

New insights into Fragile X syndrome

Relating genotype to phenotype at the molecular level

Irina Pozdnyakova¹ and Lynne Regan^{1,2}

¹ Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

² Department of Chemistry, Yale University, New Haven, CT, USA

Keywords

Drosophila Fragile X related protein; Fragile X syndrome; KH domains; NMR; stability

Correspondence

L. Regan, Department of Molecular Biophysics and Biochemistry, Yale University, PO Box 208114, New Haven, CT 06520-8114, USA
E-mail: lynne.regan@yale.edu

(Received 6 October 2004, revised 30 November 2004, accepted 13 December 2004)

doi:10.1111/j.1742-4658.2004.04527.x

Lack of functional Fragile X mental retardation protein (FMRP) is the primary cause of the Fragile-mental retardation syndrome in humans. In most cases, the disease results from transcriptional silencing of fragile mental retardation gene 1, *fmr1*, which encodes FMRP. However, a single missense mutation (I304N) in the second KH domain of FMRP gives rise to a particularly severe case of Fragile X syndrome. A *Drosophila* homolog of FMRP has been identified, *Drosophila* Fragile X related protein (dFXRP). The corresponding missense mutation in dFXRP, the I307N, has pronounced effects on the *in vivo* activity of the protein. The effect of the point mutation on the structure and function of FMRP is unclear, and published data are contradictory. No *in vitro* structural or stability studies have been performed on dFXRP. Here we show that a construct that contains only the tandem KH1-KH2 domains is a stable, well-folded unit suitable for detailed structural and functional characterization. Using this KH1-KH2 construct we explicitly test a hypothesis that has been proposed to explain the effect of the Ile → Asn mutation: that it causes complete unfolding of the protein. Here we show that the I307N point mutation does not completely unfold the KH domain. The KH1-KH2 construct bearing I307N substitution is stable in isolation and adopts a native-like fold. Thus our data favor alternative explanations for the *in vivo* observed loss of dFXRP activity associated with I307N mutation: (a) the point mutation might affect intra and/or inter-molecular interactions of dFXRP; or (b) it might impair dFXRP's interactions with its RNA target(s).

Fragile X syndrome is the most common cause of inherited mental retardation in humans. It is an X-linked disorder that occurs with a frequency of approximately 1 in 3000 males [1]. A number of cognitive and physical abnormalities are associated with the syndrome: mental retardation, behavioral problems, facial dysmorphism, connective tissue abnormalities, testicular enlargement in males, and premature ovarian failure in females [2,3].

The syndrome takes its name from the physical 'fragility' of the chromosome. Chromosome breakage occurs at the tip of the X chromosome at the FRAXA

locus, Xq27.3. The fragility is associated with an expansion of CGG repeats upstream of the *fmr1* gene, which also causes increased DNA methylation, silencing of gene expression, and a substantial reduction in the levels of Fragile X mental retardation protein (FMRP) [4–10].

The sequence of FMRP hints at its function [11]. Nuclear localization and nuclear export signals (NLS and NES), tandem K-homology domains (KH1 and KH2) and an RGG box, suggest that FMRP may shuttle in and out of the nucleus and that it may have an RNA binding activity. Nevertheless, the actual

Abbreviations

dFMRP, *Drosophila* Fragile X related protein; FMRP, Fragile X mental retardation protein; NLS, nuclear localization signal; NES, nuclear export signal; KH1 and KH2, K-homology domains; FXRP, Fragile X related protein; RNP, ribonuclear protein.

cellular role of FMRP is still far from clear. It has been suggested that the protein may play a role in some, or all, of the following: (a) nuclear–cytoplasmic shuttling of RNA; (b) translational control; (c) dendritic transport of RNA; (d) dendrite-specific translation and regulation [12].

An individual with an especially severe manifestation of Fragile X syndrome provides an important insight into the molecular basis of the disease. In this case there is no CGG repeat expansion upstream of *fmr1* but there is a single point mutation (I304N) within the KH2 domain [13]. This finding has important implications for Fragile X research, because it directly implicates lack of FMRP function as the primary cause of the syndrome.

