Androgen receptor gene CAG and GGC repeat lengths in idiopathic male infertility

A.Ferlin¹, L.Bartoloni¹, G.Rizzo¹, A.Roverato², A.Garolla¹ and C.Foresta¹,³

¹University of Padova, Department of Medical and Surgical Sciences, Centre for Male Gamete Cryopreservation, Padova and
²University of Modena and Reggio Emilia, Department of Social, Cognitive and Quantitative Sciences, Modena, Italy
³To whom correspondence should be addressed at: University of Padova, Department of Medical and Surgical Sciences, Centre for Male Gamete Cryopreservation, Via Ospedale 105, 35128 Padova, Italy. E-mail carlo.foresa@unipd.it

The androgen receptor (AR) has two polymorphic sites in exon 1, characterized by different numbers of CAG and GGC repeats resulting in variable lengths of polyglutamine and polyglycine stretches. Longer CAG repeats result in a reduced AR transcriptional activity, whereas the role of the GGC triplets is less clear. A relationship between decreased spermatogenesis and moderate expansion in the CAG tract has been found in some studies, but not in others. Furthermore, the joint distribution of CAG and GGC repeats in male infertility has never been reported before. We analysed CAG and GGC repeat lengths in a group of 163 men with idiopathic infertility compared with 115 fertile normozoospermic men. No difference was found between patients and controls in the mean and median values, and in distribution of CAG and GGC, when considered separately. However, the analysis of the joint distribution of CAG and GGC showed that the distribution of particular haplotypes is significantly different between patients and controls. In particular, two CAG/GGC haplotypes seem to increase susceptibility to infertility (CAG = 21/GGC = 18 and CAG >21/GGC >18, relative risk 2.47 and 1.6), while one haplotype (CAG >23/GGC ≤16, relative risk 0.09) seems to confer a protective effect against the disease. These data show a combined effect of CAG and GGC repeat numbers on AR function and the first evidence of a relationship of particular CAG/GGC haplotypes with male infertility.

Key words: androgen receptor/CAG/GGC/male infertility/triplets

Introduction

Androgens and a functioning androgen receptor (AR) are essential for development and maintenance of the male phenotype and spermatogenesis. Consistent with this, mutations in the AR gene cause a variety of defects related to androgen insensitivity with the less severe phenotype represented by isolated male infertility (Aiman et al., 1979; Quigley et al., 1995; Hiort et al., 2000).

The AR is encoded by a gene located on chromosome Xq11–12 (Lubahn et al., 1988) and consists of eight exons. The AR is a member of the steroid/nuclear receptor superfamily of ligand-activated transcription factors, and like the other members of this family, its structure includes an N-terminal transactivation domain (TAD, part of exon 1), a central DNA-binding domain (DBD, exons 2 and 3), and a C-terminal ligand-binding domain (LBD, exons 4 to part of exon 8). The binding of androgens to the LBD causes nuclear translocation of the ligand–receptor complex and triggers a series of molecular events culminating in the interaction of DBD with cognate response elements and activation of androgen-responsive genes (Quigley et al., 1995).

The receptor exhibits two polymorphic sites in exon 1, characterized by different numbers of CAG and GGC repeats resulting in variable lengths of polyglutamine and polyglycine stretches in the N-terminal TAD of the AR protein, that seem to modulate AR function. The number of CAG and GGC repeats ranges from about 10 to 35 (with a mean of 21–23) and 4 to 24 (with a mean of 16–17) respectively in normal men. Longer CAG repeat lengths result in reduced AR transcriptional activity both in vivo and in vitro (Chamberlain et al., 1994; Choong et al., 1996), and there is evidence that an inverse correlation between CAG number and androgenicity exists. Expansion to >40 of this repeat is related to Kennedy syndrome, a rare motoneuron disorder which is also associated with androgen insensitivity, testicular atrophy, reduced sperm production, and infertility (Brooks et al., 1995; Kazemi-Esfarjani et al., 1995). On the other hand, shorter AR polyglutamine tracts, and thus a more transcriptionally active AR, have been associated with increased prostate cancer risk (Giovannucci et al., 1997; Hakimi et al., 1997; Ingles et al., 1997; Stanford et al., 1997; Kantoff et al., 1998; Platz et al., 1999; Hsing et al., 2000), although this is still a controversial matter. Moderate expansion of the AR polyglutamine tract, while remaining within the normal polymorphic range, may also alter the AR function. In vitro experiments using AR constructs harbouring 15, 20 and 31 CAG repeats have shown an inverse relationship between the length of the repeat and the transactivation capacity (Tut et al., 1997). This is consistent with the finding that polymorphisms in CAG tract length correlate with sperm concentration in normal men (von Eckardstein et al., 2001). However, previous studies examining CAG repeat number in infertile men have reported conflicting results, with some (Giwerzman et al., 1998; Dadze et al., 2000; Sasagawa 2001; Van Golde et al., 2002; Rajpert-De Meyts et al., 2002; Lund et al., 2003) showing no expansion, and others (Tut et al., 1997; Dowsing et al., 1999; Yoshida et al., 1999; Mifsud et al., 2001; Patrizio et al., 2001; Wallerand et al., 2001; Casella et al., 2003; Menguai et al., 2003) reporting increased length (but still within the normal range) with respect to fertile control men. In particular, studies involving Singaporean, Australian, North American and Japanese subjects found...
an association between CAG length and male infertility, whereas this was not evident in studies from Europe. These discordant results may reflect the patient ethnicity selection, as the distribution of the number of CAG repeats is lowest in African-Americans, intermediate in whites, and highest in Asians. Furthermore, additional bias may be related to sample size restrictions, choice of the control group and inclusion patient criteria.

