



## A low resolution study of the ultrastructure of Fragile X related proteins

Ljiljana Sjekloća, Petr V Konarev, John Eccleston, Ian A Taylor, Dmitri I Svergun, Annalisa Pastore

### ► To cite this version:

Ljiljana Sjekloća, Petr V Konarev, John Eccleston, Ian A Taylor, Dmitri I Svergun, et al.. A low resolution study of the ultrastructure of Fragile X related proteins. *Biochemical Journal*, 2009, 419 (2), pp.347-357. 10.1042/BJ20082197 . hal-00479125

**HAL Id: hal-00479125**

**<https://hal.science/hal-00479125>**

Submitted on 30 Apr 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**A low resolution study of the ultrastructure of Fragile X related proteins**

Ljiljana Sjekloća<sup>1</sup>, Petr V. Konarev<sup>2</sup>, John Eccleston<sup>1</sup>, Ian A. Taylor<sup>1</sup>, Dmitri I. Svergun<sup>2</sup>,  
Annalisa Pastore<sup>1\*</sup>

<sup>1</sup>National Institute for Medical Research, The Ridgeway, London NW7 1AA, UK

<sup>2</sup>EMBL, Hamburg, Germany & Institute of Crystallography of RAS, Moscow, Russia

\*To whom correspondence should be addressed

**Running title:** Biophysical studies of FXR proteins

**Key words:** Biophysics, fragile X-related proteins, fragile X mental retardation; small angle  
X-ray scattering, structure

Accepted Manuscript

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20082197

## Abstract

Fragile X related proteins form a family implicated in RNA metabolism. Their sequence is composed of conserved N-terminal and central regions which contain Tudor and KH domains and of a divergent C-terminus with motifs rich in Arg and Gly. The most interesting member of the family is probably FMRP, since absence or mutation of this protein in human causes fragile X syndrome, the most common cause of inherited mental retardation. Understanding the structural properties of FMRP is essential for correlating it to its functions. The structures of isolated domains of FMRP have been reported, but nothing is yet known about the spatial arrangement of the different modules, partly because of difficulties in producing both the full-length protein and its multi-domain fragments in quantities, purities and monodispersity amenable for structural studies. Here, we describe how we have produced overlapping recombinant fragments of human FMRP and its paralogues which encompass the evolutionary conserved region. We have studied their behaviour in solution by complementary biochemical and biophysical techniques, identified the regions which promote self-association and determined their overall 3D shape. Our study paves the way to further studies and rationalises the existing knowledge on the self-association properties of these proteins.

## Abbreviations

AUC: analytical ultracentrifugation; CD: circular dichroism; DLS: dynamic light scattering; EOM: ensemble optimization method; ES: electron spray; FMRP: fragile X mental retardation protein; FXR proteins: fragile X related proteins; FXTAS: Fragile X associated tremor ataxia syndrome; HLH: helix-loop-helix; HSQC: hetero single quantum coherence; KH: K-homology; SEC-MALLS: size exclusion chromatography-multi angle laser light scattering; MALDI-TOF: mass assisted laser desorption ionization-time of flight; NDF: N-terminal domain of FMRP; NES: nuclear export signal; NLS: nuclear localization signal; NMR: nuclear magnetic resonance; SAXS: small angle X-ray scattering; SDS-PAGE: sodium dodecyl sulphate-poly-acrylamide gel electrophoresis; SEC: size exclusion chromatography; TROSY: transverse relaxation optimized spectroscopy; Trx: thioredoxin.

## Introduction

Fragile X mental retardation syndrome is a neurodevelopmental X-linked disorder which accounts for mental retardation, autistic behaviour and cranium-facial malformations in 1 over 4000 males and in 1 over 8000 females [1]. It is caused by silencing of the *fmr1* gene with consequent lack of expression of the gene product fragile X mental retardation protein (FMRP). There are also rare but well documented cases of fragile X patients who express truncated [2] or mutated [3] versions of FMRP. Mutations in the 5' un-translated region of *fmr1* gene can lead to fragile X associated tremor ataxia syndrome (FXTAS), a late onset progressive neurodegenerative disorder in which level of *fmr1* mRNA is drastically increased [4, 5]. Two fragile X related (FXR) autosomal genes, *fxr1* and *fxr2*, have been identified [6]. They encode proteins with high sequence homology but distinct cellular localization and unknown function. Orthologous genes are present in other vertebrates and in some invertebrates, forming all together the fragile X related (FXR) protein family [7].

FMRP is an RNA binding protein known to recognize various coding and non coding RNAs [8-10]. FMRP homodimerizes and interacts with FXR1P and FXR2P [6, 11] as well as with other cytoplasmic and nuclear proteins [12]. Accumulating evidence suggests that FMRP is involved in gene expression at a transcriptional and post-transcriptional level, with an active role in regulating local protein synthesis in dendritic synapses [13, 14]. Under stress conditions, high levels of transfected FMRP in STEK *fmr1* KO cell lines induce formation of granules by trapping mRNAs into repressed mRNP particles [15].

FMRP, like all the proteins of the FXR family, is highly modular and consists of the N-terminal Domain of FMRP (NDF) which contains two repeats of Tudor motifs, a linker with a potential helix-loop-helix motif (HLH) which contains also a nuclear localization signal (NLS), a central region with two KH motifs (KH1 and KH2) and a nuclear export signal (NES), and of a C-terminus which contains RGG motifs [16, 17]. Each of these regions is able to bind RNA individually [18-21]. While the amino terminal and central portions of FXR proteins are highly conserved, a significant variability is observed in the C-terminus [6].

The three dimensional structures of isolated domains of FMRP have been reported. The two Tudor repeats of NDF, each made of five  $\beta$ -strands, have been shown to pack against each other and to be both needed to form an independent folded unit [22]. The KH1 domain [23] adopts an  $\alpha\beta$  fold, characteristic of KH domains of type I [24]. The presence of a 66 residue long FMRP-specific loop in KH2 has probably delayed its structural determination until the structure of a construct containing KH1 and KH2 in tandem was solved by removing the loop [25]. Besides these studies, very little is presently known about the way the individual modules are assembled in space. This information is nevertheless the prerequisite for understanding how FMRP and its homologues may recognise their protein and RNA partners. A possible reason for this lack of structural information on the full-length protein and/or on constructs containing more than 1-2 domains could be the tendency to aggregate and/or to degrade of FMRP and its orthologues. Difficulties in production of full length recombinant human FMRP have for instance been reported both in bacterial expression hosts [17] and in *in vitro* transcription-translation systems [26]. The problem may become particularly severe when protein production needs to be in the quantities, concentrations and purities necessary for structural studies.

Here, we describe a characterization based on complementary biochemical and biophysical techniques of multi-domain regions of FXR human proteins. We have studied in parallel FMRP, FXR1 and FXR2 since the high degree of homology (70-80% identity and 80-90% similarity) between these proteins allows us to extend results obtained on any of them to the whole family. Identification of even subtle differences between them could in turn be important for understanding their functional specificity, currently still unclear. After careful analysis of the domain boundaries and extensive studies of the aggregation properties

of different protein fragments, we were able to produce suitable quantities of three overlapping multi-domain constructs and to characterise their stoichiometries and behaviour. We show that different domains of the FXR proteins contribute to self-association in different concentration ranges. We propose that the self-association properties observed in solution, which are similar to those reported for the full-length proteins, may be related to the ability of these proteins to homo and hetero-assemble *in vivo* and to their tendency of forming granules [15, 27, 28]. We also establish the spatial assembly of the evolutionary conserved region. Our work paves the way to further determination of the ultra-structure of FXR proteins in isolation and in complex with RNA and other protein partners.

## Experimental procedures

### Bioinformatic analysis

Sequence analysis was carried out using the following web-tools: ProtParam, (<http://www.expasy.ch/tools/protparam.html>); for calculating physical and chemical parameters; FoldIndex (<http://biportal.weizmann.ac.il/fldbin/findex>); for evaluating the probability of a given protein sequence to be intrinsically unfolded; PeptideCutter (<http://www.expasy.org/tools/peptidecutter>); for detecting putative cleavage sites for enzymes and chemicals; and PESTFind (<http://www.at.embnat.org/toolbox/pestfind>); for searching potential proteolytic cleavage sites.

### Protein expression and purification

The cDNA clones encoding human FMRP (Swiss Prot accession number Q06787), FXR1P (P51114), and FXR2P (P51116) were sub-cloned to obtain different recombinant protein fragments. PCR amplicons encoding FMRP Nt-KH1 and FXR1P Nt-KH1 were cloned into a modified pET-24 plasmid (Novagen) with *NcoI* and *NotI* sites and containing a N-terminal His<sub>6</sub>-thioredoxin (Trx)-tag with a cleavage sequence for tobacco etch virus (TEV) protease ENLYFQ\*GA (the asterisk indicates the cleavage site; the extra residues left after cleavage are underlined). A differently modified pET-24 vector with a non removable C-terminal His<sub>6</sub>-tag was used for production of FXR1P Nt-KH2.