In spite of its key importance, the effect of the I304N mutation on the structure and function of FMRP is still unclear, and there have been several contradictory reports in the literature. Dreyfuss and colleagues initially concluded, using total brain RNA, that the I304N mutation severely impairs FMRP's RNA binding ability [14]. Using different assays, other groups have reported that the I304N mutation causes no diminution of RNA-binding [15]. Pastore and colleagues suggested that the I304N mutation introduces a polar residue into the hydrophobic core and thus unfolds the protein [16] (Fig. 1). They explored this possibility by introducing an asparagine residue at the equivalent position in the isolated KH1 domain of FMRP (I241N). The KH1 domain carrying this mutation was completely unfolded. Although at first sight this result provides an attractive explanation for the severe effect of this mutation, it is important to note that the clinically relevant mutation, I304N, occurs in



Fig. 1. NMR structure of the KH1 domain from FMRP (PDB code 2FMR). The KH1 domain has a typical KH β/α fold. The helices pack against a three-stranded antiparallel β -sheet forming the hydrophobic core of the protein. The conserved isoleucine residue is shown in orange.

the KH2 domain not the KH1 domain. Darnell and colleagues speculated that the effect of the I304N mutation might be to disrupt a 'hydrophobic platform' involved in RNA recognition. They had observed such a mode of KH domain–RNA interaction in the cocrystal structure of the KH3 domain of NOVA bound to an *in vitro* selected RNA [17,18]. Yet another hypothesis is that the I304N mutation causes FMRP to associate into ribonuclear protein (RNP) particles of abnormal density, perhaps by disruption of protein–protein interactions [15]. To summarize, three main hypotheses have been proposed to explain the deactivating effect of I304N mutation: (a) the I304N mutation causes unfolding of FMRP; (b) it impairs binding of target RNAs by FMRP; (c) it precludes association of FMRP in RNP particles.

Thus, despite the fact that the I304N mutation causes Fragile X syndrome, and that investigations of the I304N mutation play a key role in Fragile X research, there is no consensus in the Fragile X field on either the structural or functional consequences of the mutation.

Because of the limitations in working with humans, Fragile X studies have expanded to other organisms. In vertebrates there are two autosomal proteins that are highly homologous to FMRP, Fragile X related proteins 1 and 2 (FXR1 and FXR2), which are often found in association with each other, and with FMRP, in RNP complexes [19]. Their tissue distribution largely overlaps with that of FMRP. In more primitive organisms, *Drosophila melanogaster* for example, there is a single protein, dFXRP (also referred to as dFMRP). dFXRP contains all the key sequence features of FMRP, FXR1P and FXR2P: KH, RGG, NLS and NES motifs. Phylogenetic sequence analysis led to the suggestion that dFXRP is the ancestral progenitor of the vertebrate FMRP and FXR proteins, and that in flies this single protein performs the functions carried out by the entire FMRP family in vertebrates [20].

Both dFXRP-null mutations and overexpression of dFXRP result in altered synaptic development and function in *Drosophila*. The dFXRP-null mutant flies display pronounced synaptic overgrowth, overelaboration of synaptic terminals, increased branching and an increased number of synaptic boutons [21] morphological differences which correlate well with the dendritic spine overgrowth observed in Fragile X individuals. The tissue distribution patterns and subcellular localization of dFXRP resembles the combined expression pattern of mammalian FMRP and FXR proteins [20,22], adding support to the hypothesis that dFXRP is a functional homolog of human FMRP and FXR proteins.

Dreyfuss and colleagues introduced the I307N point mutation in dFXRP (the equivalent of I304N in human FMRP) [20]. Wild-type dFXRP and dFXRP(I307N) were overexpressed in *Drosophila* eye using eye-specific promoters. Overexpression of the wild-type dFXRP results in a rough eye phenotype, which has been shown to be a consequence of the induction of apoptosis. Overexpression of dFXRP(I307N), however, causes a much less severe phenotype, and much milder changes in the photoreceptor pattern (Fig. 2). Thus, the wild-type dFXRP activity exhibited when

over-expressed in the fly eye is significantly diminished by the I307N mutation.

