The solution structure of the pentatricopeptide repeat protein PPR10 upon binding *atpH* RNA

Benjamin S. Gully[1], Nathan Cowieson[2], Will A. Stanley[1,3], Kate Shearston[1,3], Ian D. Small[3], Alice Barkan[4] and Charles S. Bond[1,*]

[1]School of Chemistry and Biochemistry, The University of Western Australia, Crawley, Western Australia, Australia, [2]SAXSWAXS beamline, Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria, Australia, [3]Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, Crawley, Western Australia, Australia and [4]Institute of Molecular Biology, University of Oregon, Eugene, Oregon, USA

ABSTRACT

The pentatricopeptide repeat (PPR) protein family is a large family of RNA-binding proteins that is characterized by tandem arrays of a degenerate 35-amino-acid motif which form an α-solenoid structure. PPR proteins influence the editing, splicing, translation and stability of specific RNAs in mitochondria and chloroplasts. *Zea mays* PPR10 is amongst the best studied PPR proteins, where sequence-specific binding to two RNA transcripts, *atpH* and *psaJ,* has been demonstrated to follow a recognition code where the identity of two amino acids per repeat determines the base-specificity. A recently solved ZmPPR10:*psaJ* complex crystal structure suggested a homodimeric complex with considerably fewer sequence-specific protein–RNA contacts than inferred previously. Here we describe the solution structure of the ZmPPR10:*atpH* complex using size-exclusion chromatography-coupled synchrotron small-angle X-ray scattering (SEC-SY-SAXS). Our results support prior evidence that PPR10 binds RNA as a monomer, and that it does so in a manner that is commensurate with a canonical and predictable RNA-binding mode across much of the RNA–protein interface.

INTRODUCTION

The pentatricopeptide repeat (PPR) superfamily (1) is characterized by a degenerate 35 amino acid repeat and is specific to eukaryotes although most numerous and varied in the plant kingdom (2), with over 400 identified in *Arabidopsis thaliana* (3). PPR proteins function in RNA stabilization (4–6), editing (7), maturation and post-transcriptional modifications (8,9) and are essential for organelle biogenesis (3) and translational control (10–14). Mutations can result in cytoplasmic male sterility (15,16), impaired seed development (17) and diverse embryonic defects (18).

PPR proteins can be broadly divided into two families, those consisting of tandem arrays of canonical P-class motifs and those composed of repeats of triplets of P-, L- and S-class motifs (3). Regardless of repeat class, conservation of key amino acids from each motif has been correlated with the sequence of known RNA targets to establish a statistically and experimentally validated predictive code for RNA recognition by PPR proteins (19–21). This code is consistent with a modular interaction where a base is coordinated by the sixth residue of one PPR repeat, and the first residue of the next repeat. Such a modular, predictable, mode of single-stranded RNA binding makes PPR proteins a potential scaffold for the design of biotechnologically useful proteins (22,23).

Predictions of an α-solenoid tertiary structure for PPR tracts (1,24) have been confirmed recently by a number of crystal structures including a segment of the mitochondrial RNA polymerase (25), Arabidopsis proteinaceous ribonuclease P (26), the P-class PPR protein THA8L (27) and a synthetic consensus P-class PPR protein (28). Crystal structures of two PPR proteins, *Brachypodium* THA8 (29) and *Zea mays* PPR10 (ZmPPR10) (30), in complex with RNA have partially confirmed the modular RNA-binding mode, but revealed non-modular and idiosyncratic interactions that had not been anticipated based on prior data. In addition, these structures implicate a dimeric quaternary structure in RNA binding.

ZmPPR10 localizes to plastids where it binds *cis*-elements of two polycistronic transcripts (6,13), the minimal binding sites identified as 17 and 18 nucleotide tracts in the *atpI-atpH* and *psaJ-rpl33* intercistronic regions. PPR10 binding was shown to stabilize the RNA *in vivo*, protecting the transcripts from exoribonucleases in either direction, and to activate mRNA translation by remodeling lo-

cal RNA structure ([6]). Analytical ultracentrifugation experiments unambiguously demonstrated the association of PPR10 monomers into dimers at high concentration and dissociation of PPR10 dimers into monomers on binding an oligonucleotide representing *atpH* RNA, resulting in a 1:1 protein:RNA stoichiometry ([19]), with a dissociation constant that has been measured at ca. 1 nM ([6]).

In the recent crystal structure of ZmPPR10 in complex with an 18 nucleotide RNA fragment of *psaJ* (PDB 4M59 ([30])), the protomer consists of a continuous array of $20\frac{1}{2}$ P-class PPR repeats forming a right-handed superhelix. A dimer is generated by the association of two anti-parallel molecules, and one RNA molecule is bound at each end of the dimer by elements of both monomers. Notably, reported attempts to crystallize PPR10 with *atpH* resulted in apparent dissociation of the dimer and failure to crystallize, suggesting the possibility of a different, higher-affinity, mode of RNA binding compared to *psaJ*, for which a dissociation constant of >200 μM has been measured ([13]). Here we use size-exclusion chromatography-coupled synchrotron small-angle X-ray scattering (SAXS) ([31]) to produce clear structural evidence for a monomeric PPR10:*atpH* complex in solution, demonstrating a potential structural rearrangement with respect to the PPR10:*psaJ* crystal structure, that is more consistent with the proposed PPR–RNA recognition code and with other prior data concerning PPR10–RNA interactions.

