FXPs are functional. Thus, mutations implemented to boost expression of FMRP and FXR2P do not appear to impact their RNA-binding specificity or their ability to repress translation. We were intrigued by the results of our binding studies to poly-$G_{17}U$, as work by Darnell et. al demonstrated that the C-termini of the FXPs had differing affinity for an *in vitro* selected G-quadruplex forming RNA (sc1): FMRP bound with high affinity, FXR2P showed lower affinity and non-specific binding, while FXR1P showed no binding [11]. The results we observe may be due to the fact that sc1 was selected using FMRP, while poly-$G_{17}U$ may form a generic G-quadruplex structure recognized by all three proteins [40]. Additionally, our results may have been impacted by assessing the binding of the full-length proteins. Future assays testing the RNA-binding specificity of the full-length proteins should yield insightful results as it has been proposed that the multiple RNA-binding domains of the FXPs function cooperatively [31,33].

Our purification protocol opens the door for compelling research on the FXP family, particularly FXR1P and FXR2P, which have not been studied as extensively as FMRP. As an example, we observed high affinity binding of FXR1P isoform 2 to poly-$G_{17}U$, despite previous results suggesting only the muscle-specific isoforms exhibit high affinity binding for G-quadruplex forming RNAs [26]. Additionally, the ability of FXR2P to bind to G-quadruplexes is not well-documented, yet our results suggest this could be a worthwhile avenue for further research. Finally, the trend we observed for the extent of translation inhibition of *Renilla* luciferase mRNA, which has potential G-

quadruplex structures, matches the trend for the FXPs' binding affinity for the G-quadruplex forming poly-$G_{17}U$ RNA. It would therefore be interesting to test if the RNA-binding affinities are correlated with the extent of translation regulation for other mRNAs. Our results highlight the utility of our protocol for purifying and comparing the functions of the FXPs *in vitro.*

In summary, we have identified a rapid, simple, and inexpensive purification protocol for the human FXP family, while many of our techniques can be broadly applied. We found the MBP tag very efficient at enhancing the expression and solubility of our proteins, which may be useful to researchers working with large eukaryotic proteins, or proteins with disordered regions [51-52]. By disrupting ribosomal stalling proline-rich motifs within FMRP and FXR2P we drastically boosted recombinant expression while reducing the production of TPs. This technique, or co-expression with EF-P, may assist in the recombinant expression of eukaryotic proteins, 10% of which possess polyproline motifs [60]. An ammonium sulfate precipitation followed by a heparin column allowed the FXPs to be obtained in high yields, free of *E. coli* protein and nucleic acid contamination. Additionally, this procedure removed the vast majority, and in some cases all, C-terminally TPs. We found the heparin column to be a quick and effective method for removing nucleic acid contamination from nucleic acid binding proteins. All three proteins demonstrated RNA-binding activity through their binding to G-quadruplex and kissing complex forming RNAs and were successful in repressing the translation of *Renilla* luciferase mRNA. We hope this procedure will mitigate obstacles faced in studying the important roles of the FXP family in translational regulation, and in doing so, promote diverse research questions. Moreover, the

techniques described will aid researchers in recombinantly expressing and purifying proteins with poor expression, proline-rich regions, disordered regions, or nucleic acid binding properties.

Materials and Methods

2.8 Creation of Fragile X Protein Expression Vectors

An *E. coli* codon optimized sequence was purchased in pUC57 from GeneWiz for FXR2P, and FMRP was purchased as a gene block from IDT. The human sequence (not optimized for *E. coli*) for FXR1P was purchased from Addgene. The genes coding for the human fragile X proteins (FMRP isoform 1 NCBI Reference Sequence: NP_002015.1, FXR1P isoform 2: NP_001013456.1, and FXR2P NP_004851.2) were introduced through Ligation Independent Cloning into the pMCSG9 vector (DNASU plasmid repository) which provides an N-terminal His_6-MBP sequence, T7 promoter, ColE1 origin of replication, and ampicillin resistance [61]. For FXR2P only, the TEV protease cleavage site located after the MBP tag was replaced with an HRV 3C protease cleavage site. The resulting plasmids were transformed into DH5α *E. coli* cells (ThermoFisher), purified, and the sequences verified by Sanger sequencing (GeneWiz). After confirming the cloning process, the plasmids were transformed into chemically competent Rosetta 2(DE3)pLysS *E. coli* cells for protein expression (Novagen, chloramphenicol resistance).

Primers to insert codon optimized FMRP into pMCSG9

Forward: 5'- TACTTCCAATCCAATGCCATGGAAGAACTGGTGGTTGAAGTGCGTG-3’

Reverse: 5’-TTATCCACTTCCAATGTTACGGCACACCATTGACCAGCGG-3’

Primers to insert FXR1P into pMCSG9

Forward: 5'-TACTTCCAATCCAATGCCGCGGAGCTGACGGTGGAGGTT-3’

Reverse: 5’-TTATCCACTTCCAATGTTAATCACATCTTTTGCCTAGCCC-3’

Primers to insert codon optimized FXR2P into pMCSG9

Forward: 5'-TACTTCCAATCCAATGCCATGGGCGGTCTGGCGAGC-3'

Reverse: 5'-TTATCCACTTCCAATGTTAGCTCACACCATTCACCATGCTACC-3'

Primers to replace TEV site of FXR2P pMCSG9 with an HRV 3C cleavage site

Forward: 5'- CTGGAAGTTCTGTTCCAGGGTCCGATGGGCGGTCTGGCGAGC-3'

Reverse: 5'- GCTACCACCACCACCAGTCTGCGCGTCTTTCAGGG-3'

2.9 Creation of FMRP/FXR2P Mutants and Shortened FXR2P Constructs

Mutations were selected by comparing the amino acid sequence of the human fragile X protein to the same protein in other species in order to avoid mutating highly conserved residues. If possible, prolines were mutated into an amino acid present in another species at the corresponding position (Supplementary Figure 3). Site-directed

mutagenesis or sequence deletions were achieved through PCR with designed primers (listed below) on FMRP/FXR2P in pMCSG9. DpnI digestion, PCR purification, T4 PNK treatment, and ligation were performed sequentially after PCR to produce the desired plasmids. The resulting plasmids were transformed into DH5α *E. coli* cells (ThermoFisher), purified, and the sequences verified by Sanger sequencing (GeneWiz). Plasmids containing the desired mutations were transformed into chemically competent Rosetta 2(DE3)pLysS *E. coli* cells (Novagen) to produce desirable cell stocks.

Primers to make $FXR2P_{1-387}$

Forward: 5'-TAACATTGGAAGTGGATAACGGATCCG-3'

Reverse: 5'-TTGACGCAGTTGCTCGTCAATC-3'

Primers to make $FXR2P_{1-515}$

Forward: 5'-TAACATTGGAAGTGGATAACGGATCCG-3'

Reverse: 5'-ATCCGGGTCTTTCAGCACG-3'

Primers to mutate prolines

Codons that introduce a mutation are underlined.

FMRP P451S

Forward: 5’ TCTCCGAACCGTACCGATAAAGAAAAGTC 3’

Reverse: 5’ CGGACGAGAGCTTGCACCGATTTG 3’

FMRP P451N

Forward: 5’ AATCCGAACCGTACCGATAAAGAAAAGTC 3’

Reverse: 5’ CGGACGAGAGCTTGCACCGATTTG 3’

FXR2P P400 (P474S & P492S & P494S)

Forward: 5' GACCGGTGGTCGTGGCCGTGGTAGCCCGAGCGCGCCGCGTCCG 3'

Reverse:

5' GGACGACGACGGCTTTCTTCACCACGGGTGCTCGGATCACGGTCACCC 3'

FXR2P P500 (P529S & P538S & P540S)

Forward: 5' GAGCCGGGCGAAAGCCCGAGCGCGAGCGCGCGTCG 3'

Reverse: 5' GCTATCCACCGGGCTTTCCGGTTCGCTGGTGTCCAGC 3’

FXR2P P394T

Forward: 5'- ATTGGCCTGGGTTTTCGTACCCCGGGTAGCGGCCGTG-3'

Reverse: 5'- TTGACGCAGTTGCTCGTCAATCTGCAGACGCTCCAG-3'

FXR2P P474S

Forward: 5'- ACCCGTGGTGAAGAAAGCCGTCG-3'

Reverse: 5'- GCTCGGATCACGGTCACCCGGACC-3'

FXR2P P492S & P494S

Forward:5'-CCGAGCGCGCCGCGTCCGACCAGC-3'

Reverse: 5'- GCTACCACGGCCACGACCACCG-3'

FXR2P P493N

Forward: 5'-AACCCGGCGCCGCGTCCGACCAGCC-3'

Reverse: 5'-CGGACCACGGCCACGACCACCGG-3'

FXR2P P474S & P493N

Made by taking FXR2P P493N in pMCSG9 and using the primers for FXR2P P474S to add the second mutation.

2.10 Creation of EF-P and FMRP/FXR2P Co-expression Vectors

A codon optimized sequence for Elongation Factor P (NCBI Reference Sequence: P0A6N4.2) was purchased as His_6-EF-P in pUC57 (Gene Universal). Seamless cloning was used to insert His_6-EF-P into a pDSG310 vector (a gift from Ingmar Riedel-Kruse: Addgene plasmid #115611; http://n2t.net/addgene:115611; RRID: Addgene_115611). The pDSG310 vector was selected as it has an arabinose regulated promoter (pBAD), p15A origin of replication, and kanamycin resistance, which are all distinct from those of the pMCSG9 vector used for FMRP and FXR2P

[62]. Using primers that enabled seamless cloning, PCR was used to prepare the His_6-EF-P DNA for insertion into pDSG310, and the backbone of pDSG310 was likewise prepared. The insert and pDSG310 backbone were digested with BbsI and ligated, and the ligated plasmid was transformed into DH5α *E. coli* cells which produced colonies containing viable EF-P containing plasmids. The resulting plasmids were purified, and the sequences verified by Sanger sequencing (GeneWiz). To make cells co-expressing EF-P and FMRP or FXR2P, chemically competent Rosetta 2(DE3)pLysS *E. coli* cells (Novagen) were transformed with 1:1 EF-P plasmid: FMRP/FXR2P plasmid (200 ng each). A control cell stock containing only EF-P in Rosetta 2(DE3)pLysS *E. coli* cells was also produced.

Primers to insert EF-P into pDSG310

Forward: 5' CGTCGAGAAGACTACTAGATGCACCATCATCATCATC 3'

Reverse: 5' CGTCGAGAAGACTTATTTCACGCGGCTCACATATTC 3'

Primers to linearize pDSG310

Forward: 5' CGTCGAGAAGACTGAAATAATAATACTAGAGCCAGGCATCAAATAAAAC 3'

Reverse: 5' CGTCGAGAAGACATCTAGTATTTCTCCTCTTTCTCTAGTAGCTAGC 3'

2.11 Expression Tests of Fragile X Proteins, Mutants, & Co-expression with EF-P

Overnight cultures containing ampicillin (100 μg/mL) and chloramphenicol (25 μg/mL; these are the concentrations of antibiotics used in all cultures) were inoculated with the appropriate Rosetta 2(DE3)pLysS *E. coli* cells from a glycerol stock. For co-expression tests with EF-P, kanamycin was also added (50 μg/mL). Three milliliter overnight cultures were incubated ~16-20 hours at 37°C, ~215 RPM. The following day, LB broth containing ampicillin and chloramphenicol (and kanamycin for cells co-expressing EF-P) was inoculated with the corresponding overnight culture at a ratio of 0.005:1 overnight culture: LB broth. The cultures were incubated at 37°C, ~215 RPM until the OD_{600} reached ~0.4-0.6, although an OD_{600} of up to 0.8 was allowed in some cases. At this time, the cultures were split into equal volumes (3 mL each) to create an uninduced and induced sample. To the induced samples, Isopropyl β-d-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.4 mM. For co-expression with EF-P, arabinose was also added to a final concentration of 0.1% to induce the expression of EF-P. The samples were then incubated for an additional 3 hours at 37°C, ~215 RPM (2 hours for FMRP and mutants Figure 2). For the initial testing of FXR2P mutants only, (Supplementary Figure 4) expression at 14 °C for ~18 hours was performed. These cultures were cooled for 10 min at ~0°C to slow cell growth prior to the addition of IPTG.

After expression, samples were prepared for analysis by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). The OD_{600} of a ¼th dilution of each culture was obtained and used to determine the OD_{600} of the stock solution. For each culture, 500 μL of sample was centrifuged at 16,100 RCF to pellet the cells. After removing the supernatant, the pellet was resuspended in 25 μL of resuspension buffer (10 mM Tris pH 7.5, 20 mM NaCl) per 0.5 OD_{600} unit. After resuspending, 60 μL of the cell pellet resuspension was combined with 15 μL of 5X SDS-PAGE loading dye. The samples were boiled at ~95 °C for 10 minutes, then spun down for a few seconds. The samples were mixed, and 10 μL of each sample was loaded onto a 10% SDS polyacrylamide gel that was run for 10 min at 100V, followed by 180V until the dye front ran off the gel (~55 minutes). The protein bands were visualized by staining the gels with Coomassie Brilliant Blue.

Comparison of full-length FMRP and FXR2P expression with and without mutations was performed by analyzing band intensities for each full-length FXP in ImageJ. The background intensity was subtracted from each sample by subtracting the band intensity in the corresponding uninduced samples from the band intensity in the induced samples. For FMRP P451S only, there appears to be leaky expression as a band corresponding to FMRP appears present in the uninduced sample. Therefore, for FMRP P451S only, the uninduced sample band intensity for FMRP P451N was subtracted instead of the FMRP P451S uninduced intensity. Fold change of expression was then calculated by normalizing the band intensity of each full-length FXP mutant to the band intensity for the respective unmutated full-length FXP. The same analysis

was performed to analyze full-length FXR2P expression with and without EF-P co-expression. After subtracting the background occurring in the uninduced sample, the fold change of expression was calculated by normalizing the band intensity of full-length FXR2P when co-expressed with EF-P to the band intensity of full-length FXR2P without EF-P expression.

2.12 Expression and Purification of Recombinant Fragile X Proteins

For each purification, two 4 L flasks were prepared with 1 L of LB broth with ampicillin and chloramphenicol (2 liters of cell culture), and each flask was inoculated with 5 mL of an overnight culture (0.005:1 overnight culture: LB broth). The cells were incubated at 37°C, ~215 RPM until the OD_{600} reached ~ 0.4-0.5; this step generally took 4-5 hours. The cultures were then cooled for 20 minutes by transferring to an incubator at 14°C, ~150 RPM. The expression of each protein was induced with the addition of IPTG to a final concentration of 0.4 mM. The induction was carried out for 13-15 hours at 14°C, ~150 RPM.

After induction the cells were split into six 500 mL centrifuge flasks and pelleted by centrifuging at 4,420 RCF, 4 °C, 15 min (Beckman J2-HC Centrifuge). Pellets were transferred to a 50 mL polypropylene Falcon tube and weighed; the typical weight was 6.4-7.4 grams from 2 L of culture. The pellet was then resuspended in lysis buffer with no salt (50 mM Tris pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT), and 1 mM phenylmethylsulfonyl fluoride (PMSF)) to a total volume of ~50 mL. The resuspended cells were sonicated (Branson Digital Sonifier) on ice with

ten 8 second pulses at an amplitude of 60% interspersed with 1 min pauses. The crude lysate was then clarified by centrifuging at 50,271 RCF, 30 min, 4 °C (Beckman Coulter Optima LE-80K Ultracentrifuge). The clarified lysate was placed in a beaker on a stir plate at 4 °C and concentrated ammonium sulfate (5 mM HEPES pH 7.5) at 4 °C (concentration is temperature dependent, but ~3.8 M at 0°C) was added to a final concentration of 20% for FXR1P and FXR2P and 25% for FMRP.

After ammonium sulfate addition, the clarified lysate was allowed to sit at 4˚C for at least an hour. Subsequently, the solution was centrifuged at 5,087 RCF, 15 min, 4˚C (Sigma 4K15C) and the pellet was resuspended in 60-75 mL of lysis buffer. The resuspension was dialyzed in 2L of dialysis buffer (50 mM Tris pH 7.5, 1 mM EDTA, and 1 mM DTT) for at least 16 hours. After dialyzing, the protein solution was centrifuged at 5,087 RCF, 30 min, 4˚C in a swinging bucket centrifuge (Sigma 4K15C) to remove insoluble protein.

Fifty milliliters of the supernatant was loaded into a 50 mL superloop (Amersham Biosciences) and bound to a 5 mL heparin column (HiTrap Heparin HP 1X5 mL, GE Healthcare) by fast protein liquid chromatography (ÄKTApurifier, Amersham Pharmacia Biotech) that had been pre-equilibrated with at least 5 column volumes (CV) of 0 M salt buffer (25 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT, 25% glycerol). The typical mass of protein loaded onto the column ranged from 17-130 mg. After collecting the flow-through the column was washed with 2 CV of 0 M salt buffer. The proteins were subsequently eluted with a linear salt gradient that started with 0 M salt buffer (25 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT, 25% glycerol) and increased the salt concentration over 30 CV, ending with 1 M salt buffer (25 mM Tris pH 7.5, 1 mM EDTA,

1 mM DTT, 25% glycerol, 1 M NaCl). Each of the FXPs eluted within a unique range of salt concentrations. Pure FMRP fractions eluted from ~500-600 mM NaCl, with the peak max at ~560 mM, FXR1P at ~560-700 mM NaCl, peak max at ~640 mM, and FXR2P over a large range, however the most full-length with the least TPs eluted from ~700-830 mM NaCl with the peak max at ~730 mM (Supplementary Figures 5-7). It is important to note that KCl can be used instead of NaCl. This appears to elute the proteins at lower salt concentrations: ~400-500 mM for FMRP and ~500 mM for FXR1P and FXR2P. Additionally, a step gradient has also been successful for separating the FXPs from contaminant proteins and TPs.

After analyzing the elution fractions by SDS-PAGE, the desired fractions were either pooled and concentrated by centrifugation at 2,493 RCF, 4°C through a 15 mL 50 kDa MW cutoff concentrator (Amicon Ultra -15 Centrifugal Filters, Millipore Sigma) or individual elution fractions from the heparin column were stored. In either case, the final sample(s) was centrifuged at 16,100 RCF for 10 min at 4˚C to remove any precipitated protein immediately prior to concentration measurements and storage. The concentration of the supernatant was determined from the absorbance at 280 nM (A280 values) (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). The pure FXPs were then stored at 4°C, -15°C, or -80°C. The best temperature for storing the proteins appears to be -80°C as precipitation appears to occur over time at 4°C and -15°C. For samples stored at any temperature, precipitation may occur over time or upon thawing. We therefore recommend that samples are centrifuged to remove precipitated protein and the protein concentration remeasured prior to use in assays.

The A260/280 ratio of stored samples is typically ~0.53-0.62, indicating the nucleic acid contamination has been removed (https://www.biotek.com/resources/docs/PowerWave200_Nucleic_Acid_Purity_Assessment.pdf). Protein yield ranged from ~1.30-8.59 mg; lower yields around 1-2 mg occurred when the fractions from the heparin column were pooled and concentrated. A loss in concentration occurs during concentration steps as discussed previously. The yield reflects the mass of protein at the end of the purification after removing precipitated protein. Storage buffers for use as blanks for concentration readings and for use in assays were created by mixing 0 M and 1 M salt buffers to create buffers with salt concentrations matching that of the final stored protein samples. The concentration of salt in the protein sample was determined from the elution plots from the FPLC machine.

2.13 Mass Spectrometry Analysis of the Fragile X Proteins

The identities of the FXPs were confirmed through mass spectrometry. The band corresponding to each protein was excised from an SDS-PAGE gel and trypsin digested. The samples were subsequently analyzed by liquid chromatography-mass spectrometry (1.5-hour Reverse phase C18 gradient).

2.14 G-quadruplex RNA-binding of the Fragile X Proteins

To confirm that the FXPs were functional, we tested their binding to G-quadruplex forming RNA through fluorescence anisotropy. To ensure the RNA-binding we observed was not due to the tag we used, we also assessed the binding of His_6-MBP.

The purified FXPs were centrifuged at 16,100 RCF, 10 min, 4°C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). Poly-$G_{17}U$ and CR1 labeled with a 3' fluorescein (Dharmacon) were diluted to 5X concentrations (~25 nM) and these solutions were kept in the dark during the experiment. The RNAs in these 5X solutions were renatured in renaturation buffer (50 mM Tris pH 7.5, 75 mM KCl, 2 mM $MgCl_2$) by heating at 68 °C for 5 minutes, then slow cooled from 68 °C to ~28°C for ~1 hour in a water bath. Water, binding buffer, protein storage buffer, protein, and the 5X RNA solution were added in the order listed and mixed together for a final reaction volume of 200 μL. The final reactions contained 20 mM Tris pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 1 μM BSA, 1 mM DTT, 100 ng/μL tRNA (to prevent non-specific binding), and ~5 nM RNA. The protein concentrations tested were 0, 1, 2, 5, 10, 20, 50, 100, and 200 nM (for FMRP, only one trial for 1 and 2 nM points). It is important to note that for each protein concentration tested the total volume of protein + protein storage buffer remained constant. In each trial the binding buffer was adjusted to account for the Tris pH 7.5 and DTT that were contributed from

the protein storage buffer. For trials with His_6-MBP we assumed the storage buffer contributions were negligible since the protein was diluted in FXP storage buffer prior to use, and the FXP storage buffer was used as the protein storage buffer in the anisotropy reactions. Reactions were thoroughly mixed and incubated in the dark at room temperature for 1 hour. After incubation, each reaction was added into a 96-well non-binding plate (Greiner Bio-One) for fluorescence anisotropy using a multimode microplate reader (SPARK TECAN). Samples were excited at 485 nm and emission was measured at 535 nm. To determine binding affinities, the anisotropy data from each binding assay were normalized to initial values without protein, plotted, and fit to a quadratic equation as previously described

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

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 [46]. Three independent trials were performed (except for FMRP 1 and 2 nM points) to determine standard deviations.

RNA sequences:

18 nucleotides CR1: 5'-GCUAUCCAGAUUCUGAUU-Fluorescein-3'

18 nucleotides poly- $G_{17}U$: 5'-GGGGGGGGGGGGGGGGGU-Fluorescein-3'

2.15 *In vitro* Transcription and Fluorescein Labeling of ΔKC2 RNA

The sequence for ΔKC2 was PCR amplified from a pGEM3Z plasmid (a gift from Eileen Chen) which contains a T7 promoter sequence. Nine 100 μL transcription reactions were set up with 90 μL of the PCR-generated DNA template, 4 mM NTPs, 1X transcription buffer (40 mM Tris pH 8.0, 20 mM $MgCl_2$, 2 mM Spermidine, 0.1 % Triton X-100), 5 mM DTT, and ~0.27 μg of T7 RNA polymerase. Each reaction was treated with 2 units of RQ1 DNase (Promega) for 30 min at 37°C, followed by gel purification on a 10% denaturing polyacrylamide gel. It is important to note that our ΔKC2 RNA contains 10 extra nucleotides at the 5' end from cloning into pGEM3Z relative to the sequence used by Darnell et al [39].

To label the RNA, 0.5 nmoles of RNA was 3' oxidized for 90 minutes at room temperature (0.5 mM KIO_4, 100 mM NaOAc pH 5.2) then incubated with fluorescein 5-thiosemicarbizide (FTSC) at 4°C overnight (100 mM NaOAc pH 5.2, 1.5 mM FTSC). The RNA was then purified using a Monarch RNA Clean-up Kit (New England BioLabs).

<u>Primers to PCR amplify ΔKC2 RNA</u>

Forward: 5'-GCAACTGTTGGGAAGGGCGATCG-3'

Reverse: 5'-AGACGCACATACCAGCCGCTAGC-3'

2.16 RNA-binding of Fragile X proteins by Electrophoretic Mobility Shift Assay

To confirm the FXPs were functional, we tested their binding to ΔKC2 RNA through an electrophoretic mobility shift assay. Based on the results from fluorescence anisotropy, we used poly-$G_{17}U$ and CR1 RNAs as positive and negative controls respectively.

The purified FXPs were centrifuged at 16,100 RCF, 10 min, 4°C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). Fluorescein-labeled Poly-$G_{17}U$, CR1, and ΔKC2 were diluted to 10X concentrations (1 μM) and these solutions were kept in the dark during the experiment. The 10X RNA solutions were renatured in renaturation buffer (50 mM Tris pH 7.5, 100 mM KCl, 5 mM $MgCl_2$) by heating at 68 °C for 5 minutes, then slow cooled from 68 °C to ~29°C for ~1 hour in a water bath. Water, 10X binding buffer, protein storage buffer, fragile X protein, and the 10X RNA solution were added in the order listed and mixed together for a final reaction volume of 26 μL. The final reactions contained 50 mM Tris pH 7.5, 145 mM KCl, 5 mM $MgCl_2$, 1 μM BSA, 10 mM DTT, 50 ng/μL tRNA (to prevent non-specific binding), ~100 nM fluorescein-labeled RNA, and for reactions containing protein, 250 nM of protein. For each protein concentration tested the total volume of protein + protein storage buffer remained constant. In each reaction the binding buffer was adjusted to account for the Tris pH 7.5, KCl, and DTT

that were contributed from the protein storage buffer. The reactions were thoroughly mixed and incubated in the dark at room temperature for 1 hour. After incubation, 3 μL of loading dye (xylene cyanol in 50% glycerol) was added to each reaction. A 0.8% agarose gel (SeaKem GTG agarose) was prepared in 1X TBE buffer (100 mM Tris pH 8.3, 100 mM borate, 2 mM Na_2EDTA). After loading 13 μL of each sample, the gel was run at 4°C for 2 hours at 66 V in 1X TBE buffer. The gel was then scanned using a laser scanner (Typhoon FLA 9500, GE Healthcare) and the gel was analyzed in ImageJ.

RNA sequences:

18 nucleotides CR1: 5'-GCUAUCCAGAUUCUGAUU-Fluorescein-3'

18 nucleotides poly-$G_{17}U$: 5'-GGGGGGGGGGGGGGGGGU-Fluorescein-3'

72 nucleotides ΔKC2: 5'-GGGCGAAUUCGGGAUUCCGACCAGAAGGGGCU
AAGGAAUGGUGGGACGAGCUAGCGGCUGGUAUGUGCGUCU-Fluorescein-3'

2.17 Analysis of *In vitro* Translation Regulation by the Fragile X Proteins

The purified FXPs and His_6-MBP were centrifuged at 16,100 RCF, 10 min, 4°C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c

spectrophotometer). Water, 5' capped *Renilla* luciferase mRNA with a 25-nucleotide 3' poly(A) tail, protein storage buffer, and the corresponding FXP or His_6-MBP were combined in the order listed, mixed, and allowed to incubate for 10 minutes at room temperature. Subsequently, we added 2X rabbit reticulocyte lysate that was treated with micrococcal nuclease to reduce endogenous mRNAs and reduce background translation. The reactions were then allowed to incubate at 30°C for 1.5 hours. The final 50 μL reactions contained 10 nM *Renilla* luciferase mRNA and 500 nM FXP or His_6-MBP. After incubation, 45 μL of each reaction was combined with 5 μL of 30 μM colenterazine to achieve a final concentration of 3 μM colenterazine. Each reaction was then added into a 384-well plate (Greiner Bio-One) and the luminescence determined using a multimode microplate reader (SPARK TECAN). For each *in vitro* translation reaction with protein added, the raw luminescence values were compared to the raw luminescence values of the reaction with protein storage buffer only added in order to account for any effect on translation resulting from salts or other reagents in the protein storage buffers.

2.18 Accession Codes

FMRP: Q06787

FXR1P: P51114

FXR2P: P51116

MBP: P0AEX9

Chapter 2, in full, is a reprint of the material as it appears in A Simple Procedure for Bacterial Expression and Purification of the Fragile X Protein Family, Scientific Reports, 2020. Edwards, Madison; Xu, Mingzhi; Joseph, Simpson, Nature, 2020. The dissertation author was the primary investigator and author of this paper.

References

[1] 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., Eur. Mol. Biol. Organ. J. (1995).

[2] Y. Zhang, J.P. O'Connor, M.C. Siomi, S. Srinivasan, A. Dutra, R.L. Nussbaum, G. Dreyfuss, The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2., EMBO J. 14 (1995) 5358.

[3] Siomi, M. C., Zhang, Y., Siomi, H. & Dreyfuss, G. 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**, 3825–32 (1996).

[4] C.E. Bakker, Y. de Diego Otero, C. Bontekoe, P. Raghoe, T. Luteijn, A.T. Hoogeveen, B.A. Oostra, R. Willemsen, Immunocytochemical and Biochemical Characterization of FMRP, FXR1P, and FXR2P in the Mouse, Exp. Cell Res. 258 (2000) 162–170. https://doi.org/10.1006/excr.2000.4932. [5] Sidorov, M. S., Auerbach, B. D. & Bear, M. F. Fragile X mental retardation protein and synaptic plasticity. *Mol. Brain* **6**, 15 (2013).

[6] Chen, E. & Joseph, S. Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie* **114**, 147–154 (2015).

[7] R.J. Hagerman, E. Berry-Kravis, H.C. Hazlett, D.B. Bailey, H. Moine, R.F. Kooy, F. Tassone, I. Gantois, N. Sonenberg, J.L. Mandel, P.J. Hagerman, Fragile X syndrome, Nat. Rev. Dis. Prim. 3 (2017) 17065. https://doi.org/10.1038/nrdp.2017.65.

[8] M. Pieretti, F. Zhang, Y.-H. Fu, S.T. Warren, B.A. Oostra, C.T. Caskey, D.L. Nelson, Absence of expression of the FMR-1 gene in fragile X syndrome, Cell. 66 (1991) 817–822. https://doi.org/10.1016/0092-8674(91)90125-I.

[9] 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. https://doi.org/10.1093/hmg/1.6.397.

[10] S.A. Irwin, B. Patel, M. Idupulapati, J.B. Harris, R.A. Crisostomo, B.P. Larsen, F. Kooy, P.J. Willems, P. Cras, P.B. Kozlowski, R.A. Swain, I.J. Weiler, W.T. Greenough, Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: A quantitative examination, Am. J. Med. Genet. 98 (2001) 161–167. https://doi.org/10.1002/1096-8628(20010115)98:2<161::AID-AJMG1025>3.0.CO;2-B.

[11] Darnell, J. C., Fraser, C. E., Mostovetsky, O. & Darnell, R. B. Discrimination of

common and unique RNA-binding activities among Fragile X mental retardation protein paralogs. *Hum. Mol. Genet.* **18**, 3164–3177 (2009).

[12] W. Guo, E.D. Polich, J. Su, Y. Gao, D.M. Christopher, A.M. Allan, M. Wang, F. Wang, G. Wang, X. Zhao, Fragile X Proteins FMRP and FXR2P Control Synaptic GluA1 Expression and Neuronal Maturation via Distinct Mechanisms, Cell Rep. 11 (2015) 1651–1666. https://doi.org/10.1016/j.celrep.2015.05.013.

[13] S. Cavallaro, S. Paratore, F. Fradale, F.M.S. de Vrij, R. Willemsen, B.A. Oostra, Genes and pathways differentially expressed in the brains of Fxr2 knockout mice, Neurobiol. Dis. 32 (2008) 510–520. https://doi.org/10.1016/J.NBD.2008.09.008. [14] Bontekoe, C. J. M. *et al.* Knockout mouse model for Fxr2: a model for mental retardation. *Hum. Mol. Genet.* **11**, 487–498 (2002).

[15] Kirkpatrick, L. L., McIlwain, K. A. & Nelson, D. L. Alternative Splicing in the Murine and Human FXR1 Genes. *Genomics* **59**, 193–202 (1999).

[16] F. Tamanini, L.L. Kirkpatrick, J. Schonkeren, L. van Unen, C. Bontekoe, C. Bakker, D. Nelson, H. Galjaard, B. a Oostra, a T. Hoogeveen, The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins., Hum. Mol. Genet. (2000). https://doi.org/10.1093/hmg/9.10.1487.

[17] L. Davidovic, S. Sacconi, E.G. Bechara, S. Delplace, M. Allegra, C. Desnuelle, B. Bardoni, Alteration of expression of muscle specific isoforms of the fragile X related protein 1 (FXR1P) in facioscapulohumeral muscular dystrophy patients., J. Med. Genet. 45 (2008) 679–85. https://doi.org/10.1136/jmg.2008.060541.

