SCIENTIFIC REPORTS
nature research

OPEN

A simple procedure for bacterial expression and purification of the fragile X protein family

Madison Edwards, Mingzhi Xu & Simpson Joseph✉

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 and 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.

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 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[10]. It is estimated that 1/5000 males and 1/4000–8000 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 (Fig. 1). Additionally, all three proteins possess RNA-binding domains of interest: three well-conserved K homology (KH) domains,

Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0314, USA. ✉email: sjoseph@ucsd.edu

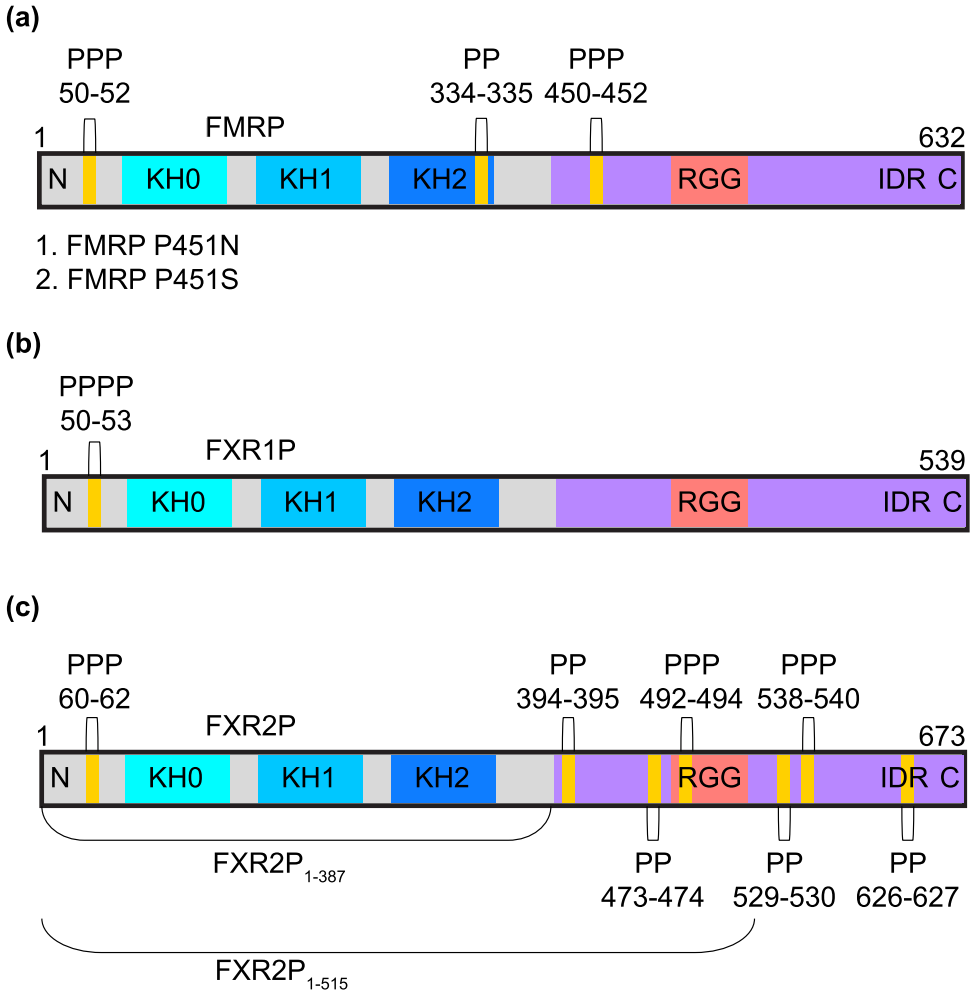

Figure 1. Domains and diprolyl/polyproline stretches of the fragile X protein family. (**A**) FMRP isoform 1, (**B**) FXR1P isoform 2, and (**C**) FXR2P with relevant protein domains and motifs labeled. Diprolyl/polyproline locations are displayed, and shortened FXR2P constructs and FXR2P/FMRP mutants 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].

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 (Supplementary Fig. 1). 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.

