Frameshift PQBP-1 mutants K192S^{fs*7} and R153S^{fs*41} implicated in X-linked intellectual disability form stable dimers

Running title: Dimerisation of PQBP-1 XLID mutants

Shah Kamranur Rahman[a,1], Hitoshi Okazawa (岡澤 均)[b], Yu Wai Chen (陳友維)[a,2,*]

[a]Randall Centre for Cell and Molecular Biophysics, King's College London, Guy's Campus, London SE1 1UL, United Kingdom

[b]Department of Neuropathology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

[1]Present address: Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom; email: shah.rahman@lshtm.ac.uk

[2]Present address: Department of Applied Biology and Chemical Technology, Hong Kong Polytechnic University, Hunghom, Hong Kong.

*To whom correspondence should be addressed. Email: yu-wai.chen@polyu.edu.hk, tel. number.: +852 3400 8660

Abstract

Polyglutamine tract-binding protein-1 (PQBP-1) is a nuclear intrinsically disordered protein playing important roles in transcriptional regulation and RNA splicing during embryonic and postembryonic development. In human, its mutations lead to severe cognitive impairment known as the Renpenning syndrome, a form of X-linked intellectual disability (XLID). Here, we report a combined biophysical study of two PQBP-1 frameshift mutants, K192S^{fs*7} and R153S^{fs*41}. Both mutants are dimeric in solution, in contrast to the monomeric wild-type protein. These mutants contain more folded contents and have increased thermal stabilities. Using small-angle X-ray scattering data, we generated three-dimensional envelopes which revealed their overall flat shapes. We also described each mutant using an ensemble model based on a native-like initial pool with a dimeric structural core. PQBP-1 is known to repress transcription by way of interacting with the C-terminal domain of RNA polymerase II, which consists of 52 repeats of a consensus heptapeptide sequence YSPTSPS. We studied the binding of PQBP-1 variants to the labelled peptide which is phosphorylated at positions 2 and 5 (YpSPTpSPS) and found that this interaction is significantly weakened in the two mutants.

Keywords: Renpenning syndrome; intrinsically disordered protein; SAXS, natively unfolded; dimerization

Abbreviations: PQBP-1: Polyglutamine tract-binding protein-1; CTD: C-terminal domain; PRD: polar-amino-acid-rich domain; XLID: X-linked intellectual disability; AUC: analytical ultracentrifugation; CD: circular dichroism; SAXS: small angle X-ray scattering; EOM: ensemble optimisation method; NSD: normalized spatial discrepancy

INTRODUCTION

Polyglutamine tract-binding protein-1 (PQBP-1) is expressed mainly in neurons throughout the brain, with abundant levels in cerebellar cortex and hippocampus (Kunde et al., 2011; Qi et al., 2005; Waragai et al., 1999). The expression level of PQBP-1 in the brain is the highest at birth which gradually decreases towards adulthood, suggesting its importance in early development. The cellular interactors of PQBP-1 infer that it plays important roles in transcription regulation and RNA processing. Among these, PQBP-1 binds to the phosphorylated C-terminal domain (CTD) of RNA polymerase II (pol II), which is a platform with which many transcription regulatory proteins interact (Egloff and Murphy, 2008; Okazawa et al., 2002; Waragai et al., 1999). PQBP-1 also interacts with many components of the splicing machinery (Iwasaki and Thomsen, 2014; Mizuguchi et al., 2014; Wang et al., 2013; Waragai et al., 2000). The presence of PQBP-1 in RNA granules and early spliceosome further confirms its critical roles in pre-mRNA processing and transportation (Kunde et al., 2011; Wang et al., 2013).

PQBP-1 is 265 amino acids in size and consists mainly of three domains: the WW domain, the polar-amino-acid-rich domain (PRD) and the C-terminal region (Fig. 1A). WW, the only folded domain, is a small three-stranded antiparallel β-sheet structural motif commonly found in a diverse range of proteins of cell signaling and transcription regulation (Hu et al., 2004; Otte et al., 2003). This domain, in other contexts, has been implicated in many human diseases including muscular dystrophy, cancer, hypertension, Alzheimer's and Huntington's diseases (Hu et al., 2004). It is the WW domain (residues 48-81) of PQBP-1 that mediates its interaction with RNA pol II CTD (Okazawa et al., 2002; Waragai et al., 1999). The isolated PQBP-1 WW domain is a transcription activator, whereas the full-length protein represses transcription (Komuro et al., 1999).

The PRD (residues 104-163) has five consecutive copies of the heptapeptide motif DRXH (D/E) KX, then three copies of (D/E) R, a small stretch of intervening amino acids and seven copies of (D/E) R. This low complexity region of high charge density mediates binding to polyglutamine tracts (Waragai et al., 2000). The C-terminal region (residues 190-265) contains a stretch of highly conserved amino acids (Sudol et al., 2012) that interacts with spliceosome protein U5-15kD (Mizuguchi et al., 2014; Takahashi et al., 2010; Waragai et al., 2000).

The mutations in PQBP-1 have deleterious effects that lead to severe cognitive impairment and results in the Renpenning syndrome, a type of X-linked intellectual disability (XLID) (Germanaud et al., 2011). Its clinical features include mental retardation, microcephaly, short stature, facial dysmorphy, spastic paraplegia and midline defects. (Stevenson et al., 2005). In this work, the mutant protein products implicated in two particularly severe clinical manifestations were studied. The first was a 4-basepair deletion (c.459_462delAGAG) which produces a 192-residue mutant protein, p.Arg153Serfs*41 (denoted R153S^{fs*41} here); the second was a 2-basepair deletion (c.575_576delAG) that produces a 197-residue mutant protein, p.Lys192Serfs*7 (denoted K192S^{fs*7}) (Germanaud et al., 2011; Kalscheuer et al., 2003; Lenski et al., 2004). Both frameshift mutants are shortened in protein sizes by approximately 25% compared to the wild-type PQBP-1 (Fig. 1A,B). While K192S^{fs*7} is effectively PQBP-1 lacking its CTD, R153S^{fs*41} has a disrupted PRD and the rest replaced by a long non-native tail (Fig. 1B). It is worth noting that a natural splice variant, PQBP-1b/c, has a similar sequence composition to K192S^{fs*7}: consisting of residues 1-192 but has a variant tail of 32 residues (27). R153S^{fs*41} is found to be dispersed throughout the cell, presumably due to losing its nuclear localisation signal sequence (Kalscheuer et al., 2003). In order to understand the molecular basis of these diseases, we investigated the structural properties exhibited by these two mutant proteins. The natively-unfolded nature of

the wild type (Rees et al., 2012; Takahashi et al., 2009) led us to employing solution biophysical methods: analytical ultracentrifugation (AUC), circular dichroism (CD) spectroscopy and small angle X-ray scattering (SAXS). We further studied if the interactions between these mutants and RNA pol II are affected by measuring the change in fluorescence polarisation of a labelled peptide.

