BJ www.biochemj.org

BJ Structure

Biochem. J. (2009) **419**, 347–357 (Printed in Great Britain) doi:10.1042/BJ20082197

online data

A study of the ultrastructure of Fragile-X-related proteins

Ljiljana SJEKLOĆA*, Petr V. KONAREV†‡, John ECCLESTON*, Ian A. TAYLOR*, Dmitri I. SVERGUN†‡ and Annalisa PASTORE*[1]

*National Institute for Medical Research, The Ridgeway, London NW7 1AA, U.K., †European Molecular Biology Laboratory, Notkestrasse 85, Hamburg D-22603, Germany, and ‡Institute of Crystallography of Russian Academy of Sciences, Moscow 119333, Russia

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 arginine and glycine residues. The most widely studied member of the family is probably FMRP (fragile X mental retardation protein), since absence or mutation of this protein in humans causes fragile X syndrome, the most common cause of inherited mental retardation. Understanding the structural properties of FMRP is essential for correlating it with its functions. The structures of isolated domains of FMRP have been reported, but nothing is yet known with regard to the spatial arrangement of the different modules, partly because of difficulties in producing both the full-length protein and its multidomain fragments in quantities, purities and monodispersity amenable for structural studies. In the present study, 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 three-dimensional shape. The present study paves the way to further studies and rationalizes the existing knowledge on the self-association properties of these proteins.

Key words: biophysics, fragile X mental retardation, fragile-X-related protein, small-angle X-ray scattering, structure.

INTRODUCTION

Fragile X mental retardation syndrome is a neurodevelopmental X-linked disorder which accounts for mental retardation, autistic behaviour and cranial–facial malformations in one out of 4000 males and in one out of 8000 females [1]. It is caused by silencing of the *FMR1* (fragile X mental retardation 1) gene, with consequent lack of expression of the gene product FMRP (fragile X mental retardation protein). 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′ UTR (untranslated region) of the *FMR1* gene can lead to FXTAS (fragile-X-associated tremor ataxia syndrome), a late-onset progressive neurodegenerative disorder in which the level of *FMR1* mRNA is drastically increased [4,5]. Two FXR (fragile-X-related) 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, all together forming the 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 (knockout) cell lines induce formation of granules by trapping mRNAs into repressed mRNP (messenger ribonucleoprotein) particles [15].

FMRP, like all of the proteins of the FXR family, is modular and consists of the NDF (N-terminal domain of FMRP) which contains two repeats of Tudor motifs, a linker with a potential HLH (helix-loop-helix) motif which also contains an NLS (nuclear localization signal), a central region with two KH motifs (KH1 and KH2) and an NES (nuclear export signal), and of a C-terminus which contains RGG motifs [16,17]. Each of these regions is able to bind RNA individually [18–21]. Although the N-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 up of five β-strands, have been shown to pack against each other and to both be 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 (KH1KH2Δ) 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 recognize 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 degrade shown by 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

Abbreviations used: AUC, analytical ultracentrifugation; DLS, dynamic light scattering; EOM, ensemble optimization method; *FMR1*, fragile X mental retardation 1; FMRP, fragile X mental retardation protein; FXR, fragile-X-related; FXTAS, fragile-X-associated tremor ataxia syndrome; HLH, helix-loop-helix; HSQC, heteronuclear single-quantum coherence; MALDI–TOF, matrix-assisted laser-desorption ionization–time-of-flight; MALLS, multi-angle laser light scattering; mRNP, messenger ribonucleoprotein; NDF, N-terminal domain of FMRP; NES, nuclear export signal; Ni-NTA, Ni^{2+}-nitrilotriacetate; NLS, nuclear localization signal; SAXS, synchrotron radiation X-ray scattering; SEC, size-exclusion chromatography; TCEP, tris-(2-carboxyethyl)phosphine; TEV, tobacco etch virus; TROSY, transverse relaxation optimized spectroscopy; Trx, thioredoxin.

[1] To whom correspondence should be addressed (email apastor@nimr.mrc.ac.uk).

quantities, concentrations and purities necessary for structural studies.

In the present study, we describe a characterization based on complementary biochemical and biophysical techniques of multidomain 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 multidomain constructs and to characterize 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. The present study paves the way to further determination of the ultrastructure of FXR proteins in isolation and in complex with RNA and other protein partners.

EXPERIMENTAL

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://bioportal.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.embnet.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 (Swiss-Prot accession number P51114) and FXR2P (Swiss-Prot accession number P51116) were subcloned 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 an N-terminal His_6-Trx (thioredoxin) tag with a cleavage sequence for TEV (tobacco etch virus) protease ENLYFQ*GA (the asterisk indicates the cleavage site and the extra residues left after cleavage are underlined). A differently modified pET-24 vector with a non-removable C-terminal His_6 tag was used for production of FXR1P Nt-KH2.

