The two most common alleles of the coding GGN repeat in the androgen receptor gene cause differences in protein function

Felix F Brockschmidt, Markus M Nöthen and Axel M Hillmer

Department of Genomics, Life and Brain Center, University of Bonn, Sigmund-Freud-Straße 25, 53127 Bonn, Germany

(Requests for offprints should be addressed to A M Hillmer; Email: axel.hillmer@uni-bonn.de)

Abstract

Genetic studies have reported association of a polyglycine-encoding GGN repeat in exon 1 of the androgen receptor (AR) gene with common human traits. The polyglycine tract is located in the transactivating domain of the AR protein, suggesting an effect of repeat length on receptor function. Here, we compare the functional characteristics of the two most common alleles (23 and 24 repeats) and two extreme alleles (10 and 27 repeats) in a reporter gene assay in HeLa cells. A correlation between the repeat length and AR activity was observed. This is attributable to both a higher protein concentration, determined by ELISA, and a higher per-protein activity of long repeat alleles. Interestingly, protein concentration does not correlate with transcript quantity, determined by real-time PCR assays, and no influence of repeat length on protein stability could be detected in translation inhibition assays. This may suggest that repeat length affects translation efficiency. In conclusion, our data provide evidence of functional differences between the two most common alleles of the AR GGN repeat, supporting its potential role in the development of human traits.

Journal of Molecular Endocrinology (2007) 39, 1–8

Introduction

The androgen receptor (AR) is a member of the nuclear receptor superfamily and mediates the effects of androgens (Mangelsdorf et al. 1995). Upon binding of testosterone or 5α-dihydrotestosterone (DHT), the conformation of AR alters and it binds as a homodimer to androgen response elements, executing its function as a transcription factor (Lee & Chang 2003).

Given its central function in androgen-related pathways, genetic variation in the AR gene has been tested for association with a variety of common human traits (Quigley et al. 1995, Edwards et al. 1999, Hsing et al. 2000, Avila et al. 2001, Ellis et al. 2001, Chang et al. 2002, Aschim et al. 2004, Ferlin et al. 2004, Ruhyel et al. 2004, Zeegers et al. 2004, Hillmer et al. 2005, Sasaki et al. 2005, Cox et al. 2006, Lundin et al. 2006). The most widely tested variants are two repeat sequences, CAG and GGN repeats, in the coding region of AR that encode variable lengths of polyglutamine (polyQ) and polyglycine (polyG) tracts in the N-terminal part of the protein. The effect of repeat lengths on receptor function has been investigated for both of these repeats. A large body of evidence suggests that a long CAG repeat results in reduced AR transactivity (Mhatre et al. 1993, Chamberlain et al. 1994, Kazemi-Esfarjani et al. 1995, Gao et al. 1996, Nakajima et al. 1996, Buchanan et al. 2004, Wang et al. 2004). Fewer studies have been performed for the GGN repeat. The GGN-encoded polyG tract is located in one of the two N-terminal transactivating domains of the AR protein. In accordance with this, functional data suggest an impact of the GGN repeat on transactivating activity of AR (Gao et al. 1996, Ding et al. 2005, Werner et al. 2006). While these studies have analyzed several repeat lengths, they did not compare the functional properties of the alleles of 23 and 24 glycine repeats which are the most common repeats found in the European population and which have been reported to confer opposite effects on the development of human traits (Ellis et al. 2001, Aschim et al. 2004, Ferlin et al. 2004, Hillmer et al. 2005). Recently, Lundin et al. (2007) reported the first functional comparison of AR with polyG23 and polyG24 using a prostate-specific antigen promoter-driven reporter gene in COS-1 cells. PolyG23 showed higher transactivating activity than polyG24 at various DHT and R1881 concentrations which was not due to higher protein concentrations.

In the present study, we analyzed the transactivating activity of AR containing the two most common repeat lengths (23 and 24 repeats) as well as two extreme repeat lengths 10 and 27 repeats. We used a GRE2-TATA promoter-driven reporter gene assay in HeLa cells as an experimental model. To further understand the cause of functional differences, we quantified AR protein and AR messenger and determined per-protein activity.
Results

