Trinucleotide repeats in the human androgen receptor: a molecular basis for disease

C S Choong1 and E M Wilson2
1Department of Pediatric Endocrinology, Mater Misericordiae Children’s Hospital, South Brisbane, Queensland 4101, Australia
2The Laboratories for Reproductive Biology and the Departments of Pediatrics and Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599–7500, USA

(REquests for offprints should be addressed to EM Wilson)

THE ANDROGEN RECEPTOR

Sequence similarities among the steroid receptor family of nuclear receptors

The androgen receptor (AR) is a member of the superfamily of nuclear hormone receptors (Zhou et al. 1994). This family of ligand-dependent transcription factors are key regulatory proteins in diverse physiological processes, including embryogenesis, development and homeostasis. Steroids are lipophilic hormones derived from hydroxylation and side chain cleavage of cholesterol. Current theory suggests that steroids diffuse through cell membranes to interact with their cognate receptors, although recent data suggests a transport role of the serum binding proteins (Rosner 1991). Upon binding hormone, steroid receptors undergo allosteric changes which enable them to bind high affinity sites in chromatin and modulate gene expression (Yamamoto 1985) (Fig. 1). Cloning of the cDNAs of the glucocorticoid (Hollenberg et al. 1985), oestrogen (Green et al. 1986), mineralocorticoid (Arriza et al. 1987) and progesterone (Misrahi et al. 1987) receptors revealed structural homology in three functional domains; a highly conserved DNA binding domain, a steroid binding domain, and an amino- or NH2-terminal domain that is important for transcriptional activity (Evans 1988, Mangelsdorf et al. 1995).

Functional domains of AR

The human AR gene localizes to the X chromosome at Xq11–12 (Brown et al. 1989) and is encoded in 8 exons (Lubahn et al. 1988b). The calculated molecular weight of AR is 98 999 Daltons (Da). The major AR mRNA species is about 10 kilobases (kb) with a less abundant mRNA of 7 kb detected in some human tissues (Lubahn et al. 1988a). Differences in the published AR cDNA sequence relate primarily to two polymorphic trinucleotide repeats encoding polyglutamine and polyglycine within the first exon (Chang et al. 1988a, Lubahn et al. 1988b, Tilley et al. 1989). Functional organization of AR is similar to that of the other members of the steroid receptor family (Lubahn et al. 1988a, Kuiper et al. 1989, Rundlett et al. 1990, Jenster et al. 1991, Zhou et al. 1994). The entire amino-terminal domain is encoded by the first exon, the two zinc fingers of the DNA binding domain by exons 2 and 3, the hinge region which contains the nuclear targeting signal by part of exon 4, and the steroid binding domain by exons 4–8 (Fig. 2). Amino acid sequence homology with other members of the steroid receptor family is greatest in the DNA binding domain.

Role of the androgen receptor in virilization

In the 46 XY human embryo, differentiation of the testes occurs at approximately six weeks in the presence of testis determining factor which is encoded by a gene residing in the sex determining region of the short arm of the Y chromosome (Sinclair et al. 1990). The reproductive tissues undergo virilization in the presence of testosterone and 5α-dihydrotestosterone in an ordered sequence. Production of Müllerian inhibitory factor by the testis induces regression of the Müllerian ducts (Jost et al. 1973). During puberty further androgenization occurs with the rise in serum testosterone (Mooradian et al. 1987). The critical
The role of AR in male sexual differentiation is now clearly established by the study of naturally occurring mutations in the AR gene of patients with androgen insensitivity syndrome and has been the subject of a number of recent reviews (McPhaul et al. 1993, Sultan et al. 1993, Patterson et al. 1994, Brown 1995, Quigley et al. 1995, Gottlieb et al. 1997).

There are now over 200 reported mutations of the human AR gene identified in patients with androgen insensitivity syndrome and somatic mutations in prostatic cancer tissue (Gottlieb et al. 1997). With few exceptions the mutations in the human AR gene associated with androgen insensitivity are limited to the DNA and steroid binding domains and splice site mutations. The distribution is uneven through these domains and there is an increase in density of mutations in exons 5 and 7 (McPhaul et al. 1993, Quigley et al. 1995). Notably these two exons represent the most highly conserved regions of AR when compared with other members of the steroid receptor subfamily and may contain critical structural elements of the ligand binding pocket. The functional characterization of AR mutations have defined specific regions of the receptor involved in ligand activation and binding and receptor dimerization. Important areas for further investigation include the effect of steroid binding domain mutations on the recently described receptor coactivator association (Yeh & Chang 1996).

Within the amino-terminal domain comparatively few mutations associated with androgen insensitivity have been reported. Examples include premature stop codons (Zoppi et al. 1993), frame shifts (Batch et al. 1992, Hiort et al. 1994), missense mutations (Gottlieb et al. 1997) and expansions of
the homopolymeric CAG trinucleotide repeat (La Spada et al. 1991). This was initially thought to be due to the lower sensitivity of screening techniques such as denaturing gradient gel electrophoresis in the presence of the high GC content of the amino-terminus. However our recent findings of greater sequence divergence within the amino-terminal domain between primates compared with the DNA and steroid binding domains suggests that the amino-terminal domain may be more functionally tolerant to amino acid substitution than either the DNA or steroid binding domains. Thus naturally occurring mutations within the amino-terminus of AR may be less likely to present with an androgen insensitive phenotype.

Evolution of the primate androgen receptor trinucleotide repeats

We undertook sequence analysis of the AR DNA in several primate species to investigate the evolution and functional consequences of polymorphic variation within the AR. AR nucleotide and amino acid sequence were compared among the primates, including the most primitive prosimian lemur. The amino-terminal CAG repeat whose expansion is associated with spinal and bulbar muscular atrophy (SBMA) (see below) decreases exponentially among the primate species with increasing evolutionary distance from human (Table 1). A less polymorphic GGC repeat together with the preceding (GGT)3GGG(GGT)2 sequence code for 24 consecutive glycine residues in human AR. Like the CAG repeat, the GGC repeat shortens inversely and exponentially among primates with increasing evolutionary distance from human, undergoing concurrent GGC and GGG transition to GGT as the GGC repeat expands (Choong et al. 1998).

However, in contrast to the CAG repeat whose expansion influenced AR mRNA levels (Choong et al. 1996a), the shorter GGC repeats in lower primates did not appear to alter AR mRNA expression. The amino-terminal 1–53 and 360–429 amino acid residues in human AR are completely conserved among the primate species studied. As these regions have not been found critical for AR transactivation (Zhou et al. 1993), their high conservation may reflect the complete sequence conservation of the steroid binding domain among the primates, as we demonstrated previously that these NH2-terminal regions interact with the androgen-bound AR steroid binding domain during AR dimerization (Langley et al. 1995). Overall, however, exon A coding for the NH2-terminal region and exon D coding for the hinge region show the greatest divergence of amino acid sequence whereas exons G and H in the steroid binding domain are the most highly conserved (Table 2).

