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

Philippa Davies*, Kate Watt*, Sharon M Kelly1, Caroline Clark2, Nicholas C Price1 and Iain J McEwan

School of Medical Sciences, Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK
1Institute of Biomedical and Life Sciences, University of Glasgow, Joseph Black Building, Glasgow G12 8QQ, Scotland, UK
2School 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. 2004a). 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,
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 thioredoxin 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-helical secondary structure (Nagai et al. 2007). Interestingly, the inclusion of poly-proline sequences, C-terminus 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’ CCGCGCAGATCTATGGAAGTGCA-AGATCTATGGAAGTGCA-3’ and ARC 5’ GCCGGCGGATCCATGTCCCCCGAAGG 3’ and ARC 5’ GCCGGCGGATCCATGTCCCCCGAAGG 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

(GT TAGGGCTGGGAAGGG 3’).
above) and cloned into the expression plasmids pET-19b 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 attenuation at 600 nm of 0.6–0.9 was reached. Expression was induced with 1 mM isopropyl β-D-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 Ni²⁺-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/ml trypsin or 4 ng/ml endo Gln-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, Chancellors 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 (F₀; 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 NaH₂PO₄, 6 mM Na₂HPO₄, 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 (1S)-(+) -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 (Sreevama & Woody
1993). In each case, the quality of the fitting procedure was judged by the very low value (≤0.01) 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 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.

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. 2004a,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 λmax 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 λmax 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
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 \( \lambda_{\text{max}} \) 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 (\( \Delta Q \)) 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-NTD\( \Delta Q \) 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 \( \alpha \)-helix content (to 16%) and a small decrease in the \( \beta \)-structures (to 23% strand and 22% turn). By contrast, in the case of the AR-NTDQ polypeptide, the \( \alpha \)-helix content was reduced (to 11%) with small increases in \( \beta \)-structures (to 28% strand and 25% turns). In the presence of TFE, there was a further increase in the \( \alpha \)-helix content of AR-NTDQ45 compared with the wild-type polypeptide and again this was accompanied by decreases in both non-ordered and \( \beta \)-structure (Fig. 2D, Table 1). A similar increase in \( \alpha \)-helix was observed in the presence of TMAO (data not shown). Although there was an increase in \( \alpha \)-helix content for the AR-NTDQ 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 \( \alpha \)-helical structure in aqueous solution, but also a lower propensity to form such structure in the presence of TFE.

The values of \( \lambda_{\text{max}} \) 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 \( \lambda_{\text{max}} \) was 342 ± 0.5 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 \( \lambda_{\text{max}} \) for AR-NTDQ 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 \( \alpha \)-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-NTD\( \Delta Q \) 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 \( \mu \)g), Q45 (6 \( \mu \)g) and \( \Delta Q \) (13 \( \mu \)g). *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 \( \mu \)g/ml) in buffer (black curve) or 6 M urea (grey curve) after excitation at 278 nm. The locations of the tryptophan and tyrosine emissions are indicated. (D–G) As for (B) and (C), but analysis of AR-NTDQ45 (D and E) and AR-NTD\( \Delta Q \) (F and G).

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

www.endocrinology-journals.org

Downloaded from Bioscientifica.com at 01/27/2019 12:53:44PM via free access
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 aC2 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 aC2.

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 225 and lysine 289. Analysis of the products of the partial trypsin digest for ARN-Q45 using aC2 revealed an additional series of fragments containing the expanded poly-Q repeat: 34, 28 and 23 kDa. The absence of these fragments with a-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 1 Summary of secondary structure content

<table>
<thead>
<tr>
<th>Protein</th>
<th>α-Helix (%)</th>
<th>β-Strand (%)</th>
<th>Turn (%)</th>
<th>Non-ordered (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARN-Q20/Buffer</td>
<td>14</td>
<td>27</td>
<td>24</td>
<td>34</td>
</tr>
<tr>
<td>ARN-Q45/Buffer</td>
<td>16</td>
<td>23</td>
<td>22</td>
<td>39</td>
</tr>
<tr>
<td>ARN-ΔQ/Buffer</td>
<td>11</td>
<td>28</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>ARN-Q20/TFE</td>
<td>58</td>
<td>12</td>
<td>9</td>
<td>22</td>
</tr>
<tr>
<td>ARN-Q45/TFE</td>
<td>66</td>
<td>9</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>ARN-ΔQ/TFE</td>
<td>30</td>
<td>21</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>

*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.

### Table 2 Summary of steady-state fluorescence emission spectra

<table>
<thead>
<tr>
<th>Protein</th>
<th>λ_{max} (nm)</th>
<th>Tryptophan emission*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Buffer (nm)</td>
<td>Urea (nm)</td>
</tr>
<tr>
<td>ARN-Q20</td>
<td>342 ± 2</td>
<td>350</td>
</tr>
<tr>
<td>ARN-Q45</td>
<td>342 ± 0.5</td>
<td>351</td>
</tr>
<tr>
<td>ARN-ΔQ</td>
<td>343 ± 3</td>
<td>347</td>
</tr>
</tbody>
</table>

*λ_{max} 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-helical 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-helical secondary structure was less pronounced in...
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.
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 modeling 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 \( \alpha \)-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 \( \alpha \)-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 \( \alpha \)-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 \( \beta \)-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 \( \alpha \)-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 \( \alpha \)-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 \( \alpha \)-helical (Nagai et al. 2007, present study) and \( \beta \)-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](https://www.endocrinology-journals.org/)

**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: \( \alpha \)-ab3510, amino acids 1–21; \( \alpha \)-N441, amino acids 299–315; and \( \alpha \)-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-helical secondary structure (Reid et al. 2002, Kumar et al. 2004a) and shows molten-globule-like characteristics (Lavery & McEwan 2008). Significantly, the AF1 domain/NTD of the oestrogen (ER; Warmark et al. 2001), glucocorticoid (GR; Kumar et al. 2004b) 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 ERα-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


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.


