Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain

Philippa Davies*, **Kate Watt***, **Sharon M Kelly**[1], **Caroline Clark**[2], **Nicholas C Price**[1] and **Iain J McEwan**

School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK

[1]Institute of Biomedical and Life Sciences, University of Glasgow, Joseph Black Building, Glasgow G12 8QQ, Scotland, UK

[2]School of Medicine, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK

(Correspondence should be addressed to I J McEwan; Email: iain.mcewan@abdn.ac.uk)

*(P Davies and K Watt contributed equally to this work)

Abstract

Poly-amino acid repeats, especially long stretches of glutamine (Q), are common features of transcription factors and cell-signalling proteins and are prone to expansion, resulting in neurodegenerative diseases. The amino-terminal domain of the androgen receptor (AR-NTD) has a poly-Q repeat between 9 and 36 residues, which when it expands above 40 residues results in spinal bulbar muscular atrophy. We have used spectroscopy and biochemical analysis to investigate the structural consequences of an expanded repeat (Q45) or removal of the repeat (ΔQ) on the folding of the AR-NTD. Circular dichroism spectroscopy revealed that in aqueous solution, the AR-NTD has a relatively limited amount of stable secondary structure. Expansion of the poly-Q repeat resulted in a modest increase in α-helix structure, while deletion of the repeat resulted in a small loss of α-helix structure. These effects were more pronounced in the presence of the structure-promoting solvent trifluoroethanol or the natural osmolyte trimethylamine *N*-oxide. Fluorescence spectroscopy showed that the microenvironments of four tryptophan residues were also altered after the deletion of the Q stretch. Other structural changes were observed for the AR-NTDQ45 polypeptide after limited proteolysis; in addition, this polypeptide not only showed enhanced binding of the hydrophobic probe 8-anilinonaphthalene-1-sulphonic acid but was more sensitive to urea-induced unfolding. Taken together, these findings support the view that the presence and length of the poly-Q repeat modulate the folding and structure of the AR-NTD.

Journal of Molecular Endocrinology (2008) **41,** 301–314

Introduction

The androgen receptor (AR) is thought to be the sole mediator of the actions of the steroid hormones testosterone and dihydrotestosterone in male reproductive tissues, such as the testes, epididymis and prostate (reviewed in Choong & Wilson (1998) and McEwan (2004)). Mutations in the AR can lead to a disruption of male development (androgen insensitivity) and to a neuromuscular degenerative disorder termed spinal bulbar muscular atrophy (SBMA; also called Kennedy's disease; Quigley *et al.* 1995, Choong & Wilson 1998, Palazzolo *et al.* 2008 and references therein). The AR is a member of the nuclear receptor superfamily, and is organized into distinct structural and functional domains. The C-terminal ligand-binding domain and the central DNA-binding domain show significant identity both between ARs of different species and with other members of the nuclear receptor superfamily (see Choong & Wilson 1998, Lavery & McEwan 2005 and references therein). By contrast, the sequence of the N-terminal domain (NTD) of the protein is more divergent, and is characterized by homopolymer tracts of glutamine, glycine and proline residues (reviewed in Choong & Wilson (1998), Choong *et al.* (1998), Lavery & McEwan (2005) and McEwan *et al.* (2007)). Regions within the N-terminus of the human and rat receptors, termed AF1, that are important for transactivation have been delineated by deletion analysis (Simental *et al.* 1991, Jenster *et al.* 1995), the use of fusion proteins (Jenster *et al.* 1995, McEwan & Gustafsson 1997) and the study of point mutations (Chamberlain *et al.* 1996, Betney & McEwan 2003). The AF1 domain is structurally flexible and has the propensity to form α-helix structure (Reid *et al.* 2002, Kumar *et al.* 2004*a*). The transactivation domain makes multiple protein–protein interactions with general transcription factors and co-regulatory proteins (reviewed in Lavery & McEwan (2005)).

To date, nine neurodegenerative diseases have been described, including Huntington's disease and SBMA,

0952–5041/08/041–301
DOI: 10.1677/JME-08-0042
Online version via http://www.endocrinology-journals.org

where the underlying defects involve expansion of the triplet codon CAG, which codes for the amino acid - glutamine (Q; Choong & Wilson 1998, Paulson *et al.* 2000, Poletti 2004, Gatchel & Zoghbi 2005). The proteins - involved lack significant sequence identity outside the poly-Q repeat, suggesting a common pathology associated directly with the repeat, which behaves as a 'gain-of-function' mutation. These disorders are also characterized by the different populations of neurones affected (Ross 1995). The AR-NTD has three poly-Q repeats, the largest of which is polymorphic and between 9 and 36 residues in length. Expansion of this repeat to between 38 and 65 glutamines results in motor neurone cell death in the brain stem and spinal cord leading to progressive muscle wasting (reviewed in La Spada & Taylor (2003) and Poletti (2004)). Despite extensive research, the pathology of this disease remains unclear. Studies from animal models and cell culture have suggested that hormone activation, aggregation, fragmentation and/or subcellular localization could all play a role in disease aetiology or progression (reviewed in La Spada & Taylor (2003), Poletti (2004) and Palazzolo *et al.* (2008)). However, recently, evidence has been presented for the formation of 'amyloid ion channels' and disruption of ionic homeostasis and protein-misfolding diseases (Quist *et al.* 2005). A consideration of protein sequences of the AR and other triplet disease proteins reveals that repeats prone to expansion are limited to higher primates (Djian *et al.* 1996, Choong & Wilson 1998). Thus, it could be speculated that the presence of these repeats may have had an evolutionary advantage and may play a role in the normal function of the target protein.

The structural nature of poly-Q repeats and the consequences of expansion have been the subject of a number of studies and considerable debate. A variety of different conformations have been observed or modelled with isolated peptides having variable poly-Q repeat lengths and flanking amino acids. These include random coil (Altschuler *et al.* 1997, Starikov *et al.* 1999, Chen *et al.* 2001), β-sheet/turn (Perutz *et al.* 1994, Lathrop *et al.* 1998, Sharma *et al.* 1999, Tanaka *et al.* 2002) or α-helix (Bhattacharyya *et al.* 2006), together with more unusual types of secondary structure, such as α-sheet (Armen *et al.* 2005), μ-helix (Monoi 1995) and polyproline II helical structure (Chellgren *et al.* 2006). Fusion proteins, containing glutathione S-transferase (GST) or thirodoxin were observed to contain either random coil (Bennett *et al.* 2002, Masino *et al.* 2002) or α-helical conformations (Popiel *et al.* 2004, Nagai *et al.* 2007), while insertion of poly-Q repeats into host proteins such as sperm whale myoglobin or chymotrypsin inhibitor 2 resulted in the formation of either random coil (Finke *et al.* 2004) or a mixture of random coil and β-sheet (Tanaka *et al.* 2001, 2002) conformations. In studies where expansion of the poly-Q repeat sequence and structure were examined, a similar degree of diverse conformations have been reported. These range from there being no change in peptide or protein conformation (Chen *et al.* 2001, Bennett *et al.* 2002, Chellgren *et al.* 2006) to an increase in β-turn structure (Tanaka *et al.* 2001) or α-helix secondary structure (Nagai *et al.* 2007). Interestingly, the inclusion of poly-proline sequences, C-terminal to a poly-Q repeat, resulted in an increase in random coil (Popiel *et al.* 2004, Bhattacharyya *et al.* 2006). This may be significant, as recent reports have correlated sequence context and the nature of flanking amino acid sequences with modulation of the cytotoxicity of poly-Q repeats (Chai *et al.* 2001, Dehay & Bertolotti 2006, Duennwald *et al.* 2006). To date, the only studies on the structure of a full-length (or nearly full-length) poly-Q disease protein are for ataxin-3, which causes spinocerebellar ataxia 3 or Machado-Joseph's disease. The poly-Q repeat was found to be flexible and solvent exposed (Masino *et al.* 2003) and expansion of the repeat from Q27 to Q78 resulted in reduced α-helix content and formation of β-sheet fibrils (Bevivino & Loll 2001). However, Chow *et al.* (2004) found no major differences in secondary structure content for ataxin-3 with 15, 28 or 50 glutamines.

What emerges from the above studies is that the poly-Q repeat may adopt multiple conformations and that the amino acid sequences flanking the repeat are likely to influence these conformations. Clearly, to obtain a better understanding of the structure and function of poly-Q repeats, it will be important to study the repeat in the context of the target protein. In the present study, we used biophysical and biochemical approaches to investigate the conformation of the AR-NTD with variable (0, 20 and 45) poly-Q repeat lengths. These studies revealed that removal of the large poly-Q repeat led to a reduction in local structural flexibility and a reduction in α-helix content. Increasing the poly-Q repeat caused modest changes in secondary structure content and the generation of a unique pattern of fragments after limited proteolysis with endo Glu-C protease. Collectively, the findings suggest that poly-Q length modulates the conformation and folding of the AR-NTD, which in turn is likely to impact upon receptor function.