Full-length dFXRP is 684 amino acids long and cannot be purified in significant quantities. We have therefore developed the isolated KH domains as an experimentally tractable system. We have tested the proposed hypothesis of the deactivating-by-unfolding effect of I307N mutation using protein constructs containing the KH region of dFXRP.

Results and Discussion

The mammalian FMRP, FXRP1, FXRP2 and fly dFXRP proteins are homologous and the KH motifs are especially highly conserved (70% identity) (Fig. 3). For our studies of the effect of the I307N mutation, which occurs in the second KH domain, we have chosen to use the tandem KH1-KH2 construct rather than the isolated KH2 domain, because inter-domain interactions between two arrayed KH modules have been noted, for example in the bacterial transcription factor NusA [23]. In this example, the crystal structure reveals the presence of an elaborate H-bonding network between the two C-terminal helices of one KH domain and the β -sheet of the preceding KH motif. The contiguous arrangement of the KH1-KH2 domains in NusA has been proposed to provide an extended RNA-binding surface. These considerations suggest that tandem KH1-KH2 domains might be a minimal biologically relevant system for studies of the structural role of I307N mutation.

We cloned and expressed the KH1-KH2 and the KH1-KH2(I307N) domains of dFXRP. Both constructs express well, are folded and can be readily purified from the soluble fraction of *Escherichia coli*.

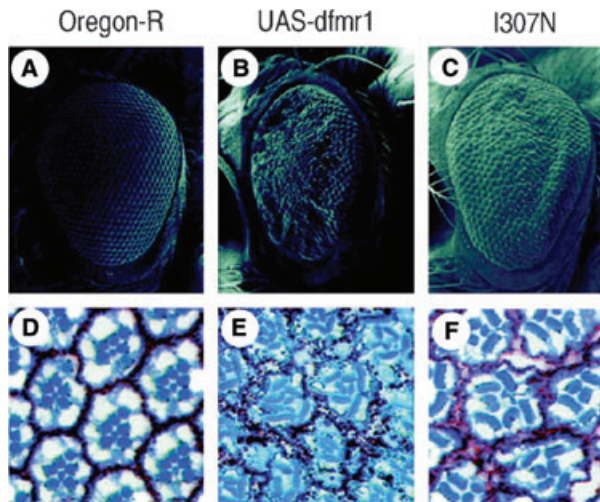


Fig. 2. Fly eye phenotypes associated with overexpression of dFXRP(wild-type) and dFXRP(I307N). (A) A normal fly eye. (B) dFXRP(wild-type) overexpressed in fly eye: note the rough eye phenotype. (C) dFXRP(I307N) overexpressed in fly eye: the effect on eye morphology is much less severe. (D–F) Cross-sections of ommatidia from eyes of (A), (B) and (C). (Adapted from [20])

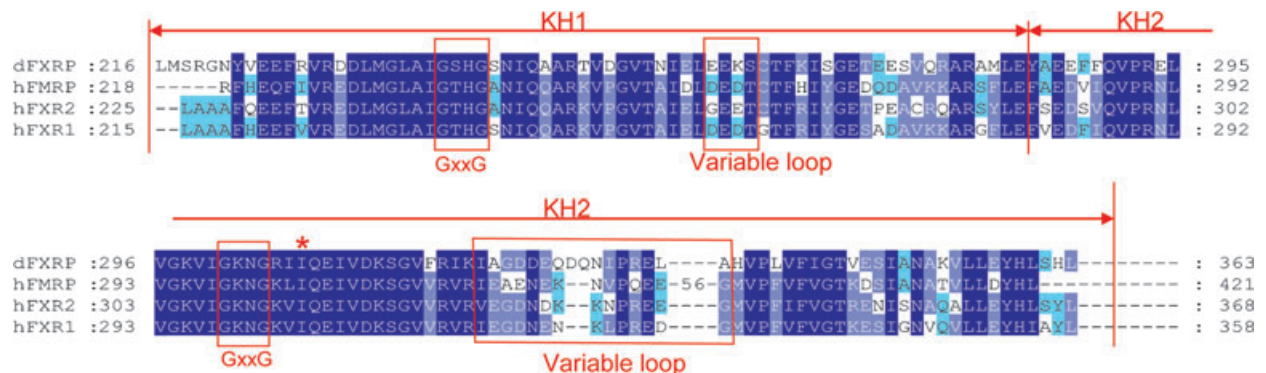


Fig. 3. Multiple-sequence alignment of KH domains from dFXRP and the human Fragile X proteins (in this figure named hFMRP, hFXR1, hFXR2). Conserved residues are colored in dark blue and semiconserved residues are colored in lighter shades of blue. The domain boundaries, the conserved signature motif GxxG and a variable loop between β -strands β 2 and β 3 which was proposed to have functional significance are indicated. The conserved isoleucine residue is marked with an asterisk. Residue numbering corresponds to the full-length proteins.

