Published in final edited form as:

Neuroscience. 2011 May 5; 181C: 48–66. doi:10.1016/j.neuroscience.2011.02.029.

A non-canonical start codon in the *Drosophila fragile X* gene yields two functional isoforms

Rebecca W. Beerman and Thomas A. Jongens

Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

Abstract

Fragile X syndrome is caused by the loss of expression of the fragile X mental retardation protein (FMRP). As a RNA binding protein, FMRP functions in translational regulation, localization, and stability of its neuronal target transcripts. The *Drosophila* homologue, dFMR1, is well conserved in sequence and function with respect to human FMRP. Although dFMR1 is known to express two main isoforms, the mechanism behind production of the second, more slowly migrating isoform has remained elusive. Furthermore, it remains unknown whether the two isoforms may also contribute differentially to dFMR1 function. We have found that this second dFMR1 isoform is generated through an alternative translational start site in the *dfmr1* 5'UTR. This 5'UTR coding sequence is well conserved in the *melanogaster* group. Translation of the predominant, smaller form of dFMR1 (dFMR1-S^N) begins at a canonical start codon (ATG), whereas translation of the minor, larger form (dFMR1-L^N) begins upstream at a non-canonical start codon (CTG). To assess the contribution of the N-terminal extension toward dFMR1 activity, we generated transgenic flies that exclusively express either dFMR1-S^N or dFMR1-L^N. Expression analyses throughout development revealed that dFMR1-S^N is required for normal dFMR1-L^N expression levels in adult brains. *In situ* expression analyses showed that either dFMR1-S^N or dFMR1-L^N is individually sufficient for proper dFMR1 localization in the nervous system. Functional studies demonstrated that both dFMR1-S^N and dFMR1-L^N can function independently to rescue *dfmr1* null defects in synaptogenesis and axon guidance. Thus, *dfmr1* encodes two functional isoforms with respect to expression and activity throughout neuronal development.

Keywords

dFMR1; *Drosophila*; fragile X mental retardation; non-canonical start codon; attB/attP transgenic system

1.1

The most common inherited form of intellectual disability, fragile X syndrome (FXS), is caused by the loss of the fragile X mental retardation protein, FMRP (Verkerk et al., 1991, Ashley et al., 1993a). In addition to below average IQ, fragile X patients also exhibit an

© 2011 IBRO. Published by Elsevier Ltd. All rights reserved.

Address correspondence to: Tom Jongens, PhD, 415 Curie Blvd. / 510 Clinical Research Building, Philadelphia, PA 19104, Phone: (215) 573-9322, jongens@mail.med.upenn.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

increased frequency of autism, sleep disorders, and attention deficit hyperactivity disorders among other neurological abnormalities (D'Hulst and Kooy, 2009). The clinical features associated with FXS underscore the importance of FMRP for the proper function and development of many neurological processes.

Characterization of FMRP in several different experimental model systems has defined FMRP as an important neuronal component for post-transcriptional regulation of specific transcripts. FMRP has three RNA binding domains that have all been shown to bind RNA: two K homology domains (KH) (homologous to the RNA binding domain in hnRNP K), and one RGG box (enriched in arginine and glycine residues) (Ashley et al., 1993a, Brown et al., 1998). FMRP was originally shown to function in translational regulation (Ashley et al., 1993a, Zhang et al., 1995, Khandjian et al., 1996, Siomi et al., 1996, Corbin et al., 1997, Lagerbauer et al., 2001, Li et al., 2001, Khandjian et al., 2004) and more recently FMRP has also been shown to function in transcript localization and transcript stability (Miyashiro et al., 2003, Rackham and Brown, 2004, Zalfa et al., 2007, Dichtenberg et al., 2008, Estes et al., 2008, Wang et al., 2008).

The use of animal model systems has provided great insight into how the absence of this one protein leads to the wide range of symptoms seen in fragile X patients. With respect to sequence, FMRP is highly conserved in mice (97% identical sequence) and *Drosophila* (up to 85% similarity in the RNA binding domains) (Ashley et al., 1993b, Wan et al., 2000). With respect to conservation of function, the phenotypes associated with the loss of *Fmr1* in the mouse and fly recapitulate many aspects of FXS in human patients, allowing for molecular dissection of the relevant pathways (D'Hulst and Kooy, 2009, Gatto and Broadie, 2009a). For example, abnormal axonal branching in the *Drosophila* FXS models (analogous to abnormal dendritic elaborations observed in post-mortem analysis of fragile X patients) has been directly linked to mis-regulation of FMRP target transcripts, such as the *microtubule associated protein 1b* (*map1b*) or *Futsch* (in flies) and the actin-binding protein, *chickadee* (*Drosophila* homologue of *profilin*) (Hinton et al., 1991, Zhang et al., 2001, Reeve et al., 2005).

Although FXS (or FXS-like phenotypes in animal model systems) is caused by the repression of one gene, alternative splicing of the *FMR1* transcript actually produces multiple protein isoforms (Ashley et al., 1993b). Consequently, in order to understand how FMRP expression prevents FXS, it is also important to study how the different FMRP isoforms may contribute to FMRP function individually or collaboratively. To this point, previous studies have already established that these isoforms are likely to be functionally distinct. For example, expression of full length FMRP is cytoplasmic, whereas an alternatively spliced product, isoform 4 (iso4), is localized to the nucleus because exon 14 (which encodes a nuclear export sequence) is excluded (Eberhart et al., 1996). Given that several of the alternatively spliced isoforms of *FMR1* also alter the protein coding sequence, and are differentially expressed in different tissues it is possible that these sequence changes could also affect FMRP activity (Xie et al., 2009). However, the biological requirement for each of the alternatively spliced *FMR1* isoforms has yet to be formally tested *in vivo*. Alternative splicing also occurs in *dfmr1* and is predicted to result in several protein isoforms, some of which were shown to be important for several specific neuronal functions of dFMR1 *in vivo* (Schenck et al., 2002, Banerjee et al., 2010).

Alternative start codons provide an additional (but less common) mechanism for generating multiple protein isoforms. Based on the reading frame of upstream ATGs, 3% of human mRNA could encode for amino (N)-terminal extensions as a product of upstream translation initiation sites (Kochetov et al., 2005). In rare cases, the alternative initiation codons are non-ATG start codons (Kozak, 1997, Touriol et al., 2003, Chang and Wang, 2004). The

well-studied human fibroblast growth factor 2, *FGF2*, contains three upstream non-canonical CTG start codons. Selection of the alternative start codons over the canonical start codon is regulated under various environmental conditions (such as heat shock) and the resulting N-terminal extensions confer different functional attributes to the FGF2 isoforms, such as nuclear localization or protein-protein binding affinities (Touriol et al., 2003, Yu et al., 2007). However, few investigations have probed the extent to which isoforms derived from alternative non-canonical start codons independently contribute to protein expression and function *in vivo* in transgenic animals (Miles et al., 2003). In our current study we show that the second most highly expressed isoform of dFMR1 is produced through the use of an alternative non-canonical start codon and we aimed to study how the presence or absence of the N-terminal extension affects the expression and/or activities of dFMR1.

2.1 Experimental procedures

2.1.1 DNA constructs

To generate the *dfmr1* GFP reporter constructs, the pUASpEGFPc1 (Drosophila Genomics Resource Center, DGRC #1240) was engineered to include extra restriction sites into which various segments of the *dfmr1* transcript could be inserted. First, linker DNA oligos were generated by Integrated DNA technology (IDT) and ligated into pUASpEGFPc1. Specifically, an AscI site was added to the 5' end of pUASP promoter (between the pre-existing NarI and PciI restriction sites) and a PacI site was added to the 3' end of EGFP (between the pre-existing PspXI and BbvCI sites, which also eliminates the K10 3'UTR), to form pUASpEGFPc1.A.P. The region spanning from the pre-existing StuI to the 5' end of EGFP were PCR amplified to add a NotI and MluI site upstream of the start codon in EGFP. The region spanning from the 5' to the 3' end of the EGFP coding sequence (just upstream of the stop codon) were PCR amplified to introduce the 5' MluI site and 3' BamHI, NheI and PacI sites (the NheI site contains a stop codon in frame with the EGFP coding sequence). The first PCR amplified region was digested with StuI and MluI and the second region was digested with MluI and PacI and both were ligated into pUASpEGFPc1.A.P (digested with StuI and PacI) to form N.M.pUASpEGFPc1.B.N.A.P. The modified N.M.pUASpEGFPc1.B.N.A.P construct was then cloned into attB-P[acman]-CmR (DGRC #1244) using the AscI and PacI sites to form pUASP-EGFP-attB. The different transcript elements (5'UTR, coding region, and/or 3'UTR) could then be subcloned into pUASP-EGFP-attB. The *dfmr1* RA 5'UTR (568bp) was PCR amplified from the *dfmr1* genomic rescue construct to include NotI and MluI sites on its 5' and 3' ends respectively. The 5'UTR+TAG-GFP construct was made by PCR amplification of the *dfmr1* RA 5'UTR with the same NotI forward primer as described above, and an MluI containing reverse primer that included a TAG stop codon just upstream of the MluI site. The 5'UTR CAG-GFP construct was generated using point mutagenesis to change the CTG→CAG using the Stratagene quikchange II XL site directed mutagenesis kit before the *dfmr1* 5'UTR was cloned into pUASP-EGFP-attB. The *dfmr1* coding sequence (~2kb) was PCR amplified from AF305881 cDNA or the *Rin* coding sequence (~2kb) was PCR amplified to include BamHI and NheI sites onto their 5' and 3' ends respectively. The α 1 tubulin 3'UTR (276bp, CG1913) was PCR amplified from ovary cDNA or the *dfmr1* RA 3'UTR (~1.8kb) was PCR amplified from a *dfmr1* genomic construct (previously published) to include NheI and PacI sites onto their 5' and 3' ends respectively (Dockendorff et al., 2002). All PCR amplifications were carried out using Stratagene Pfu Turbo polymerase and amplified regions were sequenced in a carrier vector (Invitrogen DNA pCR®-Blunt) or in an intermediate stage vector to verify that no mutations were introduced during the PCR amplification steps.

The *dfmr1* WTR point mutant constructs were generated by digesting the *dfmr1* WTR genomic constructs from pBluescript (pBS) using BamHI and PspXI restriction sites to

clone into pUASP-EGFP-attB (digestion of pUASP-EGFP-attB with BamH1 and PspX1 will eliminate all of the pUASP-EGFP elements). In order to insert point mutations into the WTR construct, a 2.8kb 5' genomic fragment was PCR amplified to add on an EcoR1 site just upstream of an endogenous Hpa1 site and to add on a BamH1 site just downstream of an endogenous RsrII site to ligate into Invitrogen DNA pCR®-Blunt. Point mutagenesis (as schematized in Fig. 3) was carried out directly on the DNA fragments in DNA pCR®-Blunt or using an EcoR1-BamH1 fragment cloned into PALTER-1 (Promega) using the Stratagene quikchange II XL site directed mutagenesis kit according to the manufacturer's instructions. After sequencing the 2.8kb 5' genomic region to ensure correct point mutagenesis, the fragment was cloned back into the WTR genomic construct in pBluescript using Hpa1 and RsrII sites, followed by cloning into pUASP-EGFP-attB (to use for injections).