The functional consequences of variations in the GGC repeat are less clear, even if deletion of the polyglycine tract reduces AR transcriptional activity in transient transfection assay (Gao et al., 1996). Epidemiological investigations on the association between the number of GGC repeats and prostate cancer risk have produced inconsistent results. Even if short GGC repeat length seems to increase the risk of the disease (Irvine et al., 1995; Hakimi et al., 1997; Stanford et al., 1997; Chang et al., 2002; Chen et al., 2002), no clear conclusions could be made. Only two studies reported the distribution of GGC lengths among the infertile men and found no difference from that in the general population (Tut et al., 1997; Lundin et al., 2003). Moreover, the effect of combined CAG and GGC repeats in spermatogenesis is largely unknown, and a description of CAG and GGC distribution in Italian men has never been reported.

In this study we analysed CAG and GGC repeat lengths in a group of idiopathic infertile men compared to normal fertile subjects. These variables were analysed separately and jointly in order to investigate whether particular haplotypes may specifically be associated with male infertility and could have a role in regulating AR function.

Materials and methods

Subjects

Patients and controls were prospectively recruited for this study with the approval of the Hospital Ethical Committee and informed consent was obtained from each subject.

We recruited 163 men, which satisfied the following criteria: (i) infertility with a sperm count <20×10^6/ml, excluding obstructive azoospermia; (ii) absence of known causes of testicular damage. A complete medical history and a physical examination were undertaken. Semen analysis was repeated at least twice and was performed according to World Health Organization (1999) guidelines. When semen analysis repeatedly revealed azoospermia (absence of sperm) or oligoazooospermia with sperm concentrations <5×10^6/ml, a bilateral testicular fine needle aspiration cytology (FNAC) (Foresta et al., 1992) was performed. In all subjects we had ultrasound examination of the testes (for testis morphology and volume) and plasma determination of FSH, LH, prolactin, and testosterone concentrations. Exclusion of known causes of male infertility was done by careful history (excluding for example cryptorchidism, varicocele, orchiitis, or testicular trauma), sperm antibody determination, endocrine profile (excluding for example hypogonadotrophic hypogonadism, hyperprolactinemia), karyotype analysis, Y chromosome microdeletion analysis (Foresta et al., 1997; Ferlin et al., 2003), cystic fibrosis transmembrane regulator gene (CFTR) mutation. AR gene mutations were excluded by PCR and direct sequencing, using a set of 11 oligonucleotide primers covering exons 1–8 (Lubahn et al., 1989). None of the patients studied had clinical features, by history or physical examination, of androgen resistance or neurological symptoms.

Infertile men were divided into three subgroups according to the severity of the spermatogenic defect. Forty-five (27.6%) had azoospermia and a bilateral testicular cytological appearance of Sertoli cell-only syndrome (SCOS), 87 (53.4%) had severe oligoazooospermia (spem concentration <5×10^6/ml) and a testicular cytological appearance of severe hypospermatogenesis (SH), and 31 (19.0%) had moderate oligoazooospermia (MO, sperm concentration 5–20×10^6/ml). SCOS is defined as the complete absence in both testes of germ cells, whereas SH is defined as a strong reduction of germ cells, which are however in normal relative proportions, and therefore without maturation arrest (Foresta et al., 1992).

A total of 115 men with proven fertility and normozoospermia served as controls. All patients and controls were of Caucasian origin and came from different Italian regions, but 92 of 115 controls and 136 of 163 patients came from the North East of Italy.