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

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 (Figs. 1C and 2C). As the ring structure of proline makes it a poor peptide bond donor and acceptor, two or more consecutive prolines can cause

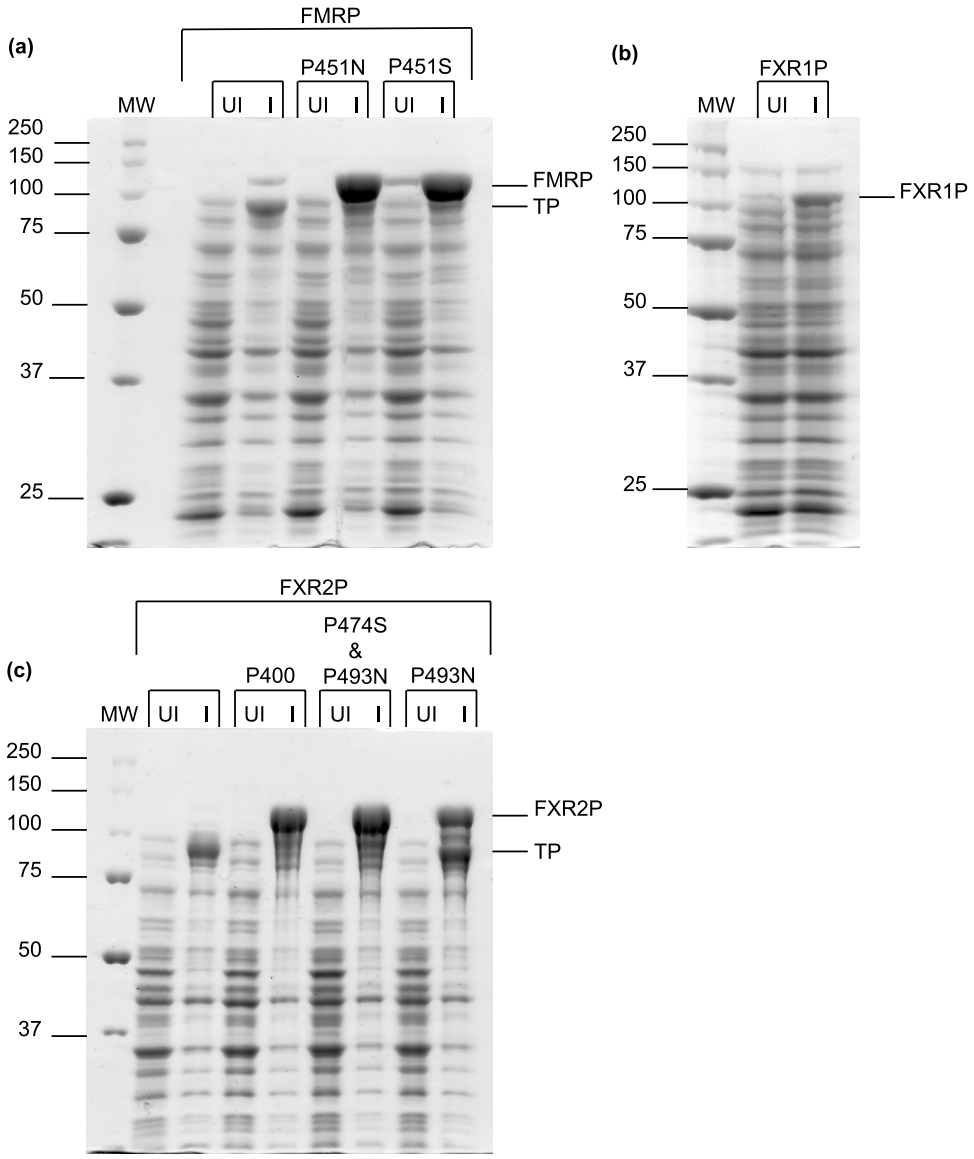

Figure 2. Mutations enhance the expression of FMRP and FXR2P. (**A**) His_6-MBP-FMRP (~115 kDa) and mutants, (**B**) His_6-MBP-FXR1P (~104 kDa), and (**C**) His_6-MBP-FXR2P (~117 kDa) and mutants. Lanes show the comparison between uninduced samples (UI) and samples induced with IPTG (**I**). FMRP P451N and FMRP P451S mutants exhibited a sevenfold 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 and 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).

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 (Supplementary Fig. 2). 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 (Fig. 2A,C).

