



## Polyglutamine tract binding protein-1 is an intrinsically unstructured protein

Masaki Takahashi<sup>a</sup>, Mineyuki Mizuguchi<sup>a,\*</sup>, Hiroyuki Shinoda<sup>a</sup>, Tomoyasu Aizawa<sup>b</sup>, Makoto Demura<sup>b</sup>, Hitoshi Okazawa<sup>c</sup>, Keiichi Kawano<sup>b</sup>

<sup>a</sup> Faculty of Pharmaceutical Sciences, University of Toyama, 2630, Sugitani, Toyama 930-0194, Japan

<sup>b</sup> Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

<sup>c</sup> Department of Neuropathology, Medical Research Institute and 21st Century Center of Excellence Program (COE) for Brain Integration and Its Disorders, Tokyo Medical and Dental University, 1-5-45, Yushima, Bunkyo-ku, Tokyo 113-8510, Japan

### ARTICLE INFO

#### Article history:

Received 30 September 2008

Received in revised form 27 February 2009

Accepted 3 March 2009

Available online 17 March 2009

#### Keywords:

Unstructured protein

Protein structure

PQBP-1

Polyglutamine

U5-15kD

### ABSTRACT

Polyglutamine tract binding protein-1 (PQBP-1) is a nuclear protein that interacts with disease proteins containing expanded polyglutamine repeats. PQBP-1 also interacts with RNA polymerase II and a spliceosomal protein U5-15kD. In the present study, we demonstrate that PQBP-1 is composed of a large unstructured region and a small folded core. Intriguingly, the large unstructured region encompasses two functional domains: a polar amino acid rich domain and a C-terminal domain. These findings suggest that PQBP-1 belongs to the family of intrinsically unstructured/disordered proteins. Furthermore, the binding of the target molecule U5-15kD induces only minor conformational changes into PQBP-1. Our results suggest that PQBP-1 includes high content of unstructured regions in the C-terminal domain, in spite of the binding of U5-15kD.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Polyglutamine diseases are autosomal dominant, typically late-onset, fatal neurodegenerative disorders, in which the expansion of coding CAG generates long stretches of polyglutamines. Polyglutamine tract expansion is non-pathogenic up to a threshold length, but abnormal expansions cause diseases such as Huntington's disease [1], Kennedy's disease [2], dentatorubro-pallidoluysian atrophy (DRPLA) [3–6] and spinocerebellar ataxias [7–10]. In most cases, the mutant proteins containing expanded polyglutamine tracts translocate into the nucleus and form intra-nuclear aggregates or inclusion bodies, which are believed to cause neuronal cell loss.

A growing body of evidence indicates that polyglutamine disease proteins interact directly with nuclear proteins. These nuclear proteins include the TATA-binding protein, the *Drosophila* eyes-absent protein, the CREB-binding protein, p53, the nuclear receptor co-repressor, mSin3A, and TAFII130 [11–16]. One hypothesis suggests that expanded polyglutamines result in aberrant interactions with nuclear proteins and thereby lead to transcriptional dysregulation [12,17–23]. This hypothesis

is supported by the following observations. The TATA-binding protein is localized to nuclear inclusions in spinocerebellar ataxia (SCA) type-3 and TAFII130 is localized to inclusions in DRPLA [11,16]. The CREB-binding protein coaggregates with androgen receptor and huntingtin [13,15]. Mutant huntingtin interacts with p53 and the CREB-binding protein, which results in transcriptional repression [14].

The polyglutamine tract binding protein-1 (PQBP-1) is one of the molecules involved in the pathology of polyglutamine diseases [24]. PQBP-1 is a 265 amino acid protein, which contains a WW domain (WWD), a polar amino acid rich domain (PRD) and a C-terminal domain (CTD) (Fig. 1). The WWD interacts with the C-terminal domain of the RNA polymerase II large subunit (Pol II) [25]. The PRD of PQBP-1 binds to the polyglutamine tracts of various proteins, such as huntingtin and the neuron-specific transcription factor Brn-2 [26]. The CTD of PQBP-1 binds to U5-15kD [27], a component of the U5 small ribonucleoprotein particle, which is one of the components of the spliceosome [28].

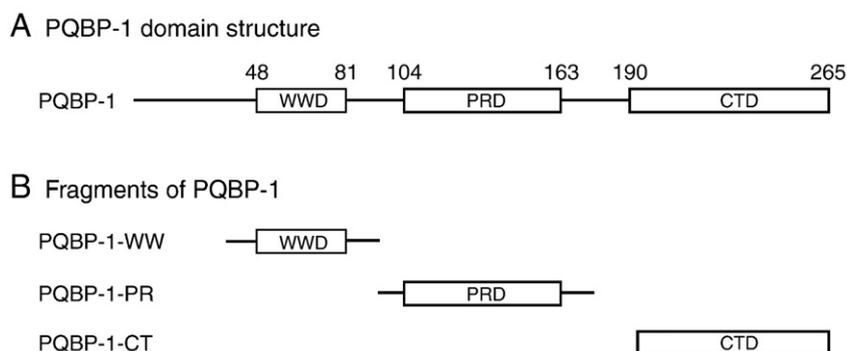
PQBP-1 is expressed predominantly in the brain regions susceptible to SCA-1 [26]. The SCA-1 disease protein, ataxin-1, interacts with PQBP-1 and the PQBP-1-ataxin-1 complex reduces the level of phosphorylated Pol II, leading to transcriptional dysregulation. Moreover, PQBP-1 colocalizes with ataxin-1 in nuclear inclusion bodies [25]. These observations are consistent with the above-mentioned hypothesis that transcriptional perturbation underlies polyglutamine-mediated pathology.

Very little is known about the structure of PQBP-1, although this information would help to understand the detailed functions of this protein. In this article, we report the results of biophysical and biochemical studies aimed at understanding the structural characteristics

**Abbreviations:** ANS, 1-anilino-8-naphthalenesulfonate; CD, circular dichroism; CTD, C-terminal domain; DRPLA, dentatorubro-pallidoluysian atrophy; DTT, dithiothreitol; GdnHCl, guanidine hydrochloride; HSQC, heteronuclear single-quantum correlation; Pol II, RNA polymerase II large subunit; PONDR, predictors of natural disordered regions; POODLE-L, prediction of order and disorder by machine learning-L; PQBP-1, polyglutamine tract binding protein-1; PRD, polar amino acid rich domain; SCA, spinocerebellar ataxia; WWD, WW domain

\* Corresponding author.

E-mail address: [mineyuki@pha.u-toyama.ac.jp](mailto:mineyuki@pha.u-toyama.ac.jp) (M. Mizuguchi).



**Fig. 1.** PQBP-1 domain structure (A) and PQBP-1 fragments (B). (A) The WW, PRD and CTD correspond to residues 48–81, residues 104–163 and residues 190–265, respectively. (B) PQBP-1-WW, PQBP-1-PR and PQBP-1-CT consist of residues 36–94, 94–176 and 193–265 of PQBP-1, respectively.

of PQBP-1. We demonstrate that PQBP-1 is composed of a large unstructured region and a small folded core. The large unstructured region encompasses two functional domains, the PRD and the CTD.