The *dfmr1* C-terminally TAP tagged WTR point mutant constructs were made as follows: First BspE1 restriction sites were PCR amplified onto both sides of the TAP tag (containing protein A and the calmodulin binding peptide, (Rigaut et al., 1999)) and was inserted into exon 12 of the genomic *dfmr1* WTR construct in pBS such that base pair 310–320 of exon 12 will be translated as SGSG (rather than SGQQ), followed by the TAP sequence and results in the deletion of the last 20 amino acids from the expressed dFMR1 protein, referred to as WTR-cTAP in pBS vector (A. Pepper, unpublished results). The WTR-L^N mutation is the same as that in WTR-L^N-cTAP and was generated as described above. WTR-S^N-cTAP is different from WTR-S^N in that WTR-S^N-cTAP carries a stop codon just 9 codons upstream of the normal ATG start codon (changing an existing AAA→TAA), made via the same method as described above and produces only dFMR1-S^N at levels comparable to those produced from WTR-S^N (data not shown). The GFP-DCP1 constructs were a gift from Elisa Izaurralde (Eulalio et al., 2007b).

2.1.2 Bioinformatics predictions and sequence alignment

Alignment of *dfmr1* 5'UTR nucleotide sequences from the twelve *Drosophila* species was carried out using phastCons alignment from the UCSC genome browser (April 2006 Assembly) and Sequencher v.4.7. Translation and alignment of the putative N-terminal amino acid sequences was carried out using Clustal W from MacVector version v.9.5.2. Residues within the dFMR1 N-terminal extension predicted to be competent for post-translational modifications were identified by scores ranked above threshold values by Pfam and additional proteomics tools from Expert Protein Analysis System (ExPASy) from Swiss Institute of Bioinformatics (Gasteiger et al., 2003, Finn et al., 2010).

2.1.3 Fly strains

Wild-type (WT) is *w¹¹¹⁸* or isogenized *w¹¹¹⁸* 31 (*iso31*) (gift of Amita Sehgal, stock #5905 from Bloomington Stock Center (Ryder et al., 2004), and *dfmr1* null flies were obtained from *w¹¹¹⁸;dfmr1³/TM6C,Tb,Sb* stock (Dockendorff et al., 2002). The following *attP40* transgenic strains were made by Genetic Services Inc.: *WTR*, *WTR-L^N*, and *WTR-L^N.ATG*. We generated *WTR-S^N* flies by injecting the *WTR-S^N* construct into flies carrying one copy of integrase driven by the *vasa* promoter (*yw vas-phi* (2A)) on the X chromosome and one copy of *attP40* on the second chromosome (Bischof et al., 2007, Markstein et al., 2008). Correct integration at the *attP40* locus was tested for each *attP40* transgenic line by genomic PCR using the same primers sets as described in ((Venken et al., 2006) data not shown and). The transgenic *attP40* lines were outcrossed to *w¹¹¹⁸* lines for two generations before crossing them into *w¹¹¹⁸* balancer stocks and then into the *dfmr1³/TM6C,Tb,Sb* stock. Isogenized *dfmr1³* stocks and isogenized *attP WTR* stocks were generated by outcrossing into the *iso31* strain for at least 8 generations before crossing them together, to increase the probability that any phenotypic difference observed was due to the transgene itself and not background elements within the transgenic strains.

2.1.4 S2 cell culture and transfection

Drosophila Schneider S2 cells were grown in serum free media (Invitrogen Sf-900 II SFM supplemented with GIBCO L-glutamine (45ml of 200mM per 500mL media) and GIBCO Penicillin-Streptomycin antibiotics (2.5ml of 10000U/ml per 500mL media). Cells were plated at 2×10^6 cells/ml in 2ml media in 6 well plates. DNA constructs were transfected into S2 cells using Invitrogen's Cellfectin using 1.5 μ g pMT-GAL4 (GAL4 driven by metallothionein promoter, DGRC #1042) and 1.7 μ g of various pUASP-EGFP-attB constructs according to the manufacturers' instructions. The following day, the media was replaced with 2ml media supplemented with CuSO₄ (500 μ M final concentration). On the third day, the cells were pelleted (spin 5min at $1000 \times g$), and lysed with 400 μ l ice-cold RIPA buffer (150mM NaCl, 1% Nonidet P-40, 0.5% Deoxycholate, 0.1% SDS, 50mM Tris-HCl (pH 8.0), 10mM NaF, 0.4mM EDTA (pH8.0), 10% Glycerol) supplemented with 1mM sodium orthovanadate and 1 $\times$ protease inhibitors (Sigma P2714) just before use. Lysed cells were spun 20min at $16000 \times g$ at 4°C and supernatant (protein lysate) was transferred to new tube. S2 cell treatment were the same for Fig. 9 except that no pMT-GAL4 was required. S2 cell treatments for Fig. 6 were the same as described above for Fig. 2, except our lab had subsequently switched to using S2 cells using Invitrogen's Schneider's *Drosophila* Media supplemented with 10% heat inactivated Fetal Bovine Serum (FBS) (Invitrogen cat#26140079), penicillin/streptomycin antibiotics (5ml/500ml media GIBCO cat# 15140), L-glutamine (GIBCO GlutaMAX cat# 35050 5mL/500mL media), and HEPES (Invitrogen cat#15630080 5ml of 1M HEPES/500mL media). Additionally DNA constructs were transfected into S2 cells using Qiagen's Effectene using 1 μ g of pMT-GAL4 and 1 μ g of various pUASP-EGFP-attB constructs according to the manufacturers' instructions.

2.1.5 Western analyses

All tissues were dissected in 1X PBS. Protein extraction and western procedure methods have been previously described in (Pepper et al., 2009), except we also used NuPAGE® Novex® 7% Tris-Acetate precast gels and NuPAGE® and Tris-Acetate SDS Running Buffer (20X) (Invitrogen) for better separation of dFMR1 isoforms. Each lane was loaded with approximately one ovary (from two ovary pairs), 2 larval brains, 2 pharate adult brains, 2 (15–20 hour) adult heads, 10 μ l of S2 cell lysate, or 20 μ g protein lysate from whole embryos or whole pupae. The following primary antibodies were used: mouse anti-dFMR1 (6A15, 1:5000, (Wan et al., 2000)), rat anti-catenin, alpha (DCAT1, 1:4000, Developmental Studies Hybridoma Bank (DSHB)), mouse anti-beta tubulin (E7, 1:10,000, DSHB), chicken anti-GFP (GFP-1020, 1:8000, Aves Labs), rabbit anti-S6kinase, phospho Thr 398 (#9209, 1:200, Cell Signaling). The following secondary antibodies were used: HRP-conjugated goat anti-mouse (1:2000, Jackson ImmunoResearch), HRP-conjugated goat anti-rat (1:2000, Jackson ImmunoResearch), HRP-conjugated goat anti-chicken (1:3000, Jackson ImmunoResearch), and HRP-conjugated goat anti-rabbit (1:2000, Jackson ImmunoResearch). Either Thermo Scientific SuperSignal West Pico Chemiluminescent Substrate or Roche Lumi-Light Western Blotting Substrate was used to detect the HRP signal from the secondary antibody. Quantification was carried out using Kodak ID image analysis software (v.4.5.0). The net intensity of the dFMR1 band was normalized to the net intensity of the corresponding loading control signal before calculating the ratio to *WTR*. Each set of proteins: dFMR1-L^N, dFMR1-S^N, or loading control intensities were often measured from different exposures of the same blot to ensure that the signal was within the linear range for quantification. Average ratios to *WTR* were pooled together from at least three biological replicates and graphed as an average $\pm$ standard error of the mean.

The western analysis in Figure 1 was carried out exactly the same as described above, except the LI-COR Odyssey Infrared Imaging System was used to quantify the ratio between signal intensities of the dFMR1 isoforms. Consequently, the Odyssey Blocking

Buffer was used to block the membrane and dilute the antibodies, 1XPBS +0.2% Tween 20 was used to wash the blots, and IRdye® 800-conjugated goat anti-mouse (1:5000 Rockland Immunochemicals) was used as the secondary antibody. The signal intensities were quantified using the Odyssey imaging system software (v.2.0) according to the manufacturer's instructions.

2.1.6 Immunofluorescence

Pharate adult brains, larval brains, ovaries, and 3–5 day old adult brains were dissected in 1X PBS, and fixed and treated as previously described in (Pepper et al., 2009). Dissection and staining of third instar larvae was carried out essentially as described in (Pepper et al., 2009), except that fixation was carried out for 30 minutes, and all samples were mounted in Vectashield mounting medium (Vector laboratories). S2 cell immunofluorescence was carried out as follows: Following transfection, cells were transferred to 4-well chamber slides (Lab-Tek) pre-treated with 1mg/ml poly-L-lysine (15min at RT and rinsed briefly with water) and allowed to adhere to the slide for 1hr 29°C. Cells were fixed with 2% paraformaldehyde in 1X PBS or 10min at RT. Cells were rinsed 3 × 5 min with 1X PBS, permeabilized with 0.1% Triton-X100 in 1X PBS for 2 min, rinsed 3 times with 1X PBS and blocked with Blocker (500 ml of Blocker: 0.5g BSA, 5g PVP-10, 5g PVP-40, 0.5g PVP-360, filled to 500ml with 1X PBS) for 10 min at RT. Primary antibody was incubated with cells overnight at 4°C. Then cells were washed 2 × 5 min in Blocker, followed by 5 min wash in 1X PBS. Secondary antibody was incubated with cells for 30 min RT. Then cells were washed 2 × 10 min with Blocker followed by 10 min wash with 1X PBS, briefly rinsed with water, chambers were removed and KPL mounting media was added. The following primary antibodies were used: mouse anti- fasciclin II (ID4, 1:5, DSHB), mouse anti-Orb (6H4, 1:30, DSHB), Cy5-conjugated goat anti-HRP (1:200, Jackson ImmunoResearch), mouse anti-discs large (4F3, 1:1000, DSHB), Texas Red-conjugated goat anti-HRP (1:50, Jackson ImmunoResearch), rat anti-DE cadherin (DCAD2, 1:20, DSHB), chicken anti-protein A (1:10,000 Immunology Consultants Laboratory), and mouse anti-dFMR1 (6A15, 1:1000, (Wan et al., 2000)). The following secondary antibodies were used: FITC-conjugated goat anti-mouse IgG (1:250, Jackson ImmunoResearch), FITC-conjugated goat anti-mouse IgG1 (1070-02, 1:350, SouthernBiotech), Texas Red-conjugated goat anti-mouse IgG_{2a} (1080-07, 1:350, SouthernBiotech), Cy5-conjugated goat anti-rat (1:200, Jackson ImmunoResearch), and Texas Red-conjugated goat anti-chicken (1:500, Rockland Immunochemicals). All images were captured using Leica TCS SP confocal microscope using software version 2.6.1 except Fig. 8B, for which the Leica SP5 confocal microscope using software version LAS AF 2.02 was used. Representative images in Fig. 10 are maximum projections from a stack of sections taken at 1µm sections, and images from Figs. 4B, 4C, 8, and 9 are an average of four scans from a single section. Confocal images in Figs. 4B, and 8 were each taken with the same settings, and images in Figs. 9 and 10 were each taken with different settings to optimize each image.