Protein expression in *E. coli* BL21 (DE3) STAR strain was induced with isopropyl β-D-thiogalactopyranoside (1 mM) for 5 h at 25°C. Large scale purification of the recombinant proteins was performed as follows. Cells were harvested, resuspended in 20 mM Tris-HCl (pH 8), 150 mM NaCl, 10 mM imidazole, 0.2% Igepal CA-630 (Sigma-Aldrich), 1 mM phenylmethylsulfonyl fluoride, supplemented with the Complete EDTA-free protease inhibitors mix (Roche), and lysed by freezing-thawing and ultrasound sonication. The soluble fraction of centrifugation-clarified cell lysate was applied onto nickel-nitrilotriacetic acid (Ni-NTA) metal affinity chromatography matrix (Qiagen) from which the recombinant His<sub>6</sub>-Trx or C-terminal His<sub>6</sub>-tagged proteins were eluted with 50 mM Tris-HCl (pH 8.0), 300 mM imidazole, 2 mM β-mercaptoethanol and dialysed into 50 mM Tris-HCl (pH 8.0), 1 mM DTT, 0.5 mM EDTA. His<sub>6</sub>-Trx tag was cleaved-off by digestion with TEV protease at 23°C. The proteins were further purified by additional passage on Ni-NTA, preparative gel filtration (HiLoad Superdex 200 HR 16/60), anion exchange chromatography (MonoQ HR 5/5), and analytical size exclusion chromatography (SEC) (Superdex 200 HR 10/30 GL or Superdex 75 HR 10/30 GL). Protein identity was confirmed by electro-spray (ES) and mass assisted laser desorption ionization-time of flight (MALDI-TOF) spectrometry and N-terminal protein sequencing. Protein purity was assessed by Coomassie blue-stained sodium dodecyl sulphate poly-acrylamide gel electrophoresis (SDS-PAGE) and mass spectrometry. Protein concentrations were estimated by measurements of absorbance at 280 nm using theoretical extinction coefficients calculated by ProtParam. The proteins were concentrated using an ultra-filtration device (VIVASpin, Millipore).



$^{15}\text{N}$  labelled proteins were prepared as described above but by growing the cells in M9 minimal media supplemented with  $^{15}\text{N}$  labelled ammonium sulphate.

#### *Spectroscopic measurements*

Circular dichroism (CD) spectra of the purified recombinant proteins were recorded at 20°C using a Jasco J-715 spectropolarimeter (JASCO Co., Tokyo, Japan), over a wavelength range from 260 to 195 nm in 1 mm path length Quartz cuvettes (Hellma GmbH & Co. KG, Müllheim, Germany). The purified recombinant peptides were used at protein concentrations of 5  $\mu\text{M}$  in a buffer containing 50 mM Tris-HCl (pH 8.0), 1 mM  $\beta$ -mercaptoethanol.

Nuclear magnetic resonance (NMR) 1D and 2D  $\{^1\text{H}, ^{15}\text{N}\}$ -HSQC and TROSY spectra were recorded at 25°C and 800 MHz using a VARIAN INOVA (VARIAN, Palo Alto, CA) spectrometer equipped with a cryo-probe. The range of protein concentrations explored was from 80 to 230  $\mu\text{M}$ . Spectra were recorded for proteins in 50 mM Tris-HCl (pH 7.5), 5 mM TCEP either in the absence or in the presence of 150 mM NaCl.

#### *Experiments to test monodispersity*

Analytical SEC was performed at 23°C using either Superdex 200 10/30 GL or Superdex 75 10/30 GL columns equilibrated with 50 mM Tris-HCl (pH 8.0) and either 5 mM TCEP or 10 mM  $\beta$ -mercaptoethanol. We tested samples in a concentration range between 10-900  $\mu\text{M}$ . The injection volume was 100  $\mu\text{l}$  and the flow rate 0.5 ml/min.

Dynamic light scattering (DLS) measurements were performed at 20°C or 10°C in a concentration range of 2.5  $\mu\text{M}$  -1.4 mM for the two Nt-KH1 constructs and 2.5-114  $\mu\text{M}$  for FXR1P Nt-KH2, using a Viscotek model 802 instrument (Viscotek Corporation, Houston, TX). Results were analyzed by integrated Viscotek OmniSIZE software. All DLS measurements were carried out without previous dilution or filtration.

Molecular weights and molecular weight distributions were determined using on line multi-angle laser light scattering coupled with size exclusion chromatography (SEC-MALLS) [29]. Samples were applied to a Superdex 200 10/300 GL column equilibrated in 50 mM Tris-HCl (pH 8.0), 5 mM TCEP at a flow rate of 0.5 ml/min. The column was mounted on a Jasco HPLC controlled by the Chrompass software package. The scattered light intensity of the column eluent was recorded at sixteen angles using a DAWN-HELEOS laser photometer (Wyatt Technology Corp., Santa Barbara, CA). The loaded concentrations were 230  $\mu\text{M}$  and 140  $\mu\text{M}$  for FMRP and FXR1P Nt-KH1 respectively. Two concentrations (33  $\mu\text{M}$  and 70  $\mu\text{M}$ ) were tested for FXR1P Nt-KH2. The concentration of the eluent was determined from the refractive index change ( $\text{dn}/\text{dc} = 0.186$ ) using an OPTILAB-rEX differential refractometer equipped with a Peltier temperature-regulated flow cell, maintained at 25°C (Wyatt Technology Corp., Santa Barbara, CA). The wavelength of the laser in the DAWN-HELEOS and the light source in the OPTILAB-rEX was 658 nm. The weight-averaged molecular weight of material contained in chromatographic peaks was determined using the ASTRA software version 5.1 (Wyatt Technology Corp., Santa Barbara, CA). Briefly, at 1 s intervals throughout the elution of peaks the scattered light intensities together with the corresponding protein concentrations were used to construct Debye plots ( $K_c/R_\theta$  versus  $\sin^2(\theta/2)$ ). The weight-averaged molecular weight was then calculated at each point in the chromatogram from the intercept of an individual plot. An overall average molecular weight and polydispersity term for each species was calculated by combining and averaging the data from the individual measurements.

Analytical ultracentrifuge (AUC) equilibrium measurements were carried out on an Optima XL-A (Beckman Coulter Inc., Fullerton, CA) instrument equipped with absorbance optics using six sector cells and scanning at 280 nm. Each protein sample was examined at three different protein concentrations (7, 14 and 21  $\mu\text{M}$ ) which were obtained by diluting the protein eluted from Superdex 200 10/30 GL column into the same buffer to obtain an

absorbance at 280 nm of 0.2, 0.4 and 0.6. The experiments were performed at 20°C. The solutions were allowed to reach equilibrium at 18,000 rpm, 22,500 rpm, 25,000 rpm, and 40,000 rpm. The data were globally fit using Beckman/Origin software. Sedimentation velocity measurements were carried out with the same instrument but using two-sector cells. Protein solutions with an absorbance of ~1.0, corresponding to 30  $\mu$ M (FMRP Nt-KH1) and 33  $\mu$ M (FXR1P Nt-KH2) were centrifuged at 20°C and 40,000 rpm. Two hundred scans were recorded at 5 min intervals. The data were analyzed as a distribution of sedimentation coefficients using SEDFIT [30].

### SAXS experiments

Synchrotron radiation X-ray scattering (SAXS) data were collected on the EMBL X33 beamline at the DORIS III storage ring, DESY, Hamburg [31]. Protein solutions of FXR1P (Nt-KH1 and Nt-KH2) and FMRP (Nt-KH1) in 50 mM Tris-HCl (pH 8.0), 5 mM TCEP were measured for solute concentrations of 12-250  $\mu$ M, 12-126  $\mu$ M and 18-420  $\mu$ M, respectively. The experiments were carried out at 10°C. A MAR345 image plate at sample-detector distance 2.7 m and wavelength  $\lambda = 0.15$  nm, covering the momentum transfer range  $0.12 < s < 4.9$  nm<sup>-1</sup> ( $s = 4\pi \sin(\theta)/\lambda$  where  $2\theta$  is the scattering angle), was used. For each measurement, two 2 min exposures were compared to check for radiation damage. No radiation effects were observed. The data were averaged after normalization to the intensity of the incident beam. The scattering of the buffer was subtracted and the difference data were extrapolated to zero solute concentration following standard procedures.