In creating mutations, we aimed to preserve the original protein sequence as much as possible (Supplementary Fig. 3). 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 (Supplementary Fig. 4). However, we found

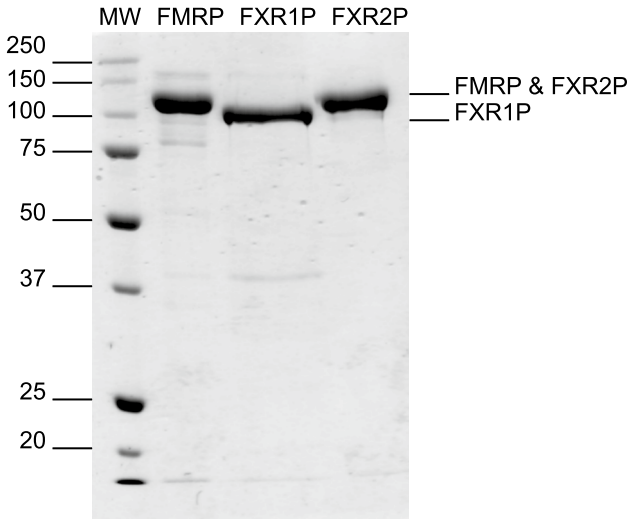

Figure 3. Purified fragile X proteins. Approximately 2 µg of each fusion protein was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). His_6-MBP-FMRP P451S (~ 115 kDa), His_6-MBP-FXR1P (~ 104 kDa), and His_6-MBP-FXR2P P474S and P493N (~ 117 kDa). The identity of each protein was confirmed by mass spectrometry.

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, Supplementary Fig. 4). This ultimately led us to select a FXR2P double mutant (FXR2P P474S and P493N) for future purification attempts (Fig. 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 (Fig. 2A). FXR1P did not need to be mutated as it contains only one polyproline motif, however we did observe a truncation for FXR1P (Figs. 1B and 2B). 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.

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 h 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 (Supplementary Figs. 5–7). 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 (Supplementary Fig. 8). 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 (Supplementary Fig. 6). Additionally, the FXR2P TPs began eluting prior to the full-length protein, while the full-length was predominantly eluted at high salt concentrations (Supplementary Fig. 7). 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 (Fig. 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.

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.

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 (Fig. 4A-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 (Fig. 4D). 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].

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 (Fig. 5A). 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].

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 (Fig. 5B). Similar to our anisotropy results for poly-$G_{17}U$, we observed the greatest translation inhibition by FXR2P, and the least by FXR1P (FXR2P $2.06\% \pm 0.54$ percent luciferase activity, FMRP $11.7\% \pm 2.1$, and FXR1P $20.2\% \pm 4.3$). As expected, the His-MBP tag did not inhibit the translation of *Renilla* luciferase mRNA ($103.8\% \pm 7.4\%$ activity). Our results suggest that our purified FXPs maintain their ability to regulate translation.