MATERIALS & METHODS

Construction of expression plasmids

The plasmids for producing the two XLID mutants were generated by the QuikChange mutagenesis method (Agilent Technologies) from the expression plasmid pHMGW-PQBP-1 (Rees et al., 2012) with the following sets of primers

p.Lys192Serfs*7:

5' CCTTTCGGCTTACTGCTCTTGCTCTTGGGATAG **3'**

5' CTATCCCAAGAGCAAGAGCAGTAAGCCGAAAGG **3'**

p.Arg153Serfs*41:

5' CGTTCCCTGTCTCGCTCTCTCTGTCTAC **3'**

5' GTAGACAGAGAGAGCGAGACAGGGAACG **3'**

Expression and purification

We expressed the fusion proteins in *Escherichia coli* strain Rosetta 2 DE3 (Merck). We cultured bacterial cells at 37 °C till its absorbance at 600 nm reached 0.6. Protein expression was induced by adding 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by further incubation at 18 °C for 18 hours. We then pelleted and resuspended the cells in binding buffer (20 mM Na_2HPO_4 pH 7.4, 150 mM NaCl). After cell breakage, the cleared lysate was loaded on a 5 mL HisTrap column (GE Healthcare) and purification was carried out as per manufacturer's instructions. The purified sample was dialyzed in PreScission buffer (20 mM Tris-HCl pH 7.0, 150 mM NaCl, 1 mM 1,4-dithiothreitol (DTT), 1 mM

ethylenediaminetetraacetic acid (EDTA)) and digested overnight with PreScission protease at 4 °C. The digested samples were loaded onto a HiLoad 26/600 Superdex 75 pg column (GE Healthcare) and eluted in phosphate-buffered saline (PBS, pH 7.4) at 4 °C.

Analytical ultracentrifugation

We performed sedimentation velocity experiments on a Beckman Coulter model XL-A analytical ultracentrifuge equipped with UV scanning optics. K192S^{fs*7} samples (420 µL) were prepared in PBS at concentrations of 0.06, 0.49, and 1.38 mg mL^{-1}, then the set of three samples were loaded into 12 mm double-sector cells with quartz windows along with a buffer reference (PBS). Likewise, R153S^{fs*41} samples at concentrations of 0.07, 0.29, and 0.63 mg mL^{-1} were used. Centrifugation speed was 50,000 rpm with 300 scans monitored and collected continuously at 280 nm at 20 °C.

For each sample, data from the scans were fitted to a continuous sedimentation coefficient distribution $c(s)$ in inhomogeneous solvent model with the program SEDFIT (Schuck, 2000). From the program SEDNTERP (Hayes D., 1995), we computed the buffer density (ρ = 1.0053 g mL^{-1}), buffer viscosity (η = 1.0189 cP), and partial specific volume at 20 °C of K192S^{fs*7} ($\overline{v}$ = 0.713505 cm^3 g^{-1}) and R153S^{fs*41} ($\overline{v}$ = 0.718124 cm^3 g^{-1}). We calculated the molecular mass M corresponding to each peak appearing in $c(s)$ distribution from the formula:

$$M = (sRT)/[D\ (1 - \overline{v}\,\rho\,)]$$

Where R is gas constant, T is temperature, D the diffusion coefficient of the sedimenting species characterised by sedimentation coefficient s. The Stokes radii are obtained by integration in SEDFIT.

Circular dichroism spectroscopy

CD spectra of samples at 0.30 mg mL^{-1} in CD buffer (20 mM Tris-HCl pH 7.4) were measured on a Chirascan-Plus spectrometer (Applied Photophysics Ltd., Leatherhead, UK). We collected spectra from 260–190 nm in a 1.0 mm rectangular cell at 20 °C, with a step size of 1 nm, 1 s acquisition time per point with a spectral bandwidth of 2 nm.

For the variable temperature experiment, we measured CD spectra at 0.2 mg mL^{-1} of each protein sample in a 0.5 mm-pathlength cell, with 1.42 s acquisition time. The samples were cooled to 8 °C, then heated to 90 °C at the rate of 1 °C per minute with a step size of 2 °C and then cooled to 20 °C. Melting data at 225 nm was extracted for analysis. We used the protocols of Greenfield (2006) (Greenfield, 2006) to fit the thermal denaturation data with monomer, 2-state dimer, or 3-state dimer (with monomeric or dimeric intermediates) folding/unfolding models. From non-linear regression curve fitting (Supplementary information S3), melting temperature (T_m) and standard enthalpy change of folding ($\Delta H°$) were determined. With a single melting experiment, the change in heat capacity under constant pressure, ΔC_p, was assumed to be zero to simplify the calculations. This condition is not real and will lead to an overestimation of T_m as well as an underestimation of ΔH at where the equilibrium constant, $K = 1$ (Greenfield, 2004). Nevertheless, the thermodynamic properties so quantified could be descriptively compared. Using the integrated form of the van't Hoff equation, $K = K_{Tm} \exp [\ \Delta H / R\ (1/T_m - 1/T)\ ]$, the equilibrium constants of folding at T = 298K were calculated, assuming ΔH to be independent of temperature. Then we calculated the standard free energy change of folding and standard entropy change of folding.

The data of the wild-type protein (Rees et al., 2012) was re-analysed with a 2-state monomer unfolding model (Supplementary Fig. S3.1).

SAXS experiments and data processing

SAXS data were collected at the European Molecular Biology Laboratory P12 beamline (wavelength used was 0.124 nm) at the PETRA III storage ring at Deutsches Elekronen Synchrotron (DESY), Hamburg, Germany. Scattering data were collected at 10 °C on samples in PBS and at the following concentrations: K192S^{fs*7}, at 0.73, 4.4, 5.2, and 9.2 mg mL^{-1}; R153S^{fs*41}, at 1.0, 3.0, 6.1 and 7.3 mg mL^{-1}, on a Pilatus 2M detector at a distance of 3.0 m. Each image was averaged from up to 20 frames of exposure time 0.045 s.