### Table 1. Triplet length of the major trinucleotide repeats

<table>
<thead>
<tr>
<th>Repeat Strain</th>
<th>Repeat amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTG (TTG) Leu</td>
<td>CAG-Ia,b Gln</td>
</tr>
<tr>
<td>Human AR</td>
<td>54–57</td>
</tr>
<tr>
<td>Human</td>
<td>4</td>
</tr>
<tr>
<td>Chimp</td>
<td>3</td>
</tr>
<tr>
<td>Baboon</td>
<td>1</td>
</tr>
<tr>
<td>Macaque</td>
<td>1</td>
</tr>
<tr>
<td>Lemur</td>
<td>1</td>
</tr>
<tr>
<td>Rat</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CAG-II Gln</td>
</tr>
<tr>
<td></td>
<td>58–78</td>
</tr>
<tr>
<td></td>
<td>18 n=1</td>
</tr>
<tr>
<td></td>
<td>20 n=1</td>
</tr>
<tr>
<td></td>
<td>21 n=2</td>
</tr>
<tr>
<td></td>
<td>23 n=1</td>
</tr>
<tr>
<td></td>
<td>8 n=5</td>
</tr>
<tr>
<td></td>
<td>10 n=2</td>
</tr>
<tr>
<td></td>
<td>7 n=5</td>
</tr>
<tr>
<td></td>
<td>9 n=1</td>
</tr>
<tr>
<td></td>
<td>7 n=4</td>
</tr>
<tr>
<td></td>
<td>9 n=1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>CAG-III Gln</td>
</tr>
<tr>
<td></td>
<td>193–197</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>GGC/CCC/CCT Pro</td>
</tr>
<tr>
<td></td>
<td>372–379</td>
</tr>
<tr>
<td></td>
<td>24 (18)</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>17 (11) n=1</td>
</tr>
<tr>
<td></td>
<td>19 (13) n=1</td>
</tr>
<tr>
<td></td>
<td>13 (6)</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>13 (6)</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>6 (1) n=4</td>
</tr>
<tr>
<td></td>
<td>7 (2) n=1</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>GGT/GGC Gly (GGC)</td>
</tr>
<tr>
<td></td>
<td>449–472/453–472</td>
</tr>
</tbody>
</table>

aAR trinucleotide repeat lengths for CAG-I, CAG-II and GGC and amino acid repeat lengths for CAG-III, Leu, Pro and Gly. Gln repeat length is CAGn+1 at CAG-I for human, chimp, baboon and macaque and at CAG-II for macaque because CAA coding for Gln follows CAG-I and precedes CAG-II in these species.
bCAG-I is polymorphic in the human population (range 11–31, mean 22 ± 2) (Edwards et al. 1992). cThe (GGN)n is polymorphic (range 8–18) (Lubahn et al. 1988b; Chang et al. 1988b; Faber et al. 1989; Tilley et al. 1989; Irvine et al. 1995; Sleddens et al. 1993). GGC repeat length is indicated in parentheses. dThe number (n) of unrelated animals assayed for repeat length. eHuman AR amino acid residue numbers according to Lubahn et al. (1988b).
<table>
<thead>
<tr>
<th>Exon</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>Nucleotide total</th>
<th>Homology (%)</th>
<th>Amino acid total</th>
<th>Homology (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Chimp</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>99·8</td>
<td>3</td>
</tr>
<tr>
<td>Baboon</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>98·6</td>
<td>8</td>
<td>99·1</td>
</tr>
<tr>
<td>Macaque</td>
<td>27</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>39</td>
<td>98·5</td>
<td>9</td>
<td>99·0</td>
</tr>
<tr>
<td>Lemur</td>
<td>138</td>
<td>5</td>
<td>7</td>
<td>17</td>
<td>13</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>191</td>
<td>93·0</td>
<td>49</td>
<td>94·7</td>
</tr>
<tr>
<td>Rat</td>
<td>208</td>
<td>13</td>
<td>8</td>
<td>39</td>
<td>15</td>
<td>10</td>
<td>5</td>
<td>6</td>
<td>304</td>
<td>89·0</td>
<td>86</td>
<td>90·6</td>
</tr>
</tbody>
</table>

Primate sequences are submitted to Genbank under the following accession numbers: *Pan troglodytes* U94177 (chimp); *Papio hamadryas* U94176 (baboon); *Macaca fascicularis* U94179 (macaque); *Eulemur fulvus collaris* U94178 (lemur).

$a$Sequence changes in exon A exclude CAG and GGC trinucleotide repeat expansions or contractions.
The CAG and GGC repeats expanded predominantly after branching of the old world monkeys. These mutations support the time of diversification among the most primitive primates and comparison of lemur AR DNA sequence with that of rat and higher primates supports their transitional status among evolving primates.

OVERVIEW OF THE TRINUCLEOTIDE REPEATS AND NEURODEGENERATIVE DISEASES

The human genome contains arrays of tandem nucleotide repeats, the length of which may be polymorphic across the population (Craig-Holmes & Shaw 1971, Weber 1990, Karlin & Burge 1996). Trinucleotide repeat sequences are characterized by their ability to undergo dynamic mutation or nucleotide alteration over a compressed time span of only a few generations. Products of dynamic trinucleotide mutations may diverge further from the original DNA on a subsequent meiosis, with the probability of mutation being directly related to the number of repeating units (Sutherland & Richards 1995). Proposed mechanisms contributing to nucleotide repeat instability include trans-acting factors such as DNA enzymes involved in replication and repair (Strand et al. 1993, Jaworski et al. 1995). Linkage analyses of flanking markers suggest that cis acting DNA elements within the vicinity of tandem repeats contribute to instability – the founder effect (Kunst & Warren 1994, Richards et al. 1996).

In order to investigate a founder effect, Rubinsztein et al. (1995c) examined a Δ2642 deletion of four consecutive codons within the Huntington’s disease gene in a range of human and non-human primate populations. The Δ2642 deletion was associated with human chromosomes containing 20 or more CAG repeats, suggesting that expanded repeat sequences increase and result in the overrepresentation of the Δ2642 deletion on Huntington’s disease chromosomes (Rubinsztein et al. 1995c).

Uninterrupted perfect repeats are themselves inherently unstable (Sutherland & Richards 1993) and probably exert powerful cis acting destabilizing effects on DNA repeat sequences. One mechanistic model that has been proposed to explain nucleotide repeat expansion postulates that pausing and slippage of the elongated strand may occur during polymerization, resulting in deletion or addition of tandem repeats (Richards & Sutherland 1994, Kang et al. 1995a,b). In long repeats two breaks may occur during polymerization, resulting in a fragment consisting solely of tandem repeats with the ability to slide during polymerization thus facilitating the addition of additional copies of the repeat to the original sequence (Richards & Sutherland 1994).

The phenomenon of a dynamic mutation was first reported in association with an hereditary neurological disorder, the fragile X syndrome, which is a form of familial X-linked mental retardation (Kremer et al. 1991). Discovery of AR containing an expanded polymorphic CAG trinucleotide repeat as the candidate gene for X-linked spinal and bulbar muscular atrophy (La Spada et al. 1991) heralded the subsequent identification of trinucleotide expansions in association with a heterogeneous group of neurodegenerative disorders including most notably Huntington’s disease (Anonymous 1993). Thirteen neurodegenerative diseases associated with tandem repeat mutations have now been identified (Tables 3 and 4) and more will probably be discovered. These mutations can be broadly classed into two groups: repeat expansions located within noncoding regions of the gene and CAG trinucleotide repeat expansions within protein coding regions that result in a polyglutamine tract.

CAG TRIPLET REPEATS AND DISEASE

The group of disorders associated with CAG trinucleotide repeat expansions constitute a clinically heterogeneous cluster of neurodegenerative diseases with a phenotypic spectrum that includes progressive muscular atrophy and androgen insensitivity in SBMA (La Spada et al. 1992), dementia and choreoathetosis in Huntington’s disease (Kremer et al. 1994), ataxia, ophthalmparesis and weakness in the autosomal dominant spinocerebellar ataxias (Harding 1993), cerebellar ataxia, myoclonic epilepsy and dementia in Dentato-rubropallidoluysian atrophy (Miwa 1994), and generalized epilepsy without myoclonus and schizophrenia in the Haw River syndrome (Burke et al. 1994). The latter two conditions share the identical causative mutation and the basis for phenotypic differences between the two syndromes remains unresolved (Table 4).