Materials and methods

Construction of expression plasmids

DNA fragments coding for the human AR-NTD, amino acids 1–537, were amplified by PCR using the following primers: ARN 5′ GCGCGCAGATCTATGGAAGTGCAGTTAGGGCTGGGAAGGG 3′ and ARC 5′ GCGCGCGGATCCCATGTCCCCGTAAGGTCCGG 3′. PCR products were generated using a Pwo PCR Kit (Hybaid, Basing Stoke, Hamps, UK) and digested with BamHI and BglII restriction enzymes (sites underlined in sequences

above) and cloned into the expression plasmids pET-19bm and pGEX-2TK digested with BamHI. Plasmids pSVARo (Brinkmann *et al.* 1989) and pSVAR70 (Jenster *et al.* 1994; both kindly provided by Dr Brinkmann, Erasmus University, Rotterdam, The Netherlands) were used as the template DNA to generate AR-NTDQ20 and AR-NTDΔQ plasmids respectively. Expression plasmids encoding an expanded poly-Q repeat (AR-NTDQ45) were generated using the NTD sequence amplified using the genomic DNA derived from a patient with SBMA. Positive clones were identified by restriction digest screening and the presence of the inserts confirmed by DNA sequencing.

Expression and purification of recombinant proteins from *Escherichia coli*

E.coli cells, strain BLR (DE3) containing the appropriate recombinant expression plasmid were grown overnight and diluted 1:20 in 2×TY media, containing ampicillin and 0·5% (w/v) glucose and grown at 37 °C until an attenuance at 600 nm of 0·6–0·9 was reached. Expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 90 min at 37 °C and the cells collected and lysed using a combination of freeze thawing and treatment with 0·5 mg/ml lysozyme. Supernatants were applied to Ni^{2+}-nitriloacetate -agarose columns and eluted with imidazole: 0·2 M imidazole, 0·5 M NaCl, 20 mM Tris–HCl (pH 7·9) and 5% glycerol. Purified recombinant proteins were dialysed from elution buffer into dialysis buffer (25 mM Hepes (pH 7·9), 100 mM sodium acetate, 5% glycerol and 1 mM dithiothreitol) and snap frozen in liquid nitrogen and stored at −80 °C. Purified proteins were analysed by SDS-PAGE and protein concentration was measured by the method of Bradford (1976) using BSA standards.

Structural analysis of the AR-NTD: partial proteolysis assay

Purified AR-NTD recombinant proteins were partially digested with proteases to compare protein tertiary structure. Totally 25 pmol protein was incubated with 0·1 ng/μl trypsin or 4 ng/μl endo Glu-C in proteolysis buffer in a total reaction volume of either 10 or 25 μl for the times indicated at room temperature (25 °C). Reactions were stopped by addition of an equal volume of 2×SDS sample buffer and heating at 75 °C for 5 min. Exactly 15 μl of each sample was analysed by SDS-PAGE and either silver stained or transferred to nitrocellulose for western blot analysis using antibodies against amino acids 1–21 (ab3510, Abcam, Cambridge, UK), amino acids 299–315 (sc441, Santa Cruz Biotechnology, Santa Cruz, CA, USA) or recognizing the expanded poly-Q repeat (1C2, Chemicon, Chandlers Ford, Hamps, UK).

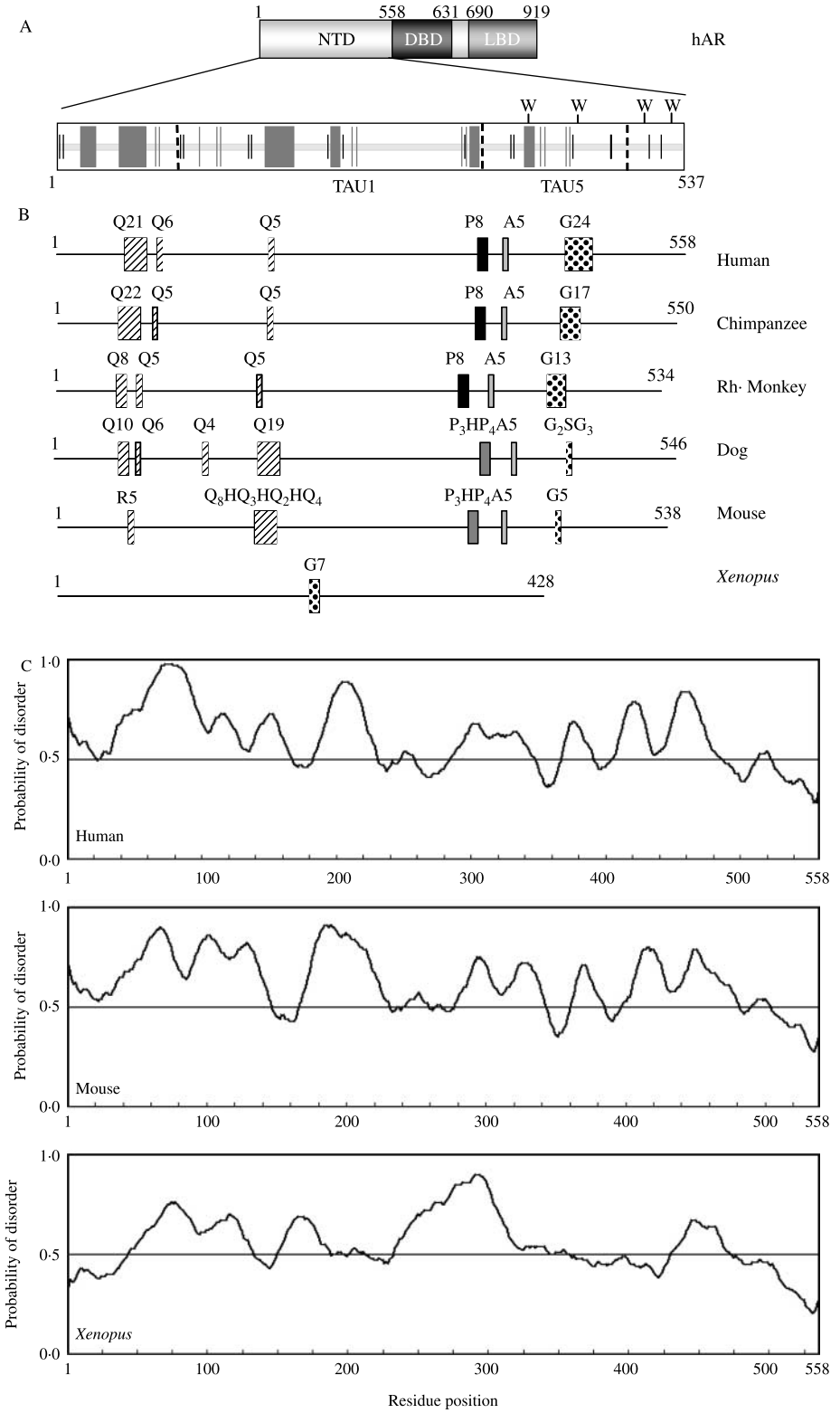
Structural analysis of the AR-NTD: fluorescence spectroscopy

The steady-state fluorescence emissions from tyrosine and tryptophan residues in AR-NTD polypeptides were measured using a Shimadzu RF-1501 spectrofluorimeter. Totally 25 μg/ml protein was diluted in dialysis buffer to a total volume of 1 ml. Measurements were taken using excitation wavelengths of 278 and 295 nm and under conditions of induced folding or unfolding with the osmolyte TMAO and urea respectively. All measurements were made in duplicate or triplicate and corrected for the contributions of buffers and solutes.

Fluorescence studies with the hydrophobic probe 8-anilinonaphthalene-1-sulphonic acid (ANS; Fluka) were performed essentially as previously reported (Cardamone & Puri 1992, Bailey *et al.* 2001, Lavery & McEwan 2008). Binding of ANS to the AR-NTD polypeptides during unfolding was monitored by recording the fluorescence over the range of 420–530 nm after excitation at 370 nm; the bandwidths for both excitation and emission were set at 10 nm. AR-AF1 polypeptides, at final concentrations of 1 μM, were incubated with ANS (50 μM) in PBS or PBS and urea (0·2–6 M) for 30 min at room temperature in the darkness. Experiments were performed at least three times. In order to compare the unfolding of the various protein samples, the results are presented as a ratio of fluorescence in the presence of urea (F) over the sample in the absence of urea (F_0; mean ± S.D) for at least three independent experiments. All measurements were corrected for both buffer and background (non-bound) ANS fluorescence.