The ScÅtter program (v.2.3F) was employed for data processing (Rambo and Tainer, 2013). The data collected for the higher concentration samples showed signs of aggregation and thus their low-angle ranges (first 100~150 data points) were discarded. For each protein, buffer-corrected data were scaled to the highest concentration set, then averaged and normalised. The reduced data was analysed with the ATSAS suite of programs (Petoukhov et al., 2012). The forward scattering I_0 and the radius of gyration R_g were determined using the Guinier method implemented in ATSAS Data Analysis module, with data restricted to the limit of $q \times R_g = 1.1$ for intrinsically disordered proteins (Borgia et al., 2016). The molecular mass of these samples was estimated by comparing their I_0 with that of the reference, bovine serum albumin (BSA). Indirect Fourier transform on the scattering curves were calculated with GNOM (versions 4.5a and 5.0) to yield the pair distribution function, $P(r)$. GNOM was first run in automatic mode (AutoGnom) to give an estimation of the maximum particle dimension D_{max}, using a subset of the data. After the initial AutoGnom run, the analysis was extended to the full range of data up to $q = 4$ nm^{-1}. D_{max} was manually varied in steps in both directions in order to achieve a good quality $P(r)$ function (smooth curve with minimal

oscillation, no negative values and gentle approach to D_{max}) which also satisfies the constraints that $P(r)$ function is 0 at $r = 0$ and $r = D_{max}$.

SAXS structural analysis and modelling

The normalised Kratky plot (implemented in ScÅtter) was employed to compare the overall conformational states of the proteins (Durand et al., 2010). Here, $[(q\,R_g)^2 \times I_q / I_0]$ was plotted against $(q\,R_g)$, where I_q is the scattering intensity at the scattering vector, q ($q = 4\pi \sin(\theta) / \lambda$, where 2θ is the scattering angle and λ is the wavelength), and I_0 is the scattering intensity extrapolated to zero scattering angle.

The quantitative estimates of the degree of the dynamics and conformational heterogeneity were analysed using an ensemble optimisation method (EOM, version 2.0) (Bernadó et al., 2007; Tria et al., 2015) hosted by the ATSAS online web server (https://www.embl-hamburg.de/biosaxs/atsas-online/). Briefly, an initial pool of 10,000 "native-like" structures were generated, with or without a structured WW domain dimer (WW_2). Two models of WW_2 dimeric domains were prepared with MODELLER v.9 (Sali and Blundell, 1993) using PDB IDs 2DWV and 3LE4 as templates, respectively (Supplementary information S2A,B). The parts of the polypeptide chain outside the WW domain were constructed from the protein sequence represented by Cα atoms. The conformations of each residue are selected from a database of known coil parameters derived from high-resolution crystal structures. The "native-like" models were constrained to have a Cα distribution following those of disordered proteins. For each pool of initial structures, a genetic algorithm was applied to select for those leading to minimal discrepancy (assessed by the reduced χ^2 value) between the calculated scattering of the ensemble model and the experimental data.

Molecular envelope models of the two mutants were constructed using DAMMIF (20 models). The range of data used was restricted to 2.0 and 2.2 nm^{-1} for $K192S^{fs*7}$ and $R153S^{fs*41}$, respectively, according to the empirical $q_{max} \times R_g < 7\sim8$ / rule (Petoukhov et al., 2012). Simulated annealing was employed to minimize the discrepancy between calculated and experimental data. The success of reconstruction was assessed by the normalized spatial discrepancy (NSD). The most representative model was recycled into DAMMIN for a final round of refinement. Bead models were rendered using the graphics program QuteMol (Tarini et al., 2006) with the respective sphere radii.

The experimental data, $P(r)$ function output, DAMMIN model and EOM model of the wild type and the two mutants have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB: SASDED2, SASBDB: SASDET6 and SASBDB: SASDEU6).

.

Peptide binding studies

The RNA pol II CTD peptide (YpSPTpSPS) was purchased from lifetein LLC (Somerset, NJ, USA) with a fluorophore, 5-carboxytetramethylrhodamine (5-TAMRA) attached to its N-terminus. The concentration of the peptide was 10 μM for all experiments. For each protein sample, six concentrations covering up to 300 μM were used. The proteins (100 μL) and peptides (10 μL) were mixed at room temperature for 30 minutes before measurements were taken. The readings of fluorescence polarisation were recorded on a FluoroMax-4 (Horiba Scientific, Kyoto, Japan). The excitation wavelength was 594 nm whereas the emission wavelength was 577 nm. The data were analysed using a single-site specific binding model, by non-linear regression to the equation

$$P = P_{max}\, x / (K_d + x) + \text{background}$$

where P is the measured fluorescence polarization, x is the concentration of the proteins (PQBP-1 or its mutants), P_{max} is the maximum polarization corresponding to all specific binding sites occupied, K_d is the equilibrium dissociation constant of the protein-peptide complex and "background" is a constant value. Curve fitting and error calculations were performed with GraphPad Prism 7.

RESULTS

PQBP-1 mutants are dimeric species

In analytical ultracentrifugation (AUC) studies, K192S^{fs*7} showed a major peak at $S^0_{20,w}$ = 2.76 (±0.87) S from distribution analysis, which accounts for 80% of the sedimenting species and corresponds to a molecular mass of 46.2 (±0.9) kDa (Fig. 2A). A minor peak at S = 4.8 which corresponds to a molecular mass of ~120 kDa was also registered. For R153S^{fs*41}, we observed a major peak at $S^0_{20,w}$ = 2.95 (±0.70) S accounting for 90% of the sedimenting species and corresponding to a molecular mass of 43.5 (±0.8) kDa (Fig. 2B). Therefore, both K192S^{fs*7} and R153S^{fs*41} are dimers (Table 1). The frictional coefficient ratios (f / f_0) of K192S^{fs*7} and R153S^{fs*41} are 1.8 and 1.6 respectively, thus these mutants behave like elongated molecules similar to the wild type (f / f_0) = 1.74) (Rees et al., 2012).

The molecular masses obtained from SAXS analyses for both mutants were 43-46 (±5) kDa (Table 1); which agreed well with the AUC results.