MATERIALS AND METHODS

Protein expression and purification

The *Zm*PPR10:pMAL-TEV plasmid generated as described previously ([13]) was used to transform *E. coli* Rosetta[TM] 2 (DE3) (Novagen). Cells were grown in 2YT media containing 1% (w/v) D-glucose, 50 mM Tris-HCl at pH 7.5 with kanamycin at 50 μg ml^{-1} and chloramphenicol at 50 μg ml^{-1}. The culture was grown at 37°C to an optical density (600 nm) of 0.6 and then cooled on ice for 5 min. Expression was then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside and the culture shaken overnight at 16°C for protein expression. The bacterial pellet was resuspended in Buffer A (50 mM Tris-HCl at pH 8.0, 500 mM KCl, 10% (v/v) glycerol, 1 mM dithiothreitol) supplemented with 0.13 mM phenylmethanesulfonyl fluoride (Roche), 1 mini Complete protease inhibitor tablet (Roche), 0.5 μl Benzonase (Sigma-Aldrich) and 1 mg ml^{-1} lysozyme, and lysed under high pressure using a Emulsiflex C5 homogenizer (Avestin). Lysate supernatant was batch loaded onto amylose resin (GE healthcare) and left at 4°C for 1 h with agitation. The column was washed with buffer A, and fusion protein eluted with buffer A supplemented with 50 mM maltose. Fusion protein was cleaved overnight at 4°C with tobacco etch virus protease and then subjected to size-exclusion chromatography (BioLogic DuoFlow, Bio-Rad) with a HiLoad 16/60 Superdex 200 prep grade column (GE Healthcare) developed in buffer A. Peak fractions were confirmed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (NuPage® Novex 4–12%, Bis-Tris gel, Invitrogen) stained with Coomassie brilliant blue. Purified proteins were concentrated in a Vivaspin® 20 (MWCO 30

000) centrifugal concentrator (GE) and quantified by absorption at 280 nm (Nanodrop Lite, ThermoScientific). RNA concentration was analysed by absorption at 260 nm. PPR10:*atpH* complex was prepared by briefly incubating a 1:1 molar ratio of purified protein with RNA (5′-GUAUCCUUAACCAUUUC-3′ [IDT]).

Circular dichroism analysis

Purified PPR10 protein at a concentration of 0.1 mg ml^{-1} was dialysed into 10 mM potassium phosphate at pH 8.0, 100 mM KF, and analysed at 20°C in a 1 mm path length quarz cuvette on a JASCO J-810 spectropolarimeter. Measurements from 260 to 185 nm with wavelength steps of 1 nm were conducted in triplicate to facilitate the generation of error bars denoting standard deviation. Estimations of α-helix content were made using characteristic signals at 208 and 222 nm:$\%\alpha_{208} = [(|[\Theta]_{208}| - 4000) / (33\,000 - 4000)]$ x 100 and $\%\alpha_{222} = [(|[\Theta]_{222}| - 3000) / (36\,000 - 3000)]$ x 100 ([32,33]).

Small-angle X-ray scattering analysis

SAXS measurements were conducted at the SAXS/WAXS beamline of the Australian Synchrotron with continuous data collection on a 1 M Pilatus detector ([34]). SEC-SY-SAXS was controlled by a Shimadzu HPLC system. For the PPR10:*atpH* complex, scattering data were collected from 50 μl of a 6 mg ml^{-1} sample loaded onto a Superdex200 Precision column (GE healthcare) at 0.1 ml min^{-1} in 50 mM Tris-HCl at pH 7.3, 250 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol. For the MBP-PPR10:*atpH* complex, 90 μl of a 6 mg ml^{-1} sample was run through the Precision column at 0.5 ml min^{-1} in Buffer A. For experiments where timed shutter opening was critical, the delay between the FPLC UV detector and X-ray detector was calibrated using glucose isomerase solution as a standard. Scattering data from 2 s exposures were background corrected and averaged using scatterBrain. Data were corrected for fouling by aggregates while the shutter was open, by using linear interpolation of background from averaged frames prior to the peak to averaged frames after the peak, in an analogous approach to that used in US-SOMO ([35]). A high-q cutoff of 0.21 was used due to increasing noise beyond this resolution. The radius of gyration (R_g), maximum dimension (D_{max}) and $P(r)$ distribution plots of the samples were determined with the ATSAS software ([36]) using PRIMUS ([37]) and GNOM ([38]) respectively. Initially, the automated D_{max} determination by GNOM was used. However, we chose to manually increase D_{max} as it was underestimated judging by the appearance of the $P(r)$ plot (discussed below).