## 2. Materials and methods

### 2.1. Prediction of ordered and disordered regions in PQBP-1

We predicted ordered and disordered regions in PQBP-1 using predictors of natural disordered regions (PONDR) (<http://www.pondr.com>). The default predictor VL-XT was used [29,30]. The prediction of disorder was also performed with another predictor named the prediction of order and disorder by machine learning-L (POODLE-L) (<http://mbs.cbrc.jp/poodle/>) [31].

### 2.2. Expression plasmids

PQBP-1 cDNA was subcloned between the EcoRI and XhoI sites of pGEX-6P-1 (GE Healthcare Bio-Sciences, Buckinghamshire, UK). PQBP-1(36–94) cDNA was subcloned into pGEX-2TK predigested with BamHI and EcoRI (GE Healthcare Bio-Sciences) [25]. PQBP-1(94–176) and PQBP-1(193–265) cDNAs were subcloned between the BamHI and Sall sites of pGEX-6P-1. U5-15kD cDNA was subcloned between the EcoRI and Sall sites of pGEX-6P-1. We checked the sequence of the inserted DNA using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA).

### 2.3. Protein expression and purification

All proteins were expressed in *Escherichia coli* strain BL21(DE3), and were detected in a soluble fraction after cell lysis by sonication. GST-fusion proteins were purified with glutathione Sepharose 4B (GE Healthcare Bio-Sciences), and the proteins were separated from GST by digestion with protease. PreScission protease was used for cleavage of GST-PQBP-1, GST-PQBP-1(94–176), GST-PQBP-1(193–265) and GST-U5-15kD. Thrombin was used for cleavage of GST-PQBP-1(36–94). After removing thrombin with benzamidine Sepharose 4 Fast Flow (GE Healthcare Bio-Sciences), PQBP-1(36–94) was purified by gel-filtration chromatography. Protein purity was checked by SDS-PAGE with Coomassie brilliant blue staining. Protein concentration was determined by UV absorption at 280 nm.

### 2.4. Circular dichroism spectroscopy

Circular dichroism (CD) spectra were recorded on a JASCO J-805 spectropolarimeter. An optical cuvette with a 1-mm path length was used. The temperature of the measuring cell was maintained at 25 °C. Spectra were collected in a spectral range of 200–250 nm. The proteins were dissolved in 50 mM sodium phosphate (pH 7.0), and the protein

concentrations were 21–25 μM. Spectra of PQBP-1 and PQBP-1(36–94) were measured in the presence of 1 mM dithiothreitol (DTT).

### 2.5. NMR spectroscopy

M9 medium or C.H.L. medium (Chlorella Industry, Tokyo, Japan) were used for uniformly <sup>15</sup>N-labeled protein and <sup>15</sup>N/<sup>2</sup>H-labeled protein. The fragments of PQBP-1 were labeled with <sup>15</sup>N. The full-length PQBP-1 was labeled with <sup>15</sup>N and <sup>2</sup>H. The protein solutions were concentrated to 0.3–0.5 mM with an Amicon Ultra device (Millipore, Bedford, MA). The NMR samples contained 50 mM sodium phosphate (pH 7.0), 1 mM DTT, 50 μM 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt and 5% D<sub>2</sub>O.

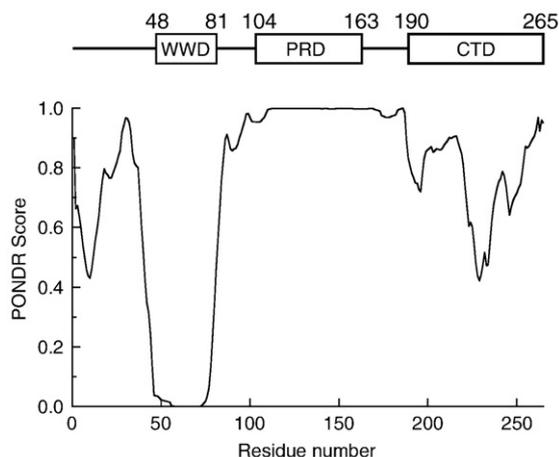
<sup>1</sup>H-<sup>15</sup>N heteronuclear single-quantum correlation (HSQC) spectra were acquired at 15 °C on a Bruker DMX-500 spectrometer. The obtained NMR data were processed with NMRPipe [32] and analyzed with NMRView [33].

### 2.6. Limited proteolysis

Proteolysis of PQBP-1, PQBP-1(94–176) and PQBP-1(193–265) was carried out by trypsin digestion (Sigma, St. Louis, MO) at 25 °C for 10 min. The buffer solution contained 20 mM sodium phosphate (pH 8.0) and 1 mM DTT. The proteolysis was also performed in the absence of DTT, and provided the same results. The enzyme/substrate weight ratio was 1:500 or 1:10. The proteolysis reaction was stopped by acidification with trifluoroacetic acid to pH 2.0. The reaction mixture was analyzed by mass spectrometry. Mass spectra were recorded on an autoflex-T1 mass spectrometer (Bruker Daltonics, Bremer, Germany) with the matrix-assisted laser desorption ionization time-of-flight technique. The proteolysis mixture was also analyzed by reverse phase HPLC with a COSMOSIL 5C4-AR-300 column (Nacalai Tesque, Kyoto, Japan). Several fragments were separated by HPLC, followed by mass spectrometry and N-terminal sequencing. The N-terminal sequencing by automated Edman degradation was performed with a PPSQ-20 protein sequencer (Shimadzu, Kyoto, Japan).

### 2.7. Surface plasmon resonance

Surface plasmon resonance was measured with a BiAcCore J (BiAcCore, Uppsala, Sweden). Approximately 3000 resonance units of U5-15kD were immobilized on a CM5 sensor chip (BiAcCore) by the method of amine coupling. The experiments were performed in 50 mM sodium phosphate (pH 7.0) and 1 mM DTT at 25 °C. PQBP-1 solutions were injected for 3 min at a flow rate of 60 μl/min, and PQBP-1 was dissociated for 3 min. The concentration of PQBP-1 ranged from 0.6 to 4.1 μM. Surface regeneration was achieved by the injection of 10 mM glycine buffer (pH 2.0) at a flow rate of 60 μl/min. Association and dissociation rate constants were determined with BIAviewer software.



**Fig. 2.** PONDR analysis of the PQBP-1 amino acid sequence. A PONDR score  $>0.5$  predicts the region to be disordered, while a score  $<0.5$  predicts the region to be ordered. The PQBP-1 domain structure is shown above the panel.

### 2.8. Tryptophan fluorescence

Fluorescence spectra were acquired on an F-4500 fluorescence spectrophotometer (Hitachi, Tokyo, Japan) at 25 °C. The concentration of PQBP-1 was 21–25  $\mu\text{M}$ . Buffer solution contained 50 mM sodium phosphate (pH 7.0) and 1 mM DTT. The excitation wavelength was 295 nm, and the emission spectra were recorded between 310 and 410 nm.