PQBP-1 mutants dimerisation contribute to high thermal stability

Like the wild-type protein, the circular dichroism (CD) spectra of both mutants (Fig. 3A,C) are characteristic of samples that do not have regular secondary structures (Rees et al., 2012). The spectra lack features and have negative maxima at 200 nm (Drake et al., 1988; Siligardi and Drake, 1995). The temperature dependences of CD spectra of both mutants are also typical of disordered proteins showing broad melting curves with a shallow transition (Fig. 3A-D). For both mutants, the CD spectra at higher temperatures differ significantly from those at lower temperatures. At the wavelength we monitored denaturation (225 nm), thermal unfolding was reversible (Fig. S.3.2). The biggest CD signal changes for both mutants were

registered at around 225-230 nm. For both mutants, the negative maximum signal (~198-200 nm) increased significantly after thermal unfolding and refolding by cooling (Fig. S.3.2).

With a set of single-concentration temperature-melting data, we attempted semi-quantitative 2-state and 3-state unfolding analyses described in Supplementary section S3. The K192S^{fs*7} data fit well to a 2-state dimer folding model (Table 2). Attempts to fitting with either of the 3-state models were unsuccessful. The data of the R153S^{fs*41} mutant were best interpreted with a 3-state folding/unfolding mechanism (Tables 2 and S3.1).

With a $\Delta G°$ of –2 kJ mol^{-1}, the native state of the wild-type protein is only marginally more stable than its unfolded state. Both mutants are significantly more stable in their respective native states with more negative $\Delta G°$s (Table 2). The melting temperature of K192S^{fs*7} (Fig. 3B) is similar to that of the wild type and close to the physiological body temperature. For R153S^{fs*41}, its T_{m1} is lower and T_{m2} is higher than the body temperature (Fig. 3D). The thermal stability of the K192S^{fs*7} mutant is mainly contributed by increased enthalpy. Both the standard enthalpy and entropy changes, $\Delta H°$ and $\Delta S°$, involved in the formation of the R153S^{fs*41} dimer are much larger (eight to ten times) than those of the K192S^{fs*7} mutant.

PQBP-1 mutants are more compact than the wild type

Analysis of the low-angle region of scattering data of K192S^{fs*7} yielded a radius of gyration (R_g) of 3.83 (±0.08) nm and that of R153S^{fs*41} gave 3.63 (±0.05) nm (see Supplementary section S4 for Guinier plots). These R_g values are similar to that of the monomeric wild type — considering that the mutants are dimeric, this suggests that they are more densely packed.

The $P(r)$ analyses yielded R_g values (of both real and reciprocal spaces) that agree well with those obtained from the respective Guinier analyses (Table 1).

The ratio R_g / R_s is an indicator of the compactness of a molecule in solution. K192S^{fs*7} and R153S^{fs*41} have R_g / R_s ratios of 0.91 and 0.97 (Table 1), respectively, which are closer to the value of a compact spherical species (0.8) than that of an unstructured random coil (1.5). They both are more compact than the wild type (R_g / R_s = 1.03) (Table 1). K192S^{fs*7} has a substantially larger R_s value (4.23 nm). Our interpretation is that the mutation left it with a highly-charged (29 out of 45 residues, 64%) C-terminal tail (Fig. 1B), which promotes an extensive hydration shell, i.e. a large hydrodynamic radius. On the contrary, the non-native tail of R153S^{fs*41} (R_s = 3.71 nm) has a balanced distribution of charged and uncharged residues (Fig. 1B) that favours formation of local folded structures.

The normalized Kratky plots (Fig. 4A) for both mutants peaked at approximately $q \times R_g$ = 2.1 followed by a valley; revealing that both mutants are partially unfolded but are more folded and more compact than the wild-type protein (Fig. 4A).

The $P(r)$ functions of mutants K192S^{fs*7} and R153S^{fs*41} have asymmetric shapes (Fig. 4B), with maxima at low r followed by a long tail, which is consistent with an elongated and partially unfolded molecule, like the wild-type protein. This is consistent with the AUC results. For both mutants, the peak of their distance distribution functions are at 3.5 nm which is similar to their respective R_g and R_s values (Table 1). The maximum molecule dimension is 14 (±1) nm for K192S^{fs*7} and 13 (±1) nm for R153S^{fs*41} (Table 1, Fig. 4B). Both values are similar to that of the wild-type protein.

The knowledge that these two mutant proteins are partially folded justified the use of an ensemble optimisation method (EOM) to gain insights. We obtained the best models from

a "native-like" initial pool of structures that included two protomer sequences and a fixed dimeric model of the WW domain, although EOM did not discriminate between the two different modes of dimerisation (see Methods section). The following analysis was based on dimer Model-1 (supplementary section S2A). After optimisation, the ensemble model of K192S^{fs*7} has a χ^2 value of 1.18 (11 structures) and for R153S^{fs*41}, 1.14 (14 structures), showing that the ensemble models describe their respective experimental scattering data satisfactorily (Fig. 5A,B). The average R_g of the K192S^{fs*7} ensemble is 3.89 nm, whereas that of the R153S^{fs*41} ensemble is 3.72 nm. Both are considerably smaller than the average R_g of the initial pools of 4.3–4.4 nm (Fig. 5E,F); i.e. the optimised ensembles are more compact than the initial native pool populations. The distribution of R_g values of both K192S^{fs*7} and R153S^{fs*41} EOM models consist of one major peak whereas for the wild-type EOM model, the distribution of R_g consists of one main peak with a heterogeneous higher population (Rees et al., 2012). The average D_{max} of the K192S^{fs*7} ensemble is 13.9 nm, and that of the R153S^{fs*41} ensemble is 12.5 nm (Figs. 5C,D). The distribution of these values follows one major peak in both mutants similar to the overall distribution of the wild-type protein (Rees et al., 2012). The average R_g and D_{max} of each ensemble model agree well with the experimentally determined values (Table 1).