2.1.7 Phenotypic scoring of NMJ and axon extension defects

NMJ bouton numbers were counted blindly (genotypes were obscured from the counter) using the "cell counter" tool in ImageJ software (NIH) to facilitate scoring. "n" refers to the number of hemisegments counted per genotype. Type 1b boutons were differentiated from type 1s boutons by strength of anti-DLG staining surrounding the anti-HRP (stronger in type 1b) and size of boutons along the branches (generally larger for type 1b). Additionally, given the increase in complexity for the NMJ at muscle 6/7 relative to muscle 4, analysis of muscle 6/7 was carried out using the isogenized strains. Statistical analyses were carried out using InStat v.3.1a. Scoring β-lobe fusion in the mushroom bodies was scored as abnormal if the two lobes touched at the midline and normal if the two β-lobe termini could be distinguished as two separate ends.

2.1.8 Heat shock treatment

Incubated 1–2 day old wild-type (*iso31*) flies (10–15 per vial) in water bath at 39°C (heat shock) or 25°C (control) for 1 hr. Flies were frozen on dry ice at 0', 5', 10', 15', 30', and 60' post 1 hr treatment before the heads were removed on dry ice and processed for western analyses as described above.

2.1.9 5'RACE

1–2 ml of wild-type (*w¹¹¹⁸*) flies were collected, frozen on dry ice, and then immersed in liquid nitrogen, after which mass collection of heads was obtained through shaking frozen flies through sieves, and grinding the (100μl) heads in liquid nitrogen with a mortar and pestle. Total RNA was then extracted from the head powder using TRIzol (Ambion) following manufacturer's instructions, followed by 2 × 30 min DNase treatments with DNA-free (Ambion). To extract ovary RNA, WT ovaries were hand dissected in 1X PBS and total RNA was extracted with the Qiagen RNeasy kit with one on-column DNase treatment. 5'RACE was carried out using total head or ovary RNA with the Invitrogen Gene Racer kit according to manufacturer's instructions (using GeneRacer oligo dT primer for cDNA synthesis). First PCR amplification was carried out using Stratagene Pfu turbo polymerase with GeneRacer 5' primer and the following *dfmr1* specific primer: 5' CACGGGCAGCGTGAAGTGGTAGAA 3'. A second nested round of PCR was carried out using GeneRacer 5' Nested primer and the following *dfmr1* specific primer: 5' GTCCACATCCACAAAGATGCCATCAT 3'. The only prominent band ~500bp was excised and the DNA was extracted and cloned into Invitrogen DNA pCR®-Blunt vector and then multiple clone inserts were sequenced.

2.1.10 QT-PCR

Flies were collected and frozen on dry ice at approximately 15–20 hours after eclosion. Approximately 15–20 heads were pooled together for each biological replicate. RNA extraction protocol, QT-PCR procedure and analysis methods have been previously described, with one exception: Stratagene Brilliant II SYBR QPCR Low ROX Master Mix was used (Pepper et al., 2009). *dfmr1* primers sets (that detect all *dfmr1* transcripts except *dfmr1-RB* as the primers are located in the last exon and 3'UTR) have been previously described in (Pepper et al., 2009) and 28S primer sequences were previously described in (Papacit et al., 2004). Three biological replicates were carried out for each genotype (with one outlier removed for wild-type mRNA) and three technical replicates were carried out for each biological sample.

3.1 Results

3.1.1 *dfmr1* is expressed as two major isoforms

Western analyses for dFMR1 consistently shows the presence of two major dFMR1 isoforms throughout development, including embryonic (and in the cell line derived from late stage embryos, S2 cells), larval, and pupal stages, and in adult fly tissues, including adult heads and ovaries (Fig. 1A). Quantification of the levels of each dFMR1 isoform consistently showed an unequal expression ratio between the isoforms; the slower migrating dFMR1 is expressed at ~0.1–0.2 fold relative to the faster migrating dFMR1 (Figs. 1A and 1B). Although the slower migrating form of dFMR1 seemed to be ubiquitous, the origin of this additional dFMR1 isoform has proven elusive (Schenck et al., 2002, Siomi et al., 2002). Given that the *dfmr1* 5'UTR is relatively long (586nt in ensembl *dfmr1* RA isoform and 422nt in *dfmr1* embryo EST clone LD09557), we hypothesized that an upstream alternative translational start site in the *dfmr1* 5'UTR could account for the slower migrating dFMR1 isoform. The *dfmr1* 5'UTR is sufficient in length to provide enough sequence to encode for

the theoretical 7kD of amino acid sequence that separate the two isoforms, however, the 5'UTR does not contain any additional ATG start codons. Therefore, if the *dfmr1* 5'UTR contained an alternative start codon it would be non-canonical. Although somewhat rare, non-canonical start codons have been identified in several organisms including plants, yeast, flies, and mammals (Kozak, 1997, Touriol et al., 2003, Chang and Wang, 2004).

3.1.2 Alternative translational start site sequence is well conserved among drosophilids

Sequence alignment of the putative *dfmr1* 5'UTR sequences from the 12 sequenced *Drosophila* species' genomes led to the identification of one potential CTG start codon that is conserved in 10/12 *Drosophila* species. Furthermore, in 8/10 species, this potential CTG codon is preceded by sequence that conforms to the *Drosophila* translation start consensus sequence C/A A A C/A A T G (Fig. 2A and (Cavener, 1987, Liu and Xue, 2005)). If this CTG start codon were to be used for translation initiation, the reading frame would be maintained without interrupting stop codons in 6 of the 12 most closely related *Drosophila* species in the *melanogaster* group (Figs. 2A and 2B). Starting with CTG, the *D. melanogaster* 5'UTR sequence would provide an extra 144nt of coding sequence (all within the 3rd exon) that would translate to an additional 48 amino acids at the N-terminus of dFMR1 (Figs. 1 and 2). Such an extension of amino acid sequence is sufficient to provide the extra molecular weight of the slower migrating dFMR1 isoform.

3.1.3 *dfmr1* contains an upstream alternative start codon

To determine whether the *dfmr1* 5'UTR did encode an alternative start codon, we tested whether fusing the *dfmr1* RA 5'UTR in frame with and upstream of a GFP reporter would yield two GFP isoforms. Transient transfection of this GFP reporter construct into S2 cells, followed by western analysis revealed that expression from this reporter does produce two GFP isoforms (~27kDa and ~35kDa), whereas expression of the same GFP construct without the *dfmr1* 5'UTR produced the expected smaller isoform alone (Figs. 2C and 2D). Expression of the *dfmr1* 5'UTR reporter construct with a stop codon (TAG) inserted just upstream of and in frame with the normal GFP start codon produced only the 27kDa GFP isoform (Figs. 2C and 2D), suggesting that translation initiation within the 5'UTR was the origin of the 35kDa GFP isoform. The 35kDa GFP isoform was also eliminated by introducing a point mutation into the *dfmr1* 5'UTR that changed the potential CTG start codon to CAG, suggesting that the conserved CTG in the *dfmr1* 5'UTR, can function as an alternative translational start site in this reporter construct (Figs. 2C and 2D).

The theoretical N-terminal peptide translated from the *Drosophila melanogaster dfmr1* 5'UTR contains many charged and hydrophilic residues, but no clear regions of homology to other known domains or proteins were predicted (using Basic Local Alignment Search Tool for protein (blastp) from the National Center for Biotechnology Information (NCBI) and Fig. 2). Bioinformatics based protein modification prediction programs identified a few potential phosphorylation, glycosylation (O-GlcNAc), and acetylation sites in this N-terminal peptide, but the probabilities were modest and no additional signal sequences were predicted to exist (using Pfam and additional proteomics tools from Expert Protein Analysis System (ExPASy) from Swiss Institute of Bioinformatics (Fig. 2 and (Gasteiger et al., 2003, Finn et al., 2010)).

Although analysis of the predicted dFMR1 N-terminal peptide sequence gave few clues as to how this N-terminal extension might affect dFMR1 function, the fact that the DNA sequence coding for the theoretical peptide is well conserved in the most closely related *Drosophila* species (some of which diverged over 10 million years ago) suggested biological relevance and warranted further investigation. To better characterize the different dFMR1 isoforms *in vivo*, we used a transgenic approach to generate fly strains that would express

only one dFMR1 isoform. A genomic *dfmr1* wild-type rescue (WTR) construct was previously determined to be sufficient to rescue all *dfmr1* null phenotypes tested ((Dockendorff et al., 2002, Banerjee et al., 2010) and Fig. 3A). The vector sequence surrounding this genomic *dfmr1* sequence was engineered for site-specific insertion using the attB/attP system to ensure that all constructs were inserted into the same genomic location (*attP40*, an intergenic location on the second chromosome) (Venken et al., 2006, Bischof et al., 2007, Markstein et al., 2008). Site-specific insertion of the constructs allowed for direct comparisons among transgenic lines. Our initial analyses utilized three transgenic lines: a WTR strain (expresses both dFMR1 isoforms from the *attP40* site), a WTR- S^N strain (expresses only the short dFMR1 isoform (dFMR1- S^N) due to a point mutation at the alternative start codon, CTG→CAG) and a WTR- L^N strain (expresses only the long dFMR1 isoform (dFMR1- L^N) due to a point mutation at the normal start codon, ATG→TTG (methionine to leucine amino acid substitution)) (Fig. 3A). After crossing these WTR lines into the *dfmr1*³ background (*dfmr1* null allele, located on the third chromosome), we assessed how dFMR1- S^N and dFMR1- L^N separately contribute to dFMR1 expression, localization, and function in flies (Dockendorff et al., 2002). All of the following experiments with the WTR strains were conducted in the *dfmr1*³ background, such that WTR;*dfmr1*³ is abbreviated by "WTR" and indicates that the flies are homozygous for the WTR transgene and WTR/+;*dfmr1*³ is abbreviated by "WTR/+" and indicates that the flies are heterozygous for the WTR transgene, etc.