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

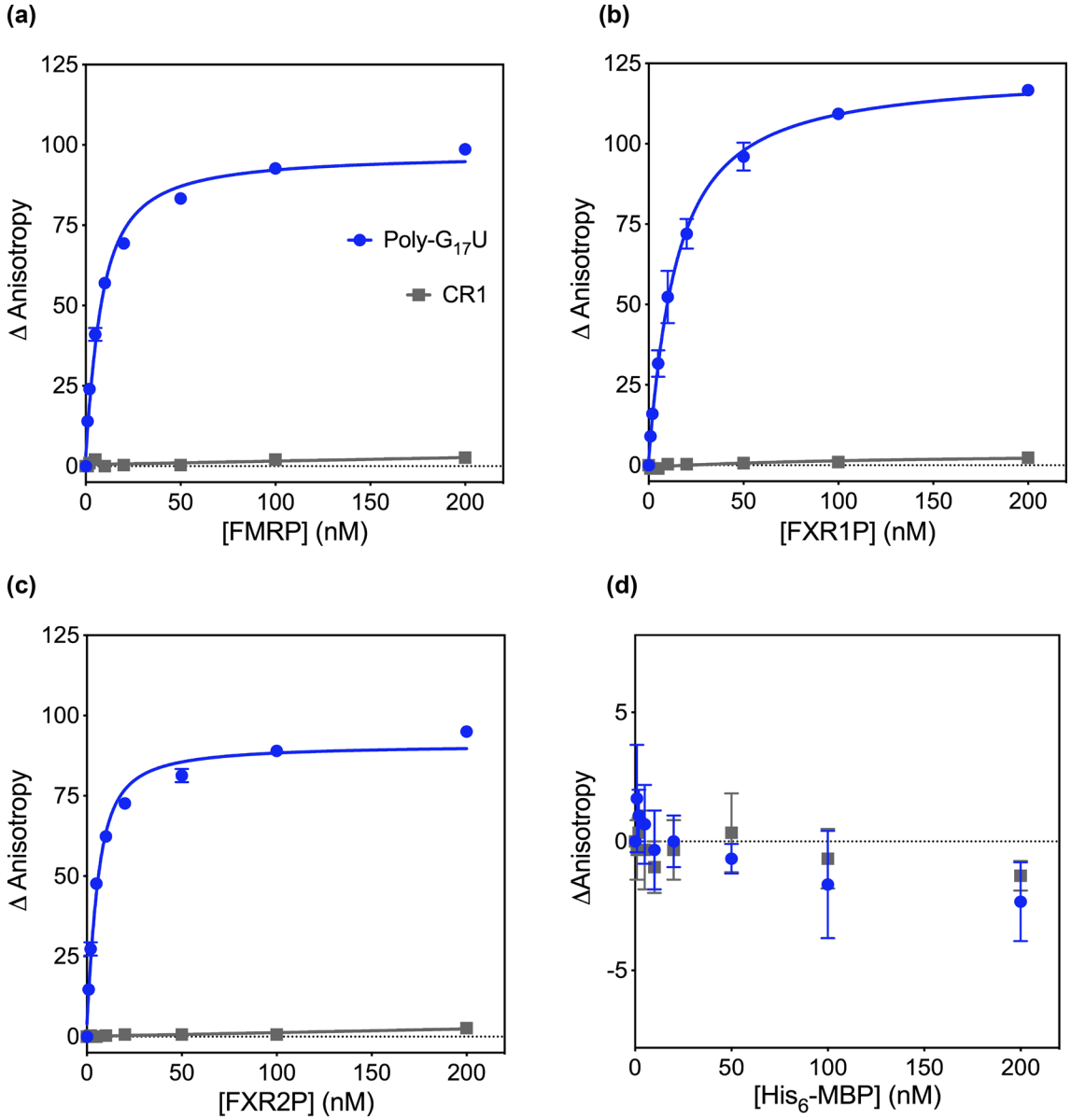

Figure 4. The fragile X proteins bind poly-$G_{17}U$. The RNA-binding capabilities of (**A**) FMRP, (**B**) FXR1P, and (**C**) FXR2P were assessed by fluorescence anisotropy. All three proteins bound to poly-$G_{17}U$ with high affinity: FMRP $K_D = 5.6 \pm 0.6$, FXR1P $K_D = 11.7 \pm 1$, and FXR2P $K_D = 3.1 \pm 0.4$. 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.