Structural analysis of the AR-NTD: circular dichroism (CD) spectroscopy

Purified AR-NTDQ20, AR-NTDQ45 and AR-NTDΔQ were dialysed against 4 mM NaH_2PO_4, 6 mM Na_2HPO_4, 100 mM sodium sulphate, 1 mM dithiothreitol (pH 7) for CD analysis. The far u.v. CD spectra for each protein were measured at 20 °C on a Jasco J-810 spectropolarimeter calibrated with (1*S*)-(+)-10-camphorsulphonic acid. Far u.v. CD spectra (190–260 nm) were measured at concentrations in the range of 0·39–1·27 mg/ml using a cell of 0·02 cm path length. The far u.v. CD spectra for the polypeptides were also measured in the presence of 50% (v/v) trifluoroethanol (TFE) or 2 M trimethylamine *N*-oxide (TMAO). The CD spectra were measured in duplicate using two different preparations of recombinant protein with similar results. The spectra obtained were the average of eight individual spectra, with less than 5% error between each measurement. The proportions of each secondary structure type were estimated from the CD data using the CDSSTR procedure (Sreerama & Woody



1993). In each case, the quality of the fitting procedure was judged by the very low value ($\leq 0{\cdot}01$) of the normalized root mean square deviation and the excellent superposition of the experimental and reconstructed spectra.

Results

Species comparison of poly-amino acid repeats in the AR-NTD

The AR-NTD contains several amino acid repeats, two of which, a poly-Q and poly-G repeat, show polymorphisms in the human population (Fig. 1A and B). Strikingly, these two repeats appear to be a feature of higher primates (Djian *et al.* 1996, Choong & Wilson 1998, Choong *et al.* 1998), and are either shorter, interrupted or absent in other mammalian species and *Xenopus* and fish ARs (Fig. 1B and data not shown). This has been observed for other poly-Q repeats that are prone to expansion and associated with neurodegenerative disease (Djian *et al.* 1996). One consequence of these amino acid repeats could be to contribute to the predicted intrinsic disorder of the AR-NTD (reviewed in Lavery & McEwan (2005)). Figure 1C shows the naturally unstructured regions predicted by the neural network RONN (Yang *et al.* 2005) for the human, mouse and *Xenopus* AR-NTD. Indeed, the repeats, particularly for the human and mouse AR, do coincide with peaks of disordered structure (Fig. 1C). Interestingly, the poly-A sequence, where present, is predicted to be ordered. However, they are clearly not the only determinant, as *Xenopus* AR-NTD, which lacks the majority of these amino acid repeats, is still predicted to be 50–55% unstructured.

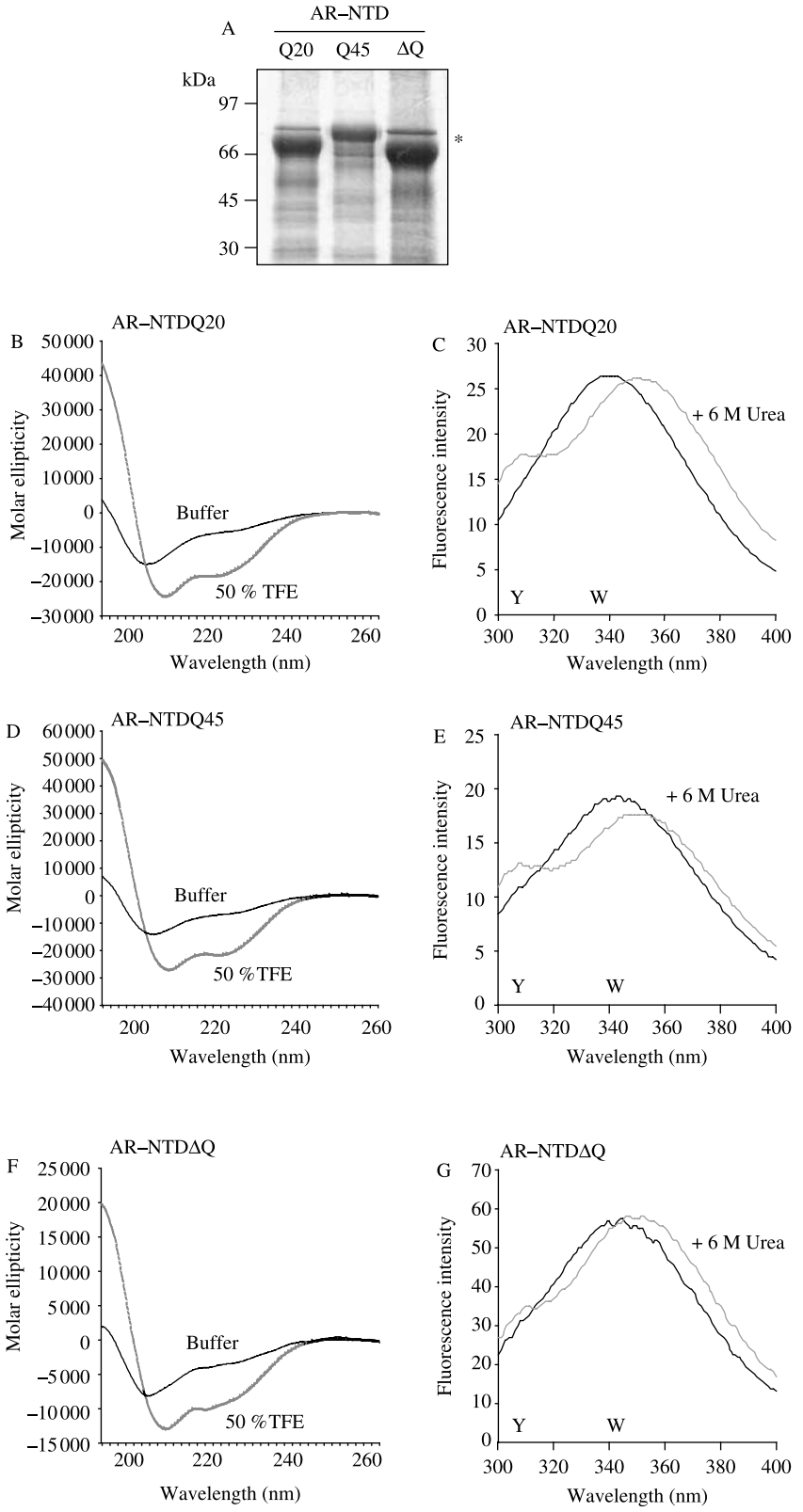
Spectroscopic analysis of the AR-NTD with Q0, Q20 or Q45 repeats

To measure directly the conformational consequences of changes in the poly-Q amino acid repeat length on the human AR-NTD, we have expressed and purified recombinant polypeptides with 0, 20 or 45 residues. The AR-NTD is coded by a single exon and a repeat of 45Q was recovered from the genomic DNA of a patient with Kennedy's disease. After expression in *E. coli* and purification from the soluble bacterial extract by Ni-affinity chromatography, the proteins were judged to be at least 85% pure on SDS-PAGE (Fig. 2A).

Figure 2B shows the CD spectra for wild-type AR-NTDQ20 in buffer and in the presence of a hydrophobic solvent, TFE, that has been reported to stabilize α-helical structure. The spectrum in buffer is typical for a protein with a relatively limited amount of stable secondary structure, showing a minimum at around 200 nm. The estimated secondary structure content indicates a polypeptide with 14% α-helix, 27% β-strand, 24% turns and 34% non-ordered structure (Table 1). In the presence of 50% (v/v) TFE, a solvent that helps stabilize secondary structure, the spectrum is more typical of a protein with significant α-helix content, with minima at 208 and 222 nm, and a calculated value of 58% α-helix (Fig. 2B, Table 1). There is concomitant reduction of the non-ordered structure (to 22%) and β-strand (to 12%; Table 1). There is a similar change in the spectrum in the presence of the natural osmolyte TMAO, which is thought to stabilize proteins in a native conformation (data not shown). It should be noted that in the presence of TMAO, reliable CD data could only be obtained down to 210 nm because of absorbance of the osmolyte; this precludes detailed analysis of the secondary structure under these conditions. Therefore, similar to our previous studies with the isolated AF1 domain (Reid *et al.* 2002, Kumar *et al.* 2004*a*,*b*), the AR-NTD has the propensity to form α-helical structure in a hydrophobic environment or in the presence of the osmolyte TMAO.