3.1.4 dFMR1- S^N is expressed at appropriate levels without dFMR1- L^N

For all experiments, we compared the transgenic control WTR strain to wild-type controls with respect to dFMR1 levels and dFMR1 function. And secondly, we compared the mutant WTR- S^N and WTR- L^N strains to the WTR strain. Quantitative RTPCR (QT-PCR) results from adult heads showed that *dfmr1* transcript levels were similar amongst the different transgenic lines and similar to endogenous *dfmr1* transcript levels (Fig. 3B). In light of recent findings that dFMR1 levels vary throughout development, dFMR1 protein levels were measured from the transgenic strains by western analyses of head lysates specifically at 15–20 hr post eclosion (Fig. 3C and (Tessier and Broadie, 2008)). Expression levels of dFMR1- S^N and dFMR1- L^N from heterozygous WTR/+ heads were roughly equal to those from its wild-type heterozygous counterpart, *dfmr1*³/+, showing that the control transgenic line produced dFMR1 at close to endogenous levels and maintained the unequal expression ratio between dFMR1- L^N and dFMR1- S^N (Fig. 3C). Similarly, WTR- S^N /+ expressed dFMR1- S^N at levels roughly equal to the levels of dFMR1- S^N from WTR/+. Given that removal of the alternative start codon did not result in increased dFMR1- S^N levels, the *dfmr1* alternative start codon does not likely function to repress translation initiation at the downstream start codon. Analyses of dFMR1 expression from 3rd instar larval (L3) brains yielded results similar to those seen in adult heads (Fig. 3D). Interestingly expression levels of dFMR1- L^N from WTR- L^N /+ were consistently lower than levels of dFMR1- L^N from WTR/+ in adult heads and larval brains (this observation will be examined more closely later) (Figs. 3C and 3D).

Because the relatively low levels of dFMR1- L^N from WTR- L^N flies would complicate any functional comparisons between flies expressing only dFMR1- L^N or only dFMR1- S^N , we also generated a fourth *attP40* genomic *dfmr1* WTR transgenic line that over-expressed dFMR1- L^N . The dFMR1- L^N over-expression line, WTR- L^N .ATG, contains two point mutations: one mutation changes the alternative start codon (CTG) to ATG and the other changes the normal ATG start codon to TTG (Fig. 3A). As visualized by western analyses, WTR- L^N .ATG flies expressed dFMR1- L^N at levels roughly equal to dFMR1- S^N in WTR flies from 15–20 hr adult heads and L3 larval brains (Figs. 3C and 3D). Furthermore, the higher expression levels of dFMR1- L^N from WTR- L^N .ATG flies (~1.5–2 fold over-expression

relative to dFMR1- L^N from *WTR* flies) demonstrated that the mutation from ATG to TTG itself was not the cause of decreased dFMR1- L^N levels from *WTR-L^N* flies. Curiously, longer exposures of the anti-dFMR1 western blots did show a very low level of dFMR1- S^N expression from *WTR-L^N* flies, presumably from the TTG start site, however, there was never any visible dFMR1- S^N production from the *WTR-L^N.ATG* flies (an "over-exposed" representative larval brain western is depicted in Fig. 3E, however, this was also observed in long exposures from westerns with adult heads and ovaries (data not shown)).

Given that dFMR1 is also important for female germ-line development, the *WTR* transgenic lines were analyzed for dFMR1 expression during oogenesis (Costa et al., 2005). In accordance with the nervous system characterizations, expression of either dFMR1- S^N or dFMR1- L^N alone, from *WTR-S^N/+* or *WTR-L^N/+* respectively, was sufficient to produce near endogenous levels of each isoform in western analyses using ovary lysates (Fig. 4A). And dFMR1- L^N levels from *WTR-L^N.ATG/+* ovaries were comparable to dFMR1- S^N levels from *WTR/+* ovaries (Fig. 4A).

We also measured the viability and fertility of the different transgenic strains. The *WTR*, *WTR-S^N*, and *WTR-L^N.ATG* strains inserted at *attP40* and the *WTR* strain from (Dockendorff et al., 2002) all exhibited similar percentages of male and female fertility and all strains partially rescued the reduced viability observed in *dfmr1* null animals (data not shown). The *WTR-L^N* transgenic strain, which expresses dFMR1 at low levels, exhibited a slightly different profile with lower rates of female fertility and viability, but similar levels of male fertility (data not shown).

3.1.5 dFMR1- L^N levels are lower in specific tissues in the absence of dFMR1- S^N

In the female germ line in ovaries, the expression of dFMR1- L^N from *WTR-L^N* animals was nearly equal to that of dFMR1- L^N from *WTR* animals (Fig. 4A). However, in the larval and 15 hr adult brains, dFMR1- L^N expression from *WTR-L^N* animals was consistently lower than dFMR1- L^N levels from *WTR* animals (Fig. 3). Although dFMR1- L^N and dFMR1- S^N are predicted to co-exist temporally (based on western analyses throughout development), we were intrigued by the fact that dFMR1- L^N levels (in the absence of dFMR1- S^N) are lower in larval and adult brains. This developmental and tissue specific difference in isolated dFMR1- L^N levels was different from the comparative expression profile of dFMR1- S^N from *WTR-S^N* flies (dFMR1- S^N levels from *WTR-S^N* flies were essentially equal to dFMR1- S^N levels from *WTR* flies). We hypothesized that dFMR1- S^N might stabilize or promote production of dFMR1- L^N . To test that hypothesis, we measured dFMR1- L^N levels from trans-heterozygous animals (*WTR-S^N/WTR-L^N*, one copy of each transgene) from pharate adult heads (just before eclosion) and adult heads (15–20 hours post-eclosion) using western analyses (Figs. 5A and 5B respectively). The presence of dFMR1- S^N does indeed result in higher levels of dFMR1- L^N in adult heads from trans-heterozygous animals, in comparison to levels of dFMR1- L^N from *WTR-L^N/+* ($p = 0.0286$, Fig. 5A). However, at a slightly earlier time point in development, the levels of dFMR1- L^N from *WTR-L^N/+* pharate adult heads show no further increase in the transheterozygous animals (Fig. 5B). Consequently, these data suggest that dFMR1- S^N is required to modulate dFMR1- L^N levels at specific stages of development, however the requirement of dFMR1- S^N for normal dFMR1- L^N levels is normally masked by the co-expression of both dFMR1- L^N and dFMR1- S^N *in vivo*.

3.1.6 Unequal expression ratio between dFMR1- L^N and dFMR1- S^N is maintained through post-transcriptional control

In addition to differential regulation of dFMR1- L^N levels in isolation, the unequal ratio between dFMR1- L^N and dFMR1- S^N also suggests that dFMR1- L^N is also regulated in the presence of dFMR1- S^N (Fig. 1). To determine whether the preferential selection of the ATG

start site could be explained by higher levels of a shorter, non-CTG containing *dfmr1* transcript, we carried out 5'RACE for the *dfmr1* transcript. Using total RNA from heads and ovaries from wild-type flies, 5'RACE analysis yielded one dominant *dfmr1* 5' start site (Fig. 6A). Subsequent sequencing of the predominant 5'RACE product identified a 5' start site upstream of the alternative CTG start codon, initiating transcription slightly downstream of the 5' end of the LD09557 *dfmr1* cDNA and 5' end of *dfmr1* RA (Fig. 6A). That *dfmr1* carries one predominant 5' transcriptional start site corroborates previous Northern analyses of *dfmr1* that showed only two predominant forms due to alternative polyadenylation sites (Wan et al., 2000, Zhang et al., 2001). Therefore, the mechanism that leads to higher dFMR1-S^N levels (with respect to dFMR1-L^N) most likely occurs downstream of transcription at the level of start codon selection or protein stability.

Additional insight into understanding why dFMR1-S^N levels might be higher relative to dFMR1-L^N levels came from further analysis of the GFP reporter construct used to test the alternative start codon hypothesis in Fig. 2. We were puzzled by the fact that the alternative start codon from the *dfmr1* 5'UTR in the GFP reporter construct led to almost equal levels of both GFP isoforms, but the expression ratio of the endogenous dFMR1 isoforms is always skewed. To test whether the *dfmr1* sequence itself might contribute *in cis* to the skewed ratio, additional GFP reporter constructs were made. Western analysis for steady state expression levels of GFP from transiently transfected S2 cells revealed that replacement of the *α-tubulin* 3'UTR (276bp) with the *dfmr1* 3'UTR (1.8kb) in the GFP reporter construct had almost no effect on the ratio between the GFP isoforms (Figs. 6B and 6C). However, addition of the dFMR1 coding sequence (downstream of the GFP start site) yielded a skewed ratio more similar to the endogenous ratio (Figs. 6B, 6C, and 1). Interestingly, the expression ratio appears to be more dependent on the length of the coding sequence rather than the sequence itself, because addition of a different coding sequence of similar length, *Rasputin*, yielded a similar skewing of the ratio of GFP isoforms (Figs. 6B and 6C).

3.1.7 Heat shock does not alter the steady-state ratio of expression of dFMR1 isoforms

Previous studies have also shown that cellular stresses, such as heat shock, can alter start site preference for non-canonical alternative start codons (Vagner et al., 1996). A common cellular response to different environmental stresses is global translational repression by transiently blocking universally required translation initiation components and formation of large and presumably translationally repressed mRNA and protein complexes termed stress granules (Yamasaki and Anderson, 2008). Consequently, the finding that human FMRP and more recently, dFMR1 are localized to stress granules has sparked interesting speculation about how FMRP might contribute to such global translational repression under different environmental conditions (Mazroui et al., 2002, Linder et al., 2008, Didiot et al., 2009, Farny et al., 2009). Therefore we measured the expression ratios of dFMR1-L^N : dFMR1-S^N at several time points after a 1hr heat shock in wild-type flies by western analyses with adult head lysates (Fig. 7). The ratio of dFMR1-L^N : dFMR1-S^N did not change in the heat-shocked flies relative to mock treated flies, suggesting that regulation of dFMR1-L^N levels is not a component of the *Drosophila* heat-shock response.