Coordinates were manipulated with PDB-MODE ([39]). Rigid-body modeling was performed using SASREF ([40]). The starting model was derived from a single chain of PPR10 extracted from PDB 4M59 ([30]), broken into seven 'domains' each consisting of three PPR motifs. RNA was included in the domain definition by placing the appropriate trinucleotide adjacent to the predicted RNA-binding residues of each domain. For the MBP-PPR10 data, PDB entry 1MBP was used as an additional protein domain. Two distance constraints per domain were added to ensure physically reasonable modeling, as the PPR motifs in PPR10
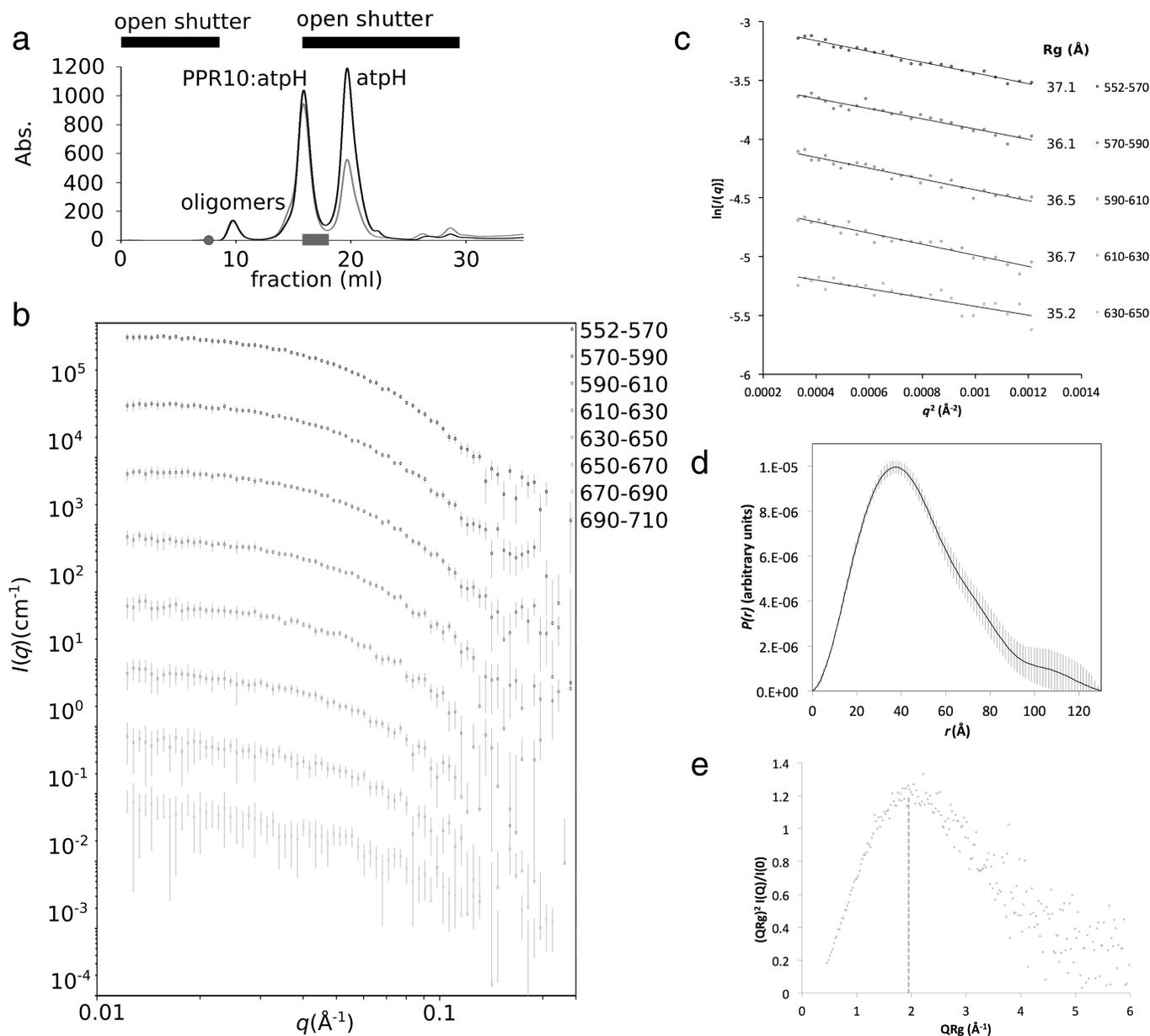
Figure 1. SEC-SY-SAXS of PPR10:atpH. **(a)** Size exclusion chromatography trace of PPR10:*atpH* showing peaks corresponding to oligomers, PPR10:*atpH* complex and unbound *atpH* RNA. Black bars indicate the part of the chromatogram for which SAXS was collected by opening the beamline shutter. The grey circle indicates position where the first data frame was recorded, and the bar, the region corresponding to frames 552–710. **(b)** Scattering profiles of the PPR10:*atpH* complex averaged across 20 frame segments, corresponding to the second half of the PPR10:*atpH* peak (grey bar in (a)). **(c)** Guinier analysis of the low-*q* scattering data for the five most strongly scattering 20 frame segments indicates a radius of gyration of 36 Å. **(d)** $P(r)$ distribution for frames 552–710 indicates a maximum dimension of 125 Å. **(e)** Dimensionless Kratky plot is diagnostic of a globular but somewhat extended protein.

form a continuous array. CRYSOL (41) and MOLEMAN2 (42) were used to evaluate model quality by quantifying how representative the atomic models are of the data and $P(r)$ distribution plots. Molecular graphics were created with Py-MOL (Schroedinger LLC).

RESULTS

We have generated a plausible model of the solution structure of ZmPPR10 in complex with the 17 nucleotide minimal binding site of the maize *atpH* transcript informed

by a combination of SAXS data, circular dichroism spectroscopy, and pre-existing bioinformatic, biochemical and crystallographic data. The secondary structure content of PPR10 was estimated to be predominantly α-helical (84%) from circular dichroism as expected for the anti-parallel α-helical topology of a PPR motif, and comparable with calculations based on the crystal structures (71% α-helix). Preliminary SAXS measurements on PPR10 alone and after incubation with *atpH* RNA yielded complex uninterpretable scattering profiles due to significant aggregation. We there-
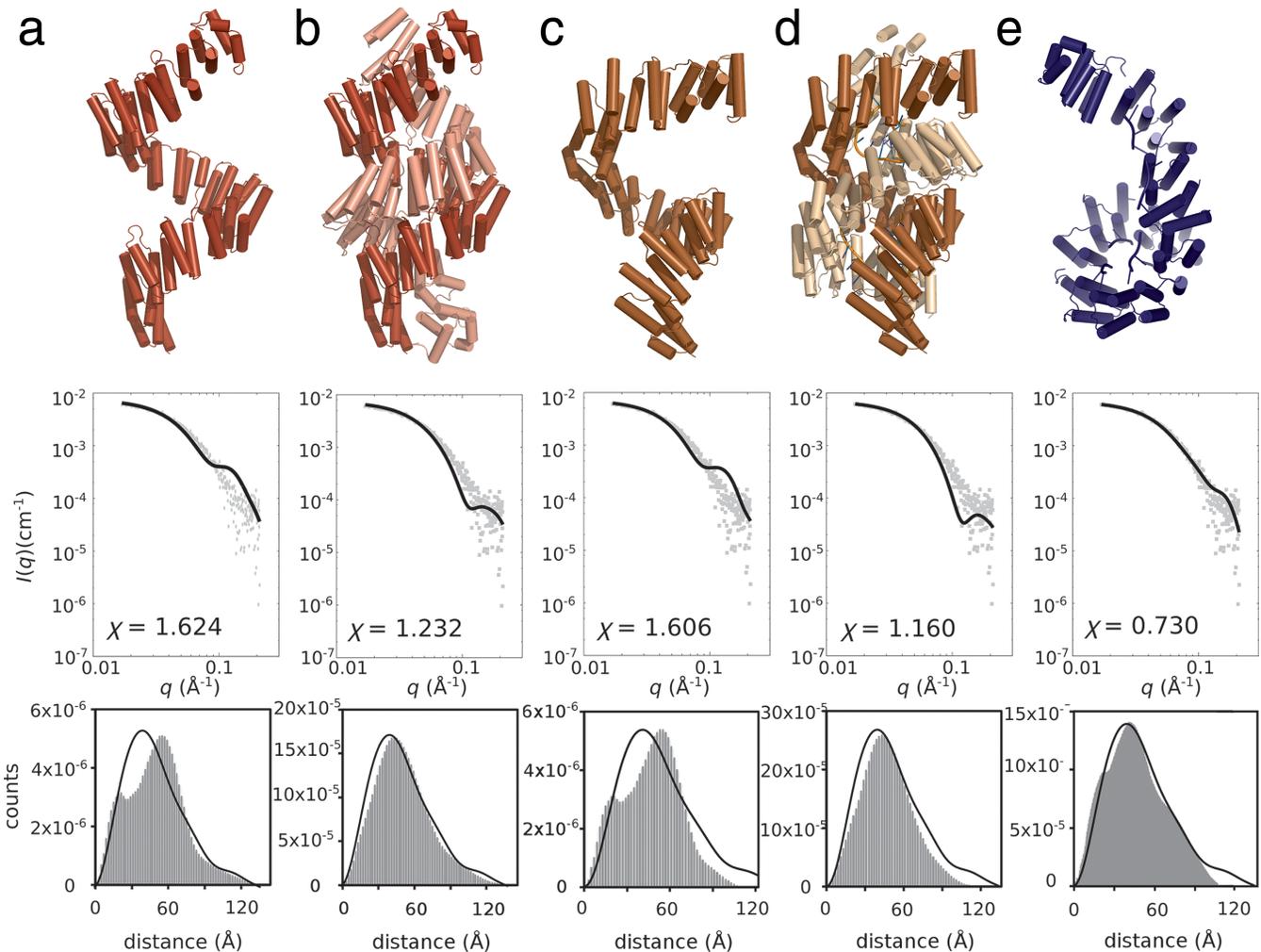
Figure 2. Validation of candidate molecular models against PPR10:*atpH* scattering data. Top row: Molecular model. Middle row: Optimized fit to data. Bottom row: Reconstructed pair distribution plots from the respective models (solid histogram) compared with the experimentally derived *P(r)* distribution (black line). **(a)** apoPPR10 monomer extracted from PDB 4M57; **(b)** apoPPR10 dimer from PDB 4M57; **(c)** PPR10:*psaJ* monomer extracted from PDB M459; **(d)** PPR10:*psaJ* dimer from PDB 4M59; **(e)** The best-fitting SASREF model. Fitting parameters indicate a considerably better fit for the SASREF model.