To investigate the folding of the AR-NTD, we measured the intrinsic fluorescence emission spectra under different conditions. The AR-NTD contains four tryptophan amino acids at 396, 432, 493 and 517 (Fig. 1A). The λ_{max} for tryptophan emission is dependent upon the local environment surrounding the aromatic residues and exposure to solvent. Normally, the emission from tyrosine residues is not observed as there can be quenching due to a distance-dependent energy transfer to neighbouring tryptophan residues and/or to proton transfer from the excited state (see Eftink & Ghiron (1976)). The steady-state emission spectrum for AR-NTD will therefore provide information on the local structural environment surrounding the tryptophan residues and will allow the potential 'long-range' influence of the poly-Q repeat to be studied. Figure 2C shows the emission spectrum in buffer and in the presence of the denaturant urea after excitation at 278 nm. The λ_{max} for tryptophan in buffer was 342 ± 2 nm ($n=4$), which is markedly red shifted to 350 nm in the presence of urea. This is consistent with the tryptophan residues becoming more solvent exposed as the polypeptide chain is unfolded; in addition, a peak was

Figure 1 The human AR domain structure and amino acid analysis. (A) Schematic representation of the human AR domain organization: NTD, N-terminal domain; DBD, DNA-binding domain; and LBD, ligand-binding domain. Below is a representation of the predicted secondary structure content for the NTD and the location of the transactivation units (TAU) 1 and 5 (Jenster *et al.* 1995) that constitute AF-1 and the position of four tryptophan residues. Large and small bars represent α-helix and β-strand respectively. (B) Schematic representation of the AR-NTD from different species illustrating the changes in amino acid repeat lengths and location. (C) RONN (Yang *et al.* 2005) prediction of natural disorder structure for the human, mouse and *Xenopus* AR-NTD. Peaks above and below the 0·5 threshold represent disordered and ordered structure respectively.



also observed at around 305 nm due to tyrosine residues, indicating less efficient quenching by energy transfer in the unfolded protein (Fig. 2C). These data are consistent with the AR-NTDQ20 having a limited amount of stable folded structure that is lost upon treatment with urea. By contrast, in the presence of TMAO, there is a blue shift of the λ_{max} to 332 nm, suggesting the tryptophans are becoming less solvent exposed, indicating that at least this C-terminal region of the polypeptide becoming more compact (Table 2).

These experiments were repeated with AR-NTD polypeptides with an expanded (Q45) or deleted (ΔQ) poly-Q repeat to determine the effect of repeat length on AR-NTD conformation. Figure 2D and F show the CD spectra in buffer and TFE for AR-NTDQ45 and AR-NTDΔQ respectively. Table 1 summarizes the estimated secondary structure content for both polypeptides. Interestingly, expansion of the poly-Q repeat resulted in a small, but measurable increase in α-helix content (to 16%) and a small decrease in the β-structures (to 23% strand and 22% turn). By contrast, in the case of the AR-NTDΔQ polypeptide, the α-helix content was reduced (to 11%) with small increases in β-structures (to 28% strand and 25% turns). In the presence of TFE, there was a further increase in the α-helix content of AR-NTDQ45 compared with the wild-type polypeptide and again this was accompanied by decreases in both non-ordered and β-structure (Fig. 2D, Table 1). A similar increase in α-helix was observed in the presence of TMAO (data not shown). Although there was an increase in α-helix content for the AR-NTDΔQ polypeptide on addition of TFE, this was markedly less than that for Q20 and Q45 and did not appear to result from induced folding of significant amounts of the non-ordered structure (Fig. 2F, Table 1). Overall, the findings suggested that deletion of the poly-Q repeat resulted in a polypeptide, which not only had a lower proportion of α-helical structure in aqueous solution, but also a lower propensity to form such structure in the presence of TFE.

The values of λ_{max} for tryptophan emission for AR-NTDQ45 and subsequent changes in the presence of urea and TMAO were similar to those observed for the wild-type Q20 polypeptide. The λ_{max} was $342 \pm 0{\cdot}5$ nm ($n=4$) in buffer and this was red shifted to 351 nm or blue shifted to 332 nm in the presence of urea or TMAO respectively (Fig. 2E and Table 2). By contrast, the λ_{max} for AR-NTDΔQ was 343 ± 3 nm ($n=3$) in buffer, but showed much reduced shifts in urea (to 347 nm) and TMAO (to 336 nm; Fig. 2G and Table 2). Taken together, the spectroscopic data suggest that deleting the poly-Q repeat results in a less flexible polypeptide with reduced α-helical structure. Increasing the poly-Q repeat had modest effects on the secondary structure content, but did not appear to change significantly the tertiary structure of the polypeptide chain surrounding the tryptophan residues.

The stability of the AR-NTD polypeptides to denaturing conditions was further studied using the hydrophobic fluorescent probe ANS. The interaction of ANS with various proteins has been used previously to characterize hydrophobic surfaces and structural conformations (Semisotnov *et al.* 1991, Cardamone & Puri 1992, Sharma *et al.* 1998, Bailey *et al.* 2001). In aqueous solution, ANS shows only very low fluorescence when excited at 370 nm, with an emission maximum at 533 nm. However, on binding to hydrophobic patches or clusters in proteins, there is a considerable enhancement in fluorescence with a shift in the emission maximum to about 480 nm. There is also a marked enhancement in ANS fluorescence when ANS binds to 'molten globule-type' proteins (which possess limited stable tertiary structure); in such cases, the binding of the dye is greater than that to either fully folded (globular) or unfolded (random coil) protein conformations (Semisotnov *et al.* 1991, Cardamone & Puri 1992). Figure 3A shows the ANS emission spectrum in buffer and when bound to AR-NTD polypeptides with different poly-Q repeat sizes. For each AR-NTD polypeptide, there was a significant increase in ANS fluorescence, which was most dramatic with AR-NTDQ45 protein. Figure 3B shows the loss of ANS fluorescence for each AR-NTD polypeptide with increasing concentrations of urea. All three polypeptides showed non-cooperative unfolding and AR-NTDQ45 was consistently more sensitive to urea denaturation than either the wild-type (Q20) or the AR-NTDΔQ polypeptides. The marked enhancement of ANS fluorescence on binding to the AR-NTD polypeptides taken together with the lack of cooperative unfolding by increasing concentrations of urea indicates that each polypeptide has at least some 'molten globule' character with limited stable tertiary structure.

Partial proteolysis of AR-NTD polypeptides

To investigate further the structural implications of poly-Q repeat length on the conformation of the AR-NTD, sensitivity to limited proteolysis was used as a probe for global protein structure. All three polypeptides were

Figure 2 Spectroscopic analysis of the AR-NTD with different lengths of poly-Q repeat. (A) A representative Coomassie stained SDS-polyacrylamide gel showing the purified AR polypeptides: AR-NTDQ20 (9 μg), Q45 (6 μg) and ΔQ (13 μg). *A co-purifying bacterial protein. (B). Far u.v.-CD spectra for the AR-NTDQ20 (0·39 mg/ml) in aqueous buffer (black curve) or 50% TFE (grey curve). (C) Steady-state fluorescence spectra for AR-NTDQ20 (25 μg/ml) in buffer (black curve) or 6 M urea (grey curve) after excitation at 278 nm. The locations of the tyrosine and tryptophan emissions are indicated. (D–G) As for (B) and (C), but analysis of AR-NTDQ45 (D and E) and AR-NTDΔQ (F and G).

Table 1 Summary of secondary structure content

	Secondary structure[a]			
	α-Helix (%)	β-Strand (%)	Turn (%)	Non-ordered (%)
Protein				
ARN-Q20/Buffer	14	27	24	34
ARN-Q45/Buffer	16	23	22	39
ARN-ΔQ/Buffer	11	28	25	35
ARN-Q20/TFE	58	12	9	22
ARN-Q45/TFE	66	9	5	19
ARN-ΔQ/TFE	30	21	20	30

[a]Secondary structure determinations by the CDSSTR procedure of Sreerama & Woody (1993). The criteria for judging the quality of the fitting procedure are described in the Materials and methods section.

sensitive to digestion by endo Glu-C protease, with the full-length protein completely digested between 5 and 15 min (Fig. 4). In the presence of 2M TMAO, the full-length AR-NTDQ20 and NTDΔQ, and to a lesser extent NTDQ45, were completely protected or stabilized from protease digestion (Fig. 4). This is consistent with the induced folding of the AR-NTD structure in the presence of the osmolyte. It has previously been shown that TMAO does not impair protease activity (see Reid *et al.* 2002, Kumar *et al.* 2005).

Investigation of the fragments generated after partial proteolysis with endo Glu-C and trypsin using antibodies against different epitopes within the NTD or against the expanded glutamine repeat revealed further evidence of conformational changes for ARN-Q45. Partial cleavage with endo Glu-C results in digestion of full-length ARN-Q20, -Q45 and -ΔQ polypeptides as noted above. Interestingly, the α-N441 antibody, which recognizes an epitope with amino acids 299–315, detected a series of smaller fragments for ARN-Q45: 49, 30·5 and 25·5 kDa (Fig. 5A, left panel). These fragments were not visible for the wild-type NTD or the NTD lacking the poly-Q repeat. Further analysis using the α-1C2 antibody, which is selective for the expanded poly-Q repeat, identified a series of similar-sized fragments containing the expanded repeat: 57, 31 and 21 kDa (Fig. 5A, right panel). Interestingly, this antibody also detected higher molecular weight species, not seen on stained gels, that may represent dimers or trimers of the NTD with an expanded glutamine repeat (Fig. 5A, right panel). Based on the molecular size of the fragments generated, and correcting for the expansion of the poly-Q repeat, a number of potential endo Glu-C cleavage sites were identified: glutamic acids 32, 43, 150, 198 and 388. Note that the 30·5 and 25·5 kDa fragments identified with α-N441 are unlikely to contain the epitope for α-1C2.