Fig. 4. Biophysical characterization of KH constructs from *Drosophila* dFXRP. (A) CD spectra of KH1-KH2(wild-type) (red) and KH1-KH2(I307N) (blue). (B) Chemical denaturation curve of KH1-KH2(wild-type) (red) and KH1-KH2(I307N) (blue). (C) Thermal denaturation curve of KH1-KH2(wild-type) (red) and KH1-KH2(I307N) (blue).

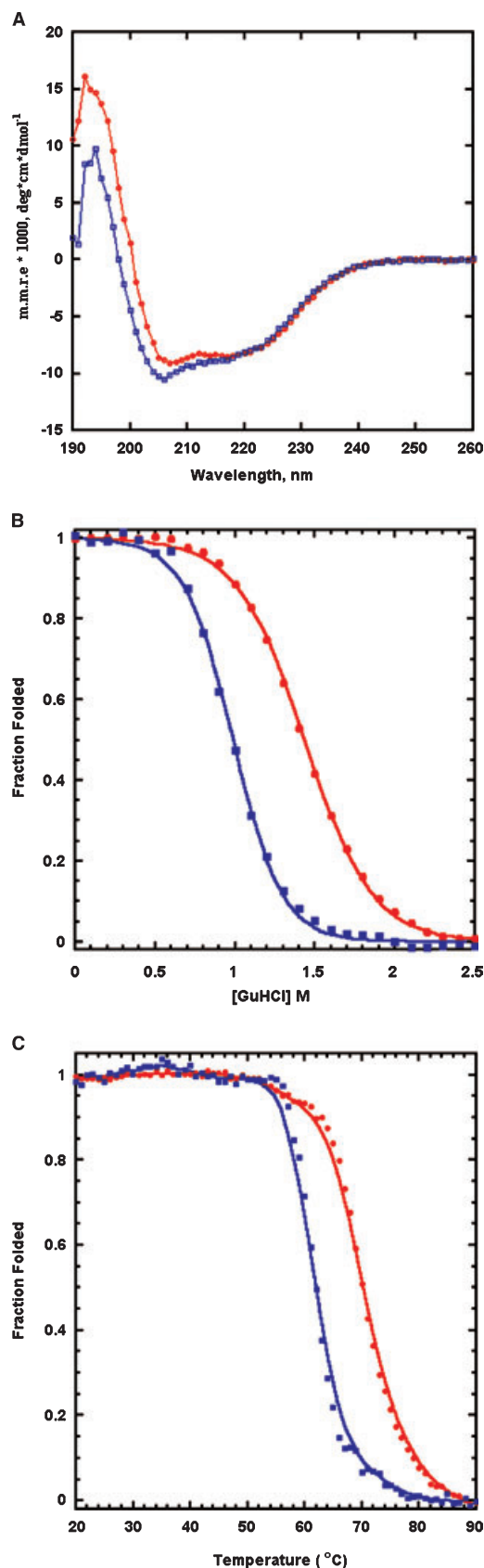
The CD spectrum of the KH1-KH2(wild-type) is dominated by the α -helical signal and is typical of that observed for K homology domains (Fig. 4A). A comparison of the KH1-KH2(wild-type) and the KH1-KH2(I307N) CD spectra shows minor differences, indicating that the effect of the mutation is not to completely unfold the protein.

The wild-type KH1-KH2 construct displays a cooperative and reversible (data not shown) chemical denaturation transition with the midpoint of GuHCl-induced unfolding 1.4 M (Fig. 4B). The calculated free energy of unfolding is $3.9 \pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ($16 \pm 0.5 \text{ kJ}\cdot\text{mol}^{-1}$). The thermal denaturation transition is cooperative (Fig. 4C) but irreversible (data not shown).