Significant differences exist among the causative genes of CAG-related diseases, including chromosomal location and predicted normal function and the neuroanatomic and subcellular localization of the gene products. With the exception of X-linked SBMA, all share the autosomal dominant mode of inheritance. Features common to all these disorders are anticipation on paternal transmission and translation of the CAG repeat sequence to form the polyglutamine repeat within the protein (Table 4).
<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene</th>
<th>Repeat motif</th>
<th>Transmission</th>
<th>Location*</th>
<th>Chromosome</th>
<th>Normal number of repeats</th>
<th>Expanded number of repeats in disease allele</th>
<th>Age of onset correlates with genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myotonic dystrophy</td>
<td>Myotonin protein kinase</td>
<td>CTG (CAG)</td>
<td>Autosomal dominant</td>
<td>3’ UTR</td>
<td>19q13.3</td>
<td>5–37</td>
<td>50–2000</td>
<td>+</td>
<td>Mahadevan et al. (1993)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Harley et al. (1993)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ashizawa et al. (1994)</td>
</tr>
<tr>
<td>Fragile X</td>
<td>FMR-1 (Fragile X mental retardation-1 gene product)</td>
<td>CCG (CGG)</td>
<td>X-linked</td>
<td>5’ UTR</td>
<td>Xq27.3</td>
<td>5–52</td>
<td>≥200</td>
<td>+</td>
<td>Kremer et al. (1991)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Devys et al. (1993)</td>
</tr>
<tr>
<td>FRAXE</td>
<td>FMR-2</td>
<td>CCG (CGG)</td>
<td>X-linked</td>
<td>Not known</td>
<td>Xq27–28</td>
<td>6–25</td>
<td>≥200</td>
<td>Not known</td>
<td>Hamel et al. (1994)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gecz et al. (1996)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gu et al. (1996)</td>
</tr>
<tr>
<td>FRA11B</td>
<td>CBL-2 protooncogene</td>
<td>CCG (CGG)</td>
<td>Autosomal dominant</td>
<td>Not known</td>
<td>11q</td>
<td>8–11</td>
<td>≥200</td>
<td>Not known</td>
<td>Jones et al. (1995)</td>
</tr>
<tr>
<td>Freidrich’s ataxia</td>
<td>Frataxin</td>
<td>GAA</td>
<td>Autosomal recessive</td>
<td>Intron 1</td>
<td>9q13</td>
<td>8–22</td>
<td>120–1700</td>
<td>+</td>
<td>Campuzano et al. (1996)</td>
</tr>
<tr>
<td>Progressive myoclonus epilepsy type 1</td>
<td>Cystatin B</td>
<td>CCCGCCGCC CGCG CGCG</td>
<td>Autosomal recessive</td>
<td>5’ UTR</td>
<td>21q22.3</td>
<td>2–3</td>
<td>12, 13</td>
<td>Not known</td>
<td>Pennachio et al. (1996)</td>
</tr>
<tr>
<td></td>
<td>Poly (A) binding protein 2</td>
<td>(GCG) polyalanine tract</td>
<td>Autosomal dominant</td>
<td>Polyalanine tract/exon 1</td>
<td>14q11</td>
<td>6</td>
<td>8–13</td>
<td>+</td>
<td>Brais et al. (1998)</td>
</tr>
</tbody>
</table>

*Location of the repeat is indicated if known. UTR: untranslated.
TABLE 4. Neurodegenerative disorders associated with CAG trinucleotide repeat and polyglutamine expansions

<table>
<thead>
<tr>
<th>Disease</th>
<th>Gene</th>
<th>Repeat motif</th>
<th>Transmission</th>
<th>Location</th>
<th>Chromosome</th>
<th>Normal number of repeats</th>
<th>Expanded number of repeats in disease allele</th>
<th>Age of onset correlates with genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spinal and bulbar muscular atrophy</td>
<td>Androgen receptor</td>
<td>CAG</td>
<td>X-linked</td>
<td>Exon 1</td>
<td>Xq11–12</td>
<td>11–31</td>
<td>40–66</td>
<td>+</td>
<td>La Spada et al. (1991)</td>
</tr>
<tr>
<td>Spinocerebellar ataxia type 6</td>
<td>Alpha 1A-voltage-dependent calcium channel</td>
<td>CAG</td>
<td>Autosomal dominant</td>
<td>Not known</td>
<td>Not known</td>
<td>4–16</td>
<td>21–27</td>
<td>Not known</td>
<td>Zhuchenko et al. (1997)</td>
</tr>
</tbody>
</table>
The position of the polyglutamine repeat within the protein coding region varies between different genes. For AR and huntingtin, the repeat expansion is within the amino-terminal region, whereas in atrophin, the abnormal expansion lies close to the carboxyl-terminus. The range in length of the polymorphic glutamine (Gln) tracts in the normal population is between 7 and 36 repeats. The expansion is limited in this group of disorders with the longest reported disease alleles in spinocerebellar atrophy type 1 with 81 triplet repeats. The shortest is spinocerebellar atrophy type 6 with 21 CAG repeats. It is noteworthy that 21 triplet repeats are within the normal range polymorphism for a number of the CAG repeat diseases (Table 4).

X-linked spinal bulbar and muscular atrophy, Kennedy’s syndrome

X-linked SBMA first reported in two families by Kennedy et al. in 1968 is a rare inherited neurodegenerative disease characterized by progressive neuromuscular weakness resulting from loss of motor neurons in the brain stem and spinal cord. Disease onset, initially in the proximal musculature, develops in the third to fifth decades of life and is often preceded by muscular cramps on exertion, tremor of the hands and elevated muscle creatine kinase. Weakness of the facial muscles and tongue is prominent.

Expansion of the CAG polyglutamine repeat within the amino-terminus of AR was subsequently recognized as the causative mutation (La Spada et al. 1991). Consistent with other CAG repeat diseases such as Huntington’s disease (Andrew et al. 1993, MacMillan et al. 1993, Snell et al. 1993), age of onset of symptoms appears to correlate inversely with CAG repeat length (Igarashi et al. 1992, La Spada et al. 1992, MacLean et al. 1995a). However, in SBMA the correlation between the molecular abnormality and either the severity of muscle weakness, or rate of progression of neuromuscular weakness (Shimada et al. 1995) and the CAG repeat length remains controversial. Expansion of AR CAG repeat sequences in X-linked SBMA pedigrees is associated with greater instability in male meioses compared with that of the female leading to a bias towards paternal transmission of mutant alleles with expanded CAG repeats (La Spada et al. 1992, Shimada et al. 1995). A similar finding was observed with the Huntington’s disease allele (Andrew et al. 1993, Snell et al. 1993). However, the use of transgenic mice with yeast artificial chromosomes expressing the AR CAG repeat expansion within its genomic context provided evidence that AR CAG repeat instability increased in maternal transmission and with maternal age and may be influenced by cis-acting elements (La Spada et al. 1998). A study of 95 Japanese patients revealed an average gain in CAG repeat length on paternal transmission of 1-4 repeats (Shimada et al. 1995), significantly less than that observed in other CAG repeat diseases where a range of 9 is seen in Huntington’s disease (Andrew et al. 1993, Duyao et al. 1993), 3-3 in spinocerebellar atrophy type 1 (Chung et al. 1993) and 4-2 in Dentato-rubropallidoluysian atrophy (Koide et al. 1994). This may account for the lesser degree of anticipation seen in SBMA compared with Huntington’s disease pedigrees. In the latter disorder large differences in age of onset between parent and child do not always parallel differences in repeat length between generations (Andrew et al. 1993). The weaker correlation in Huntington’s disease between the number of CAG repeats and the age of onset when there is maternal as opposed to paternal origin suggests a sex-specific modifying effect (Snell et al. 1993).