fore employed size-exclusion chromatography-coupled synchrotron small-angle X-ray scattering (SEC-SY-SAXS; Figure 1). This method allowed us to separate the various complexes inline, identify their content from their A260/A280 ratios and examine the solution structure of specific SEC fractions (Figure 1a). A significant challenge to this method was encountered due to the earliest eluting aggregated proteins adhering to the sample capillary at the point of X-ray exposure, thus tainting all further images. By carefully timing the opening of the X-ray shutter to only measure regions of interest and appropriate background segments it was possible to measure high quality SAXS on the protein:RNA complex. Scattering data were obtained for a monodisperse peak with an A260/A280 ratio and molecular size commensurate with a monomeric PPR10:*atpH* complex (Figure 1b). Data collection and analysis are summarized in Tables 1 and 2, respectively.

PPR10:*atpH* displayed a stable scattering profile for a globular protein throughout the SEC peak. Guinier anal-

ysis of the low-*q* data indicated a consistent radius of gyration (R_g) close to 36 Å for the most concentrated fractions (Table 2, Figure 1c). An upturn at low-*q* became evident in the scattering profile during later fractions, possibly due to adhesion of protein on the capillary (Figure 1b), rendering Guinier analysis of these lower concentration samples impossible. Thus, data from the peak fraction (frames 552–570) were averaged and used for subsequent analysis.

The *P(r)* distribution (Figure 1d) is indicative of a globular structure with an extended tail: an asymmetric distribution with a maximum at 40 Å with a shoulder at 20 Å and a steady decline through the longer distances with shoulders at 80 and 110 Å and a D_{max} of ca. 125 Å. A dimensionless Kratky plot contains a peak at 1.95 with a height of around 1.2 indicating a somewhat elongated, but generally globular protein (with reference to (43)) that is suitable for Porod analysis (Figure 1e). The molecular mass (M_r) of the scattering particle, determined from the Porod invariant (V_p),
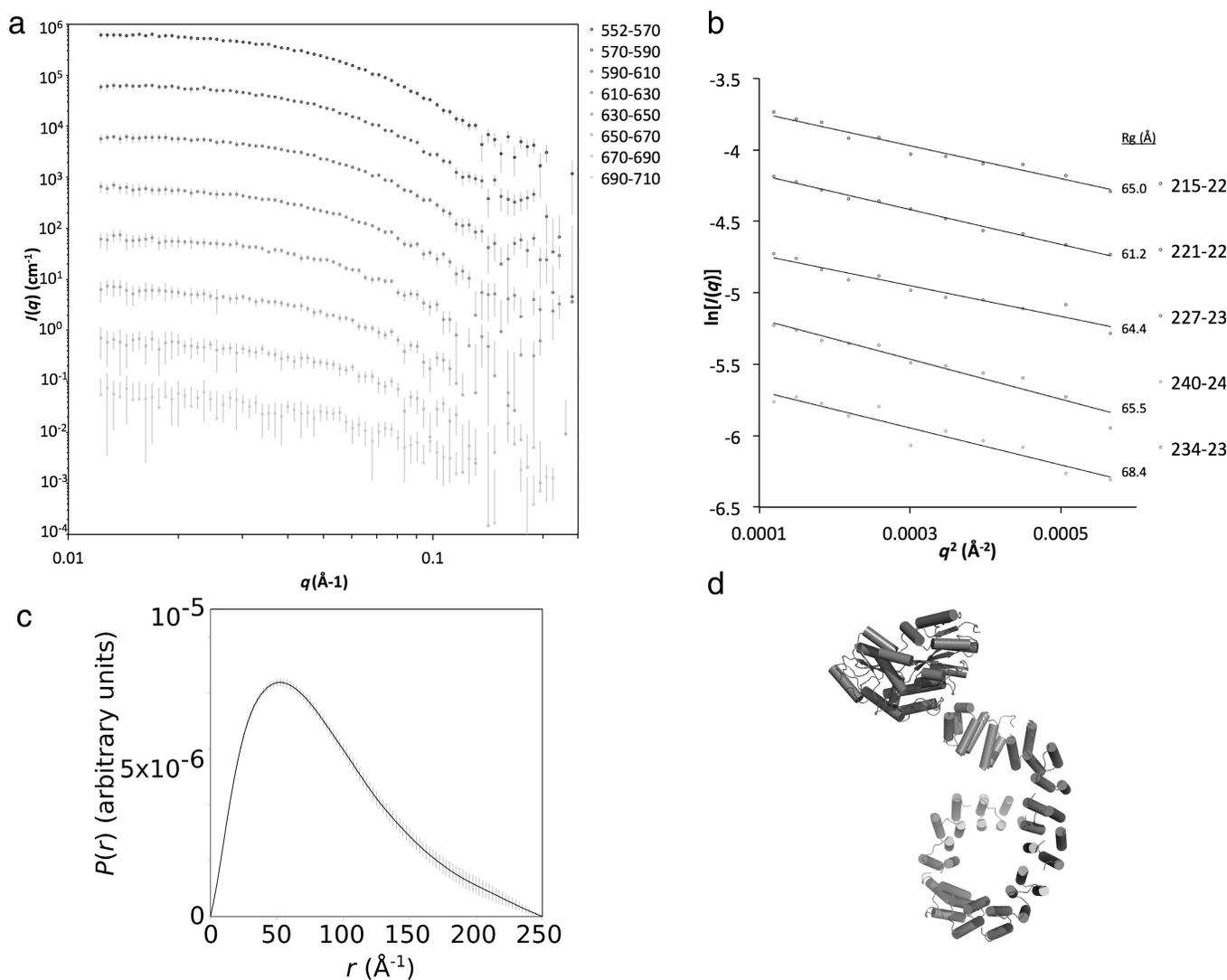
Figure 3. SEC-SY-SAXS analysis of MBP-PPR10 fusion protein:*atpH* complex. **(a)** Scattering profiles of averaged and background subtracted five-frame region of the peak fraction of MBP-PPR10:*atpH*. **(b)** Guinier analysis and **(c)***P(r)* distribution for the 215–220 averaged scattering profile. **(d)** The best-fitting SASREF model for the MBP fusion protein.