Protein expression in *Escherichia coli* BL21 (DE3) STAR strain was induced with 1 mM IPTG (isopropyl β-D-thiogalactoside) 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 8.0), 150 mM NaCl, 10 mM imidazole, 0.2 % Igepal CA-630 (Sigma–Aldrich) and 1 mM PMSF, supplemented with the Complete EDTA-free protease inhibitors mix (Roche), and lysed by freeze–thawing and ultrasound sonication. The soluble fraction of centrifugation-clarified cell lysate was applied on to Ni-NTA (Ni^{2+}-nitrilotriacetate) metal-affinity chromatography matrix (Qiagen) from which the recombinant His_6-Trx or C-terminal His_6-tagged proteins were eluted with 50 mM Tris/HCl (pH 8.0), 300 mM imidazole and 2 mM 2-mercaptoethanol and dialysed into 50 mM Tris/HCl (pH 8.0), 1 mM DTT (dithiothreitol) and 0.5 mM EDTA. The His_6-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 5/5GL), and analytical SEC (size-exclusion chromatography; Superdex 200 HR 10/30 GL or Superdex 75 HR 10/30 GL). Protein identity was confirmed by ES (electro-spray) and MALDI–TOF (matrix-assisted laser-desorption ionization–time-of-flight) spectrometry and N-terminal protein sequencing. Protein purity was assessed by Coomassie-Blue-stained SDS/PAGE and MS. Protein concentrations were estimated by measurements of absorbance at 280 nm using theoretical molar absorption coefficients calculated by ProtParam. The proteins were concentrated using an ultrafiltration device (VIVASpin).

^{15}N-labelled proteins were prepared as described above, but by growing the cells in M9 minimal media supplemented with ^{15}N-labelled ammonium sulfate.

Spectroscopic measurements

CD spectra of the purified recombinant proteins were recorded at 20 °C using a Jasco J-715 spectropolarimeter (Jasco), over a wavelength range from 260 to 195 nm in 1-mm-pathlength quartz cuvettes (Hellma). The purified recombinant peptides were used at protein concentrations of 5 μM in a buffer containing 50 mM Tris/HCl (pH 8.0) and 1 mM 2-mercaptoethanol.

1D or 2D NMR (^{1}H,^{15}N)-HSQC (heteronuclear single-quantum coherence) and TROSY (transverse relaxation optimized spectroscopy) spectra were recorded at 25 °C and 800 MHz using a Varian Inova spectrometer (Varian) equipped with a cryoprobe. The range of protein concentrations explored was from 80 to 230 μM. Spectra were recorded for proteins in 50 mM Tris/HCl (pH 7.5) and 5 mM TCEP [tris-(2-carboxyethyl)phosphine] 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 2-mercaptoethanol. We tested samples in a concentration range between 10 and 900 μM. The injection volume was 100 μl and the flow rate was 0.5 ml/min.

DLS (dynamic light scattering) measurements were performed at 20 °C or 10 °C in a concentration range of 2.5 μM–1.4 mM for the two Nt-KH1 constructs and 2.5–114 μM for FXR1P Nt-KH2, using a Viscotek model 802 instrument (Viscotek). Results were analysed by integrated Viscotek OmniSIZE software. All DLS measurements were carried out without previous dilution or filtration.

Molecular masses and molecular-mass distributions were determined using on-line MALLS (multi-angle laser light scattering) coupled with SEC (SEC–MALLS) [29]. Samples were applied to a Superdex 200 10/300 GL column equilibrated in 50 mM Tris/HCl (pH 8.0) and 5 mM TCEP at a flow rate of 0.5 ml/min. The column was mounted on a Jasco HPLC machine controlled by the Chrompass software package. The scattered light intensity of the column eluent was recorded at 16 angles using a DAWN-HELEOS laser photometer (Wyatt Technology). The loaded concentrations were 230 μM and 140 μM for FMRP and FXR1P Nt-KH1 respectively. Two concentrations (33 μM

and 70 μM) were tested for FXR1P Nt-KH2. The concentration of the eluent was determined from the refractive index change ($dn/dc = 0.186$) using an OPTILAB-rEX differential refractometer equipped with a Peltier temperature-regulated flow cell, maintained at 25 °C (Wyatt Technology). The wavelength of the laser in the DAWN-HELEOS and the light source in the OPTILAB-rEX was 658 nm. The mass-averaged molecular mass of material contained in chromatographic peaks was determined using the ASTRA software version 5.1 (Wyatt Technology). 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_θ against $\sin^2(\theta/2)$]. The mass-averaged molecular mass was then calculated at each point in the chromatogram from the intercept of an individual plot. An overall average molecular mass and polydispersity term for each species was calculated by combining and averaging the data from the individual measurements.

AUC (analytical ultracentrifugation) equilibrium measurements were carried out on an Optima XL-A (Beckman Coulter) 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 μM) which were obtained by diluting the protein eluted from the 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 rev./min, 22 500 rev./min, 25 000 rev./min and 40 000 rev./min. The data were globally fitted 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 μM (FMRP Nt-KH1) and 33 μM (FXR1P Nt-KH2), were centrifuged at 20 °C and 40 000 rev./min. Two hundred scans were recorded at 5 min intervals. The data were analysed as a distribution of sedimentation coefficients using SEDFIT [30].

SAXS (synchrotron radiation X-ray scattering) experiments

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) and 5 mM TCEP were measured for solute concentrations of 12–250 μM, 12–126 μM and 18–420 μM respectively. The experiments were carried out at 10 °C. A MAR345 image plate at a sample-detector distance 2.7 m and wavelength $\lambda = 0.15$ nm, covering the momentum transfer range $0.12 < s < 4.9$ nm^{-1} ($s = 4\pi \sin(\theta)/\lambda$ where 2θ 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 molecular masses were evaluated by comparison of the forward scattering with that from reference solutions of BSA (molecular mass 66 kDa). The excluded volume of the hydrated particle (the Porod volume) was computed as previously described [35] (eqn 1):

$$V = 2\pi^2 I(0) / \int_0^\infty s^2 I_{exp}(s) ds \quad (1)$$

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 $I_{exp}(s)$ to minimize the discrepancy (eqn 2):

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

where N is the number of experimental points, c is a scaling factor and $I_{calc}(s_j)$ 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 number 2BKD) and the crystal structure of FMRP KH1 (PDB number 2QND). The scattering from these models was calculated using the program CRYSOL [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 EOM (ensemble optimization method) [40] which allows for co-existence 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 pulldown assay was performed to check for potential interaction of FMRP Nt-KH1 and FXR1P Nt-KH2–His_6. Pure recombinant FXR1P Nt-KH2–His_6 was immobilized on Ni-NTA–agarose beads, extensively washed with equilibrating buffer [20 mM Tris/HCl (pH 7.4), 0.1 mM EDTA and 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 gels by Coomassie Blue staining. Both FXR1P Nt-KH2–His_6 and FMRP Nt-KH1 were used at concentrations of 10 μM.