Determination of the CAG and GGC repeat number

Genomic DNA was extracted from peripheral blood leukocytes using DNA isolation kit (Roche, Italy). The AR exon 1 was amplified from genomic DNA in two different PCR reactions, giving overlapping amplicons. Both reactions are performed under the same conditions (standard conditions with 8% dimethylsulphoxide) and with the same cycle (94°C for 1 min, 58°C for 1 min, 72°C for 1 min, repeated 37 times). The CAG repeat is contained in the amplicon produced with primers A0 GTGGTGTCCCGCAAGTTCC and A5 GCCACACTCTCCTCCCAAGGCAATT. It is sequenced with the primer A2 GCTTGAAGGTGCTGTTCCTC, using standard conditions for auto-mated sequencing. The GGC repeat is amplified with primers A3n CAGCAAGAGACTGCCCAAGC and A10 CCAGAACACAGAGTG ACTCTGCCC, and it is sequenced with primer A8 GGACTGGGATA-GGCACACTGCTCAACC. Primers A2, A5, A8 and A10 are from Lubahn et al. (1989), whereas we designed the new primers A0 and A3n. Sequence analyses were performed by using the gap4 software of the Staden package (Staden, 1996) available at the UK Human Genome Mapping Project webpage (http://www.hgmp.mrc.ac.uk/).

Statistical analysis

Differences in CAG and GGC mean repeat length were tested by the Wilcoxon’s rank sum test. Differences among frequencies were calculated with both χ²-test and Fisher’s exact test. Relative risks and the corresponding 95% confidence intervals were calculated on the basis of the asymptomatic normal distribution of these quantities (Agresti, 1990). Fisher’s exact test was used to analyse independence in two-way contingency tables. P <0.05 was considered statistically significant. Computations were performed by using Open-source statistical software “R”.

Results

Overall, the mean number of CAG and GGC repeats in exon 1 of the AR gene was 21.7 ± 2.8 (range 9–29) and 17.2 ± 1.9 (range 4–22) respectively in infertile men, and 21.6 ± 3.3 (range 9–31) and 17.0 ± 1.7 (range 8–21) in proven fertile control men (Table I). These differences were not statistically significant. There was also no difference when the median value of CAG and GGC numbers were compared. The subgrouping of infertile patients by category of spermatogenic disorders (SCOS, SH, MO) also showed no differences in terms of mean and median values of CAG and GGC repeat number with respect to controls (Table I). To explore the possibility that AR sensitivity could be determined by the total number of CAG and GGC repeats, we calculated the sum of triplets in each subject (CAG + GGC). This number was not different between controls (38.8 ± 3.4, range 27–45, median 39) and infertile men (39.0 ± 3.5, range 21–47, median 39) (data not shown).

Based on the median value of repeats in normal fertile men, we compared the distribution of men according to the length of CAG and GGC triplets, considering short repeats those ≤22 and ≥17 respectively, and long repeats those ≥23 and ≥18 respectively, and we observed no statistically significant difference. Also the distribution of men according to the combination of short/long CAG and GGC repeats did not show any difference between patients and controls. We also analysed whether infertile men may have longer CAG repeats by comparing the proportion of men with ≥23, ≥24, ≥25, ≥26 and ≥27 repeats. In this case we also found no differences between patients and controls (data not shown). The same analysis was performed for GGC repeats (≥18, ≥19 and ≥20) and no difference was found (data not shown).

Next we considered the joint distribution of CAG and GGC. For this analysis the data were collected in two-way tables (Tables II and III) reporting frequencies for each CAG/GGC haplotype. We considered

418
in a single category the CAG numbers ≤20 and those ≥24, corresponding to the first and third quartiles of the distribution for controls. In the GGC distribution, most of the observations belong to categories 17 and 18, and therefore we considered in a single category the GGC numbers ≤16 and those ≥19. The analysis of association between the two variables showed that the hypothesis of independence can be accepted for fertile controls ($P = 0.19$), whereas there is strong evidence ($P < 0.01$) that for infertile patients the two variables are not independent. This is true also in the different groups of infertile patients. A detailed analysis of the percentages of the two-way tables showed that the main differences between categories 17 and 18 are due to the low number of patients in each group. The calculated relative risk of these haplotypes with respect to the infertility condition is shown in Table IV, which also shows that the prevalence of the cells corresponding to CAG ≥21/GGC ≥18 is significantly more frequent in patients with SH with respect to controls (33.3 versus 20.9%, $P < 0.05$), but it does not reach statistical significance in infertile men considered as a whole or in infertile men with SCOS or MO.