Data manipulation was performed using the program package PRIMUS [32]. The forward scattering  $I(0)$  and the radius of gyration  $R_g$  were evaluated using the Guinier approximation [33] assuming that at very small angles ( $s < 1.3/R_g$ ) the intensity is represented as  $I(s) = I(0) \exp(-1/3(R_g s)^2)$ . These parameters were also computed from the entire scattering patterns using the program GNOM [34], providing the distance distribution functions  $p(r)$  and the maximum particle dimensions  $D_{\max}$ . The MWs were evaluated by comparison of the forward scattering with that from reference solutions of bovine serum albumin (66 kDa MW). The excluded volume of the hydrated particle (the Porod volume) was computed as previously described [35]:

$$V = 2\pi^2 I(0) / \int_{-\infty}^{\infty} s^2 I_{\text{exp}}(s) ds \quad (1)$$

Low resolution *ab initio* models of the three constructs were generated by the program DAMMIN [36] which represents the protein by an assembly of densely packed beads. Simulated annealing was employed to build a compact interconnected configuration of beads inside a sphere with the diameter  $D_{\max}$  that fits the experimental data  $I_{\text{exp}}(s)$  to minimize the discrepancy:

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[ \frac{I_{\text{exp}}(s_j) - cI_{\text{calc}}(s_j)}{\sigma(s_j)} \right]^2 \quad (2)$$

where  $N$  is the number of experimental points,  $c$  is a scaling factor and  $I_{\text{calc}}(s_j)$  and  $\sigma(s_j)$  are the calculated intensity and the experimental error at the momentum transfer  $s_j$ , respectively.

An alternative model of FXR1P Nt-KH1 was reconstructed by molecular modelling using the NMR structure of FMRP NDF (PDB entry 2bkd) and the crystal structure of FMRP KH1 (PDB entry 2qnd). The scattering from these models was calculated using the program CRY SOL [37]. The HLH motif (residues 135-217) for which we have no structural information was represented by an interconnected chain of dummy residues [38]. A simulated annealing protocol implemented in the program BUNCH [39] was employed to generate native-like conformations without steric clashes. Possible flexibility of the HLH motif was assessed by the ensemble optimization method (EOM) [40] which allows for coexistence of different conformations contributing to the experimental scattering pattern.

These conformers were selected by a genetic algorithm from a pool containing  $10^5$  randomly generated models. Genetic algorithms were employed to find the subsets of the conformers which fit the experimental data best. The obtained subsets were analysed to yield the  $R_g$  distributions in the optimum ensembles. A similar approach was applied for the FMRP Nt-KH1 construct which has the same length as FXR1P Nt-KH1 and high sequence identity (83%) and similarity (93%).

#### *Protein interaction in vitro assay*

A His-tag pull-down assay was performed to check for potential interaction of FMRP Nt-KH1 and FXR1P Nt-KH2-His<sub>6</sub>. Pure recombinant FXR1P Nt-KH2-His<sub>6</sub> was immobilized on Ni-NTA-agarose beads, extensively washed with equilibrating buffer 20 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 100 mM NaCl, and incubated with purified FMRP Nt-KH1 at 4°C for 1 h. The matrix was washed extensively first with the same buffer and successively increasing the NaCl concentration to 300 mM. The bound proteins were eluted with 0.3 M imidazole in equilibrating buffer and visualized on SDS-PAGE gel by Coomassie blue staining. Both FXR1P Nt-KH2-His<sub>6</sub> and FMRP Nt-KH1 were used at concentrations of 10  $\mu$ M.

## **Results**

### ***Identification of potential degradation hot spots in FXR proteins***

Obtaining properly folded homogenous protein samples is the prerequisite for structural studies. We analyzed the amino acid sequences to understand the reasons of poor solubility and stability of recombinant FMRP and its homologues. We used FoldIndex software for predictions of natively unfolded sequences [41]. This suggests that the C-termini of FXR-proteins (ca. from residue 430 to 632) have a high likelihood to be disordered. This is in agreement with experimental observations [20, 42] and with low sequence-conservation of this region among FXR-proteins [6]. A second natively unfolded region is present only in FMRP and corresponds to residues 321-372 within a loop of KH2 encoded by exons 11 and 12. In eukaryotes, natively unfolded regions are often involved in transient interactions and/or post-translational modification sites. In prokaryotic expression hosts, the absence of interaction partners together with potential local misfolding is likely to render them susceptible to protease degradation.

The amino acid sequences of FMRP, FXR1P, and FXR2P were also examined for the presence of potential PEST sequences. These are hydrophilic stretches with high content of prolines, glutamates, serines, and threonines, which target proteins for ubiquitin-mediated degradation in eukaryotic cells. These sequences could in principle contribute to protein degradation also in *E. coli*, as PEST mediated degradation has been reported also in *L. monocytogenes* [43]. We identified potential PEST sequences in the C-termini of all FXR proteins (see **Figure S1 and Table S1 of Suppl. Mat.**). They are more numerous in FXR2P.

This analysis may explain some of the difficulties encountered in producing full-length FXR proteins. We thus decided to concentrate our structural characterization on the conserved N-terminus for which most interactions have been reported (for review see [12]).

### ***Production of multi-domain recombinant fragments***

With the ultimate attempt to fully characterize the FXR proteins, we systematically produced overlapping constructs for all three human homologues covering two or more tandem domains (**Figure 1**). The construct boundaries were chosen according to the specific motif consensus sequences and to the multiple alignment of the FXR family (**Figure S1 of Suppl. Mat.**). All expressed constructs resulted in the soluble fraction but had very different behaviours towards degradation and solubility.

Nt-NES, the longest construct which covers the whole conserved region, both from FMRP and FXR2P, though well expressed and soluble in different *E. coli* strains, was



invariably degraded. Degradation of Nt-NES from FMRP was so severe that the degradation fragments could not even be separated and analyzed. The presence of severe degradation already at the level of the cellular lysate indicates that it should occur intracellularly, prior to protein isolation from bacterial cells. On the contrary, Nt-NES of FXR2P could be purified but only together with a degraded species which became dominant within a few days. MALDI-TOF analysis showed that degradation of this construct is caused by proteolytic cleavage by a thermolysin-like metallo-protease between residues Tyr288 and Leu289 (that is near the end of KH1) and, to a lesser extent, by a trypsin-like protease between residues Arg255 and Lys256 (i.e. in the middle of KH1). Addition of specific protease inhibitors diminished but did not abolish degradation to the 1-288 fragment. The two observed degradation sites were identified as poor PEST sequences (**Table S1**).

The corresponding Nt-NES region from FXR1P could be obtained as a soluble and stable full-length protein. However, the construct proved to be anyway unsuitable for structural characterization because of concentration- and salt-dependent aggregation and precipitation (data not shown).

The longest stable recombinant fragment (not degrading and not precipitating) was Nt-KH2 of FXR1P (**Figure S2 of Suppl. Mat.**). The corresponding constructs from FMRP and FXR2P could not be produced in a stable form since their degradation was too severe. We could however produce without detectable degradation the shorter Nt-KH1 fragment from both proteins, which includes only the region up to the end of KH1. This strongly suggests that in recombinant FMRP and FXR2P, either KH2 is not properly folded or that the interface between KH1 and KH2 is flexible and constitutes a region easily attacked by proteases.

A construct  $\Delta$ NDF-NES (132-449) of FMRP, which starts after NDF, degraded completely shortly after purification. The equivalent constructs from FXR1 and FXR2 were stable towards degradation but precipitated visibly and quantitatively upon concentration.

The constructs from FXR1P are overall more stable and better behaving than the corresponding fragments of FMRP or FXR2P.

This lengthy and labour intensive work led to the production of four well behaving stable long constructs, Nt-KH2 from FXR1P and Nt-KH1 from all three proteins. We concentrated on three of them for further studies, choosing the unique Nt-KH2 from FXR1P and Nt-KH1 from FMRP and from the better behaving FXR1P. We dropped for the time being Nt-KH1 of FXR2P since in all preliminary studies it behaved very similarly to FXR1P.

### ***Spectroscopic analysis of FXR multidomain constructs***

The three well-behaving fragments were characterised by CD and NMR techniques to ensure that they are correctly folded. The far-UV CD spectra of FXR1 Nt-KH2 and Nt-KH1 and of FMRP Nt-KH1 are typical of well folded proteins with features of a mixed  $\alpha\beta$  fold, as we expect from what we know about the structures of the Tudor and KH domains which must contribute to the signal (**Figure 2A**). The secondary structure content as estimated from the spectra is, as expected, very similar for the two Nt-KH1 constructs (ca. 21%  $\alpha$ , 27%  $\beta$ , 20% turn and 32% random) and increases for FXR1P Nt-KH2 (28%  $\alpha$ , 22%  $\beta$ , 20% turn and 30% random). This increase is roughly consistent with the increment expected by assuming the secondary structure content observed in KH1 $\Delta$ KH2.