RESULTS

Identification of potential degradation hot spots in FXR proteins

Obtaining properly folded homogeneous protein samples is the prerequisite for structural studies. We analysed 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

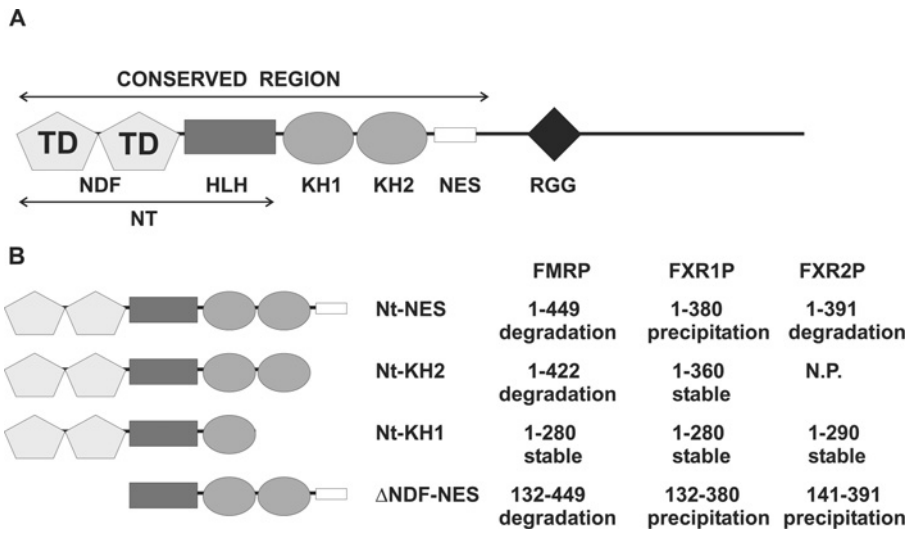

Figure 1 Choice and production of the multidomain constructs used in the present study

(**A**) Schematic representation of the architecture of FXR proteins. The different sequence motifs are indicated with different geometric shapes. (**B**) Summary of the constructs produced and their properties. N.P., not produced.

[41]. This suggests that the C-termini of FXR proteins (approx. from residue 430 to 632) have a high likelihood of being 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 a high content of proline, glutamate, serine and threonine residues, which target proteins for ubiquitin-mediated degradation in eukaryotic cells. These sequences could, in principle, also contribute to protein degradation in *E. coli*, as PEST-mediated degradation has also been reported in *L. monocytogenes* [43]. We identified potential PEST sequences in the C-termini of all FXR proteins (see Supplementary Figure S1 and Supplementary Table S1 at http://www.BiochemJ.org/bj/419/bj4190347add.htm). 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 a review see [12]).

Production of multidomain recombinant fragments

With the ultimate goal 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 (Supplementary Figure S1). 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, although 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 analysed. 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. In contrast, 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 metalloprotease between residues Tyr^{288} and Leu^{289} (which is near the end of KH1) and, to a lesser extent, by a trypsin-like protease between residues Arg^{255} and Lys^{256} (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 (Supplementary 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 unsuitable for structural characterization because of concentration- and salt-dependent aggregation and precipitation (results not shown).

The longest stable recombinant fragment (not degrading and not precipitating) was Nt-KH2 of FXR1P (Supplementary Figure S2 at http://www.BiochemJ.org/bj/419/bj4190347add.htm). 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 ΔNDF-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.