[18] E. Khandjian, B. Bardoni, F. Corbin, A. Sittler, S. Giroux, D. Heitz, S. Tremblay, C. Pinset, D. Montarras, F. Rousseau, J.-L. Mandel, Novel isoforms of the fragile X related protein FXR1P are expressed during myogenesis, Hum. Mol. Genet. 7 (1998) 2121–2128. https://doi.org/10.1093/hmg/7.13.2121.

[19] J.F. Coy, Z. Sedlacek, D. Bächner, H. Hameister, S. Joos, P. Lichter, H. Delius, A. Poustka, Highly conserved 3′ UTR and expression pattern of FXR1 points to a divergent gene regulation of FXR1 and FMR1, Hum. Mol. Genet. 4 (1995) 2209–2218. https://doi.org/10.1093/hmg/4.12.2209.

[20] E.J. Mientjes, R. Willemsen, L.L. Kirkpatrick, I.M. Nieuwenhuizen, M. Hoogeveen-Westerveld, M. Verweij, S. Reis, B. Bardoni, A.T. Hoogeveen, B.A. Oostra, D.L. Nelson, Fxr1 knockout mice show a striated muscle phenotype: implications for Fxr1p function in vivo, Hum. Mol. Genet. 13 (2004) 1291–1302. https://doi.org/10.1093/hmg/ddh150.

[21] Kirkpatrick, L. L., McIlwain, K. A. & Nelson, D. L. Comparative Genomic Sequence Analysis of the FXR Gene Family: FMR1, FXR1, and FXR2. *Genomics* **78**, 169–177 (2001).

[22] Myrick, L. K., Hashimoto, H., Cheng, X. & Warren, S. T. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. *Hum. Mol. Genet.* **24**, 1733–1740 (2015).

[23] Gantois, I. & Kooy, R. F. Targeting fragile X. *Genome Biol.* **3**, reviews1014 (2002).

[24] Järvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. (2016). doi:10.1186/s12964-016-0132-3

[25] Calabretta, S. & Richard, S. Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends in Biochemical Sciences* **40**, 662–672 (2015).

[26] E. Bechara, L. Davidovic, M. Melko, M. Bensaid, S. Tremblay, J. Grosgeorge, E.W. Khandjian, E. Lalli, B. Bardoni, Fragile X related protein 1 isoforms differentially modulate the affinity of fragile X mental retardation protein for G-quartet RNA structure, Nucleic Acids Res. 35 (2006) 299–306. https://doi.org/10.1093/nar/gkl1021.

[27] S. Adinolfi, A. Ramos, S.R. Martin, F. Dal Piaz, P. Pucci, B. Bardoni, J.L. Mandel, A. Pastore, The N-Terminus of the Fragile X Mental Retardation Protein Contains a Novel Domain Involved in Dimerization and RNA Binding †, Biochemistry. 42 (2003) 10437–10444. https://doi.org/10.1021/bi034909g.

[28] Bardoni, B., Schenck, A. & Louis Mandel, J. A Novel RNA-binding Nuclear Protein That Interacts With the Fragile X Mental Retardation (FMR1) Protein. *Hum. Mol. Genet.* **8**, 2557–2566 (1999).

[29] Schenck, A., Bardoni, B., Moro, A., Bagni, C. & Mandel, J. L. A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8844–9 (2001).

[30] Majumder, P., Chu, J.-F., Chatterjee, B., Swamy, K. B. S. & Shen, C.-K. J. Co-regulation of mRNA translation by TDP-43 and Fragile X Syndrome protein FMRP. *Acta Neuropathol.* **132**, 721–738 (2016).

[31] Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J. & Warren, S. T. 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 (2014).

[32] 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, T. Tuschl, FMRP targets distinct mRNA sequence elements to regulate protein expression., Nature. 492 (2012) 382–6. https://doi.org/10.1038/nature11737.

[33] T. Maurin, K. Lebrigand, S. Castagnola, A. Paquet, M. Jarjat, A. Popa, M. Grossi, F. Rage, B. Bardoni, HITS-CLIP in various brain areas reveals new targets and new modalities of RNA binding by fragile X mental retardation protein., Nucleic Acids Res. 46 (2018) 6344–6355. https://doi.org/10.1093/nar/gky267.

[34] S.A. Whitman, C. Cover, L. Yu, D.L. Nelson, D.C. Zarnescu, C.C. Gregorio, Desmoplakin and talin2 are novel mRNA targets of fragile X-related protein-1 in cardiac muscle, Circ. Res. (2011). https://doi.org/10.1161/CIRCRESAHA.111.244244.

[35] L. Davidovic, N. Durand, O. Khalfallah, R. Tabet, P. Barbry, B. Mari, S. Sacconi, H. Moine, B. Bardoni, A Novel Role for the RNA–Binding Protein FXR1P in Myoblasts Cell-Cycle Progression by Modulating p21/Cdkn1a/Cip1/Waf1 mRNA Stability, PLoS Genet. 9 (2013) e1003367. https://doi.org/10.1371/journal.pgen.1003367.

[36] J. Garnon, C. Lachance, S. Di Marco, Z. Hel, D. Marion, M.C. Ruiz, M.M. Newkirk, E.W. Khandjian, D. Radzioch, Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level., J. Biol. Chem. 280 (2005) 5750–63. https://doi.org/10.1074/jbc.M401988200.

[37] E. Fernández, K.W. Li, N. Rajan, S. De Rubeis, X. Fiers, A.B. Smit, T. Achsel, X.C. Bagni, Cellular/Molecular FXR2P Exerts a Positive Translational Control and Is Required for the Activity-Dependent Increase of PSD95 Expression, (n.d.). https://doi.org/10.1523/JNEUROSCI.4800-14.2015.

[38] X.-L. Xu, R. Zong, Z. Li, M.H.U. Biswas, Z. Fang, D.L. Nelson, F.-B. Gao, FXR1P but not FMRP regulates the levels of mammalian brain-specific microRNA-9 and microRNA-124., J. Neurosci. 31 (2011) 13705–9. https://doi.org/10.1523/JNEUROSCI.2827-11.2011.

[39] 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/10.1101/gad.1276805.

[40] 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) 489–499. https://doi.org/10.1016/S0092-8674(01)00566-9.

[41] C. Schaeffer, B. Bardoni, J.L. Mandel, B. Ehresmann, C. Ehresmann, H. Moine, The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif., EMBO J. 20 (2001) 4803–13. https://doi.org/10.1093/emboj/20.17.4803.

[42] L. Sjekloća, P. V Konarev, J. Eccleston, I.A. Taylor, D.I. Svergun, A. Pastore, A study of the ultrastructure of fragile-X-related proteins., Biochem. J. 419 (2009) 347–57. https://doi.org/10.1042/BJ20082197.

[43] Evans, T. L. & Mihailescu, M. I. Recombinant bacterial expression and purification of human fragile X mental retardation protein isoform 1. *Protein Expr. Purif.* (2010). doi:10.1016/j.pep.2010.06.002

[44] Sjekloća, L., Pauwels, K. & Pastore, A. On the aggregation properties of FMRP - A link with the FXTAS syndrome? *FEBS J.* (2011). doi:10.1111/j.1742-4658.2011.08108.x

[45] J. V. Sopova, E.I. Koshel, T.A. Belashova, S.P. Zadorsky, A. V. Sergeeva, V.A. Siniukova, A.A. Shenfeld, M.E. Velizhanina, K. V. Volkov, A.A. Nizhnikov, D. V. Kachkin, E.R. Gaginskaya, A.P. Galkin, RNA-binding protein FXR1 is presented in rat brain in amyloid form, Sci. Rep. 9 (2019) 18983. https://doi.org/10.1038/s41598-019-55528-6.

[46] Athar, Y. M. & Joseph, S. RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein. *J. Mol. Biol.* (2020). doi:10.1016/J.JMB.2020.04.021

[47] J. Zhang, G. Wang, W.-W. He, M. Losh, E. Berry-Kravis, W.E. Funk, Expression and Characterization of Human Fragile X Mental Retardation Protein Isoforms and Interacting Proteins in Human Cells, Proteomics Insights. 10 (2019) 117864181882526. https://doi.org/10.1177/1178641818825268.

[48] B. Tsang, J. Arsenault, R.M. Vernon, H. Lin, N. Sonenberg, L.-Y. Wang, A. Bah, J.D. Forman-Kay, Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation., Proc. Natl. Acad. Sci. U. S. A. 116 (2019) 4218–4227. https://doi.org/10.1073/pnas.1814385116.

[49] S. Ceman, W.T. O'Donnell, M. Reed, S. Patton, J. Pohl, S.T. Warren, Phosphorylation influences the translation state of FMRP-associated polyribosomes, Hum. Mol. Genet. 12 (2003) 3295–3305. https://doi.org/10.1093/hmg/ddg350.

[50] Andersen, K. R., Leksa, N. C. & Schwartz, T. U. Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**, 1857–61 (2013).

[51] Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32 (2004).

[52] Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).

[53] L. Peil, A.L. Starosta, J. Lassak, G.C. Atkinson, K. Virumae, M. Spitzer, T. Tenson, K. Jung, J. Remme, D.N. Wilson, Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P, Proc. Natl. Acad. Sci. 110 (2013) 15265–15270. https://doi.org/10.1073/pnas.1310642110.

[54] A.L. Starosta, J. Lassak, L. Peil, G.C. Atkinson, K. Virumäe, T. Tenson, J. Remme, K. Jung, D.N. Wilson, Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site, Nucleic Acids Res. (2014). https://doi.org/10.1093/nar/gku768.

[55] Riddihough, G. Tagging truncated proteins with CAT tails. *Science (80-.).* **347**, 38.13-40 (2015).

[56] Mills, E. W. & Green, R. Ribosomopathies: There's strength in numbers. *Science (80-.).* **358**, eaan2755 (2017).

[57] Buskirk, A. R. & Green, R. Getting past polyproline pauses. *Science* (2013). doi:10.1126/science.1233338

[58] 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 β-turn in the RGG motif of FMRP, Proc. Natl. Acad. Sci. 112 (2015) E5391–E5400. https://doi.org/10.1073/pnas.1515737112.

[59] Chen, E., Sharma, M. R., Shi, X., Agrawal, R. K. & Joseph, S. Fragile X mental retardation protein regulates translation by binding directly to the ribosome. *Mol. Cell* **54**, 407–417 (2014).

[60] Buskirk, A. R. & Green, R. Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, (2017).

[61] M.I. Donnelly, M. Zhou, C.S. Millard, S. Clancy, L. Stols, W.H. Eschenfeldt,

F.R. Collart, A. Joachimiak, An expression vector tailored for large-scale, high-throughput purification of recombinant proteins, Protein Expr. Purif. 47 (2006) 446. https://doi.org/10.1016/J.PEP.2005.12.011.

[62] Glass, D. S. & Riedel-Kruse, I. H. A Synthetic Bacterial Cell-Cell Adhesion Toolbox for Programming Multicellular Morphologies and Patterns. *Cell* (2018). doi:10.1016/j.cell.2018.06.041

[63] F. Madeira, Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019., Nucleic Acids Res. 47 (2019) W636–W641. https://doi.org/10.1093/nar/gkz268.

Chapter 3:

The Fragile X Proteins Differentially Regulate Translation of Reporter mRNAs with G-quadruplex Structures

Abstract

Fragile X Syndrome, as well as some manifestations of autism spectrum disorder, results from improper RNA regulation due to a deficiency of fragile X mental retardation protein (FMRP). FMRP and its autosomal paralogs, fragile X related proteins 1 & 2 (FXR1P/2P), have been implicated in many aspects of RNA regulation, from protein synthesis to mRNA stability and decay. The literature on the fragile X related proteins' (FXPs) role in mRNA regulation and their potential mRNA targets is vast. Therefore, we developed an approach to investigate the function of FXPs in translational control using three potential mRNA targets. Briefly, we first selected top mRNA candidates found to be associated with the FXPs and whose translation are influenced by one or more of the FXPs. We then narrowed down the FXPs' binding site(s) within the mRNA, analyzed the strength of this binding *in vitro*, and determined how each FXP affects the translation of a minimal reporter mRNA with the binding site. Overall, all FXPs bound with high affinity to RNAs containing G-quadruplexes, such as Cyclin Dependent Kinase Inhibitor p21 and FMRP's own coding region. Interestingly, FMRP inhibited the translation of each mRNA distinctly and in a manner that appears to correlate with its binding to each mRNA. In contrast, FXR1P/2P inhibited all mRNAs tested. Finally, although binding of our RNAs was due to the RGG (arginine-glycine-glycine) motif-containing C-terminal region of the FXPs, this region was not sufficient to cause inhibition of translation.

Introduction

Fragile X Syndrome (FXS) is the predominant type of inherited intellectual disability and autism spectrum disorder, as well as the first genetic disorder to link RNA regulation to human cognitive function [1,2]. Patients with this disorder may experience seizures, hyperactivity, anxiety, and poor language development [1]. On a cellular level, patients with FXS possess a greater density of dendritic spines, and increased numbers of long and immature-shaped spines [3]. It is estimated that 1 in 5,000 males and 1 in 4,000 to 8,000 females possess the full FXS mutation [1]. FXS predominantly results from a CGG trinucleotide repeat expansion in the 5' untranslated region of the *FMR1* gene [1,4]. The expanded repeats are hypermethylated causing transcriptional silencing of the *FMR1* gene, leading to a deficiency or absence of fragile X mental retardation protein (FMRP) [1,4–6]. FMRP's role in RNA regulation and translation repression has been studied extensively, particularly as it relates to FXS [4]. Beyond FXS and autism spectrum disorder, genes of FMRP mRNA targets are enriched for psychiatric disorders: a recent article found genes with a high probability of being FMRP targets were enriched for association with schizophrenia, bipolar disorder, and major depressive disorder [7]. If we extend our discussion of RNA regulation to the entire fragile X protein family, the impacts of RNA misregulation on human health go beyond neuronal development and cognition.

FMRP is one of three RNA-binding, ribosome-associating proteins involved in translational regulation that comprise the fragile X protein (FXP) family; the other two members are referred to as fragile X-related protein 1 (FXR1P) and fragile X-related

protein 2 (FXR2P) [8–11]. While less studied, FMRP's autosomal paralogs, FXR1P and FXR2P, are noteworthy for their role in translational regulation as their deficiency also leads to developmental abnormalities [8,9,12]. For example, FXR2P-deficient mice have impaired dendritic maturation of new neurons, with new neurons possessing shorter and less complex dendrites compared to wild-type mice [13]. These mice revealed decreased neural connectivity as new neurons with shorter dendrites connected to fewer presynaptic neurons [13]. On the other hand, FXR1P is unique among the FXPs as in humans, FXR1P mRNA demonstrates alternative splicing and is abundant in heart and skeletal muscle tissue [8,14–16]. Furthermore, FXR1P expression is altered in myoblasts from patients with facioscapulohumeral muscular dystrophy [15].

The FXPs are multidomain proteins with high amino acid identity over the first 58-70% of their sequences, but lower identity thereafter **(Figure 3.1A)** [17]. All three proteins possess RNA-binding domains of interest: three K homology (KH) domains within the highly conserved N-terminal region, and an arginine-glycine-glycine (RGG) motif with poor conservation located in the divergent C-termini **(Figure 3.1B-D)** [12,18,19]. In addition, all three proteins possess two Agenet domains which have been shown to bind methylated lysines [20,21]. A less explored feature of the FXPs is their C-terminal intrinsically disordered region (IDR) which constitutes ~30-43% of the entire protein sequence but has lower sequence conservation (Figure 1A) [17]. IDRs are enriched in RNA-binding proteins compared to the entire human proteome and can support protein aggregation, phase transitions, and bind to RNA both specifically and non-specifically [22,23]. Within the IDR of FXR1P and FXR2P are arginine-rich motifs

that likely impart these paralogs with unique RNA-binding capabilities. The isoform of FXR1P that we study, isoform 2, has one such region, while FXR2P has two. Due to their similarity to the nucleolar-targeting signal (NoS) of the protein Rev of human immunodeficiency virus type 1, these sequences are referred to as NoS1 and NoS2 respectively [24]. In other proteins, these motifs have been shown to support RNA recognition, where the few non-arginine amino acids mediate specific binding [25].

Many studies have attempted to identify and validate the mRNA targets of FMRP, while several papers have identified targets of FXR1P and FXR2P [13,26–32]. Although there appears to be overlap in the mRNA targets of the FXP family, there is evidence that each protein has unique mRNA targets [12,13,26,29,31,33]. To validate, analyze, or compare the mRNA targets of the FXP family determined from pull-down methods, researchers often test the direct binding of each protein to its mRNA targets *in vitro*. These studies allow researchers to identify binding sites within a target mRNA or test binding to *in vitro* selected RNAs, which can lead to the identification of sequence motifs or structural features the proteins may recognize *in vivo* [34–36]. Such studies have identified G-quadruplexes and kissing complexes as RNA features recognized by FMRP [12,34–36].

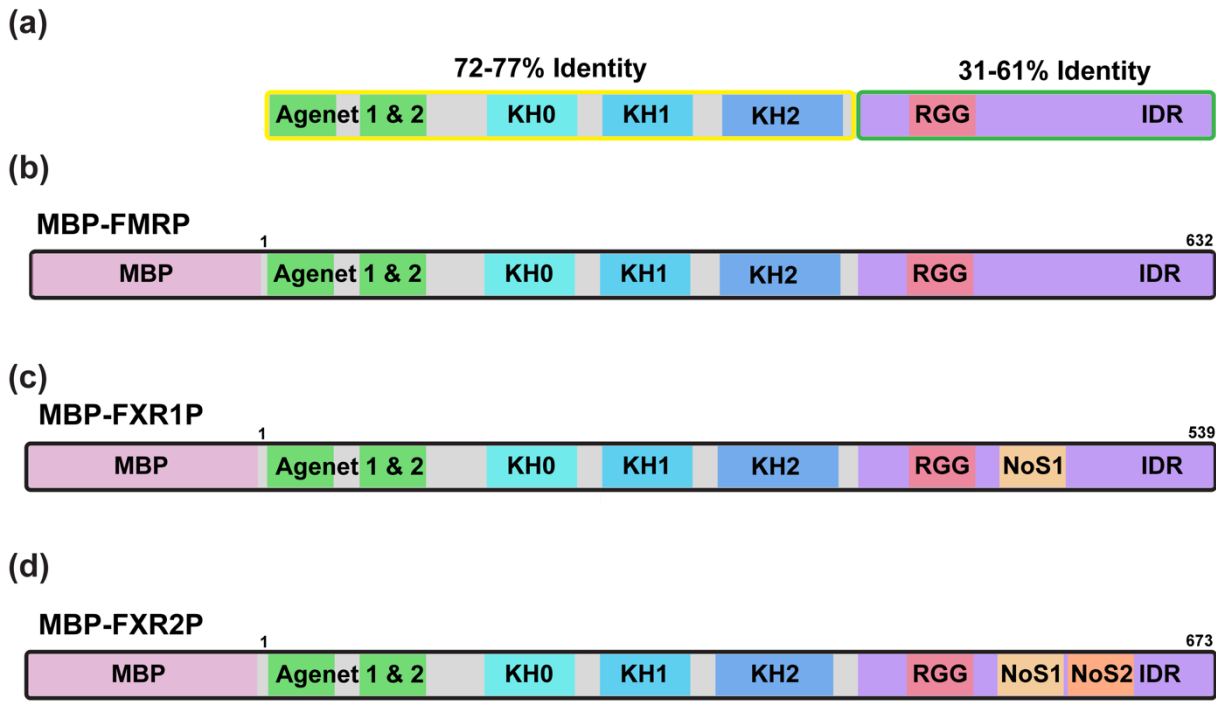


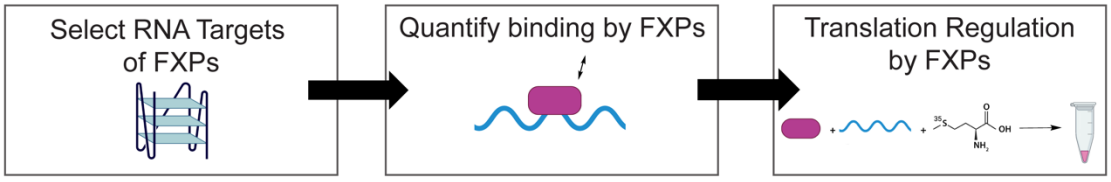
Figure 3.1: Schematics of the fragile X proteins. (A) The fragile X proteins are multidomain proteins that are well-conserved through the sequence RQIG of each protein (outlined in yellow), but their conservation diverges after this point (outlined in green) [17]. (B-D) Schematics of protein constructs used in this article showing relevant domains. We used human FMRP isoform 1, human FXR1P isoform 2, and human FXR2P. All three FXP constructs used in our lab are purified with an N-terminal maltose-binding protein tag (MBP) which is shown.

One immense challenge to determining the mRNA targets of FMRP has been confounding influences such as the biological material or cell type used for assays, the assay implemented, the analyses applied, the presence of FXP interaction partners, and even the particular isoform of the FXP, to name just a few [26]. As just one example, it has been reported that FXR1P can either repress or activate translation of its target mRNAs in non-neuronal cells depending on the cellular context, while different isoforms of FXR1P appear to inhibit translation through distinct mechanisms such as destabilizing mRNA transcripts or through regulating the process of translation itself [30,37]. In an attempt to disentangle our understanding of FMRP's mRNA targets, Suhl *et al.* performed a comparison of three large FMRP mRNA target studies and found an overlap of only ~3.2%, indicating that the methods used significantly impacted the results [26]. From this overlap between three FMRP mRNA target studies (Brown, Darnell, and Ascano-RIP), Suhl *et al.* provided a list of the top 40 targets of FMRP that are associated with FXS, autism, and mental retardation/intellectual disability [26]. Despite the many confounding factors clouding research into FXS therapeutics, by identifying an enriched FMRP recognition motif (WGGA clusters that could form G-quadruplexes), as well as top targets shared amongst three unique studies, Suhl *et al.* provided researchers with a place to focus their attention for therapeutic targets [26].

To further investigate reporter mRNA targets of the FXPs, we developed a method to authenticate the numerous targets identified *in vivo* or through pull-down assays and the few targets worth pursuing for therapeutics **(Figure 3.2A).** First, we narrowed our focus to the most biologically relevant target RNAs with which one or more of the FXPs have been shown to (1) interact with through RNA pull-down assays,

(2) exert an influence over *in vivo*, and (3) contain potential G-quadruplexes **(Figure 3.2B).** We then validated and quantified direct binding of the FXPs to these RNAs *in vitro* by determining binding affinities using a fluorescence anisotropy assay. At this stage we confirmed that the targets are capable of forming G-quadruplexes with an N-methyl mesoporphyrin IX (NMM) assay [38]. The validated candidates (and corresponding negative controls) were then incorporated into our minimal reporter mRNA to observe if translation was altered in the presence of the FXPs **(Figure 3.2C).** Our systematic studies revealed that FMRP, FXR1P, and FXR2P all bind with high affinity to G-quadruplex RNAs. Interestingly, FMRP inhibited translation of our minimal reporter mRNA with the P21 γ G-quadruplex, whereas FXR1P/2P inhibited translation independent of G-quadruplex structures in the mRNA. Thus, FMRP's translation inhibition correlated with its mRNA binding affinity, whereas FXR1P/2P inhibited translation independently of their mRNA binding affinities.

(a)



(b)

P21 γ: 5'-GGGGGUGUGUGGUGGUGGGGGUGGGCUUAUCUGGGAUGGGGU-3'

Mutant γ: 5'-GGGCCUCUCUCCUCCUCCCCCUCCCCUUAUCUCCCAUCCCCU-3'

N19: 5'-GGGGAAGAGGACAAGGAGGAAGAGGACGUGGAGGAGGCUUCU-3'

ARE: 5'-GGGAUUAUUUAUUAUUUAUUUAUUAUUUAUUUAUUUA-3'

(c)

Figure 3.2: Systematic analysis of FXP RNA targets. (A) The workflow we used to characterize the FXPs' interaction with, and influence on selected RNA targets. (B) Sequences selected for the three target RNAs and one negative control RNA (Mutant γ). (C) The NanoFX mRNAs consist of a 5'-methylguanosine cap, a Kozak sequence, a constant coding sequence which encodes a FLAG tag and four methionines, and a 30-nucleotide poly-A tail. For the reporters that contain a 3' UTR, it is a constant 39-nucleotides but contains various target or control RNA sequences. Figure created with BioRender.

Results

3.1 Selection of target mRNA sequences

The mRNA targets of the FXPs have been identified by RNA pull-down assays, such as cross-linking immunoprecipitation (CLIP), which report RNA-protein interactions that may not have a strong initial interaction, and may not be physiologically relevant [26]. Such errors can be introduced through the crosslinking method utilized, particularly as the specificity of crosslinking is not fully understood [26,39]. This problem is prevalent as many studies have reported results with little overlap and have identified few targets with a validated association with FMRP [26]. For the initial testing of our approach, we further narrowed our focus to two targets: G-quadruplexes and AU-rich regions. We selected mRNAs that possess the potential to form G-quadruplexes as this structure has steadfastly been reported as a target of FMRP, and we have recently shown it is a target of the entire FXP family [12,17,35,36,40]. A very recent study supports our decision to focus on G-quadruplexes: mRNA transcripts in FMRP-null mouse neuronal cells that depend on FMRP for efficient transportation in neuronal cells were enriched for G-quadruplex sequences in their 3' UTRs; similar results were also observed in the neurons derived from FXS patients [41]. Additionally, FMRP has been found to bind to G-quadruplexes in both the 5' and 3' UTRs of its target mRNAs, often as a means to inhibit their translation [42]. Second, we selected an AU-rich region, as this was a reported target of FXR1P [31,41,42].

For our biologically relevant RNA targets we selected the G-quadruplex forming regions of Cyclin Dependent Kinase Inhibitor p21 (P21 γ) and FMR1 (N19) mRNAs from previous publications, as well as the AU-rich region of Tumor Necrosis Factor α (TNF-α) (Figure 2B) [30,31,36,43,44]. P21 γ was discovered when Davidovic *et al*. searched for mRNA targets of FXR1P that could be linked to the extreme muscular phenotypes that are observed in FXR1P's absence [30]. They observed that the cell-cycle progression regulator p21 was increased in Fxr1 knockdown Facio-Scapulo Humeral Dystrophy human myoblasts leading to premature exit of the cell-cycle [30]. Furthermore, FXR1P was found to bind p21 mRNA through a G-quadruplex in its 3' UTR (referred to as P21 γ), with the FXR1P/G-quadruplex complex leading to a reduced half-life of the p21 mRNA [30]. For our P21 γ RNA sequence we reduced their P21 γ sequence to the nucleotides listed as 918-955 of murine Cyclin Dependent Kinase Inhibitor p21 as reported in Davidovic *et al*. [30]. To create a negative control RNA which we refer to as Mutant γ, we simply replaced all the guanines in our P21 γ sequence with cytosines, which precludes the formation of G-quadruplexes in Mutant γ.

The second G-quadruplex of interest, referred to as N19, is located within the coding region of FMRP's own mRNA [36]. FMRP was found to bind to a 100-nucleotide region that codes for FMRP's RGG motif with a specific and high affinity interaction [36]. Additionally, when this region was inserted into the 5' UTR of a reporter gene, FMRP inhibited its translation, leading researchers to question if FMRP negatively regulates its own expression *in vivo* [36]. From this publication, we selected a portion

(38 nucleotides of the 100-nucleotide region) of Human FMRP's mRNA that was reported to be necessary but not sufficient for FMRP's binding [36]. FMRP was originally believed to inhibit its own translation by binding to this region; however, a follow-up study revealed that *FMR1* mRNA G-quadruplexes are potent exonic splicing enhancers [42,45]. It appears that the binding of FMRP to G-quadruplexes within its mRNA regulates the production of different isoforms of FMRP [42,45]. It is important to note that an additional study found that FXR1P (only the long isoform e) was also able to bind to this RNA specifically [46].

Finally, we selected the 3' UTR AU-rich element (ARE) of Human TNF-α which is highly conserved among mammals [43]. In fact, the region is so conserved that we were able to select a portion that was present in all three papers we reviewed on this mRNA [31,43,44]. All three articles cite this mRNA as a target of FXR1P, with two reporting that FXR1P represses TNF-α translation [31,43,44]. As we had selected three biologically relevant mRNAs that (1) are bound by one or more of the FXPs, and (2) whose translation are regulated by one or more of the FXPs, we were able to move to the next step of our analysis.

3.2 Confirmation of G-quadruplex formation in target RNAs

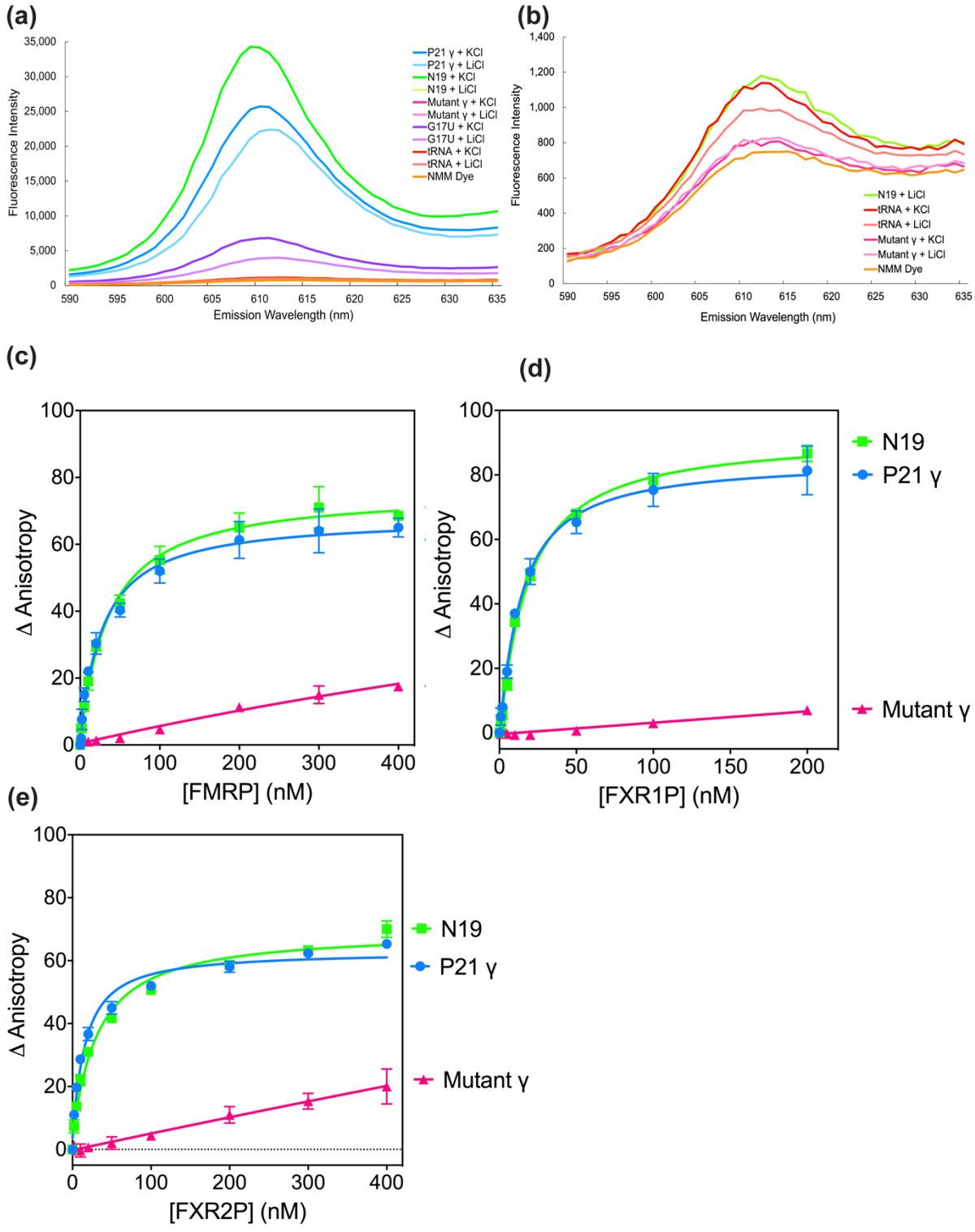
Since we had reduced the mRNA target sequences to 37 (ARE) or 42 nucleotides (P21 γ, Mutant γ, & N19), we first used QGRS mapper and RNAfold to confirm that our reduced G-quadruplex sequences were still predicted to form G-quadruplex structures [47,48]. Once we had synthesized these RNAs, we used an N-

methyl mesoporphyrin IX (NMM) assay to detect G-quadruplex formation and found that our P21 γ and N19 sequences were capable of forming G-quadruplexes [38] **(Figure 3.3A-B).** It is worth mentioning that the P21 γ G-quadruplex appears very stable, as it can form even after refolding in the presence of LiCl.