3.1.8 dFMR1-S^N and dFMR1-L^N are individually competent for proper dFMR1 localization in the nervous system

In addition to characterizing the bulk expression levels of dFMR1 from the different transgenic lines and reporter constructs by western analyses, cellular localization of dFMR1-L^N and dFMR1-S^N were also examined to test whether one isoform promoted or inhibited dFMR1 localization. We characterized the localization of the dFMR1 isoforms within the central nervous system in adult and larval brains using confocal microscopy and whole-mount staining with anti-dFMR1. dFMR1-S^N and dFMR1-L^N localization in the adult

brains from *WTR-S^N/+* and *WTR-L^N.ATG/+* strains respectively showed normal enrichment of dFMR1 within the central brain in comparison to WT, *dfmr1³/+* and *WTR/+* control strains ((Zhang et al., 2001) and Fig. 8A)). The pattern of dFMR1-L^N staining from *WTR-L^N/+* brains is less clear because dFMR1-L^N levels were undetectable above background staining as compared to *dfmr1* null brains (Fig. 8A).

dFMR1-L^N and dFMR1-S^N also displayed normal dFMR1 enrichment patterns from *WTR-S^N* and *WTR-L^N* larval brains respectively (Fig. 8B). Specifically, both strains exhibited wild-type dFMR1 enrichment in the soma through the midline and lateral edges of the thoracic ganglion of the brain and showed minimal overlap with the neuropil (networks of dendrites and axons, visualized with anti-HRP) as compared to WT and *WTR* larval brains ((Zhang et al., 2001) and Fig. 8B). Additionally, dFMR1 expression within the peripheral nervous system at neuromuscular junctions (NMJ) from 3rd instar larvae showed proper dFMR1 enrichment in the post-synaptic muscles from each *WTR* strain ((Zhang et al., 2001) and Fig. 8C). Similarly, either dFMR1-S^N or dFMR1-L^N was sufficient for proper dFMR1 distribution during oogenesis as well. dFMR1 was enriched in the presumptive oocyte in egg chambers from *WTR/+*, *WTR-S^N/+*, and *WTR-L^N.ATG/+* when dFMR1 expression was examined by whole mount immunofluorescent staining with anti-dFMR1 using confocal microscopy ((Costa et al., 2005) and Fig. 4B). Given that the individual staining patterns of dFMR1-S^N and dFMR1-L^N in the central and peripheral nervous system and female germ line are similar to those of controls, it is unlikely that the N-terminal extension either precludes or promotes proper dFMR1 localization in these tissues.

dFMR1 has also been shown to co-localize with the sub-cellular messenger ribonucleoprotein (mRNP) aggregates known as processing bodies (P bodies) in primary neuronal cells (Barbee et al., 2006). P bodies contain many proteins involved in mRNA decay (such as the decapping protein 1, DCP1) and are linked to other mRNP aggregates including stress granules and neuronal granules (Eulalio et al., 2007a). Therefore we also examined co-localization of P body components with the different dFMR1 isoforms by transiently co-transfecting a C-terminal tandem affinity purification (TAP) tagged (Rigaut et al., 1999) version of the *WTR* constructs (see experimental procedures) with a GFP tagged DCP1 (Eulalio et al., 2007b) construct into S2 cells. TAP tagged isoforms of dFMR1-L^N or dFMR1-S^N were each found competent to co-localize with DCP1 in S2 cells (Fig. 9). Although these results do not unequivocally prove that either isoform in isolation is competent for dFMR1 localization to P bodies *in vivo*, they do suggest that both isoforms are independently sufficient to promote normal dFMR1 localization to P bodies in S2 cells.

3.1.9 Either dFMR1-L^N or dFMR1-S^N is sufficient to promote normal development of synapses at larval neuromuscular junctions

Although dFMR1 levels and distribution are not affected by the absence or presence of the N-terminal peptide, characterizing dFMR1 activity in the different *WTR* transgenic lines could reveal a unique biological requirement for either isoform, as demonstrated from studying the multiple FGF2 isoforms (Yu et al., 2007). Additionally, dFMR1-L^N could function to enhance or antagonize the activity of dFMR1-S^N. Alternatively the extra N-terminal sequence may impact dFMR1-L^N function more subtly such that the functions of either isoform in isolation are indistinguishable. We sought to test these hypotheses by examining the *WTR* strains in developmental processes known to require dFMR1 activity. One particularly well-studied model system used to study synapse formation, is the 3rd instar larval NMJ, which is also known to require dFMR1 for proper development and function (Zhang et al., 2001, Gatto and Broadie, 2008). Normally, the development of specific synaptic junctions between the pre-synaptic neurons and the post-synaptic muscle are tightly regulated such that the neuronal innervation at the NMJ in each larva is highly uniform (Hoang and Chiba, 2001). One quantifiable aspect of NMJ morphology is the

number of boutons (essentially synapses visualized by the pre-synaptic membrane marker anti-HRP and the clustering of a component of the post-synaptic receptor, marked by anti-discs-large (DLG)) present at specific NMJs. *dfmr1* null larvae display an abnormally high number of boutons compared to wild-type larval NMJ at abdominal hemisegment A3, muscle 4 ((Zhang et al., 2001) and Figs. 10A and 10B). The bouton numbers were indistinguishable at A3, muscle 4 among all of the *WTR* strains and WT, but *dfmr1*³ larvae displayed higher numbers of type 1b boutons compared to each of the other strains ($p < 0.001$, $n \geq 14$) (Figs. 10A and 10B). At A3, muscle 6/7, the innervation by type 1b boutons form a more complex pattern, but similarly require dFMR1 for proper morphology ((Zhang et al., 2001) and Fig. 10C)). Again, the transgenic *WTR*, *WTR-S^N*, and *WTR-L^N.ATG* larval strains displayed type 1b bouton numbers significantly different from the *dfmr1* null larvae ($p < 0.001$, $n \geq 25$), and indistinguishable from the WT strain (Fig. 10C). Expression of endogenous levels of dFMR1-L^N from *WTR-L^N* was insufficient to completely rescue the *dfmr1* null phenotype at A3, muscle 6/7 ($n \geq 25$) and presented bouton numbers in-between those from *WTR* and *dfmr1* null strains. That *WTR-L^N* is sufficient to rescue the *dfmr1* null phenotype in muscle 4, but not muscle 6/7 suggests proper NMJ development at muscle 6/7 depends on dFMR1 levels rather than the type of isoform because *WTR-L^N.ATG* rescues the *dfmr1* phenotype in junctions at muscle 4 and 6/7. Thus, with respect to NMJ development, both isoforms can function independently to promote normal synaptogenesis at the NMJ.

3.1.10 dFMR1-L^N and dFMR1-S^N display indistinguishable roles in axon guidance in the adult mushroom bodies

In addition to playing an important role in regulating synapse development in the larval NMJ, dFMR1 is also required for proper development of neuronal processes in the central nervous system. We have assessed the contribution of the different dFMR1 isoforms in dFMR1 function in axon guidance by examining mushroom body morphology in pharate adult brains. The mushroom body is important for learning and memory and consists of 5 lobes, each composed of bundles of neurons that receive sensory input from the olfactory centers (Heisenberg, 2003). Normally during larval and pupal development, the neurons within each lobe are guided to specific three-dimensional coordinates, however in *dfmr1* null animals, there is an abnormally high frequency of β -lobe crossover through the midline, revealing a failure of proper axon guidance during development (Michel et al., 2004, Pan et al., 2004). Whole mount staining of pharate adult brains, using anti-fasciclin II (FasII) as a marker of the α - and β -lobes of the mushroom body was used to characterize the gross morphology of the α -lobes in the different *WTR* strains. The β -lobe morphology was normal in WT and also in all *WTR* strains in contrast to the high frequency of β -lobe crossover typical in the *dfmr1* null strain (Fig. 10D). In light of these results, we conclude that dFMR1-S^N or dFMR1-L^N (even at low endogenous levels) is sufficient to rescue the *dfmr1* null phenotype of β -lobe crossover, indicating again that both isoforms can function independently in neuronal development.

4.1 Discussion

dFMR1 has been known to exist as two main protein isoforms since it was first visualized by western analysis by Wan et al. (Wan et al., 2000). Since that initial observation, several additional post-transcriptional alterations have been identified at the level of the *dfmr1* transcript, including alternative splicing and alternative poly(A) site selection, and post-translational modifications at the level of the dFMR1 protein (including phosphorylation and methylation) (Wan et al., 2000, Zhang et al., 2001, Schenck et al., 2002, Siomi et al., 2002, Stetler et al., 2006, Banerjee et al., 2010). However, previous studies ruled out those modifications (including alternative splicing at the 8/9 exon junction, phosphorylation, and

downstream methionine codons) as potential mechanisms for generating the two predominant dFMR1 isoforms (Schenck et al., 2002, Siomi et al., 2002). The identification of an alternative, non-canonical start codon in the *dfmr1* 5'UTR as the mechanism of production of the second isoform facilitates two different lines of research: 1) *dfmr1* provides another *in vivo* case study of a relatively rare event wherein translation initiation occurs at both a non-canonical start codon and a canonical start codon ((Sugihara et al., 1990, Boyd and Thummel, 1993), and 2) the identification will allow a more precise dissection of the activity and function of dFMR1. The work presented here initiates a characterization of *dfmr1* in both of those research contexts.

4.1.1 dFMR1-L^N levels likely reflect regulation of alternative start codon usage

With respect to the regulation of translation initiation, the downstream ATG codon is preferred throughout development and into adulthood (Fig. 1). In fact the preference for the downstream start codon is so strong, that in the *WTR-L^N* transgenic line, containing the ATG→TTG mutation, the TTG was still occasionally utilized as an initiation codon, even though it is situated in a relatively weak Kozak sequence context (Fig. 2 and 3E and (Liu and Xue, 2005)). However, our western analyses with the *dfmr1* 5'UTR-GFP reporter constructs showed an almost equal ratio between the CTG initiated GFP isoform and the ATG initiated GFP isoform (Fig. 2). Although the GFP reporter expression construct contains a different ATG initiation context from the GFP coding sequence (ideal Kozak sequence) compared with the endogenous *dfmr1* transcript (weak Kozak sequence), changing the length of the coding region itself recapitulated the endogenous expression ratio between dFMR1-L^N and dFMR1-S^N (Fig. 6 and (Liu and Xue, 2005)). The altered ratio between the GFP isoforms suggests that the sequence or length of the open reading frame might contribute to the higher levels of dFMR1-S^N relative to dFMR1-L^N *in vivo*. The change in expression ratios is likely due to an effect on codon preference rather than protein stability conferred by the N-terminal sequence because the presence of the N-terminal extension in the *WTR.ATG-L^N* flies yielded high levels of dFMR1-L^N (Fig. 3). Additionally, the correlation between *cis* sequences and translational start site selection has been recognized in other transcripts. Usage of the downstream ATG alternative start codon in vascular endothelial growth factor (*VEGF*) mRNA has been shown to be dependent on the inclusion (or exclusion) of particular exonic sequences within the *VEGF* coding sequence itself (Bornes et al., 2004). Similarly, addition of a portion of the luciferase coding sequence downstream of the multiple CTG codons in *Drosophila E74A* also altered the preference for different start codons (Boyd and Thummel, 1993). The *dfmr1* 5'UTR-GFP reporter data do not suggest a specific coding sequence to be important for maintaining the uneven dFMR1 isoform ratio. However, start codon usage was affected when elements of the coding region were altered in each construct (*VEGF*, *E74A*, and *dfmr1*) and could indicate a shared, yet unknown mechanistic aspect for start-site selection perhaps through altered secondary structure.