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

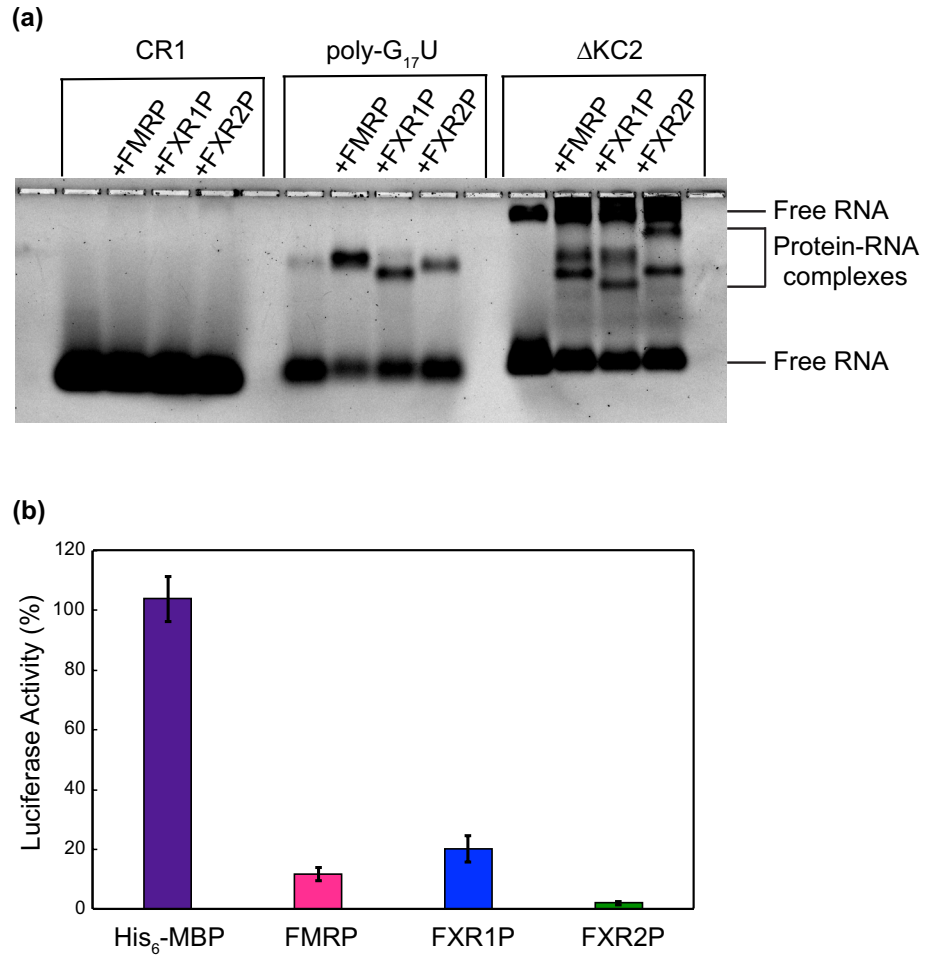

Figure 5. The fragile X proteins bind ΔKC2 and inhibit translation of renilla luciferase mRNA. (**A**) Binding to ΔKC2 was assessed using an agarose EMSA with CR1 and poly-$G_{17}U$ included as negative and positive controls, respectively. (**B**) The addition of fragile X proteins led to a reduction 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.

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.

Methods

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′.

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 Fig. 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 $FXR2P_{1\text{-}387}$. Forward: 5′-TAACATTGGAAGTGGATAACGGATCCG-3′.
Reverse: 5′-TTGACGCAGTTGCTCGTCAATC-3′.

Primers to make $FXR2P_{1\text{-}515}$. 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 and P492S and P494S). Forward: 5′ GACCGGTGGTCGTGGCCGTGGTAGCCCGAGCGCGCCGCGTCCG 3′
Reverse: 5′ GGACGACGACGGCTTTCTTCACCACGGGTGCTCGGATCACGGTCACCC 3′

FXR2P P500 (P529S and P538S and 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 and P494S. Forward:5′-CCGAGCGCGCCGCGTCCGACCAGC-3′
Reverse: 5′-GCTACCACGGCCACGACCACCG-3′

FXR2P P493N. Forward: 5′-AACCCGGCGCCGCGTCCGACCAGCC-3′
Reverse: 5′-CGGACCACGGCCACGACCACCGG-3′

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

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 #115,611; https://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′ CGTCGAGAAGACTACTAGATGCACCATCAT CATCATC 3′.
Reverse: 5′ CGTCGAGAAGACTTATTTCACGCGGCTCACATATTC 3′.

Primers to linearize pDSG310. Forward: 5′ CGTCGAGAAGACTGAAATAATAATACTAGAGCCAGG CATCAAATAAAAC 3′.
Reverse: 5′ CGTCGAGAAGACATCTAGTATTTCTCCTCTTTCTCTAGTAGCTAGC 3′.

Expression tests of fragile X proteins, mutants, and 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 h 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 h at 37 °C, ~ 215 RPM (2 h for FMRP and mutants Fig. 2). For the initial testing of FXR2P mutants only, (Supplementary Fig. 4) expression at 14 °C for ~ 18 h 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 min, 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 100 V, followed by 180 V until the dye front ran off the gel (~ 55 min). 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.

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 L 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 h. The cultures were then cooled for 20 min 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 h 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 g 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 s 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-80 K 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 h. 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 Figs. 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.

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-h Reverse phase C18 gradient).

G-quadruplex RNA-binding of the fragile × 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 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 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 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[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′.

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 min 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).

Primers to PCR amplify ΔKC2 RNA.. Forward: 5′-GCAACTGTTGGGAAGGGCGATCG-3′.
Reverse: 5′-AGACGCACATACCAGCCGCTAGC-3′.

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 min, then slow cooled from 68 °C to ~29 °C for ~1 h 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 h. 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 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.