NMR  $\{^1\text{H}, ^{15}\text{N}\}$  hetero-correlated HSQC spectra recorded at 25°C and 800 MHz of the same constructs at different concentrations show interesting features. The spectrum of FXR1P Nt-KH2 is always too broad in the whole range of concentrations explored (30-230  $\mu\text{M}$ ), also in TROSY experiments (data not shown). The spectra of Nt-KH1 from FXR1 and FMRP are similarly broad when in the presence of 150 mM NaCl. They are instead relatively sharp at protein concentrations as high as 230  $\mu\text{M}$  when collected from samples with virtually no salt other than the buffer (**Figure 2B**). This is well consistent with previous

observations that both the NDF and KH1 give better spectra at low ionic strengths [22, 23]. Interestingly, the sharper resonances observed in the two Nt-KH1 constructs belong to KH1, as it can be seen from comparing the spectra with those of the isolated KH1 domain [23]. Resonances from NDF (as for instance the indole protons of the three only tryptophans present in the sequence) can also be observed but only when lowering significantly the vertical scale of the 2D plot. These results suggest that the linker or part of it between NDF and KH1 (HLH) is flexible, allowing KH1 to move semi-independently from the rest of the molecule.

### ***Assessment of the degree of self-association of the samples***

To test systematically the degree of monodispersity of the three stable constructs in a wide range of protein concentrations, we investigated their properties by different but complementary techniques. The measurements were carried out in the absence of salt since FXR1P Nt-KH2 precipitates already at very low concentrations of either NaCl or KCl. The Nt-KH1 constructs behaved very similarly at three different concentrations of NaCl (0, 50, or 150 mM NaCl).

A preliminary qualitative analysis by SEC suggested that both Nt-KH1 constructs elute from the column as monomeric proteins at micromolar concentrations (data not shown). Under similar conditions, Nt-KH2 elutes instead at a MW compatible with a mixture of monomer and dimer. In order to characterise these solution molecular weights in more detail SEC-MALLS detection was employed, a technique that is able to determine MW, stoichiometry and organization of assemblies present in solution [37]. The SEC-MALLS data demonstrate that the measured solution molecular weights of FMRP-Nt-KH1 ( $37,100 \text{ Da} \pm 0.6\%$ ), and FX1RP-Nt-KH1 ( $34,800 \text{ Da} \pm 0.6\%$ ) are comparable with monomeric MWs. Moreover, they display negligible concentration dependency, at least up to concentrations of  $\sim 150 \mu\text{M}$  (**Figure 3A,B**). SEC-MALLS analysis of FXR1P-Nt-KH2 reveals that inclusion of the KH2 domain results in significant concentration dependency of the weight-averaged molecular weight of the protein (**Figure 3C**). In this case, the weight-averaged MW of a sample with an average elution peak concentration of  $2.5 \mu\text{M}$  is  $45,000 \text{ Da} \pm 0.8\%$  i.e. only marginally bigger than the value expected for the monomer ( $42,000 \text{ Da}$ ). At higher protein concentration ( $6.6 \mu\text{M}$ ) a shift in the peak to an earlier eluting position is apparent and the observed MW increases significantly up to  $49,000 \pm 0.6\% \text{ Da}$ , indicating self-associative behaviour. On the simple assumption of a monomer/dimer equilibrium, the weight-averaged MW of  $49,000 \text{ Da}$  determined at  $6.6 \mu\text{M}$  equates to a sample composition of  $\sim 10\%/90\%$  dimer/monomer.

This behaviour was further confirmed by equilibrium and sedimentation velocity AUC performed at three different molar concentrations (in the range  $7\text{--}21 \mu\text{M}$ ). The equilibrium AUC results could be fit well assuming the presence of a monomeric species only for the Nt-KH1 samples (i.e. MWs of  $32 \pm 1 \text{ Da}$  and  $29 \pm 2 \text{ Da}$  for the FMRP and FX1RP constructs respectively which should be compared with their theoretical MWs of  $32,349 \text{ Da}$  and  $31,570 \text{ Da}$ ) (**Figure 4A**) but not for FXR1P Nt-KH2, for which the curves could not be fit at any concentration, in agreement with the presence of a mixture of species. When analysed by sedimentation velocity AUC, both Nt-KH1 constructs sediment with coefficients of  $2.5 \text{ Svedberg (S)}$ , which correspond to MWs around  $31 \text{ kDa}$  well in agreement with the values expected for the monomer. At the same concentration ( $33 \mu\text{M}$ ), FXR1P Nt-KH2 sediments as an unresolved mixture of two components with MW values around  $40 \text{ kDa}$  and  $80 \text{ kDa}$  and with sedimentation coefficients of  $3 \text{ S}$  and  $\sim 4.5 \text{ S}$  respectively (**Figure 4B**).

Finally, we used DLS to explore the construct behaviour over a wider range of concentrations ( $5 \mu\text{M}\text{--}1.4 \text{ mM}$ ). Both Nt-KH1 constructs were monomeric up to  $\sim 150 \mu\text{M}$ . At higher concentrations, we observed an increase of the hydrodynamic radii (from  $\sim 3$  to  $\sim 4 \text{ nm}$ ) towards values consistent with monomer/dimer mixtures (**Figure 5A**). This indicates

that self-association may occur also for these samples at sufficiently elevated concentrations. Nt-KH2 of FXR1P behaves as a monomeric species only up to  $\sim 5$ – $10\ \mu\text{M}$ . At higher concentrations, we observe the progressive increase of the hydration radius up to a value indicative of large MW aggregates (**Figure 5B**). Concomitantly, the polydispersity increases.

Taken together, these results indicate that the FXR proteins have a strong tendency to self-associate via dimer formation through the contribution of different regions of the molecules: the presence of KH2 seems to be sufficient to promote aggregation already at relatively low concentrations suggesting a pivotal role of this domain in the process.

### Shape reconstruction by SAXS

Finally, we used SAXS experiments both to determine the shape of our constructs and to gain information about the HLH domain (residues 135–217), for which no structure is available. For FXR1P Nt-KH1, we obtained a MW of  $34 \pm 4\ \text{kDa}$  when studying it at  $12\ \mu\text{M}$ – $35\ \mu\text{M}$ . The excluded (Porod) volume of the particle in solution, a parameter which for globular proteins is expected to be ca. twice the MW in kDa, is  $57 \pm 5\ \text{nm}^3$  (**Figure 6A and Table 1**). This further confirms that the protein is a monomer at the tested concentrations. The experimental values of the gyration radius ( $R_g$ ) and of the maximal size of the particle ( $D_{\text{max}}$ ) ( $3.77 \pm 0.04\ \text{nm}$  and  $13.0 \pm 0.5\ \text{nm}$ , respectively) point to an elongated structure. The long-tailed shape of the distance distribution function  $p(r)$  for FXR1P Nt-KH1 is also consistent with an elongated shape of the protein (**Figure S3**). A typical low resolution shape of FXR1P Nt-KH1 was reconstructed *ab initio* by DAMMIN [36] (**Figure 6B**). It fits the experimental data with a discrepancy  $\chi_{\text{ab}}$  of 1.48. The model reveals three distinct regions: a longer domain with an overall shape similar to that of the N-terminal NDF domain, a middle one presumably accounting for the 81 residues of the HLH motif (sequence 135–217), and a third one which likely corresponds to the KH1 module. At higher concentration ( $250\ \mu\text{M}$ ) FXR1P Nt-KH1 sample revealed unspecific aggregation.

The overall parameters of FMRP Nt-KH1 at  $18\ \mu\text{M}$ – $40\ \mu\text{M}$  are close to those of FXR1P Nt-KH1 (**Table 1**). The FMRP Nt-KH1 models are slightly more extended than those of FXR1P Nt-KH1, in agreement with the slightly larger values of  $R_g$  and  $D_{\text{max}}$ , but the overall shape is similar (**Figure 6C**). At higher concentrations of FMRP Nt-KH1 the presence of larger aggregates was observed.

Nt-KH2 samples measured in a concentration range  $12$ – $126\ \mu\text{M}$  revealed strong concentration dependence in the scattering patterns already at low concentrations (**Table 1**). The higher concentration samples ( $48\ \mu\text{M}$  and  $126\ \mu\text{M}$ ) clearly show high MW and  $R_g$  values consistent with the presence of aggregates, possibly tetramers and even larger. The parameters measured at  $12\ \mu\text{M}$  and  $24\ \mu\text{M}$  reasonably correspond to a monomer to dimer mixture. Indeed, the experimental MW and an excluded volume calculated for FXR1P Nt-KH2 at these concentrations (**Table 1**) have intermediate values between the ones of the monomers and the dimers (the theoretical MW of the Nt-KH2 monomer is  $42\ \text{kDa}$ ). The increase of overall parameters and excluded volume at  $24\ \mu\text{M}$  (compared to  $12\ \mu\text{M}$ ) points out to the progressive shift of the monomer/dimer equilibrium towards the dimers. The elongated overall shape corresponding to the scattering data at a low concentration (**Figure S3B**) resembles the shapes obtained for the shorter constructs (**Figure 6B,C**). Since all the measured solutions of Nt-KH2 appeared to be mixtures of different oligomers, no attempts of more detailed modelling were made. However, the very presence of Nt-KH2 dimers even at low solute concentrations confirms that the dimerization of FXR1P Nt-KH2 should occur through KH2 domain.