4.1.2 The role of the N-terminal extension in dFMR1 function

With respect to *dfmr1* function, we have found that dFMR1-S^N itself may play a role in start site selection as expression of dFMR1-S^N from trans-heterozygous adult flies (*WTR-S^N/WTR-L^N*) resulted in increased dFMR1-L^N levels relative to *WTR-L^N/+* alone (Fig. 5). Although dFMR1-S^N may affect dFMR1-L^N levels indirectly, dFMR1-S^N might directly bind dFMR1-L^N at certain developmental time points which we would observe as altered dFMR1-L^N stability in the absence of dFMR1-S^N, given that dFMR1 has been shown to bind to itself (Wan et al., 2000, Zhang et al., 2001, Siomi et al., 2002). In fact, Siomi et al. indirectly showed that dFMR1-S^N specifically complexes with dFMR1-L^N because a TAP tagged *dfmr1* construct that did not contain the upstream CTG alternative start codon pulled down endogenous dFMR1-L^N and dFMR1-S^N in S2 cells through immunoprecipitation

(Siomi et al., 2002). Because dFMR1 can also bind to its own *dfmr1* transcript, the change in dFMR1-L^N levels in the absence of dFMR1-S^N could also reflect a role for dFMR1-S^N in promoting CTG start site selection of its own transcript (Zhang et al., 2001).

In addition to studying how the levels of the dFMR1 isoforms are regulated, we also directly examined the functional competence of each dFMR1 isoform independently in different transgenic strains. The conservation of the nucleotide sequence encoding the alternative reading frame in *dfmr1* within the *melanogaster* group suggests that the production of dFMR1-L^N is likely to be biologically functional (Fig. 2). Thus far, we have shown that both dFMR1-L^N and dFMR1-S^N possess full activity when expressed at equal levels by all tests administered. The presence of an N-terminal extension has been shown to promote differential localization in several genes containing non-canonical start codons (for example the N-terminal extension in FGF2 promotes nuclear localization and in VEGF prevents its secretion (Bugler et al., 1991, Huez et al., 2001) and (Touriol et al., 2003)). However, the presence of the N-terminal extension in dFMR1-L^N did not alter dFMR1 localization in any of the cell-types examined (including the central and peripheral nervous systems, and female germ line) (Figs. 4 and 8).

Additionally both dFMR1-L^N and dFMR1-S^N were independently capable of functioning to promote normal development in two processes known to be *dfmr1* dependent: axonal growth in the larval peripheral nervous system at the NMJ and in the adult central brain in the mushroom body (Fig. 10 and (Zhang et al., 2001, Michel et al., 2004). Interestingly, endogenous low levels of dFMR1-L^N (low compared to dFMR1-S^N) were sufficient to rescue the *dfmr1* null synaptic bouton defects at muscle 4 and the β -lobe crossover defect in the mushroom body, but higher levels of dFMR1-L^N (at least equal to endogenous dFMR1-S^N) were required to rescue the *dfmr1* synaptic bouton defects in muscle 6/7. Previous studies have shown that dFMR1 activity is dosage dependent, given that over-expressing dFMR1 often results in opposite phenotypes relative to loss of function *dfmr1* alleles (including (Dockendorff et al., 2002, Morales et al., 2002, Pan et al., 2004, Gatto and Broadie, 2009b)). Consequently, this differential requirement for precise dFMR1 dosage in different muscles likely reflects a mechanistic difference in the developmental program of synaptogenesis in muscle 4 versus 6/7 (indeed these different muscles are innervated by different nerve branches as well (Hoang and Chiba, 2001)).

4.1.3 Conclusion

With the development and characterization of a transgenic system in which either isoform can be analyzed individually *in vivo*, we are poised to investigate many additional aspects of dFMR1 expression and function in an isoform specific manner. For example, it will be interesting to identify the specific protein partners or mRNA species that bind to each isoform, especially in light of the fact that several residues within the annotated N-terminus of dFMR1 have been implicated in protein-protein interactions (Reeve et al., 2008). Furthermore, this model system will facilitate future studies that could reveal novel dFMR1 activities previously concealed by studying dFMR1 as one factor rather than as the two proteins that coexist *in vivo*.

Abbreviations

FMRP	fragile X mental retardation protein 1
dFMR1	<i>Drosophila</i> fragile X mental retardation protein 1
FXS	fragile X syndrome
NMJ	neuromuscular junction

HRP	horse-radish peroxidase
DCP1	decapping protein 1
FGF2	fibroblast growth factor 2
VEGF	vascular endothelial growth factor

Acknowledgments

We thank Steve Liebhaber and Eric Moss for sharing several critical insights and discussions that helped to shape our studies and manuscript. We thank Balpreet Bhogal and Anita Pepper for valuable discussions and technical support throughout the development of the project and an additional thank you to Anita Pepper for making the *dfmr1* WTR-cTAP construct. The *attP40* flies were a gift from Michelle Markstein and Norbert Perrimon, the phi-integrase expressing flies were a gift from Konrad Basler and the S2-NP cells were a gift from Sara Cherry. Additionally we thank Klaus Kaestner and his lab for advising and sharing equipment for the QT-PCR experiments and Gideon Dreyfuss and his lab for guidance and sharing equipment for the Odyssey imaging system.

References

- Ashley CT Jr, Wilkinson KD, Reines D, Warren ST. FMR1 protein: conserved RNP family domains and selective RNA binding. *Science*. 1993a; 262:563–566. [PubMed: 7692601]
- Ashley CT, Sutcliffe JS, Kunst CB, Leiner HA, Eichler EE, Nelson DL, Warren ST. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nat Genet*. 1993b; 4:244–251. [PubMed: 8358432]
- Banerjee P, Schoenfeld BP, Bell AJ, Choi CH, Bradley MP, Hinchey P, Kollaros M, Park JH, McBride SM, Dockendorff TC. Short- and long-term memory are modulated by multiple isoforms of the fragile X mental retardation protein. *J Neurosci*. 2010; 30:6782–6792. [PubMed: 20463240]
- Barbee SA, Estes PS, Cziko AM, Hillebrand J, Luedeman RA, Collier JM, Johnson N, Howlett IC, Geng C, Ueda R, Brand AH, Newbury SF, Wilhelm JE, Levine RB, Nakamura A, Parker R, Ramaswami M. Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. *Neuron*. 2006; 52:997–1009. [PubMed: 17178403]
- Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A*. 2007; 104:3312–3317. [PubMed: 17360644]
- Bornes S, Boulard M, Hieblot C, Zanibellato C, Iacovoni JS, Prats H, Touriol C. Control of the vascular endothelial growth factor internal ribosome entry site (IRES) activity and translation initiation by alternatively spliced coding sequences. *J Biol Chem*. 2004; 279:18717–18726. [PubMed: 14764596]
- Boyd L, Thummel CS. Selection of CUG and AUG initiator codons for *Drosophila* E74A translation depends on downstream sequences. *Proc Natl Acad Sci U S A*. 1993; 90:9164–9167. [PubMed: 8415672]
- Brown V, Small K, Lakkis L, Feng Y, Gunter C, Wilkinson KD, Warren ST. Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein. *J Biol Chem*. 1998; 273:15521–15527. [PubMed: 9624140]
- Bugler B, Amalric F, Prats H. Alternative initiation of translation determines cytoplasmic or nuclear localization of basic fibroblast growth factor. *Mol Cell Biol*. 1991; 11:573–577. [PubMed: 1986249]
- Cavener DR. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Res*. 1987; 15:1353–1361. [PubMed: 3822832]
- Chang KJ, Wang CC. Translation initiation from a naturally occurring non-AUG codon in *Saccharomyces cerevisiae*. *J Biol Chem*. 2004; 279:13778–13785. [PubMed: 14734560]
- Corbin F, Bouillon M, Fortin A, Morin S, Rousseau F, Khandjian EW. The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. *Hum Mol Genet*. 1997; 6:1465–1472. [PubMed: 9285783]

- Costa A, Wang Y, Dockendorff TC, Erdjument-Bromage H, Tempst P, Schedl P, Jongens TA. The *Drosophila* fragile X protein functions as a negative regulator in the orb autoregulatory pathway. *Dev Cell*. 2005; 8:331–342. [PubMed: 15737929]
- D'Hulst C, Kooy RF. Fragile X syndrome: from molecular genetics to therapy. *J Med Genet*. 2009; 46:577–584. [PubMed: 19724010]
- Dicthenberg JB, Swanger SA, Antar LN, Singer RH, Bassell GJ. A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. *Dev Cell*. 2008; 14:926–939. [PubMed: 18539120]
- Didiot MC, Subramanian M, Flatter E, Mandel JL, 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; 20:428–437. [PubMed: 19005212]
- Dockendorff TC, Su HS, McBride SM, Yang Z, Choi CH, Siwicki KK, Sehgal A, Jongens TA. *Drosophila* lacking *dfmr1* activity show defects in circadian output and fail to maintain courtship interest. *Neuron*. 2002; 34:973–984. [PubMed: 12086644]
- Eberhart DE, Malter HE, Feng Y, Warren ST. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum Mol Genet*. 1996; 5:1083–1091. [PubMed: 8842725]
- Estes PS, O'Shea M, Clasen S, Zarnescu DC. Fragile X protein controls the efficacy of mRNA transport in *Drosophila* neurons. *Mol Cell Neurosci*. 2008; 39:170–179. [PubMed: 18655836]
- Eulalio A, Behm-Ansmant I, Izaurralde E. P bodies: at the crossroads of post-transcriptional pathways. *Nat Rev Mol Cell Biol*. 2007a; 8:9–22. [PubMed: 17183357]
- Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol*. 2007b; 27:3970–3981. [PubMed: 17403906]
- Farny NG, Kedersha NL, Silver PA. Metazoan stress granule assembly is mediated by P-eIF2alpha-dependent and -independent mechanisms. *Rna*. 2009; 15:1814–1821. [PubMed: 19661161]
- Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceri G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A. The Pfam protein families database. *Nucleic Acids Res*. 2010; 38:D211–D222. [PubMed: 19920124]
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003; 31:3784–3788. [PubMed: 12824418]
- Gatto CL, Broadie K. Temporal requirements of the fragile X mental retardation protein in the regulation of synaptic structure. *Development*. 2008; 135:2637–2648. [PubMed: 18579676]
- Gatto CL, Broadie K. The fragile X mental retardation protein in circadian rhythmicity and memory consolidation. *Mol Neurobiol*. 2009a; 39:107–129. [PubMed: 19214804]
- Gatto CL, Broadie K. Temporal requirements of the fragile x mental retardation protein in modulating circadian clock circuit synaptic architecture. *Front Neural Circuits*. 2009b; 3:8. [PubMed: 19738924]
- Heisenberg M. Mushroom body memoir: from maps to models. *Nat Rev Neurosci*. 2003; 4:266–275. [PubMed: 12671643]
- Hinton VJ, Brown WT, Wisniewski K, Rudelli RD. Analysis of neocortex in three males with the fragile X syndrome. *Am J Med Genet*. 1991; 41:289–294. [PubMed: 1724112]
- Hoang B, Chiba A. Single-cell analysis of *Drosophila* larval neuromuscular synapses. *Dev Biol*. 2001; 229:55–70. [PubMed: 11133154]
- Huez I, Bornes S, Bresson D, Creancier L, Prats H. New vascular endothelial growth factor isoform generated by internal ribosome entry site-driven CUG translation initiation. *Mol Endocrinol*. 2001; 15:2197–2210. [PubMed: 11731620]
- Khandjian EW, Corbin F, Woerly S, Rousseau F. The fragile X mental retardation protein is associated with ribosomes. *Nat Genet*. 1996; 12:91–93. [PubMed: 8528261]
- Khandjian EW, Huot ME, Tremblay S, Davidovic L, Mazroui R, Bardoni B. Biochemical evidence for the association of fragile X mental retardation protein with brain polyribosomal ribonucleoparticles. *Proc Natl Acad Sci U S A*. 2004; 101:13357–13362. [PubMed: 15329415]