RNA sequences. 18 nucleotides CR1: 5′-GCUAUCCAGAUUCUGAUU-Fluorescein-3′.
18 nucleotides poly-$G_{17}U$: 5′-GGGGGGGGGGGGGGGGGU-Fluorescein-3′.
72 nucleotides ΔKC2: 5′-GGGCGAAUUCGGGAUUCCGACCAGAAGGGGCUAAGGAAUGGUGGGACG AGCUAGCGGCUGGUAUGUGCGUCU-Fluorescein-3′.

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 min 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 h. 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.

Received: 27 May 2020; Accepted: 7 September 2020
Published online: 28 September 2020

References

1. Siomi, M. C. *et al.* FXR1, an autosomal homolog of the fragile X mental retardation gene. *EMBO J.* **14**(11), 2401–2408 (1995).
2. Zhang, Y. *et al.* The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J.* **14**, 5358 (1995).
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**, 3825–3832 (1996).
4. Bakker, C. E. *et al.* Immunocytochemical and biochemical characterization of FMRP, FXR1P, and FXR2P in the Mouse. *Exp. Cell Res.* **258**, 162–170 (2000).
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. Hagerman, R. J. *et al.* Fragile X syndrome. *Nat. Rev. Dis. Prim.* **3**, 17065 (2017).
8. Pieretti, M. *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817–822 (1991).
9. Sutcliffe, J. S. *et al.* DNA methylation represses FMR-1 transcription in fragile X syndromeDNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum. Mol. Genet.* **1**, 397–400 (1992).
10. Irwin, S. A. *et al.* Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: A quantitative examination. *Am. J. Med. Genet.* **98**, 161–167 (2001).
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. Guo, W. *et al.* Fragile X proteins FMRP and FXR2P control synaptic GluA1 expression and neuronal maturation via distinct mechanisms. *Cell Rep.* **11**, 1651–1666 (2015).
13. Cavallaro, S. *et al.* Genes and pathways differentially expressed in the brains of Fxr2 knockout mice. *Neurobiol. Dis.* **32**, 510–520 (2008).
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. Tamanini, F. *et al.* The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins. *Hum. Mol. Genet.* https://doi.org/10.1093/hmg/9.10.1487 (2000).
17. Davidovic, L. *et al.* Alteration of expression of muscle specific isoforms of the fragile X related protein 1 (FXR1P) in facioscapulohumeral muscular dystrophy patients. *J. Med. Genet.* **45**, 679–685 (2008).
18. Khandjian, E. *et al.* Novel isoforms of the fragile X related protein FXR1P are expressed during myogenesisNovel isoforms of the fragile X related protein FXR1P are expressed during myogenesis. *Hum. Mol. Genet.* **7**, 2121–2128 (1998).
19. Coy, J. F. *et al.* Highly conserved 3′ UTR and expression pattern of FXR1 points to a divergent gene regulation of FXR1 and FMR1. *Hum. Mol. Genet.* **4**, 2209–2218 (1995).
20. Mientjes, E. J. *et al.* Fxr1 knockout mice show a striated muscle phenotype: Implications for Fxr1p function in vivo. *Hum. Mol. Genet.* **13**, 1291–1302 (2004).
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**, 1014 (2002).
24. Järvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis)order in RNA regulation. (2016). https://doi.org/10.1186/s12964-016-0132-3
25. Calabretta, S. & Richard, S. Emerging roles of disordered sequences in RNA-binding proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015).
26. Bechara, E. *et al.* 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**, 299–306 (2006).
27. Adinolfi, S. *et al.* The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding†. *Biochemistry* **42**, 10437–10444 (2003).
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. USA* **98**, 8844–8849 (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. Ascano, M. *et al.* FMRP targets distinct mRNA sequence elements to regulate protein expression. *Nature* **492**, 382–386 (2012).
33. Maurin, T. *et al.* 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**, 6344–6355 (2018).
34. Whitman, S. A. *et al.* Desmoplakin and talin2 are novel mRNA targets of fragile X-related protein-1 in cardiac muscle. *Circ. Res.* https://doi.org/10.1161/CIRCRESAHA.111.244244 (2011).
35. Davidovic, L. *et al.* 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**, e1003367 (2013).
36. Garnon, J. *et al.* Fragile X-related protein FXR1P regulates proinflammatory cytokine tumor necrosis factor expression at the post-transcriptional level. *J. Biol. Chem.* **280**, 5750–5763 (2005).