The KH1-KH2(I307N) protein also exhibits cooperative chemical and thermal denaturation transitions (Fig. 4B,C). Again, only the chemical denaturation transition is reversible. The thermodynamic stability of KH1-KH2(I307N) is slightly reduced relative to KH1-KH2(wild-type), but the mutation certainly does not completely unfold the protein. The calculated free energy of unfolding of KH1-KH2(I307N) is $3.5 \pm 0.1 \text{ kcal}\cdot\text{mol}^{-1}$ ($14.7 \pm 0.5 \text{ kJ}\cdot\text{mol}^{-1}$).

To further investigate the structural effects of the I304N mutation, we performed ^{15}N - ^1H HSQC experiments on ^{15}N uniformly labeled KH1-KH2(wild-type) and KH1-KH2(I307N). Both proteins show the presence of a well-defined globular fold, with many well-dispersed cross-peaks in the HSQC spectra (Fig. 5). We observe that several cross-peaks in the spectrum of KH1-KH2(I307N) are shifted relative to the corresponding peaks in the KH1-KH2(wild-type) spectrum. These differences are few, and are likely to be a result of local perturbations in the vicinity of the introduced mutation, with the overall fold conserved.

The data presented here represent the first biophysical characterizations of the KH domains of dFXRP. The structure and stabilities of KH1-KH2(wild-type) and KH1-KH2(I307N) were compared, and show that the I307N mutation does not cause complete unfolding of the KH2 domain. Our results demonstrate that the I307N mutation, which significantly reduces the *in vivo* activity of dFXRP in *Drosophila* eye, does so not by disrupting the KH domain tertiary fold but, most



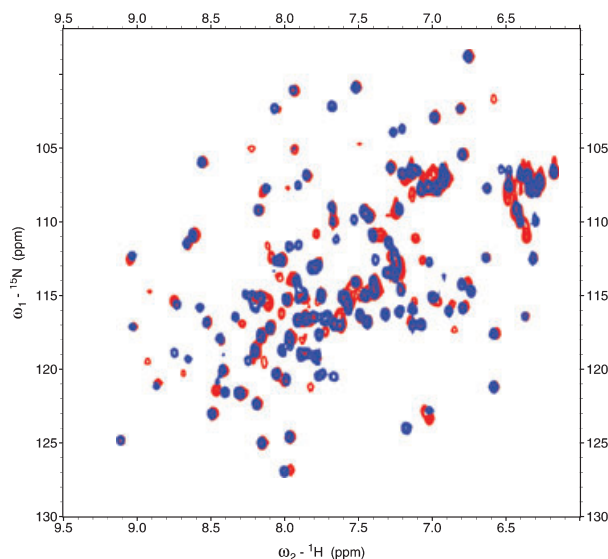


Fig. 5. The overlay of HSQC spectra of ^{15}N uniformly labeled KH1-KH2(wild-type) (red) and KH1-KH2(I307N) (blue). The data were recorded at 600 MHz on 1.2 mM protein samples at 25 °C.

likely, by specifically effecting dFXRP interactions with its biological partner(s).

Experimental procedures

Cloning, protein expression and purification

The DNA coding the KH domains of dFXR, KH1-KH2 (amino acids 284–363), was PCR amplified from the full-length dFXRP clone (a generous gift of H Siomi, Institute for Genome Research, University of Tokushima, Japan). An *NcoI* restriction site, followed by His-tag and Tobacco Etch Virus (TEV) protease cleavage site were incorporated at the 5' end of the construct and a *BamHI* restriction site at the 3' end. The PCR product was digested with *NcoI* and *BamHI* restriction enzymes (New England Biolabs, Beverly, MA, USA) and cloned into pET15b vector (Novagen, Madison, WI, USA). Ile307Asn point mutation was introduced into the wild-type KH1-KH2 construct using QuikChange Mutagenesis kit (Stratagene, La Jolla, CA, USA).