- Kochetov AV, Sarai A, Rogozin IB, Shumny VK, Kolchanov NA. The role of alternative translation start sites in the generation of human protein diversity. *Mol Genet Genomics*. 2005; 273:491–496. [PubMed: 15959805]
- Kozak M. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *Embo J*. 1997; 16:2482–2492. [PubMed: 9171361]
- Laggerbauer B, Ostareck D, Keidel EM, Ostareck-Lederer A, Fischer U. Evidence that fragile X mental retardation protein is a negative regulator of translation. *Hum Mol Genet*. 2001; 10:329–338. [PubMed: 11157796]
- Li Z, Zhang Y, Ku L, Wilkinson KD, Warren ST, Feng Y. The fragile X mental retardation protein inhibits translation via interacting with mRNA. *Nucleic Acids Res*. 2001; 29:2276–2283. [PubMed: 11376146]
- Linder B, Plottner O, Kroiss M, Hartmann E, Laggerbauer B, Meister G, Keidel E, Fischer U. Tdrd3 is a novel stress granule-associated protein interacting with the Fragile-X syndrome protein FMRP. *Hum Mol Genet*. 2008; 17:3236–3246. [PubMed: 18664458]
- Liu Q, Xue Q. Comparative studies on sequence characteristics around translation initiation codon in four eukaryotes. *J Genet*. 2005; 84:317–322. [PubMed: 16385165]
- Markstein M, Pitsouli C, Villalta C, Celniker SE, Perrimon N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet*. 2008; 40:476–483. [PubMed: 18311141]
- Mazroui R, Huot ME, Tremblay S, Filion C, Labelle Y, Khandjian EW. Trapping of messenger RNA by Fragile X Mental Retardation protein into cytoplasmic granules induces translation repression. *Hum Mol Genet*. 2002; 11:3007–3017. [PubMed: 12417522]
- Michel CI, Kraft R, Restifo LL. Defective neuronal development in the mushroom bodies of *Drosophila* fragile X mental retardation 1 mutants. *J Neurosci*. 2004; 24:5798–5809. [PubMed: 15215302]
- Miles CG, Slight J, Spraggon L, O'Sullivan M, Patek C, Hastie ND. Mice lacking the 68-amino-acid, mammal-specific N-terminal extension of WT1 develop normally and are fertile. *Mol Cell Biol*. 2003; 23:2608–2613. [PubMed: 12640141]
- Miyashiro KY, Beckel-Mitchener A, Purk TP, Becker KG, Barret T, Liu L, Carbonetto S, Weiler IJ, Greenough WT, Eberwine J. RNA cargoes associating with FMRP reveal deficits in cellular functioning in *Fmr1* null mice. *Neuron*. 2003; 37:417–431. [PubMed: 12575950]
- Morales J, Hiesinger PR, Schroeder AJ, Kume K, Verstreken P, Jackson FR, Nelson DL, Hassan BA. *Drosophila* fragile X protein, DFXR, regulates neuronal morphology and function in the brain. *Neuron*. 2002; 34:961–972. [PubMed: 12086643]
- Olsen AS, Triemer DF, Sanders MM. Dephosphorylation of S6 and expression of the heat shock response in *Drosophila melanogaster*. *Mol Cell Biol*. 1983; 3:2017–2027. [PubMed: 6419058]
- Pan L, Zhang YQ, Woodruff E, Broadie K. The *Drosophila* fragile X gene negatively regulates neuronal elaboration and synaptic differentiation. *Curr Biol*. 2004; 14:1863–1870. [PubMed: 15498496]
- Papacit M, Orengo D, Juan E. Sequences upstream of the homologous cis-elements of the *Adh* adult enhancer of *Drosophila* are required for maximal levels of *Adh* gene transcription in adults of *Scaptodrosophila lebanonensis*. *Genetics*. 2004; 167:289–299. [PubMed: 15166155]
- Pepper AS, Beerman RW, Bhogal B, Jongens TA. Argonaute2 suppresses *Drosophila* fragile X expression preventing neurogenesis and oogenesis defects. *PLoS ONE*. 2009; 4:e7618. [PubMed: 19888420]
- Rackham O, Brown CM. Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *Embo J*. 2004; 23:3346–3355. [PubMed: 15282548]
- Reeve SP, Bassetto L, Genova GK, Kleyner Y, Leyssen M, Jackson FR, Hassan BA. The *Drosophila* fragile X mental retardation protein controls actin dynamics by directly regulating profilin in the brain. *Curr Biol*. 2005; 15:1156–1163. [PubMed: 15964283]
- Reeve SP, Lin X, Sahin BH, Jiang F, Yao A, Liu Z, Zhi H, Broadie K, Li W, Giangrande A, Hassan BA, Zhang YQ. Mutational analysis establishes a critical role for the N terminus of fragile X mental retardation protein FMRP. *J Neurosci*. 2008; 28:3221–3226. [PubMed: 18354025]

- Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol.* 1999; 17:1030–1032. [PubMed: 10504710]
- Ryder E, Blows F, Ashburner M, Bautista-Llacer R, Coulson D, Drummond J, Webster J, Gubb D, Guntton N, Johnson G, O'Kane CJ, Huen D, Sharma P, Asztalos Z, Baisch H, Schulze J, Kube M, Kittlaus K, Reuter G, Maroy P, Szidonya J, Rasmuson-Lestander A, Ekstrom K, Dickson B, Hugentobler C, Stocker H, Hafen E, Lepesant JA, Pflugfelder G, Heisenberg M, Mechler B, Serras F, Corominas M, Schneuwly S, Preat T, Roote J, Russell S. The DrosDel collection: a set of P-element insertions for generating custom chromosomal aberrations in *Drosophila melanogaster*. *Genetics.* 2004; 167:797–813. [PubMed: 15238529]
- Schenck A, Van de Bor V, Bardoni B, Giangrande A. Novel features of dFMR1, the *Drosophila* orthologue of the fragile X mental retardation protein. *Neurobiol Dis.* 2002; 11:53–63. [PubMed: 12460546]
- Siomi MC, Higashijima K, Ishizuka A, Siomi H. Casein kinase II phosphorylates the fragile X mental retardation protein and modulates its biological properties. *Mol Cell Biol.* 2002; 22:8438–8447. [PubMed: 12446764]
- Siomi MC, 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.* 1996; 16:3825–3832. [PubMed: 8668200]
- Stetler A, Winograd C, Sayegh J, Cheever A, Patton E, Zhang X, Clarke S, Ceman S. Identification and characterization of the methyl arginines in the fragile X mental retardation protein Fmrp. *Hum Mol Genet.* 2006; 15:87–96. [PubMed: 16319129]
- Sugihara H, Andrisani V, Salvaterra PM. *Drosophila* choline acetyltransferase uses a non-AUG initiation codon and full length RNA is inefficiently translated. *J Biol Chem.* 1990; 265:21714–21719. [PubMed: 2123874]
- Tessier CR, Broadie K. *Drosophila* fragile X mental retardation protein developmentally regulates activity-dependent axon pruning. *Development.* 2008; 135:1547–1557. [PubMed: 18321984]
- Touriol C, Bornes S, Bonnal S, Audigier S, Prats H, Prats AC, Vagner S. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biol Cell.* 2003; 95:169–178. [PubMed: 12867081]
- Vagner S, Touriol C, Galy B, Audigier S, Gensac MC, Amalric F, Bayard F, Prats H, Prats AC. Translation of CUG- but not AUG-initiated forms of human fibroblast growth factor 2 is activated in transformed and stressed cells. *J Cell Biol.* 1996; 135:1391–1402. [PubMed: 8947560]
- Venken KJ, He Y, Hoskins RA, Bellen HJ. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science.* 2006; 314:1747–1751. [PubMed: 17138868]
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* 1991; 65:905–914. [PubMed: 1710175]
- Wan L, Dockendorff TC, Jongens TA, Dreyfuss G. Characterization of dFMR1, a *Drosophila melanogaster* homolog of the fragile X mental retardation protein. *Mol Cell Biol.* 2000; 20:8536–8547. [PubMed: 11046149]
- Wang H, Dictenberg JB, Ku L, Li W, Bassell GJ, Feng Y. Dynamic association of the fragile X mental retardation protein as a messenger ribonucleoprotein between microtubules and polyribosomes. *Mol Biol Cell.* 2008; 19:105–114. [PubMed: 17978095]
- Xie W, Dolzhanskaya N, LaFauci G, Dobkin C, Denman RB. Tissue and developmental regulation of fragile X mental retardation 1 exon 12 and 15 isoforms. *Neurobiol Dis.* 2009; 35:52–62. [PubMed: 19362146]
- Yamasaki S, Anderson P. Reprogramming mRNA translation during stress. *Curr Opin Cell Biol.* 2008; 20:222–226. [PubMed: 18356035]
- Yu PJ, Ferrari G, Galloway AC, Mignatti P, Pintucci G. Basic fibroblast growth factor (FGF-2): the high molecular weight forms come of age. *J Cell Biochem.* 2007; 100:1100–1108. [PubMed: 17131363]

- Zalfa F, Eleuteri B, Dickson KS, Mercaldo V, De Rubeis S, di Penta A, Tabolacci E, Chiurazzi P, Neri G, Grant SG, Bagni C. A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat Neurosci.* 2007; 10:578–587. [PubMed: 17417632]
- Zhang Y, O'Connor JP, Siomi MC, Srinivasan S, Dutra A, Nussbaum RL, Dreyfuss G. The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *Embo J.* 1995; 14:5358–5366. [PubMed: 7489725]
- Zhang YQ, Bailey AM, Matthies HJ, Renden RB, Smith MA, Speese SD, Rubin GM, Broadie K. *Drosophila* fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell.* 2001; 107:591–603. [PubMed: 11733059]

Figure 1.

dFmr1 is expressed as two protein isoforms throughout development. (A) Western analyses of protein lysates collected from various developmental time points from wild-type flies: whole embryos 0–3hr, 3–6hr, 6–9hr, 9–12hr, 12–15hr, whole pupae at 2 days and 4 days, 3rd instar larval brains, 1 day-old adult heads, ovaries from adult females, and S2 cells. Western blots were probed with α -dFMR1 and β -tubulin as loading controls. (B) Graph depicts skewed ratio between dFMR1 isoforms from the western blot shown in A, and is representative of other experiments carried out, but not included in this quantification (i.e., slower migrating isoform is expressed at ~0.1–0.2 fold lower than faster migrating form).