### 2.9. ANS binding

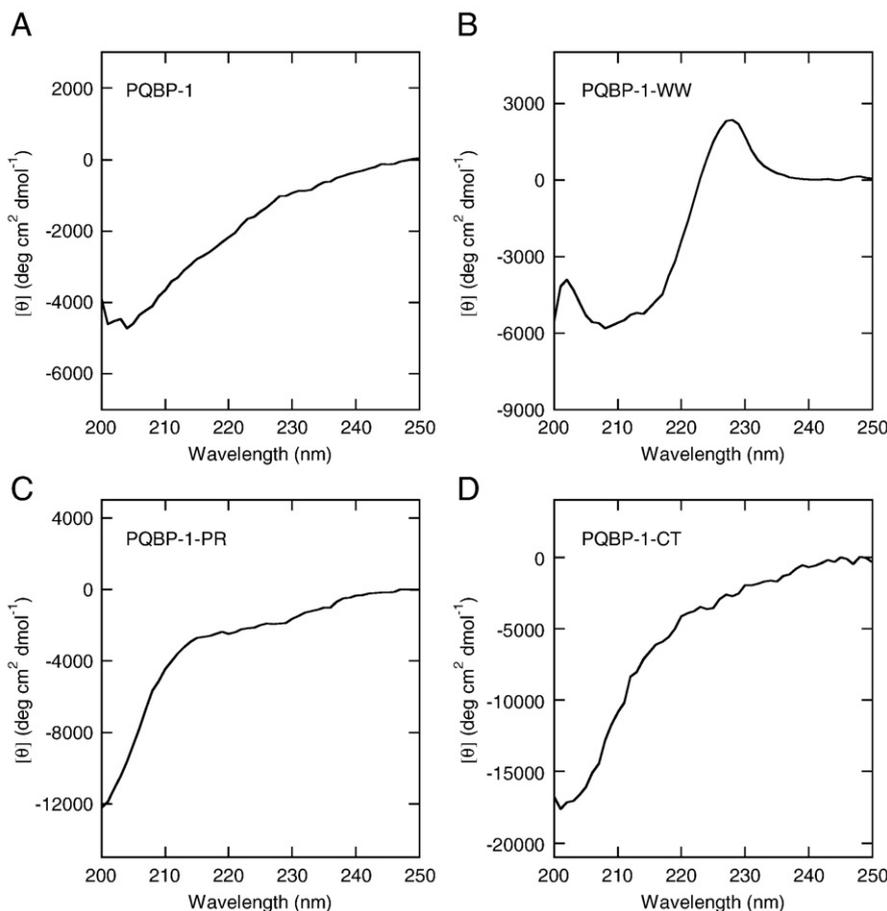
The hydrophobic fluorescent molecule, 1-anilino-8-naphthalene-sulfonate (ANS), was used as a probe for solvent-exposed hydrophobic clusters [34,35]. ANS fluorescence spectra were recorded on an F-4500 fluorescence spectrophotometer (Hitachi) at 25 °C. Buffer solutions contained 50 mM sodium phosphate (pH 7.0) and 1 mM DTT. ANS was excited at 380 nm, and the emission spectra were recorded between 410 and 650 nm.

### 2.10. GST pull-down assays

Purified GST-fusion proteins (1.8 nmol of GST-PQBP-1 or 31 nmol of GST-PQBP-1-CT) were incubated with Glutathione Sepharose 4B beads (GE Healthcare Bio-Sciences) for 1 h at 4 °C. To remove unbound GST-fusion proteins, the beads were washed with 50 mM phosphate buffer (pH 7.0) containing 1 mM DTT. The beads were incubated with 2.8 nmol of U5-15kD for 1 h at 4 °C and washed with 50 mM phosphate buffer (pH 7.0). After centrifugation, the supernatant was removed, and the beads were mixed with 2 $\times$  SDS-PAGE sample buffer. Proteins bound to the beads were separated on an SDS-PAGE gel. The gels were stained with Coomassie Brilliant Blue. GST protein expressed by empty pGEX-6P-1 vector was used as a negative control.

### 2.11. Analytical gel-filtration

To probe the binding of U5-15kD to PQBP-1, we performed analytical gel-filtration experiments. The sample contained 23  $\mu\text{M}$  U5-15kD and various concentrations of PQBP-1 (0–46  $\mu\text{M}$ ). Gel-



**Fig. 3.** Far-UV CD spectra of PQBP-1 (A), PQBP-1-WW (B), PQBP-1-PR (C) and PQBP-1-CT (D). All spectra were recorded at 25 °C.

filtration experiments were performed on a Superdex™ 200 column (GE Healthcare) equilibrated with 50 mM phosphate (pH 7.0) and 1 mM DTT. The experiments were performed at 4 °C with a flow rate of 0.4 ml/min.

### 3. Results

#### 3.1. Prediction of ordered and disordered regions in PQBP-1

We used PONDR to analyze the intrinsic disorder tendencies in PQBP-1 [29,30]. Any region with a PONDR score greater than 0.5 was considered disordered (Fig. 2). The PONDR score of residues 50–75 was almost zero, indicating the highly ordered structure of the WWD. Intriguingly, PONDR predicted a long disordered region over residues 100–265, which includes the PRD and CTD. Another predictor, POODLE-L, also predicted that the PRD and CTD are largely disordered [31].

#### 3.2. Spectroscopic analysis of full-length PQBP-1

We spectroscopically characterized PQBP-1 to gain insight into its structural properties. The secondary and tertiary structure was investigated by CD, NMR, and fluorescence spectroscopy.

To determine the secondary structure of the full-length PQBP-1, we recorded the far-UV CD spectrum (Fig. 3A). The CD spectrum displays a minimum at 204 nm with negative ellipticity of  $-4700 \text{ deg cm}^2 \text{ dmol}^{-1}$ , which is characteristic of a protein in largely unfolded conformations, whereas the spectral features indicative of  $\alpha$ -helices or  $\beta$ -sheets are absent (Fig. 3A).

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum also revealed the largely unfolded conformation of full-length PQBP-1, as judged by the numerous sharp resonances of backbone amides which overlap within a narrow  $^1\text{H}$  chemical shift range from 7.8 to 8.8 ppm (Fig. 4A). In addition to the sharp resonances, however, the HSQC spectrum for the full-length PQBP-1 showed several well-dispersed peaks typical of a folded conformation. The HSQC thus indicates that the full-length PQBP-1 contains a small structured region as well as a large unstructured region.