(44), corresponds to a monodisperse, 1:1 PPR10:*atpH* complex (Tables 1 and 2).

As the monomeric state of the complex as determined by AUC, REMSA and now SAXS contrasts with the reported dimeric crystal structure, we next sought to build a molecular model—taking into account the limitations of data quality—that best reflects the scattering data. We first examined the crystal structures of ZmPPR10 in the absence (4M57) and presence (4M59) of *psaJ* RNA. Initially, we generated candidate models of the observed dimers, and monomers extracted from the crystal structures, and compared their minimized fit with the observed scattering profiles, and compared their distance distributions with the observed *P(r)* distribution (Figure 2). The relatively high χ values combined with the lack of common features in the distance distribution (e.g. a maximum at 60 Å compared to 40 Å) led us to conclude that there were notable differences be-

tween the solution structure and the crystallographic models.

Armed with this knowledge we generated rigid body models from PPR10 broken into seven domains, with RNA fragments attached, and refined against the scattering data. Highly consistent results were obtained over multiple independent modeling runs for PPR10:*atpH* (Figure 2e) presented as a monomeric complex. The model has an excellent χ fitting parameter (0.73). The best-fitting model is a right-handed superhelical array with the C-terminal region forming a toroid. However, the pseudosymmetric nature of the PPR10 molecule—a superhelical array—made it impossible to objectively assign either end as the N- or C-terminus.

To address this problem, we repeated the data collection procedure with an N-terminal MBP-PPR10 fusion protein. Scattering data were analysed (Figure 3) revealing increases of R_g to 61 Å and D_{max} to 250 Å, as expected from addition of an MBP domain. Rigid body modeling of this complex
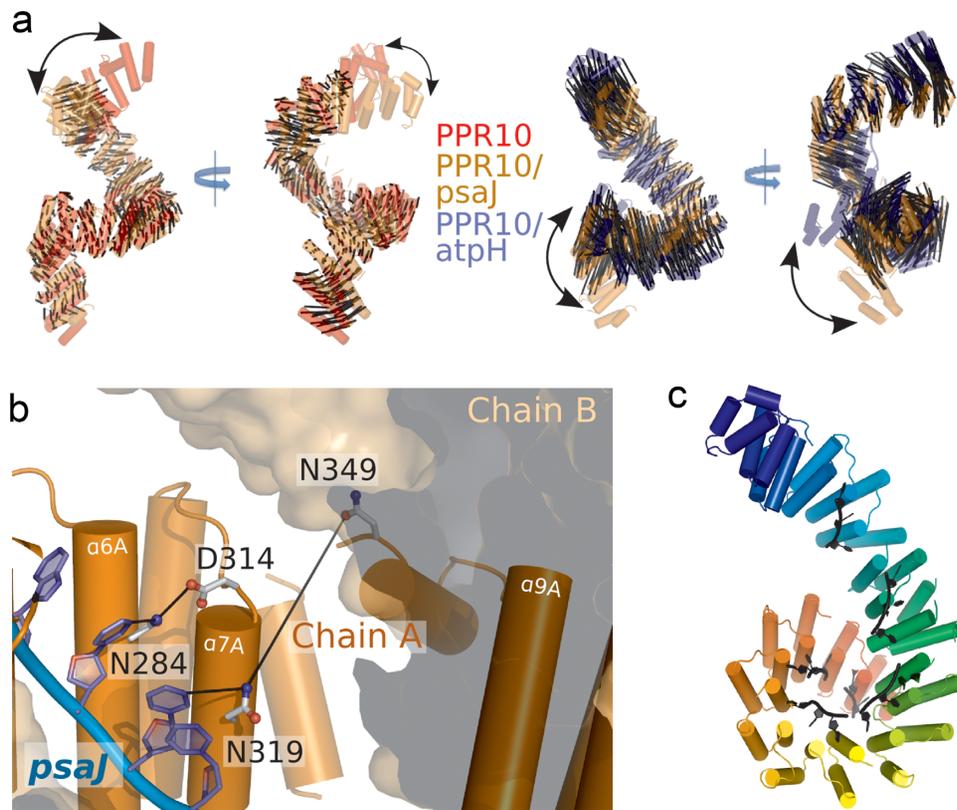
Figure 4. Global conformation of PPR10. **(a)** Dynamic structure of a PPR10 monomer:A single PPR10 molecule extracted from the RNA-free dimer structure (4M57 chain A, red cartoon), the PPR10:*psaJ* dimer (4M59 chain A, gold) and the monomeric PPR10:*atpH* solution structure (this work, blue) can be superimposed effectively (RMSD ranging from 7 to 11 Å including Cα atoms from most of the structure: black lines link superimposed Cα atoms between two structures). The N-terminal region (oriented at the top) differs between apo- and *psaJ*-bound proteins, while only the C-terminus differs significantly between *psaJ*- and *atpH*-bound proteins. Global differences between superposed structures are indicated by black arrows. **(b)***PsaJ* RNA coordination by PPR10 motifs 6 and 7 in the crystal structure (4M59; chain A orange cartoon; chain B, beige/gray surface). Residues Asn284, Asp314 and Asn319, Asn349 (positions 6 and 1′ for motifs 6 and 7 respectively) are known to dictate RNA specificity, however Asn349 is displaced from RNA binding by the dimeric interface. **(c)** The solution structure of PPR10:*atpH* illustrates that the arrangement of the PPR motifs (colored blue to red from N-terminus) is conducive to ssRNA binding (black cartoon) as a monomer (Note: breaks in RNA chain are a consequence of the rigid body modeling approach used).