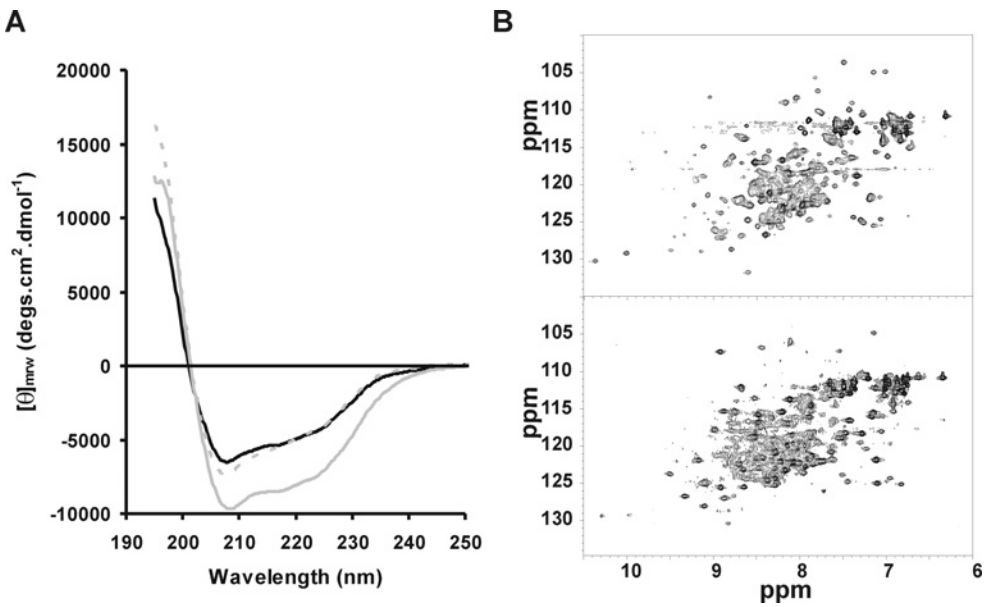

Figure 2 Spectroscopic studies of FXR proteins

(**A**) Far-UV CD spectra in molar ellipticity of the Nt-KH1 of FMRP (black continuous line), FXR1P (grey broken line) and FXR1P Nt-KH2 (grey continuous line). The measurements were carried out at 20 °C using 5 μM concentrations. (**B**) NMR HSQC spectra of Nt-KH1 from FXR1P (top panel) and FMRP (bottom panel). The spectra were recorded at 800 MHz and 25 °C, using concentrations of 60 μM (FXR1P) and 230 μM (FMRP).

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. For the present study we did not use 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 characterized by CD and NMR techniques to ensure that they were 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 was, as expected, very similar for the two Nt-KH1 constructs (approx. 21 % α, 27 % β, 20 % turn and 32 % random) and increased for FXR1P Nt-KH2 (28 % α, 22 % β, 20 % turn and 30 % random). This increase was roughly consistent with the increment expected by assuming the secondary structure content observed in KH1KH2Δ.

NMR (^{1}H,^{15}N)-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 was always too broad in the whole range of concentrations explored (30–230 μM), and also in TROSY experiments (results not shown). The spectra of Nt-KH1 from FXR1 and FMRP were similarly broad when in the presence of 150 mM NaCl. They were instead relatively sharp at protein concentrations as high as 230 μM when collected from samples with virtually no salt other than the buffer (Figure 2B). This was 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 belonged 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 only three tryptophan residues 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 systematically test 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 and 150 mM NaCl).

A preliminary qualitative analysis by SEC suggested that both Nt-KH1 constructs are eluted from the column as monomeric proteins at micromolar concentrations (results not shown). Under similar conditions, Nt-KH2 is eluted instead at a molecular mass compatible with a mixture of monomer and dimer. In order to characterize these solution molecular masses in more detail, SEC–MALLS detection was employed, a technique that is able to determine molecular mass, stoichiometry and organization of assemblies present in solution [37]. The SEC–MALLS data demonstrate that the measured solution molecular masses of FMRP Nt-KH1 (37 100 $\pm$ 0.6 % Da), and FX1RP Nt-KH1 (34 800 $\pm$ 0.6 % Da) were comparable with monomeric molecular masses. Moreover, they display negligible concentration-dependency, at least up to concentrations of $\sim$ 150 μM (Figures 3A and 3B). SEC–MALLS analysis of FXR1P Nt-KH2 reveals that inclusion of the KH2 domain results in significant concentration-dependency of the mass-averaged molecular mass of the protein (Figure 3C). In

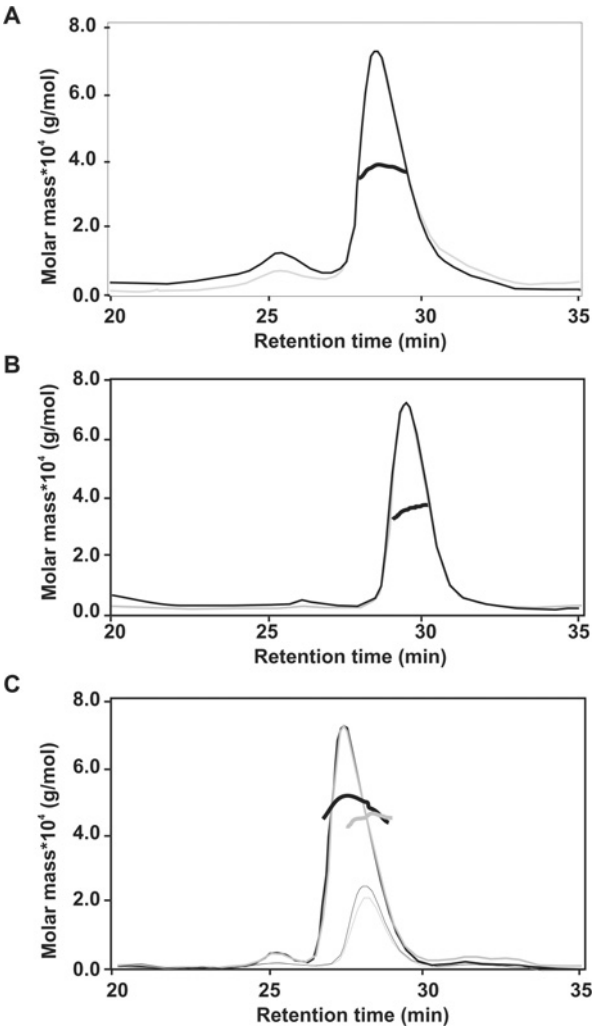

Figure 3 Solution molecular mass of FMRP Nt-KH1, FXR1P Nt-KH1 and FXR1P Nt-KH2 determined by SEC–MALLS

(**A** and **B**) are the elution profiles produced by application of 100 μl of 150 μM FMRP Nt-KH1 and 75 μM FXR1P Nt-KH1 to a Superdex 200 10/300 GL column. The grey and black 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 mass-averaged molecular mass determined at 1-s intervals throughout the elution of chromatographic peaks. (**C**) The elution profiles produced by application of 33 μM and 70 μM FXR1P Nt-KH2 to the column. Thin (33 μM) and thick (70 μM) black curves are scattered light intensity at 90°; thin and thick grey lines are the corresponding chromatograms monitored by the differential refractive index. Grey and black overlaid points are the individual mass-averaged molecular mass measurements for the 33 μM and 70 μM samples respectively.