SBMA is the exception with an X-linked pattern of inheritance compared with the autosomal dominant inheritance observed for all other CAG repeat diseases. This difference may provide clues to the pathophysiology of the neurological disease, for although females express AR (Takeda et al. 1990, Ruizveld de Winter et al. 1991, MacLean et al. 1995b), the symptoms and signs of SBMA are limited to males (Amato et al. 1993). Random X chromosome inactivation (Lyon 1972, Gartler & Riggins 1983) or lower androgen levels (Bardin & Lipsott 1967, Kirshner & Bardin 1972) may account for lack of disease expression in females. Mild electromyographic abnormalities have, however, been reported in the absence of neurological and biochemical abnormality in female heterozygote carriers (Sobue et al. 1993). AR is widely expressed in many tissues (Sar et al. 1990) including the central nervous and muscular system (Sar et al. 1990, Takeda et al. 1990, Doyu et al. 1994, Doumit et al. 1996). The absence of any neuromuscular deficit or degeneration in patients with complete androgen insensitivity (Quigley et al. 1995) suggests one hypothesis that a neurotoxic gain of function by the CAG expansion mutation may be triggered by the presence of androgens.

Androgen insensitivity and the CAG repeat

In Kennedy’s initial description of SBMA, one of the affected individuals had a mastectomy scar after being treated for gynaecomastia (Kennedy et al. 1968). Arbizu et al. (1983) subsequently emphasized the frequent presence of signs indicating the development of androgen insensitivity. Men with
SBMA usually manifest varying degrees of partial androgen insensitivity including gynaecomastia, testicular atrophy, oligospermia, azoospermia and elevated serum gonadotrophins (Harding et al. 1982, Arbizu et al. 1983, Sobue et al. 1994, MacLean et al. 1995a). Diabetes mellitus is also a rare association (Harding et al. 1982, Arbizu et al. 1983). The likelihood of gynaecomastia correlates with increasing repeat length (MacLean et al. 1995a, Shimada et al. 1995). The delayed presentation of androgen resistance in SBMA compared with that associated with other AR gene mutations remains unexplained. The clinical findings suggest attrition of AR function with age and a pathophysiology probably distinct from that involved in the neurodegenerative process.

Several previous reports indicate that expansion of the Gln repeat in the AR NH 2-terminal region results in a structurally altered protein with reduced transcriptional capacity (Mhatre et al. 1993, Chamberlain et al. 1994, Trifiro et al. 1994, Tut et al. 1997). The findings that [3H]R1881 equilibrium binding affinity and protein expression by immunoblot are usually normal in this setting support a Gln tract mediated inhibition of the transactivation function (Chamberlain et al. 1994). Others report a reduced equilibrium binding affinity of [3H]R1881 in suprapubic skin fibroblasts from a normal Kd of 0·19 nM to 0·24-11·7 nM in SBMA patients (MacLean et al. 1995a). The change in affinity correlated with the severity of testicular atrophy and gynaecomastia (MacLean et al. 1995a). Gln lengths greater than 28 have recently been associated with an increased risk of impaired spermatogenesis (Tut et al. 1997). Our analysis using AR constructs with expanded CAG repeats in the range associated with SBMA raises an alternative hypothesis that reduced AR mRNA and protein expression (Choong et al. 1996a) may be responsible for the partially androgen insensitive phenotype.

**CAG Expansion in the Amino-Terminus of AR Affects AR Expression**

We evaluated the effect of variation of CAG repeat length on AR RNA expression in vitro and found an inverse correlation between the length of the CAG trinucleotide repeats and the extent of AR mRNA expression and AR protein levels (Choong et al. 1996a). The CAG repeat in human AR immediately precedes a single CAA codon for Gln (Lubahn et al. 1988a) making the number of Gln residues CAGn+1. Equivalent amounts of the full-length human AR expression vector DNA with 0, 14, 23, 43 and 65 CAG repeats in the region of exon 1 that codes for the NH 2-terminal domain were transiently expressed in monkey kidney COS cells and RNA was extracted for Northern blot analysis. AR mRNA levels were inversely related to CAG repeat length. CAG repeat lengths of 43 and 65, which are associated with SBMA, resulted in a significant reduction in AR mRNA levels when compared with wild-type AR with 23 repeats or with 0 or 14 CAG repeats. A 4- to 5-fold reduction was observed between 0 and 65 CAG repeats and a 2- to 3-fold reduction between the wild-type length of 23 CAG repeats to 65 CAG repeats associated with SBMA. When the data from several experiments were normalized by rehybridization with a glyceraldehyde-3-phosphate dehydrogenase probe, there was a linear inverse relationship between CAG repeat length and levels of AR mRNA expression. Similarly, levels of AR mRNA expression in CV1 cells used in transcription assays were inversely correlated with CAG repeat length, although absolute levels were reduced relative to COS cells. Analysis of immunoblot and [35S]methionine incorporation into expressed AR protein in COS cells showed a direct relationship between reduced AR mRNA and protein levels with reduction in AR signal intensity as the CAG repeat expanded to 43 and 65. The rate of AR protein degradation using pulse/chase labelling with [35S]methionine (Zhou et al. 1995) was not altered by variations in the length of the polyglutamine tract encoded by the CAG repeat sequences.

COS cells transfected with increasing CAG repeat-length AR expression vector DNA using calcium phosphate precipitation were analysed for AR binding affinity and capacity for [3H]R1881, a radiolabelled synthetic androgen. Scatchard plot analysis revealed equivalent equilibrium binding affinity for AR encoded by sequence containing 0, 23, 43, or 65 CAG repeats but binding capacity was reduced for AR with sequence containing 43 or 65 CAG repeats. The half-time of [3H]R1881 dissociation at 37 °C of 3·0 ± 0·5 h determined as previously described (Zhou et al. 1995) was not influenced by the 44 or 66 Gln repeat length. Thus binding capacity appeared to be the only parameter of androgen binding that was altered by expansion of the Gln repeat with the apparent equilibrium binding constant remaining unaffected. The decrease in binding capacity most likely reflected the observed decrease in AR mRNA and protein expression with CAG expansion.

Androgen-induced transcriptional activity of AR with increasing CAG repeats was tested by transient cotransfection in monkey kidney CV1 cells using a luciferase reporter vector containing the androgen
responsive mouse mammary tumour virus promoter. Luciferase activity induced by 0.01 and 0.1 nM dihydrotestosterone was indistinguishable from recombinant AR of increasing Gln repeat lengths when expressed as fold induction relative to background activity determined in the absence of dihydrotestosterone. Background levels of luciferase activity detected in the absence of androgen were lower with CAG repeat expansion probably reflecting decreased levels of AR expression. Since the fold induction of transcriptional activation was similar to that of wild-type AR, the results suggest that increased Gln repeat lengths do not alter the inherent transcriptional activity of AR. In our studies of a family with partial androgen insensitivity, a glutamic acid to lysine mutation in the second codon of the AR gene reduces AR protein expression by alteration of the nucleotide context for translation initiation with no inherent decrease in transcriptional activation (Choong et al. 1996b). The association of the partial androgen insensitivity phenotype with reduced levels of AR expression due either to CAG repeat expansion or partial disruption of the initiation codon suggest that a threshold level of AR is required for full reproductive capacity.