Plasmids containing corresponding inserts were transformed into BL21 Gold (DE3) cells. The bacteria were grown in Luria–Bertani medium at 37 °C. The cells were induced with 1 mM isopropyl thio- β -D-galactoside after the cell culture reached attenuation of 0.6–0.7 at 600 nm. The growth was continued for additional 5 h at 30 °C. The cells were harvested by centrifugation and resuspended in 50 mM Tris/HCl buffer containing 100 mM NaCl, 5% glycerol, 1 mg mL⁻¹ lysozyme and complete EDTA-free protein inhibitor cocktail (Roche, Basel, Switzerland). After a 30-min incubation on ice, the cells were lysed by sonication. The soluble fraction of the whole cell lysate was incubated with Ni-NTA matrix

(Qiagen Inc., Valencia, CA, USA) and His-tagged proteins were then eluted with imidazol-containing buffer. The N-terminal His-tag was removed by cleavage with TEV protease (Invitrogen, Carlsbad, CA, USA) followed by gel filtration chromatography (High load Superdex RH-75 column, Amersham Biosciences, Uppsala, Sweden) yielding pure proteins of interest. Protein concentration was determined spectroscopically by measuring UV absorbance at 280 nm. The extinction coefficients for each studied construct were calculated from amino acid composition [24] using EXPASY PROTPARAM tool. Protein purity was confirmed by SDS/PAGE.

Thermodynamic stability studies

CD spectra were recorded at 20 μM protein concentration (50 mM phosphate buffer pH 7.0, 100 mM NaCl, 1 mM dithiothreitol) at 25 °C in a 0.1-cm path-length cuvette using AVIV spectrophotometer Model 215 (AVIV Instruments Inc.). Thermal and chemical denaturation transitions were monitored by CD absorption at 222 nm. Thermal scans were performed in the forward and reverse direction from 15 °C to 95 °C in 1 °C steps with equilibration time of 1 min at each temperature.

Chemical denaturation was induced by small additions of GuHCl (Ultra pure grade, ICN Biomedicals Inc., Aurora, OH, USA). The titrations were performed in automatic mode. At each titration point (0.1-M stepwise increase in denaturant concentration) the ellipticity was monitored after 10-min equilibration time (established to be sufficient for achieving an equilibrium) with stirring; denaturant solution contained the protein of interest at the same concentration as the titrate solution of folded protein in the cell, thus protein concentration was kept constant during the course of the experiment. Measurements were performed in 1-cm path-length cuvette.

Unfolding transitions were analyzed using a two-state model to determine $\Delta G_{\text{U}}(\text{H}_2\text{O})$ and m -values. The transition-midpoints were calculated as $\Delta G_{\text{U}}(\text{H}_2\text{O})/m$. The experimental unfolding curve was fitted (in KALEIDAGRAPH) to the following expression derived for a two-state process:

$$Y_{\text{obs}} = \{Y_{\text{U}} + Y_{\text{F}}[\exp((\Delta G_{\text{U}}(\text{H}_2\text{O}) - m[\text{G}_{\text{U}}\text{HCl}]/RT)]/[1 + \Delta G_{\text{U}}(\text{H}_2\text{O}) - m[\text{G}_{\text{U}}\text{HCl}]/RT]\}$$

where Y_{obs} , Y_{U} , and Y_{F} are the observed spectroscopic signal, denatured-protein baseline, and folded-protein baseline, respectively. From the fit, $\Delta G_{\text{U}}(\text{H}_2\text{O})$, the free energy of unfolding in aqueous solution, and m , the dependence of the free energy on denaturant concentration are calculated.

NMR spectroscopy

NMR spectra were recorded on Varian Inova (Varian Inc., Palo Alto, CA, USA) 600 MHz with 0.5–1 mM ^{15}N -labeled proteins in 50 mM Tris/HCl pH 7.2, 100 mM NaCl, 10% D₂O. The data were recorded at 25 °C, processed with

NMRPIPE [25] and analyzed with SPARKY (Goddard, University of California, San Francisco, CA, USA).

Acknowledgements

We thank Dr H. Siomi for kindly sending us the full-length dFXRP clone. This work was supported in part by grants to L.R. from the March of Dimes and the Fragile X Foundation.