3.3 Binding of FXPs to target RNAs by fluorescence anisotropy

We subsequently utilized fluorescence anisotropy assays to determine if our RNAs were still bound by the FXPs, and to rank the FXPs binding affinity for each RNA by determining equilibrium dissociation constants (K_D values) **(Figure 3.3C-E).** This step proved essential, as we found binding to the AU-rich region of Tumor Necrosis Factor-α was much weaker than for the G-quadruplex RNAs, and thus a less ideal RNA for our *in vitro* translation studies, or as a therapeutic target **(Figure 3.4).** However, we obtained binding affinities similar to those reported in the literature (0.96 μM), bolstering our confidence in our methods (data not shown) [44]. Additionally, the FXPs showed little to no binding to the Mutant γ sequence, making it a potential negative control RNA. After this step, we narrowed our focus to the G-quadruplex targets, P21 γ and N19, as the binding affinities of the FXPs for these RNAs were much stronger than for the ARE.

Figure 3.3: The FXPs bind G-quadruplex-forming RNAs P21 γ and N19 with high affinity. (A) The NMM assay demonstrates that P21 γ and N19 (4 μM) form G-quadruplexes like the positive control RNA (G17U), indicated by the increase in fluorescence intensity after refolding in buffer with KCl. All other RNAs do not show fluorescence intensity after refolding in buffer with KCl, similar to the negative control (*E. coli* tRNA). One representative trial of two trials is shown. (B) Zoomed in on the graph Figure 3A to show the comparison between traces. FMRP (C), FXR1P (D), and FXR2P (E) bind P21 γ with high affinity (FMRP K_D = 27 ± 4 nM, FXR1P K_D = 12 ± 1 nM, and FXR2P K_D = 14 ± 2 nM) as well as N19 (FMRP K_D = 34 ± 4 nM, FXR1P K_D = 15 ± 1 nM, and FXR2P K_D = 29 ± 4 nM). All proteins show poor or no binding to Mutant γ (FMRP > 2 μM, FXR1P and FXR2P no binding). Results are from 3 or more independent trials with error bars displaying the standard deviation.


(a)
Fluorescence Intensity
Emission Wavelength (nm)
P21 γ + KCl
P21 γ + LiCl
N19 + KCl
N19 + LiCl
Mutant γ + KCl
Mutant γ + LiCl
G17U + KCl
G17U + LiCl
tRNA + KCl
tRNA + LiCl
NMM Dye
(b)
Fluorescence Intensity
Emission Wavelength (nm)
N19 + LiCl
tRNA + KCl
tRNA + LiCl
Mutant γ + KCl
Mutant γ + LiCl
NMM Dye
(c)
Δ Anisotropy
[FMRP] (nM)
(d)
Δ Anisotropy
[FXR1P] (nM)
N19
P21 γ
Mutant γ
(e)
Δ Anisotropy
[FXR2P] (nM)
N19
P21 γ
Mutant γ

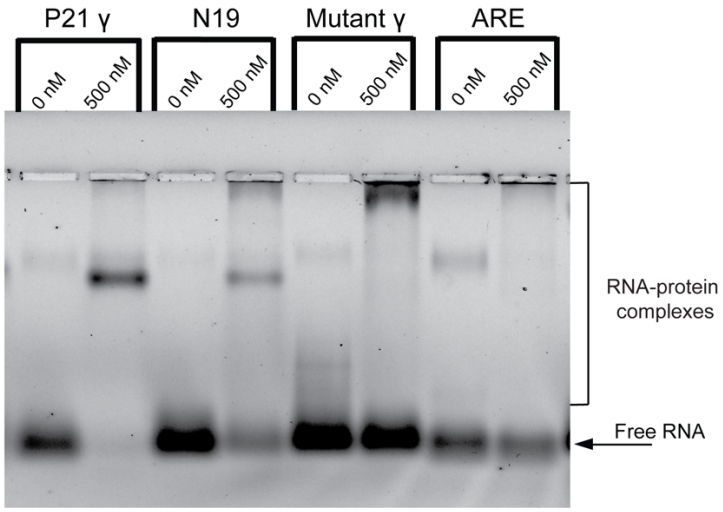


Figure 3.4: The FXPs bind G-quadruplex-forming RNAs P21 γ and N19, but do not bind tightly to Mutant γ and ARE. (A) 100 nM of each 3' fluorescein labeled target RNA sequence was run on a 0.8% native agarose gel in the presence and absence of FMRP (500 nM). This electrophoretic mobility shift assay confirmed that FMRP favors the G-quadruplex-forming RNAs over the Mutant γ and ARE sequences. This can be observed by the reduction of the free RNA band for P21 γ and N19 with a corresponding appearance of an RNA-protein complex band when FMRP is added.

3.4 Creation of minimalistic reporter mRNAs containing 3' UTR target RNA sequences

When assessing a potential mRNA target of FMRP, or its paralogs, researchers (including our own lab) have often added the target mRNA sequence into the 5' or 3' (UTRs) of reporters such as *Renilla,* firefly or NanoLuc luciferases [30,36,43,49]. However, the use of large reporter RNAs can convolute the interpretation of data from *in vitro* translation experiments (IVT), impairing the ability to specifically determine if a protein's association with the target mRNAs alters their translatability. With the length of these reporters (NanoLuc – 557 nucleotides, *Renilla* – 997 nts, and firefly – 1702 nts) comes an increasing probability for the presence of a FMRP target RNA sequence or motif located within the reporter itself. Especially alarming is the fact that our bioinformatic analysis indicates that all the luciferases' mRNA sequences possess potential G-quadruplex forming sites, (NanoLuc- 3, *Renilla* – 3, and firefly – 5) a well-recognized RNA structural target of FMRP. Moreover, preliminary data from our lab corroborates that the FXPs can bind to the reporters themselves. Although researchers may attempt to control for such binding sites, we preferred to create a minimal reporter mRNA (118 nucleotides), to reduce the chance of extraneous binding sites occurring within the mRNA and to eliminate potential G-quadruplex formation sites.

We incorporated our target and control RNA sequences into the 3' UTR of our minimalistic reporter mRNA and added a 5' methylguanosine cap, a Kozak sequence, and a 30 nucleotide 3' poly(A) tail to mimic a eukaryotic mRNA **(Figure 3.2C).** We also created a control reporter that lacks a 3' UTR sequence. Each reporter mRNA encodes

a FLAG peptide which allows for purification from other proteins in our system, and several methionines, which enable detection. The design of our reporter was influenced by other research groups who have had success monitoring the translation of a small reporter RNA [50,51]. However, in contrast to these groups, we found more success monitoring the production of our peptide through radiolabeling, as opposed to immunoblot analysis [50,51]. We refer to our minimalistic reporter RNAs as NanoFX mRNAs.

After creating our NanoFX mRNAs, we verified that the 5' capping step was successful by ensuring that eIF4E bound these NanoFX mRNAs **(Figure 3.5A).** We also verified that the G-quadruplexes could still form within the context of a NanoFX mRNA using the NMM assay. Indeed, both P21 γ and N19 NanoFX mRNAs formed G-quadruplex structures **(Figure 3.6A).** Next, we confirmed the FXPs' ability to bind the target sequences within the context of a NanoFX mRNA by EMSAs. For the NanoFX mRNAs, we observed a similar binding affinity trend to that from our fluorescence anisotropy assays: tightest binding to the P21 γ NanoFX mRNA, followed by the N19 NanoFX mRNA, with similar or perhaps lower binding to the Mutant γ version, and no binding to our NanoFX mRNA with no 3' UTR **(Figure 3.7A-C)**. The two bands present in each lane of this native gel represent the folded RNA monomers to which the FXPs bind (shifted further), and an intermolecular dimer (present close to wells) **(Figure 3.5B-C).** Interestingly, we found that for the N19 NanoFX mRNA, refolding appears to lead to increased binding by FMRP, which is likely true for all the FXPs **(Figure 3.6B).** This suggests that refolding may need to occur for the G-quadruplex to form within the N19 NanoFX mRNA. On the other hand, the P21 NanoFX mRNA was bound similarly

by FMRP regardless of refolding. We subsequently proceeded to test our NanoFX mRNAs in our IVT assay, making sure to refold the N19 NanoFX mRNA prior to this assay.

3.5 Translation of NanoFX mRNAs are inhibited in the presence of FXPs

Translation efficiency of the NanoFX mRNAs in the presence vs. absence of the FXPs was assessed by tracking incorporation of ^{35}S-methionine into the resulting peptide. We performed our IVT reactions in rabbit reticulocyte lysate (RRL) lacking methionine (-MET), then isolated the peptide from other proteins within the system using anti-FLAG M2 beads. The translation efficiency was quantified by liquid scintillation counting and visualized by phosphorimaging **(Figure 3.8A).** Overall, we observed different trends for FMRP compared to its paralogs, FXR1P and FXR2P **(Figure 3.8B-D).** While FMRP regulated each NanoFX mRNA distinctly, and in a manner that appears to correlate with its binding to each NanoFX mRNA, FXR1P/2P inhibited all NanoFX mRNAs tested. Whereas FXR1P appeared to inhibit each NanoFX mRNA to a similar extent, FXR2P inhibited the P21 γ NanoFX mRNA the most. This data suggests that FMRP may function as a specific regulator of translation, whereas FXR1P/2P may, at least in some contexts, function as global translation regulators.

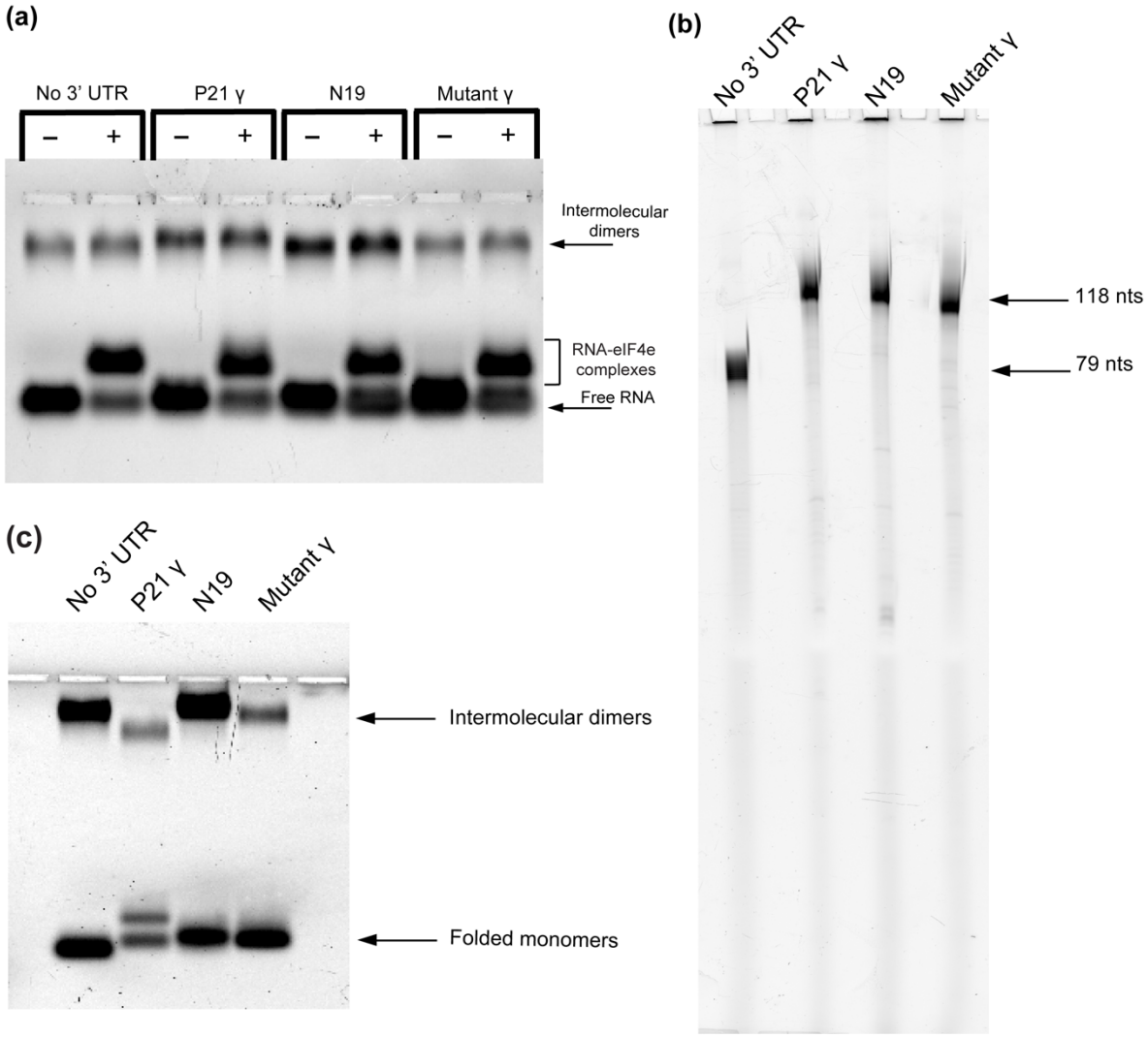
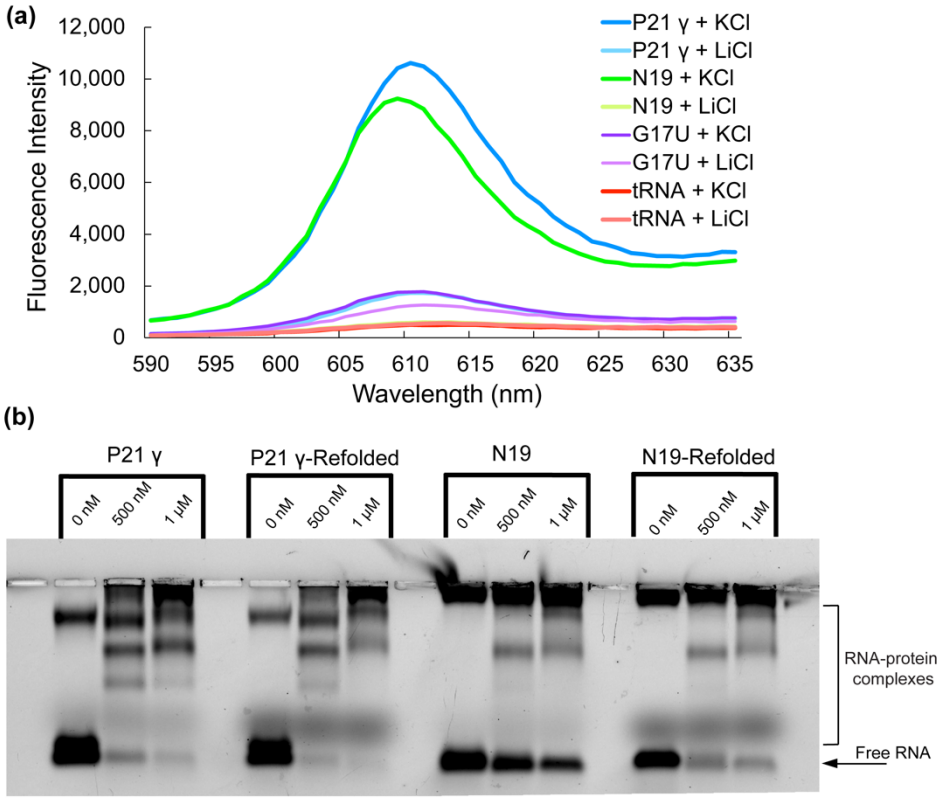


Figure 3.5: Quality control analysis of the NanoFX mRNAs. (A) Successful 5' capping was verified by testing the binding of 250 nM of eukaryotic initiation factor E (eIF-4E) to 50 nM of each 3' fluorescein labeled NanoFX mRNA using an electrophoretic mobility shift assay on a 0.6% native agarose gel. (B) All NanoFX mRNAs appear as a single band of the correct size on a 15% Urea-PAGE denaturing gel. (C). The 3' fluorescein labeled NanoFX mRNAs appear as two bands on a 0.8% native agarose gel. As the NanoFX mRNAs appears as one band on denaturing gels, the lower band in this native agarose gel is likely the correctly folded monomers, while the higher bands are intermolecular dimers.



(c)

	N19 No Refold	N19 Refold
FMRP	113.1 ± 9.0%	85 ± 16%
FXR1P	84 ± 24%	60.2 ± 6.5%
FXR2P	101 ± 23%	81.7 ± 6.3%

Figure 3.6: Refolding is important for G-quadruplex formation in the N19 NanoFX mRNA. (A) The P21 γ and N19 NanoFX mRNAs (2 μM) form G-quadruplexes like the positive control RNA (G17U), indicated by the presence of increased fluorescence intensity after refolding in buffer with KCl. All other RNAs do not show fluorescence intensity increases after refolding in buffer with KCl, similar to the negative control (E. coli tRNA). One representative trial of two trials is shown. (B) Refolding of N19 NanoFX mRNA prior to addition of FMRP appears to lead to enhanced binding, whereas refolding the P21 γ NanoFX mRNA does not appear to be as important for binding. This 0.8% native agarose gel was run in 1X TBE for 2 hours at 66V. (C) Refolding of the N19 NanoFX mRNA reporter prior to IVT assays leads to greater translation inhibition by the FXPs. For FMRP, a shift from slight enhancement to inhibition is observed. It appears that refolding the N19 NanoFX mRNA is critical for the formation of the G-quadruplex. Additionally, the presence of a G-quadruplex appears to influence the extent of translation inhibition. Our findings suggest that the stability of a G-quadruplex structure within an mRNA *in vivo* may influence the extent of regulation by the FXPs. Results are from three trials where in each trial, values were normalized to reactions with protein storage buffer added instead of FXPs.

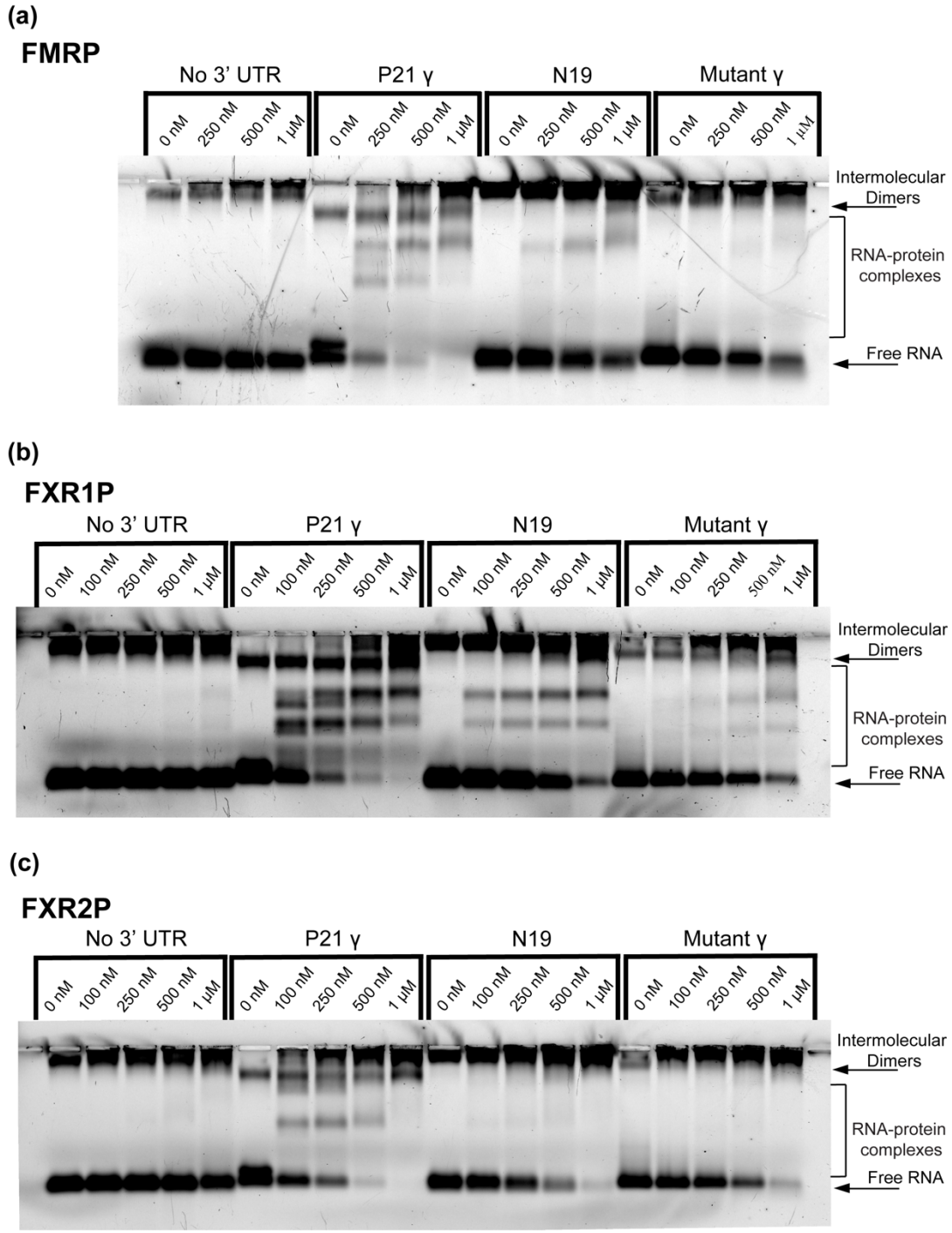
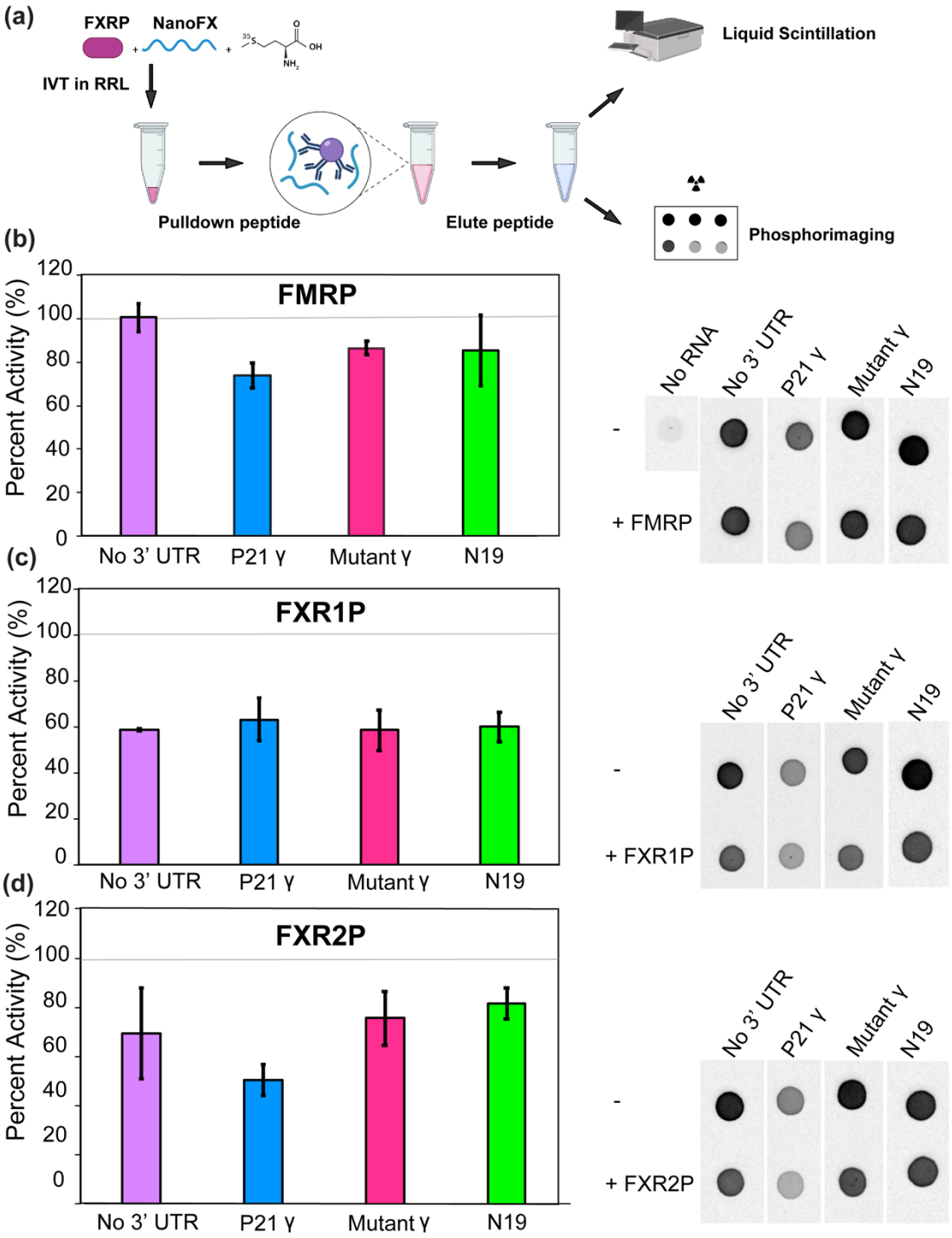


Figure 3.7: The FXPs bind to NanoFX mRNAs with 3' UTR target sequences. (A-C) The FXPs show very little to no binding to the no 3' UTR NanoFX mRNA, stronger binding to the G-quadruplex NanoFX mRNAs (P21 γ & N19), and weak to moderate binding to Mutant γ NanoFX mRNA. The concentration of each FXP was increased from 0 to 1 μM, as indicated above the lanes.

Figure 3.8: *In vitro* translation assay workflow and results for FXPs. (A) To validate the FXPs' influence on target RNAs, we tested the FXPs' ability to alter their translatability in rabbit reticulocyte lysate (RRL) lacking methionine (-MET). Radioactive MET is incorporated into our reporter peptide, which we purified from other proteins in the system using a FLAG affinity purification. We quantified the abundance of our reporter peptide through liquid scintillation counting. The results were subsequently visualized through phosphorimaging. Figure created with BioRender. (B) FMRP does not affect the translation of the no 3' UTR NanoFX mRNA but inhibits all NanoFX mRNAs with 3' UTRs (no 3' UTR 100 ± 6%, P21 γ 74 ± 6%, Mutant γ 86 ± 3%, and N19 85 ± 16%). Results are from three independent trials where, in each trial, values were normalized to reactions with FXP protein storage buffer added instead of FXPs (n =3). Error bars display the standard deviation. (C) FXR1P inhibits all NanoFX mRNAs to a similar extent (no 3' UTR 59 ± 1%, P21 γ 63 ± 9%, Mutant γ 59 ± 9%, and N19 60 ± 7%). Results are from at least three independent trials where in each trial, values were normalized to reactions with FXP protein storage buffer added instead of FXPs. Error bars display the standard deviation. (D) FXR2P has an inhibitory effect on all NanoFX mRNAs, but inhibition of the P21 γ NanoFX mRNA is the greatest (no 3' UTR 69 ± 18%, P21 γ 51 ± 6%, Mutant γ 76 ± 11%, and N19 82 ± 6%). Results are from four independent trials, with one outlier value for P21 γ NanoFX mRNA removed (n=4). For all FXPs, phosphorimaging results were taken in all trials, with a representative image shown.


(a)
FXRP
NanoFX
IVT in RRL
Liquid Scintillation
Pulldown peptide
Elute peptide
Phosphorimaging
(b)
FMRP
Percent Activity (%)
No 3' UTR
P21 γ
Mutant γ
N19
No RNA
-
+ FMRP
(c)
FXR1P
+ FXR1P
(d)
FXR2P
+ FXR2P

3.6 NanoFX mRNAs are inhibited by FXPs in a concentration dependent manner

To further investigate how the FXPs regulate mRNAs *in vitro*, we tested the translation of the P21 γ NanoFX and no 3' UTR NanoFX mRNAs the presence of lower quantities of protein. Our titration revealed that all three FXPs appear to inhibit the translation of the P21 NanoFX in a concentration dependent manner, with the greatest inhibition at 500 nM FXPs. The highest concentration we tested was 500 nM FXPs as this is a biologically relevant concentration and due to the limitations of our chosen system **(Figure 3.9A-C).** For instance, although we controlled for the addition of reagents to our system, RRL is inhibited by addition of excessive salt, and the FXPs must be stored in high salt storage buffers, precluding us from testing higher protein concentrations. For the no 3' UTR NanoFX, inhibition was seen at higher concentrations, but we did not observe a continual decrease of inhibition as more protein was added.

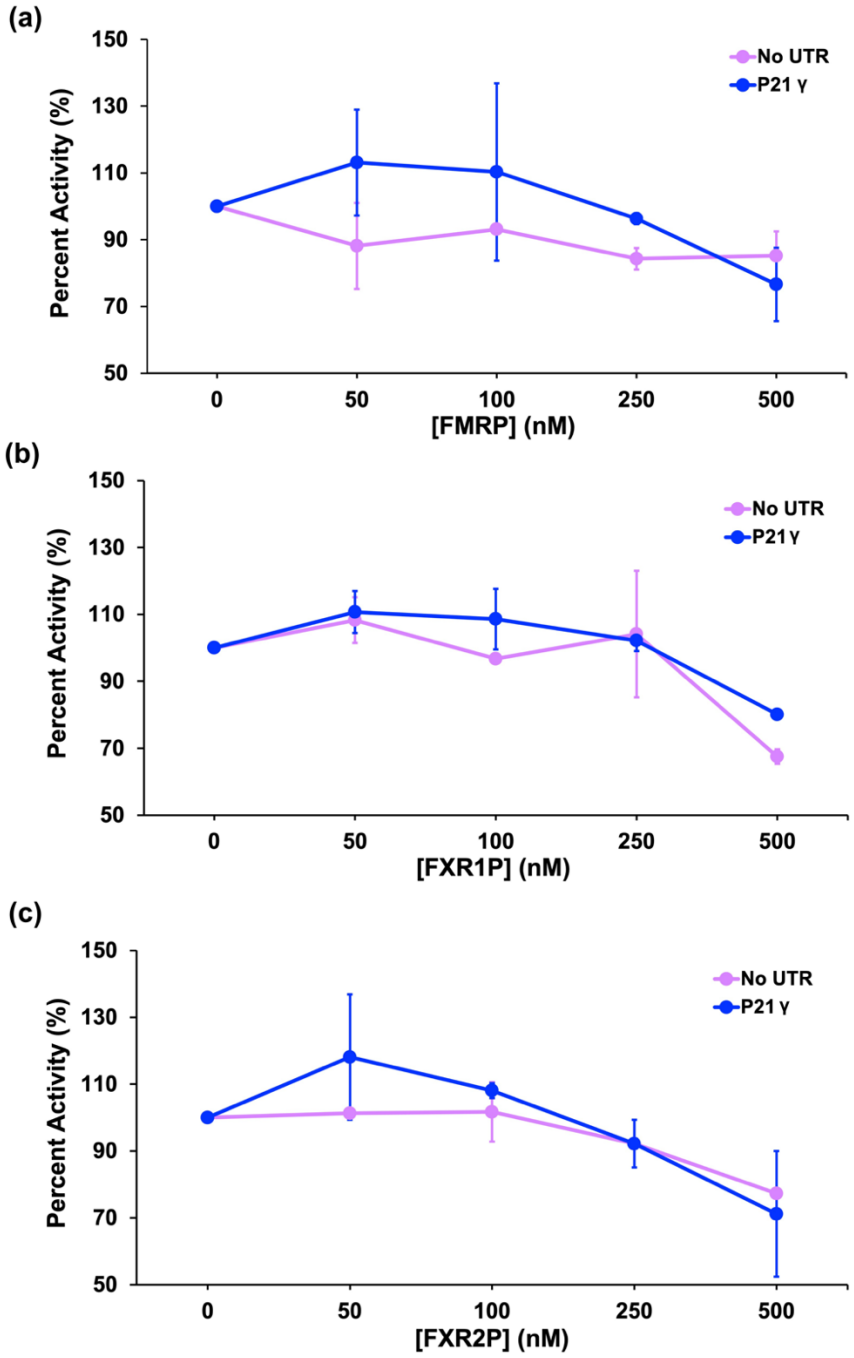


Figure 3.9: The FXPs regulate P21 γ NanoFX mRNA in a concentration dependent manner. The translation regulation of P21 γ and no 3' UTR NanoFX mRNAs in the presence of various concentrations of FMRP (A), FXR1P (B), and FXR2P (C) were tested. It appears that the inhibition of the mRNAs is greater with increased FXP concentrations, except for FMRP with no 3' UTR NanoFX mRNA, where the slight inhibition is fairly consistent. However, this aligns with our results (Figure 5) that indicated that FMRP significantly inhibits translation of P21 γ but not no 3' UTR NanoFX mRNA at 500 nM. Results are from two trials, except for FMRP at data points 0 and 500 nM, for which data is from four trials. Error bars display the standard deviation.