**Reduced AR expression accounts for androgen insensitivity in X-linked SBMA**

Disruption of AR functional activity results in the syndrome of androgen insensitivity which, in its complete form, is characterized by a female external phenotype at birth and loss of normal male sexual development in 46 XY genetic males (Quigley et al. 1995). Partial androgen insensitivity is characterized by incomplete masculinization and results from gene mutations that cause less severe disruption of AR function. AR missense mutations usually interfere with androgen or DNA binding (French et al. 1990, McPhaul et al. 1993, Brown 1995). SBMA patients undergo normal prenatal and pubertal development and are fertile, indicating the presence of a functional AR. As adults, usually after the age of 30 years (Brooks & Fischbeck 1995) and in proportion to CAG repeat length, features of partial androgen resistance develop including gynaecomastia and elevated serum gonadotrophin levels (Shimada et al. 1995). Gynaecomastia in the adult male is a sensitive indicator of disturbed androgen/oestrogen ratios or of impaired androgen action mediated by AR (Wilson et al. 1980). The concentration of AR begins to decrease in the normal male between the ages of 20–30 years, as shown strikingly in human foreskin fibroblasts (Roehrborn et al. 1987). One hypothesis suggested by our findings is that adult onset androgen insensitivity observed in SBMA results from reduced AR mRNA and protein production compounded by normal AR attrition and the normal age-dependent decline in androgen levels (Zumoff et al. 1982, Nahoul & Roger 1990). During embryonic development and puberty, higher AR expression and androgen production may compensate for the inhibitory effects of the first exon expanded CAG repeat on AR mRNA expression allowing normal male sex development to occur.

Previously results using tissue from affected individuals support the hypothesis that reduced AR mRNA and protein expression is the molecular basis for androgen insensitivity in SBMA patients. Equilibrium androgen binding affinity was unchanged but B_max was reduced in SBMA patients with gynaecomastia (Warner et al. 1992). Absence of AR immunostaining in scrotal skin from affected subjects (Matsuura et al. 1992) contrasted strong immunostaining in age-matched controls indicating reduced AR protein expression in affected subjects. Reverse transcriptase-polymerase chain reaction (RT-PCR) analysis of spinal cord showed 4- to 7-fold reduced AR mRNA levels compared with normal controls and undetectable AR protein by immunoblot analysis (Nakamura et al. 1994). This difference is greater than that observed in our studies described above and may be due to the use of the strong heterologous cytomegalovirus promoter in our AR recombinant expression vectors. Reduced androgen binding capacity with normal affinity was reported in SBMA cultured genital skin fibroblasts (Danek et al. 1994). Several clones of a stable neuronal cell line prepared with a 65 CAG repeat expansion had reduced AR protein levels compared with wild-type controls (Brooks et al. 1997).

It is important to note that subjects with complete androgen insensitivity have normal muscle strength and neuronal function indicating that reduced AR expression cannot account for the progressive muscle weakness associated with SBMA.

**Molecular mechanisms for reduced AR RNA expression with CAG expansion**

Decreased AR mRNA expression with CAG expansion may result from a common mechanism involving transcriptional interference. Pertinent to our findings for the CAG repeat in AR, the CTG triplet repeat in the mouse growth inhibitory factor/metallothionein III gene promoter represses transcriptional activity in a direction- and position-independent manner (Imagawa et al. 1995). One possible mechanism for transcriptional interference by trinucleotide repeat expansion is preferential
nucleosome assembly. In myotonic dystrophy and Huntington’s disease, electron microscopy reveals increased nucleosome formation at the expanded CTG (CAG) triplet repeat that could interfere with transcription complex progression (Wang et al. 1994). Although CTG repeat expansion failed to enhance nucleosome formation, preferential nucleosome assembly in the repeat region was thought to contribute to repeat expansion (Godde & Wolfe 1996). The repeat length associated with most triplet repeat disease phenotypes correlates with the length of DNA required for nucleosome formation (146 bp).

A number of other mechanisms by which expanded triplet repeat sequences could result in transcriptional interference have been described. For example hypermethylation of CGG triplet repeats causes reduced expression of the fragile X mental retardation gene (Pieretti et al. 1991), RNA polymerase pausing proportional to repeat length (Kang et al. 1995b), formation of hairpin structures by triplet repeats (Nadel et al. 1995, Wells 1996) and intermolecular DNA triplex formation (Ohshima et al. 1996). With respect to the AR, our studies raise the possibility that reduced mRNA levels associated with the CAG trinucleotide repeat expansion could result from abnormal association with RNA binding proteins.

Experimental data from Ikeda et al. (1996) indicate accelerated apoptosis in transfected cells and reduced expression of the Machado-Joseph disease protein type 1 fragments containing the expanded polyglutamine repeat. Scherzinger et al. (1997) found the expression of glutathione-S-transferase (GST)-fusion huntingtin protein containing 83 and 122 polyglutamine repeats to be markedly reduced compared with that of proteins with less than 35–40 Gln repeats. Scherzinger et al. (1997) proposed that the reduced expression of the truncated protein may be due to accelerated cell death. Contrary to this hypothesis we noted equivalent levels of glyceraldehyde-3-phosphate dehydrogenase expression in total RNA extracted from a standard number of cells. We did not observe a difference in cell density at the time of RNA extraction. However, we did not exclude the possibility that accelerated apoptosis contributes to reduced AR mRNA expression observed in our studies.

**HYPOTHESES FOR NEUROTOXICITY OF THE CAG EXPANSION**

**Altered RNA metabolism may contribute to CAG repeat associated disease pathogenesis**

Although the CAG repeat gene products are heterogeneous in their cellular location, function and presence in coding sequences, expression as mRNA is a common feature making RNA metabolism a possible site of altered function. The known interaction between mRNAs and small nuclear ribonucleoproteins and RNA-binding proteins suggests a possible site of action (Dreyfuss et al. 1988).

McLaughlin et al. (1996) isolated a 63 kDa RNA-binding protein from human and rat cortical and striatal extracts that specifically interacts with the CAG repeat sequence in cRNA transcribed in vitro from huntingtin and AR constructs. The efficiency of the interaction depended on the length of the CAG repeat (McLaughlin et al. 1996). Our finding of an inverse correlation between CAG repeat length and AR mRNA expression (Choong et al. 1996a) may be a consequence of an inappropriate RNA binding-protein interaction with specific mRNA which may disturb cellular dynamics or alter the regulation, transport and expression of CAG containing RNA.

In support of this hypothesis, recent work by Davis et al. (1997) demonstrated increased nuclear retention of mutant RNA transcripts of the myotonic dystrophy protein kinase containing the CTG expansion in differentiated cultured myoblasts. The mutant transcripts are polyadenylated and spliced, form stable clusters that are tightly linked to the nuclear matrix, and do not appear to have decreased stability in experiments where actinomycin D is used to block RNA synthesis (Davis et al. 1997). Hamshere et al. (1997) observed both nuclear retention and reduced polyadenylated mutant expanded myotonic dystrophy protein kinase transcripts. These findings for myotonic dystrophy have not been tested for the CAG repeat diseases.

**Protein–protein interactions**

The mechanism whereby an expanded polyglutamine repeat leads to neurodegenerative disease and the underlying mechanism for the susceptibility of neuronal tissue to the pathogenic consequences of CAG expansion are unknown. The mode of inheritance and ubiquitous expression of the mutant proteins in tissues from affected individuals (Strong et al. 1993) and the gene-specific localization of pathology suggests a gain of toxic function dependent on specific neuronal factors (Paulson & Fischbeck 1996).

Polyglutamine tracts are functionally important domains in a number of ubiquitous transcription factors (Mitchell & Tjian 1989, Gerber et al. 1994). The AR is the only gene product of the CAG trinucleotide repeat diseases with a clearly defined...
functional role as a transcription factor (Zhou et al. 1994). One hypothesis is that the expanded AR CAG repeat interacts with as yet unspecified cofactors to cause aberrant transcriptional activation of genes within the neuron resulting in accelerated degeneration.

A number of protein–protein interactions with potential pathogenic significance have been proposed and demonstrated by a variety of techniques. Using the yeast two hybrid protein interaction assay and in vitro binding with glutathione S-transferase fusion proteins, the huntingtin-associated protein (HAP)-1 (Li et al. 1995) was discovered. This protein is selectively expressed in the brain, with enhanced binding in the presence of expanded huntingtin Gln repeat lengths as seen in Huntington’s disease patients. The polyglutamine region of huntingtin has also been shown to interact with calmodulin in a calcium-dependent fashion (Bao et al. 1996).