To test for the transactivating activity of AR with different polyG lengths, AR expression plasmids encoding for 10, 23, 24, and 27 polyG repeats were cotransfected with an androgen responsive reporter plasmid in HeLa cells. Cells were incubated with and without DHT and assayed for reporter gene activity. Absence of reporter gene activity in the negative control (plasmid without AR) indicated that no endogenous AR expression interfered with the assay (Fig. 2A). This was supported by the real-time PCR analysis which detected no AR transcript in HeLa cells without exogenous AR (Fig. 3). All four AR constructs showed DHT-inducible transactivating activity (Fig. 2A). In accordance with two previous reports (Gao et al. 1996, Werner et al. 2006), longer polyG resulted in higher transactivating activity after incubation with DHT. Interestingly, the correlation appeared to be nonlinear. The polyG23 construct, which is 13 repeats longer than the polyG10 construct, resulted in a 1.5-fold higher activity than polyG10, whereas polyG24 activated the reporter gene 1.3-fold more than polyG23 but was only one repeat longer (Fig. 2A). PolyG27 showed a 1.3-fold stronger transactivating activity than polyG23. There was no clear correlation of protein of polyG10 versus polyG23 and polyG24 (Fig. 2B). The correlation was again nonlinear. The effect between polyG23 and polyG24 was greater than polyG10 versus polyG27 (Fig. 2B). The effect was less clear in the absence of DHT. AR protein levels of all repeat lengths were clearly lower in the absence of hormone. Quantifying AR activity and AR protein from the same samples made it possible to define the per-protein transactivation activity (Fig. 2C). Longer polyG repeats showed higher per-protein activity than shorter alleles. The data suggest a more linear correlation since larger differences of polyG lengths resulted in larger differences in per-protein activity (polyG10 versus polyG23) and small differences of repeat lengths resulted in a negligible difference in per-protein activity (polyG23 versus polyG24).

Is the higher protein quantity of AR with longer repeats due to higher transcript levels? To address this question, AR messenger of HeLa transfection experiments was quantified. There was no clear correlation between repeat length and AR transcript quantity in the presence of DHT (Fig. 3). However, polyG10 and polyG27 demonstrated lower and higher messenger concentrations respectively. Transcript levels were clearly higher in the absence of hormone than in the presence of hormone, whereby shorter repeats show lower AR transcript levels than longer repeats.

To test whether polyG specific differences in AR quantity are due to differences in protein stability, translation of transfected HeLa cells was inhibited by cycloheximide (CHX) and AR protein was quantified by ELISA at different time points. By inhibition of translation for 8 h, AR quantity declined, whereas ß-actin concentration did not change significantly (Fig. 4A). Although there was a significant effect of different AR polyG lengths on AR quantity after 2-5 h of translation inhibition, no consistent differences over three inhibition times were observed (Fig. 4B).

Discussion

GGN23 and GGN24 are the two most frequent alleles of the GGN repeat in whites (together 86%; Fig. 1; Lumbroso et al. 1997, Hillmer et al. 2005). They are of epidemiological interest considering the common disease–common variant hypothesis (Lander 1996). In associated studies, the two alleles showed opposite effects (Ellis et al. 2001, Aschim et al. 2004, Ferlin et al. 2004, Hillmer et al. 2005), indicating the importance of their functional comparison. Our study compares the functional properties of these frequently occurring alleles. In addition, we analyzed the two extreme alleles to explore the maximum of functional difference caused by GGN variability under the hypothesis of a correlation between repeat length and AR function. Comparing the transactivation capacity of polyG10, polyG23, polyG24, and polyG27 (encoded by the respective GGN alleles) showed that longer repeats have a stronger transactivating capacity than shorter repeats. This is in accordance with the observation of Gao et al. (1996) that AR with a deleted polyG had a reduced transactivating activity when compared with polyG23, as well as with the analysis of Werner et al. (2006) showing that transactivation of polyG0 < polyG10 < polyG16 < polyG25. In contrast, Ding et al. (2005) found an inverse correlation between the polyG length and transactivation capacity when analyzing polyG19–polyG23, and Lundin et al. (2007) reported polyG23 having a higher transactivation capacity than polyG10, polyG24, and polyG27. The different promoters of the reporter genes and different cell lines used in the studies might account for the discrepancies. This implies that the cell type-specific environment and promoters of AR responsive genes are of importance for the functional outcome of AR. Interestingly, the difference in activity between polyG23 and polyG24 in the present study was similar.
to the differences between polyG10 and polyG23, and polyG24 and polyG27 (Fig. 2A), although polyG23 and polyG24 are separated by only one repeat.