3.7 FXPs inhibit the translation of NanoFX mRNAs similarly over time

Next, we tested how the FXPs regulate translation during our assay by performing a time course assay. We compared the FXPs' regulation to a protein control, maltose-binding protein (MBP), a protein we have previously demonstrated does not bind RNA or regulate translation [17]. Samples taken at various time points throughout the assay revealed that translation of the P21 NanoFX reporter occurs continually in the presence of all proteins tested **(Figure 3.10A).** For all reactions, the amount of peptide being produced increases throughout the experiment, with the rate of production slowing towards the end of the experiment. After 50 minutes, the scintillation counts remain constant or begin to decline for all samples. To further validate our results, we also performed a time course to analyze the regulation of *Renilla* luciferase mRNA, an mRNA we have previously demonstrated is inhibited by the FXPs, but not MBP [17]. In this assay, we compared translation in the presence of the FXPs to MBP and to each protein's respective buffer: the FXPs clearly inhibited translation with respect to their buffer controls, whereas MBP did not **(Figure 3.10B).** Although the production of protein increased over time for all samples, there was significantly less protein produced at each time point with the FXPs compared to MBP or their buffer controls.

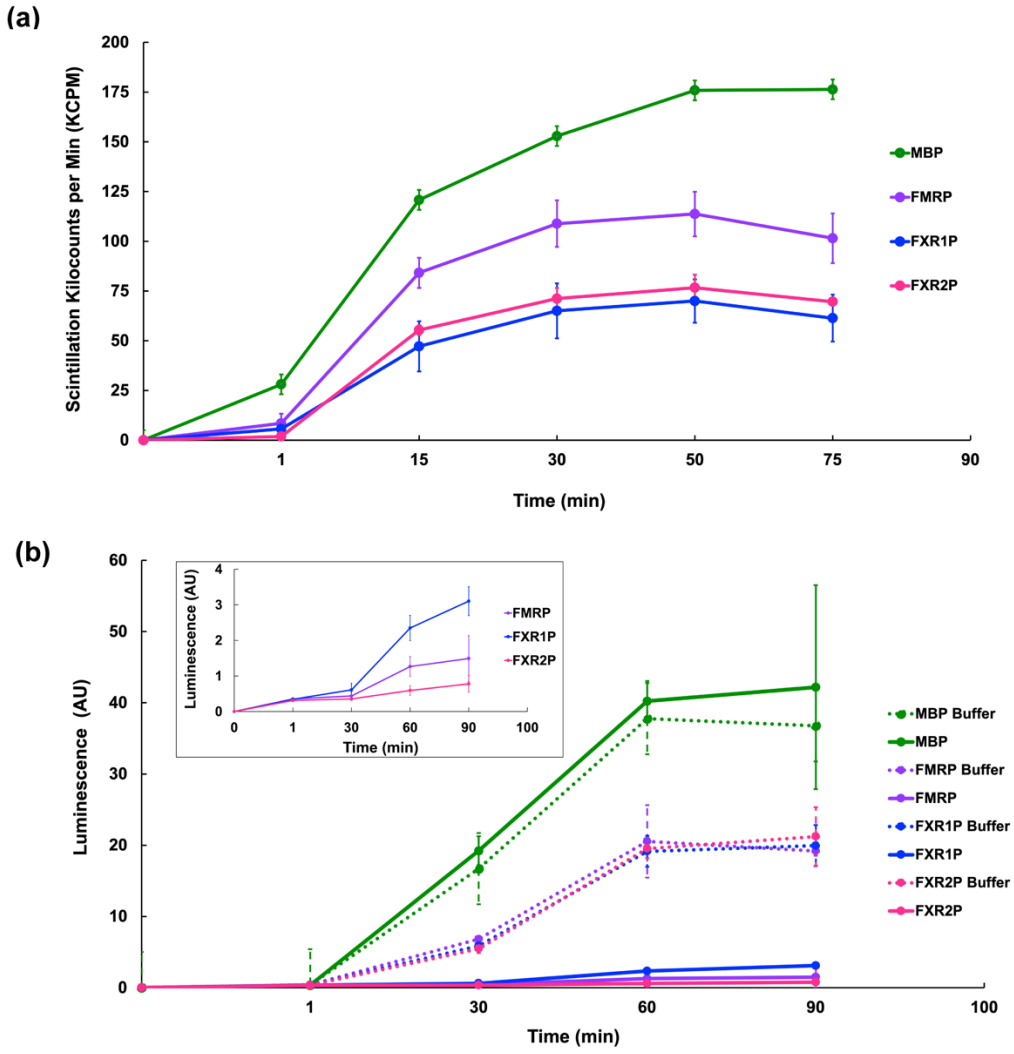


Figure 3.10: The FXPs inhibit the translation of NanoFX and *Renilla* luciferase mRNA over time. (A). We compared the translation of the P21 γ NanoFX mRNA in the presence of the FXPs as well as MBP, a protein we have previously demonstrated does not regulate translation. For all samples in this assay, the production of reporter peptide increases steadily over time. In the beginning, the reporter peptide is produced quickly, with the rate of production decreasing over time. (B). We also tested translation regulation of *Renilla* Luciferase mRNA over time *in vitro* where we observed that the production of protein increases over time. As this was a quicker assay to perform, we also compared the translation of the FXPs to MBP and to their respective buffers. Whereas the FXPs clearly inhibited translation with respect to their buffer controls, MBP did not. All results are from two trials, with error bars displaying the standard deviation of trials.

3.8 NanoFX mRNAs are stable in IVT system

Finally, we tested whether the addition of the FXPs leads to degradation of the no 3' UTR and P21 γ NanoFX mRNAs in RRL by purifying the mRNA from the RRL system after incubation. Although we previously observed that the FXPs generally reduce the translation of the NanoFX mRNAs compared to their buffer controls, we did not observe degradation in the presence of the FXPs for either NanoFX mRNA **(Figure 3.11).** We therefore conclude that the inhibition we observed in our system is not due to degradation of the NanoFX mRNAs. Overall, our IVT experiments were successful for determining how the FXPs regulate each mRNA target sequence *in vitro* without the influence of other RNA recognition elements (obvious or cryptic) that are present in the larger luciferase reporters.

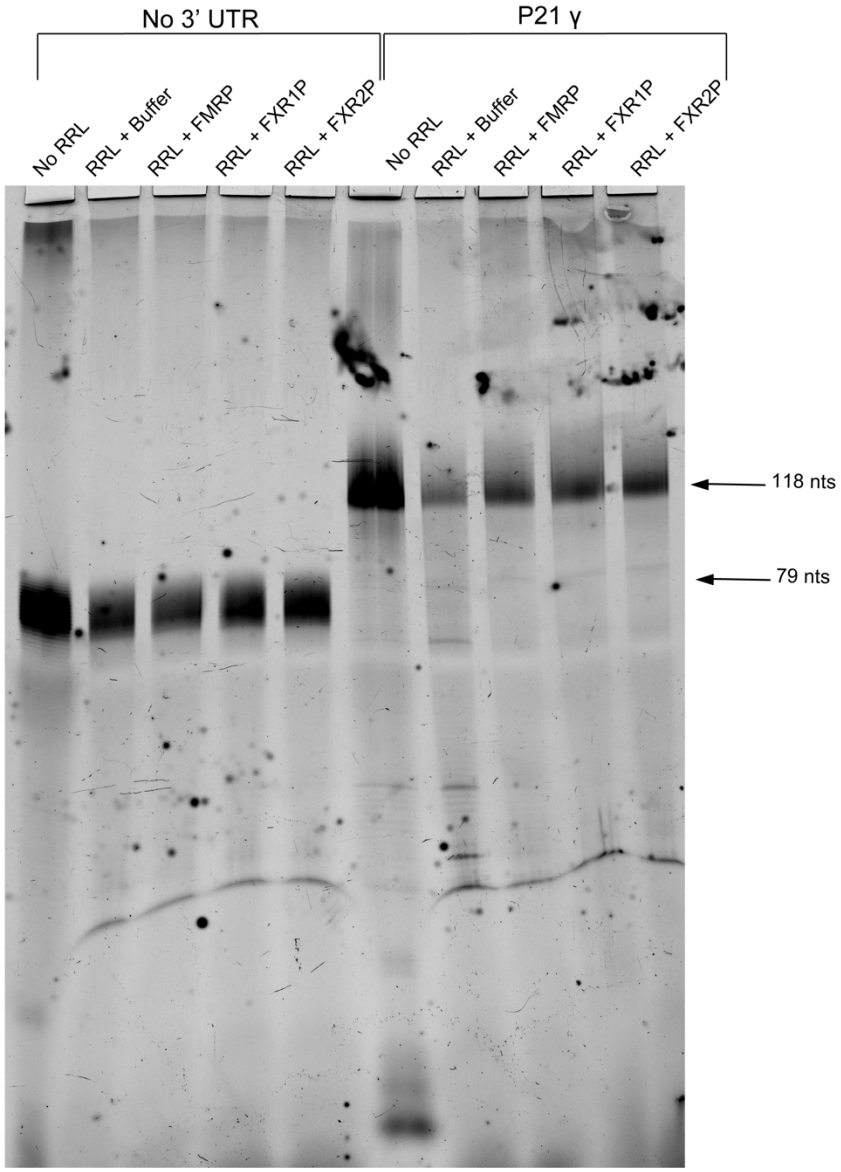


Figure 3.11 The FXPs do not cause degradation of the NanoFX mRNAs *in vitro*. (A). To test if the NanoFX mRNAs experienced degradation during our *in vitro* translation assays, we purified 3' fluorescein-labeled no 3' UTR and P21 γ NanoFX mRNAs after a typical *in vitro* translation reaction. The purified mRNAs were run alongside of a control, a sample of the mRNA that was not incubated in RRL. We also tested whether the addition of the FXPs led to differential degradation of the mRNAs. By quantifying the band intensity of each mRNA after translation in the presence of FXPs to the mRNA with buffer added using ImageJ, we did not observe differential degradation between FXPs. The band intensities normalized to the buffer control are: no 3' UTR NanoFX mRNA with FMRP = 96%, FXR1P = 98%, FXR2P = 104%, and P21 γ mRNA with FMRP = 127%, FXR1P = 136%, FXR2P = 127%. The percentages for P21 γ NanoFX are greater than 100% because less of the mRNA was recovered from the sample with buffer only than in the presence of the FXPs. Two trials were performed, with an image and values from a representative trial presented above.

3.9 The RGG motif-containing C-termini of FMRP binds to target GQ RNAs

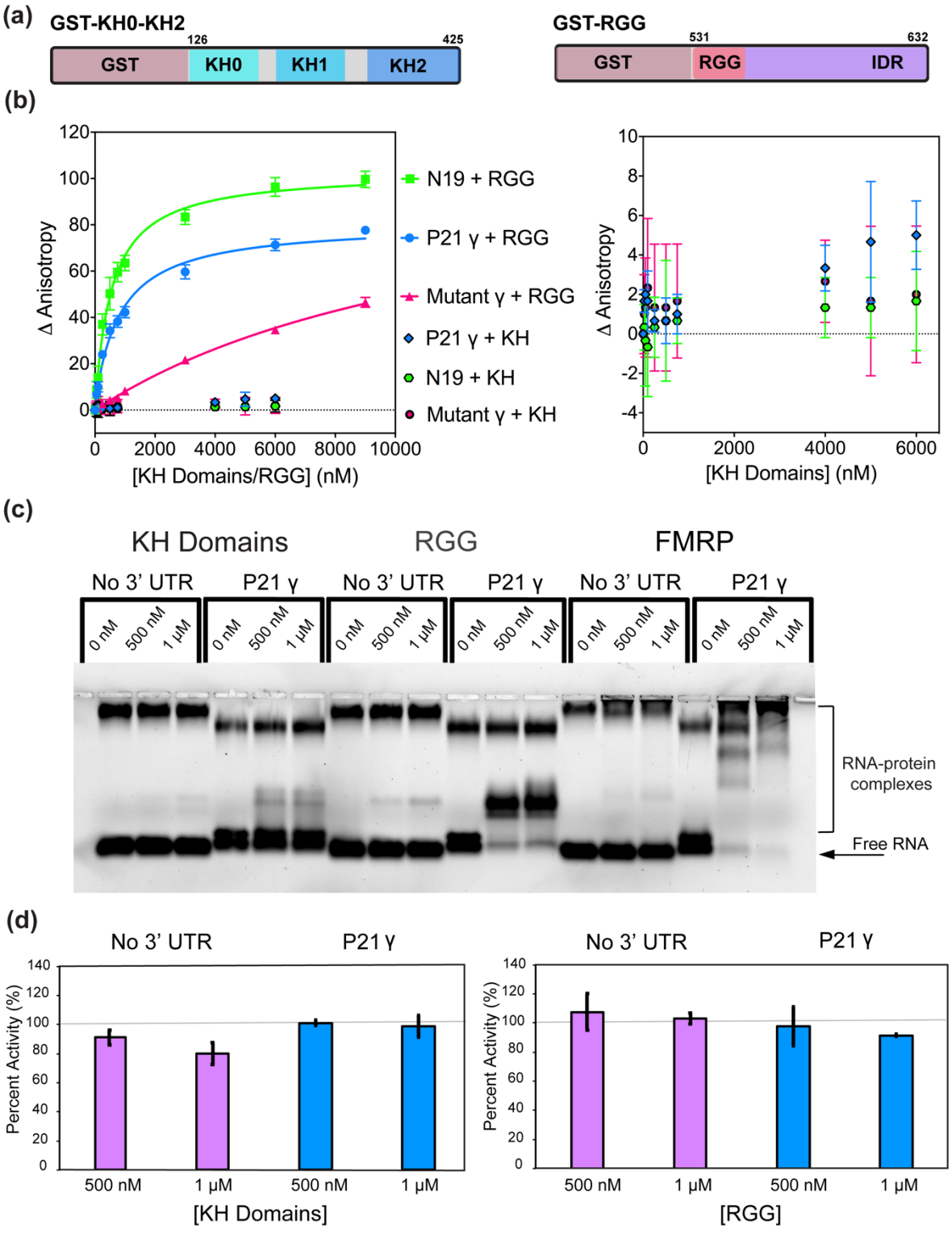
For the final part of our study, we were curious how the different RNA-binding domains of the FXPs contribute to the proteins' ability to bind and regulate our mRNA targets. Many *in vitro* studies of FMRP have focused on only a section of this protein, as the full-length protein is challenging to purify, so we were also interested in comparing the results of the full-length protein to those of its RNA-binding domains. As the KH domains and the RGG motif of FMRP have been of primary interest to researchers, we decided to focus on the KH0-KH2 domains and the RGG motif through C-terminal end of FMRP **(Figure 3.12A).**

From our fluorescence anisotropy assays, we were not surprised to find that the RGG motif of FMRP was able to bind RNA sequences whereas the KH domains were not **(Figure 3.12B).** This aligns with the plethora of research demonstrating that the RGG motif binds to G-quadruplex structures [40,41]. However, when comparing the binding affinities of the RGG motif to full-length FMRP, we observed greatly reduced binding for the RGG motif. Thus, other regions of FMRP must contribute to its RNA-binding ability, perhaps the region between the KH domains and the RGG motif that was not included in either of our constructs but is present in the full-length protein.

3.10 The KH domains and C-termini do not inhibit translation as strongly as full-length FMRP

The fluorescence anisotropy binding data were consistent with our binding studies with the NanoFX mRNAs. Again, the KH domains did not bind to either of the NanoFX mRNAs tested while the RGG motif bound to the G-quadruplex P21 γ NanoFX mRNA, yet not as tightly as the full-length protein **(Figure 3.12C).** When testing the effect of these constructs on *in vitro* translation, we observed that the RGG motif did not inhibit the NanoFX mRNA with no 3' UTR **(Figure 3.12D).** This aligns with the fact it did not bind this NanoFX mRNA and is consistent with the results for full-length FMRP. The RGG motif only slightly inhibited the P21 γ NanoFX mRNA, even at 1 μM protein concentration, which is double the protein concentration used for the studies with full-length FMRP. The KH domains did not inhibit the P21 γ NanoFX mRNA as predicted, but we observed some inhibition for the no 3' UTR NanoFX mRNA. Since the full-length FMRP did not inhibit the no 3' UTR NanoFX mRNA, this suggests it is ideal to analyze the function of the full-length protein whenever possible, as this should yield the most biologically relevant results. Nevertheless, our studies with the KH domains and RGG motif yielded useful insight into individual contributions to the full protein's function: the RGG motif was responsible for binding G-quadruplex RNAs yet was unable to lead to the same levels of inhibition as observed for the full-length protein. Thus, the various domains in FMRP must cooperate to inhibit the translation of its mRNA targets.

Figure 3.12 RNA-binding and IVT studies with FMRP's KH domains and RGG motif. (A) Schematic of the two glutathione S-transferase-tagged FMRP constructs used to compare the functions of FMRP's two classes of RNA-binding domains. (B) The RGG motif principally binds to the G-quadruplex RNAs (N19 K_D = 550 ± 50 nM, P21 γ K_D = 830 ± 70 nM, Mutant γ K_D = 13 ± 2 μM), while the KH domains did not bind any of the RNAs tested. The graph on the right shows the zoomed in view of the binding data for the construct with the KH domains. Error bars display the standard deviation of three or more independent trials. (C) The KH domains do not bind the target RNA sequences within the context of the NanoFX mRNAs. Similar to full-length FMRP, the RGG motif binds the P21 γ NanoFX mRNA, but not the no 3' UTR NanoFX mRNA. The concentration of each FXP were increased from 0 to 1 μM, as indicated above the lanes. (D) The KH domains only slightly inhibited the no 3' UTR NanoFX mRNA (91 ± 5% at 500 nM and 80 ± 8% at 1 μM) but did not significantly alter translation of the P21 γ NanoFX mRNA (101 ± 2.0% at 500 nM and 99 ± 7% at 1 μM). Similar to full-length FMRP, the RGG motif did not inhibit the no 3' UTR NanoFX mRNA (108 ± 13% at 500 nM and 103 ± 4% at 1 μM). The RGG motif slightly inhibited P21 γ NanoFX mRNA, but only at 1 μM (98 ± 13% at 500 nM and 92 ± 1% at 1 μM). Results are from three independent trials (n=3) where in each trial, values were normalized to reactions with protein storage buffer added instead of protein. Error bars display the standard deviation.


(a)
GST-KH0-KH2
126
425
GST
KH0
KH1
KH2
GST-RGG
531
632
GST
RGG
IDR
(b)
Δ Anisotropy
[KH Domains/RGG] (nM)
N19 + RGG
P21 γ + RGG
Mutant γ + RGG
P21 γ + KH
N19 + KH
Mutant γ + KH
Δ Anisotropy
[KH Domains] (nM)
(c)
KH Domains
RGG
FMRP
No 3' UTR
P21 γ
0 nM
500 nM
1 μM
RNA-protein complexes
Free RNA
(d)
No 3' UTR
P21 γ
Percent Activity (%)
500 nM
1 μM
[KH Domains]
[RGG]

3.11 Discussion

Our studies show that G-quadruplexes within biologically relevant mRNAs P21 γ and N19 are bound by all FXPs with high affinity: equilibrium dissociation constants (K_D) in the nanomolar range. We confirmed that the FXPs still bound these target sites within the NanoFX mRNAs before testing them in our IVT system. Both FMRP and FXR2P have the highest binding affinity for P21 γ mRNA and exert the greatest influence over its translation in the context of a NanoFX mRNA. Additionally, refolding the P21 γ NanoFX mRNA does not seem to greatly influence FMRP's ability to bind this RNA, perhaps because it forms a more stable G-quadruplex structure. Thus, of the two target RNA targets that we analyzed, P21 γ appears to show the greatest potential for therapeutics. As FMRP reduced P21 γ's translation, this suggests researchers could investigate inhibitors of the cell-cycle regulator Cyclin Dependent Kinase Inhibitor p21 (*p21*) as a treatment to reverse the effects of FXS. Our results support previous researchers' speculations that an excess of *p21* due to a lack of FMRP may lead to errors in p21-dependent cell cycle exit of neuronal progenitors during neurogenesis [52].

Our research also provided us with further insight into the differences between each of the FXPs. For example, our finding that FMRP may regulate mRNAs in a manner distinct from the other FXPs has led us to question whether the R-rich regions present only in FXR1P/2P lead to enhanced or less specific RNA-binding and inhibition **(Figure 3.13).** Interestingly, while studies have reported that only the long, muscle specific isoforms of FXR1P can bind to G-quadruplexes, we did not find this to be true

[46]. Our neuronal Human isoform 2/b of FXR1P bound to and regulated the translation of G-quadruplex NanoFX mRNAs. In line with previous research where FXR1P was found to preferably bind to a G-quadruplex over an AU-rich region within the 3' UTR of P21 mRNA, we observed the same RNA-binding preferences for FMRP and FXR1P in our studies as well (FXR2P was not tested with AU-rich RNA, although we predict it would behave similarly) [30].

We further analyzed the functions of the KH domains and RGG motifs, which drew our attention to potential new avenues of research. By comparing the functions of FMRP to its KH domains and RGG motif, we found that the RGG motif appears to be predominantly responsible for FMRP's binding to G-quadruplexes, yet other regions of the protein must contribute to the greatly increased binding affinity of the full-length protein, and its greater translation inhibition. Therefore, it would be interesting to investigate contributions that the region between the KH domains and RGG motif make to the overall protein's function. In agreement with this finding, previous work in our lab demonstrated that the C-terminal region of FMRP plays an essential role in the inhibition of translation, yet an N-terminally truncated FMRP led to greater inhibition [49]. This result, corroborated by our findings, indicates that the unstructured region between the KH domains and the RGG motif, may cooperate with the RGG motif/C-terminus of FMRP to regulate translation [49]. Another possibility is that the RNA-binding capabilities of the RGG motif are augmented by the FXPs' ability to dimerize. As the ability of the FXPs to dimerize has been attributed to their N-terminal regions, this could explain why our RGG motif construct bound RNA and inhibited translation, but not to the same extent as full-length FMRP [53].

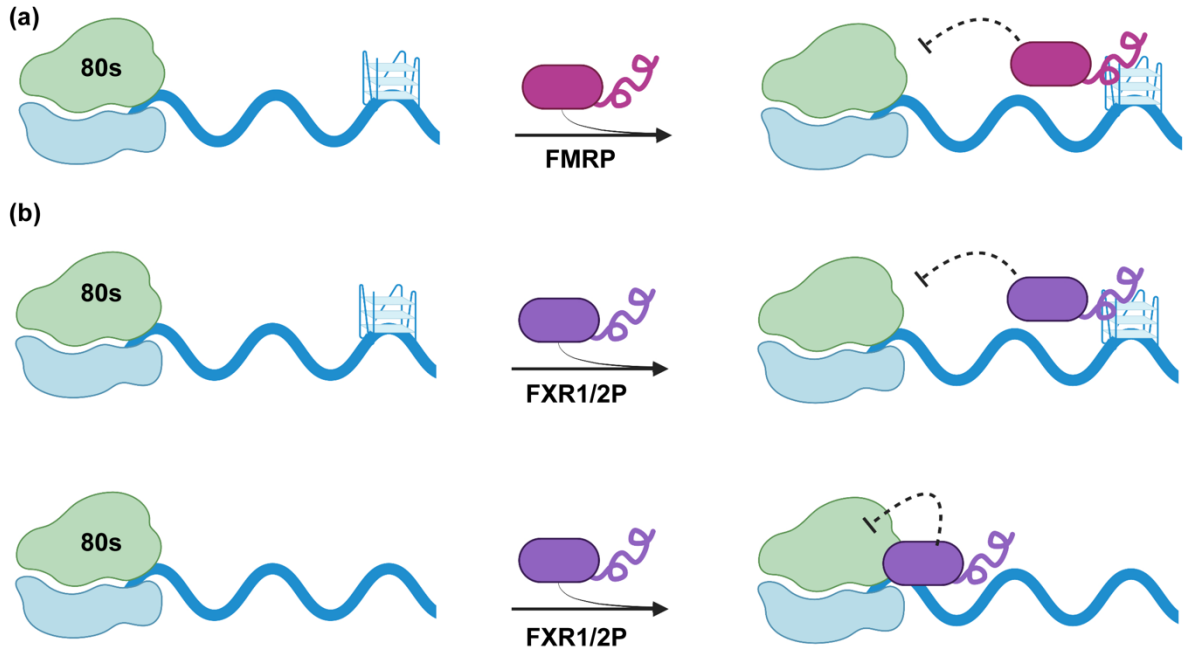


Figure 3.13. The fragile X proteins differentially regulate translation of mRNAs. (A). FMRP's inhibition of translation correlates with its binding to mRNA. FMRP was found to bind with the highest affinity to mRNAs with G-quadruplex structures, and this binding occurs primarily due to the RGG motif/C-terminal domain. Therefore, FMRP may bind to G-quadruplexes within mRNAs to inhibit translation. (B). FXR1P/2P likewise bind with the highest affinity to mRNAs with G-quadruplex structures. However, their inhibition of translation does not correlate with their binding to mRNA. Rather, FXR1/2P appear to globally inhibit translation. We propose that FXR1/2P may be able to bind target mRNA structures to inhibit translation and/or bind to the ribosome to globally inhibit translation.

A second interesting finding was that the KH domains did not bind any RNAs we tested, however, we cannot rule out the possibility that the KH domains can bind RNA sequences or structures not tested in this study. We also observed slight translation inhibition of our no 3' UTR NanoFX mRNA by the KH domains, leading us to consider that they may support the overall function of the full-length FXPs in inhibiting translation. Our findings support using the full-length versions of these proteins *in vitro* to gain a more thorough understanding of their biological functions. We have previously published a method to purify the full-length versions of all the FXPs, so researchers may now easily and accurately test how FMRP and its paralogs behave, which have been shown to be important for proper neural, muscle, and cardiac development as described previously [17].

Our work also provided insight into the stability of the RNA structures we tested. We were able to shorten the G-quadruplex targets to 42 nucleotide sequences, and still observed high affinity binding and translation inhibition. Furthermore, we observed that in contrast to N19, the G-quadruplex in P21 γ appears to be very stable, forming even after refolding in the presence of LiCl. The sensitivity of the N19 G-quadruplex structure may explain why Schaeffer *et al.* found that their 35 nucleotide sequence was necessary but not sufficient for FMRP's binding, whereas we were able to get binding with only 38 out of the 100 nucleotides from their reported functional binding site [36]. Interestingly, we observed that FMRP appeared to bind the N19 NanoFX mRNA more tightly after a refolding step. In line with this finding, all FXPs had a slight inhibitory effect on the N19 NanoFX mRNA after a refolding step **(Figure 3.6C).** This finding indicates that the FXPs regulatory effects may, in certain situations, depend on the

structure of the RNA in question. Additionally, the fact that the FXPs did not inhibit N19 NanoFX mRNA without refolding demonstrates that reagents from our protein purification were not the cause of the observed inhibition of translation. It is tempting to speculate, and future studies may wish to address, whether the folded state of a G-quadruplex RNA is a mechanism through which the cell controls how an mRNA should be regulated. Recent work has demonstrated that G-quadruplexes can function as regulatory elements in neurological disorders, and thus may themselves serve as effective therapeutic targets for FXS and other disorders [42].

Finally, our results illustrate how our approach can yield valuable information for FXS therapeutics (or other disorders). As researchers pursue costly and lengthy therapeutics, many seemingly viable targets for FXS treatments will not meet the necessary standards. It is therefore imperative that biochemists continue performing stringent assays to validate the overabundance of potential FMRP targets identified from RNA pull-down assays. Such validations will refine our list of potential targets, providing researchers with greater chances of success in identifying treatments for FXS.

Materials and Methods

3.12 Purification of Fragile X Proteins

The fragile X proteins (FMRP, FXR1P, & FXR2P) were purified as described previously using KCl as the salt in elution buffers [17]. A step gradient of KCl was used instead of a linear gradient to elute the fragile X proteins. The FMRP KH0-KH2 domains and the RGG motif were purified by Youssi Athar as described previously [38].

3.13 *In vitro* Transcription of P21 γ, Mutant γ, & N19 RNAs and NanoFX mRNAs

Oligos containing a T7 promoter sequence were ordered from Integrated DNA Technologies (IDT). Oligos for the short 3' UTR sequences were first gel purified, then annealed to the 18T7T primer prior to transcription reactions. Oligos for the NanoFX mRNAs (except for the no 3' UTR NanoFX mRNA) were PCR amplified prior to transcription.

For each 100 µL transcription reaction the following reactants were used: 1000 pmol of oligo annealed to 1000 pmol 18T7T primer (for the 3' UTR sequences and no 3' UTR NanoFX mRNA) or 10 µL of the PCR-generated DNA template (for the three NanoFX mRNAs with a 3' UTR), 4 mM NTPs, 1X transcription buffer (40 mM Tris pH 8.0, 20 mM $MgCl_2$, 2 mM Spermidine, 0.1% Triton X-100), 5 mM DTT, and ~ 0.27 µg of T7 RNA polymerase. Each reaction was treated with 2 units of RQ1 DNase

(Promega) for 30 min at 37 °C, followed by gel purification on 15% denaturing polyacrylamide gels.

DNA Templates to Produce Target RNA Sequences

All DNA templates to produce the target RNA sequences were transcribed using the 18T7T primer: 5'-TAATACGACTCACTATAG-3'

P21 γ: 5'-ACCCCATCCCAGATAAGCCCACCCCCACCACCACACACCCCCTATAGTGAGTCGTATTA-3'

Mutant γ: 5'-AGGGGATGGGAGATAAGGGGAGGGGGAGGAGGAGAGAGGCCCTATAGTGAGTCGTATTA-3'

N19: 5'-AGAAGCCTCCTCCACGTCCTCTTCCTCCTTGTCCTCTTCCCCTATAGTGAGTCGTATTA-3'

ARE: 5'-TAAATAAATAAATAATAAATAAATAATAAATAATCCCTATAGTGAGTCGTATTA 3'

Resulting Target RNA Sequences

P21 γ: 42 nucleotides 5'-GGGGGUGUGUGGUGGUGGGGGUGGGCUUAUCUGGGAUGGGGU-Fluorescein-3'

Mutant γ: 42 nucleotides 5'- GGGCCUCUCUCCUCCUCCCCCUCCCCUUAUCUCCCAUCCCCU -Fluorescein-3'

N19: 42 nucleotides: 5′- GGGGAAGAGGACAAGGAGGAAGAGGACGUGGAGGAGGCUUCU-Fluorescein-3′.