Table 1. SAXS data collection and refinement statistics

	PPR10:*atpH*	MBPPPR10:*atpH*
Data collection		
Instrument	Australian synchrotron SAXS/WAXS beamline	Australian synchrotron SAXS/WAXS beamline
Beam geometry (μm)	120	120
Wavelength (Å)	1.033	1.033
q range (A^{-1})	0.01–0.21	0.01–0.21
Exposure per frame (s)	2	2
Flow-rate (ml min^{-1})	0.1	0.5
Concentration (mg ml^{-1})	6–0	6–0
Temperature (K)	288	288
Structural parameters		
$I(0)$ (cm^{-1}) [from $P(r)$]	0.007 $\pm$ 0.00	0.01 $\pm$ 0.00
R_g (Å) [from $P(r)$]	36.19 $\pm$ 1.08	63.2 $\pm$ 0.01
$I(0)$ (cm^{-1}) (from Guinier)	0.008 $\pm$ 0.00	0.009 $\pm$ 0.00
R_g (Å) (from Guinier)	37.39 $\pm$ 2.30	61.20.78 $\pm$ 1.87
D_{max} (Å)	125	250
Molecular-mass determination		
Porod invariant (V_p)(Å^3)	165,328	203,698
Estimated M_r from V_p (Da)	97,251	161,900
Calculated M_r (Da)	87,882	135,607

Table 2. Guinier and Porod analysis of peak fractions from PPR10:atpH and MBP-PPR10:atpH SEC-SY-SAXS

Frame no.	$I(0)$	Radius of gyration, R_g (Å)	Guinier range (q)	Maximum dimension, D_{max} (Å)	$P(r)$ Range (q)	Porod invariant, V_p (Å^3)	Molecular mass, M_r, from V_p (Da)	Oligomeric state from M_r
552–570	0.007 ± 0.000	37.394 ± 2.299	0.014–0.034	130.1	0.014–0.25	165,328	97,251	1.11
570–590	0.006 ± 0.000	36.055 ± 3.811	0.014–0.034	126.2	0.014–0.25	162,893	95,819	1.09
590–610	0.005 ± 0.000	36.456 ± 3.967	0.014–0.034	127.6	0.014–0.25	165,752	97,501	1.11
610–630	0.004 ± 0.000	36.773 ± 5.734	0.014–0.034	126.4	0.014–0.25	158,006	92,944	1.06
630–650	0.003 ± 0.000	35.034 ± 9.020	0.014–0.034	123.9	0.014–0.25	187,843	110,495	1.26
650–670	0.002 ± 0.000	38.532 ± 5.436	0.014–0.034	130.0	0.014–0.25	162,642	95,671	1.09
215–220	0.012 ± 0.000	67.459 ± 6.188	0.012–0.025	250.0	0.012–0.35	210,905	124,061	0.94
221–226	0.012 ± 0.000	61.199 ± 1.868	0.012–0.025	250.0	0.012–0.35	203,698	135,607	1.03
227–233	0.007 ± 0.000	55.053 ± 1.385	0.012–0.025	250.0	0.012–0.35	204,390	143,800	1.09
234–239	0.005 ± 0.000	54.975 ± 2.248	0.012–0.025	250.0	0.012–0.35	184,072	129,100	0.98
240–245	0.003 ± 0.000	55.947 ± 3.515	0.012–0.025	250.0	0.012–0.35	160,763	112,200	0.85

Forward scatter, $I(0)$, radius of gyration, R_g determined using PRIMUS. Maximum dimension, D_{max}, and Porod volume, V_p, and molecular mass, M_r, determined using GNOM. Oligomeric state (n) determined from the ratio of $M_r(V_p)$ to M_r (theoretical).

produced a satisfactory solution ($\chi = 0.78$) for a protein with MBP at the N-terminus and a toroidal arrangement at the C-terminus, thus supporting our model of the untagged protein, which is used in the following comparative analysis.

Alignment of monomeric PPR10 structures from three different available structural contexts (a monomer extracted from the apo-PPR10 dimer (PDB 4M57); a monomer extracted from the *psaJ*-bound dimer (PDB 4M59) and the monomeric *atpH*-bound solution structure, this work) reveals two notable regions of structural rearrangement (Figure 4a) that are also discernable from the scattering profiles and distance distributions (Figure 2). While the C-terminal region of the apo- and *psaJ*-bound monomers aligns well, the four N-terminal PPR motifs are arranged differently. Conversely, the *psaJ*- and *atpH*-bound conformations only differ significantly around the four C-terminal PPR motifs.