Glyceraldehyde-3-phosphate dehydrogenase is a ubiquitous multifunctional enzyme essential to glycolysis. It interacts selectively with the normal polyglutamine proteins of Huntington’s disease and dentorubral pallidoulusyan atrophy (Burke et al. 1996, Roses 1996) as well as the polyglutamine region of the spinocerebellar atrophy type 1 gene product ataxin-1 and the AR (Koshy et al. 1996). Other potential huntingtin interacting proteins include a ubiquitin-conjugating enzyme (Kalchman et al. 1996) and a human homologue of S. cerevisiae Sla2p, shown to be important in cytoskeletal integrity (Kalchman et al. 1997, Wanker et al. 1997).

Goldberg et al. (1996) reported accelerated degradation of huntingtin by apoptotic extracts and by the key proteolytic enzyme apopain, the human counterpart of the nematode cysteine protease Sla2p, shown to be important in cytoskeletal integrity (Kalchman et al. 1997, Wanker et al. 1997).

Goldberg et al. (1996) reported accelerated degradation of huntingtin by apoptotic extracts and by the key proteolytic enzyme apopain, the human counterpart of the nematode cysteine protease death-gene product, CED-3. The rate of cleavage of huntingtin was coincident with activation of apopain and increased with the length of the polyglutamine repeat (Goldberg et al. 1996). Post mortem examination of neural tissue from patients with Huntington’s disease suggests that accelerated apoptosis occurs with increased non-random distribution of DNA fragmentation in cells in the corpus striatum compared with normals (Dragunow et al. 1995). In the neostriatum of individuals with Huntington’s disease, the pattern of distribution of TUNEL (TdT-mediated dUTP-digoxigenin nick end-labeling)-positive apoptotic neurons and glia was similar to that seen during normal development of the nervous system (Portera-Cailliau et al. 1995). One hypothesis suggested by these observations is that amino-terminal protein fragments with expanded polyglutamine tracts that result from accelerated degradation of huntingtin by proteases such as apopain, associate with glyceraldehyde-3-phosphate dehydrogenase, thereby disrupting energy metabolism and promoting the neuroatrophy (Vonsattel et al. 1985, Myers et al. 1991) observed in Huntington’s disease.

It is conceivable that the polyglutamine region interacts with a number of cellular proteins. In some of these interactions a variation in the polyglutamine length could affect the tenacity of the interaction. Altered protein–protein interactions with the polyglutamine region have been demonstrated by several methods, lending support to a gain of function theory.

Perutz et al. (1994) suggested that Gln repeats function as polar zippers, for example, joining specific transcription factors bound to separate DNA segments. Glutamine expansions may cause disease by increased nonspecific affinity between such factors or by gradual precipitation of the affected proteins in neurons (Perutz et al. 1994). Another proposed neuropathic mechanism is cellular toxicity resulting from enhanced aggregation of proteins with an expanded polyglutamine tract secondary to transglutaminase cross-linking (Green 1993). Accumulation over time of these toxic products or, more likely, a sequestration of critical proteins from their normal function, could account for the relationship between polyglutamine length and age of onset of the neurodegenerative phenotype.

Scherzinger et al. (1997) have shown that proteolytic cleavage of a GST-huntingtin fusion protein leads to formation of insoluble high molecular weight protein aggregates when the protein contains a polyglutamine expansion which is in the pathological range. When analysed by SDS-polyacrylamide gel electrophoresis, insoluble high molecular weight proteins were detected with proteins containing the expanded polyglutamine tracts of 81 or 122 repeats. The critical length for formation of aggregates in these experiments was suggested to be between 35 and 40 repeats. Electron micrographs of these aggregates revealed a fibrillar structure reminiscent of β-amyloid fibrils in Alzheimer’s disease and resembled the intranuclear inclusion bodies detected in symptomatic mice transgenic for the Huntington’s disease mutation. Recombinant proteins migrated at a size corresponding closely to the predicted molecular weight when still part of the fusion protein. However, when the GST tag was removed by trypsin, HD51 migrated with two protein products at 37 and 60 kDa.

Transgenic experiments in mice overexpressing the expanded polyglutamine domain of either ataxin 1 (Burright et al. 1995) or the isolated...
polyglutamine tract of the spinocerebellar atrophy type 3 gene into Purkinje cells (Ikeda et al. 1996) suggest that accelerated apoptosis of specific neurons contributes to neurodegeneration. Of particular note, no phenotype was observed in mice overexpressing the entire mutated expanded spinocerebellar atrophy type 3 gene (Ikeda et al. 1996). However, mice transgenic for the 5’ end of the human Huntington’s disease gene carrying (CAG)_{115} to (CAG)_{150} repeat expansions demonstrated a progressive neurological syndrome that exhibited many of the characteristic features of Huntington’s disease (Mangiarini et al. 1996). The transgene was ubiquitously expressed with some variability in tissue levels. The movement disorder in the transgenic Huntington’s disease mice was distinct from that reported for mice transgenic for the spinocerebellar atrophy type genes (Burright et al. 1995, Ikeda et al. 1996). Nor was there evidence for muscular atrophy as seen in SBMA, indicating that the effect of the polyglutamine repeat is specific to the protein in which it occurs.

Using antibodies directed to the amino-terminus of huntingtin just proximal to the expanded CAG repeat (amino acids 1–17), Davies et al. (1997) demonstrated that mice transgenic for exon 1 of the human Huntington’s disease gene carrying (CAG)_{115} to (CAG)_{150} repeat expansions developed neuronal intranuclear inclusions containing the proteins huntingtin and ubiquitin prior to the onset of the neurological disorder. Immunoreactivity in symptomatic transgenic mice was specifically concentrated in a densely stained circular intranuclear inclusion within the neuronal but not glial cells. Antibodies which recognize regions outside the transgene did not stain the inclusion, indicating that endogenous mouse huntingtin protein did not co-localize with the ectopic transgenic protein. Immunocytochemical analysis demonstrated localization of ubiquitin in the nuclear inclusion. While accumulation of reaction product was also evident within the cytoplasm, there was no co-localization of ubiquitin in these cytoplasmic vesicles. More frequent indentations and an apparent increased density of pores in the nuclear membrane was also associated with the intranuclear inclusion. The temporal expression of the intranuclear inclusion bodies preceded onset of the neurological disorder and loss of brain and body weight in symptomatic transgene animals (Davies et al. 1997), supporting a causal role in the etiology of the neurological disorder. Detection of neuronal intranuclear inclusion bodies by immunocytochemistry using antiserum to the huntingtin amino-terminal epitope (amino acids 1–17) in post mortem brain tissue from Huntington’s disease patients suggests that these inclusions play a role in the pathophysiology of Huntington’s disease (DiFiglia et al. 1997). The neuronal distribution of the intranuclear inclusion bodies, the higher incidence of these aggregations in brains from juvenile onset Huntington’s disease patients, and their absence in a presymptomatic adult support this hypothesis (DiFiglia et al. 1997).

A recent report by Ordway et al. (1997) of late onset neurological disease associated with neuronal intranuclear inclusion in transgenic mice expressing a mutant hypoxanthine phosphoribosyltransferase protein containing an expanded CAG repeat of 146 residues encoded within exon 3 suggests that the neurotoxicity of the polyglutamine repeat relates to its presence in smaller or truncated proteins.