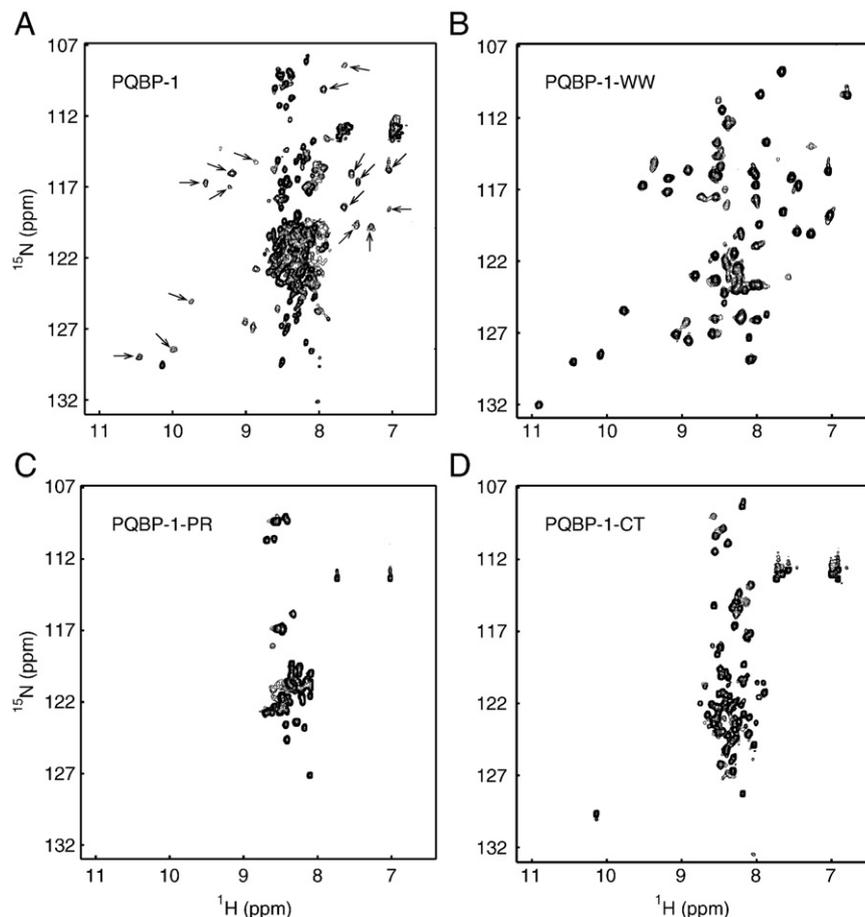
The presence of the folded region in the full-length PQBP-1 was supported by the tryptophan fluorescence and ANS-binding assay.

PQBP-1 has three tryptophan residues in the WWD and one in the CTD. The fluorescence spectrum of PQBP-1 in the native state showed a fluorescence maximum at 340 nm, whereas addition of 6 M guanidine hydrochloride (GdnHCl) led to a shift to a longer wavelength, 350 nm, and a reduction in the intensity (Fig. S1). These results indicate the presence of tryptophan residues buried inside of PQBP-1 in the native state.

We used the fluorescence probe ANS to detect the hydrophobic clusters exposed to solvent. A slight but significant increase in ANS fluorescence and a blue shift of the emission maximum was observed when PQBP-1 was added to ANS, indicating the presence of exposed hydrophobic clusters in PQBP-1 (Fig. S1).

#### 3.3. Spectroscopic analysis of PQBP-1 fragments

To analyze the contribution of each functional region to the structural properties of full-length PQBP-1, we constructed several truncated fragments on the basis of the domain structure (Fig. 1B). The fragment PQBP-1-WW consists of residues 36–94, PQBP-1-PR



**Fig. 4.** The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of PQBP-1 (A), PQBP-1-WW (B), PQBP-1-PR (C) and PQBP-1-CT (D). All spectra were recorded at 15 °C. (A) The arrows indicate the resonances whose chemical shifts are the same as those in the spectrum of PQBP-1-WW.

consists of residues 94–176, and PQBP-1-CT consists of residues 193–265, respectively. The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of these fragments were recorded, and were compared with that of the full-length PQBP-1.

The  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of PQBP-1-PR and PQBP-1-CT showed the sharp resonances of backbone amides clustered within 7.8–8.8 ppm ( $^1\text{H}$  dimension), indicating highly flexible, unfolded conformations (Fig. 4C and D). In contrast, the backbone amide resonances of PQBP-1-WW are dispersed from 7.0 to 9.8 ppm, which is characteristic of folded proteins (Fig. 4B). Furthermore, several resonances of PQBP-1-WW appeared at the same chemical shifts as in the HSQC of full-length PQBP-1 (Fig. 4A and B). Therefore, the comparison of the HSQC spectra of full-length PQBP-1 and its fragments revealed that the PRD and CTD are largely unstructured in PQBP-1, and the WW domain is involved in the structured region of PQBP-1.

The far-UV CD spectrum of PQBP-1-PR and -CT lacked the typical signature of secondary structure, exhibiting instead only a negative signal at 200 nm (Fig. 3C and D). In contrast, the far-UV spectrum of PQBP-1-WW has a prominent positive maximum at 228 nm and a negative one at 208 nm (Fig. 3B). Since these spectral features of PQBP-1-WW are quite similar to those of other WW domains [36], PQBP-1-WW probably adopts a small  $\beta$ -sheet structure as in the case of other WW domains. These CD data support the conclusion derived from NMR that PQBP-1 comprises the folded region in the WW domain and the large unstructured regions in the PRD and CTD.

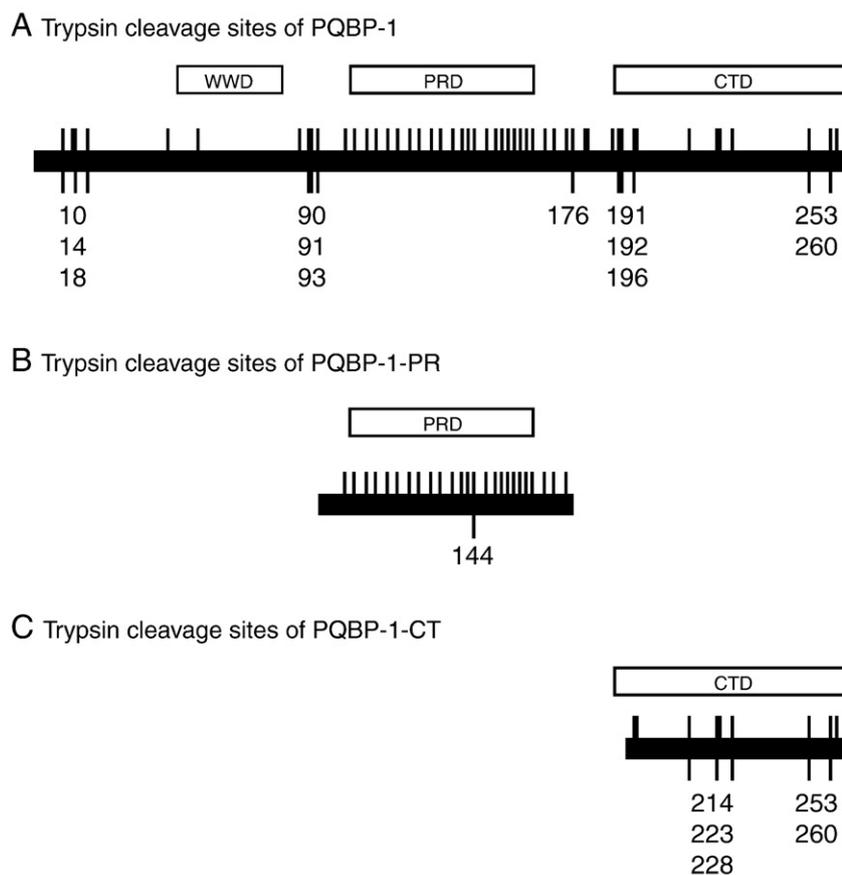
### 3.4. Limited proteolysis

Although the NMR and CD data demonstrated that the PRD and CTD of PQBP-1 are unstructured, it is unclear whether these domains

have no residual structure. Accordingly, in addition to the spectroscopic studies described above, we performed limited proteolysis to examine whether the PRD and CTD include residual structure. In general, limited proteolysis demonstrates the cleavage sites in proteins by proteases, and thus reveals the regions protected from proteolysis, namely the structured regions of proteins. The intrinsically unstructured proteins are extremely sensitive to proteolysis, resulting in complete degradation by proteases under the usual conditions of limited proteolysis [37]. We reasoned, nevertheless, that protection from an early cleavage event in proteolysis indicates the residual structure in the unstructured proteins [38].