this case, the mass-averaged molecular mass of a sample with an average elution peak concentration of 2.5 μM is 45 000 $\pm$ 0.8 % Da, i.e. only marginally bigger than the value expected for the monomer (42 000 Da). At a higher protein concentration (6.6 μM) a shift in the peak to an earlier eluting position was apparent and the observed molecular mass increased significantly up to 49 000 $\pm$ 0.6 % Da, indicating self-associative behaviour. On the simple assumption of a monomer/dimer equilibrium, the mass-averaged molecular mass of 49 000 Da determined at 6.6 μM equated 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 7–21 μM). The equilibrium AUC results could be fitted well, assuming the presence of a monomeric species only for the Nt-KH1 samples (i.e. molecular masses of 32 $\pm$ 1 kDa and 29 $\pm$ 1 kDa for the FMRP and FX1RP constructs respectively, which should be compared with their theoretical molecular masses of 32 349 Da and 31 570 Da) (Figure 4A), but not for FXR1P Nt-KH2, for which the curves could not be fitted at any concentration, in agreement with the presence of a mixture of species. When analysed by sedimentation velocity AUC, both Nt-KH1 constructs sedimented with coefficients of 2.5 S, which corresponded to molecular masses of approx. 31 kDa, well in agreement with the values expected for the monomer. At the same concentration (33 μM), FXR1P Nt-KH2 sedimented as an unresolved mixture of two components with molecular mass values of approx. 40 kDa and 80 kDa, and with sedimentation coefficients of 3 S and $\sim$4.5 S respectively (Figure 4B).

Finally, we used DLS to explore the construct behaviour over a wider range of concentrations (5 μM–1.4 mM). Both Nt-KH1 constructs were monomeric up to $\sim$150 μM. At higher concentrations, we observed an increase of the hydrodynamic radii (from $\sim$3 to $\sim$4 nm) towards values consistent with monomer/dimer mixtures (Figure 5A). This indicated that self-association may also occur for these samples at sufficiently elevated concentrations. Nt-KH2 of FXR1P behaved as a monomeric species only up to $\sim$5–10 μM. At higher concentrations, we observed the progressive increase of the hydration radius up to a value indicative of large-molecular-mass aggregates (Figure 5B). Concomitantly, the polydispersity increased.

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 (amino acid residues 135–217), for which no structure is available. For FXR1P Nt-KH1, we obtained a molecular mass of 34 $\pm$ 4 kDa when studying it at 12–35 μM. The excluded (Porod) volume of the particle in solution, a parameter which for globular proteins are expected to be approx. twice the molecular mass in kDa, is 57 $\pm$ 5 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_{max}) (3.77 $\pm$ 0.04 nm and 13.0 $\pm$ 0.5 nm respectively) point to an elongated structure. The long-tailed shape of the distance distribution function p(r) for FRX1P Nt-KH1 is also consistent with an elongated shape of the protein (Supplementary Figure S3 at http://www.BiochemJ.org/bj/419/bj4190347add.htm). 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 χ_{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 probably corresponds to the KH1 module. At higher concentrations (250 μM) the FXR1P Nt-KH1 sample revealed unspecific aggregation.

The overall parameters of FMRP Nt-KH1 at 18–40 μM 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_{max}, but the overall shape is similar (Figure 6C). At higher