Solution models of PQBP-1 mutants

Since both mutants have increased compactness (i.e. they contain more folded structures), we attempted *ab initio* modeling to visualise their average conformations. The 20 DAMMIF models of K192S^{fs*7} aligned with an average normalized spatial discrepancy (NSD) value of 0.62 indicating a successful shape reconstruction. The R153S^{fs*41} models also agreed with a mean NSD value of 0.68. However, attempts to impose two-fold symmetry gave worse

results. After the final round of refinement, the respective models fit well with the experimental data showing χ^2 values of 1.06 for K192S^{fs*7} and 1.16 for R153S^{fs*41} (Fig. 6A,B). The molecular envelopes show in general, a flat shape with a broad face (Fig. 6C,D). The longest molecular spans of the models for K192S^{fs*7} and R153S^{fs*41} were 14 nm and 13 nm respectively, in good agreement with the D_{max} estimated from their respective $P(r)$ functions (both 13 nm).

PQBP-1 mutants bind weaker to RNA polymerase II

The CTD of RNA polymerase II consists of 52 repeats of the consensus heptapeptide sequence YSPTSPS that is subjected to phosphorylation during the initiation and elongation steps of transcription. We studied the binding of PQBP-1 variants to the heptapeptide phosphorylated at positions 2 and 5 (YpSPTpSPS) (Verdecia et al., 2000). Wild-type PQBP-1 binds the peptide with a dissociation constant, K_d of 154 (±10) μM. Previously, the interaction between other WW domains with various phosphopeptides have been studied. The Pin1 WW domain binds to the same peptide studied here with a K_d of 34 μM (Verdecia et al., 2000). Pin1 WW domain also binds to a peptide from Cdc25 phosphatase with a K_d of 117 μM; and to a peptide from human τ protein with a K_d of 230 μM (Wintjens et al., 2001). Our result showed similar weak binding affinity and is consistent with the nature of a reversible interaction that is required for effective transcription regulation (Tapia et al., 2010). The mutants K192S^{fs*7} and R153S^{fs*41} bind substantially weaker with K_d values of 456 (±16) μM and 338 (±35) μM, respectively (Fig. 7).

DISCUSSION

In this study, the structural properties of two frameshift mutants were compared with those of wild-type PQBP-1. Solution models built from SAXS analysis revealed that both mutants are flat dimeric molecules with the same maximum dimensions as the wild type, suggesting that the two protomers lie lengthways alongside each other. In the case of K192S^{fs*7}, we observed a tendency to form even higher oligomers (4–6 protomers, Fig. 2A). The two clinically important mutants in this study show many similar structural properties: they are partly unfolded but have more structure and are more compact than the wild-type protein. The apparent increase in structure of both mutants is presumably due to the removal of the C-terminal region, which is largely disordered (Nabeshima et al., 2014; Takahashi et al., 2009). Despite their similar sizes (K192S^{fs*7} has 197 residues; R153S^{fs*41} has 192), their sequences and natures are very different owing to R153S^{fs*41} having a long non-native tail. The thermodynamic properties of K192S^{fs*7} is more like the wild-type protein whereas those of R153S^{fs*41} is completely different (Table 2). The fitting of thermal melting data was performed to test if simple folding models can be used to describe the mechanisms of "folding" (here refers to the formation of the physiological dimer from the denatured monomers). K192S^{fs*7} showed 2-state melting whereas R153S^{fs*41} denatured via an intermediate. Our analysis did not discriminate between the two (monomeric or dimeric intermediate) unfolding models as they showed similar R^2 values of non-linear regression (Table S3.1).

The melting profiles of the mutants showed only limited cooperativity. In PQBP-1, the melting curve has contributions from the WW domain and the rest of the protein. Even the isolated WW domain does not show a steep melting transition (Tapia et al., 2010). PQBP-1 is best described as of the molten globule state, consisting of near-native secondary structures with a loosely-packed hydrophobic core (Rees et al., 2012). The shallow melting curves of

PQBP-1 and the two mutants indicate that their "folding" represents an increase of compactness of the molten globular states. The T_m of the WW domain is 45 °C (Tapia et al., 2010); therefore its folding is concomitant with the rest of the protein. In the case of the mutants, the "folding" further includes the dimerisation events, which can be considered an intermolecular packing of secondary structures.

The RNA pol II peptide binding studies may shed some light on the dimerisation interface. Presumably, both mutants have their peptide-binding sites blocked on dimerisation. The protein sequence of K192S^{fs*7} has only six non-native residues at its tail (Fig. 1B), therefore, it is conceivable that the PQBP-1 sequence up to residue 191 has the propensity to dimerise. The full-length protein being monomeric then argues for the C-terminal region playing an auto-inhibitory role against homo-dimerisation. Indeed, the N-segment (residues 1–219) and C-segment (residues 220–265) interact with each other weakly, with a dissociation constant of > 10^{-4} M (Nabeshima et al., 2014).

Assuming that the regions responsible for dimerisation are in common on both mutants, we speculate that the WW domain itself is involved. Some WW domains are known to form hetero- or homo-dimers (Ohnishi et al., 2007; Senturia et al., 2010). Here, the ensemble models which had a dimeric WW domain incorporated showed the best fit to the SAXS data. Dimerisation of the PQBP-1 WW domain is known in one other XLID mutant, which has a missense mutation (Pucheta-Martinez et al., 2016; Sudol et al., 2012; Tapia et al., 2010). This is the result of disulphide bond formation of the introduced cysteine (Y65C) residue.

In earlier works, it has been found that the loss of the C-terminal region in most of the PQBP-1 frameshift mutants will abolish its interaction to the spliceosomal protein U5-15kD (Mizuguchi et al., 2014). Here we show that mutant proteins dimerise and their interactions with the transcription machinery are affected. In addition, toxic gain-of-function associations with other cellular molecules may also be present because of the higher stabilities of the

mutants. In the case of R153S^{fs*41}, it is further complicated by having its nuclear localisation compromised (Kalscheuer et al., 2003) — thus the mutant dimers may also interact with cytoplasmic molecules. These results suggest that inhibition of dimer formation may be a strategy worthy of pursuing for therapeutic intervention of the Renpenning syndrome.

CONCLUSIONS

We performed structural and biophysical characterisation of two disease proteins involved in X-linked intellectual disability. Our main finding is that the mutant proteins form stable dimers unlike the wild type which are monomers. We obtained the solution models (molecular envelope and ensemble representation) of these partially disordered proteins from X-ray scattering data and discussed their global structural changes relative to the wild type. The mutated proteins have much reduced interactions with the cell transcription machinery, and thus gene expression is affected in these developmental diseases.

Author Contributions

S.K.R. and Y.W.C. designed and performed the research, analysed the data and wrote the manuscript. S.K.R. and H.O. contributed experimental resources.