To obtain more detailed models of FXR1P Nt-KH1 and FMRP Nt-KH1, we used the high resolution structures of NDF and of the KH1 domain and treated them as rigid bodies. The HLH region was represented as a flexible chain of dummy residues. For these two constructs, multiple runs of BUNCH [39] starting from random initial conformations yielded

to a similar overall shape and a good data fitting with the *ab initio* shapes ( $\chi$  values 1.52, 1.75 for FXR1P Nt-KH1 and FMRP Nt-KH1 respectively).

Taken together, these results suggest that: the average distances between the NDF and KH1 remain similar in both constructs. Our data also strongly suggest the presence of some flexibility in the HLH region, as shown by using the EOM software [40], which accounts for multiple configurations. Comparison the  $R_g$  distribution derived from the optimized ensembles with that obtained from randomly generated models (**Figure S3C**). The former distribution is nearly as broad as the latter one, suggesting that the HLH motif contains flexible regions in agreement with the NMR data.

### ***Nt-KH1 does not mediate heterodimerization of FMRP and FXR1P***

Ascertained that recombinant Nt-KH1 of either FMRP or of FXR1P is not able to homodimerize at low concentration, we investigated if Nt-KH1 could mediate heterodimerization of the two proteins *in vitro*. Pull down assays were performed using FXR1P Nt-KH2 and FMRP Nt-KH1 at concentrations in which they are monomeric (10  $\mu$ M). We did not observe interaction (data not shown).

### **Discussion**

We have produced and extensively characterized overlapping regions of human members of the FXR family spanning the conserved N-terminus. We focused on this region in view to future structural studies and because it is crucial for interaction with several cellular partners [12]. As in previous reports, we encountered difficulties, mostly related to protein degradation and aggregation, in producing multiple domain fragments in quantities suitable for biophysical and structural studies. Interestingly, the degradation patterns observed for the Nt-NES constructs resemble previous observations on *in vitro* produced full length FXR proteins (Figures 5A and 7B in [6]. We suggest that a possible cause of degradation is the presence of PEST motifs along the sequences. A putative PEST motif was already identified in FMRP in the region which includes the phosphorylation site at Ser500 [44]. Chemical modification of this serine did not however make any difference for the stability of full-length FMRP transfected in murine fibroblasts.

We have explored a wide range of concentrations using different techniques to investigate the self-association process (**Figure 7**). Most of our constructs show tendency to self-associate in agreement with previous work both on shorter FMRP constructs [20] and on full-length recombinant murine FMRP produced in baculovirus [18]. While it is known that truncated versions of a protein may be prone to aggregation, the observation of similar properties for different fragments and the notion that also the full-length proteins have similar behaviour strongly suggests that what we observe *in vitro* may reflect cellular functions. FXR proteins are well known to homo-associate *in vivo* [6, 11]. In *Drosophila* neurons, the mobility of mRNA is controlled by FMRP in a concentration dependent manner [45]. Under stress conditions, high levels of transfected FMRP induce formation of granules in which mRNAs are trapped into repressed mRNP particles [15]. These properties could be controlled by the tendency of FXR proteins to self-associate and be modulated by the local concentration of FMRP and its cellular partners. Our observations could also be relevant for understanding the formation of intranuclear inclusions observed in neurons of patients with FXTAS [46], a misfolding disease associated with Fragile X that has features similar to those of Parkinson's disease. We propose that the behaviour we observe *in vitro* may be typical of what we could name 'complex-orphan proteins' that is proteins which exist in the cell as part of large molecular assemblies. When produced in isolation, they will reproduce an elevated tendency to self-associate.

An important result of our work is that we do not observe only protein dimerization but a more complex pattern of self-association which proceeds in a continuous way from the



monomer to large MW aggregates via formation of dimeric species. Self-association is promoted by different regions of the proteins which start to participate to the process in different ranges of concentrations. In a previous paper, we had described dimerization of shorter constructs spanning the N-terminus of FMRP [20]. In the context of longer constructs, KH2 seems to be the first motif to promote dimer formation and to provide seeding for further aggregation. This is in agreement with the notion that FMRP dimerization is abolished in a I304N mutant, which causes structural and functional destabilization of the KH2 fold [3, 23, 25, 47, 48]. A link between KH2 and dimerization is also suggested by the crystal structure of KH1KH2 $\Delta$  in which the dimer interface is formed only by KH2 [25]. More in general, KH domains of different proteins are known to mediate homodimerization *in vitro* [49].

Outside the conserved region explored here, it is likely that *in vivo* the elements downstream to KH2 also contribute to self-association. The NES region of FMRP, for instance, originally predicted as a coiled coil motif which could promote homo- and hetero-dimerization [26], contains three leucines which have decisive importance for protein localization and could participate to self-association. In our hands, the Nt-NES fragment of FXR1P has a strong tendency, higher than that observed for shorter constructs, to form unspecific aggregates, although further sequence analysis has not supported the presence of a standard dimerization motif in this region [50].

SAXS techniques allowed us to grasp information on the shape of our constructs. The gyration radii of all three constructs indicate an elongated shape consistent with an extended structure of the contiguous domains. The region less well defined in our analysis remains the HLH motif for which we do not have any direct structural information at high resolution. This region is potentially very important since it has been shown to contain the NLS signal of FMRP, although this could not be mapped at the level of single amino acids [16]. The flexibility observed by NMR and SAXS could be at variance with what would be expected from sequence analysis which does not identify intrinsically unfolded sequences in this region. A fascinating hypothesis could be that nuclear signalling is achieved by the induction of a conformational change of HLH mediated by interactions with either distal regions of the same molecule or another partner. We would therefore be observing in our constructs only one of the functionally relevant states. More work will be needed to test this possibility.

In conclusion, our work represents one step further towards understanding the structure of FMRP, thanks to the use of multidomain fragments of FXR proteins which were rigorously checked for structural integrity and monodispersity. Our findings imply that the determinants of FXR protein self-association are distributed along different regions of the molecules, rationalising only apparently contradictory previous reports. Conclusive evidence at atomic details of the molecular mechanism through which FXR homo- and hetero-association occur can now only come from structural studies of the full length protein appropriately post-translationally modified. We suggest that, although difficult, this pursuit might be achieved by focusing on FXR1P which is less prone to degradation and does not contain the cumbersome insertion present in KH2 of FMRP.

### Acknowledgment

We wish to thank Dr Steven Howell for mass spectrometry and for MALDI-TOF experiments, Dr Stephen R. Martin for help with circular dichroism spectroscopy, and the NIMR NMR centre and Dr Geoff Kelly in particular for technical support and help in collecting the NMR data. We are also grateful to Dr Human Rezaei for helpful scientific discussion.



## References

- 1 Sherman, S. L., Marsteller, F., Abramowitz, A. J., Scott, E., Leslie, M. and Bregman, J. (2002) Cognitive and behavioral performance among FMR1 high-repeat allele carriers surveyed from special education classes. *Am. J. Med. Genet.* **114**, 458-465
- 2 Lugenbeel, K. A., Peier, A. M., Carson, N. L., Chudley, A. E. and Nelson, D. L. (1995) Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat. Genet.* **10**, 483-485
- 3 De Boule, K., Verkerk, A. J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B. A. and Willems, P. J. (1993) A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet.* **3**, 31-35
- 4 Feng, Y., Zhang, F., Lokey, L. K., Chastain, J. L., Lakkis, L., Eberhart, D. and Warren, S. T. (1995) Translational suppression by trinucleotide repeat expansion at FMR1. *Science* **268**, 731-734
- 5 Hagerman, R. J., Leavitt, B. R., Farzin, F., Jacquemont, S., Greco, C. M., Brunberg, J. A., Tassone, F., Hessel, D., Harris, S. W., Zhang, L., Jardini, T., Gane, L. W., Ferranti, J., Ruiz, L., Leehey, M. A., Grigsby, J. and Hagerman, P. J. (2004) Fragile-X-associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation. *Am. J. of Hum. Gen.* **74**, 1051-1056
- 6 Zhang, Y., O'Connor, J. P., Siomi, M. C., Srinivasan, S., Dutra, A., Nussbaum, R. L. and Dreyfuss, G. (1995) The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J.* **14**, 5358-5366
- 7 Guduric-Fuchs, J., Mohrlen, F., Frohme, M. and Frank, U. (2004) A fragile X mental retardation-like gene in a cnidarian. *Gene* **343**, 231-238
- 8 Ashley, C. T., Jr., Wilkinson, K. D., Reines, D. and Warren, S. T. (1993) FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* **262**, 563-566
- 9 Darnell, J. C., Jensen, K. B., Jin, P., Brown, V., Warren, S. T. and Darnell, R. B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**, 489-499
- 10 Sung, Y. J., Conti, J., Currie, J. R., Brown, W. T. and Denman, R. B. (2000) RNAs that interact with the fragile X syndrome RNA binding protein FMRP. *Biochem. Biophys. Res Comm.* **275**, 973-980
- 11 Tamanini, F., Van Unen, L., Bakker, C., Sacchi, N., Galjaard, H., Oostra, B. A. and Hoogeveen, A. T. (1999) Oligomerization properties of fragile-X mental-retardation protein (FMRP) and the fragile-X-related proteins FXR1P and FXR2P. *Biochem. J.* **343 Pt 3**, 517-523
- 12 Bardoni, B., Davidovic, L., Bensaid, M. and Khandjian, E. W. (2006) The fragile X syndrome: exploring its molecular basis and seeking a treatment. *Expert Rev. Mol. Med.* **8**, 1-16
- 13 Bagni, C. and Greenough, W. T. (2005) From mRNP trafficking to spine dysmorphogenesis: the roots of fragile X syndrome. *Nature Rev.* **6**, 376-387
- 14 Garber, K., Smith, K. T., Reines, D. and Warren, S. T. (2006) Transcription, translation and fragile X syndrome. *Curr. Opin. Genet. Dev.* **16**, 270-275
- 15 Mazroui, R., Huot, M. E., Tremblay, S., Filion, C., Labelle, Y. and Khandjian, E. W. (2002) Trapping of messenger RNA by Fragile X Mental Retardation protein into cytoplasmic granules induces translation repression. *Hum. Mol. Genet.* **11**, 3007-3017
- 16 Eberhart, D. E., Malter, H. E., Feng, Y. and Warren, S. T. (1996) The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum. Mol. Genet.* **5**, 1083-1091
- 17 Siomi, H., Siomi, M. C., Nussbaum, R. L. and Dreyfuss, G. (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* **74**, 291-298