Results obtained from sub-analysis including those subjects coming from North East of Italy only (92 controls and 136 patients) were identical to those obtained with the whole groups of subjects (data not shown). Plasma concentration of LH, testosterone and prolactin were not different between patients and controls, whereas FSH was higher in SCOS–SH patients ($17.2 ± 8.8$ versus $2.7 ± 1.2$ IU/l, $P < 0.01$).

### Discussion

Diagnostic advances have progressively decreased in recent years the percentage of men classified as having idiopathic infertility. In particular, accumulating evidence suggests that a genetic component may be responsible for about 15% of severe infertile male subjects (Foresta et al., 2002). Such genetic factors include for example chromosomal aberrations, Y chromosome long arm microdeletions involving one or more "azoospermia factors" (AZF) (Foresta et al., 2001), mutations in the CFTR gene, and mutations in the AR (Hiort et al., 2000). All these conditions are well known causes of male infertility, directly affecting spermatogenesis or vas deferens. The role of CAG triplets in exon 1 of AR in male infertility is less clear, and the combined effect of CAG and GGC repeat numbers is completely unknown.

In this study we analysed for the first time both CAG and GGC triplets in a large group of idiopathic severely oligozoospermic men and found that there is no difference with respect to normozoospermic fertile controls when these variables are analysed separately. However, we found significant differences when the joint distribution of CAG and GGC and the specific combinations of CAG and GGC were analysed. In particular, infertile men more frequently presented the combination CAG = 21/GGC = 18 and very rarely the combination CAG = 23/GGC = 16. In addition, patients with SH showed more frequently the combination CAG ≥21/GGC ≥18.

Our findings suggest several important considerations. Firstly, this is the first analysis of both CAG and GGC in Italy. Comparison with other studies where the method of direct sequencing has been used shows that CAG and GGC allele distribution in Italy agrees with findings in other Caucasian populations (Lumbroso et al., 1997; Correa-Cerro et al., 1999; Sasaki et al., 2003), even if subtle differences in the pattern of distribution (such as a higher frequency of 17 than 16 GGC) could suggest ethnic Italian differences. Second, there is no association between CAG length and male infertility, and

### Table I. CAG and GGC values in infertile men and controls

<table>
<thead>
<tr>
<th></th>
<th>CAG</th>
<th>GGC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of alleles</td>
<td>Range</td>
</tr>
<tr>
<td>Controls (n = 115)</td>
<td>17</td>
<td>9–31</td>
</tr>
<tr>
<td>Infertile men (n = 163)</td>
<td>15</td>
<td>9–29</td>
</tr>
<tr>
<td>SCOS (n = 45)</td>
<td>10</td>
<td>13–27</td>
</tr>
<tr>
<td>SH (n = 87)</td>
<td>12</td>
<td>9–27</td>
</tr>
<tr>
<td>MO (n = 31)</td>
<td>10</td>
<td>16–29</td>
</tr>
</tbody>
</table>

No statistically significant ($P > 0.05$) differences were observed. SCOS = Sertoli cell-only syndrome; SH = severe hypospermatogenesis; MO = moderate oligozoospermia.

### Table II. Joint distribution of CAG and GGC percentages (number in parenthesis) for the 115 fertile control men

<table>
<thead>
<tr>
<th></th>
<th>CAG</th>
<th>GGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤16</td>
<td>3.5 (4)</td>
<td>11.3 (13)</td>
</tr>
<tr>
<td>17</td>
<td>14.8 (17)</td>
<td>9.1 (17)</td>
</tr>
<tr>
<td>18</td>
<td>2.6 (3)</td>
<td>14.8 (17)</td>
</tr>
<tr>
<td>&gt;19</td>
<td>32.2 (37)</td>
<td>11.2 (13)</td>
</tr>
</tbody>
</table>

### Table III. Joint distribution of CAG and GGC percentages (number in parenthesis) for the 163 infertile men

<table>
<thead>
<tr>
<th></th>
<th>CAG</th>
<th>GGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤16</td>
<td>1.8 (3)</td>
<td>9.2 (15)</td>
</tr>
<tr>
<td>17</td>
<td>12.9 (21)</td>
<td>8.0 (13)</td>
</tr>
<tr>
<td>18</td>
<td>1.8 (3)</td>
<td>3.7 (6)</td>
</tr>
<tr>
<td>≥19</td>
<td>27.6 (45)</td>
<td>3.7 (6)</td>
</tr>
</tbody>
</table>