We used trypsin as a proteolytic enzyme in this study, taking advantage of the numerous cleavage sites within the PRD and CTD sequences; there are 21 and 11 trypsin cleavage sites in the PRD and CTD, respectively (Fig. 5). In addition, we performed proteolysis of PQBP-1-PR and PQBP-1-CT as well as full-length PQBP-1. Detailed analysis of trypsin cleavage products using mass spectrometry showed that trypsin cleaved at the N-terminal region of PQBP-1, the loop region between WW domain and PRD, and the whole portion of CTD (Fig. 5A and C). However, trypsin cleaved PQBP-1-PR only at residue 144, and thus PRD is resistant to proteolysis (Fig. 5A and B). Moreover, a large amount of a fragment encompassing PRD (residues 94–176) was detected by HPLC after the trypsin cleavage of PQBP-1 (Fig. S2A). Assuming that the regions protected from proteolysis indicate the presence of residual structure, the PRD retains a certain amount of residual structure in PQBP-1.

In addition to the limited proteolysis under the mild proteolytic conditions (the enzyme/substrate weight ratio is 1:500), we performed the proteolysis experiments using a high concentration

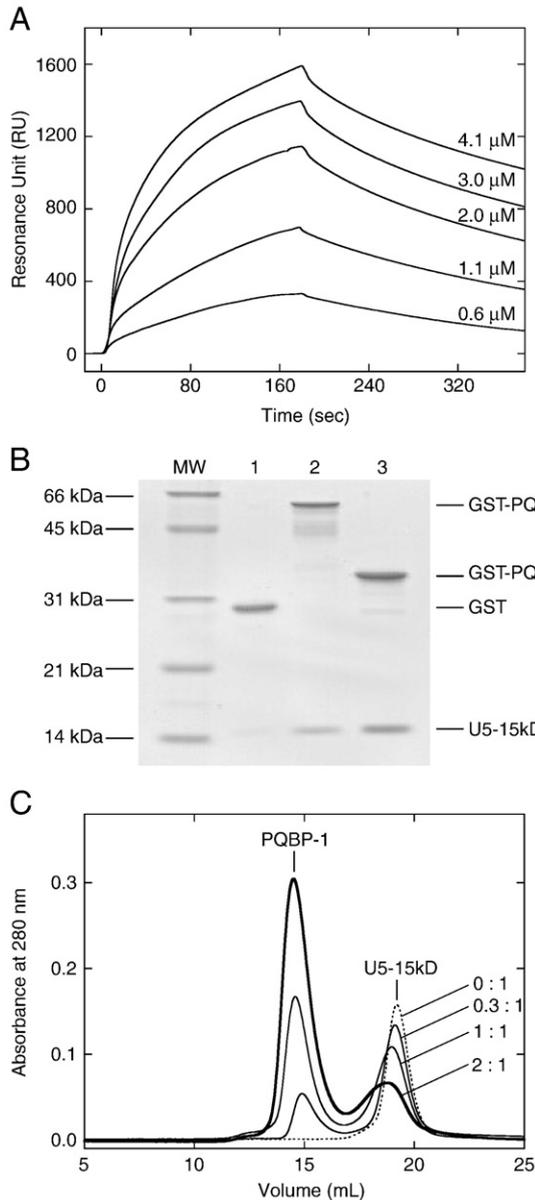


**Fig. 5.** Trypsin cleavage sites of PQBP-1 (A), PQBP-1-PR (B) and PQBP-1-CT (C). The vertical lines above the thick lines designate the predicted trypsin cleavage sites. The prediction was performed with Peptide Cutter (<http://br.expasy.org/tools/peptidecutter/>). Boxes indicate the positions of the WW, PRD and CTD. The vertical lines below the thick lines indicate the proteolytic cleavage sites after the 10 min incubation at 25 °C. The proteolytic cleavage sites were identified by mass spectrometry. The enzyme/substrate weight ratio is 1:500.

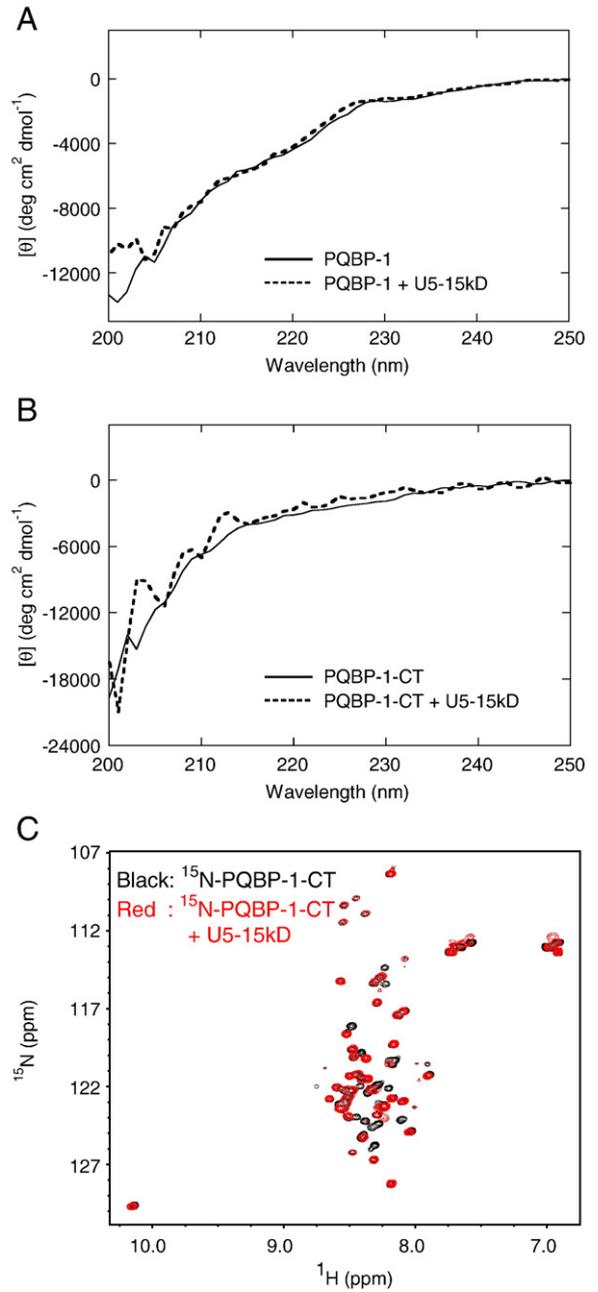
of trypsin. When the enzyme/substrate weight ratio was 1:10, trypsin cleavages occurred in many sites of PQBP-1-PR (Fig. S2B). Therefore, lack of detection of the predicted peptides under the mild proteolytic conditions indicates the protection from proteolysis (Fig. 5B).