DISCUSSION

This study shows how the coupling of size-exclusion chromatography with synchrotron SAXS allows productive study of challenging macromolecular complexes, particularly when studying a well-characterized system with ample orthogonal information, from genomic, biochemical and crystallographic sources.

We have determined that the solution structure of ZmPPR10 complexed with a fragment of *atpH* is not compatible with the dimeric form observed in the PPR10:*psaJ* crystal structure, thus confirming previous analytical ultracentrifugation analyses. The nature of recognition of *atpH* and *psaJ* has been well characterized (6,13,19), and the PPR–RNA recognition code has been validated biochemically for the N-terminus/5′ end of the PPR–RNA duplex by mutating residues Asn284, Asp314 and Asn319, Asn349 in PPR motifs 6 and 7, and observing a predictable change in binding specificity (19). The dimeric conformation observed in the *psaJ*-bound PPR10 structure is not fully compatible with this result, as one of the key residues in motif 7 does not engage the RNA (Figure 4b). Indeed the structure of the helix bearing Asn349 is highly disturbed, presenting considerably different positions in the two subunits of the crystal structure: in both cases Asn319 is well over 10 Å distant from the RNA. It seems likely that the presence of the other subunit of the dimer close to this region is a cause of the disruption of this binding site (Figure 4b).

In the SAXS structure, the more highly overwound superhelix of the C-terminal region would sterically hinder dimerisation of the protein (Figure 4c). In this tight, toroidlike part of the structure there are intramolecular interactions between residues of the C-terminal repeats and repeats earlier in the sequence, which in the dimeric crystal structure are involved in intermolecular interactions instead. Our results are consistent with the 1:1 stoichiometry established by analytical ultracentrifugation for a PPR10-*atpH* complex (19). That a similar architecture would apply to the PPR10-*psaJ* complex in solution is supported by the comigration of PPR10-*atpH* complexes and PPR10-*psaJ* complexes in native gels and by the extensive sequence conservation across the length of the *psaJ* and *atpH* RNAs.

A question is raised about whether all of these unambiguously observed configurations can be correct. The answer may lie in the necessarily convoluted dynamic process that is required for PPR10 to dissociate from a dimeric state to a monomeric state and to bind sequence specifically to RNA in a cellular context involving other binding partners and RNA secondary structures (6).

It is unclear whether the dimeric PPR10:*psaJ* structure observed crystallographically represents an intermediate state in the dynamic process of RNA recognition by PPR10, or if it represents a dead-end complex formed at the high concentrations required for crystallisation. At a late point in our analysis, a paper became available that raises further doubt regarding the dimeric nature of the complex (45). Here, the authors of the PPR10 crystal structures determine that an N-terminal truncation and a point mutation that were essential for crystallogenesis in fact drive dimerization. Our wild-type PPR10 protein makes a more physiologically relevant model for the interaction between PPR10 and target RNA, taking into account the miniscule cellular concentration of PPR10 (with respect to the micromolar dissociation constant for dimerisation) and the large excess of 'nonspecific' single-stranded RNA. While it remains plausible that dimeric apoPPR10 binds to an RNA transcript in low-affinity mode causing a structural rearrangement that dissociates the dimer and allows the high-affinity monomeric complex to form, it is likely that in the cell a monomeric PPR10 interacts with bulk single-stranded RNA before exchanging partners to achieve high-affinity binding to its target RNA motif as a monomeric protein:RNA complex similar to the solution structure presented here.

Despite their well-described modularity and relatively predictable RNA-binding specificity, this example of a native PPR protein reveals a potentially convoluted process

of RNA binding. If the promise for PPR proteins to be applied as ssRNA-binding tools in biotechnology is to be maximized, a fuller understanding of the detailed structural changes that occur on RNA binding is essential.