The quantification of the AR protein in the presence of hormone indicates that the higher transactivation activity of AR with long polyG is due in part to a higher AR protein level (Fig. 2B and C). This is particularly the case for the comparison of polyG23 and polyG24 with a negligible per-protein transactivating difference but a detectable difference in protein quantity (Fig. 2). This difference in protein quantity is neither the result of a higher AR messenger level (Fig. 3), nor is it attributable to a higher protein stability of AR polyG24 relative to AR polyG23 (Fig. 4B). In general, the quantity of AR protein of the four polyG variants is positively correlated with the polyG length in the presence of hormone (Fig. 2B). In addition to this effect, per-protein activity increases with the extension of polyG (Fig. 2C). The correlation between the quantity of AR messenger and AR protein is less clear however (Figs 2B and 3), which may suggest that the difference in protein quantity between the four polyG lengths is attributable to different translation efficiencies. Hormone binding clearly stabilizes AR protein since the presence of DHT leads to high protein levels at low RNA concentration (Figs 2B and 3). This effect has been described previously (Gao et al. 1996). In our transient transfection experiments, we observed a significantly higher quantity of messenger in the absence of hormone. This was unexpected since AR expression is driven by a simian virus 40 (SV40) promoter. However, a negative feedback loop in the androgen-signaling pathway might account for this phenomenon. Interestingly, in the absence of hormone, we observed a higher quantity of AR transcript with long GGN tract. This might be due to the decrease of free energy resulting from increasing GGN repeat length (Ding et al. 2005) which may result in an increase in structural order, thereby stabilizing the RNA.

In the present study, we have focused on the two most common alleles of the GGN repeat. However, less common GGN alleles might also contribute to disease susceptibility as suggested for other rare sequence variations contributing to complex traits (Pritchard & Cox 2002, Cohen et al. 2004). Our data imply that rare GGN alleles can have stronger effects on AR function. This may result in higher penetrance with regard to expression of disease.

Since the CAG repeat also affects AR function (Mhatre et al. 1993, Chamberlain et al. 1994, Kazemi-Esfarjani et al. 1995, Gao et al. 1996, Nakajima et al. 1996, Buchanan et al. 2004, Wang et al. 2004), the combination of certain CAG and GGN alleles might have additive effects as suggested by the detection of functional differences between five analyzed CAG/GGN combinations (Gao et al. 1996). In whites, there is no significant linkage disequilibrium between the two repeats (Kittles et al. 2001), resulting in a variety of haplotypes with potential differences in AR function. The variance of the allelic distribution across populations (Kittles et al. 2001) may contribute to population-specific differences in the susceptibility for associated traits.

As androgens mediate a wide range of developmental and physiological responses through AR, (Lee & Chang 2003), it is conceivable that variability in AR can have an impact on a set of human traits. It has been proposed that slight changes in AR function caused by different GGN alleles may predispose to the development of congenital abnormalities of male genital organs (Aschim et al. 2004). Further, certain CAG/GGN
haplotypes have been suggested to increase susceptibility to infertility (Ferlin et al. 2004), and an association between the GGN repeat and semen volume has been reported (Lundin et al. 2006). A decreased risk of breast cancer has been described for women with long GGN alleles (Suter et al. 2003), whereas other studies could not detect association (Dunning et al. 1999, Kadouri et al. 2001). Shorter GGN alleles have been associated with an increased risk of prostate cancer (Hsing et al. 2000, Chang et al. 2002), while longer alleles have been associated with endometrial cancer (Sasaki et al. 2005) which would be in accordance with the differing effects of androgens on the endometrium and the prostate (androgens exert an inhibitory effect on endometrial cell proliferation, whereas they have a mitogenic effect in the prostate). However, the association findings with prostate cancer remain controversial and no effect was shown in a large meta-analysis (Zeegers et al. 2004). For the combination of short CAG and GGN repeats and the GGN23 allele respectively, association with androgenetic alopecia has been reported (Ellis et al. 2001, Hillmer et al. 2005). Our finding of a functional difference between the two most common alleles of the GGN repeat supports a direct role of the repeat in the development of human traits. The unraveling of tissue-specific consequences of different GGN alleles should help to understand the pathophysiology of associated traits.

Materials and methods

Genotyping of GGN repeat

EDTA-anticoagulated venous blood samples were collected from 188 German male blood donors from the blood transfusion center of the University Hospital Bonn, and lymphocyte DNA was isolated by salting out with saturated NaCl solution (Miller et al. 1988). The study was approved by the ethics committee of the University of Bonn, and informed consent was obtained from all participants. Genotyping of the GGN repeat of AR was performed as described elsewhere (Hillmer et al. 2005).