### 3.5. Binding of PQBP-1 to U5-15kD

Binding of U5-15kD is one of the functions of PQBP-1. To determine the affinity of PQBP-1 to U5-15kD, we decided to analyze this interaction by surface plasmon resonance (Fig. 6A). We determined the apparent association rate constant and dissociation rate constant



**Fig. 6.** (A) BIAcore sensorgrams showing the interaction of PQBP-1 with U5-15kD. Solutions containing various concentrations of PQBP-1 (0.6, 1.1, 2.0, 3.0 and 4.1  $\mu\text{M}$ ) were injected over immobilized U5-15kD. (B) GST pull-down assays for analyzing the interaction between PQBP-1 and U5-15kD. U5-15kD was mixed with GST (lane 1 as negative control), GST-PQBP-1 (lane 2) and GST-PQBP-1-CT (lane 3), respectively. Proteins bound to glutathione sepharose beads were analyzed by an SDS-PAGE stained with Coomassie Brilliant Blue. MW indicates molecular weight markers. (C) Gel-filtration chromatograms of U5-15kD in the absence or presence of PQBP-1. All samples contained 23  $\mu\text{M}$  U5-15kD. The chromatogram of free U5-15kD is indicated by dotted line. Solid lines indicate the chromatograms of U5-15kD in the presence of PQBP-1. The molar ratio (PQBP-1:U5-15kD) for each chromatogram is indicated in the figure.



**Fig. 7.** (A) Far-UV CD spectra of PQBP-1 in the free (solid line) and the bound (dotted line) state with U5-15kD. (B) CD spectra of PQBP-1-CT in the free (solid line) and the bound (dotted line) state with U5-15kD. (A, B) The contribution of U5-15kD was subtracted from the CD spectrum in the bound state. (C) Superposition of the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of PQBP-1-CT in the absence (black) and presence (red) of U5-15kD. The molar ratio (PQBP-1:U5-15kD) is 1:1. The spectra were recorded at 15  $^\circ\text{C}$ .

for the complex between PQBP-1 and U5-15kD. PQBP-1 exhibited a fast association rate with U5-15kD ( $k_a = 4.1 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ ) and a slow dissociation rate ( $k_d = 2.5 \times 10^{-3} \text{ s}^{-1}$ ). Hence, the dissociation constant ( $K_D = k_d / k_a$ ) was calculated to be 0.6  $\mu\text{M}$ .

The binding was confirmed by GST pull-down assay and gel-filtration chromatography. The GST pull-down assay shows that either GST-PQBP-1 or GST-PQBP-1-CT binds to U5-15kD (Fig. 6B). We performed a quantitative analysis of densitometry of band intensities in a stained gel (Fig. 6B): 8 pmol of U5-15kD binds to 16 pmol of GST-PQBP-1 (lane 2 in Fig. 6B), and 54 pmol of U5-15kD binds to 71 pmol of GST-PQBP-1-CT (lane 3 in Fig. 6B). From this, we do not consider the possibility that only a small portion of U5-15kD binds to PQBP-1. In the

gel-filtration experiments, PQBP-1 and U5-15kD eluted at about 15 ml and 19 ml, respectively (Fig. 6C). In the presence of PQBP-1, U5-15kD eluted at slightly lower elution volume than in the free state. Furthermore, as the molar ratio of PQBP-1 increased, peak broadening of U5-15kD was observed. From these findings, we conclude that PQBP-1 weakly binds to U5-15kD.

We examined the possibility that the binding of U5-15kD might induce conformational change in PQBP-1. Fig. 7A shows the far-UV CD spectra of PQBP-1 in the free and the bound state with U5-15kD. PQBP-1 in the free and the bound form exhibited similar far-UV CD spectra typical of an unfolded protein. In addition, the CD spectrum of PQBP-1-CT in the free form was compared with that in the bound state (Fig. 7B), since PQBP-1 interacts with U5-15kD via the CTD. PQBP-1-CT in the presence of U5-15kD exhibited the CD spectrum typical of a largely unstructured conformation. Taken together, PQBP-1 includes high content of unstructured regions in the CTD, in spite of the binding of U5-15kD.

In addition to the CD studies, we used NMR to investigate the conformation of PQBP-CT. Fig. 7C shows the HSQC of <sup>15</sup>N-labeled PQBP-1-CT in the absence or presence of U5-15kD. Comparison of the two spectra shows that several resonances of PQBP-1-CT are strongly attenuated in the presence of U5-15kD. However, the attenuation, induced by the interaction, is limited to a low proportion of peaks. Moreover, the observable resonances in the presence of U5-15kD remain at the same chemical shifts as in the HSQC of the free PQBP-1-CT. These findings suggest that the binding of U5-15kD induces only minor conformational changes into PQBP-1-CT. Therefore, the binding of U5-15kD does not induce extensive tight packing interactions in PQBP-1.

#### 4. Discussion

PQBP-1 is predominantly localized to the nucleus [26], and is important for nuclear functions such as transcription and RNA processing [24–26]. For example, PQBP-1 WWD plays an important role in the regulation of transcription activity by interacting with the carboxy-terminal domain of Pol II [25], and PQBP-1 CTD plays an important role in the regulation of RNA processing by interacting with the spliceosomal protein U5-15kD [27]. These functions of PQBP-1 are perturbed by the binding of PRD with expanded polyglutamine tracts of disease proteins [24].

Our view of proteins is dominated by the notion that a well-defined three-dimensional structure is a prerequisite for their function. In the last decade, however, this structure–function paradigm has been re-evaluated based on the growing evidence that many proteins and protein domains are unfolded under physiological conditions. These proteins are referred to as “intrinsically unstructured” or “intrinsically disordered” [39]. The intrinsically unstructured/disordered proteins possess a non-rigid structure under physiological conditions; the molecule is extended, highly flexible, and has little secondary and tertiary structure. Indeed, it is now estimated that 35–51% of eukaryotic proteins have at least one long (>50 residues) unstructured region in their functional state [40]. Therefore, the group of intrinsically unstructured proteins deserves to be recognized as a new structural category within the protein kingdom [41].

In this work, we analyzed the structure of PQBP-1 and its fragments, and identified regions that contribute to the conformational properties of full-length PQBP-1. The <sup>1</sup>H–<sup>15</sup>N HSQC NMR as well as far-UV CD spectroscopy demonstrate that the PRD and CTD of PQBP-1 lack extensive contributions of ordered secondary and tertiary structure. The PRD and CTD encompass 23% and 29% of the total amino acid residues of PQBP-1, respectively, while the WWD with rigid structure encompasses only 13% of the total sequence of PQBP-1. Therefore, our results emphasize that PQBP-1 essentially contains large unstructured regions. PQBP-1 can thus be classified as a member of the emerging class of intrinsically unstructured proteins.

Numerous proteins lacking intrinsic globular structures have been identified, and unstructured regions seem to confer functional advantages in these proteins [39–41]. The functional advantage of lacking intrinsic globular structure, as opposed to a rigid one, may reside in the plasticity conferred by structural flexibility, which (1) allows binding to multiple targets with high specificity and low affinity and (2) allows the protein to overcome steric restrictions, thereby enabling larger surface interactions [39–41]. For PQBP-1, the flexibility may be the prerequisite to overcome steric restrictions when PQBP-1 interacts with many different targets, such as Pol II via WWD and U5-15kD via CTD.