Trypsin treatment similarly resulted in digestion of the full-length NTD polypeptides and major N-terminal fragments of 38, 49 and 33 kDa for ARN-Q20, -Q45 and ΔQ respectively (Fig. 5B). The fragment was most prominent for the NTD with the expanded repeat, but was clearly visible with all three polypeptides. Based on the molecular size of the fragments generated, and the theoretical cleavage pattern, a number of potential trypsin cleavage sites were identified, between lysine 235 and lysine 289. Analysis of the products of the partial trypsin digest for ARN-Q45 using α-1C2 revealed an additional series of fragments containing the expanded poly-Q repeat: 34, 28 and 23 kDa. The absence of these fragments with α-ab3510 suggests that the first 21 amino acids have been cleaved off. Taken together the antibody epitope results suggest that the expanded glutamine repeat alters the conformation of the AR-NTD and selective sensitivity to endo Glu-C digestion.

Table 2 Summary of steady-state fluorescence emission spectra

	λ_{max} Tryptophan emission[a]		
	Buffer (nm)	Urea (nm)	TMAO (nm)
Protein			
ARN-Q20	342±2 (n=4)	350	332
ARN-Q45	342±0·5 (n=4)	351	332
ARN-ΔQ	343±3 (n=3)	347	336

[a]λ_{ex} was 278 nm.

Discussion

It is increasingly clear that the amino acids flanking the poly-Q repeat in proteins involved in neurodegenerative diseases may influence both the structure and the pathology of the repeat. Therefore, analysis of the repeats in the context of the native protein will be important for the full understanding of the effect of the poly-Q stretch on protein structure and function. In order to understand the structural effects of the largest poly-Q repeat on the conformation and folding of the human AR-NTD in more detail, we studied polypeptides with 0 (ΔQ), 20 (wild-type) or 45 (SBMA) glutamine residues. The normal range for this glutamine stretch is from 9 to 36, while larger repeats (from 38 to 65) are associated with SBMA (Paulson *et al.* 2000, Poletti 2004, Gatchel & Zoghbi 2005). Using CD spectroscopy, we show that the AR-NTD with an expanded repeat increases the α-helix content and reduces β-structures. By contrast, deleting the poly-Q repeat resulted in an AR-NTD polypeptide with decreased α-helical structure and a concomitant increase in β-strand and turn secondary structure. While all three AR-NTD polypeptides had the propensity to adopt a more helical conformation in a hydrophobic environment or in the presence of a natural osmolyte (TMAO), the tendency to form α-helix secondary structure was less pronounced in

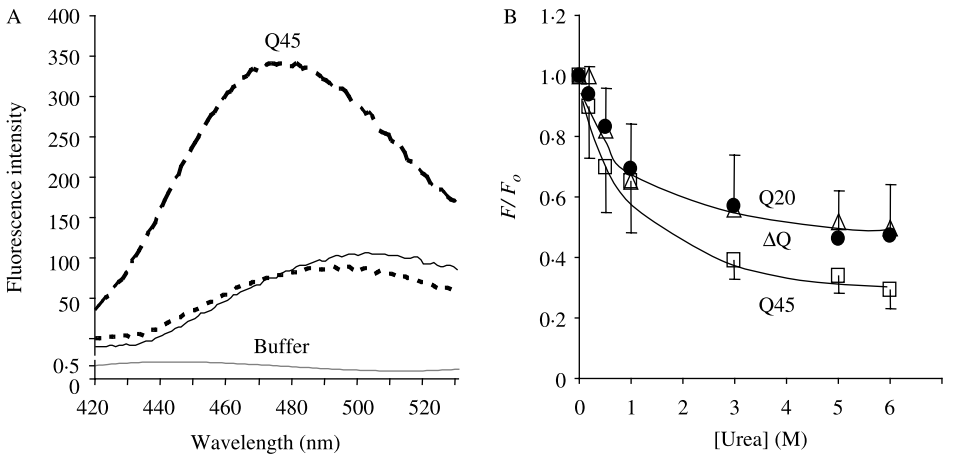


Figure 3 Urea-induced unfolding of ANS–protein complexes. (A) AR-NTD polypeptides at a concentration of 50 μg/ml (1 μM) were incubated with 50 μM ANS for 30 min in the dark. After excitation at 370 nm, the fluorescence emission spectrum in the range 420–530 nm was measured. The experiment was repeated thrice and a representative spectrum is shown for each protein. (B) Unfolding of the AR-NTD polypeptides was monitored by measuring the fluorescence emission of ANS at 480 nm in the absence (F_o) and presence of urea (F) after excitation at 370 nm. The ratio F/F_o was then plotted against urea concentration: AR-NTDDQ (open triangles), AR-NTDQ20 (filled circles) and AR-NTDQ45 (open squares). The means ± S.D. shown represent the results of up to four independent experiments.

the absence of the poly-Q repeat. All three polypeptides formed a more protease-resistant conformation in the presence of TMAO and the tryptophan residues became less solvent exposed. However, in the absence of the poly-Q repeat, the local structure surrounding the tryptophan residues was again less affected by the addition of the structure-stabilizing TMAO or of the unfolding agent urea. Collectively, these data support the argument that altering the poly-Q repeat length impacts upon the folding of the AR-NTD.

The hydrophobic probe ANS has been used successfully to investigate structural features of proteins. In the present study, we show that binding to AR-NTD polypeptides markedly enhances ANS fluorescence. This, together with the non-cooperative unfolding with increasing urea concentration, indicates that each polypeptide has at least some 'molten globule' character. The results were most dramatic for AR-NTD with an expanded poly-Q repeat, which overall appeared less stable and more susceptible to the urea denaturation, suggesting a less stable tertiary structure. Taken together, the biophysical and biochemical data suggest that the presence and possibly the length of the poly-Q repeat alter both the secondary structure and the folding properties of the AR-NTD.

What are the functional consequences of the above conformational changes? In our hands, expansion of the poly-Q repeat did not significantly alter either the binding of the p160 co-activator, NCOA1, or the ligand–dependent interaction of the AR-NTD with the ligand-binding domain. However, the deletion of the poly-Q stretch impaired both sets of interactions (Davies and McEwan unpublished observations). These results suggest that the alteration of secondary structure content and/or the reduced flexibility of the AR-NTDΔQ polypeptide has an impact on the ability of the protein to make specific protein–protein interactions. Interestingly, Buchanan *et al.* (2004) identified a double-point mutation in a prostate tumour that interrupted the poly-Q repeat (Q12LQ6LQ) which showed increased transcriptional activity but significant disruption of the N/C-terminal AR interaction. Furthermore, they observed impaired N/C-terminal receptor intersections with both short and long glutamine stretches and no major changes in p160 binding.

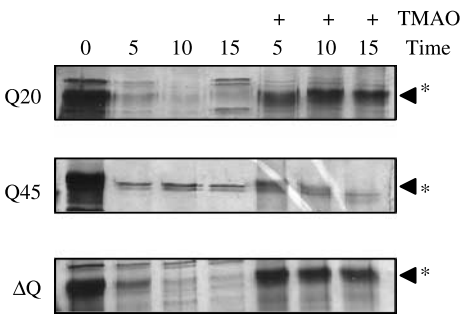


Figure 4 Folding of the AR-NTD polypeptides in the presence of TMAO. AR-NTD polypeptides (25 pmol) were digested with 4 ng/μl endo Glu-C for 0–15 min in the absence or presence of 2 M of the natural osmolyte TMAO (+). Proteins were resolved by SDS-PAGE and detected by silver staining. The full-length protein is indicated by the filled arrow head. *A co-purifying bacterial protein.