Acknowledgements

S.K.R. was funded by a King's College London Graduate School studentship. Anne Tuukkanen and Stephen Boivin at the EMBL (DESY, Hamburg, Germany) helped with sample characterisation and data acquisition. Dmitri Svergun, also at EMBL, gave critical comments of the text. Tam Bui of the Bimolecular Spectroscopy Centre, King's College London helped with collection of CD data. Chin-Pang Benny Yiu advised on the interpretation of the CD data. We are grateful to the free use of the ATSAS online server (http://www.embl-hamburg.de/biosaxs/atsas-online/). BioStuct-X funded most of the costs

for the SAXS experiments (project IDs: 652, 3772 and 7287). Y.W.C. also acknowledges support from the Innovation and Technology Commission of Hong Kong, the Hong Kong Polytechnic University and the Life Science Area of Strategic Fund 1-ZVH9.

Declarations of interest: none

References

Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., Svergun, D.I., 2007. Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129, 5656-5664.

Borgia, A., Zheng, W., Buholzer, K., Borgia, M.B., Schuler, A., Hofmann, H., Soranno, A., Nettels, D., Gast, K., Grishaev, A., Best, R.B., Schuler, B., 2016. Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. J Am Chem Soc 138, 11714-11726.

Drake, A.F., Siligardi, G., Gibbons, W.A., 1988. Reassessment of the electronic circular dichroism criteria for random coil conformations of poly(L-lysine) and the implications for protein folding and denaturation studies. Biophys Chem 31, 143-146.

Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., 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.

Egloff, S., Murphy, S., 2008. Cracking the RNA polymerase II CTD code. Trends Genet 24, 280-288.

Germanaud, D., Rossi, M., Bussy, G., Gerard, D., Hertz-Pannier, L., Blanchet, P., Dollfus, H., Giuliano, F., Bennouna-Greene, V., Sarda, P., Sigaudy, S., Curie, A., Vincent, M.C., Touraine, R., des Portes, V., 2011. The Renpenning syndrome spectrum: new clinical insights supported by 13 new PQBP1-mutated males. Clin Genet 79, 225-235.

Greenfield, N.J., 2004. Analysis of circular dichroism data. Methods Enzymol 383, 282-317.

Greenfield, N.J., 2006. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc 1, 2527-2535.

Hayes D., L.T., Philo J. 1995. Sednterp: Sedimentation Interpretation Program. The University of New Hampshire.

Hu, H., Columbus, J., Zhang, Y., Wu, D., Lian, L., Yang, S., Goodwin, J., Luczak, C., Carter, M., Chen, L., James, M., Davis, R., Sudol, M., Rodwell, J., Herrero, J.J., 2004. A map of WW domain family interactions. Proteomics 4, 643-655.

Iwasaki, Y., Thomsen, G.H., 2014. The splicing factor PQBP1 regulates mesodermal and neural development through FGF signaling. Development 141, 3740-3751.

Kalscheuer, V.M., Freude, K., Musante, L., Jensen, L.R., Yntema, H.G., Gecz, J., Sefiani, A., Hoffmann, K., Moser, B., Haas, S., Gurok, U., Haesler, S., Aranda, B., Nshedjan, A., Tzschach, A., Hartmann, N., Roloff, T.C., Shoichet, S., Hagens, O., Tao, J., Van Bokhoven, H., Turner, G., Chelly, J., Moraine, C., Fryns, J.P., Nuber, U., Hoeltzenbein, M., Scharff, C., Scherthan, H., Lenzner, S., Hamel, B.C., Schweiger, S., Ropers, H.H., 2003. Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. Nat Genet 35, 313-315.

Komuro, A., Saeki, M., Kato, S., 1999. Npw38, a novel nuclear protein possessing a WW domain capable of activating basal transcription. Nucleic Acids Res 27, 1957-1965.

Kunde, S.A., Musante, L., Grimme, A., Fischer, U., Muller, E., Wanker, E.E., Kalscheuer, V.M., 2011. The X-chromosome-linked intellectual disability protein PQBP1 is a component of neuronal RNA granules and regulates the appearance of stress granules. Hum Mol Genet 20, 4916-4931.

Lenski, C., Abidi, F., Meindl, A., Gibson, A., Platzer, M., Frank Kooy, R., Lubs, H.A., Stevenson, R.E., Ramser, J., Schwartz, C.E., 2004. Novel truncating mutations in the polyglutamine tract binding protein 1 gene (PQBP1) cause Renpenning syndrome and X-linked mental retardation in another family with microcephaly. Am J Hum Genet 74, 777-780.

Mizuguchi, M., Obita, T., Serita, T., Kojima, R., Nabeshima, Y., Okazawa, H., 2014. Mutations in the PQBP1 gene prevent its interaction with the spliceosomal protein U5-15 kD. Nat Commun 5, 3822.

Nabeshima, Y., Mizuguchi, M., Kajiyama, A., Okazawa, H., 2014. Segmental isotope-labeling of the intrinsically disordered protein PQBP1. FEBS Lett 588, 4583-4589.

Ohnishi, S., Guntert, P., Koshiba, S., Tomizawa, T., Akasaka, R., Tochio, N., Sato, M., Inoue, M., Harada, T., Watanabe, S., Tanaka, A., Shirouzu, M., Kigawa, T., Yokoyama, S., 2007. Solution structure of an atypical WW domain in a novel beta-clam-like dimeric form. FEBS Lett 581, 462-468.

Okazawa, H., Rich, T., Chang, A., Lin, X., Waragai, M., Kajikawa, M., Enokido, Y., Komuro, A., Kato, S., Shibata, M., Hatanaka, H., Mouradian, M.M., Sudol, M., Kanazawa, I., 2002. Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death. Neuron 34, 701-713.

Otte, L., Wiedemann, U., Schlegel, B., Pires, J.R., Beyermann, M., Schmieder, P., Krause, G., Volkmer-Engert, R., Schneider-Mergener, J., Oschkinat, H., 2003. WW domain sequence activity relationships identified using ligand recognition propensities of 42 WW domains. Protein Sci 12, 491-500.

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

Pucheta-Martinez, E., D'Amelio, N., Lelli, M., Martinez-Torrecuadrada, J.L., Sudol, M., Saladino, G., Gervasio, F.L., 2016. Changes in the folding landscape of the WW domain provide a molecular mechanism for an inherited genetic syndrome. Sci Rep 6, 30293.