The intrinsically unstructured proteins resemble the denatured states of globular proteins. However, the denatured states are not true random coils [42], and the random coil state is hardly compatible with the highly effective functioning of the intrinsically unstructured proteins [39–41]. In fact, the intrinsically unstructured proteins often possess only limited and transient, but significantly populated, secondary structure within an otherwise disordered polypeptide [37]. While the NMR and CD data demonstrate that the PRD and CTD have highly flexible, unfolded conformations, the PRD is not fully unfolded as indicated by the limited proteolysis experiments. Thus, the PRD possesses the residual structure, which may be functionally significant and may govern the preferential accessibility of the PRD destined for the polyglutamine tracts. On the other hand, the trypsin proteolysis does not indicate the residual structure in the CTD.

The present results demonstrated that the CTD of PQBP-1 lacks stable secondary and tertiary structure in the free state. So far, there have been numerous examples of proteins and protein domains that are unstructured in the free state but which become folded into rigid structures upon binding to the target molecules [39]. In contrast, the binding of the target molecule is not accompanied by global folding of the CTD of PQBP-1. Similar conclusions have been reported by Miron et al. [43] and Liu et al. [44]: the binding of the target molecule induces only minor conformational changes into the intrinsically unstructured proteins, the carboxy-terminal domain of xeroderma pigmentosum complementation group C protein [43] and a transcriptional activator ApLLP [44]. The disordered polypeptide chain has a greater capture radius than a rigid globular structure, and allows the protein to quickly find its specific binding site [45].

Our data suggest that PQBP-1 includes high content of unstructured regions in the CTD, in spite of the binding of U5-15kD. But how can U5-15kD recognize the specific binding site on PQBP-1? One possibility is that the folding occurs in a limited portion of the CTD, which includes key residues for the U5-15kD-binding. Another possibility is that U5-15kD recognizes a short extended segment, which includes the key residues arranged in a certain pattern.

Once PQBP-1 captures U5-15kD using the CTD, the spliceosome may be located in close proximity to Pol II, because PQBP-1 binds to Pol II via its WWD. When PQBP-1 connects the spliceosome and Pol II, PQBP-1 may act as a spacer molecule between these large multi-protein complexes. PQBP-1 may be important to regulate distance and to avoid steric hindrance between the transcription machinery and the splicing machinery.

#### Acknowledgments

This study was supported by Grants-in-Aids for Scientific Research on Priority Areas from the Ministry of Education, Culture, Science and Technology of Japan. This study was partially supported by the Program for the Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN).

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbapap.2009.03.001.