Figure 2. *dfmr1* 5'UTR contains an alternative translational start site. (A) Alignment of *dfmr1* 5'UTR sequences. The conserved CTG and the normal ATG start codon are in bold and underscored. Shaded sequence is conserved with the *D. melanogaster dfmr1* 5'UTR. The nucleotide positions are numbered individually for each sequence relative to the ATG start codon. (B) Alignment of theoretical dFMR1 N-terminal peptide translated from CTG (methionine or leucine) to ATG (methionine) in 6/12 *Drosophila* species. The number next to each sequence indicates the number of additional amino acids within the putative N-terminal peptide for each species. Residues predicted to be competent for phosphorylation, acetylation, and/or glycosylation are highlighted in pink, blue, or grey respectively. (C) Schematic of GFP reporter constructs used for transient transfection, comprised of pUASP promoter, GFP coding region, and α -tubulin 3'UTR in all constructs, and *dfmr1* 5'UTR in some constructs. (D) Representative western blot for GFP using protein lysates extracted from S2 cells transiently transfected with the following constructs: (-) no transfection; (GFP) GFP- α -tubulin 3'UTR; (5'UTR-GFP) *dfmr1* 5'UTR-GFP- α -tubulin 3'UTR; (5'UTR+TAG-GFP) *dfmr1* 5'UTR-GFP- α -tubulin 3'UTR with TAG stop codon inserted just upstream of GFP start codon; and (5'UTR CAG-GFP) *dfmr1* 5'UTR-GFP- α -tubulin 3'UTR with point mutation that changes CTG→CAG.

Figure 3.

Transgenic fly strains express dFMR1 isoforms at endogenous levels. (A) Schematic Figure of the *dfmr1* genomic rescue construct denoting the restriction sites used to subclone the construct into an *attB* vector and the sequence around the point mutations introduced to generate constructs that will only produce one isoform of dFMR1. Black boxes and lines represent the *dfmr1* exons and introns respectively and gray lines denote 5' and 3' genomic sequence also contained in the construct. The alternative CTG and canonical ATG start codons are in bold and ellipses denote nucleotides that are absent due to space constraints. Each point mutant construct is illustrated: *WTR-S*^N contains a CTG→CAG transition, *WTR-L*^N contains an ATG→TTG transition, and *WTR-L*^{N.ATG} contains two transitions:

CTG→ATG and ATG→TTG. (B) *dfmr1* transcript levels were graphed as an average ratio to *dfmr1* levels from *WTR/+* flies after normalization to 28S transcript levels using QT-PCR from total head RNA of 15–20hr flies. Error bars are s.e.m. calculated from three biological replicates (with one outlier removed for WT). (C,D) Representative western blots for dFMR1 protein in 15–20 hr adult heads (C) and 3rd instar larval brains (D). Graphed below each representative blot is the quantification of average dFMR1-S^N (white bar) or dFMR1-L^N (black bar) levels (normalized to α -catenin as a loading control) relative to dFMR1-S^N or dFMR1-L^N from *WTR/+* heads. Quantification of dFMR1-L^N relative to dFMR1-S^N from *WTR/+* is graphed as a gray column. Error bars are s.e.m. calculated from at least three biological replicates. (E) Representative shorter (top panel) and longer (middle panel) exposures of a western blot for dFMR1 protein from 3rd instar larval brains. α -catenin protein levels are shown as a loading control.

Figure 4. dFMR1-S^N and dFMR1-L^N are expressed and localized similarly during oogenesis. (A) Representative western blot of dFMR1 protein from ovaries of 3–5 day old females. Graphed below is the quantification of average dFMR1-S^N (white bar) or dFMR1-L^N (black bar) levels (normalized to loading control) relative to dFMR1-S^N or dFMR1-L^N from *WTR/+* ovaries. Quantification of dFMR1-L^N relative to dFMR1-S^N from *WTR/+* is graphed as a gray column. Error bars are s.e.m. calculated from at least three biological replicates. (B,C) Representative images of whole mount immunostaining with α -dFMR1 (green), α -DE-cadherin (blue) and α -Orb (red) in egg chambers, with all images taken together with the

same settings (B) and images taken with similar settings on a different day (C). Scale bars are 25 μ m.

Figure 5. dFMR1-S^N is required for normal dFMR1-L^N expression levels in heads of newly eclosed adults. Representative western blots of dFMR1 protein in 15–20 hr adult heads (A) and pharate adult heads (B). Graphed below each representative blot is the quantification of average dFMR1-S^N (white bar) or dFMR1-L^N (black bar) levels (normalized to α-catenin as a loading control) relative to dFMR1-S^N or dFMR1-L^N from *WTR*/+ heads. Error bars are s.e.m. calculated from at least three biological replicates, **p* = 0.0286 (Mann-Whitney test).

Figure 6.

Uneven expression ratio between dFMR1-L^N and dFMR1-S^N is maintained through post-transcriptional regulation. (A) 5' RACE cDNA from head and ovary RNA extracts were amplified by RT-PCR using nested primers and the products were visualized using an ethidium bromide stained agarose gel. One main *dfmr1* 5' start transcription site was identified (indicated by arrows). Additional control lanes in which the PCR reaction was carried out with no cDNA or only with the forward nested primer "F" are also shown. Below the RT-PCR is a summary of the sequencing results from the bands labeled with the arrows and the 5' sequence from two other previously described *dfmr1* transcripts (*dfmr1* RA, and embryo EST clone LD09557) are included above the 5'RACE sequence results for reference.

Numbers in superscript refer to the nucleotide position after the transcription start, which is denoted as "1." (B) A schematic of GFP reporter constructs used for transient transfection in S2 cells is placed adjacent to each representative western blot for GFP protein. Approximate masses of GFP (in kilodaltons) if ATG start codon from GFP is used and approximate lengths of various 3'UTR sequences (in base pairs) are denoted within the schematic. (C) Quantification of (B) as ratio between long-GFP/short-GFP isoforms graphed as the mean ratio from three biological replicates and error bars are s.e.m.

Figure 7.

Ratio of dFMR1-L^N to dFMR1-S^N levels remains constant after heat shock. Representative western blot for dFMR1 protein in 1–2 day old adult heads from WT flies at various time points (0 min, 5min, 10min, 15min, 30 min, or 60 min) after a 1hr treatment at 25°C (control) or 39°C (heat shock). Levels of β -tubulin are used as a loading control and levels of S6 kinase (phosphorylated at Thr 398) were used to show that the heat-shock treatment did indeed elicit a heat shock response (S6 Kinase is dephosphorylated at time 0 min after heat shock, but not in the control) (Olsen et al., 1983). The graph illustrates the ratio of dFMR1-L^N signal to dFMR1-S^N signal at each time point from control flies (white bars) and heat-shocked flies (black bars) and is representative of another experiment carried out, but not included in this quantification.

Figure 8. *dfmr1-L^N* and *dfmr1-S^N* are similarly localized in the central and peripheral nervous system. (A) Representative images of whole mount immunostaining with α -dFMR1 (green) in 2–5 day male brains. Scale bar is 25 μ m. (B) Image box set apart shows dFMR1 staining throughout a WT 3rd instar larval brain, with the thoracic ganglion encircled. Below are representative images of whole mount immunostaining with α -dFMR1 (green) and α -HRP (red) specifically from the thoracic ganglion. Scale bar is 25 μ m. Inset boxes magnify a few cells from the whole image to illustrate the cytoplasmic localization of dFMR1 (note that *dfmr1-L^N* levels in inset box for *WTR-L^N* were increased from original image using levels adjustments to emphasize *dfmr1-L^N* localization). (C) Representative images of whole

mount immunostaining with α -dFMR1 (green, enriched in post-synaptic muscle) and α -HRP (red, highlighting pre-synaptic innervation) in 3rd instar larval NMJ at A3, muscle 6/7. Scale bar is 25 μ m.

Figure 9.

Both dFMR1- L^N and dFMR1- S^N associate with P body component DCP1 in S2 cells. Representative confocal images from S2 cells transiently co-transfected with carboxyl-terminal TAP tagged *dfmr1* WTR constructs (visualized with anti-protein A, red) and GFP-DCP1 (visualized without antibodies, green). The WTR-cTAP construct produces both dFMR1- L^N and dFMR1- S^N , whereas WTR- S^N -cTAP only produces dFMR1- S^N , and WTR- L^N -cTAP only produces dFMR1- L^N . The constructs are described in more detail in the experimental procedures section. Scale bar is 10 μ m.

Figure 10. *dFMR1-L^N* is not required for *dFMR1* function in specifying proper morphology within the peripheral and central nervous systems. (A) Representative images of whole mount immunostaining with α -DLG (green) and α -HRP (magenta) in 3rd instar larvae at NMJ A3, muscle 4. Scale bar is 25 μ m. (B,C) Quantification of average bouton number (type 1b from A3 muscle 4 in (B) and type1b from A3 muscle 6/7 (C), error bars are s.e.m. *** $p < 0.001$ (Kruskal-Wallis test (nonparametric ANOVA) analysis followed by Dunn's multiple comparisons test, indicated a statistically significant probability that there is a difference between genotypes with asterisks and *dfmr1¹*. (D) Representative images of whole mount immunostaining with α -fasciclin II (green) in pharate adult brains. Percentage of abnormal brains (with β -lobe fusion) and number of brains (n) analyzed for the percentage calculation are listed below each representative image. Scale bar is 50 μ m for all images except for *WTR-L^N* (100 μ m).