Qi, Y., Hoshino, M., Wada, Y., Marubuchi, S., Yoshimura, N., Kanazawa, I., Shinomiya, K., Okazawa, H., 2005. PQBP-1 is expressed predominantly in the central nervous system during development. Eur J Neurosci 22, 1277-1286.

Rambo, R.P., Tainer, J.A., 2013. Accurate assessment of mass, models and resolution by small-angle scattering. Nature 496, 477-481.

Rees, M., Gorba, C., de Chiara, C., Bui, T.T., Garcia-Maya, M., Drake, A.F., Okazawa, H., Pastore, A., Svergun, D., Chen, Y.W., 2012. Solution model of the intrinsically disordered polyglutamine tract-binding protein-1. Biophys J 102, 1608-1616.

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815.

Schuck, P., 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78, 1606-1619.

Senturia, R., Faller, M., Yin, S., Loo, J.A., Cascio, D., Sawaya, M.R., Hwang, D., Clubb, R.T., Guo, F., 2010. Structure of the dimerization domain of DiGeorge critical region 8. Protein Sci 19, 1354-1365.

Siligardi, G., Drake, A.F., 1995. The importance of extended conformations and, in particular, the PII conformation for the molecular recognition of peptides. Biopolymers 37, 281-292.

Stevenson, R.E., Bennett, C.W., Abidi, F., Kleefstra, T., Porteous, M., Simensen, R.J., Lubs, H.A., Hamel, B.C., Schwartz, C.E., 2005. Renpenning syndrome comes into focus. Am J Med Genet A 134, 415-421.

Sudol, M., McDonald, C.B., Farooq, A., 2012. Molecular insights into the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1. FEBS Lett 586, 2795-2799.

Takahashi, M., Mizuguchi, M., Shinoda, H., Aizawa, T., Demura, M., Okazawa, H., Kawano, K., 2009. Polyglutamine tract binding protein-1 is an intrinsically unstructured protein. Biochim Biophys Acta 1794, 936-943.

Takahashi, M., Mizuguchi, M., Shinoda, H., Aizawa, T., Demura, M., Okazawa, H., Kawano, K., 2010. Polyglutamine tract-binding protein-1 binds to U5-15kD via a continuous 23-residue segment of the C-terminal domain. Biochim Biophys Acta 1804, 1500-1507.

Tapia, V.E., Nicolaescu, E., McDonald, C.B., Musi, V., Oka, T., Inayoshi, Y., Satteson, A.C., Mazack, V., Humbert, J., Gaffney, C.J., Beullens, M., Schwartz, C.E., Landgraf, C., Volkmer, R., Pastore, A., Farooq, A., Bollen, M., Sudol, M., 2010. Y65C missense mutation in the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1 affects its binding activity and deregulates pre-mRNA splicing. J Biol Chem 285, 19391-19401.

Tarini, M., Cignoni, P., Montani, C., 2006. Ambient occlusion and edge cueing to enhance real time molecular visualization. IEEE Trans Vis Comput Graph 12, 1237-1244.

Tria, G., Mertens, H.D., Kachala, M., Svergun, D.I., 2015. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCrJ 2, 207-217.

Verdecia, M.A., Bowman, M.E., Lu, K.P., Hunter, T., Noel, J.P., 2000. Structural basis for phosphoserine-proline recognition by group IV WW domains. Nat Struct Biol 7, 639-643.

Wang, Q., Moore, M.J., Adelmant, G., Marto, J.A., Silver, P.A., 2013. PQBP1, a factor linked to intellectual disability, affects alternative splicing associated with neurite outgrowth. Genes Dev 27, 615-626.

Waragai, M., Junn, E., Kajikawa, M., Takeuchi, S., Kanazawa, I., Shibata, M., Mouradian, M.M., Okazawa, H., 2000. PQBP-1/Npw38, a nuclear protein binding to the polyglutamine tract, interacts with U5-15kD/dim1p via the carboxyl-terminal domain. Biochem Biophys Res Commun 273, 592-595.

Waragai, M., Lammers, C.H., Takeuchi, S., Imafuku, I., Udagawa, Y., Kanazawa, I., Kawabata, M., Mouradian, M.M., Okazawa, H., 1999. PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival. Hum Mol Genet 8, 977-987.

Wintjens, R., Wieruszeski, J.M., Drobecq, H., Rousselot-Pailley, P., Buee, L., Lippens, G., Landrieu, I., 2001. 1H NMR study on the binding of Pin1 Trp-Trp domain with phosphothreonine peptides. J Biol Chem 276, 25150-25156.

Table 1. Solution properties of PQBP-1 wild type and mutants compared.

	K192S^{fs*7}	R153S^{fs*41}	wild type
Hydrodynamic radius, R_s, nm			
AUC	4.23 (±0.07)	3.71 (±0.07)	3.67 (±0.02) [a]
Shape parameters			
SAXS R_g (Guinier), nm	3.83 (±0.08)	3.63 (±0.05)	3.77 (±0.03) [b]
SAXS R_g ($P(r)$ real), nm	3.78 (±0.01)	3.61 (±0.01)	N.A.
SAXS R_g ($P(r)$ recip.), nm	3.76	3.59	N.A.
R_g / R_s [c]	0.91 (±0.02)	0.97 (±0.02)	1.03 (±0.01) [b]
AUC f / f_0	1.8	1.6	1.7 [a]
Molecular mass, kDa			
Calculated	23.25	21.91	30.63 [a]
AUC	46.2 (±0.9)	43.5 (±0.8)	31.0 (±0.5) [a]
SAXS [d]	45 (±5)	43-46 (±5)	30 (±3) [a]
D_{max}, nm	13 (±1)	13 (±1)	13 [a]

[a] The values of the wild type were from Rees *et al* (2012).

[b] The R_g values from Guinier analysis of the wild type were recalculated in this work with data to $q \times R_g = 1.1$.

[c] The R_g values used in these calculation were those obtained from Guinier analysis.

[d] The molecular masses obtained from SAXS are expressed in ranges because they are dependent on the method and data range used.

AUC: analytical ultracentrifugation; SAXS: small angle X-ray scattering; N.A.: not available.

Table 2. Thermodynamic properties of PQBP-1 wild type and two mutants.