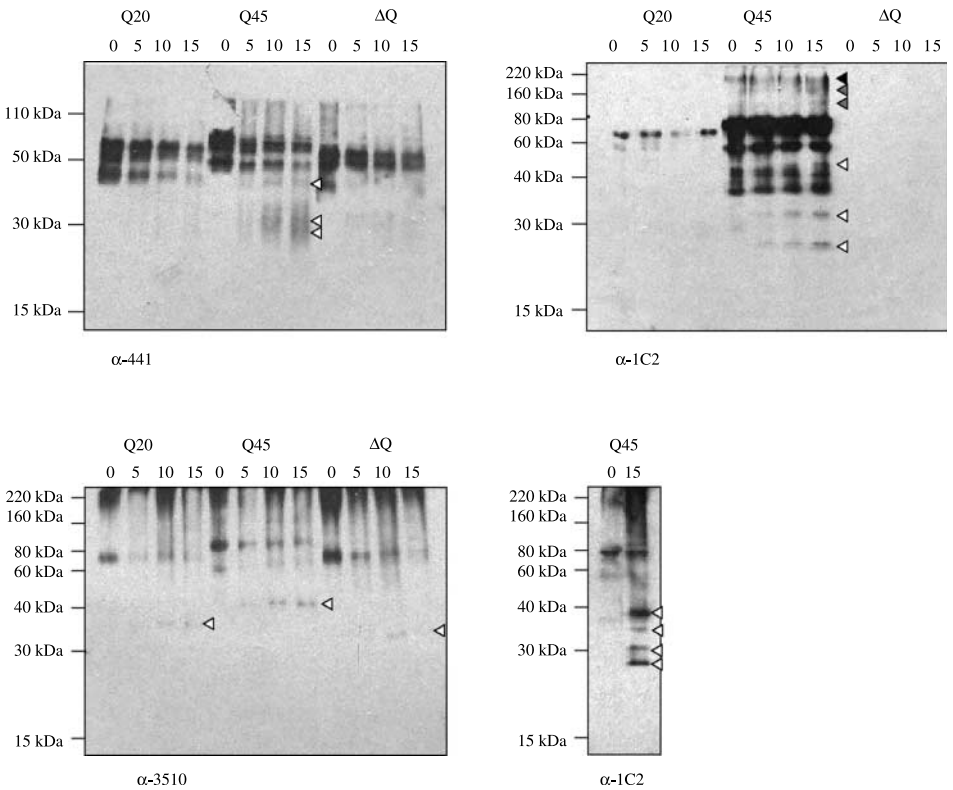


Figure 5 Expansion of the poly-Q repeat results in a conformational change. (A) The products of a partial endo Glu-C digestion of AR-NTD polypeptides with 0, 20 or 45 glutamines were resolved by SDS-PAGE and the fragments detected using antibodies against the NTD (N441) or the poly-Q repeat (1C2). Fragments ranging in size from 21 to 57 kDa for the ARN-Q45 polypeptide are indicated (open arrow head). Larger species, possibly representing dimers or trimers of ARN-Q45, detected with antibody 1C2 are also indicated (filled arrow heads). Note the full-length ARN-Q20 was also detected by antibody 1C2 on this blot, but was not usually observed; no signal is seen with ARN-ΔQ and this antibody. The results shown are representative of at least three to four independent blots. (B) The products of a partial trypsin digestion of AR-NTD polypeptides with 0, 20 or 45 glutamines were resolved by SDS-PAGE and the fragments detected using antibodies against the first 21 amino acids (ab3510) or the poly-Q repeat (1C2). Fragments ranging in size from 23 to 49 kDa for the ARN-Q45 polypeptide and fragments of 38 and 33 kDa for ARN-20Q and -ΔQ respectively are indicated (open arrow heads).

They speculated that this double mutation resulted in a more stable conformation in the AR-NTD. The stabilization of structure or at least a reduction in local structural plasticity may also result from the removal of the poly-Q repeat, with concomitant impact on protein–protein interactions. Expansion of the poly-Q repeat in the AR is proposed to lead to a receptor protein that is compromised for transcriptional activity (Mhatre *et al.* 1993, Chamberlain *et al.* 1994, Jenster *et al.* 1994, Nakajima *et al.* 1996, Tut *et al.* 1997, Irvine *et al.* 2000, Callewaert *et al.* 2003, Wang *et al.* 2004). Conversely, a reduction in poly-Q length has been associated with a more active AR (Chamberlain *et al.* 1994, Irvine *et al.* 2000, Callewaert *et al.* 2003, Wang *et al.* 2004). However, results from other studies have argued that transcriptional activity is not directly affected by poly-Q length, but that repeat expansion results in reduced receptor protein levels (Choong *et al.* 1996, Neuschmid-Kaspar *et al.* 1996, Brooks *et al.* 1997). Interestingly, the work of Beilin *et al.* (2000) suggests that the effect of poly-Q length may be cell specific.

A common feature of the poly-Q repeat diseases is the observation of intracellular aggregates comprising the poly-Q protein and other cellular proteins. However, the structural nature of the expanded glutamine repeats and the relationship of aggregation to cellular pathology are less well understood. Molecular modelling originally suggested that β-structures (strand and turn) could form the basis of nucleation and aggregation reviewed in Masino & Pastore (2001). Experimental evidence supporting this model suggests that short repeats are random coil, while larger repeats form β-sheets, with a conversion from intra- to intermolecular interactions and aggregate formation

(Masino & Pastore 2001 and references therein). However, the majority of studies seem to find no change in the conformation with expansion of the glutamine stretch (see Bennett *et al.* 2002, Chen 2003, Chow *et al.* 2004, Chellgren *et al.* 2006). The exception is a study from Nagai *et al.* (2007) who observed an increase in α-helix content with increasing poly-Q length of peptides fused to thioredoxin. Interestingly, Guo *et al.* (2007), investigating the structure of the glutamine-rich domain of the enzyme histone deacetylase 4, also revealed an α-helical conformation that readily and reversibly assembled into higher order structures and formed stable tetramers. The authors suggested that the structural properties of the glutamine-rich domain, together with destabilization of the tetramer structure, may lead to amyloid fibre formation. However, Bevivino & Loll (2001) reported that the α-helix content of maltose-binding protein–ataxin-3 fusion, with an expanded glutamine repeat, was reduced compared with wild-type protein. Significantly, Palazzolo *et al.* (2008) have recently reported, for AR-NTD polypeptides fused to GST, that expansion of the poly-Q repeat led to an increase in anti-parallel β-sheet. The formation of such structures could either directly or indirectly through misfolding result in the formation of oligomeric forms of the AR. It is interesting therefore that expanding the large poly-Q repeat from 20 to 45 in the AR-NTD, in the absence of a fusion protein partner, resulted in a small increase in α-helical secondary structure and a significant increase in the propensity to form a helical conformation. We would speculate that the ability of the AR-NTD to form α-helix secondary structure may be important in the pathology of SBMA and the formation of protein aggregates, as well as underlying the normal function of this domain. In studies where a change in conformation has been observed for expanded poly-Q repeats both α-helical (Nagai *et al.* 2007, present study) and β-sheet (Masino & Pastore 2001, Palazzolo *et al.* 2008), secondary structures have been reported. This suggests that different types of secondary structure may be compatible with self-association and oligomer formation. The formation of a given type of secondary structure in turn may depend on a combination of factors: destabilization of protein structure by repeat expansion; the nature of the amino acids flanking or surrounding the repeat; and the cellular environment.

Expansion of the poly-Q repeat also resulted in changes in protease sensitivity. Previously, proteolysis of the AR with expanded glutamine repeats has been associated with cell death. Pinsky and co-workers reported a difference in trypsin digestion for the full-length AR with a repeat of 44Q compared with normal receptor (Abdullah *et al.* 1998). Notably, they observed

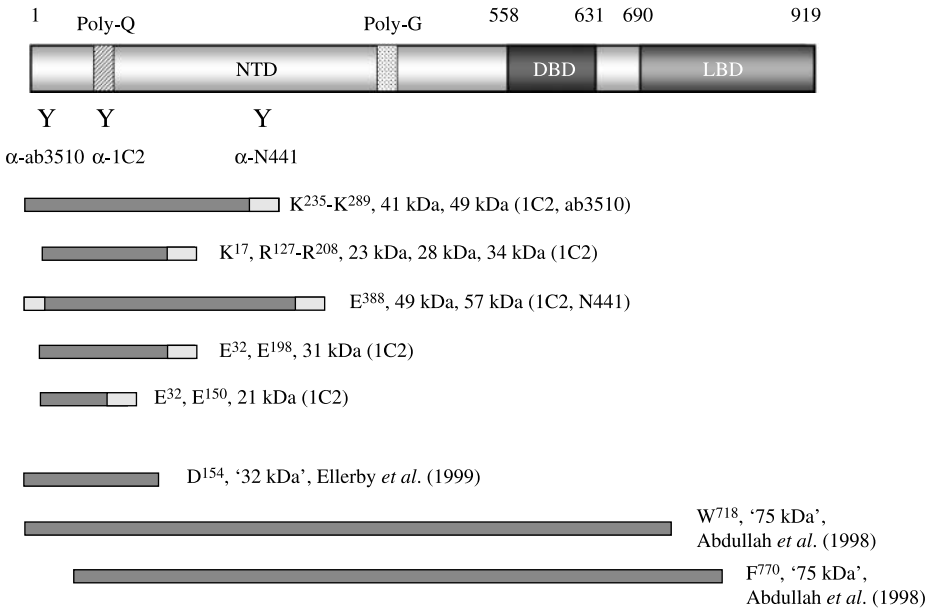


Figure 6 Summary of proteolytic cleavage of the AR and poly-Q repeat expansion. The domain structure of the full-length wild-type human AR is represented and the poly-Q (21 residues) and poly-G (23 residues) repeats indicated. The epitopes recognized by the antibodies used in the present study are indicated below: α-ab3510, amino acids 1–21; α-N441, amino acids 299–315; and α-1C2 against the poly-Q repeat. Fragments observed for proteolytic cleavage of the AR full-length or NTD (present study) and likely cleavage sites, as determined from the molecular weights and the theoretical cleavage of AR-NTD using PeptideCutter (www.expasy.ch), are shown below. Fragment molecular weights are for the receptor proteins with expanded poly-Q repeats of 44 (Abdullah *et al.* 1998), 45 (present study) and 50 (Ellerby *et al.* 1999).

a 75 kDa fragment when this mutant AR was expressed in COS cells, suggesting a conformational change upon expansion of the poly-Q repeat. Ellerby *et al.* (Ellerby *et al.* 1999) observed a 32 kDa fragment for a receptor with a repeat of 50Q and identified a caspase 3 cleavage site at aspartic acid 154. Interestingly, this caspase site is close to the putative endo Glu-C cleavage site(s) identified in the present study and suggesting that poly-Q expansion alters the structure in the region C-terminal of the repeat (Fig. 6).