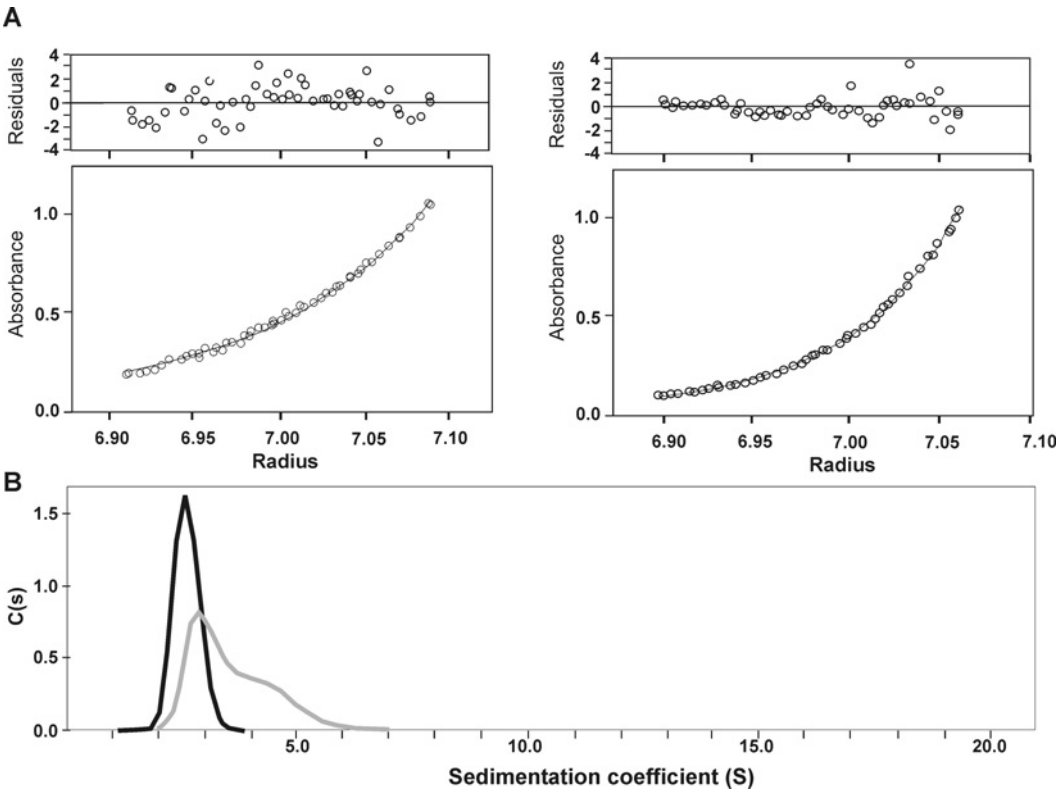

Figure 4 AUC studies

(**A**) Equilibrium AUC data for FMRP Nt-KH1 (left-hand panel) and FXR1P Nt-KH1 (right-hand panel) at 25 000 rev./min and 21 μM protein concentration. (**B**) Plot of the distribution of sedimentation coefficients [C(s)] against the sedimentation velocity (S) for FMRP Nt-KH1 (black line) and FXR1P Nt-KH2 (grey line).

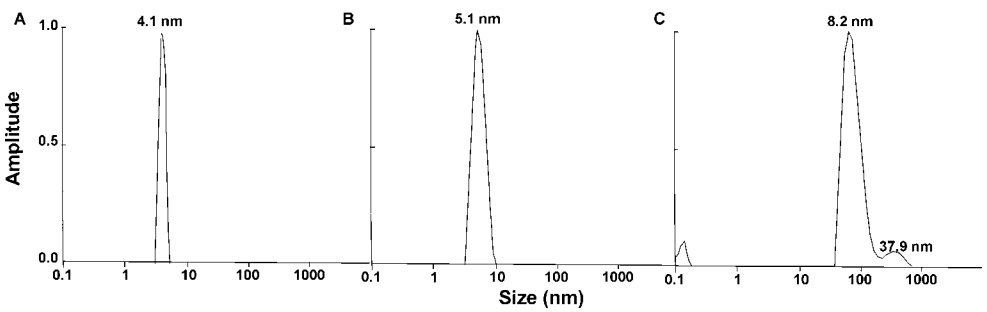

Figure 5 DLS intensity distribution

(**A**) FMRP Nt-KH1 at 1.4 mM. The polydispersity was 9.1 %. (**B** and **C**) FXR1P Nt-KH2 at 33 μM and 114 μM respectively. The polydispersity at low concentrations is 21 %. No reliable quantification could be made at high concentrations.

concentrations of FMRP Nt-KH1 the presence of larger aggregates was observed.

Nt-KH2 samples measured in a concentration range 12–126 μM revealed strong concentration-dependence in the scattering patterns already at low concentrations (Table 1). The higher concentration samples (48 μM and 126 μM) clearly show high molecular mass and R_g values consistent with the presence of aggregates, possibly tetramers and even larger. The parameters measured at 12 μM and 24 μM reasonably correspond to a monomer-to-dimer mixture. Indeed, the experimental molecular mass 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 molecular mass of the Nt-KH2 monomer is 42 kDa). The increase of overall parameters and excluded volume at 24 μM (compared with 12 μM) points 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 (Supplementary Figure S3B) resembles the shapes obtained for the shorter constructs (Figures 6B and 6C). Since all of 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 confirmed that the dimerization of FXR1P Nt-KH2 should occur through the 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

Table 1 Synopsis of the SAXS results in the range of concentrations which could be analysed into details

V_p, excluded volume of the hydrated particle; χ_{ab} and χ_{rb} are the values for the fitted curves from *ab initio* models and from rigid body modelling using BUNCH respectively.

SAXS result	FMRP Nt-KH1	FXR1P Nt-KH1	FXR1P Nt-KH2	FXR1P Nt-KH2	FXR1P Nt-KH2
Concentration range	18–40 μM	12–35 μM	12 μM	24 μM	36 μM
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
Molecular mass (kDa)	38 $\pm$ 6	34 $\pm$ 4	57 $\pm$ 5	77 $\pm$ 5	125 $\pm$ 10
V_p (nm^3)	52 $\pm$ 5	57 $\pm$ 5	110 $\pm$ 10	135 $\pm$ 10	228 $\pm$ 15
χ_{ab}	1.37	1.48	–	–	–
χ_{rb}	1.75	1.52	–	–	–