ARE: 37 nucleotides 5'-GGGAUUAUUUAUUAUUUAUUUAUUAUUUAUUUAUUUA-Fluorescein-3'

DNA Templates to Produce NanoFX mRNAs

No 3' UTR NanoFX mRNA: 5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTACATCATCATATCGCCGTCATGGTCTTTGTAGTCCATGGTGGCGGCCTATAGTGAGTCGTATTA-3'

P21 γ NanoFX mRNA: 5'-ACCCCATCCCAGATAAGCCCACCCCCACCACCACACACCTTACATCATCATATCGCCGTCATGGTCTTTGTAGTCCATGGTGGCGGCCTATAGTGAGTCGTATTA-3'

Mutant γ NanoFX mRNA: 5'-AGGGGATGGGAGATAAGGGGAGGGGGAGGAGGA GAGAGGTTACATCATCATATCGCCGTCATGGTCTTTGTAGTCCATGGTGGCGGC CTATAGTGAGTCGTATTA-3'

N19 NanoFX mRNA: 5'-AGAAGCCTCCTCCACGTCCTCTTCCTCCTTGTCCTCTT CTTACATCATCATATCGCCGTCATGGTCTTTGTAGTCCATGGTGGCGGCCTATA GTGAGTCGTATTA-3'

Primers to PCR the NanoFX mRNA DNA templates and add a 30-nucleotide Poly-A Tail

All NanoFX mRNAs with 3' UTRs were made using the same forward primer: 5'-TAATACGACTCACTATAGGC-3'

Reverse primers:

P21 γ NanoFX mRNA Reverse Primer: 5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTA CCCCATCCCAGATAAGC-3'

Mutant γ NanoFX mRNA Reverse Primer: 5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TAGGGGATGGGAGATAAGG-3'

N19 NanoFX mRNA Reverse Primer: 5’- TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGAAGCCTCCTCCACGTC-3’

The NanoFX mRNA without a 3’ UTR was produced through transcription after annealing with the 18T7T primer: 5’-TAATACGACTCACTATAG-3’

3.14 3’ Fluorescein Labeling of P21 γ, Mutant γ, & N19 RNAs and NanoFX mRNAs

RNAs used for fluorescence anisotropy, EMSAs, and mRNA stability experiments were 3’ labeled with fluorescein as described below.

To label the RNA, 0.5 nmoles of RNA was 3′ oxidized for 90 min at room temperature (100 mM KIO_4, 100 mM NaOAc pH 5.2) then incubated with fluorescein 5-thiosemicarbizide (FTSC) at 4 °C overnight (100 mM NaOAc pH 5.2, 1.5 mM FTSC). The RNA was then purified using a Monarch RNA Clean-up Kit (New England BioLabs).

3.15 5’ Capping of NanoFX mRNAs with Vaccinia Capping Enzyme

To mimic a mammalian mRNA, the NanoFX mRNAs with 3’ poly(A) tails were 5’ capped as follows.

Pure RNA was denatured by heating at 65°C for 5 minutes and refolded by cooling on ice for 5 minutes. Each 50 µL reaction consisted of 1X Capping Buffer (New

England Biolabs), 0.5 mM GTP, 0.1 mM S-adenosylmethionine (New England Biolabs), Vaccina capping enzyme (0.24 µg/ 10 µg of RNA for RNAs >100 nucleotides), RNA (~ 20-30 µg), and water. Each reaction was incubated at 37°C for 2 hours then purified using a Monarch RNA Clean-up Kit (New England BioLabs). Successful capping was verified by testing the binding of 250 nM of eukaryotic initiation factor 4E (eIF-4E) to 50 nM of each 3' fluorescein labeled NanoFX mRNA using an electrophoretic mobility shift assay on a 0.6% native agarose gel run in 1X TBE for 75 min at 66V.

3.16 N-methyl mesoporphyrin IX (NMM) assay to detect G-quadruplex formation

A stock solution of NMM (Frontier Scientific) was prepared as previously described [38]. The stock solution of NMM (8.61 mM in 0.2 N HCl) was diluted to 400 µM in 10% (v/v) DMSO to make a fresh working solution. Each RNA was refolded in the presence of KCl and LiCl (50 mM Tris pH 7.5, 2 mM $MgCl_2$, 75 mM KCl or LiCl). For 3' UTR sequences, 4 µM of RNA was used; for the NanoFX mRNAs, 2 µM of RNA was used. Refolding occurred for 5 min at 68°C followed by slow cooling to room temperature for ~1 hr. After cooling, 2 µL of 400 µM NMM solution was added to 160 µL of refolded RNA to achieve a final concentration of 5 µM NMM (1.25 µL to 100 µL for the NanoFX mRNAs). The samples were then incubated at room temperature for 10 min prior to loading 155 µL (90 µL for the NanoFX mRNAs) of each sample into a 96-well non-binding plate (Greiner Bio-One) for a fluorescence intensity scan using a

multimode microplate reader (SPARK TECAN). The samples were excited at 400 nm and the emission scanned at 560-650 nm with 5 nm bandwidths for both excitation and emission. The fluorescence intensity values were plotted as a function of wavelength from 580-640 nm. Two trials were performed for each RNA (n=2).

3.17 Assessing RNA-binding of the Fragile X Proteins through Fluorescence Anisotropy

Prior to fluorescence anisotropy assays, each protein stock was thawed from -80°C then centrifuged at 16,100 RCF, 10 min, 4 °C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). RNAs labeled with a 3′ fluorescein were diluted to 5X concentrations (~ 25 nM). The RNAs in these 5X solutions were renatured in renaturation buffer (50 mM Tris pH 7.5, 75 mM KCl, 2 mM $MgCl_2$) by heating at 68 °C for 5 min, then slow cooled from 68 °C to ~ 28 °C for ~ 1 h in a water bath. Water, binding buffer, protein storage buffer, protein, and the 5X RNA solution were added in the order listed and mixed for a final reaction volume of 200 µL. The final reactions contained 20 mM Tris pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 1 µM BSA, 1 mM DTT, 100 ng/µL tRNA (to prevent non-specific binding), and ~ 5 nM RNA. Various concentrations of protein were added for each protein to create a titration curve. It is important to note that for each protein concentration tested the total volume of protein + protein storage buffer remained constant. In each trial the binding buffer

was adjusted to account for the Tris pH 7.5 and DTT that were contributed from the protein storage buffer. Reactions were thoroughly mixed and incubated in the dark at room temperature for 1 h. After incubation, each reaction was added into a 96-well non-binding plate (Greiner Bio-One) for fluorescence anisotropy using a multimode microplate reader (SPARK TECAN). Samples were excited at 485 nm and emission was measured at 535 nm. To determine binding affinities, the anisotropy data from each binding assay were normalized to initial values without protein, plotted, and fit to a quadratic equation as previously described [38]. At least three independent trials (n=3) were performed to determine standard deviations for all proteins except for at 9000 nM for Mutant γ for the C-terminus of FMRP for which only two trials were performed. If applicable, outliers (significance level set as 0.05) were removed.

3.18 RNA-binding of Fragile X Proteins by Electrophoretic Mobility Shift Assay

To confirm the proteins could still bind to the P21 γ and N19 sequences within the context of the NanoFX mRNAs, the binding of each protein to the 3' fluorescein labeled NanoFX mRNAs was tested through an electrophoretic mobility shift assay. This assay was used as the NanoFX mRNAs are too large to be compatible with fluorescence anisotropy assays.

The purified proteins were centrifuged at 16,100 RCF, 10 min, 4 °C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein

determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). Fluorescein-labeled NanoFX mRNAs were diluted to 10X concentrations (1 μM) in water. For the RNA refolding assays only, the 1 μM RNA solutions were made in 50 mM Tris pH 7.5, 75 mM KCl, and 2 mM $MgCl_2$ and RNAs were refolded for 5 min at 68°C followed by slow cooling in a water bath to ~ 28 °C for ~ 1 h. Water, 10X binding buffer, protein storage buffer, proteins, and the 10X RNA solution were added in the order listed and mixed for a final reaction volume of 26 μL. The final reactions contained 50 mM Tris pH 7.5, 135-150 mM KCl, 5 mM $MgCl_2$, 1 μM BSA, 10 mM DTT, 50 ng/μL tRNA (to prevent non-specific binding), ~ 100 nM fluorescein-labeled RNA, and for reactions containing protein, 100-1000 nM of protein. For each protein concentration tested the total volume of protein + protein storage buffer remained constant. In each reaction the binding buffer was adjusted to account for the Tris pH 7.5, KCl, and DTT that were contributed from the protein storage buffer. The reactions were thoroughly mixed and incubated in the dark at room temperature for 1 h. After incubation, 3 μL of loading dye (xylene cyanol in 50% glycerol) was added to each reaction. A 0.8% native agarose gel (SeaKem GTG agarose) was prepared in 1X TBE buffer (100 mM Tris pH 8.3, 100 mM borate, 2 mM Na_2EDTA). After loading 13 μL of each sample, the gel was run at 4 °C for 2 h at 66 V in 1X TBE buffer. The gel was then scanned using a laser scanner (Typhoon FLA 9500, GE Healthcare) and the gel was analyzed in ImageJ.

3.19 Analysis of *in vitro* Translation Regulation of NanoFX mRNAs by the Fragile X Proteins

The purified proteins were centrifuged at 16,100 RCF, 10 min, 4 °C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). For the N19 NanoFX mRNA only, prior to setting up the reactions, the RNA was refolded in renaturation buffer (50 mM Tris pH 7.5, 75 mM KCl, 2 mM $MgCl_2$) by heating at 68 °C for 5 min, then slow cooled from 68 °C to ~ 28 °C for ~ 1 h in a water bath. 2X rabbit reticulocyte lysate (treated with micrococcal nuclease to reduce endogenous mRNAs and without methionine), water, 5′ capped NanoFX mRNAs with a 30-nucleotide 3′ poly(A) tail, protein storage buffer, the corresponding protein, and L-[^{35}S]-Methionine (PerkinElmer, 10mCi (370MBq), Specific Activity: >1000 Ci (37.0 TBq)/mMole, 50 mM Tricine, 10 mM BME), were combined in the order listed, mixed, and allowed to incubate for ~75 min at 30°C. For time trial reactions, samples were taken from a master reaction at 0, 15, 30, 50, and 75 min. The final 20 µL reactions contained 100 nM NanoFX mRNA mRNA and 500 nM protein. For titrations, protein concentrations of 50, 100, 250, and 500 nM were tested. Higher concentrations could not be tested due to limitations to how concentrated the FXPs can be, and how much added salt is tolerated by the RRL. After incubation, each reaction was added to 180 µL of 1X binding buffer (25 mM HEPES pH 7.5, 75 mM NaCl, 0.5 mM EDTA, 0.1% Triton-X, 5% Glycerol, 4 mM $MgCl_2$) and incubated with 2.5 µL of equilibrated anti-FLAG M2 magnetic beads (Sigma-Aldrich) for at least 2 hours at 4°C. After the incubation, the solution was removed from the beads, and the beads were washed

three times with 100 µL of 1X binding buffer. To elute the reporter peptide, 100 µL of FLAG peptide elution buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 4 mM $MgCl_2$, 200 ng/µL 3X FLAG peptide (Sigma-Aldrich) was added to the beads. The elution was allowed to occur for at least 45 min at 4°C. The elution was subsequently removed, and the buffer evaporated (Savant Speed Vac Plus). Once all the liquid was removed, the remaining pellet was resuspended in 10 µL of H_2O.

To quantify the amount of reporter peptide by scintillation counting, 6 µL of the resuspended pellet solution was aliquoted onto 15 mm circular Whatman papers with a pore size of 11 µm (GE Healthcare LifeSciences), which were dried under a lamp for 10 minutes. The filter paper circles were transferred to scintillation vials and 4 mL of ScintiSafe 300% (Fisher Scientific) was added. The S-35 counts were recorded by reading for 1 min/sample using a scintillation counter (LS 6500, Beckman Coulter). If applicable, outliers (significance level set as 0.05) were removed. Results shown are from at least three trials where in each trial, percent activity values were calculated by normalizing to reactions without FXPs or protein added (n=3). For titrations and time trials, two trials were performed (n=2), except for the FMRP titration where four trials were performed for the 0 and 500 nM data points.

To visualize the production of reporter peptide, 3 µL of the resuspended pellet solution was aliquoted onto a piece of Whatman paper and dried under a lamp for 10 min. After drying, the paper placed in a phosphorimaging cassette, covered with Syran wrap, and placed under a phosphorimaging screen. After exposure, the phosphorimaging screen was scanned using a laser scanner (Typhoon FLA 9500, GE Healthcare) and the image was analyzed in ImageJ.

3.20 Time Course *in vitro* Translation Regulation of Renilla Luciferase mRNA by the Fragile X Proteins

The *in vitro* translation regulation of *Renilla* Luciferase mRNA by the FXPs and His_6-MBP was performed as described previously, with several modifications reported here [17]. Reaction buffer was added to contribute 4 mM HEPES KOH pH 7.8, 20 mM potassium acetate, and 0.6 mM $MgCl_2$ to the treated rabbit reticulocyte lysate (RRL). A 100 uL master reaction was made for each protein that contained 10nM of 5' capped and 3' poly A-tailed *Renilla* luciferase mRNA 500 nM FXP or His_6-MBP Samples were taken from this master reaction at 1, 30, 60, and 90 min during an incubation at 30 °C. At each time point, 18 µL of sample was combined with 2 µL of 30 µM colenterazine to achieve a final concentration of 3 µM colenterazine. The luminescence from reactions was analyzed using a multimode microplate reader (SPARK TECAN) as described previously.

3.21 Stability of NanoFX mRNAs *in vitro*

To test the stability of NanoFX mRNAs in the RRL system, reactions were set up as described previously with 2X RRL (treated with micrococcal nuclease to reduce endogenous mRNAs), water, 5′ capped NanoFX mRNAs with a 30-nucleotide 3′ poly (A) tail and fluorescein label, protein storage buffer or 500 nM of the corresponding protein. For these assays, RRL was supplemented with methionine instead of radioactive methionine. The reagents were combined in the order listed, mixed, and

allowed to incubate for 75 min at 30°C. After reaction completion, the mRNA was purified using a Monarch RNA Clean-up Kit (New England BioLabs) and the RNA was eluted using 1% sodium dodecyl sulfate warmed to 70°C. The purified mRNAs were then run on an 8% denaturing gel and visualized using a laser scanner (Typhoon FLA 9500, GE Healthcare). As a control, 1.5 pmol of each NanoFX mRNA was loaded. The gel image was analyzed, and the band intensities quantified in ImageJ.

3.22 Accession Codes

FMRP: Q06787

FXR1P: P51114

FXR2P: P51116

MBP: P0AEX9

Chapter 3, in full, is a reprint of the material as it appears in, The Fragile X Proteins Differentially Regulate Translation of Reporter mRNAs with G-quadruplex Structures, Journal of Molecular Biology, 2021. Edwards, Madison; Joseph, Simpson, Elsevier, 2021. The dissertation author was the primary investigator and author of this paper. The dissertation author was the primary investigator and author of this paper.

References

[1] R.J. Hagerman, E. Berry-Kravis, H.C. Hazlett, D.B. Bailey, H. Moine, R.F. Kooy, F. Tassone, I. Gantois, N. Sonenberg, J.L. Mandel, P.J. Hagerman, Fragile X syndrome, Nat. Rev. Dis. Prim. 3 (2017) 17065. https://doi.org/10.1038/nrdp.2017.65.

[2] J.C. Darnell, S.J. Van Driesche, C. Zhang, K.Y.S. 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) 247–261. https://doi.org/10.1016/j.cell.2011.06.013.

[3] S.A. Irwin, B. Patel, M. Idupulapati, J.B. Harris, R.A. Crisostomo, B.P. Larsen, F. Kooy, P.J. Willems, P. Cras, P.B. Kozlowski, R.A. Swain, I.J. Weiler, W.T. Greenough, Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: A quantitative examination, Am. J. Med. Genet. 98 (2001) 161–167. https://doi.org/10.1002/1096-8628(20010115)98:2<161::AID-AJMG1025>3.0.CO;2-B.

[4] E. Chen, S. Joseph, Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins, Biochimie. 114 (2015) 147–154. https://doi.org/10.1016/j.biochi.2015.02.005.

[5] M. Pieretti, F. Zhang, Y.-H. Fu, S.T. Warren, B.A. Oostra, C.T. Caskey, D.L. Nelson, Absence of expression of the FMR-1 gene in fragile X syndrome, Cell. 66 (1991) 817–822. https://doi.org/10.1016/0092-8674(91)90125-I.

[6] 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. https://doi.org/10.1093/hmg/1.6.397.

[7] N.E. Clifton, E. Rees, P.A. Holmans, A.F. Pardiñas, J.C. Harwood, A. Di Florio, G. Kirov, J.T.R. Walters, M.C. O'Donovan, M.J. Owen, J. Hall, A.J. Pocklington, Genetic association of FMRP targets with psychiatric disorders, Mol. Psychiatry. (2020) 1–14. https://doi.org/10.1038/s41380-020-00912-2.

[8] 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., Eur. Mol. Biol. Organ. J. (1995).

[9] Y. Zhang, J.P. O'Connor, M.C. Siomi, S. Srinivasan, A. Dutra, R.L. Nussbaum, G. Dreyfuss, The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2., EMBO J. 14 (1995) 5358. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC394645/ (accessed May 21,

2020).

[10] 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. http://www.ncbi.nlm.nih.gov/pubmed/8668200 (accessed September 8, 2017).

[11] C.E. Bakker, Y. de Diego Otero, C. Bontekoe, P. Raghoe, T. Luteijn, A.T. Hoogeveen, B.A. Oostra, R. Willemsen, Immunocytochemical and Biochemical Characterization of FMRP, FXR1P, and FXR2P in the Mouse, Exp. Cell Res. 258 (2000) 162–170. https://doi.org/10.1006/excr.2000.4932.

[12] J.C. Darnell, C.E. Fraser, O. Mostovetsky, R.B. Darnell, Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs, Hum. Mol. Genet. 18 (2009) 3164–3177. https://doi.org/10.1093/hmg/ddp255.

[13] W. Guo, E.D. Polich, J. Su, Y. Gao, D.M. Christopher, A.M. Allan, M. Wang, F. Wang, G. Wang, X. Zhao, Fragile X Proteins FMRP and FXR2P Control Synaptic GluA1 Expression and Neuronal Maturation via Distinct Mechanisms, Cell Rep. 11 (2015) 1651–1666. https://doi.org/10.1016/j.celrep.2015.05.013.

[14] L.L. Kirkpatrick, K.A. McIlwain, D.L. Nelson, Alternative Splicing in the Murine and Human FXR1 Genes, Genomics. 59 (1999) 193–202. https://doi.org/10.1006/GENO.1999.5868.

[15] L. Davidovic, S. Sacconi, E.G. Bechara, S. Delplace, M. Allegra, C. Desnuelle, B. Bardoni, Alteration of expression of muscle specific isoforms of the fragile X related protein 1 (FXR1P) in facioscapulohumeral muscular dystrophy patients., J. Med. Genet. 45 (2008) 679–85. https://doi.org/10.1136/jmg.2008.060541.

[16] J.F. Coy, Z. Sedlacek, D. Bächner, H. Hameister, S. Joos, P. Lichter, H. Delius, A. Poustka, Highly conserved 3′ UTR and expression pattern of FXR1 points to a divergent gene regulation of FXR1 and FMR1, Hum. Mol. Genet. 4 (1995) 2209–2218. https://doi.org/10.1093/hmg/4.12.2209.

[17] M. Edwards, M. Xu, S. Joseph, A simple procedure for bacterial expression and purification of the fragile X protein family, Sci. Rep. 10 (2020) 15858. https://doi.org/10.1038/s41598-020-72984-7.

[18] 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) 1733–1740. https://doi.org/10.1093/hmg/ddu586.

[19] I. Gantois, R.F. Kooy, Targeting fragile X., Genome Biol. 3 (2002) reviews1014. http://www.ncbi.nlm.nih.gov/pubmed/12049671 (accessed October 3, 2017).

[20] A. Ramos, D. Hollingworth, S. Adinolfi, M. Castets, G. Kelly, T.A. Frenkiel, B. Bardoni, A. Pastore, The Structure of the N-Terminal Domain of the Fragile X Mental Retardation Protein: A Platform for Protein-Protein Interaction, Structure. 14 (2006) 21–31. https://doi.org/10.1016/J.STR.2005.09.018.

[21] M.A. Adams-Cioaba, Y. Guo, C. Bian, M.F. Amaya, R. Lam, G.A. Wasney, M. Vedadi, C. Xu, J. Min, Structural Studies of the Tandem Tudor Domains of Fragile X Mental Retardation Related Proteins FXR1 and FXR2, PLoS One. 5 (2010) e13559. https://doi.org/10.1371/journal.pone.0013559.

[22] A.I. Järvelin, M. Noerenberg, I. Davis, A. Castello, The new (dis)order in RNA regulation, (2016). https://doi.org/10.1186/s12964-016-0132-3.

[23] S. Calabretta, S. Richard, Emerging Roles of Disordered Sequences in RNA-Binding Proteins, Trends Biochem. Sci. 40 (2015) 662–672. https://doi.org/10.1016/j.tibs.2015.08.012.

[24] F. Tamanini, L.L. Kirkpatrick, J. Schonkeren, L. van Unen, C. Bontekoe, C. Bakker, D. Nelson, H. Galjaard, B. a Oostra, a T. Hoogeveen, The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins., Hum. Mol. Genet. (2000). https://doi.org/10.1093/hmg/9.10.1487.

[25] R. Tan, A.D. Frankel, A novel glutamine-RNA interaction identified by screening libraries in mammalian cells., Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 4247–52. https://doi.org/10.1073/pnas.95.8.4247.

[26] 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 WGGA sequences, Hum. Mol. Genet. 23 (2014) 5479–5491. https://doi.org/10.1093/hmg/ddu272.

[27] 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, T. Tuschl, FMRP targets distinct mRNA sequence elements to regulate protein expression., Nature. 492 (2012) 382–6. https://doi.org/10.1038/nature11737.

[28] T. Maurin, K. Lebrigand, S. Castagnola, A. Paquet, M. Jarjat, A. Popa, M. Grossi, F. Rage, B. Bardoni, HITS-CLIP in various brain areas reveals new targets and new modalities of RNA binding by fragile X mental retardation protein., Nucleic Acids Res. 46 (2018) 6344–6355. https://doi.org/10.1093/nar/gky267.

[29] S.A. Whitman, C. Cover, L. Yu, D.L. Nelson, D.C. Zarnescu, C.C. Gregorio, Desmoplakin and talin2 are novel mRNA targets of fragile X-related protein-1 in cardiac muscle, Circ. Res. (2011). https://doi.org/10.1161/CIRCRESAHA.111.244244.

[30] L. Davidovic, N. Durand, O. Khalfallah, R. Tabet, P. Barbry, B. Mari, S. Sacconi, H. Moine, B. Bardoni, A Novel Role for the RNA-Binding Protein FXR1P in Myoblasts Cell-Cycle Progression by Modulating p21/Cdkn1a/Cip1/Waf1 mRNA Stability, PLoS Genet. (2013). https://doi.org/10.1371/journal.pgen.1003367.

[31] J. Garnon, C. Lachance, S. Di Marco, Z. Hel, D. Marion, M.C. Ruiz, M.M. Newkirk, E.W. Khandjian, D. Radzioch, Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level., J. Biol. Chem. 280 (2005) 5750–63. https://doi.org/10.1074/jbc.M401988200.

[32] E. Fernández, K.W. Li, N. Rajan, S. De Rubeis, X. Fiers, A.B. Smit, T. Achsel, X.C. Bagni, Cellular/Molecular FXR2P Exerts a Positive Translational Control and Is Required for the Activity-Dependent Increase of PSD95 Expression, (n.d.). https://doi.org/10.1523/JNEUROSCI.4800-14.2015.

[33] X.-L. Xu, R. Zong, Z. Li, M.H.U. Biswas, Z. Fang, D.L. Nelson, F.-B. Gao, FXR1P but not FMRP regulates the levels of mammalian brain-specific microRNA-9 and microRNA-124., J. Neurosci. 31 (2011) 13705–9. https://doi.org/10.1523/JNEUROSCI.2827-11.2011.

[34] 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/10.1101/gad.1276805.

[35] 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. https://doi.org/10.1016/s0092-8674(01)00566-9.

[36] C. Schaeffer, B. Bardoni, J.L. Mandel, B. Ehresmann, C. Ehresmann, H. Moine, The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif., EMBO J. 20 (2001) 4803–13. https://doi.org/10.1093/emboj/20.17.4803.

[37] D. Cook, M. del Rayo Sanchez-Carbente, C. Lachance, D. Radzioch, S. Tremblay, E.W. Khandjian, L. DesGroseillers, K.K. Murai, Fragile X Related Protein 1 Clusters with Ribosomes and Messenger RNAs at a Subset of Dendritic Spines in the Mouse Hippocampus, PLoS One. 6 (2011) e26120.

https://doi.org/10.1371/journal.pone.0026120.

[38] Y.M. Athar, S. Joseph, RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein, J. Mol. Biol. (2020). https://doi.org/10.1016/J.JMB.2020.04.021.

[39] C. Lin, W.O. Miles, Beyond CLIP: advances and opportunities to measure RBP–RNA and RNA–RNA interactions, Nucleic Acids Res. 47 (2019) 5490–5501. https://doi.org/10.1093/nar/gkz295.

[40] 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 β-turn in the RGG motif of FMRP, Proc. Natl. Acad. Sci. 112 (2015) E5391–E5400. https://doi.org/10.1073/pnas.1515737112.

[41] R. Goering, L.I. Hudish, B.B. Guzman, N. Raj, G.J. Bassell, H.A. Russ, D. Dominguez, J.M. Taliaferro, FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences, Elife. 9 (2020). https://doi.org/10.7554/eLife.52621.

[42] E. Wang, R. Thombre, Y. Shah, R. Latanich, J. Wang, G-Quadruplexes as pathogenic drivers in neurodegenerative disorders, Nucleic Acids Res. 49 (2021) 4816–4830. https://doi.org/10.1093/nar/gkab164.

[43] S. Vasudevan, J.A. Steitz, AU-Rich-Element-Mediated Upregulation of Translation by FXR1 and Argonaute 2, Cell. 128 (2007) 1105–1118. https://doi.org/10.1016/J.CELL.2007.01.038.

[44] A.B. Herman, C.N. Vrakas, M. Ray, S.E. Kelemen, M.J. Sweredoski, A. Moradian, D.S. Haines, M. V. Autieri, FXR1 Is an IL-19-Responsive RNA-Binding Protein that Destabilizes Pro-inflammatory Transcripts in Vascular Smooth Muscle Cells, Cell Rep. (2018). https://doi.org/10.1016/j.celrep.2018.07.002.

[45] M.-C. Didiot, Z. Tian, C. Schaeffer, M. Subramanian, J.-L. Mandel, H. Moine, The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer., Nucleic Acids Res. 36 (2008) 4902–12. https://doi.org/10.1093/nar/gkn472.

[46] E. Bechara, L. Davidovic, M. Melko, M. Bensaid, S. Tremblay, J. Grosgeorge, E.W. Khandjian, E. Lalli, B. Bardoni, Fragile X related protein 1 isoforms differentially modulate the affinity of fragile X mental retardation protein for G-quartet RNA structure, Nucleic Acids Res. 35 (2006) 299–306. https://doi.org/10.1093/nar/gkl1021.

[47] 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–W682. https://doi.org/10.1093/nar/gkl253.

[48] A.R. Gruber, R. Lorenz, S.H. Bernhart, R. Neubock, I.L. Hofacker, The Vienna RNA Websuite, Nucleic Acids Res. 36 (2008) W70–W74. https://doi.org/10.1093/nar/gkn188.

[49] Y.M. Athar, S. Joseph, The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through Its RGG and C-Terminal Domains, Biochemistry. 59 (2020) 3813–3822. https://doi.org/10.1021/acs.biochem.0c00534.

[50] J. Karijolich, Y.-T. Yu, Converting nonsense codons into sense codons by targeted pseudouridylation, Nature. 474 (2011) 395–398. https://doi.org/10.1038/nature10165.

[51] I.S. Fernández, C.L. Ng, A.C. Kelley, G. Wu, Y.-T. Yu, V. Ramakrishnan, Unusual base pairing during the decoding of a stop codon by the ribosome., Nature. 500 (2013) 107–10. https://doi.org/10.1038/nature12302.

[52] L. Davidovic, N. Durand, O. Khalfallah, R. Tabet, P. Barbry, B. Mari, S. Sacconi, H. Moine, B. Bardoni, A Novel Role for the RNA–Binding Protein FXR1P in Myoblasts Cell-Cycle Progression by Modulating p21/Cdkn1a/Cip1/Waf1 mRNA Stability, PLoS Genet. 9 (2013) e1003367. https://doi.org/10.1371/journal.pgen.1003367.

[53] S. Adinolfi, A. Ramos, S.R. Martin, F. Dal Piaz, P. Pucci, B. Bardoni, J.L. Mandel, A. Pastore, The N-Terminus of the Fragile X Mental Retardation Protein Contains a Novel Domain Involved in Dimerization and RNA Binding †, Biochemistry. 42 (2003) 10437–10444. https://doi.org/10.1021/bi034909g.

Chapter 4:

The Fragile X Protein Disordered Regions Bind a Novel RNA Target

Abstract

The fragile X proteins (FXPs) are a family of RNA-binding proteins that regulate mRNA translation to promote proper neural development and cognition in mammals. Of particular interest to researchers is fragile X mental retardation protein (FMRP), as its absence leads to a neurodevelopmental disorder, fragile X syndrome (FXS), the leading monogenetic cause of autism spectrum disorders. A primary focus of research has been to determine mRNA targets of the FXPs *in vivo* through pull-down assays and validate them through *in vitro* binding studies; another approach has been to perform *in vitro* selection assays to identify RNA sequence and structural targets. These mRNA targets can be further investigated as potential targets for FXS therapeutics. The most well-known RNA structural target of this family of proteins is the G-quadruplex. In this article we report a novel RNA target that is about 100 nucleotides and binds to all three fragile X proteins with nanomolar dissociation constants. Furthermore, we determined that the last 102 amino acids of FMRP, which includes the RGG motif, were necessary and sufficient to bind this novel RNA target. To the best of our knowledge, this is the first example of a non-G-quadruplex structure to be bound by the RGG motif/C-termini of FMRP.

Introduction

Loss of fragile X mental retardation protein (FMRP) has been causally connected to fragile X syndrome (FXS), the most common form of inherited intellectual disability and other neurological disorders, such as schizophrenia, major depressive disorder, and bipolar disorder [1-2]. FXS occurs because of the hypermethylation of CGG trinucleotide repeat expansion in the 5' UTR region of the *FMR1* gene, leading to transcriptional silencing of FMRP in FXS patients. This loss of *FMR1* expression is also the primary monogenic cause of autism spectrum disorders [3]. An estimated 1/4000 males and 1/5000-1/8000 females have a full mutation of FXS while an estimated 1/59 of eight-year-old children in the United States have autism spectrum disorders [4-5].

FMRP, along with its paralogs, fragile-X related protein 1 (FXR1P) and fragile-X related protein 2 (FXR2P), comprise the fragile X protein (FXP) family. Similar to FMRP, FXR1/2P are also implicated in mammalian development. FXR1P is thought to undergo alternative splicing, which alters isoform balance leading to varying diseases, such as Facio-Scapulo Humeral Dystrophy, myotonic dystrophy, and diabetes [6-9]. Further, deficiency of FXR1P is neonatal lethal in mice [10]. FXR2P deficiencies have been implicated in impeding the development of neuronal dendrites, leading to shorter dendrites and thus less neural connectivity [11]. Additionally, mice that lack FXR2P demonstrate altered behaviors and atypical gene expression in their brains [12-13].

Notably, the FXPs are almost completely homologous until exon 13, with high amino acid identity for the first 58-70% of their sequence (72-77% identity), after which

they begin to diverge [14]. From this point on, the amino acid sequences have much lower similarity (31-61% identity), and this divergence has been hypothesized to contribute to their varying functions [14]. Within their sequences, the FXPs all contain three K homology (KH) domains—KH0, KH1, KH2—that are well conserved and a poorly conserved arginine-glycine-glycine (RGG) motif to help with RNA binding **(Figure 4.1A-C)** [15]. Within the N-terminal region of the FXPs are two Agenet domains which have been shown to bind methylated lysines, as well as a region allowing the FXPs to dimerize [16-18]. In the C-termini of the FXPs, an intrinsically disordered region (IDR) comprises about 30-43% of the overall sequences [14]. It is worth mentioning that IDRs have been found to be enriched in RNA-binding proteins, like the FXPs, and have the capability to bind specifically and non-specifically to RNAs [19]. Another unique trait of FXR1P/2P is the presence of arginine-rich motifs within their IDRs that have the potential to impart these paralogs with unique RNA-binding capabilities. Due to their similarity to the nucleolar-targeting signal (NoS) of the protein Rev of human immunodeficiency virus type 1, these sequences are referred to as NoS1 and NoS2 respectively [20]. In other proteins, these motifs have been shown to support RNA recognition [21].