$P$ values with respect to that found in infertile patients (Table III) are reported in Table IV.
this is in accordance with other studies from Europe (Giwercman et al., 1998; Dadze et al., 2000; Rajpert-De Meyts et al., 2002; Van Golde et al., 2002; Lund et al., 2003). The association found in Singaporean, North American and Japanese subjects (Tut et al., 1997; Yoshida et al., 1999; Mifsud et al., 2001; Patrizio et al., 2001; Casella et al., 2003) therefore needs to be further confirmed to ascertain definitively that this difference is only due to the ethnicity. Confirming previous studies (Tut et al., 1997; Lundin et al., 2003), we also found that the GGC polyglutamine tract length is not related to male infertility, and therefore his role in vivo in determining the transcriptional activity of the AR remains to be verified.

The most important result of our study came from the joint analysis of the two alleles. Two CAG/GGC haplotypes seem to increase susceptibility to the disease, while one haplotype seems to confer a protective effect against infertility. However, although we found a difference between the infertile men and the controls regarding some combinations of CAG and GGC lengths, these could be chance findings and no firm conclusion regarding the biological importance of these combinations can be drawn. Nonetheless, our findings agree with a recent report on CAG and GGC length in oesophageal cancer (Dietzsch et al., 2003). The mechanism by which haplotypes CAG = 21/GGC = 18 and CAG >21/GGC >18 appear to decrease and the haplotype CAG >23/GGC >16 to increase susceptibility to spermatogenic impairment can only be speculative. Confirmation of the association between CAG/GGC polymorphisms and the clinical phenotype would require other techniques such as transmission disequilibrium testing. However, this method requires genotyping of the mothers of all the infertile men and therefore it cannot be easily performed. In vitro transactivation studies with AR constructed with different CAG/GGC combinations and expression analyses are under way to confirm these clinical data and understand the molecular mechanisms involved in the increase or inhibition of the AR transcriptional activity.

It is noteworthy that not only are CAG and GGC repeat lengths not significantly different in infertile men with respect to controls, but they are also not different among azoospermic, severely oligozoospermic and moderately oligozoospermic subjects. However, haplotype CAG >21/GGC >18 is significantly more frequent in men with a sperm concentration <5×10^6/ml. This particular finding needs to be confirmed, but it suggests that the combination of these particular triplets, although still in the normal range, might have a more profound effect on AR activity.

In conclusion, our study suggests that some combinations of CAG and GGC triplets might modulate AR function, and shows for the first time that, instead of the length of CAG and/or GGC alone, specific CAG/GGC haplotypes are associated with an increased risk of, or may protect against, male infertility.

References


Table IV. Analysis of distribution of haplotypes CAG = 21/GGC = 18, CAG >21/GGC >18, and CAG >23/GGC >16 in infertile men and controls

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Controls (n = 115)</th>
<th>Infertile men (n = 163)</th>
<th>SCOS (n = 45)</th>
<th>SH (n = 87)</th>
<th>MO (n = 31)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAG = 21/GGC = 18</td>
<td>%</td>
<td>RR (95% CI)</td>
<td>%</td>
<td>RR (95% CI)</td>
<td>%</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
<td>%</td>
<td></td>
<td>%</td>
</tr>
<tr>
<td>Controls (n = 115)</td>
<td>5.2</td>
<td></td>
<td>20.9</td>
<td></td>
<td>6.9</td>
</tr>
<tr>
<td>Infertile men (n = 163)</td>
<td>12.9*</td>
<td>2.47 (1.03–5.93)</td>
<td>31.3</td>
<td>NS</td>
<td>0.6**</td>
</tr>
<tr>
<td>SCOS (n = 45)</td>
<td>13.3</td>
<td>NS</td>
<td>31.1</td>
<td>NS</td>
<td>0</td>
</tr>
<tr>
<td>SH (n = 87)</td>
<td>12.6</td>
<td>NS</td>
<td>33.3*</td>
<td>1.6 (1.01–2.54)</td>
<td>1.1</td>
</tr>
<tr>
<td>MO (n = 31)</td>
<td>12.9</td>
<td>NS</td>
<td>25.8</td>
<td>NS</td>
<td>0</td>
</tr>
</tbody>
</table>

*P < 0.05; **P < 0.01.

RR = relative risks; and CI = confidence intervals. SCOS = Sertoli cell-only syndrome; SH = severe hypospermatogenesis; MO = moderate oligozoospermia; NS = not significant.
appropriate use of genetic tests in infertile couples. Eur J Hum Genet 10,303±312.