REFERENCES

1. Small,I.D. and Peeters,N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.*, **25**, 46–47.
2. Barkan,A. and Small,I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.*, **65**, 415–442.
3. Lurin,C., Andres,C., Aubourg,S., Bellaoui,M., Bitton,F., Bruyere,C., Caboche,M., Debast,C., Gualberto,J., Hoffmann,B. *et al.* (2000) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089–2103.
4. Beick,S., Schmitz-Linneweber,C., Williams-Carrier,R., Jensen,B. and Barkan,A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell. Biol.*, **28**, 5337–5347.
5. Choquet,Y. (2009) 5′ and 3′ ends of chloroplast transcripts can both be stabilised by protein 'caps': a new model for polycistronic RNA maturation. *EMBO J.*, **28**, 1989–1990.
6. Prikryl,J., Rojas,M., Schuster,G. and Barkan,A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 415–420.
7. Okuda,K., Nakamura,T., Sugita,M., Shimizu,T. and Shikanai,T. (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. *J. Biol. Chem.*, **281**, 37661–37667.
8. Williams-Carrier,R., Kroeger,T. and Barkan,A. (2008) Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. *RNA*, **14**, 1930–1941.
9. Delannoy,E., Stanley,W.A., Bond,C.S. and Small,I.D. (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. *Biochem. Soc. Trans.*, **35**, 1643–1647.
10. Davies,S.M., Lopez Sanchez,M.I., Narsai,R., Shearwood,A.M., Razif,M.F., Small,I.D., Whelan,J., Rackham,O. and Filipovska,A. (2012) MRPS27 is a pentatricopeptide repeat domain protein required for the translation of mitochondrially encoded proteins. *FEBS Lett.*, **586**, 3555–3561.
11. Zoschke,R., Kroeger,T., Belcher,S., Schottler,M.A., Barkan,A. and Schmitz-Linneweber,C. (2012) The pentatricopeptide repeat-SMR protein ATP4 promotes translation of the chloroplast atpB/E mRNA. *Plant J.*, **72**, 547–558.
12. Zoschke,R., Qu,Y., Zubo,Y.O., Borner,T. and Schmitz-Linneweber,C. (2013) Mutation of the pentatricopeptide repeat-SMR protein SVR7 impairs accumulation and translation of chloroplast ATP synthase subunits in Arabidopsis thaliana. *J. Plant Res.*, **126**, 403–414.
13. Pfalz,J., Bayraktar,O.A., Prikryl,J. and Barkan,A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.*, **28**, 2042–2052.
14. Barkan,A., Walker,M., Nolasco,M. and Johnson,D. (1994) A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.*, **13**, 3170–3181.
15. Bentolila,S., Alfonso,A.A. and Hanson,M.R. (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 10887–10892.
16. Akagi,H., Nakamura,A., Yokozeki-Misono,Y., Inagaki,A., Takahashi,H., Mori,K. and Fujimura,T. (2004) Positional cloning of the rice Rf-1 gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. *Theor. App. Gen.*, **108**, 1449–1457.
17. Gutierrez-Marcos,J.F., Dal Pra,M., Giulini,A., Costa,L.M., Gavazzi,G., Cordelier,S., Sellam,O., Tatout,C., Paul,W., Perez,P. *et al.* (2007) empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. *Plant Cell*, **19**, 196–210.
18. Cushing,D.A., Forsthoefel,N.R., Gestaut,D.R. and Vernon,D.M. (2005) Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta*, **221**, 424–436.
19. Barkan,A., Rojas,M., Fujii,S., Yap,A., Chong,Y.S., Bond,C.S. and Small,I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.*, **8**, 1002910.
20. Takenaka,M., Zehrmann,A., Brennicke,A. and Graichen,K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One*, **8**, e65343.
21. Yagi,Y., Hayashi,S., Kobayashi,K., Hirayama,T. and Nakamura,T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*, **8**, e57286.
22. Filipovska,A. and Rackham,O. (2013) Pentatricopeptide repeats: modular blocks for building RNA-binding proteins. *RNA Biol.*, **10**, 1426–1432.
23. Yagi,Y., Nakamura,T. and Small,I. (2014) The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J.*, **78**, 772–782.
24. Fujii,S., Bond,C.S. and Small,I.D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 1723–1728.
25. Ringel,R., Sologub,M., Morozov,Y.I., Litonin,D., Cramer,P. and Temiakov,D. (2011) Structure of human mitochondrial RNA polymerase. *Nature*, **478**, 269–273.
26. Howard,M.J., Lim,W.H., Fierke,C.A. and Koutmos,M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16149–16154.
27. Ban,T., Ke,J., Chen,R., Gu,X., Tan,M.H.E., Zhou,X.E., Kang,Y., Melcher,K., Zhu,J.-K. and Xu,H.E. (2013) Structure of a PLS-class pentatricopeptide repeat protein provides insights into mechanism of RNA recognition. *J. Biol. Chem.*, **288**, 31540–31548.
28. Gully,B.S., Shah,K.R., Lee,M., Shearston,K., Smith,N.M., Sadowska,A., Blythe,A.J., Bernath-Levin,K., Stanley,W.A., Small,I.D. *et al.* (2015) The design and structural characterisation of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. D*, **71**, doi:10.1107/S1399004714024869.
29. Ke,J., Chen,R.Z., Ban,T., Zhou,X.E., Gu,X., Tan,M.H., Chen,C., Kang,Y., Brunzelle,J.S., Zhu,J.K. *et al.* (2013) Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.*, **20**, 1377–1382.
30. Yin,P., Li,Q., Yan,C., Liu,Y., Liu,J., Yu,F., Wang,Z., Long,J., He,J., Wang,H.W. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **504**, 168–171.
31. Beckham,S.A., Brouwer,J., Roth,A., Wang,D., Sadler,A.J., John,M., Jahn-Hofmann,K., Williams,B.R.G., Wilce,J.A. and Wilce,M.C.J. (2013) Conformational rearrangements of RIG-I receptor on formation of a multiprotein:dsRNA assembly. *Nucleic Acids Res.*, **41**, 3436–3445.
32. Greenfield,N.J. and Fasman,G.D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry*, **8**, 4108–4116.
33. Morrisett,J.D., David,J.S.K., Pownall,H.J. and Gotto,A.M. (1973) Interaction of an apolipoprotein (apoLP-alanine) with phosphatidylcholine. *Biochemistry*, **12**, 1290–1299.
34. Kirby,N.M., Mudie,S.T., Hawley,A.M., Cookson,D.J., Mertens,H.D.T., Cowieson,N. and Samardzic-Boban,V. (2013) A low-background-intensity focusing small-angle X-ray scattering undulator beamline. *J. Appl. Crystallogr.*, **46**, 1670–1680.

35. Brookes,E., Perez,J., Cardinali,B., Profumo,A., Vachette,P. and Rocco,M. (2013) Fibrinogen species as resolved by HPLC-SAXS data processing within the *UltraScan Solution Modeler* (*US-SOMO*) enhanced SAS module. *J. Appl. Crystallogr.*, **46**, 1823–1833.

36. Petoukhov,M.V., Franke,D., Shkumatov,A.V., Tria,G., Kikhney,A.G., Gajda,M., Gorba,C., Mertens,H.D.T., Konarev,P.V. and Svergun,D.I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **45**, 342–350.

37. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.*, **36**, 1277–1282.

38. Semenyuk,A.V. and Svergun,D.I. (1991) GNOM - a program package for small-angle scattering data processing. *J. Appl. Crystallogr.*, **24**, 537–540.

39. Bond,C.S. (2003) Easy editing of Protein Data Bank formatted files with EMACS. *J. Appl. Crystallogr.*, **36**, 350–351.

40. 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.

41. Svergun,D., 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.

42. Kleywegt,G. J. and Jones,T. A. (1997) Model building and refinement practice. *Methods Enzymol.*, **277**, 208–230.

43. Durand,D., Vives,C., Cannella,D., Perez,J., Pebay-Peyroula,E., Vachette,P. and Fieschi,F. (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol.*, **169**, 45–53.

44. Fischer,H., de Oliveira Neto,M., Napolitano,H.B., Polikarpov,I. and Craievich,A.F. (2010) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. *J. Appl. Crystallogr.*, **43**, 101–109.

45. Li,Q., Yan,C., Xu,H., Wang,Z., Long,J., Li,W., Wu,J., Yin,P. and Yan,N. (2014) Examination of the dimerization states of the single-stranded RNA recognition Protein Pentatricopeptide Repeat 10 (PPR10). *J. Biol. Chem.*, **289**, 31503–31512.