We have previously characterized the conformation and folding of a region of the AR-NTD, lacking the large poly-Q repeat that contains the main determinants for transcriptional regulation. The AR-AF1 domain is structurally plastic and undergoes induced folding upon specific protein–protein interactions resulting in a protease-resistant conformation and an increase in α-helix secondary structure (Reid *et al.* 2002, Kumar *et al.* 2004*a*) and shows molten-globule-like characteristics (Lavery & McEwan 2008). Significantly, the AF1 domain/NTD of the oestrogen (ER; Warnmark *et al.* 2001), glucocorticoid (GR; Kumar *et al.* 2004*b*) and progesterone (PR; Wardell *et al.* 2005) receptors share similar structural properties. Binding of the TATA-binding protein leads to an increased α-helical conformation for both the ER1α-NTD and GR-AF1, while binding of the co-regulatory protein Jun dimerization protein (JDP) to the PR DNA-binding domain resulted in induced folding the PR-NTD, consistent with an increase in α-helix secondary structure. Taken together, these findings have led to models where structural flexibility is an intrinsic property of the steroid receptor NTD/AF1, and underpins the coupled folding and receptor–protein interactions (Kumar & Thompson 2003, Lavery & McEwan 2005). It is interesting, therefore, that the present study indicates that changes in the secondary structure content and local and global folding of the AR-NTD are modulated by the presence and length of the large poly-Q repeat. Further work will be required to understand fully the relationship between AR-NTD structure, stability, protein–protein interactions and the cellular pathology observed with the expanded glutamine repeat in SBMA patients.

Declaration of interest

The authors confirm that there are no conflicts of interest associated with these studies.

Funding

We would like to express thanks for the generous support of the James Alexander Mairns Trust for funding this project, and the Biotechnology and Biological Sciences Research Council of the UK for support of the CD facility.

Acknowledgements

We are also grateful to Dr A O Brinkmann (Erasmus University, Rotterdam, The Netherlands) for the gift of plasmids.

References

Abdullah A, Trifiro MA, Panet-Raymond V, Alvarado C, de Tourreil S, Frankel D, Schipper HM & Pinsky L 1998 Spinobulbar muscular atrophy: polyglutamine-expanded androgen receptor is proteolytically resistant *in vitro* and processed abnormally in transfected cells. *Human Molecular Genetics* **7** 379–384.

Altschuler EL, Hud NV, Mazrimas JA & Rupp B 1997 Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides. *Journal of Peptide Research* **50** 73–75.

Armen RS, Bernard BM, Day R, Alonso DO & Daggett V 2005 Characterization of a possible amyloidogenic precursor in glutamine-repeat neurodegenerative diseases. *PNAS* **102** 13433–13438.

Bailey RW, Dunker AK, Brown CJ, Garner EC & Griswold MD 2001 Clusterin, a binding protein with a molten globule-like region. *Biochemistry* **40** 11828–11840.

Beilin J, Ball EM, Favaloro JM & Zajac JD 2000 Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. *Journal of Molecular Endocrinology* **25** 85–96.

Bennett MJ, Huey-Tubman KE, Herr AB, West AP Jr, Ross SA & Bjorkman PJ 2002 Inaugural article: a linear lattice model for polyglutamine in CAG-expansion diseases. *PNAS* **99** 11634–11639.

Betney R & McEwan IJ 2003 Role of conserved hydrophobic amino acids in androgen receptor AF-1 function. *Journal of Molecular Endocrinology* **31** 427–439.

Bevivino AE & Loll PJ 2001 An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel beta -fibrils. *PNAS* **98** 11955–11960.

Bhattacharyya A, Thakur AK, Chellgren VM, Thiagarajan G, Williams AD, Chellgren BW, Creamer TP & Wetzel R 2006 Oligoproline effects on polyglutamine conformation and aggregation. *Journal of Molecular Biology* **355** 524–535.

Bradford MM 1976 A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72** 248–254.

Brinkmann AO, Faber PW, van Rooij HC, Kuiper GG, Ris C, Klaassen P, van der Korput JA, Voorhorst MM, van Laar JH & Mulder E 1989 The human androgen receptor: domain structure, genomic organization and regulation of expression. *Journal of Steroid Biochemistry* **34** 307–310.

Brooks BP, Paulson HL, Merry DE, Salazar-Grueso EF, Brinkmann AO, Wilson EM & Fischbeck KH 1997 Characterization of an expanded glutamine repeat androgen receptor in a neuronal cell culture system. *Neurobiology of Disease* **3** 313–323.

Buchanan G, Yang M, Cheong A, Harris JM, Irvine RA, Lambert PF, Moore NL, Raynor M, Neufing PJ, Coetzee GA *et al.* 2004 Structural and functional consequences of glutamine tract variation in the androgen receptor. *Human Molecular Genetics* **13** 1677–1692.

Callewaert L, Christiaens V, Haelens A, Verrijdt G, Verhoeven G & Claessens F 2003 Implications of a polyglutamine tract in the function of the human androgen receptor. *Biochemical and Biophysical Research Communications* **306** 46–52.

Cardamone M & Puri NK 1992 Spectrofluorimetric assessment of the surface hydrophobicity of proteins. *Biochemical Journal* **282** 589–593.

Chai Y, Wu L, Griffin JD & Paulson HL 2001 The role of protein composition in specifying nuclear inclusion formation in polyglutamine disease. *Journal of Biological Chemistry* **276** 44889–44897.

Chamberlain NL, Driver ED & Miesfeld RL 1994 The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Research* **22** 3181–3186.

Chamberlain NL, Whitacre DC & Miesfeld RL 1996 Delineation of two distinct type 1 activation functions in the androgen receptor amino-terminal domain. *Journal of Biological Chemistry* **271** 26772–26778.

Chellgren BW, Miller AF & Creamer TP 2006 Evidence for polyproline II helical structure in short polyglutamine tracts. *Journal of Molecular Biology* **361** 362–371.

Chen YW 2003 Local protein unfolding and pathogenesis of polyglutamine-expansion diseases. *Proteins* **51** 68–73.

Chen S, Berthelier V, Yang W & Wetzel R 2001 Polyglutamine aggregation behavior *in vitro* supports a recruitment mechanism of cytotoxicity. *Journal of Molecular Biology* **311** 173–182.

Choong CS & Wilson EM 1998 Trinucleotide repeats in the human androgen receptor: a molecular basis for disease. *Journal of Molecular Endocrinology* **21** 235–257.

Choong CS, Kemppainen JA, Zhou ZX & Wilson EM 1996 Reduced androgen receptor gene expression with first exon CAG repeat expansion. *Molecular Endocrinology* **10** 1527–1535.

Choong CS, Kemppainen JA & Wilson EM 1998 Evolution of the primate androgen receptor: a structural basis for disease. *Journal of Molecular Evolution* **47** 334–342.

Chow MK, Ellisdon AM, Cabrita LD & Bottomley SP 2004 Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *Journal of Biological Chemistry* **279** 47643–47651.

Dehay B & Bertolotti A 2006 Critical role of the proline-rich region in huntingtin for aggregation and cytotoxicity in yeast. *Journal of Biological Chemistry* **281** 35608–35615.

Djian P, Hancock JM & Chana HS 1996 Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. *PNAS* **93** 417–421.

Duennwald ML, Jagadish S, Muchowski PJ & Lindquist S 2006 Flanking sequences profoundly alter polyglutamine toxicity in yeast. *PNAS* **103** 11045–11050.

Eftink MR & Ghiron CA 1976 Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. *Biochemistry* **15** 672–680.

Ellerby LM, Hackam AS, Propp SS, Ellerby HM, Rabizadeh S, Cashman NR, Trifiro MA, Pinsky L, Wellington CL, Salvesen GS *et al.* 1999 Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *Journal of Neurochemistry* **72** 185–195.

Finke JM, Cheung MS & Onuchic JN 2004 A structural model of polyglutamine determined from a host–guest method combining experiments and landscape theory. *Biophysical Journal* **87** 1900–1918.

Gatchel JR & Zoghbi HY 2005 Diseases of unstable repeat expansion: mechanisms and common principles. *Nature Reviews. Genetics* **6** 743–755.

Guo L, Han A, Bates DL, Cao J & Chen L 2007 Crystal structure of a conserved N-terminal domain of histone deacetylase 4 reveals functional insights into glutamine-rich domains. *PNAS* **104** 4297–4302.