	Wild type	K192S^{fs*7}	R153S^{fs*41}
Folding Model	monomer, 2-state N Ý U	dimer, 2-state N_2 Ý 2U	dimer, 3-state N_2 Ý 2I Ý 2U
T_m, °C	38.0 ±0.9	39.5 ±0.6	T_{m1}=17.9 ±0.8 T_{m2}=50.9 ±0.4
$\Delta H°$, kJ mol^{-1}	–50.5 ±3.5	–72.5 ±3.2	–580 ±44 ΔH_1=–155 ±18 ΔH_2=–425 ±41
$\Delta G°$, kJ mol^{-1}	–2.13 ±0.08	–31.7 ±0.1	–59.2 ±1.7 $\Delta G_1°$ = –25.1 ±0.3 $\Delta G_2°$ = –34.1 ±1.6
$\Delta S°$, J K^{-1} mol^{-1}	–162.3 ±11.7	–136.9 ±10.6	–1748 ±149 ΔS_1 = –437 ±61 ΔS_2 = –1311 ±136
R^2 of fit	0.9956	0.9983	0.9993

N, native state; I, intermediate; U, unfolded state; T_m, melting temperature; $\Delta H°$, standard enthalpy change; $\Delta G°$, standard free energy change; $\Delta S°$, standard entropy change. For the R153S^{fs*41} mutant, the analysis with a monomeric intermediate (highest R^2 value) is shown. See Table S3.1 for a comparison of three different folding/unfolding models.

Figure Legends

Figure 1. Domain architecture of PQBP-1 and its mutants. (A) WW: WW domain; PRD: polar-amino-acid-rich domain; N: nuclear localisation signal; CTR: C-terminal region. For the two mutants, black and grey bars denote regions with native and non-native sequences respectively. The grey bars on top indicate the regions which mediate interactions with other proteins or motifs (P-rich: proline-rich motifs; polyQ: polyglutamine tracts; Atx-1: ataxin-1; Htt: huntingtin; AR: androgen receptor). (B) Sequences from PQBP-1 residue 153 to C-termini of the wild type (WT) and the two mutants. The frameshift mutation occurs at the residue marked with an inverted triangle, with a C-terminal asterix marking the position encoded by the stop codon.

Acidic residues are coloured red; basic, blue and hydrophobic, green. The sequences preceding residue 153 are common to all three proteins.

Figure 2. Hydrodynamic properties of PQBP-1 mutants. Velocity sedimentation AUC results of (A) K192S^{fs*7} and (B) R153S^{fs*41} analysed with SEDFIT are presented as the concentration distribution $c(s)$ of sedimentation species at varying sample dilutions.

Figure 3. Thermodynamic properties of two PQBP-1 mutants. Left panels are the CD spectra of temperature scanning experiments for (A) K192S^{fs*7} and (C) R153S^{fs*41}. The samples were heated from 6°C (dark blue) to 90°C (brick red). Right panels are non-linear regression analyses of the temperature dependence of CD (mean residue ellipticity, $[\theta]^{MR}$) at 225 nm of the two mutants: a two-state model (red line) was used for (B) K192S^{fs*7} and a three-state model (red line) for (D) R153S^{fs*41}. See supplementary section S3 for details. The respective melting temperatures (T_m) were indicated. The goodness-of-fit is represented by the R^2 value.

Figure 4. SAXS analysis of two PQBP-1 mutants. (A) Normalised Kratky plots show that both K192S^{fs*7} (orange) and R153S^{fs*41} (light green) are partially disordered yet both have more structure than the wild-type protein (grey). (B) The $P(r)$ distribution (in arbituary units) of the two mutants and the wild type compared, with same colouring scheme as in (A).

Figure 5. EOM Analysis of PQBP-1 mutants. Top panels show fitting for K192S^{fs*7} (A) and R153S^{fs*41} (B). The calculated model intensities were in orange (K192S^{fs*7}) and in light green (R153S^{fs*41}) respectively. The experimental SAXS profiles of the mutants were shown as black open circles. The middle panels show the respective maximum dimension (D_{max}) distribution of EOM models of K192S^{fs*7} (orange, C) and R153S^{fs*41} (light green, D). The D_{max} distributions of the pool of starting structures before EOM refinement are shown in grey (C,D). The lower panels show the respective distribution of radii of gyration of EOM models of K192S^{fs*7} (orange, E) and R153S^{fs*41} (light green, F). The R_g distributions of the pool of structures before EOM refinement are shown in grey (E,F). The EOM model properties of the wild type are shown as dark grey lines (C,D,E,F).

Figure 6. *Ab initio* modelling of PQBP-1 mutants. Top panels show Dammin fitting for K192S^{fs*7} (A) and R153S^{fs*41} (B). The simulated model intensities were in orange (K192S^{fs*7}) and in light green (R153S^{fs*41}) respectively. The GNOM intensities used in modelling were shown as black open circles. Data to 2.0 nm^{-1} were used for K192S^{fs*7} and to 2.2 nm^{-1} for R153S^{fs*41}. The experimental SAXS profiles of the mutants were shown as light-grey filled circles in the background. The middle and lower panels (C,D) show the respective refined bead models of the mutants; each set related by a 90° rotation along the X-axis. A colour-filled circle is drawn with the respective radius of gyration determined by SAXS of each mutant. A broken-line open circle was drawn using the radius of gyration of the wild-type protein (3.8 nm) for comparison. The bead models were drawn to the same scale, with bead radii of K192S^{fs*7} and R153S^{fs*41} being 3Å and 2.5Å respectively.

Figure 7. Protein-peptide binding assay. Fluorescence polarisation is plotted against protein concentration. The data were analysed with a single-site specific binding model, with R^2 = 0.9964 (wild-type PQBP-1, grey), 0.9996 (K192S^{fs*7}, orange) and 0.9962 (R153S^{fs*41}, light green), respectively. All vertical error (standard error of 6 observations) bars are smaller than the symbol and are not shown. Horizontal error bars are estimated experimental errors.

SKR, HO & YWC

Highlights

- Both PQBP-1 frameshift mutants are dimeric, more compact and more stable
- *ab initio* bead models showed overall flat shapes, probably parallel arrangement
- Ensemble models showed folded core (WW domain) with the rest as unfolded regions
- The interactions with an RNA pol II-derived peptide were weakened in both mutants
- Homo-dimerisation of these mutants may link to loss of transcription functions