37. Fernández, E. *et al.* Cellular/molecular FXR2P exerts a positive translational control and is required for the activity-dependent increase of PSD95 expression. https://doi.org/10.1523/JNEUROSCI.4800-14.2015
38. Xu, X.-L. *et al.* FXR1P but not FMRP regulates the levels of mammalian brain-specific microRNA-9 and microRNA-124. *J. Neurosci.* **31**, 13705–13709 (2011).
39. Darnell, J. C. *et al.* Kissing complex RNAs mediate interaction between the fragile X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev.* **19**, 903–918 (2005).
40. Darnell, J. C. *et al.* Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**, 489–499 (2001).
41. Schaeffer, C. *et al.* The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20**, 4803–4813 (2001).
42. Sjekloća, L. *et al.* A study of the ultrastructure of fragile-X-related proteins. *Biochem. J.* https://doi.org/10.1042/BJ20082197 (2009).
43. Evans, T. L. & Mihailescu, M. I. Recombinant bacterial expression and purification of human fragile X mental retardation protein isoform 1. *Protein Expr. Purif.* https://doi.org/10.1016/j.pep.2010.06.002 (2010).
44. Sjekloća, L., Pauwels, K. & Pastore, A. On the aggregation properties of FMRP—A link with the FXTAS syndrome?. *FEBS J.* https://doi.org/10.1111/j.1742-4658.2011.08108.x (2011).
45. Sopova, J. V. *et al.* RNA-binding protein FXR1 is presented in rat brain in amyloid form. *Sci. Rep.* **9**, 18983 (2019).
46. Athar, Y. M. & Joseph, S. RNA-binding specificity of the human fragile X mental retardation protein. *J. Mol. Biol.* https://doi.org/10.1016/J.JMB.2020.04.021 (2020).
47. Zhang, J. *et al.* Expression and characterization of human fragile. *Proteomics Insights* **10**, 117864181882526 (2019).
48. Tsang, B. *et al.* Phosphoregulated FMRP phase separation models activity-dependent translation through bidirectional control of mRNA granule formation. *Proc. Natl. Acad. Sci. USA* **116**, 4218–4227 (2019).
49. Ceman, S. *et al.* Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum. Mol. Genet.* **12**, 3295–3305 (2003).
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**, 1857–1861 (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. Peil, L. *et al.* Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci.* **110**, 15265–15270 (2013).
54. Starosta, A. L. *et al.* Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic Acids Res.* https://doi.org/10.1093/nar/gku768 (2014).
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* https://doi.org/10.1126/science.1233338 (2013).
58. Vasilyev, N. *et al.* Crystal structure reveals specific recognition of a G-quadruplex RNA by a β-turn in the RGG motif of FMRP. *Proc. Natl. Acad. Sci.* **112**, E5391–E5400 (2015).
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**(1716), 20160183 (2017).
61. Donnelly, M. I. *et al.* An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. *Protein Expr. Purif.* **47**, 446 (2006).
62. Glass, D. S. & Riedel-Kruse, I. H. A synthetic bacterial cell–cell adhesion toolbox for programming multicellular morphologies and patterns. *Cell* https://doi.org/10.1016/j.cell.2018.06.041 (2018).
63. Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

Acknowledgements

We thank Reta Sarsam for cloning FXR1P into pMCSG9 and providing purified His_6-MBP for our assays. We thank Youssi Athar for cloning FMRP into pMCSG9, creating the FMRP mutant plasmids, providing *Renilla* luciferase mRNA, and sharing the mass spectrometry results of the truncated FMRP species. We thank Adam Maloney for providing 2X rabbit reticulocyte lysate. We thank Eileen Chen for cloning ΔKC2 into the pGEM3Z plasmid. This work was supported by the National Institutes of Health (R01GM114261 to S.J.) and the Cell and Molecular Genetics Training Program funded by the National Institutes of Health (T32GM007240).

Author contributions

M.E. and S.J. designed the experiments with input from M.X. M.E. and M.X. performed the experiments and all authors discussed the results. M.E. and M.X wrote the paper with input from S.J. S.J. supervised all aspects of the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-72984-7.

Correspondence and requests for materials should be addressed to S.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.