## References

- [1] The Huntington's Disease Collaborative Research Group, A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes, *Cell* 72 (1993) 971–983.
- [2] A.R. La Spada, E.M. Wilson, D.B. Lubahn, A.E. Harding, K.H. Fischbeck, Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy, *Nature* 352 (1991) 77–79.
- [3] R. Koide, T. Ikeuchi, O. Onodera, H. Tanaka, S. Igarashi, K. Endo, H. Takahashi, R. Kondo, A. Ishikawa, T. Hayashi, M. Saito, A. Tomoda, T. Miike, H. Naito, F. Ikuta, S. Tsuji, Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolusian atrophy (DRPLA), *Nat. Genet.* 6 (1994) 9–13.
- [4] S. Nagafuchi, H. Yanagisawa, E. Ohsaki, T. Shirayama, K. Tadokoro, T. Inoue, M. Yamada, Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidolusian atrophy (DRPLA), *Nat. Genet.* 8 (1994) 177–182.
- [5] S.H. Li, M.G. McInnis, R.L. Margolis, S.E. Antonarakis, C.A. Ross, Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms, *Genomics* 16 (1993) 572–579.
- [6] R.L. Margolis, S.H. Li, W.S. Young, M.V. Wagster, O.C. Stine, A.S. Kidwai, R.G. Ashworth, C.A. Ross, DRPLA gene (atrophin-1) sequence and mRNA expression in human brain, *Brain Res. Mol. Brain Res.* 36 (1996) 219–226.
- [7] Y. Kawaguchi, T. Okamoto, M. Taniwaki, M. Aizawa, M. Inoue, S. Katayama, H. Kawakami, S. Nakamura, M. Nishimura, I. Akiguchi, J. Kimura, S. Narumiya, A. Kakizuka, CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1, *Nat. Genet.* 8 (1994) 221–228.
- [8] S.M. Pulst, A. Nechiporuk, T. Nechiporuk, S. Gispert, X.N. Chen, I. Lopes-Cendes, S. Pearlman, S. Starkman, G. Orozco-Diaz, A. Lunke, P. DeJong, G.A. Rouleau, G. Aurburger, J.R. Korenberg, C. Figueroa, S. Sahba, Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2, *Nat. Genet.* 14 (1996) 269–276.
- [9] K. Sanpei, H. Takano, S. Igarashi, T. Sato, M. Oyake, H. Sasaki, A. Wakisaka, K. Tashiro, Y. Ishida, T. Ikeuchi, R. Koide, M. Saito, A. Sato, T. Tanaka, S. Hanyu, Y. Takiyama, M. Nishizawa, N. Shimizu, Y. Nomura, M. Segawa, K. Iwabuchi, I. Eguchi, H. Tanaka, H. Takahashi, S. Tsuji, Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, *DIRECT, Nat. Genet.* 14 (1996) 277–284.
- [10] G. Imbert, F. Saudou, G. Yvert, D. Devys, Y. Trottier, J.M. Garnier, C. Weber, J.L. Mandel, G. Cancel, N. Abbas, A. Dürr, O. Didierjean, G. Stevanin, Y. Agid, A. Brice, Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats, *Nat. Genet.* 14 (1996) 285–291.
- [11] M.K. Perez, H.L. Paulson, S.J. Pendse, S.J. Saionz, N.M. Bonini, R.N. Pittman, Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation, *J. Cell Biol.* 143 (1998) 1457–1470.
- [12] J.M. Boutell, P. Thomas, J.W. Neal, V.J. Weston, J. Duce, P.S. Harper, A.L. Jones, Aberrant interactions of transcriptional repressor proteins with the Huntington's disease gene product, *Hum. Mol. Genet.* 8 (1999) 1647–1655.
- [13] A. Kazantsev, E. Preisinger, A. Dranovsky, D. Goldgaber, D. Housman, Insoluble detergent-resistant aggregates form between pathological and nonpathological lengths of polyglutamine in mammalian cells, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 11404–11409.
- [14] J.S. Steffan, A. Kazantsev, O. Spasic-Boskovic, M. Greenwald, Y.Z. Zhu, H. Gohler, E.E. Wanker, G.P. Bates, D.E. Housman, L.M. Thompson, The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 6763–6768.
- [15] A. McCampbell, J.P. Taylor, A.A. Taye, J. Robitschek, M. Li, J. Walcott, D. Merry, Y. Chai, H. Paulson, G. Sobue, K.H. Fischbeck, CREB-binding protein sequestration by expanded polyglutamine, *Hum. Mol. Genet.* 9 (2000) 2197–2202.
- [16] T. Shimohata, T. Nakajima, M. Yamada, C. Uchida, O. Onodera, S. Naruse, T. Kimura, R. Koide, K. Nozaki, Y. Sano, H. Ishiguro, K. Sakoe, T. Ooshima, A. Sato, T. Ikeuchi, M. Oyake, T. Sato, Y. Aoyagi, I. Hozumi, T. Nagatsu, Y. Takiyama, M. Nishizawa, J. Goto, I. Kanazawa, I. Davidson, N. Tanese, H. Takahashi, S. Tsuji, Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription, *Nat. Genet.* 26 (2000) 29–36.
- [17] J.H. Cha, Transcriptional dysregulation in Huntington's disease, *Trends Neurosci.* 23 (2000) 387–392.
- [18] H.Y. Zoghbi, H.T. Orr, Glutamine repeats and neurodegeneration, *Annu. Rev. Neurosci.* 23 (2000) 217–247.
- [19] A. Lunke, Y. Trottier, J.L. Mandel, Pathological mechanisms in Huntington's disease and other polyglutamine expansion diseases, *Essays Biochem.* 33 (1998) 149–163.
- [20] J.S. Steffan, L. Bodai, J. Pallos, M. Poelman, A. McCampbell, B.L. Apostol, A. Kazantsev, E. Schmidt, Y.Z. Zhu, M. Greenwald, R. Kurokawa, D.E. Housman, G.R. Jackson, J.L. Marsh, L.M. Thompson, Histone deacetylase inhibitors arrest polyglutamine-dependent neurodegeneration in *Drosophila*, *Nature* 413 (2001) 739–743.
- [21] A. Wyttenbach, J. Swartz, H. Kita, T. Thykjaer, J. Carmichael, J. Bradley, R. Brown, M. Maxwell, A. Schapira, T.F. Orntoft, K. Kato, D.C. Rubinsztein, Polyglutamine expansions cause decreased CRE-mediated transcription and early gene expression changes prior to cell death in an inducible cell model of Huntington's disease, *Hum. Mol. Genet.* 10 (2001) 1829–1845.
- [22] K.B. Kegeles, A.R. Meloni, Y. Yi, Y.J. Kim, E. Doyle, B.G. Cuiffo, E. Sapp, Y. Wang, Z.H. Qin, J.D. Chen, J.R. Nevins, N. Aronin, M. DiFiglia, Huntingtin is present in the nucleus, interacts with the transcriptional corepressor C-terminal binding protein, and represses transcription, *J. Biol. Chem.* 277 (2002) 7466–7476.
- [23] H. Okazawa, Polyglutamine diseases: a transcription disorder? *Cell Mol. Life Sci.* 60 (2003) 1427–1439.
- [24] H. Okazawa, M. Sudol, T. Rich, PQBP-1 (Np/PQ): a polyglutamine tract-binding and nuclear inclusion-forming protein, *Brain Res. Bull.* 56 (2001) 273–280.
- [25] H. Okazawa, T. Rich, A. Chang, X. Lin, M. Waragai, M. Kajikawa, Y. Enokido, A. Komuro, S. Kato, M. Shibata, H. Hatanaka, M.M. Mouradian, M. Sudol, I. Kanazawa, Interaction between mutant ataxin-1 and PQBP-1 affects transcription and cell death, *Neuron* 34 (2002) 701–713.
- [26] M. Waragai, C.H. Lammers, S. Takeuchi, I. Imafuku, Y. Udagawa, I. Kanazawa, M. Kawabata, M.M. Mouradian, H. Okazawa, PQBP-1, a novel polyglutamine tract-binding protein, inhibits transcription activation by Brn-2 and affects cell survival, *Hum. Mol. Genet.* 8 (1999) 977–987.
- [27] M. Waragai, E. Junn, M. Kajikawa, S. Takeuchi, I. Kanazawa, M. Shibata, M.M. Mouradian, H. Okazawa, 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 (2000) 592–595.
- [28] K. Reuter, S. Nottrott, P. Fabrizio, R. Lüthmann, R. Ficner, Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein, *J. Mol. Biol.* 294 (1999) 515–525.
- [29] X. Li, P. Romero, M. Rani, A.K. Dunker, Z. Obradovic, Predicting protein disorder for N-, C-, and internal regions, *Genome Inform. Ser. Workshop Genome Inform.* 10 (1999) 30–40.
- [30] P. Romero, Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, A.K. Dunker, Sequence complexity of disordered protein, *Proteins* 42 (2001) 38–48.
- [31] S. Hirose, K. Shimizu, S. Kanai, Y. Kuroda, T. Noguchi, POODLE-L: a two-level SVM prediction system for reliably predicting long disordered regions, *Bioinformatics* 23 (2007) 2046–2053.
- [32] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (1995) 277–293.
- [33] B.A. Johnson, Using NMR View to visualize and analyze the NMR spectra of macromolecules, *Methods Mol. Biol.* 278 (2004) 313–352.
- [34] L. Stryer, The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites, *J. Mol. Biol.* 13 (1965) 482–495.
- [35] G.V. Semisotnov, N.A. Rodionova, V.P. Kutyschenko, B. Ebert, J. Blanck, O.B. Ptitsyn, Sequential mechanism of refolding of carbonic anhydrase B, *FEBS Lett.* 224 (1987) 9–13.
- [36] R. Kaul, A.R. Angeles, M. Jager, E.T. Powers, J.W. Kelly, Incorporating beta-turns and a turn mimetic out of context in loop 1 of the VVV domain affords cooperatively folded  $\beta$ -sheets, *J. Am. Chem. Soc.* 123 (2001) 5206–5212.
- [37] R.W. Kriwacki, L. Hengst, L. Tennant, S.I. Reed, P.E. Wright, Structural studies of p21<sup>Waf1/Cip1/Sdi1</sup> in the free and Cdk2-bound state: conformational disorder mediates binding diversity, *Proc. Natl. Acad. Sci. U. S. A.* 93 (1996) 11504–11509.
- [38] V. Csizsók, M. Bokor, P. Bánki, E. Klement, K.F. Medzihradský, P. Friedrich, K. Tompa, P. Tompa, Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2, *Biochemistry* 44 (2005) 3955–3964.
- [39] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm, *J. Mol. Biol.* 293 (1999) 321–331.
- [40] A.K. Dunker, C.J. Brown, J.D. Lawson, L.M. Iakoucheva, Z. Obradović, Intrinsic disorder and protein function, *Biochemistry* 41 (2002) 6573–6582.
- [41] P. Tompa, Intrinsically unstructured proteins, *Trends Biochem. Sci.* 27 (2002) 527–533.
- [42] D. Shortle, The denatured state (the other half of the folding equation) and its role in protein stability, *FASEB J.* 10 (1996) 27–34.
- [43] S. Miron, P. Duchambon, Y. Blouquit, D. Durand, C.T. Craescu, The carboxy-terminal domain of xeroderma pigmentosum complementation group C protein, involved in TFIIH and centrin binding, is highly disordered, *Biochemistry* 47 (2008) 1403–1413.
- [44] J. Liu, J. Song, A novel nucleolar transcriptional activator ApLLP for long-term memory formation is intrinsically unstructured but functionally active, *Bochem. Biophys. Res. Commun.* 366 (2008) 585–591.
- [45] B.A. Shoemaker, J.J. Portman, P.G. Wolynes, Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, *Proc. Natl. Acad. Sci. U. S. A.* 97 (2000) 8868–8873.