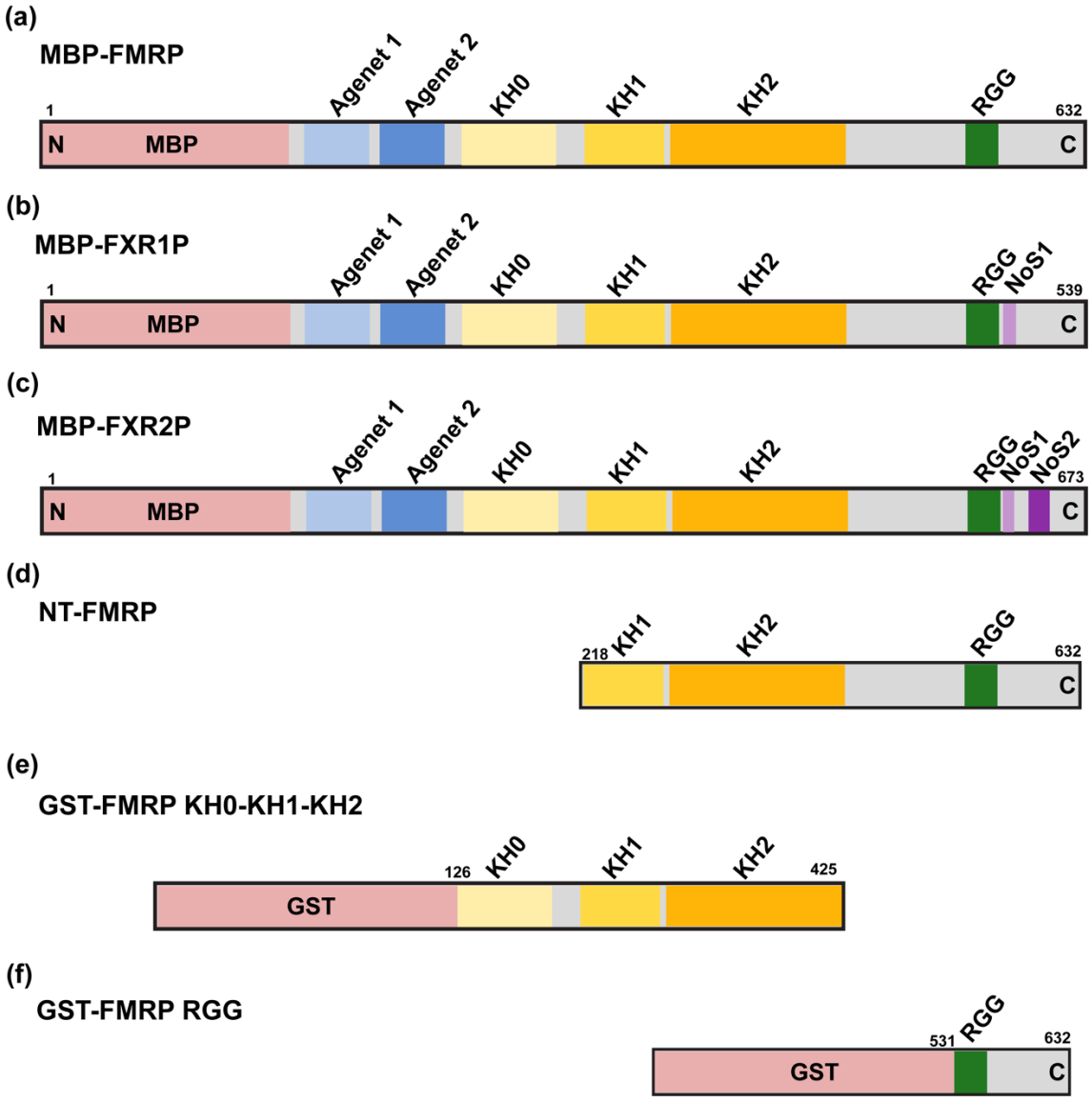


Figure 4.1: Diagram of the fragile X protein constructs. (A-C) The fragile X proteins are multidomain proteins containing tandem agenet domains, three KH domains, an RGG motif, and in FXR1/2P, nucleolar-targeting signals. All three fragile X proteins were purified as described, with a maltose-binding protein (MBP) tag [14]. Human FMRP isoform 1, human FXR1P isoform 2, and human FXR2P were used. (D-F) Schematics of FMRP protein constructs used to dissect FMRP's RNA-binding capabilities. The KH domains and the RGG motif/C-terminal region of FMRP were purified with a glutathione S-transferase tag. These constructs were purified as described previously [22].

As RNA-binding proteins, the FXP family is thought to play a role in mRNA translation regulation. For instance, FMRP's function in translational regulation is particularly important, given its established role in regulating the translation of neuronal mRNAs, and its association with around 4% of fetal neuronal mRNAs [23-24]. Previous studies have suggested that FMRP stalls ribosomal translocation on target mRNAs as a method to regulate translation [22,25]. It has been hypothesized that the loss of this regulation contributes to disease pathogenesis [25]. As a second example, *p21* expression was found to be regulated by FXR1P. When FXR1P is depleted, *p21* protein is found to be upregulated and this upregulation leads to premature cell cycle exit. Upregulation was found to occur in muscle cells, causing abnormalities and contributing to the pathogenesis of Facio-Scapulo Humeral Dystrophy [26]. In some instances, the fragile X proteins have been found to work in tandem to regulate an mRNA's translation. For example, FMRP and FXR2P were found to work in tandem to regulate neuronal maturation, with FXR2P stabilizing *GluA1* mRNA, which enhances protein expression, and FMRP promoting membrane delivery of GluA1, an AMPA receptor [11].

There is a plethora of evidence suggesting that the FXPs work with one another and other proteins to regulate translation of mRNAs within RNA granules. All three FXPs can form both homomeric and heteromeric complexes with the other FXPs [27], and the formation of heteromeric complexes can inhibit the RNA-binding abilities of the FXPs [28]. Furthermore, FMRP is involved in regulating the formation of different RNA granules, such as mRNA ribonucleoprotein complexes [29]. These complexes transport and localize mRNA to specific synaptic sites [30], inhibiting translation until

the mRNA has localized to the specified site [30-31]. Additionally, FMRP can promote stress granules formation, which include RNA-binding proteins as well as translationally repressed mRNAs that respond to cellular stress [32-33]. This has been found to be quite physiologically important, as abnormal formation of stress granules has been linked to a litany of neurodevelopmental disorders [34]. Instances such as these demonstrate the vast effects the FXPs have on mRNA and translational regulation.

RNA targets of FMRP have been successfully determined through various experiments. Notably, *in vitro* selection experiments determined that FMRP binds to a G-quadruplex structure with its RGG motif [35-37]. Structural studies have also demonstrated that upon binding, the G-quadruplex structure stabilizes the inherently disordered RGG motif [38]. Moreover, *in vivo* studies have supported the G-quadruplex as a target of the FXPs [37,39,43]. Numerous studies utilizing *in vivo* or *in vitro* enrichment have identified new mRNA targets that help elucidate FMRP's role in neural disease pathogenesis [40-43]. Pull-down assays have also been successful in identifying mRNA targets, though this presents several ambiguities: (1) are mRNA targets bound based on sequence or structure, and (2) is the interaction physiologically relevant [44-45]? Overall, the wide-ranging relationships between mRNA and the FXPs demonstrate the importance of identifying mRNA targets to facilitate understanding of the molecular mechanisms behind various neural and muscular disease pathogeneses.

While validating mRNA targets of FMRP, we discovered a novel RNA target. Our target is bound by all the FXPs, and this interaction appears to be predominantly

due to the RGG motif/C-terminal region of the FXPs, even though this RNA does not form a G-quadruplex structure. To the best of our knowledge, this is the first example of a non-G-quadruplex forming RNA that is bound by the RGG motif/C-terminus of FMRP. This is especially notable as the RGG motif is found in proteins besides the FXPs. For instance, RGG sequence motifs in the nucleic acid binding protein translocated in liposarcoma (TLS), the DEAD box RNA helicases Dbp2 and Ded1, and hnRNPA1 all bind G-quadruplex structures [46-48]. Thus, our target is a unique target of the RGG motif/C-terminus of the FXPs, and it may be a yet undiscovered target of RGG motifs of other proteins as well.

Results

4.1 A new RNA target of the FXPs is located within the Microreporter mRNA

We initially discovered our novel RNA target by testing the binding of the FXPs to two reporter RNAs we created, referred to as Microreporters 1 and 2 **(Figure 4.2A).** Microreporter 1 contained a G-quadruplex sequence from *p21* mRNA to which we found all the FXPs can bind (prior work, accepted for publication in Journal of Molecular Biology 2021). Microreporter 2 was made as a negative control and did not contain the G-quadruplex sequence. Surprisingly we saw indiscriminate binding by the FXPs to both Microreporters **(Figure 4.2B).** Next, we tested the FXPs' influence on the translation of each of the Microreporters *in vitro* in rabbit reticulocyte lysate. Translation of the Microreporters was quantified through [35-S] methionine labeling of the resulting

3X-FLAG peptide. The 3X-FLAG peptide was purified using anti-FLAG magnetic beads and the quantity of protein produced was determined by scintillation counting. We observed that the fragile X proteins repressed the *in vitro* translation of both Microreporter mRNAs similarly, regardless of the presence of a G-quadruplex forming structure in the Microreporter 1, and the absence in Microreporter 2 (Figure 2C). Conversely, our control protein, maltose-binding protein (MBP), did not significantly alter translation of either reporter, indicating a specific effect of the FXPs. These results align with our previous findings that the FXPs bind specifically to their target RNAs and repress their translation *in vitro* [14]. More importantly, both the EMSA data and the *in vitro* translation results suggest that there is a common region within both reporters to which the FXPs bind.

4.2 The non-GQ forming RNA target is identified as ΔHelix 1,2

To identify the specific region within the Microreporters to which the FXPs bound, we created an RNA containing only the common region of the Microreporter (from the 5' end through the coding sequence), then further deconstructed this common region into 15 RNA segments with overlapping regions **(Figure 4.3A).** Through electrophoretic mobility shift assays (EMSAs), we observed that FMRP bound a common 136-nucleotide region, as well as to two smaller segments of the common region, referred to as ΔHelix 1 (99 nucleotides) and ΔHelix 1,2 (94 nucleotides), respectively **(Figure 4.3B and Figures 4.4-4.5).** These smaller segments comprise most of the coding sequence of the mRNA, starting within the 3X-FLAG tag sequence,

and ending at the end of the coding sequence. To ensure FMRP was not interacting with the Microreporter's 30-nucleotide poly-A tail, we confirmed previous reports that FMRP does not bind to poly-A RNA [24]. We did not observe significant binding to either 30 or 60 nucleotide poly-A RNAs at FMRP concentrations as high as 1 μM **(Figure 4.6A).**

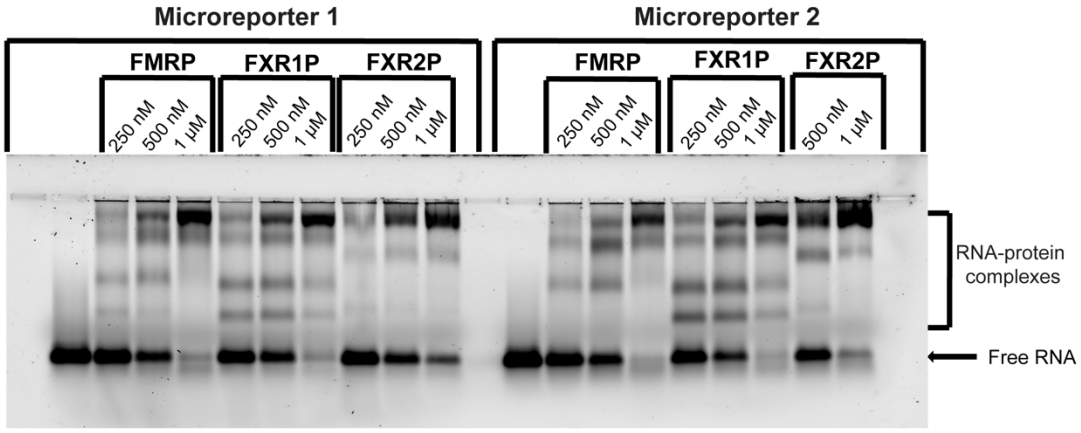
Figure 4.2: The fragile X proteins bind and inhibit the translation of Microreporters 1 & 2 similarly. (A) Microreporters (MR) were designed to mimic mammalian mRNAs, and thus are 5' capped, contain a Kozak sequence, a coding sequence, a variable 3' UTR, and a poly-A tail. The 3' UTR sequences used for each MR are listed; the MR1 3' UTR forms GQ structure, the MR2 3' UTR does not. (B) All three fragile X proteins bound similarly to both Microreporters (MR) even though the Microreporter 1 should form a GQ structure, a verified target of the fragile X proteins. All three proteins bound nearly all the free RNA at concentrations of 1 μM in this EMSA. (C) The fragile X proteins inhibited the translation of both Microreporter 1 & 2 *in vitro*; percent activity in the presence of FMRP (MR1: 63.5 ± 5.4, MR2: 65.3 ± 5.3), FXR1P (MR1: 59.0 ± 5.7, MR2: 67.9 ± 3.9), and FXR2P (MR1: 82 ± 21, MR2: 77 ± 12) are shown. In one of four trials, FXR2P did not inhibit translation of the Microreporter 1, leading to insignificant inhibition for three trials. Conversely, MBP did not inhibit the translation of either Microreporter, suggesting that the inhibition is specific to the fragile X proteins, and is not influenced by the MBP tag (Percent activity in the presence of protein Microreporter 1: 96.6 ± 8.0, Microreporter 2: 100.0 ± 9.0). Results are from three trials (except for FXR2P for which there are four trials) with the error bars showing the standard deviation between trials.

(a)

MR1 3' UTR Sequence: 5'-GGGGGUGUGUGGUGGUGGGGGUGGGCUUAUCUGGGAUGGGGU-3'

MR2 3' UTR Sequence: 5'-GGGCCUCUCUCCUCCUCCCCCUCCCCUUAUCUCCCAUCCCCU-3'

(b)



(c)

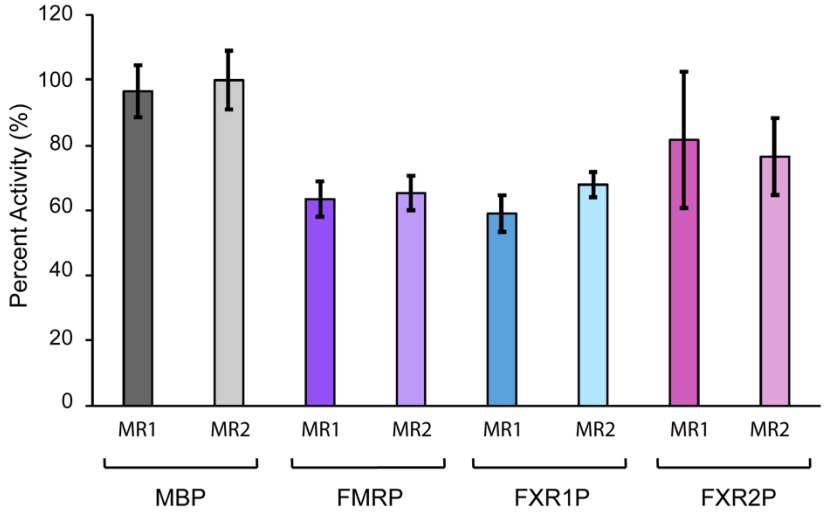
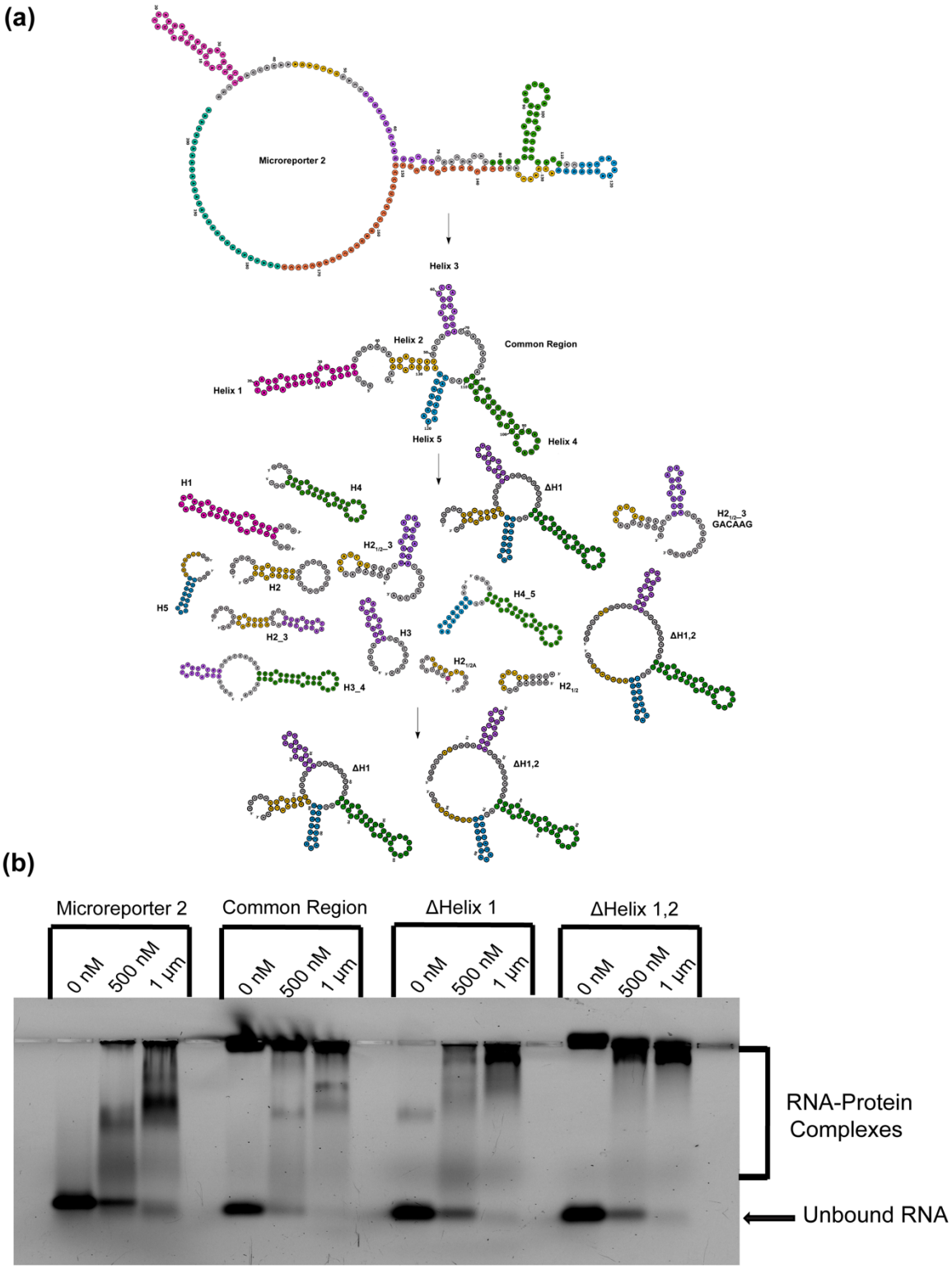
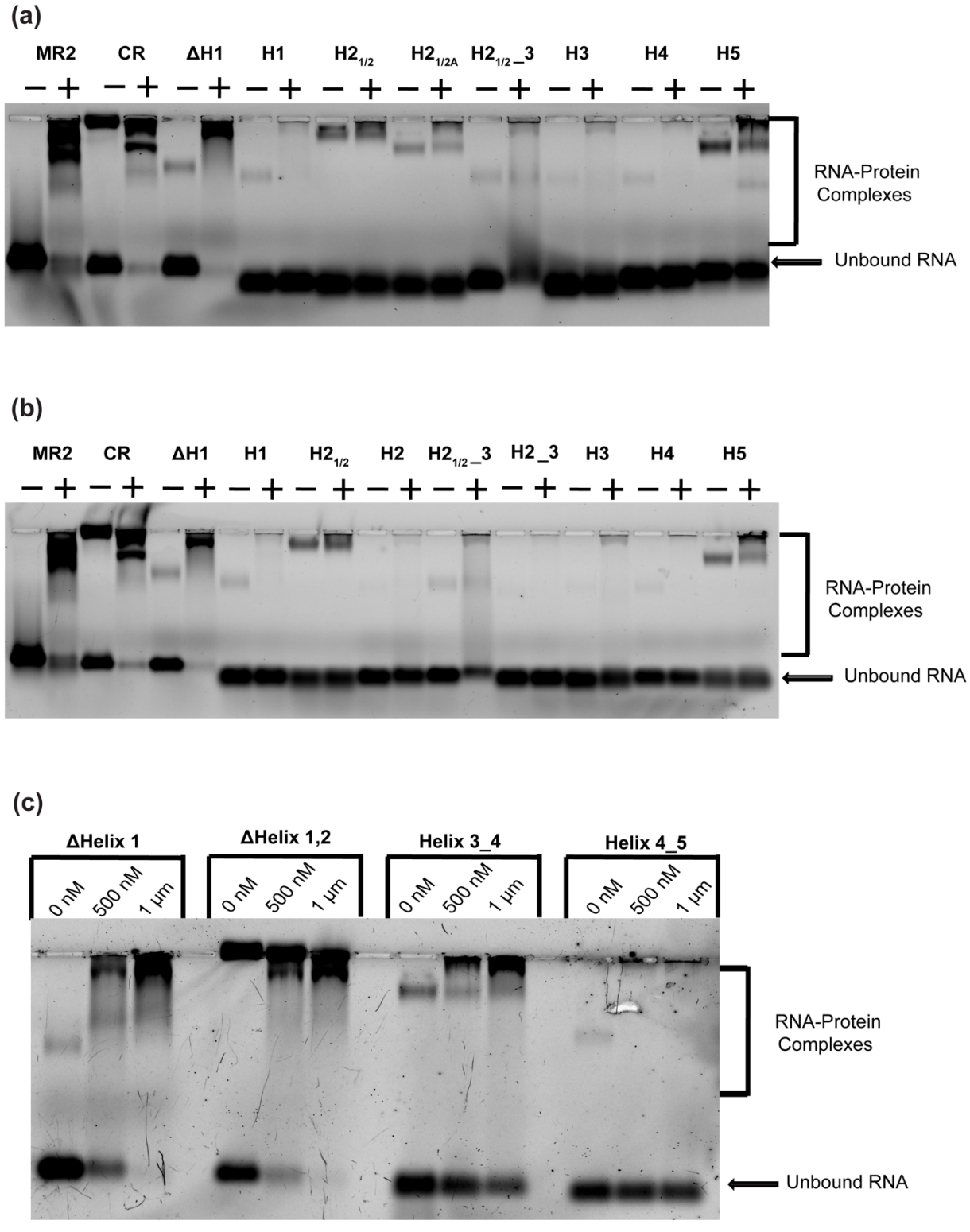


Figure 4.3: Identifying FMRP's RNA target within the Microreporter. (A) Predicted two-dimensional structures of Microreporter, the common region of the Microreporter (5' end through end of coding sequence), and 15 overlapping RNA segments that comprise the common region. ΔHelix 1 and ΔHelix 1,2 were the only segments out of the 15 tested to which FMRP bound. The structure predictions were obtained from RNAfold, with coloration added to delineate each predicted Helix loop within the common region [49]. (B) EMSA showing FMRP's binding to Microreporter 2, the common region, and to ΔHelix 1 and ΔHelix 1,2. For all four RNAs, nearly all the RNA is bound with 1 μM of FMRP. It is important to note that the common region and ΔHelix 1,2 both have a large RNA species/aggregate that is retained in the well even in samples for which no protein was added. FMRP appears to bind the smaller, likely monomeric free RNA that shifts farther (labeled unbound RNA).


(a)
Microreporter 2
Helix 3
Helix 2
Common Region
Helix 1
Helix 5
Helix 4
H1
H4
ΔH1
H2 1/2–3 GACAAG
H2 1/2–3
H2
H5
H4_5
H2_3
H3
ΔH1,2
H2 1/2A
H2 1/2
H3_4
ΔH1
ΔH1,2
(b)
Microreporter 2
Common Region
ΔHelix 1
ΔHelix 1,2
0 nM
500 nM
1 µm
RNA-Protein Complexes
Unbound RNA

While identifying the minimal binding site of the FXPs, we noticed several trends emerging. First, FMRP did not bind to any of the individual helices **(Figure 4.4A-B, see Helices 1-5).** However, for two of the RNAs containing two helices, we saw very weak binding **(Figure 4.4, see Helix $2_{1/2}$ -3 and Helix 3-4).** This led us to predict that these two RNAs must have regions necessary, but not sufficient for complete binding by the FXPs. We also noted that in both these RNAs, Helix 3 was present. Indeed, Helix 3 was also present in ΔHelix 1, which contains Helices 2-5. To further narrow down the essential helices, we created a ΔHelix 1,2, which contains most of the nucleotides of ΔHelix 1, but in which Helix 2 is unable to form. We observed that FMRP bound to ΔHelix 1 and ΔHelix 1,2 similarly, as such the formation of Helix 2 is not critical for binding **(Figure 4.4C).** Thus, through the analysis of 15 subsections of the common region of the Microreporter of 205 nucleotides, we were able to narrow down the FMRP binding site to a 94-nucleotide region we refer to as ΔHelix 1,2, which contains three Helix loop structures (Helices 3-5).

Figure 4.4: ΔHelix 1 and ΔHelix 1,2 are the smallest RNA segments to which FMRP shows noteworthy binding. (A-B). FMRP's binding to Microreporter 2 (MR2), the common region of the Microreporter (5' end through end of coding sequence), and 15 overlapping RNA segments that comprise the common region of the Microreporter were compared via an EMSA. FMRP only showed strong binding (nearly full shift at 1 μM FMRP) to segments ΔHelix 1 and ΔHelix 1,2, 99 and 94 nucleotides, respectively. Very slight binding was observed for Helix 21/2 _3 and Helix 3_4, indicating that these regions are likely necessary for binding, but not sufficient. This is consistent with the fact that ΔHelix 1 contains Helices 2-5, and ΔHelix 1,2 contains Helices 3-5 and most nucleotides from Helix 2. Several of the RNAs have a faint RNA species in addition to the free RNA that was present even in samples for which no protein was added. FMRP may interact with these RNA species, which are likely intermolecular dimers, but for our study, we were concerned with the binding to the smaller, likely monomeric, RNAs that shift farther (labeled unbound RNA).


(a)
MR2
CR
ΔH1
H1
H2 1/2
H2 1/2A
H2 1/2 _3
H3
H4
H5
RNA-Protein Complexes
Unbound RNA
(b)
MR2
CR
ΔH1
H1
H2 1/2
H2
H2 1/2 _3
H2 _3
H3
H4
H5
RNA-Protein Complexes
Unbound RNA
(c)
ΔHelix 1
ΔHelix 1,2
Helix 3_4
Helix 4_5
0 nM
500 nM
1 µm
RNA-Protein Complexes
Unbound RNA

4.3 ΔHelix 1 and ΔHelix 1,2 contain a novel FXP RNA target

Upon observing this binding, we became interested in determining the RNA structure or sequence responsible for the FXPs' binding. To the best of our knowledge, FMRP was not binding to any previously reported FMRP RNA structural or sequence targets within ΔHelix 1 or ΔHelix 1,2. Additionally, there is no potential predicted for G-quadruplex formation [49-50], and we searched for previously reported target sequences such as ACUK and GACR [45]. At first, we predicted binding of these RNAs was the result of the presence of five GACR sequences. However, we tested a 43-nucleotide subsection (named Helix $2_{1/2}$-3) containing all five GACR's and found greatly reduced binding compared to ΔHelix 1 **(Figure 4.4).** We also tested the binding of FMRP to a 3X repeat of the sequence GACAAG, and still did not observe binding **(Figure 4.5).** Finally, although we observed very weak binding to Helix $2_{1/2}$-3 and Helix 3_4, we only observed complete binding of free RNA with ΔHelix 1 and ΔHelix 1,2 which contain most of the nucleotides of Helix 2, as well as Helices 3-5 **(Figure 4.4).** As such, we were intrigued by the potential for ΔHelix 1 or ΔHelix 1,2 to contain a novel FMRP RNA sequence or structural target. Finally, to ensure that FMRP's binding to ΔHelix 1 and ΔHelix 1,2 was not due to their size, we compared these RNAs to a 95 nucleotide RNA (95KS) and to a 79 nucleotide RNA (No 3' UTR NanoFX mRNA). We observed no binding to the 79 nucleotide RNA, and negligible binding to the 95 nucleotide RNA, indicating that FMRP's binding to ΔHelix 1 and ΔHelix 1,2 is not solely due to their size (Figure 4.6B). For simplicity, we completed the remainder of our analyses with ΔHelix 1 only.

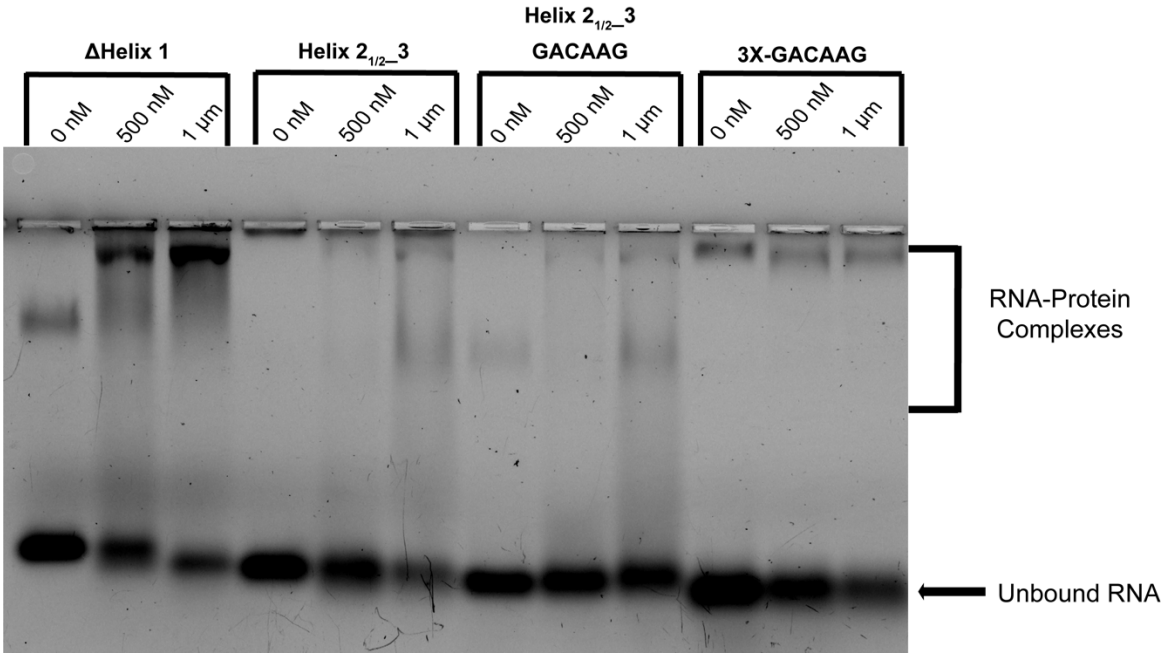


Figure 4.5: FMRP does not bind to GACAAG RNA. (A) FMRP binds to ΔHelix 1 RNA, indicated by the reduction in unbound RNA, and the appearance of a defined RNA-protein complex band. FMRP binds weakly to both Helix $2_{1/2}$ _3 and Helix $2_{1/2}$ _3 GACAAG. The addition of the GACAAG nucleotides at the end of the Helix $2_{1/2}$ _3 does not appear to increase the binding of FMRP to Helix $2_{1/2}$ _3. Additionally, FMRP does not bind to a 3X-GACAAG sequence.