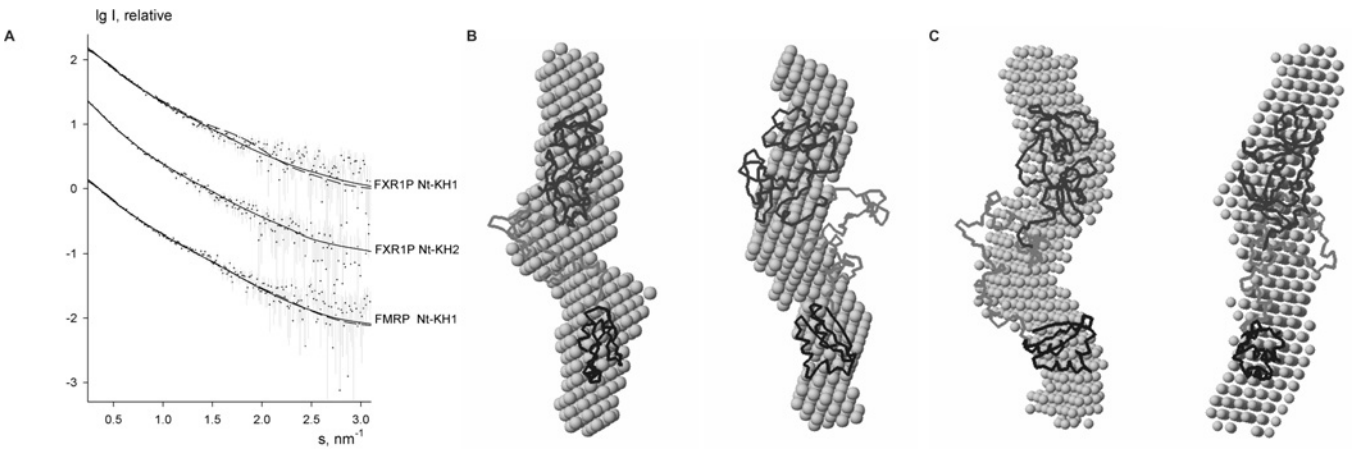

Figure 6 Determination of the overall molecular shape

(**A**) SAXS profiles from FXR1P Nt-KH1 (12 μM), FXR1P Nt-KH2 (12 μM) and FMRP Nt-KH1 (18 μM). 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 broken 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 (grey semi-transparent spheres) superimposed with the typical rigid body model obtained with BUNCH. The structures of the NDF and KH1 are displayed as darker grey C_α traces using the corresponding PDB entries (NDF is always indicated at the top, KH1 is indicated at the bottom). The 81 residues of the HLH region, for which there is no structural information, were added by BUNCH and are indicated by lighter grey C_α traces. The panel on the right is rotated counter-clockwise around the vertical axis.

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 (χ values 1.52 and 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 results 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 of the R_g distribution derived from the optimized ensembles with that obtained from randomly generated models (Supplementary 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

As we had ascertained that recombinant Nt-KH1 of either FMRP or of FXR1P was not able to homodimerize at low concentrations, we investigated whether Nt-KH1 could mediate heterodimerization of the two proteins *in vitro*. Pulldown assays were performed using FXR1P Nt-KH2 and FMRP Nt-KH1 at concentrations in which they were monomeric (10 μM). We did not observe interaction (results 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 with a 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 included the phosphorylation site at Ser^{500} [44]. Chemical modification of this serine residue did not, however, make any difference to 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 a 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]. Although it is known that truncated versions of

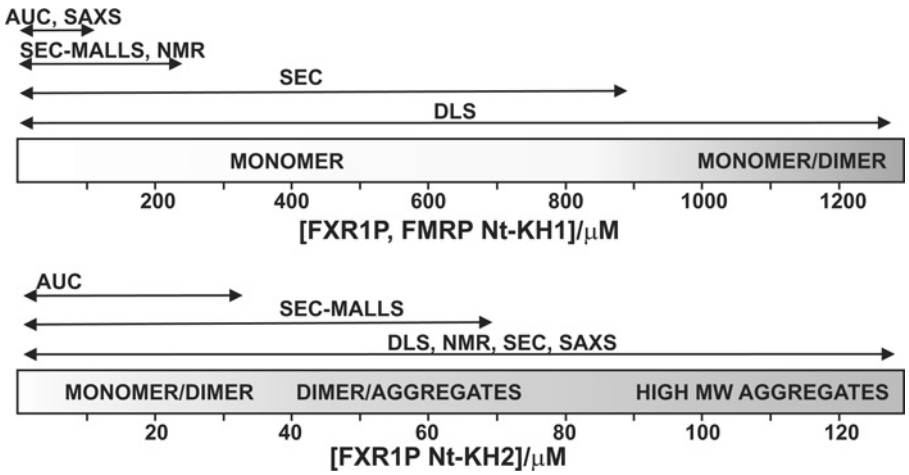

Figure 7 Summary of the results in the present study with the indication of the range of concentrations explored with each technique

MW, molecular mass.

a protein may be prone to aggregation, the observation of similar properties for different fragments and the notion also that the full-length proteins have a 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 the present study is that we do not only observe protein dimerization, but a more complex pattern of self-association which proceeds in a continuous way from the monomer to large-molecular-mass 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 KH1KH2Δ in which the dimer interface is formed only by KH2 [25]. Generally, KH domains of different proteins are known to mediate homodimerization *in vitro* [49].

Outside the conserved region explored in the present study, it is likely that, *in vivo*, the elements downstream from 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 leucine residues 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, the present study 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, rationalizing only apparently contradictory previous reports. Conclusive evidence at atomic details of the molecular mechanism through which FXRP 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.

ACKNOWLEDGEMENTS

We thank Dr Steven Howell for MS and for MALDI–TOF experiments, Dr Stephen R. Martin for help with CD 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 greatful to Dr Human Rezaei for helpful scientific discussion.

FUNDING

This work was supported under the auspices of the Medical Research Council at the National Institute for Medical Research.