Irvine RA, Ma H, Yu MC, Ross RK, Stallcup MR & Coetzee GA 2000 Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. *Human Molecular Genetics* **9** 267–274.

Jenster G, de Ruiter PE, van der Korput HA, Kuiper GG, Trapman J & Brinkmann AO 1994 Changes in the abundance of androgen receptor isotypes: effects of ligand treatment, glutamine-stretch variation, and mutation of putative phosphorylation sites. *Biochemistry* **33** 14064–14072.

Jenster G, van der Korput HA, Trapman J & Brinkmann AO 1995 Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *Journal of Biological Chemistry* **270** 7341–7346.

Kumar R & Thompson EB 2003 Transactivation functions of the N-terminal domains of nuclear hormone receptors: protein folding and coactivator interactions. *Molecular Endocrinology* **17** 1–10.

Kumar R, Betney R, Li J, Thompson EB & McEwan IJ 2004*a* Induced alpha-helix structure in AF1 of the androgen receptor upon binding transcription factor TFIIF. *Biochemistry* **43** 3008–3013.

Kumar R, Volk DE, Li J, Lee JC, Gorenstein DG & Thompson EB 2004*b* TATA box binding protein induces structure in the recombinant glucocorticoid receptor AF1 domain. *PNAS* **101** 16425–16430.

Kumar R, Serrette JM & Thompson EB 2005 Osmolyte-induced folding enhances tryptic enzyme activity. *Archives of Biochemistry and Biophysics* **436** 78–82.

Lathrop RH, Casale M, Tobias DJ, Marsh JL & Thompson LM 1998 Modeling protein homopolymeric repeats: possible polyglutamine structural motifs for huntington's disease. *Proceedings/International Conference on Intelligent Systems for Molecular Biology* **6** 105–114.

Lavery DN & McEwan IJ 2005 Structure and function of steroid receptor AF1 transactivation domains: induction of active conformations. *Biochemical Journal* **391** 449–464.

Lavery DN & McEwan IJ 2008 Structural characterization of the native N-terminal transactivation domain of the human androgen receptor: a collapsed disordered conformation underlies structural plasticity and protein-induced folding. *Biochemistry* **47** 3360–3369.

Masino L & Pastore A 2001 A structural approach to trinucleotide expansion diseases. *Brain Research Bulletin* **56** 183–189.

Masino L, Kelly G, Leonard K, Trottier Y & Pastore A 2002 Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Letters* **513** 267–272.

Masino L, Musi V, Menon RP, Fusi P, Kelly G, Frenkiel TA, Trottier Y & Pastore A 2003 Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Letters* **549** 21–25.

McEwan IJ 2004 Molecular mechanisms of androgen receptor-mediated gene regulation: structure–function analysis of the AF-1 domain. *Endocrine-Related Cancer* **11** 281–293.

McEwan IJ & Gustafsson J 1997 Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIF. *PNAS* **94** 8485–8490.

McEwan IJ, Lavery DN, Fischer K & Watt K 2007 Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nuclear Receptor Signaling* **5** 1–6.

Mhatre AN, Trifiro MA, Kaufman M, Kazemi-Esfarjani P, Figlewicz D, Rouleau G & Pinsky L 1993 Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genetics* **5** 184–188.

Monoi H 1995 New tubular single-stranded helix of poly-L-amino acids suggested by molecular mechanics calculations: I. homopolypeptides in isolated environments. *Biophysical Journal* **69** 1130–1141.

Nagai Y, Inui T, Popiel HA, Fujikake N, Hasegawa K, Urade Y, Goto Y, Naiki H & Toda T 2007 A toxic monomeric conformer of the polyglutamine protein. *Nature Structural and Molecular Biology* **14** 332–340.

Nakajima H, Kimura F, Nakagawa T, Furutama D, Shinoda K, Shimizu A & Ohsawa N 1996 Transcriptional activation by the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Journal of the Neurological Sciences* **142** 12–16.

Neuschmid-Kaspar F, Gast A, Peterziel H, Schneikert J, Muigg A, Ransmayr G, Klocker H, Bartsch G & Cato AC 1996 CAG-repeat expansion in androgen receptor in kennedy's disease is not a loss of function mutation. *Molecular and Cellular Endocrinology* **117** 149–156.

Palazzolo I, Gliozzi A, Rusmini P, Sau D, Crippa V, Simonini F, Onesta E, Bolzoni E & Poletti A 2008 The role of the polyglutamine tract in androgen receptor. *Journal of Steroid Biochemistry and Molecular Biology* **108** 245–253.

Paulson HL, Bonini NM & Roth KA 2000 Polyglutamine disease and neuronal cell death. *PNAS* **97** 12957–12958.

Perutz MF, Johnson T, Suzuki M & Finch JT 1994 Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *PNAS* **91** 5355–5358.

Poletti A 2004 The polyglutamine tract of androgen receptor: from functions to dysfunctions in motor neurons. *Frontiers in Neuroendocrinology* **25** 1–26.

Popiel HA, Nagai Y, Onodera O, Inui T, Fujikake N, Urade Y, Strittmatter WJ, Burke JR, Ichikawa A & Toda T 2004 Disruption of the toxic conformation of the expanded polyglutamine stretch leads to suppression of aggregate formation and cytotoxicity. *Biochemical and Biophysical Research Communications* **317** 1200–1206.

Quigley CA, De Bellis A, Marschke KB, el-Awady MK, Wilson EM & French FS 1995 Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocrine Reviews* **16** 271–321.

Quist A, Doudevski I, Lin H, Azimova R, Ng D, Frangione B, Kagan B, Ghiso J & Lal R 2005 Amyloid ion channels: a common structural link for protein-misfolding disease. *PNAS* **102** 10427–10432.

Reid J, Kelly SM, Watt K, Price NC & McEwan IJ 2002 Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. influence of structure-stabilizing solutes and protein–protein interactions. *Journal of Biological Chemistry* **277** 20079–20086.

Ross CA 1995 When more is less: pathogenesis of glutamine repeat neurodegenerative diseases. *Neuron* **15** 493–496.

Semisotnov GV, Rodionova NA, Razgulyaev OI, Uversky VN, Gripas' AF & Gilmanshin RI 1991 Study of the 'molten globule' intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31** 119–128.

Sharma KK, Kaur H, Kumar GS & Kester K 1998 Interaction of 1,1′-bi(4-anilino)naphthalene-5,5′-disulfonic acid with alpha-crystallin. *Journal of Biological Chemistry* **273** 8965–8970.

Sharma D, Sharma S, Pasha S & Brahmachari SK 1999 Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions. *FEBS Letters* **456** 181–185.

Simental JA, Sar M, Lane MV, French FS & Wilson EM 1991 Transcriptional activation and nuclear targeting signals of the human androgen receptor. *Journal of Biological Chemistry* **266** 510–518.

La Spada AR & Taylor JP 2003 Polyglutamines placed into context. *Neuron* **38** 681–684.

Sreerama N & Woody RW 1993 A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Analytical Biochemistry* **209** 32–44.

Starikov EB, Lehrach H & Wanker EE 1999 Folding of oligoglutamines: a theoretical approach based upon thermodynamics and molecular mechanics. *Journal of Biomolecular Structure and Dynamics* **17** 409–427.

Tanaka M, Morishima I, Akagi T, Hashikawa T & Nukina N 2001 Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *Journal of Biological Chemistry* **276** 45470–45475.

Tanaka M, Machida Y, Nishikawa Y, Akagi T, Morishima I, Hashikawa T, Fujisawa T & Nukina N 2002 The effects of aggregation-inducing motifs on amyloid formation of model proteins related to neurodegenerative diseases. *Biochemistry* **41** 10277–10286.

Tut TG, Ghadessy FJ, Trifiro MA, Pinsky L & Yong EL 1997 Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. *Journal of Clinical Endocrinology and Metabolism* **82** 3777–3782.

Wang Q, Udayakumar TS, Vasaitis TS, Brodie AM & Fondell JD 2004 Mechanistic relationship between androgen receptor polyglutamine tract truncation and androgen-dependent transcriptional hyperactivity in prostate cancer cells. *Journal of Biological Chemistry* **279** 17319–17328.

Wardell SE, Kwok SC, Sherman L, Hodges RS & Edwards DP 2005 Regulation of the amino-terminal transcription activation domain of progesterone receptor by a cofactor-induced protein folding mechanism. *Molecular and Cellular Biology* **25** 8792–8808.

Warnmark A, Wikstrom A, Wright AP, Gustafsson JA & Hard T 2001 The N-terminal regions of estrogen receptor alpha and beta are unstructured *in vitro* and show different TBP binding properties. *Journal of Biological Chemistry* **276** 45939–45944.

Yang ZR, Thomson R, McNeil P & Esnouf RM 2005 RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21** 3369–3376.

Received in final form 14 August 2008
Accepted 1 September 2008
Made available online as an Accepted Preprint 1 September 2008