**STRUCTURAL EFFECTS OF CAG REPEAT EXPANSION IN AR**

{superscript}35S-Methionine labelling of AR amino-terminal fragments with increasing CAG repeat lengths up to 64, where 64 was derived from a patient with SBMA (La Spada et al. 1992), resulted in alterations in the spacing and migration of the AR fragments on SDS gels that was not accounted for simply by changes in molecular weight (Choong et al. 1998). Spacing between the double bands which resulted from phosphorylation was greatest for a frequent polymorphism (23 CAG) in the normal human population. Correlation between CAG repeat length and increased spacing between doublet bands has been reported previously (Jenster et al. 1994). AR migration on SDS gels appeared to be influenced by both CAG repeat length and serine 94 phosphorylation, raising the possibility of a structural interaction between serine 94 phosphate and the CAG Gln repeat that persists in the presence of SDS. With isotope dilution, the bands shifted to the upper band. From the extent of band shift, phosphorylation occurs more rapidly as the CAG repeat expands. About 90% of the lower band shifts for CAG 64 whereas CAG 4 from lemur remains predominantly in the lower unphosphorylated form, suggesting that CAG repeat length alters the rate of serine 94 phosphorylation. This conclusion is supported by the lack of change in migration during isotopic dilution of AR fragments which have a serine 94 to alanine mutation and fragments in which the CAG repeat is deleted in ΔA14–83.

We and others have observed the SDS polyacrylamide gel electrophoresis mass of full-length AR containing the expanded CAG repeat is close to the calculated mass (Jenster et al. 1994, Choong et al. 1996a). However migration of truncated
AR fragments during SDS polyacrylamide gel electrophoresis was retarded suggesting formation of higher molecular weight complexes possibly due to aggregation of the truncated fragments. This observation of altered migration of the AR truncated fragment raises the possibility of a common pathogenic mechanism for neurodegeneration for the CAG triplet repeat diseases whereby protein aggregation may occur after AR proteolysis to smaller fragments encompassing an expanded CAG repeat and thus contribute to loss of neuronal cell function associated with SBMA. Full-length human AR with 44 glutamines expressed in transfected COS cells was more resistant to proteolysis by trypsin compared with wild-type AR with 20 glutamines. In these studies COS cells transfected with human AR with (CAG)44 showed increased apoptosis compared with cells expressing wild-type AR (Abdullah et al. 1998). A pathophysiological mechanism such as this is consistent with the observation that full-length AR with CAG repeat sequence expanded into the pathological range has functional activity similar to that of the wild-type receptor and supports normal male sexual development and function until adulthood. Cytoplasmic aggregation of an AR NH2-terminal fragment with overexpression of full-length AR with an expanded CAG repeat in mouse neuroblastoma NB2a/d1 cells treated with androgen suggested that protein aggregation may contribute to the phenotype of SBMA (Butler et al. 1998).

**Strategies for further investigation of androgen receptor-mediated neurodegeneration**

Our observation of altered migration of an insoluble AR truncated fragment supports a common pathogenic pathway for neurodegeneration in CAG triplet repeat diseases. We postulate that AR with an expanded CAG repeat is preferentially digested by a protease. Aggregated AR fragments may accrue with time, resulting in the formation of insoluble complexes which trigger apoptosis. Evidence for this pathway might derive from the demonstration of accelerated apoptosis in anterior spinal horn cells from individuals with spinal bulbar and muscular atrophy, aggregation of truncated AR fragments in the nuclear compartments of these neurons, and the identification of the proteolytic enzymes that interact specifically with AR in the anterior horn cells. Localization of putative proteolytic enzymes to the nuclear or cytoplasmic compartments would suggest strategies for manipulating the aberrant pathway via the AR nuclear targeting signal. In the light of our findings of reduced AR mRNA levels with expanded CAG repeat length, altered RNA metabolism is potentially another mechanism involved in the neurological dysfunction associated with the expanded CAG repeat in AR. Elucidation of pathways involving the role of CAG repeats as cis acting regulatory elements for AR mRNA, stability of AR mRNA metabolism in neuronal and nonneuronal cells, and the effects on function of other critical mRNAs or proteins, will provide insight into the pathophysiology of neurodegeneration and suggest strategies for prevention.

**FUNCTIONAL SIGNIFICANCE OF THE AR CAG REPEAT POLYMORPHISM IN PROSTATE CANCER**

The AR is critical for the normal development of the male urogenital tract. During embryogenesis AR is initially expressed in mesenchymal cells of the urogenital anlagen (Husmann et al. 1991). In the presence of androgen there is increased expression of AR positive mesenchymal cells and subsequent expression of receptor in epithelial cells (Bentvelsen et al. 1995). Differentiation of the prostatic epithelial cells is triggered by androgen (Cunha et al. 1987). The absence of prostatic development in complete androgen insensitivity syndrome (Quigley et al. 1995) and in individuals with 5α-reductase type 2 deficiency (Thigpen et al. 1992) underscores the primary role of the AR and 5α-dihydrotestosterone in prostatic development in humans. The secretory epithelium of the mature prostate expresses high levels of AR, but the interaction between the mesenchyme and epithelial compartments is not clearly defined. Maintenance of prostatic architecture and function in the adult has been shown to be dependent on androgen action (Cunha et al. 1987, Mooradian et al. 1987).

Prostate cancer is a major cause of mortality and morbidity in the aging populations of industrialized countries. Current screening methods have resulted in increased detection of early localized disease, albeit with no appreciable change in prostate cancer-related mortality thus far (Lu-Yao & Greenberg 1994). Ninety-five percent of prostate cancers are adenocarcinomas which derive from the secretory acinar. There is a correlation between the degree of prostate tumour differentiation and AR content (Chodak et al. 1992). The majority of prostate cancers at diagnosis are hormone dependent, and androgen drug ablation with orchietomy or luteinizing hormone-releasing hormone agonists and antiandrogen administration are standard therapeutic approaches. Meta-analysis of 22 ran-
domized trials demonstrated a mortality rate of 57% in patients with metastatic disease at 40 months after treatment with no significant survival difference between groups treated with surgical castration or maximal androgen blockade (Anonymous 1995). Although androgen ablation induces remission in a significant proportion of patients (Gittes 1991), the subsequent emergence of hormone independent tumours often leads to relapse (Frydenberg et al. 1997). A positive family history strongly predicts the likelihood of developing prostate cancer (Hayes et al. 1995, Narod et al. 1995, Whittemore et al. 1995) with a calculated relative risk of 2·62 (95% confidence interval (CI) 1·69-4·06) in patients with affected brothers and 1·22 (95% CI 0·77-1·94) in those with affected fathers (Narod et al. 1995). The higher relative risk in brothers observed in these case control studies suggests an X-linked genetic contribution to disease pathogenesis (Hayes et al. 1995, Narod et al. 1995).

The pathological role of AR in prostate carcinogenesis, metastases and in the evolution of hormone refractoriness is not well understood. Irvine et al. (1995) investigated possible linkage disequilibrium between the CAG and GGC microsatellites of AR and observed an association between shorter (<22 repeats) CAG alleles and longer (>16 repeats) GGC alleles and increased risk of prostate cancer in Caucasian males. A significant correlation between CAG repeat length and age at onset was observed in one study of 109 men (Hardy et al. 1996). A large prospective case control study of 587 predominantly Caucasian (>95%) subjects from the Physician’s Health Study demonstrated an inverse relationship in all age groups between CAG repeat length and risk of aggressive prostate cancer, as characterized by high histological grade, distant metastases at diagnosis and disease-related mortality (Giovannucci et al. 1997). Two further studies support an association between reduced AR CAG repeat length with increased risk of prostate cancer, with the length of the polymorphic GGC repeat being an additional potential risk factor for development of prostate cancer (Hakimi et al. 1997, Stanford et al. 1997). A survey of 11 125 patients 40 years or older with no evidence of prostate cancer from the Baltimore Veterans Affairs Medical Center reported that the normal range for prostate-specific antigen was higher in African American males than in Caucasians in all age groups. The greatest difference was in those men aged 70 years or over, where the ratio was 1·6 (Sawyer et al. 1996, 1997). African American males have previously been shown to have a distribution of CAG repeats in the AR gene which is shifted to slightly shorter repeat lengths (Irvine et al. 1995). These findings provide circumstantial evidence for increased AR expression and activity in vivo resulting from variation in the length of the CAG tandem repeat and could partially account for the higher incidence of prostate cancer (Boring et al. 1992) in African Americans compared with Caucasian males in the United States.