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 Boulle, 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., Hessl, D., Harris, S. W., Zhang, L. et al. (2004) Fragile-X-associated tremor/ataxia syndrome (FXTAS) in females with the FMR1 premutation. Am. J. Hum. Genet **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, Jr, C. T., 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 Commun. **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**, 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. Nat. 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. Nucleic Acids 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 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. and Svergun, D. (2007) Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. J. Appl. Crystallogr. **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. Crystallogr. **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. Crystallogr. **25**, 495–503

35 Porod, G. (1982) General theory. In Small-angle X-ray Scattering (Glatter, O. and Kratky, O., eds.), pp. 17–51, Academic Press, London

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., Barberato, C. and Koch, M. H. J. (1995) CRYSOL: a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. **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. In The Molecular Basis of Fragile X Syndrome (Sung, Y. J. and Denman, R. B., eds.), pp. 29–41, Research Signpost

Received 6 November 2008/6 January 2009; accepted 14 January 2009
Published as BJ Immediate Publication 14 January 2009, doi:10.1042/BJ20082197

Biochem. J. (2009) **419**, 347–357 (Printed in Great Britain) doi:10.1042/BJ20082197

online data

SUPPLEMENTARY ONLINE DATA
A study of the ultrastructure of Fragile-X-related proteins

Ljiljana SJEKLOĆA*, Petr V. KONAREV†‡, John ECCLESTON*, Ian A. TAYLOR*, Dmitri I. SVERGUN†‡ and Annalisa PASTORE*[1]

*National Institute for Medical Research, The Ridgeway, London NW7 1AA, U.K., †European Molecular Biology Laboratory, Notkestrasse 85, Hamburg D-22603, Germany, and ‡Institute of Crystallography of Russian Academy of Sciences, Moscow 119333, Russia

Figure S1 Multiple sequence alignment of FXR proteins

The sequences are shaded according to the positional conservation. The positions of the putative PEST motifs are underlined.

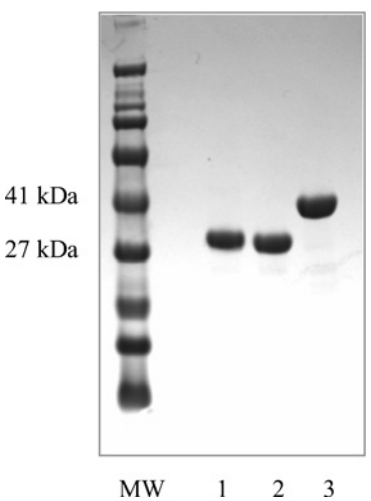

Figure S2 Gel of the three stable constructs discussed in the main text to exemplify the degree of purity

The molecular mass in kDa is indicated on the left-hand side.

[1] To whom correspondence should be addressed (email apastor@nimr.mrc.ac.uk).

Table S1 Potential PEST sequences which could promote degradation of FXR proteins

The predictions were made using PESTFind.

Protein	Potential PEST sequences (amino acid residues)	Poor PEST sequences (amino acid residues)
FMRP	484–505; 507–522	24–40; 158–179; 246–263; 276–290; 324–344
FXR1P	391–415; 470–497	24–40; 162–179; 246–263; 328–341; 578–593
FXR2P	437–459; 512–544; 557–581	1–19; 162–189; 256–273; 286–300; 338–351; 606–617

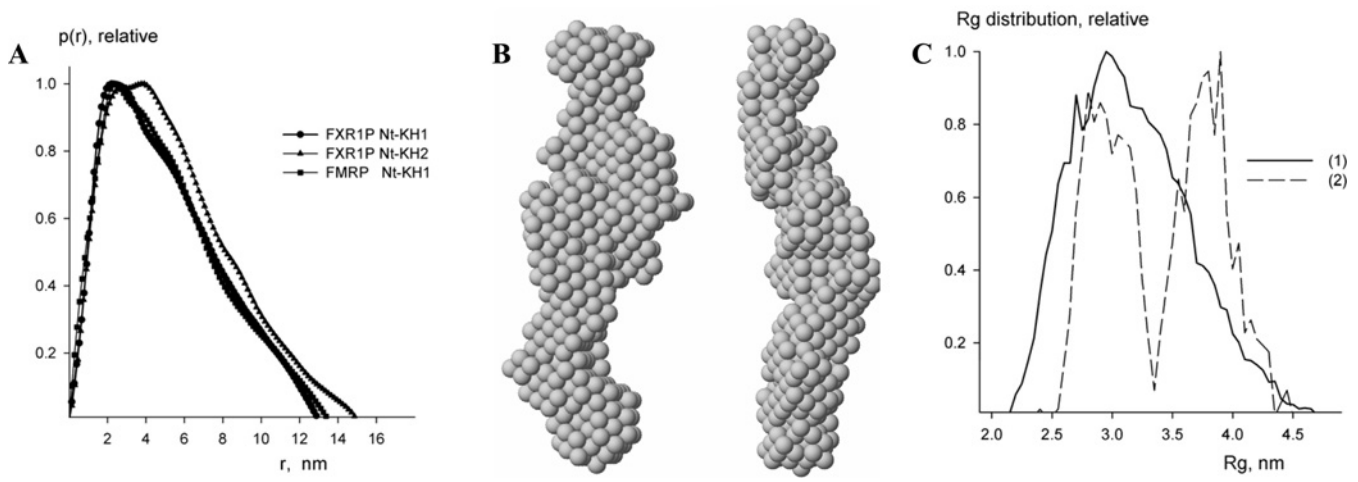

Figure S3 SAXS data

(**A**) Distance distribution functions from FXR1P Nt-KH1 (12 μM), FXR1P Nt-KH2 (12 μM) and FMRP Nt-KH1 (18 μM). (**B**) *Ab initio* bead model of FXR1P Nt-KH2 calculated with DAMMIN (grey semi-transparent spheres) The right-hand view is rotated counter-clockwise around the vertical axis. (**C**) R_g distributions from EOM for FXR1P Nt-KH1 (12 μM): initial random pool (solid line) and selected ensembles averaged over 50 independent EOM runs (broken line). The wide width of these distributions supports the flexibility of the HLH region.

Received 6 November 2008/6 January 2009; accepted 14 January 2009
Published as BJ Immediate Publication 14 January 2009, doi:10.1042/BJ20082197