References

- Morton JE, Bunday S, Webb TP, MacDonald F, Rindl PM & Bullock S (1997) Fragile X syndrome is less common than previously estimated. *J Med Genet* **34**, 1–5.
- Hagerman RJ, Staley LW, O'Connor R, Lugenbeel K, Nelson D, McLean SD & Taylor A (1996) Learning-disabled males with a fragile X CGG expansion in the upper premutation size range. *Pediatrics* **97**, 122–126.
- Lachiewicz AM, Dawson DV & Spiridigliozzi GA (2000) Physical characteristics of young boys with Fragile X Syndrome: Reasons for difficulties in making a diagnosis in young males. *Am J Med Genet* **92**, 229–236.
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang FP *et al.* (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914.
- Kooy RF, Oostra BA & Willems PJ (1998) The fragile X syndrome and other fragile site disorders. *Results Problems Cell Differ* **21**, 1–46.
- Kooy RF, Willemsen R & Oostra BA (2000) Fragile X syndrome at the turn of the century. *Mol Med Today* **6**, 193–198.
- Cummings CJ & Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. *Ann Rev Genomics Hum Genet* **1**, 281–328.
- Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT & Nelson DL (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817–822.
- Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D & Warren ST (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet* **1**, 397–400.
- Tamanini F, Kirkpatrick LL, Schonkeren J, van Unen L, Bontekoe C, Bakker C, Nelson DL, Galjaard H, Oostra BA & Hoogeveen AT (2000) The fragile X-related proteins FXR1P and FXR2P contain a functional nucleolar-targeting signal equivalent to the HIV-1 regulatory proteins. *Hum Mol Genet* **9**, 1487–1493.
- Siomi H, Siomi MC, Nussbaum RL & Dreyfuss G (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* **74**, 291–298.
- Jin P & Warren ST (2000) Understanding the molecular basis of fragile X syndrome. *Hum Mol Genet* **9**, 901–908.
- De Boule K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, de Graaff E, Oostra BA & Willems PJ (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet* **3**, 31–35.
- Siomi H, Choi M, Siomi MC, Nussbaum RL & Dreyfuss G (1994) Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* **77**, 33–39.
- Feng Y, Absher D, Eberhart DE, Brown V, Malter HE & Warren ST (1997) FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Mol Cell* **1**, 109–118.
- Musco G, Kharrat A, Stier G, Fraternali F, Gibson TJ, Nilges M & Pastore A (1997) The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat Struct Biol* **4**, 712–716. (erratum appears in *Nat Struct Biol* **4**, 840).
- Lewis HA, Chen H, Edo C, Buckanovich RJ, Yang YY, Musunuru K, Zhong R, Darnell RB & Burley SK (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure Fold Des* **7**, 191–203.
- Lewis HA, Musunuru K, Jensen KB, Edo C, Chen H, Darnell RB & Burley SK (2000) Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323–332.
- Zhang Y, O'Connor JP, Siomi MC, Srinivasan S, Dutra A, Nussbaum RL & Dreyfuss G (1995) The fragile-mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J* **14**, 5358–5366.
- Wan L, Dockendorff TC, Jongens TA & Dreyfuss G (2000) Characterization of dFMR1, a *Drosophila melanogaster* homolog of the fragile X mental retardation protein. *Mol Cell Biol* **20**, 8536–8547.
- Zhang YQ, Bailey AM, Matthies HJ, Renden RB, Smith MA, Speese SD, Rubin GM & Broadie K (2001) *Drosophila* fragile X-related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. [see comment]. *Cell* **107**, 591–603.
- Morales J, Hiesinger PR, Schroeder AJ, Kume K, Verschuuren P, Jackson FR, Nelson DL & Hassan BA (2002)

- Drosophila* fragile X protein, DFXR, regulates neuronal morphology and function in the brain. *Neuron* **34**, 961–972.
- 23 Worbs M, Bourenkov GP, Bartunik HD, Huber R & Wahl MC (2001) An extended RNA binding surface through arrayed S1 and KH domains in transcription factor NusA. *Mol Cell* **7**, 1177–1189.
- 24 Gill SC & von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319–326.
- 25 Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J & Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. [see comment]. *J Biomol NMR* **6**, 277–293.