Progression of prostate cancer is associated with genetic instability. Reported genetic aberrations include allelic loss of chromosome 8p in 50% of cases (Boring et al. 1992) and loss of heterozygosity in 20–30% of cases for chromosome regions 10q, 16p, and 17p (Cannon-Albright & Eeles 1995). Mutations in the ras oncogene, the p53 tumour suppressor gene, trisomy 7 and the MX1 gene (Eagle et al. 1995) have also been observed. The majority of somatic AR mutations were detected in hormone refractory cancer or in the setting of advanced metastatic disease (Gaddipati et al. 1994, Taplin et al. 1995, Suzuki et al. 1996). Tilley and coworkers (1996) identified mutations in 11 of 25 primary prostate tumours sampled prior to initiation of hormone therapy using single strand polymorphism screening of PCR products. Eight missense mutations were identified in exon 1. Although the functional effects of individual mutations were not characterized in vitro, the presence of AR amino acid mutations is associated with decreased intensity of immunocytochemical staining for AR in tumour cells and a significant correlation with poor response to hormone therapy (Tilley et al. 1996).

The importance of altered androgen responsiveness in tumour progression is supported by the identification of somatic mutations in AR from prostate cancer that alter ligand binding specificity (Veldscholte et al. 1990, Newmark et al. 1992, Culig et al. 1993). A somatic microsatellite mutation in the CAG trinucleotide repeat region resulting in a reduction from 24 to 18 repeats was identified in one out of forty prostate adenocarcinoma specimens of various grades obtained at radical prostatectomy and screened for mutations at this locus (Schoenberg et al. 1994). Although only the CAG repeat was analysed, the authors postulate that the high relative abundance of the mutant allele was evidence for a selective growth advantage conferred by contraction of the CAG repeat in the AR gene. Despite the reported incidence of 14-6% microsatellite instability in advanced disease (Suzuki et al. 1995), mutation at the AR CAG repeat has rarely been reported. This may reflect the fact that the DNA and ligand binding domains are the regions most commonly screened for mutations. In addition, the high GC content of the region of the AR gene that encodes the amino-terminal domain may lower the sensitivity of screening techniques.
such as denaturing gradient gel electrophoresis (De Bellis et al. 1992) for detection of CAG mutations. This does not appear to be the case, however, by single strand polymorphism screening of PCR products (Tilley et al. 1996).

Studies of hormone refractory tumours demonstrated a 4-fold AR gene amplification relative to X chromosome copy number in 30% of specimens using fluorescent in situ hybridization with an AR specific and X chromosome reference probe. These results suggest that AR is important in the development of acquired hormone resistance. Of note, significant intratumour variability was observed with individual cells expressing 5–60 AR genomic copies/cell and no evidence of AR gene amplification in paired untreated tumour samples. Both primary and recurrent tumours demonstrated strong nuclear staining, suggesting a transcriptionally active AR (Visakorpi et al. 1995). In a larger clinical study of 54 patients, the same group reported concomitant AR gene overexpression by mRNA in situ hybridization in AR amplified recurrent tumour samples. There was a more favourable response to primary therapy and longer median post recurrence survival in these patients compared with those where no amplification was observed (Koivisto et al. 1997). One implication of this work is that AR amplification with enhanced expression in hormone responsive tumours may confer a proliferative advantage by allowing cells to resume androgen dependent growth in a low androgen environment (Koivisto et al. 1997).

CONCLUSIONS

Given the structural constraints on the DNA and ligand binding domains, the effects of the CAG trinucleotide repeat length on AR expression and phosphorylation may set the background level of AR activity in response to androgen stimulation. Polymorphic variation of the CAG repeat in human AR probably contributes to the observed individual phenotypic differences in androgenization. It may also be one determinant of the magnitude of cellular responses to carcinogenic events that induce the development of prostate cancer.

Will the AR gene CAG expansion continue?

Short reiterated microsatellite sequences of 1–5 nucleotides tend to lengthen in humans relative to other primates suggesting that microsatellite evolution is directional (Rubinsztein et al. 1995a) and may result from a mutational bias towards longer repeats. In non-human primates, CAG iteration is shorter in the AR and other genes associated with human neurodegenerative diseases (Rubinsztein et al. 1995b, Djian et al. 1996). Although no universal model accounts for expanding trinucleotide repeats during mammalian evolution, older age at reproduction in human males compared with non-human primates or a larger pool of heterozygote alleles associated with a diverse population, could contribute to higher mutation rates (Rubinsztein et al. 1995a).

Natural selection tends to eliminate harmful mutations and maintain advantageous ones, yet silent mutations with no adaptive advantage can become fixed in a genome and reflect time of divergence of a species (Ayala 1995). Evolutionary expansion of the AR CAG repeat sequence may reflect natural selection rather than a neutral expanding polymorphism. CAG repeat expansion is associated with reduced AR expression and a Gln repeat–serine phosphate structural motif. Exponential expansion of the AR CAG repeat during primate evolution might continue into the range associated with androgen insensitivity and neurodegeneration. However, the deleterious effects that could result from reduced AR expression and gain-of-function defects would be expected to impair human fertility and thus arrest the expansion through natural selection.

ACKNOWLEDGEMENTS

For providing primate blood samples, we thank Michael Adams, Debbie Golden and Claude L Hughes, Department of Comparative Medicine, Bowman Gray School of Medicine, Wake Forest University for macaque, Lorraine Meller, North Carolina Zoological Park for chimpanzee, Kenneth Glander and Cathy Williams, Duke University Primate Center for lemur, and David J Handelsman, Royal Prince Alfred Hospital, University of Sydney, Australia, for baboon from the self sustaining National Health and Medical Research Council National Baboon breeding colony, supported by the Australian National Health and Medical Research Council, Bullens Primate Research Foundation and the Medical Foundation of the University of Sydney. We thank Rytas Vilgalys at Duke University for advice, Jon A Kemppainen and K Michelle Cobb for technical assistance and John R Dyer for reading the manuscript. The work was supported by grants HD16910 and HD35041 from the National Institute of Child Health and Human Development Center for Population Research, Research Award 95–193 from the Genentech Foundation for Growth and Development and by the Andrew W Mellon Foundation.

Journal of Molecular Endocrinology (1998) 21, 235–257
REFERENCES


Chung MY, Ranum LP, Duvick LA, Servadio A, Zoghbi HY & Orr HT 1993 Evidence for a mechanism predisposing to...
intergenerational CAG repeat instability in spinocerebellar ataxia type 1. Nature Genetics 5 254–258.


Faber PW, Kuiper GGJM, van Rooij HCJ, van der Korput JAGM, Brinkmann AO & Tramp J 1989 The N-terminal domain of the human androgen receptor is encoded by one large exon. Molecular and Cellular Endocrinology 61 257–262.


Hussmann DA, McPhaul MJ & Wilson JD 1991 Androgen receptor expression in the developing rat prostate is not altered by castration, flutamide, or suppression of the adrenal axis. *Endocrinology* 128 1902–1906.


Myers RH, Vonsattel JP, Paskevich PA, Kiely DK, Stevens TJ, Copple LA, Richardson JP & Bird ED 1991 Decreased neuronal and increased oligodendroglial densities in...


Rundlett SE, Wu XP & Miesfeld RL 1990 Functional characterizations of the androgen receptor confirm that the molecular basis of androgen action is transcriptional regulation. *Molecular Endocrinology* **4** 708–714.


Yeh S & Chang C 1996 Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. Proceedings of the National Academy of Sciences of the USA 93 5517–5521.


RECEIVED 6 April 1998