- 18 Brown, V., Small, K., Lakkis, L., Feng, Y., Gunter, C., Wilkinson, K. D. and Warren, S. T. (1998) Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein. *J. Biol. Chem.* **273**, 15521-15527
- 19 Schaeffer, C., Bardoni, B., Mandel, J. L., Ehresmann, B., Ehresmann, C. and Moine, H. (2001) The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J.* **20**, 4803-4813
- 20 Adinolfi, S., Ramos, A., Martin, S. R., Dal Piaz, F., Pucci, P., Bardoni, B., Mandel, J. L. and Pastore, A. (2003) The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. *Biochemistry* **42**, 10437-10444
- 21 Zalfa, F., Adinolfi, S., Napoli, I., Kuhn-Holsken, E., Urlaub, H., Achsel, T., Pastore, A. and Bagni, C. (2005) Fragile X mental retardation protein (FMRP) binds specifically to the brain cytoplasmic RNAs BC1/BC200 via a novel RNA-binding motif. *J. Biol. Chem.* **280**, 33403-33410
- 22 Ramos, A., Hollingworth, D., Adinolfi, S., Castets, M., Kelly, G., Frenkiel, T. A., Bardoni, B. and Pastore, A. (2006) The structure of the N-terminal domain of the fragile X mental retardation protein: a platform for protein-protein interaction. *Structure* **14**, 21-31
- 23 Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T. J., Nilges, M. and 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
- 24 Grishin, N. V. (2001) KH domain: one motif, two folds. *Nucl. Ac. Res.* **29**, 638-643
- 25 Valverde, R., Pozdnyakova, I., Kajander, T., Venkatraman, J. and Regan, L. (2007) Fragile X mental retardation syndrome: structure of the KH1-KH2 domains of fragile X mental retardation protein. *Structure* **15**, 1090-1098
- 26 Siomi, M. C., Zhang, Y., Siomi, H. and Dreyfuss, G. (1996) Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Mol. Cell. Biol.* **16**, 3825-3832
- 27 De Diego Otero, Y., Severijnen, L. A., van Cappellen, G., Schrier, M., Oostra, B. and Willemsen, R. (2002) Transport of fragile X mental retardation protein via granules in neurites of PC12 cells. *Mol. Cell. Biol.* **22**, 8332-8341
- 28 Dolzhanskaya, N., Merz, G., Aletta, J. M. and Denman, R. B. (2006) Methylation regulates the intracellular protein-protein and protein-RNA interactions of FMRP. *J. Cell. Sci.* **119**, 1933-1946
- 29 Wen, J., Arakawa, T., and Philo, J.S. (1996) Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Anal. Biochem.* **240**, 155-166
- 30 Schuck, P. and Rossmanith, P. (2000) Determination of the sedimentation coefficient distribution  $g^*(s)$  by least-squares boundary modeling. *Biopolymers* **54**, 328-341
- 31 Roessle, M. W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C., Svergun, D. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Cryst.* **40**, s190-s194
- 32 Konarev, P. V., Petoukhov, M. V., Volkov, V. V. and Svergun, D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J. Appl. Cryst.* **39**, 277-286
- 33 Guinier, A. (1939) La diffraction de rayons X aux tres petits angles. *Ann. Phys. (Paris)* **12**, 161-237
- 34 Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495-503
- 35 Porod, G. (1982) General theory. Small-angle X-ray scattering London, Academic Press. **edited by O. Glatter and O. Kratky.**, 17-51.

- 36 Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886
- 37 Svergun, D. I., C. Barberato, Koch, M. H. J. (1995) CRY SOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768-773
- 38 Svergun, D. I., Petoukhov, M. V. and Koch, M. H. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946-2953
- 39 Petoukhov, M. V. and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89**, 1237-1250
- 40 Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656-5664
- 41 Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics. *Prot. Sci.* **11**, 739-756
- 42 Ramos, A., Hollingworth, D. and Pastore, A. (2003) G-quartet-dependent recognition between the FMRP RGG box and RNA. *RNA* **9**, 1198-1207
- 43 Decatur, A. L. and Portnoy, D. A. (2000) A PEST-like sequence in listeriolysin O essential for *Listeria monocytogenes* pathogenicity. *Science* **290**, 992-995
- 44 Ceman, S., O'Donnell, W. T., Reed, M., Patton, S., Pohl, J. and Warren, S. T. (2003) Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum. Mol. Genet.* **12**, 3295-3305
- 45 Estes, P. S., O'Shea, M., Clasen, S. and Zarnescu, D. C. (2008) Fragile X protein controls the efficacy of mRNA transport in *Drosophila* neurons. *Mol. Cell. Neurosci.* **39**, 170-179
- 46 Greco, C. M., Hagerman, R. J., Tassone, F., Chudley, A. E., Del Bigio, M. R., Jacquemont, S., Leehey, M. and Hagerman, P. J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* **125**, 1760-1771
- 47 Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B. and Burley, S. K. (1999) Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure* **7**, 191-203
- 48 Ramos, A., Hollingworth, D. and Pastore, A. (2003) The role of a clinically important mutation in the fold and RNA-binding properties of KH motifs. *RNA* **9**, 293-298
- 49 Valverde, R., Edwards, L. and Regan, L. (2008) Structure and function of KH domains. *FEBS J.* **275**, 2712-2726
- 50 Pastore, A. (2005) A structural perspective to understand the cellular role of FMRP. *The Molecular Basis of Fragile X Syndrome*. editors **Y. J. Sung and R.B. Denman**, 29-41

## Figure legends

**Figure 1** – Choice and production of the multidomain constructs used in this paper. Above: Schematic representation of the architecture of FXR proteins. The different sequence motifs are indicated with different geometric shapes. Below: Summary of the constructs produced and their properties. N.P. stands for not produced.

**Figure 2**– Spectroscopic studies of FXR proteins. A) Far-UV CD spectra in molar ellipticity of the Nt-KH1 of FMRP (black continuous line) and FXR1P (grey dotted line) and of FXR1P Nt-KH2 (grey continuous line). The measurements were carried out at 20°C using 5  $\mu$ M concentrations. B) NMR HSQC spectra of Nt-KH1 from FXR1P (top) and FMRP (bottom). The spectra were recorded at 800 MHz and 25°C, using concentrations of 60  $\mu$ M (FXR1P) and 230  $\mu$ M (FMRP).

**Figure 3** – Solution molecular weight of FMRP Nt-KH1, FXR1P Nt-KH1 and FXR1P Nt-KH2 determined by SEC-MALLS. Panels **A** and **B** are the elution profiles produced by application of 100  $\mu$ l of 150  $\mu$ M FMRP Nt-KH1 and 75  $\mu$ M FXR1P Nt-KH1 to a Superdex 200 10/300 GL column. The blue and red lines are the chromatograms recorded by the differential refractometer and from the intensity of scattered light at 90°, respectively. In each panel, the overlaid black points are individual measurements of the weight-averaged MW determined at 1-second intervals throughout the elution of chromatographic peaks. Panel **C** shows the elution profiles produced by application of 33  $\mu$ M and 70  $\mu$ M FXR1P Nt-KH2 to the column. Thin (33  $\mu$ M) and thick (70  $\mu$ M) black curves are scattered light intensity at 90°, thin and thick grey line are the corresponding chromatograms monitored by differential refractive index. Grey and black overlaid points are the individual weight-averaged MW measurements for the 33  $\mu$ M and 70  $\mu$ M samples respectively.