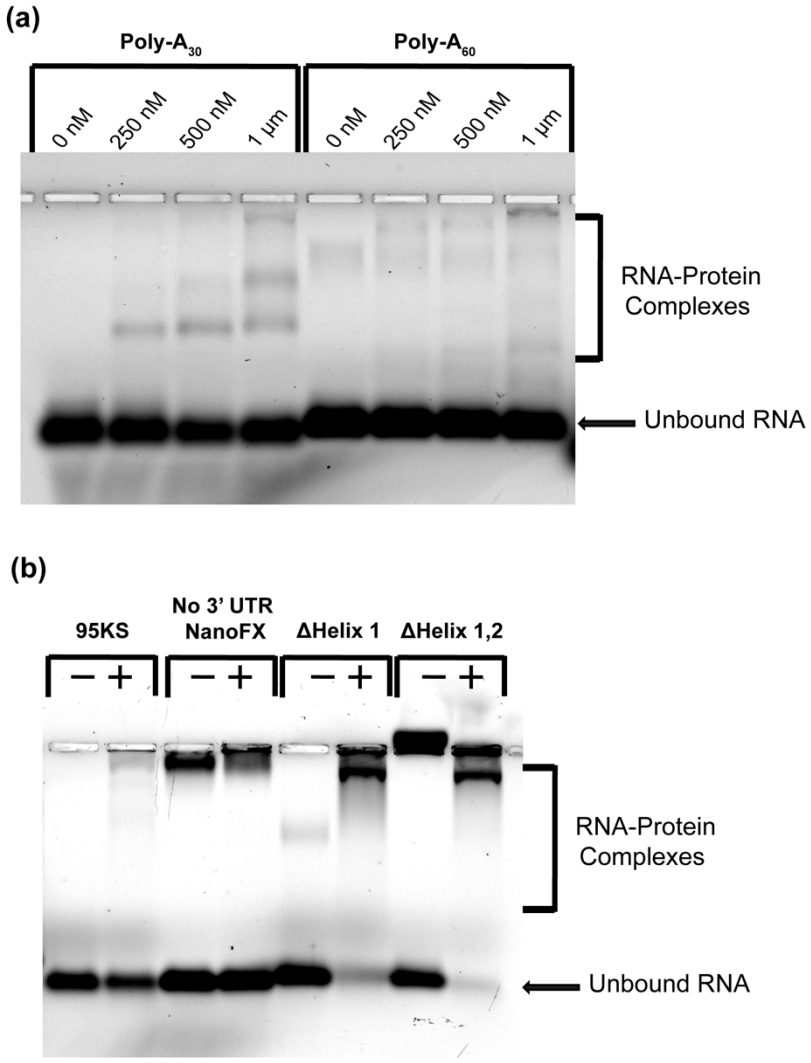


Figure 4.6: FMRP does not bind to Poly-A RNA or RNAs with similar size to ΔHelix 1 and ΔHelix 1,2. (A) FMRP's lack of binding to poly-A RNA can be observed at concentrations as high as 1 µM. (B) FMRP does not bind to a 95 nucleotide RNA (95KS) or to a 79 nucleotide RNA (No 3' UTR). Thus, binding to ΔHelix 1 and ΔHelix 1,2 (99 and 94 nucleotides, respectively) is not likely due simply to their size.

4.4 All three FXPs bind similarly to ΔHelix 1

Next, we tested if FMRP's paralogs, FXR1P and FXR2P, were also capable of binding this novel target RNA using a native EMSA. Our titration indicated that all three FXPs show similar affinities for ΔHelix 1: all three proteins appear to shift nearly all the free RNA at protein concentrations of 1 μM **(Figure 4.7).** However, the shifting pattern appears to be unique for FXR1P. While FMRP and FXR2P form a high molecular weight RNA-protein complex, as can be seen by the small shifts out of the well, FXR1P seems to form several complexes of different molecular masses. These complexes of varying sizes on our native gel are indicative of RNA-protein complexes of varying sizes. As increasing the amount of protein leads to a shift towards a higher molecular mass species, it is likely that the FXPs oligomerize and bind to the RNA, perhaps forming ribonucleoprotein (RNP) granules.

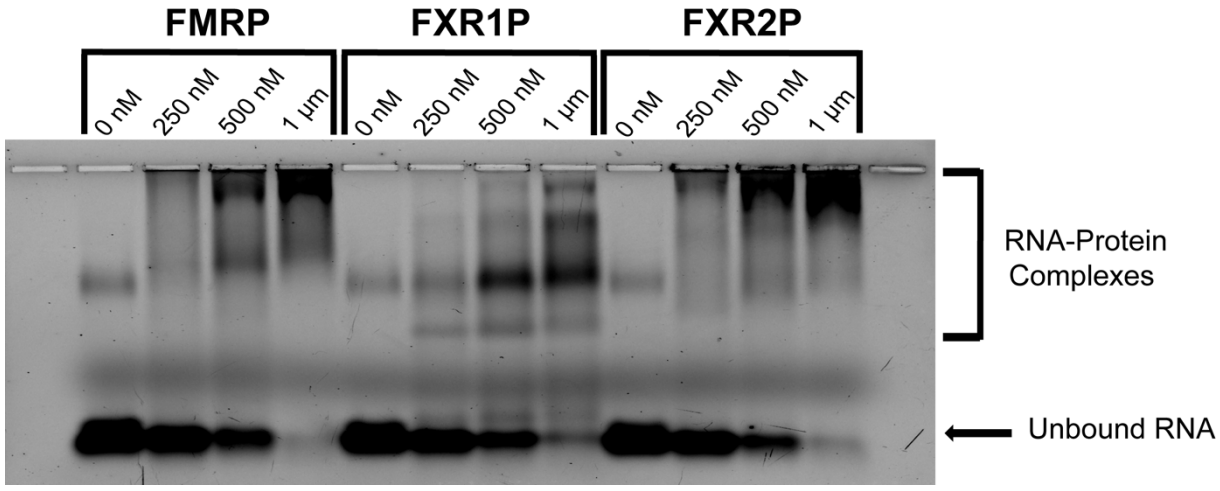
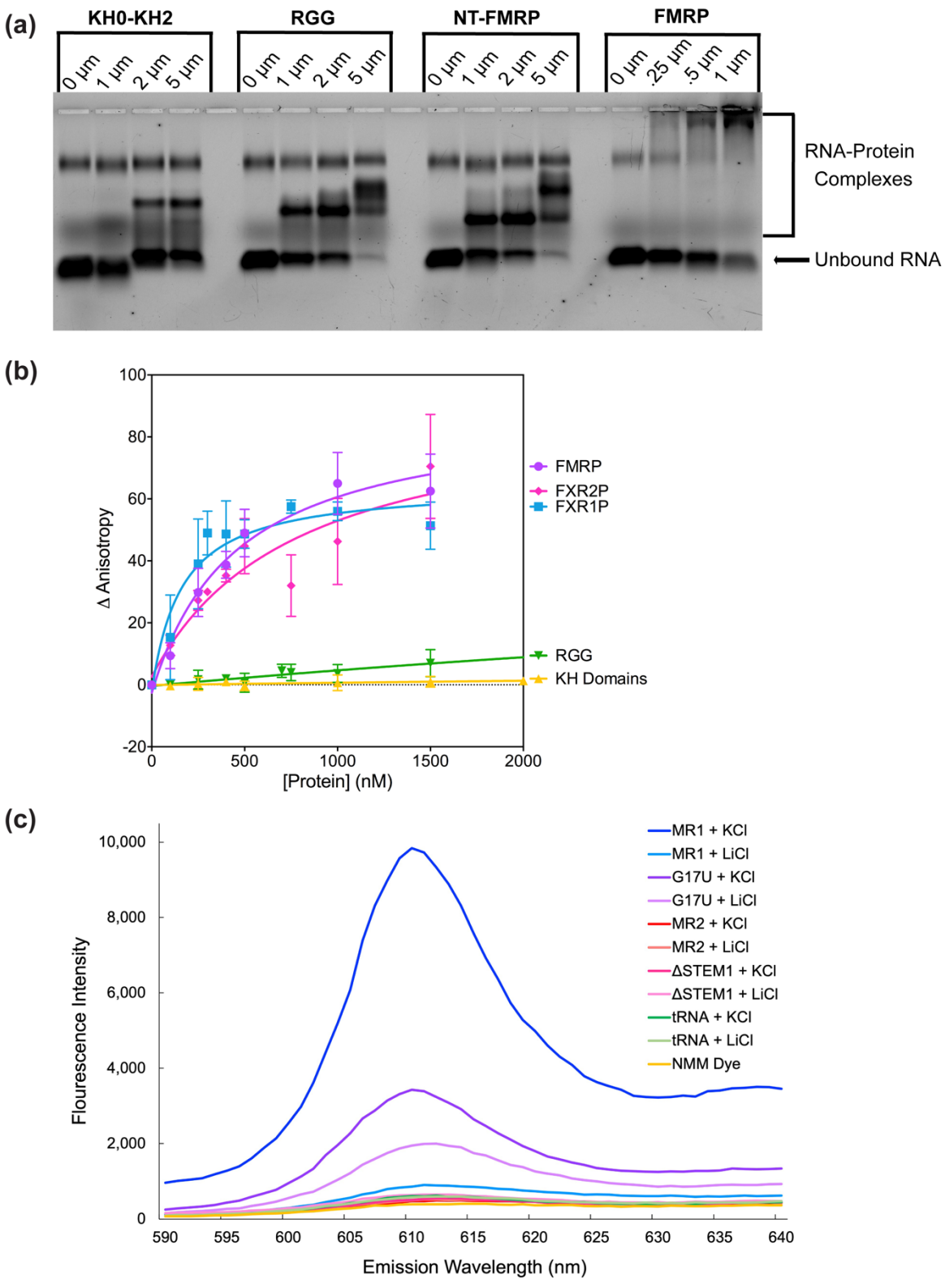


Figure 4.7: All three fragile X proteins bind ΔHelix 1 with similar affinity. (A) All three proteins bind at protein concentrations as low as 250 nM and bind nearly all the free RNA at concentrations of 1 µM in this EMSA. ΔHelix 1 has a faint RNA species in addition to the free RNA that was present even in samples for which no protein was added. FMRP may interact with this RNA species, which is likely an intermolecular dimer, but for our study, we are concerned with the binding to the smaller, likely monomeric, RNA that shifts farther (labeled unbound RNA).

4.5 The RGG/C-termini is critical for FMRP binding to ΔHelix 1

After observing that all the FXPs were able to bind ΔHelix 1, we hoped to gain more insight into the nature of this interaction by elucidating the region of FMRP responsible for binding. Using EMSAs, we compared the binding of ΔHelix 1 by full-length FMRP to various FMRP constructs containing various combinations of its RNA-binding domains. As FMRP contains two types of RNA-binding domains, the KH domains and the RGG motif, the constructs we selected were: an N-terminally truncated FMRP (missing the tandem Agenet domains and KH0), GST-KH0-KH2, and GST-RGG **(Figure 4.1).** We observed that the RGG motif-containing protein appears necessary and sufficient for binding to ΔHelix 1 **(Figure 4.8A).** This was especially apparent when comparing GST-RGG and NT-FMRP, which showed similar binding to ΔHelix 1 despite the inclusion of the KH1-KH2 domains and the entire C-terminal region in NT-FMRP. Interestingly, we observed that an RGG peptide (ARGDGRRRGGGGRGAGGR) which comprises part FMRP's RGG motif did not bind to ΔHelix 1 **(Figure 4.9A).** Thus, the full RGG motif of FMRP must have other features necessary for binding, or the amino acids of FMRP's C-terminus following the RGG motif are critical.

Figure 4.8: The RGG motif/C-terminus of FMRP is responsible for binding ΔHelix 1 RNA. (A) We observed near complete binding of ΔHelix 1 RNA at 5 μM concentrations by the RGG motif/C-termini and the NT-FMRP, whereas the KH domains showed slight binding. The strongest binding affinity was demonstrated by full-length FMRP. (B). Fluorescence anisotropy results revealed no binding of ΔHelix 1 RNA by the KH domains. The RGG motif/C-termini of FMRP showed weaker binding than the full-length fragile X proteins (FMRP KD = 450 ± 120 nM, FXR1P KD = 157 ± 59 nM, FXR2P KD = 780 ± 380 nM, and GST-RGG KD = 11 ± 6 μM). (C). A NMM assay confirmed that the Microreporter 1 (MR1) can form a G-quadruplex, similar to the positive control (G17U) if refolded in the presence of KCl, whereas the Microreporter 2 (MR2) and ΔHelix 1, similar to the negative control (tRNA), cannot.


(a)
KH0-KH2
RGG
NT-FMRP
FMRP
0 µm
1 µm
2 µm
5 µm
0 µm
1 µm
2 µm
5 µm
0 µm
1 µm
2 µm
5 µm
0 µm
.25 µm
.5 µm
1 µm
RNA-Protein Complexes
Unbound RNA
(b)
Δ Anisotropy
[Protein] (nM)
FMRP
FXR2P
FXR1P
RGG
KH Domains
(c)
Flourescence Intensity
Emission Wavelength (nm)
MR1 + KCl
MR1 + LiCl
G17U + KCl
G17U + LiCl
MR2 + KCl
MR2 + LiCl
ΔSTEM1 + KCl
ΔSTEM1 + LiCl
tRNA + KCl
tRNA + LiCl
NMM Dye

Although GST-RGG and NT-FMRP bind ΔHelix 1, their binding was much weaker than full-length FMRP: 1 μM of FMRP shifts most of the free ΔHelix 1 RNA, whereas similar results are not seen until 2-5 μM of GST-RGG and NT-FMRP are added **(Figure 4.8A).** It appears that the KH0-KH2 domains show slight binding at protein concentrations above 2 μM. We came to this conclusion as there is a band for an RNA-protein complex at 2 and 5 μM, but the majority of the RNA is shifted further. Although it is unusual that the free RNA portion appears to migrate a bit higher than in the 0 and 1 μM conditions, we believe this still corresponds to free RNA. Thus, we propose that the enhanced binding of the full-length protein is due to a coordinated effort of all the RNA-binding domains, or perhaps the dimerization capability of FMRP is important to enhance its binding to RNAs.

To further characterize the differences in binding to ΔHelix 1 by the FXPs and the domains of FMRP, we utilized a quantitative method: fluorescence anisotropy-based binding assay. Supporting our conclusions from the EMSA results, the KH domains did not show binding to ΔHelix 1 in the concentration range tested **(Figure 4.8B).** In alignment with the results from our EMSA assays, GST-RGG again showed much weaker binding to ΔHelix 1 compared to full-length FMRP. Comparing the three FXPs, we observed nanomolar binding affinities (FMRP K_D = 450 ± 120 nM, FXR1P K_D = 157 ± 59 nM, FXR2P K_D = 780 ± 380 nM).

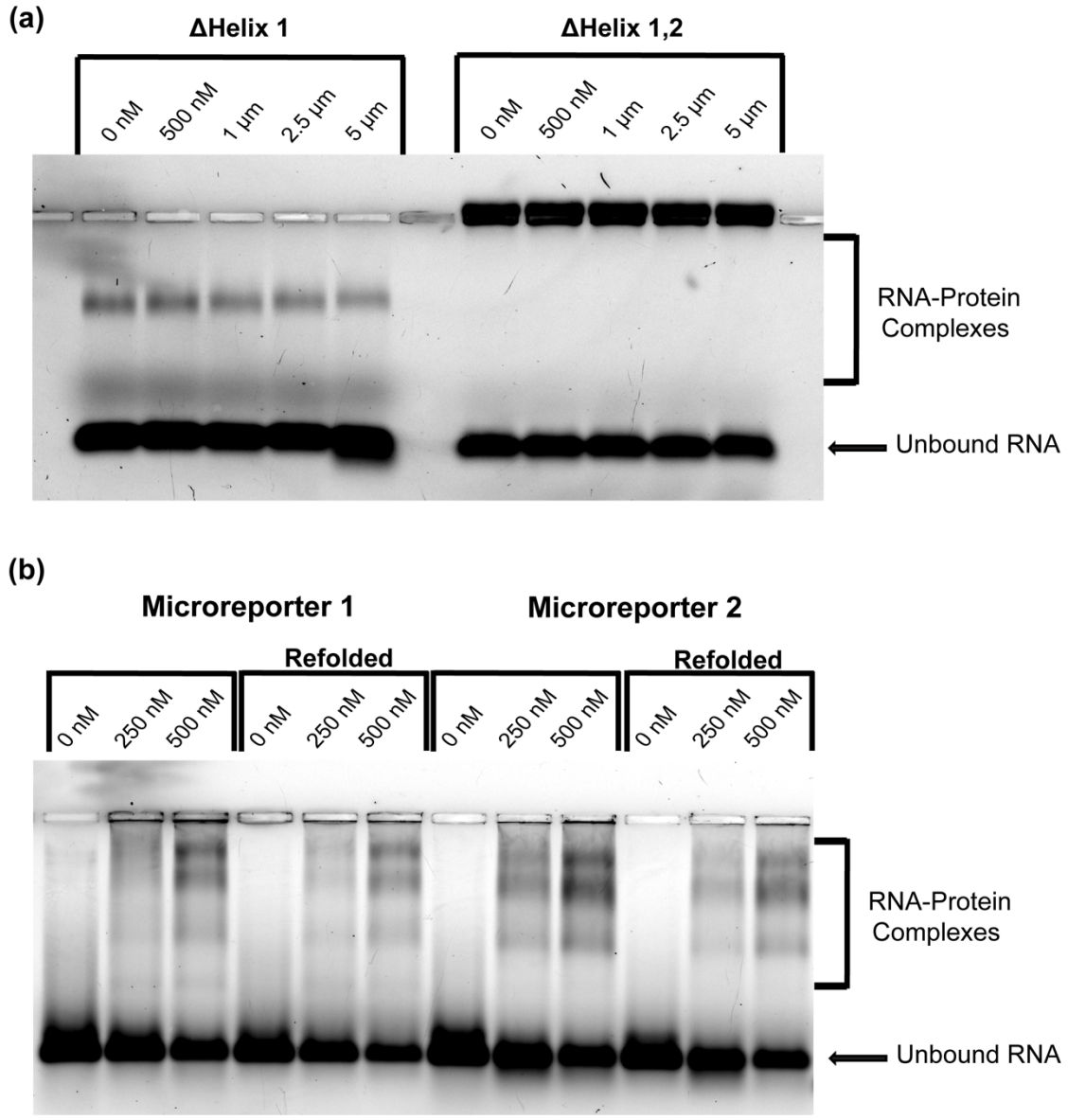


Figure 4.9: An RGG peptide does not bind ΔHelix 1 and ΔHelix 1,2; Refolding does not alter FMRP binding to Microreporters. (A) An RGG peptide which comprises part of the RGG motif of FMRP was unable to bind ΔHelix 1 or ΔHelix 1,2 RNAs. Thus, the RGG motif of FMRP must have unique features of its RGG motif allowing it to bind, or some region in the C-termini following the RGG motif is critical. (B) Refolding Microreporter 1 to form the G-quadruplex structure did not alter FMRP's binding to this RNA. No change was observed for Microreporter 2 which cannot form a G-quadruplex structure.

4.6 ΔHelix 1 is not capable of forming GQ structures

To confirm that binding to ΔHelix 1 was not due to the formation of a G-quadruplex structure, we used both RNAfold and QGRS Mapper to verify that no G-quadruplexes were predicted in ΔHelix 1 or ΔHelix 1,2 [49-50]. Next, we experimentally confirmed these predictions through a N-methyl mesoporphyrin IX (NMM) assay which detects G-quadruplex formation (cite YA and my paper). We observed that only Microreporter 1 and G17U, which contain predicted G-quadruplex structures, showed an increase in fluorescence intensity after refolding the RNAs in the presence of KCl **(Figure 4.8C).** All other RNAs tested, including ΔHelix 1, were incapable of forming the G-quadruplex structure.

4.7 Discussion

Through dissecting and analyzing the FXPs binding to various segments of the original Microreporters (205 nucleotides), we were able to determine 94 and 99-nucleotide sequences that were bound by all the FXPs (ΔHelix 1,2 and ΔHelix 1 respectively). We determined the minimal binding site to be 94-nucleotides that fold into an RNA structure with three helices (Helices 3-5). All three FXPs bound with nanomolar affinities to ΔHelix 1 and inhibited the translation of Microreporters 1 and 2 *in vitro* **(Figure 4.10).** Despite the inclusion of a G-quadruplex target sequence in Microreporter 1, the FXPs bound Microreporters 1 and 2 similarly, whether or not a refolding step was included, suggesting that their interaction with the ΔHelix 1

sequence predominates **(Figure 4.9B)**. This was an interesting finding, as the FXPs have low nanomolar affinities for G-quadruplex structures.

After determining the minimal binding site of the FXPs, we further characterized this RNA-protein interaction by determining which RNA-binding domains of FMRP are responsible for binding this target RNA. The ΔHelix 1 contains four helices in close proximity to one another, so we predicted that this target might be similar to an *in vitro* selected target of FMRP's KH2 domain: the kissing complex, Δkc2 [40]. However, based on our 3D predictions, the helices in ΔHelix 1 do not bind to one another as in a kissing complex [51]. As Δkc2 was the most similar RNA structural target we could find, and our literature search revealed only G-quadruplexes as a known target of the RGG motif of the FXPs, we predicted that the FXPs would bind to ΔHelix 1 through their KH domains. Surprisingly, our results revealed the opposite: the RGG motif/C-terminal region of FMRP was necessary and sufficient for binding to ΔHelix 1, while the KH domains did not appear to contribute much to the binding. Therefore, ΔHelix 1 and ΔHelix 1,2 appear to be unique targets of the RGG motif/C-terminus of the FXPs. It would be interesting to explore if the RNA structure of 3 helices in proximity is also a target of RGG motifs in other RNA-binding proteins.

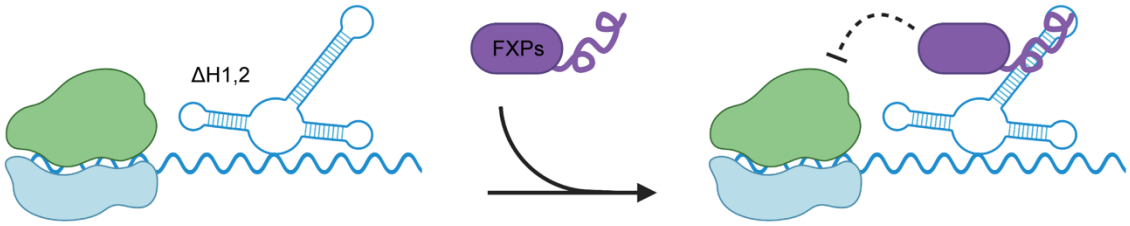


Figure 4.10: The RGG/C-termini of the FXPs binds to ΔHelix 1,2 of the Microreporters to inhibit translation. The proposed mechanism of FXP binding and translation regulation was proposed based on the results from our fluorescence anisotropy, EMSA, and IVT experiments.

As FMRP's paralogs were also capable of binding ΔHelix 1 despite their divergent RGG motifs/C-termini, it is likely that the arginine-rich motifs within their IDRs contribute to binding this RNA. This is supported by findings that the ability of arginines to form electrostatic, hydrogen-bonding, hydrophobic, π- π and cation- π interactions with nucleic acid groups can drive their interaction with RNA. The arginine-rich regions of FXR1P/2P have not been studied much to date yet may provide researchers with a means to explore the RNA-binding capabilities of arginine-rich regions in proteins [52].

We also observed that full-length FMRP shows much greater binding to ΔHelix 1 than either NT-FMRP or GST-RGG. As such, it is possible that the protein sequence prior to the KH1 domain may be critical for this increased binding. As FMRP's dimerization domain is not present in our NT-FMRP construct, perhaps the ability of the FXPs to homerize and heteromerize enhances their binding to certain RNAs. This is supported by the finding that individual RGG motifs appear to bind weakly to RNA, but when linked together, such as by oligomerization regions and domains which can increase the number of RGG instances, it can lead to a high affinity interaction [52]. Another possibility is that the KH domains contribute to binding this RNA only once the FXPs have formed homomers or heteromers. This could explain why many researchers, including our own lab, have had difficulty in identifying RNA targets of the KH domains [22]. In the case of our lab, we have yet to test the binding of a N-terminal construct of FMRP which contains the KH domains and the dimerization domain. In the future, it would be worth investigating this possibility.

We observed binding of ΔHelix 1 by the RGG motif/C-term of FMRP, while the full-length FMRP showed even higher affinity binding to this RNA, perhaps due to the

ability of full-length FMRP to oligomerize. A possible mechanism of translation inhibition for this RNA that is supported by literature is that the RGG motif/C-termini of the FXPs bind to ΔHelix 1 with weak affinity, but the oligomerization of the FXPs enhances this binding. The resulting ribonucleoprotein (RNP) granule then results in mRNA translation stalling and repression [53].

The discovery of a novel RNA target of FMRP and its paralogs opens the door to a potential new class of FXS therapeutic targets. FMRP targets, such as G-quadruplexes, that were discovered *in vitro* have later been found to be present in biologically relevant mRNAs [37]. Therefore, future researchers may wish to determine if this novel *in vitro* RNA target, or a similar structure, is also found in biologically relevant mRNAs. It is possible that the FXPs are targeting either the sequence or structure formed by the CDS of our reporter mRNAs, and it would be interesting to determine if this sequence or structure occurs in authentic mRNAs. Furthermore, this RNA target appears to be the first non-GQ target of the RGG motif/C-termini of the FXPs.

Materials and Methods

4.8 Purification of fragile X proteins

The fragile X proteins FMRP (Q06787), FXR1P (P51114), & FXR2P (P51116), were purified as described previously using a step gradient with KCl [14]. The FMRP constructs, NT-FMRP, KH0-KH2 domains, and the RGG/C-terminus were purified by Youssi Athar as described [22].

4.9 *In vitro* transcription of RNAs

All RNAs used for fluorescence anisotropy or EMSA experiments were transcribed and subsequently 3' labeled with fluorescein as described below.

Oligos containing a T7 promoter sequence were ordered from Integrated DNA Technologies (IDT). Some of the oligos were gel purified prior to annealing with the 18T7T primer. Oligos for the Microreporters, the common region, and ΔHelix 1 were PCR amplified prior to transcription.

For each 100 µL transcription reaction the following reactants were used: 1000 pmol of oligo annealed to 1000 pmol 18T7T primer or 10 µL of the PCR-generated DNA template, 4 mM NTPs, 1X transcription buffer (40 mM Tris pH 8.0, 20 mM $MgCl_2$, 2 mM Spermidine, 0.1% Triton X-100), 5 mM DTT, and ~ 0.27 µg of T7 RNA

polymerase. Each reaction was treated with 2 units of RQ1 DNase (Promega) for 30 min at 37 °C, followed by gel purification on 10-15% denaturing polyacrylamide gels.

RNA Sequences

Microreporter 1 – 205 nts

5’-GGCCGCCACCAUGGACUACAAAGACCAUGACGGCGAUUAUAAAGAUCAU
GACAUCGACUACAAGGACGACGACGACAAGGUUGUGGAGCUGAUCGAGAAGC
ACCACCACCACCACCACAAAGUGGUGAUGAACUAAGGUGUGUGGUGGUGGGG
GUGGGCUUAUCUGGGAUGGGGUAAAAAAAAAAAAAAAAAAAAAAAAAAA- 3’

Microreporter 2 – 205 nts

5’-GGCCGCCACCAUGGACUACAAAGACCAUGACGGCGAUUAUAAAGAUCAUG
ACAUCGACUACAAGGACGACGACGACAAGGUUGUGGAGCUGAUCGAGAAGCA
CCACCACCACCACCACAAAGUGGUGAUGAACUAACCUCUCUCCUCCUCCCCC
UCCCCUUAUCUCCCAUCCCCUAAAAAAAAAAAAAAAAAAAAAAAAAAAA -3’

Common Region – 136 nts

5’-GGCCGCCACCAUGGACUACAAAGACCAUGACGGCGAUUAUAAAGAUCAUG
ACAUCGACUACAAGGACGACGACGACAAGGUUGUGGAGCUGAUCGAGAAGCA
CCACCACCACCACCACAAAGUGGUGAUGAACUAA-3’

ΔHelix 1 – 99 nts

5’-GAUAAAGAUCAUGACAUCGACUACAAGGACGACGACGACAAGGUUGUGG

AGCUGAUCGAGAAGCACCACCACCACCACCACAAAGUGGUGAUGAACUAA-3’

ΔHelix 1,2 – 94 nts

5’-GGAUAAUGACAUCGACUACAAGGACGACGACGACAAGGUUGUGGAGCU

GAUCGAGAAGCACCACCACCACCACCACAAAGUGGUGAUGAACUAA-3’

Helix 1 – 38 nts

5’-GGCCGCCACCAUGGACUACAAAGACCAUGACGGCGAUU-3’

Helix 2 – 37 nts

5’-GGATAAAGATCATGACACGACGACAAACATGAACTAA-3’

Helix $2_{1/2}$ – 18 nts

5’-GUUAUAAAGAUCAUGACA -3’

Helix $2_{1/2A}$ – 19 nts

5’-GAUUAUAAAGAUCAUGACA-3’

Helix 2_3 – 43 nts

5’-GGTAAAGATCATGACATCGACTACAAGGACGAACATGAACTAA-3’

Helix $2_{1/2}$_3 – 43 nts

5’-GUUAUAAAGAUCAUGACAUCGACUACAAGGACGACGACGACAA-3’

Helix $2_{1/2}$_3 GACAAG – 47 nts

5’- GGTTATAAAGATCATGACATCGACTACAAGGACGACGACGACAAGGT-3’

Helix 3 – 30 nts

5’-GGACAUCGACUACAAGGACGACGACGACAA-3’

Helix 4 – 39 nts

5’-GCAAGGUUGUGGAGCUGAUCGAGAAGCACCACCACCACC-3’

Helix 3_4 – 65 nts

5’-GGCAUCGACUACAAGGACGACGACGACAAGGUUGUGGAGCUGAUCGAGA
AGCACCACCACCACAA-3’

Helix 5 – 27 nts

5’-GACCACCACAAAGUGGUGAUGAACUAA-3’

Helix 4_5 – 55 nts

5’-GGAAGGUUGUGGAGCUGAUCGAGAAGCACCACCACCACCACCACAAAGU
GGUGAA -3’

3X-GACAAG – 19 nts

5’-GGACAAGGACAAGGACAAG-3’

4.10 3’ Fluorescein labeling of RNAs

To label the RNA, 0.5 nmoles of RNA was 3′ oxidized for 90 min at room temperature (100 mM KIO_4, 100 mM NaOAc pH 5.2) then incubated with fluorescein 5-thiosemicarbizide (FTSC) at 4 °C overnight (100 mM NaOAc pH 5.2, 1.5 mM FTSC). The RNA was then purified using a Monarch RNA Clean-up Kit (New England BioLabs).

4.11 5’ Capping of Microreporters with Vaccinia Capping Enzyme

To mimic a mammalian mRNA, the Microreporters with a 3’ poly(A) tail were 5’ capped as follows.

Pure RNA was denatured by heating at 65°C for 5 minutes and refolded by cooling on ice for 5 minutes. Each 50 µL reaction consisted of 1X Capping Buffer (New England Biolabs), 0.5 mM GTP, 0.1 mM S-adenosylmethionine (New England Biolabs), Vaccina capping enzyme (0.24 µg/ 10 µg of RNA), RNA (~30 µg), and water. Each reaction was incubated at 37°C for two hours then purified using a Monarch RNA Clean-up Kit (New England BioLabs).

4.12 RNA-binding of fragile X proteins by electrophoretic mobility shift assay

To determine if the proteins bound to the Microreporters, and to identify the binding site within the Microreporter, the binding of each protein to 3' fluorescein labeled RNAs was tested through an electrophoretic mobility shift assay. This assay was selected as the Microreporters are too large to be compatible with fluorescence anisotropy assays.