**Figure 4** – Analytical centrifugation studies. A) Equilibrium AUC data for FMRP Nt-KH1 (left panel) and FXR1P Nt-KH1 (right panel) at 25,000 rpm and 21  $\mu$ M protein concentration. B) Plot of the distribution of sedimentation coefficients  $C(s)$  vs the sedimentation velocity  $S$  (in Svedberg units) for FMRP Nt-KH1 (black line) and FXR1P Nt-KH2 (grey line).

**Figure 5** – DLS intensity distribution. A) FMRP Nt-KH1 at 1.4 mM. The polydispersity was 9.1% and 10.4% respectively. B) and C) FXR1P Nt-KH2 at 33  $\mu$ M and 114  $\mu$ M respectively. The polydispersity at low concentration is 21%. No reliable quantification could be made at high concentration.

**Figure 6** – Determination of the overall molecular shape. A) SAXS profiles from FXR1P Nt-KH1 (12  $\mu$ M), FXR1P Nt-KH2 (12  $\mu$ M) and FMRP Nt-KH1 (18  $\mu$ M). Experimental data are displayed as dots with error bars, the scattering from typical *ab initio* models computed by DAMMIN as full lines and the calculated curves from the rigid body models computed by BUNCH as dashed lines. The plots display the logarithm of the scattering intensity as a function of momentum transfer. *Ab initio* bead models of FXR1P Nt-KH1 (**B**) and FMRP Nt-KH1 (**C**) as calculated with DAMMIN (gray semitransparent spheres) superimposed with the typical rigid body model obtained with BUNCH. The structures of the NDF and KH1 are displayed as darker grey  $C_\alpha$  traces using the corresponding PDB entries (NDF is always indicated at the top, KH1 at the bottom). The 81 residues of the HLH region, for which there is no structural information, were added by BUNCH and indicated by lighter grey  $C_\alpha$  traces. The right view is rotated counter-clockwise around the vertical axis.

**Figure 7** – Summary of the results presented in this paper with the indication of the range of concentrations explored with each technique.

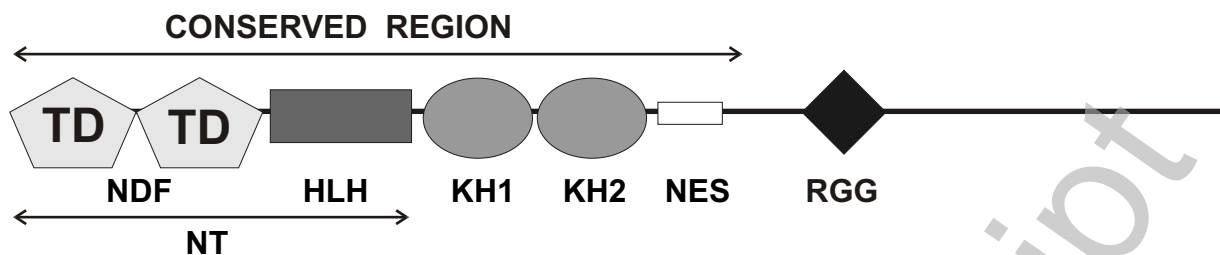
**Table 1** - Synopsis of the SAXS results in the range of concentrations which could be analysed into details.  $R_g$  stands for radius of gyration;  $D_{max}$  for maximum size of the particle,  $V_p$  for excluded volume of the hydrated particle,  $\chi_{ab}$  and  $\chi_{rb}$  values for the fit curves from *ab initio* models and from rigid body modelling using BUNCH, respectively.

	<b>FMRP Nt-KH1</b>	<b>FXR1P Nt-KH1</b>	<b>FXR1P Nt-KH2</b>	<b>FXR1P Nt-KH2</b>	<b>FXR1P Nt-KH2</b>
<b>SAXS</b>					
Concentration range	18 $\mu$ M- 40 $\mu$ M	12 $\mu$ M-35 $\mu$ M	12 $\mu$ M	24 $\mu$ M	36 $\mu$ M
$R_g$ (nm)	3.9	3.7	4.5	4.7	6.2
$D_{max}$ (nm)	13.5 $\pm$ 0.5	13.0 $\pm$ 0.5	15.0 $\pm$ 0.5	16.0 $\pm$ 0.5	25.0 $\pm$ 1.0
MW	38 $\pm$ 6	34 $\pm$ 4	57 $\pm$ 5	77 $\pm$ 5	125 $\pm$ 10
$V_p$ (nm <sup>3</sup> )	52 $\pm$ 5	57 $\pm$ 5	110 $\pm$ 10	135 $\pm$ 10	228 $\pm$ 15
$\chi_{ab}$	1.37	1.48	-	-	-
$\chi_{rb}$	1.75	1.52	-	-	-



Figure 1

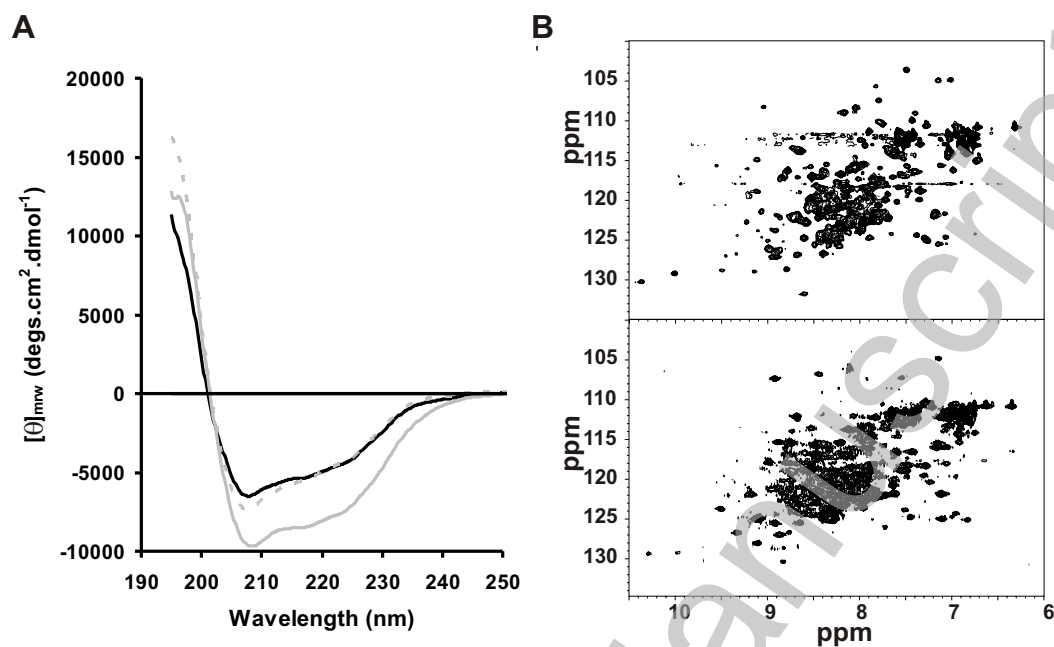
A



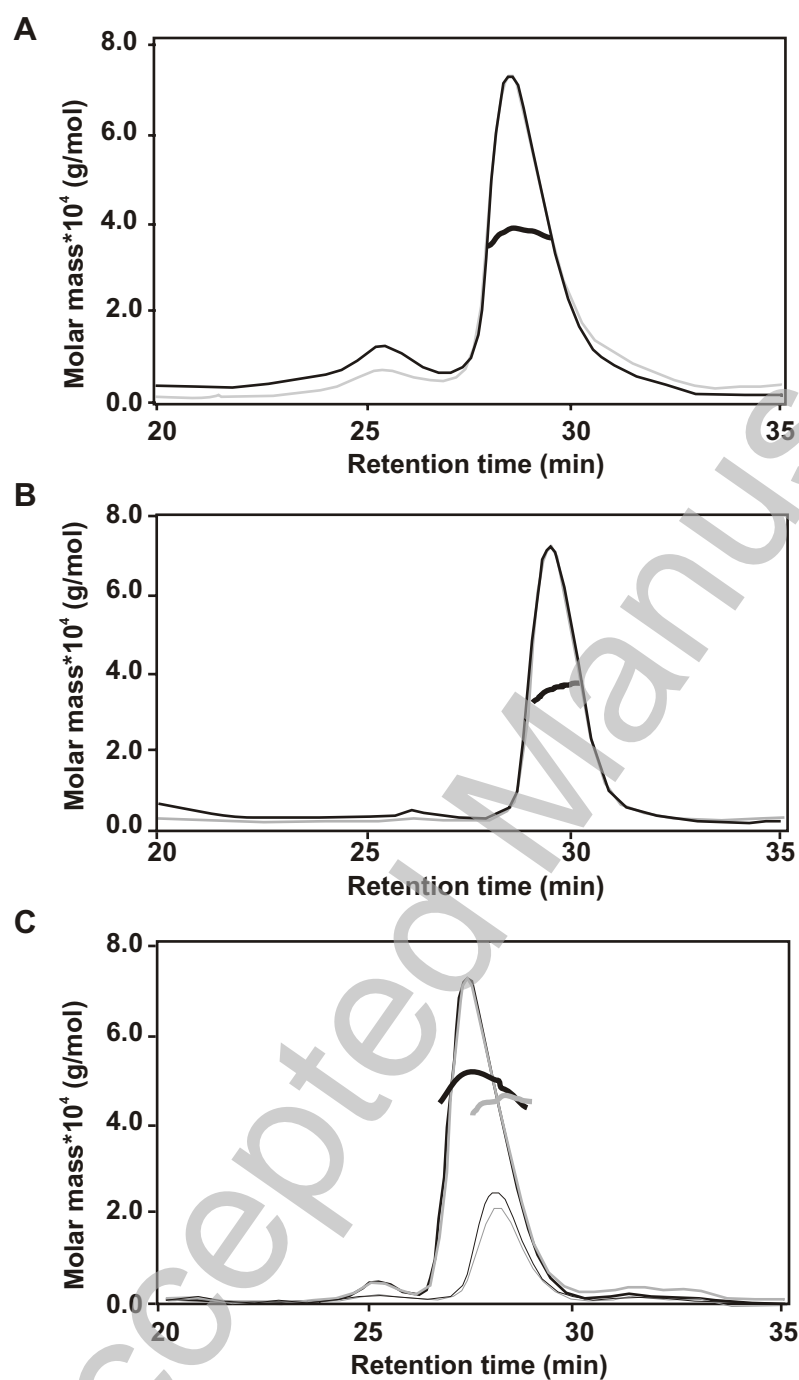
B

		FMRP	FXR1P	FXR2P
	Nt-NES	1-449 degradation	1-380 precipitation	1-391 degradation
	Nt-KH2	1-422 degradation	1-360 stable	N.P.
	Nt-KH1	1-280 stable	1-280 stable	1-290 stable
	ΔNDF-NES	132-449 degradation	132-380 precipitation	141-391 precipitation

Figure 2



**Figure 3**



**Figure 4**

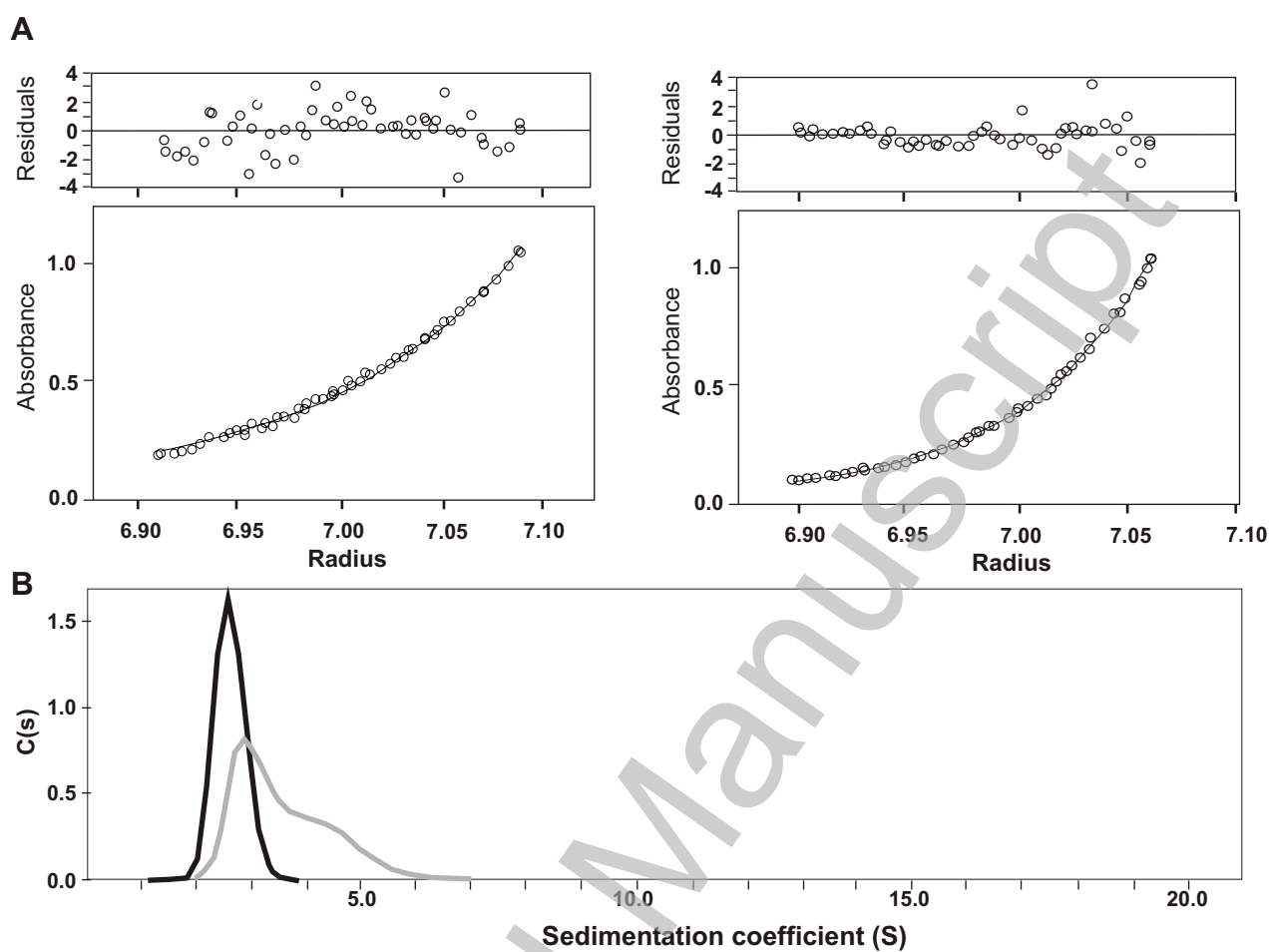
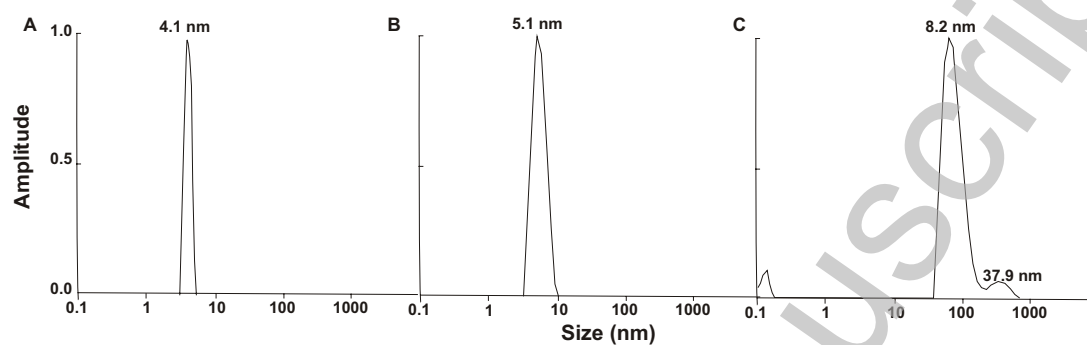
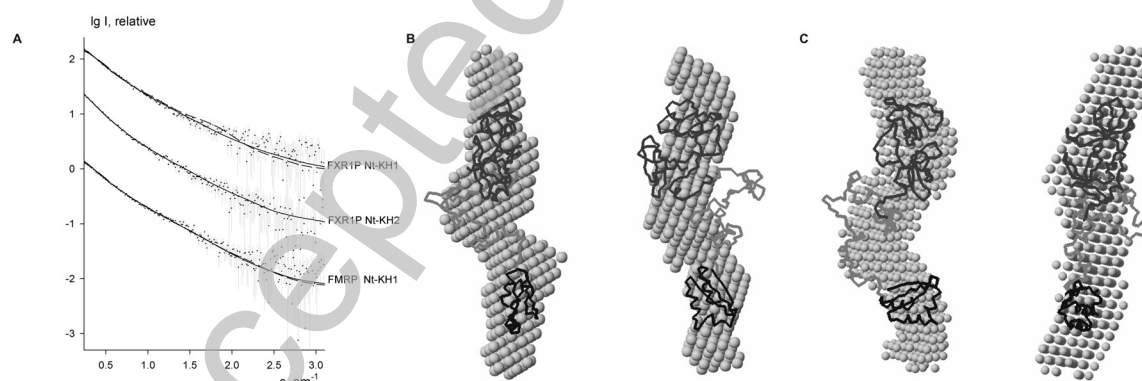


Figure 5





**Figure 6**



**Figure 7**