The purified proteins were centrifuged at 16,100 RCF, 10 min, 4 °C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). Fluorescein-labeled Microreporters were diluted to 10X concentrations (1 μM). Water, 10X binding buffer, protein storage buffer, proteins, and the 10X RNA solution were added in the order listed and mixed for a final reaction volume of 26 μL. The final reactions contained 50 mM Tris pH 7.5, 135 mM KCl, 5 mM $MgCl_2$, 1 μM BSA, 10 mM DTT, 50 ng/μL tRNA (to prevent non-specific binding), ~ 100 nM fluorescein-labeled RNA, and varying concentrations of protein. For each protein concentration tested the total volume of protein + protein storage buffer remained constant. In each reaction the binding buffer was adjusted to account for the Tris pH 7.5, KCl, and DTT that were contributed from the protein storage buffer. The reactions were thoroughly mixed and incubated in the dark at room temperature for 1 h. After incubation, 3 μL of loading dye (xylene cyanol in 50% glycerol) was added to each reaction. A 0.6 or 0.8% agarose gel (SeaKem GTG agarose) was prepared in 1X TBE buffer (100 mM Tris pH 8.3, 100 mM borate, 2 mM Na_2EDTA). After loading 13 μL of

each sample, the gel was run at 4 °C for 2-3 hrs at 66 V in 1X TBE buffer. The gel was then scanned using a laser scanner (Typhoon FLA 9500, GE Healthcare) and the gel was analyzed in ImageJ.

4.13 Analysis of *in vitro* translation regulation by the fragile X proteins

The purified proteins were centrifuged at 16,100 RCF, 10 min, 4 °C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). 2X rabbit reticulocyte lysate (treated with micrococcal nuclease to reduce endogenous mRNAs and without methionine), water, 5′ capped Microreporters with a 30-nucleotide 3′ poly(A) tail, protein storage buffer, the corresponding protein, and L-[^{35}S]-Methionine (PerkinElmer, 10mCi (370MBq), Specific Activity: >1000 Ci (37.0 TBq)/mMole, 50 mM Tricine, 10 mM BME), were combined in the order listed, mixed, and allowed to incubate for ~75 min at 30°C. The final 20 µL reactions contained either 100 nM Microreporter mRNA with ~0.6 µM protein (FXR1/2P) or 200 nM Microreporter mRNA with ~1.3 µM FMRP. After incubation, each reaction was added to 180 µL of 1X binding buffer (25 mM HEPES pH 7.5, 75 mM NaCl, 0.5 mM EDTA, 0.1% Triton-X, 5% Glycerol, 4 mM $MgCl_2$) and incubated with 2.5 uL of equilibrated anti-flag M2 magnetic beads (Sigma-Aldrich) for 2 hours at 4°C. After the incubation, the solution was removed from the beads, and the beads were washed three times with 100 µL of 1X binding buffer. To elute the reporter peptide, 100 µL of FLAG peptide

elution buffer (25 mM HEPES pH 7.5, 100 mM NaCl, 4 mM $MgCl_2$, 200 ng/µL 3X FLAG peptide (Sigma-Aldrich) was added to the beads. The elution was allowed to occur for 45 min at 4°C. The elution was subsequently removed, and the buffer evaporated (SpeedVac Plus sc110a). Once all the liquid was removed, the reimaging pellet was resuspended in 10 µL of H_2O.

To quantify the amount of reporter peptide by scintillation counting, 6 µL of the resuspended pellet solution was aliquoted onto 15 mm circular Whatman papers with a pore size of 11 µm (GE Healthcare LifeSciences), which were dried under a lamp for 10 minutes. The filter paper circles were transferred to scintillation vials and 4 mL of ScintiSafe 300% (Fisher Scientific) was added. The S-35 counts were recorded by reading for 1 min/sample using a scintillation counter (LS 6500, Beckman Coulter).

4.14 Quantifying RNA-binding of the fragile X proteins through fluorescence anisotropy

Prior to fluorescence anisotropy assays, each protein stock was thawed from -80°C then centrifuged at 16,100 RCF, 10 min, 4 °C with a benchtop centrifuge to remove any precipitated protein prior to each experiment. The supernatants containing soluble protein were obtained and the concentration of protein determined using A280 readings (Thermo Scientific NanoDrop 2000/2000c spectrophotometer). RNAs labeled with a 3′ fluorescein were diluted to 5X concentrations (~ 25 nM). Water, binding buffer, protein storage buffer, protein, and the 5X RNA solution were added in the order listed and mixed for a final reaction volume of 200 µL. The final reactions contained 20 mM

Tris pH 7.5, 75-150 mM KCl, 5 mM $MgCl_2$, 1 μM BSA, 1 mM DTT, 100 ng/μL tRNA (to prevent non-specific binding), varying concentrations of protein, and ~ 5 nM RNA. It is important to note that within each trial, for each protein concentration tested, the total volume of protein + protein storage buffer remained constant. In each trial the binding buffer was adjusted to account for the Tris pH 7.5, salt, and DTT that were contributed from the protein storage buffer. Reactions were thoroughly mixed and incubated in the dark at room temperature for 1 h. After incubation, each reaction was added into a 96-well non-binding plate (Greiner Bio-One) for fluorescence anisotropy using a multimode microplate reader (SPARK TECAN). Samples were excited at 485 nm and emission was measured at 535 nm. To determine binding affinities, the anisotropy data from each binding assay were normalized to initial values without protein, plotted, and fit to a quadratic equation as previously described[22]. Three or more independent trials were performed to determine standard deviations.

4.15 N-methyl mesoporphyrin IX (NMM) assay to detect G-quadruplex formation

A stock solution of NMM from Frontier Scientific was prepared as previously described [1]. The stock solution of NMM (8.61 mM in 0.2 N HCl) was diluted to 400 μM in 10% (v/v) DMSO to make a fresh working solution. For each RNA, 2 μM of RNA was refolded in the presence of KCl and LiCl (50 mM Tris pH 7.5, 2 mM $MgCl_2$, 75 mM KCl or LiCl). Refolding occurred for 5 min at 68°C followed by slow cooling to room temperature for ~1 hr. After cooling, 1.25 μL of 400 μM NMM solution was added to 100 μL of refolded RNA to achieve a final concentration of 5 μM NMM. The samples

were then incubated at room temperature for 10 min prior to loading 90 µL of each sample into a 96-well non-binding plate (Greiner Bio-One) for a fluorescence intensity scan using a multimode microplate reader (SPARK TECAN). The samples were excited at 400 nm and the emission scanned at 560-650 nm with 5 nm bandwidths for both excitation and emission. The fluorescence intensity values were plotted as a function of wavelength from 580-640 nm.

4.16 Accession Codes

FMRP: Q06787

FXR1P: P51114

FXR2P: P51116

MBP: P0AEX9

Chapter 4, in full, is currently being prepared for submission for publication as The Fragile X Protein Disordered Regions Bind a Novel RNA Target, Edwards, Madison; Huang, Molly; Joseph, Simpson. The dissertation author was the primary investigator and author of this material.

References

[1] Martin JP, Bell J. A pedigree of mental defect showing sex-linkage. *J Neurol Psychiatry.* 1943; **6** (3–4):154–157. doi: 10.1136/jnnp.6.3-4.154.

[2] Fatemi SH, Folsom TD. The role of fragile X mental retardation protein in major mental disorders. *Neuropharmacology.* **60**, 221-1226 (2011).

[3] R. Lozano, A. Azarang, T. Wilaisakditipakorn, R.J. Hagerman, Fragile X syndrome: A review of clinical management., Intractable Rare Dis. Res. 5 (2016) 145–57. https://doi.org/10.5582/irdr.2016.01048.

[4] Jin, Xiaowei, and Li Chen. Fragile X syndrome as a rare disease in China - Therapeutic challenges and opportunities. *Intractable & rare diseases research* vol. 4,1 (2015): 39-48. doi:10.5582/irdr.2014.01037

[5] J. Baio, L. Wiggins, D.L. Christensen, M.J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C. Robinson, Rosenberg, T. White, M.S. Durkin, P. Imm, L. Nikolaou, M. Yeargin-Allsopp, L.-C. Lee, R. Harrington, M. Lopez, R.T. Fitzgerald, A. Hewitt, S. Pettygrove, J.N. Constantino, A. Vehorn, J. Shenouda, J. Hall-Lande, K. Van, Naarden, Braun, N.F. Dowling, Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014, MMWR. Surveill. Summ. 67 (2018) 1–23. https://doi.org/10.15585/mmwr.ss6706a1.

[6] J.A. Smith, E.G. Curry, R.E. Blue, C. Roden, S.E.R. Dundon, A. Rodríguez-Vargas, D.C. Jordan, X. Chen, S.M. Lyons, J. Crutchley, P. Anderson, M.E. Horb, A.S. Gladfelter, J. Giudice, FXR1 splicing is important for muscle development and biomolecular condensates in muscle cells., J. Cell Biol. 219 (2020). https://doi.org/10.1083/jcb.201911129.

[7] Davidovic L, Durand N, Khalfallah O, Tabet R, Barbry P, Mari B, Sacconi S, Moine H, Bardoni B. A novel role for the RNA-binding protein FXR1P in myoblasts cell-cycle progression by modulating p21/Cdkn1a/Cip1/Waf1 mRNA stability. *PLoS Genet.* 2013 Mar;9(3):e1003367.

[8] Orengo JP, Chambon P, Metzger D, Mosier DR, Snipes GJ, Cooper TA. Expanded CTG repeats within the DMPK 3' UTR causes severe skeletal muscle wasting in an inducible mouse model for myotonic dystrophy. *Proc Natl Acad Sci U S A*. 2008;105(7):2646-2651. doi:10.1073/pnas.0708519105

[9] C.A. Nutter, E.A. Jaworski, S.K. Verma, V. Deshmukh, Q. Wang, O.B. Botvinnik, M.J. Lozano, I.J. Abass, T. Ijaz, A.R. Brasier, N.J. Garg, X.H.T.

Wehrens, G.W. Yeo, M.N. Kuyumcu-Martinez, Dysregulation of RBFOX2 Is an Early Event in Cardiac Pathogenesis of Diabetes., Cell Rep. 15 (2016) 2200–2213. https://doi.org/10.1016/j.celrep.2016.05.002.

[10] E.J. Mientjes, R. Willemsen, L.L. Kirkpatrick, I.M. Nieuwenhuizen, M. Hoogeveen-Westerveld, M. Verweij, S. Reis, B. Bardoni, A.T. Hoogeveen, B.A. Oostra, D.L. Nelson, Fxr1 knockout mice show a striated muscle phenotype: implications for Fxr1p function in vivo, Hum. Mol. Genet. 13 (2004) 1291–1302. https://doi.org/10.1093/hmg/ddh150.

[11] W. Guo, E.D. Polich, J. Su, Y. Gao, D.M. Christopher, A.M. Allan, M. Wang, F. Wang, G. Wang, X. Zhao, Fragile X Proteins FMRP and FXR2P Control Synaptic GluA1 Expression and Neuronal Maturation via Distinct Mechanisms, Cell Rep. 11 (2015) 1651–1666. https://doi.org/10.1016/j.celrep.2015.05.013.

[12] S. Cavallaro, S. Paratore, F. Fradale, F.M.S. de Vrij, R. Willemsen, B.A. Oostra, Genes and pathways differentially expressed in the brains of Fxr2 knockout mice, Neurobiol. Dis. 32 (2008) 510–520. https://doi.org/10.1016/J.NBD.2008.09.008.

[13] C.J.M. Bontekoe, K.L. McIlwain, I.M. Nieuwenhuizen, L.A. Yuva-Paylor, A. Nellis, R. Willemsen, Z. Fang, L. Kirkpatrick, C.E. Bakker, R. McAninch, N.C. Cheng, M. Merriweather, A.T. Hoogeveen, D. Nelson, R. Paylor, B.A. Oostra, Knockout mouse model for Fxr2: a model for mental retardation, Hum. Mol. Genet. 11 (2002) 487–498. https://doi.org/10.1093/hmg/11.5.487.

[14] Edwards, M., Xu, M. & Joseph, S. A simple procedure for bacterial expression and purification of the fragile X protein family. *Sci. Rep.* **10**, 15858 (2020).

[15] Myrick, L. K., Hashimoto, H., Cheng, X. & Warren, S. T. Human FMRP contains an integral tandem Agenet (Tudor) and KH motif in the amino terminal domain. Hum. Mol. Genet. **24**, 1733–1740 (2015).

[16] A. Ramos, D. Hollingworth, S. Adinolfi, M. Castets, G. Kelly, T.A. Frenkiel, B. Bardoni, A. Pastore, The Structure of the N-Terminal Domain of the Fragile X Mental Retardation Protein: A Platform for Protein-Protein Interaction, Structure. 14 (2006) 21–31. https://doi.org/10.1016/J.STR.2005.09.018.

[17] M.A. Adams-Cioaba, Y. Guo, C. Bian, M.F. Amaya, R. Lam, G.A. Wasney, M. Vedadi, C. Xu, J. Min, Structural Studies of the Tandem Tudor Domains of Fragile X Mental Retardation Related Proteins FXR1 and FXR2, PLoS One. 5 (2010) e13559. https://doi.org/10.1371/journal.pone.0013559.

[18] S. Adinolfi, A. Ramos, S.R. Martin, F. Dal Piaz, P. Pucci, B. Bardoni, J.L. Mandel, A. Pastore, The N-Terminus of the Fragile X Mental Retardation

Protein Contains a Novel Domain Involved in Dimerization and RNA Binding †, Biochemistry. 42 (2003) 10437–10444. https://doi.org/10.1021/bi034909g.

[19] Calabretta, S. & Richard, S. Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015).

[20] F. Tamanini, L.L. Kirkpatrick, J. Schonkeren, L. van Unen, C. Bontekoe, C. Bakker, D. Nelson, H. Galjaard, B. a Oostra, a T. Hoogeveen, The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins., Hum. Mol. Genet. (2000). https://doi.org/10.1093/hmg/9.10.1487.

[21] Tan, R. & Frankel, A. D. A novel glutamine-RNA interaction identified by screening libraries in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 4247–52 (1998).

[22] Athar, Y. M. & Joseph, S. RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein. *J. Mol. Biol.* (2020). doi:10.1016/J.JMB.2020.04.021

[23] Chen, E. & Joseph, S. Fragile X mental retardation protein: A paradigm for translational control by RNA-binding proteins. *Biochimie* **114**, 147–154 (2015).

[24] B. V, J. P, C. S, D. JC, O. WT, T. SA, J. X, F. Y, W. KD, K. JD, D. RB, W. ST, Microarray Identification of FMRP-associated Brain mRNAs and Altered mRNA Translational Profiles in Fragile X Syndrome, Cell. 107 (2001). https://doi.org/10.1016/S0092-8674(01)00568-2.

[25] N.E. Clifton, E. Rees, P.A. Holmans, A.F. Pardiñas, J.C. Harwood, A. Di Florio, G. Kirov, J.T.R. Walters, M.C. O'Donovan, M.J. Owen, J. Hall, A.J. Pocklington, Genetic association of FMRP targets with psychiatric disorders, Mol. Psychiatry. (2020) 1–14. https://doi.org/10.1038/s41380-020-00912-2.

[26] L. Davidovic, S. Sacconi, E.G. Bechara, S. Delplace, M. Allegra, C. Desnuelle, B. Bardoni, Alteration of expression of muscle specific isoforms of the fragile X related protein 1 (FXR1P) in facioscapulohumeral muscular dystrophy patients., J. Med. Genet. 45 (2008) 679–85. https://doi.org/10.1136/jmg.2008.060541.

[27] Y. Zhang, J.P. O'Connor, M.C. Siomi, S. Srinivasan, A. Dutra, R.L. Nussbaum, G. Dreyfuss, The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2., EMBO J. 14 (1995) 5358.

[28] E. Bechara, L. Davidovic, M. Melko, M. Bensaid, S. Tremblay, J. Grosgeorge, E.W. Khandjian, E. Lalli, B. Bardoni, Fragile X related protein 1 isoforms differentially modulate the affinity of fragile X mental retardation protein for G-

quartet RNA structure, Nucleic Acids Res. 35 (2006) 299–306. https://doi.org/10.1093/nar/gkl1021.

[29] Bassell GJ, Warren ST. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron*. **60**, 201–214 (2008).

[30] Kiebler M.A., Bassell G.J. Neuronal RNA granules: movers and makers. *Neuron*. **51**, 685-90 (2006).

[31] Lai A, Valdez-Sinon AN, Bassell GJ. Regulation of RNA granules by FMRP and implications for neurological diseases. *Traffic*. 2020;21(7):454-462. doi:10.1111/tra.12733

[32] Didiot M-C, Subramanian M, Flatter E, Mandel J-L, Moine H. Cells lacking the fragile X mental retardation protein (FMRP) have normal RISC activity but exhibit altered stress granule assembly. *Mol Biol Cell*. 2009;20(1):428–437. 10.1091/mbc.e08-07-0737

[33] Kiebler MA, Bassell GJ. Neuronal RNA granules: movers and makers. Neuron. 2006 Sep 21;51(6):685-90. doi: 10.1016/j.neuron.2006.08.021. PMID: 16982415.

[34] C. Gareau, E. Houssin, D. Martel, L. Coudert, S. Mellaoui, M.-E. Huot, P. Laprise, R. Mazroui, Characterization of fragile X mental retardation protein recruitment and dynamics in Drosophila stress granules., PLoS One. 8 (2013) e55342. https://doi.org/10.1371/journal.pone.0055342.

[35] J.C. Darnell, C.E. Fraser, O. Mostovetsky, R.B. Darnell, Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs, Hum. Mol. Genet. 18 (2009) 3164–3177. https://doi.org/10.1093/hmg/ddp255.

[36] Darnell, J.C., Klann E. The translation of translational control by FMRP: therapeutic targets for FXS. *Nature Neuroscience.* **16**, 1530-6 (2013).

[37] 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. https://doi.org/10.1016/s0092-8674(01)00566-9.

[38] 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/10.1038/nsmb.2064.

[39] Blice-Baum AC, Mihailescu MR. 2014. Biophysical characterization of G-quadruplex forming FMR1 mRNA and of its interactions with different fragile X mental retardation protein isoforms. RNA 20:103–114. DOI: https://doi.org/10.1261/rna.041442.113, PMID: 24249225

[40] 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/10.1101/gad.1276805.

[41] 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/10.1038/nsmb.2064.

[42] C.R. Casingal, T. Kikkawa, H. Inada, Y. Sasaki, N. Osumi, Identification of FMRP target mRNAs in the developmental brain: FMRP might coordinate Ras/MAPK, Wnt/β-catenin, and mTOR signaling during corticogenesis., Mol. Brain. 13 (2020) 167. https://doi.org/10.1186/s13041-020-00706-1.

[43] 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.

[44] V. Brown, K. Small, L. Lakkis, Y. Feng, C. Gunter, K.D. Wilkinson, S.T. Warren, Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein., J. Biol. Chem. 273 (1998) 15521–7. https://doi.org/10.1074/jbc.273.25.15521.

[45] Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J. & Warren, S. T. 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 (2014).

[46] Yagi R., Miyazaki T., Oyoshi T., G-quadruplex binding ability of TLS/FUS depends on the β-spiral structure of the RGG domain, *Nucleic Acids Research*, **46**, 5894–5901 (2018).

[47] Yan K. K., Obi I., Sabouri N., The RGG domain in the C-terminus of the DEAD box helicases Dbp2 and Ded1 is necessary for G-quadruplex destabilization, *Nucleic Acids Research*, (2021) doi: 10.1093/nar/gkab620

[48] Ghosh M., Singh M., RGG-box in hnRNPA1 specifically recognizes the telomere G-quadruplex DNA and enhances the G-quadruplex unfolding ability of UP1 domain, *Nucleic Acids Research*, **46**, 10246- 10261, (2018).

[49] Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Res.* **36**, W70–W74 (2008).

[50] Kikin O., D'Antonio L., Bagga P. S., QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences, *Nucleic Acids Research*, **34**, W676–W682 (2006).

[51] Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J., Adamiak, R.W. Automated 3D structure composition for large RNAs, *Nucleic Acids Research*, **2012**, 40(14):e112 (doi:10.1093/nar/gks339).

[52] Chong P. A., Vernon R. M., Forman-Kay J.D., RGG/RG Motif Regions in RNA Binding and Phase Separation, *Journal of Molecular Biology*, **23**, 4650-4665 (2018).

[53] Jain S, Parker R. The discovery and analysis of P Bodies. Adv Exp Med Biol. 2013;768:23-43. doi: 10.1007/978-1-4614-5107-5_3. PMID: 23224963.

Chapter 5:

Conclusion and Future Directions

Our first contribution to the study of FXS and the FXPs was the development of a rapid, simple, and inexpensive recombinant purification protocol for the entire human FXP family. Prior to our method, researchers relied on strategies such as extensive washes to remove contaminant proteins and TPs, purification from inclusion bodies, purification from eukaryotic cells, or purification of specific domains of the FXPs [1-11]. Beyond working for all three proteins in the family, our method is valuable because it allows for the purification of the full-length proteins. To understand how these proteins function *in vivo*, it is best to use full-length proteins for *in vitro* studies whenever possible, as this is more biologically relevant.

Developing this purification protocol was a challenge for multiple reasons. First, the FXPs have multiple ribosomal stalling proline-rich motifs that reduce their recombinant expression. We discovered that by disrupting ribosomal stalling proline-rich motifs within FMRP and FXR2P, we could increase recombinant expression while reducing the production of TPs. This technique, or co-expression with EF-P, may assist in the recombinant expression of eukaryotic proteins, 10% of which possess polyproline motifs [12]. Second, the large intrinsically disordered regions of the FXPs reduce their solubility *in vitro* and promote a tendency to aggregate. To overcome this challenge, we found an MBP tag greatly improved solubility, and likely also enhanced expression of the FXPs [13-14]. Third, as the proteins are RNA-binding proteins, purifying the proteins without nucleic acids was a challenge. We found that eluting from a heparin column with a salt gradient allowed us to remove contaminant proteins, TPs, and nucleic acid contamination. The purified proteins were stable and free of protein and nucleic acid contamination.

We believe our purification protocol will increase the extent of research into the FXPs by reducing barriers to study these proteins *in vitro*. We hope this procedure will mitigate obstacles faced in studying the important roles of the FXP family in translational regulation, and in doing so, promote diverse research questions. More broadly, the techniques described above will aid researchers in recombinantly expressing and purifying proteins with poor expression, proline-rich regions, disordered regions, or nucleic acid binding properties.

In the Chapter 3, we discussed how numerous studies have produced lists of countless potential mRNA targets of FMRP, resulting in a plethora of data to follow-up for FXS treatments. We developed a rigorous authentication filter to separate the few targets that are worth pursuing in preclinical trials from the numerous targets identified *in vivo* or through pull-down assays. To test our filter, we narrowed our focus to two biologically relevant target RNAs (P21 γ and N19) with which FMRP has been shown to (1) interact with through RNA pull-down assays, (2) exert an influence over *in vivo*, and (3) contain potential G-quadruplexes. We then validated and quantified direct binding of the FXPs to these RNAs *in vitro* by determining binding affinities from fluorescence anisotropy and EMSA assays. All three FXPs bound to P21 γ and N19 with nanomolar affinities, but not the negative control RNA. The validated candidates were incorporated into our NanoFX mRNAs, and we found that whereas FMRP appears to regulate the mRNAs specifically (in a way that correlates with its binding affinity for the mRNAs), FXR1P/2P appear to be global translation regulators.

We also analyzed the functions of the KH domains and RGG motifs, revealing that the RGG motif appears to be predominantly responsible for FMRP's binding to G-

quadruplexes, yet other regions of the protein must contribute to the greatly increased binding affinity of the full-length protein, and its greater translation inhibition. This is supported by previous work in our lab which demonstrated that the C-terminal region of FMRP plays an essential role in the inhibition of translation, yet an N-terminally truncated FMRP led to greater inhibition [15]. This result indicates that the unstructured region between the KH domains and the RGG motif may cooperate with the RGG motif/C-terminus of FMRP to regulate translation [15]. Another possibility is that the RNA-binding capabilities of the RGG motif are augmented by the FXPs' ability to dimerize. As the ability of the FXPs to dimerize has been attributed to their N-terminal regions, this could explain why our RGG motif construct bound RNA and inhibited translation, but not to the same extent as full-length FMRP [16].

There are several benefits to our approach. First, we have developed a pipeline through which we can validate many targets simultaneously, rapidly (1-3 months), and economically. Second, adding *in vitro* analyses to our pipeline enables us to determine if FMRP binds directly to the targets in question as opposed to an interaction, that depends on other proteins or factors. Third, we can report quantitative dissociation constants (K_D), allowing for the ranking of relative preference amongst mRNA targets. Finally, because we have determined a method to purify the full-length versions of all the FXPs, we can easily and accurately test how FMRP's paralogs behave. As described previously, this is an interesting avenue of research due to FXR1/2P's critical roles in proper neural, muscle, and cardiac development. Unfortunately, many seemingly viable targets for FXS treatment will inevitably fall short in preclinical or clinical trials. As preclinical and clinical trials are costly and take a great deal of time

and effort, it is critical that we significantly narrow down the extensive list of existing targets. Such validations will refine our list of potential targets, providing clinical researchers with greater chances of success in identifying treatments for FXS.

Finally, while pursuing an ideal, minimalistic mRNA reporter, we discovered a novel RNA target of the FXPs. Through EMSAs and florescence anisotropy, we determined that the C-terminal RGG motif-containing region of FMRP is responsible for binding this RNA. This contradicted our hypothesis that the KH domains would be primarily responsible for binding this target since it does not form a GQ structure. To the best of our knowledge, this is the first example of a non-GQ containing RNA that is bound by the C-termini/RGG motif of FMRP. Interestingly, FMRP's paralogs bind this target with similar affinities, despite their divergent C-termini.

This novel target of the FXPs could open the door to a new class of FXS therapeutic targets. In the future, it would be valuable to pursue this novel *in vitro* RNA target as many other FMRP targets discovered *in vitro* have been found to reflect on biologically relevant mRNAs [6,17]. Likewise, *in vivo* reactions may miss some mRNAs bound by FMRP as the transcript abundance, cellular conditions, and experimental conditions can preclude the discovery of authentic interactions [18]. In the future, researchers may wish to search for similar RNA structures in biologically relevant mRNAs. It would also be worth investigating if this structure forms within ribosomal RNA, as studies have found that FMRP binds to the ribosome.

Many of our findings support previous work in our laboratory. For example, our work also found (1) that the KH domains were not critical for binding the RNAs we tested, (2) the RGG-motif containing C-termini was necessary and sufficient for binding

our RNAs, (3) the RGG motif/C-termini showed weaker binding to our RNAs than the full-length FMRP [19]. Expanding on prior work, we discovered that FXR1P and FXR2P bound the RNAs we tested similarly to FMRP. We were surprised by the similarity in binding as the binding to our RNAs appears to be predominantly mediated by the C-termini of the FXPs, which have lower sequence identity. It is possible that the R-rich regions of FXR1/2P enable them to bind to similar RNAs, despite the poor conservation of the RGG motif. In the future, it would be interesting to investigate the R-rich regions of FXR1P/2P in greater detail. Perhaps they have enabled FXR1P/2P more generality in binding of RNA, allowing them to regulate mRNA translation globally.

Furthermore, we are intrigued by the function of the highly conserved KH domains in the FXPs, as our studies did not reveal a function for them. Although recognized as RNA-binding domains that present in other proteins, little is known about the RNA targets of the KH domains. However, they do appear to be essential to polyribosome association- an important function of FMRP given that 80-95% of FMRP in the brain is associated with polyribosomes [20]. Additionally, the high conservation of the KH domains, between FMRP and its paralogs, and their role in FMRP's association with polyribosomes, suggests that they confer a valuable function [20-21]. To better understand the mechanism of the FXPs, it will be important to elucidate the function of the KH domains.

Overall, our work contributes to the study of the FXPs by (1) reducing barriers to studying the proteins *in vitro*, (2) providing a minimalistic in vitro translation system to elucidate the FXPs' effects on translation, (3) expanding knowledge of the differences between FMRP and its paralogs, and (4) discovering a novel RNA target

of the FXPs that may have biological significance. We hope that our work will spur further research into this family of proteins, which have critical roles in muscular development and neurodevelopment in mammals. Our field has the potential to develop treatments for FXS and related disorders in the near future, and we strive for this outcome through our continued and combined efforts.

References

[1] Darnell, J. C., Fraser, C. E., Mostovetsky, O. & Darnell, R. B. Discrimination of common and unique RNA-binding activities among Fragile X mental retardation protein paralogs. *Hum. Mol. Genet.* **18**, 3164–3177 (2009).

[2] E. Bechara, L. Davidovic, M. Melko, M. Bensaid, S. Tremblay, J. Grosgeorge, E.W. Khandjian, E. Lalli, B. Bardoni, Fragile X related protein 1 isoforms differentially modulate the affinity of fragile X mental retardation protein for G-quartet RNA structure, Nucleic Acids Res. 35 (2006) 299–306. https://doi.org/10.1093/nar/gkl1021.

[3] J. Garnon, C. Lachance, S. Di Marco, Z. Hel, D. Marion, M.C. Ruiz, M.M. Newkirk, E.W. Khandjian, D. Radzioch, Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level., J. Biol. Chem. 280 (2005) 5750–63. https://doi.org/10.1074/jbc.M401988200.

[4] Bardoni, B., Schenck, A. & Louis Mandel, J. A Novel RNA-binding Nuclear Protein That Interacts With the Fragile X Mental Retardation (FMR1) Protein. *Hum. Mol. Genet.* **8**, 2557–2566 (1999).

[5] 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, T. Tuschl, FMRP targets distinct mRNA sequence elements to regulate protein expression., Nature. 492 (2012) 382–6. https://doi.org/10.1038/nature11737.

[6] 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/10.1101/gad.1276805.

[7] C. Schaeffer, B. Bardoni, J.L. Mandel, B. Ehresmann, C. Ehresmann, H. Moine, The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif., EMBO J. 20 (2001) 4803–13. https://doi.org/10.1093/emboj/20.17.4803.

[8] L. Sjekloća, P.V. Konarev, J. Eccleston, I.A. Taylor, D.I. Svergun, A. Pastore, A study of the ultrastructure of Fragile-X-related proteins, Biochem. J. (2009). https://doi.org/10.1042/BJ20082197.

[9] Evans, T. L. & Mihailescu, M. I. Recombinant bacterial expression and purification of human fragile X mental retardation protein isoform 1. *Protein Expr. Purif.* (2010). doi:10.1016/j.pep.2010.06.002

[10] J. V. Sopova, E.I. Koshel, T.A. Belashova, S.P. Zadorsky, A. V. Sergeeva, V.A. Siniukova, A.A. Shenfeld, M.E. Velizhanina, K. V. Volkov, A.A. Nizhnikov, D. V. Kachkin, E.R. Gaginskaya, A.P. Galkin, RNA-binding protein FXR1 is presented in rat brain in amyloid form, Sci. Rep. 9 (2019) 18983. https://doi.org/10.1038/s41598-019-55528-6.

[11] Athar, Y. M. & Joseph, S. RNA-Binding Specificity of the Human Fragile X Mental Retardation Protein. *J. Mol. Biol.* (2020). doi:10.1016/J.JMB.2020.04.021.

[12] Buskirk, A. R. & Green, R. Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372**, (2017).

[13] Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32 (2004).

[14] Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).

[15] Y.M. Athar, S. Joseph, The Human Fragile X Mental Retardation Protein Inhibits the Elongation Step of Translation through Its RGG and C-Terminal Domains, Biochemistry. 59 (2020) 3813–3822. https://doi.org/10.1021/acs.biochem.0c00534.

[16] S. Adinolfi, A. Ramos, S.R. Martin, F. Dal Piaz, P. Pucci, B. Bardoni, J.L. Mandel, A. Pastore, The N-Terminus of the Fragile X Mental Retardation Protein Contains a Novel Domain Involved in Dimerization and RNA Binding †, Biochemistry. 42 (2003) 10437–10444. https://doi.org/10.1021/bi034909g.

[17] 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. https://doi.org/10.1016/s0092-8674(01)00566-9.

[18] Suhl, J. A., Chopra, P., Anderson, B. R., Bassell, G. J. & Warren, S. T. 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 (2014).

[19] Y.M. Athar, 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, 2020. https://escholarship.org/uc/item/3kh837f4 (accessed November 4, 2021).

[20] J.C. Darnell, S.J. Van Driesche, C. Zhang, K.Y.S. 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) 247–261. https://doi.org/10.1016/j.cell.2011.06.013.

[21] 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) 3087–3096. https://doi.org/10.1093/hmg/ddg335.