Plasmids

AR expression plasmids: human AR GGN alleles of 10, 23, 24, and 27 repeats were amplified from genomic DNA by PCR using GGC-clone-F primer CAACCTTTCCACTGGCTCTGG and GGC-clone-R primer GGA-TAGGGCACTCTGCTCAC. PCR fragments were subcloned in pCR2.1 (Invitrogen) according to standard protocol and verified by sequencing. pSG5-AR expression plasmid was kindly provided by Dr John Isaacs (Cellular and Molecular Medicine, Baltimore, MD, USA). Subcloned GGN alleles of 10, 23, 24, and 27 repeats (GGN10: (GGT)3GGG(GGT)2GGC4, GGN23: (GGT)3GGG(GGT)2GGC8, GGN24: (GGT)3GGG(GGT)2GGC12, GGN27: (GGT)3GGG(GGT)2GGC16) were used in transfection assays.

Figure 2 GGN dependent AR transactivation capacity. (A–C) HeLa cells were transiently transfected with androgen responsive luciferase plasmid and AR expression plasmid with GGN repeat lengths as indicated and incubated with and without DHT, respectively. (A) Cell extracts were assayed for quantity of luciferase. Values were corrected by transfection control. (B) AR protein of cell extracts in (A) was quantified by ELISA. Values were corrected by transfection control. (C) Per-protein activity was obtained by division of AR transactivation activity by quantity of AR of the same samples. Data of three pooled experiments is shown. Values represent the mean ± S.D. Kruskal–Wallis analysis of variance: P < 0.05 was considered significant. n.s., not significant. RLU, relative luciferase units.
GGN23: (GGT)$_3$GGG(GGT)$_2$GGC$_{17}$, GGN24: (GGT)$_3$GGG(GGT)$_2$GGC$_{18}$, and GGN27: (GGT)$_3$GGG(GGT)$_3$GGC$_{20}$) were cloned in pSG5-AR by sequential restriction with CpoI and KpnI (Fermentas, St Leon-Rot, Germany) flanking the GGN repeat. Ligation was performed by the use of the Rapid DNA Ligation Kit (Roche Molecular Biochemicals) according to the manufacturer’s recommendations. Transfers were verified by sequencing and restriction with CpoI/KpnI. Restriction fragments were analyzed on a 3% agarose gel. pSG5-AR has 22 CAG repeats.

Androgen responsive firefly luciferase reporter plasmid pGRE2-TATA-Luc was kindly provided by Dr Rainer B Lanz (Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, USA). pGRE2-TATA-Luc contains the firefly luciferase gene and two glucocorticoid (and androgen) responsive elements. This reporter gene construct has been shown to be more androgen responsive than the commonly used mouse mammary tumor virus (MMTV)-Luc reporter system containing another glucocorticoid-responsive elements-harboring promoter (Alléra et al. 1998).

Renilla luciferase plasmid phRL (Promega GmbH) was used for transfection efficiency control, which contains the cytomegalovirus (CMV) immediate-early promoter/enhancer region for strong and constitutive expression of the R. luciferase.

**Tissue culture**

HeLa cells for real-time PCRs, luciferase experiments, and ELISA were cultured in Dulbecco’s modified Eagle’s medium (DMEM; CAMBREX Bio Science, Verviers, Belgium) with 10% fetal calf serum (Biochrom AG, Berlin, Germany), 1% penicillin/streptomycin (Invitrogen), and 1× Amphotericin B (Roche Molecular Biochemicals) at 37°C and 5% CO$_2$. A total of 210 000 cells per well were seeded in 12-well culture plates 1 day before transfection. On the day of transfection, cells were 80–90% confluent and the medium was replaced by serum-free OptiMEM (Invitrogen). Transfections per well were carried out using 500 ng pSG5-AR, 500 ng pGRE2-TATA-Luc, 12.5 ng phRL, and 3 μl Lipofectamine2000 (Invitrogen) in OptiMEM according to standard protocol. After 6 h of incubation, the transfection mixture was replaced by DMEM with 10% steroid hormone-stripped fetal calf serum with 1 nM DHT and carrier (ethanol) respectively. Cells were harvested after an additional 21 h of incubation. For each experiment, two transfections of each AR plasmid were incubated with carrier and three transfections were incubated with DHT. For AR stability assays, 250 000 cells were seeded in each well of a 12-well plate. Cells of each well were transected the following day with 500 ng pSG5-AR and incubated in
medium containing 1 nM DHT 6 h after transfection start. Twenty-four hours after transfection start, medium was replaced by medium containing 1 nM DHT and 100 ng/μl CHX or carrier (H2O) and incubated for 0, 2.5, 5, and 8 h. For every experiment, each AR construct incubated with and without CHX was assayed in triplicate and duplicate respectively. At least three experiments were performed for both protein- and RNA-based assays. Cells were washed with PBS and harvested in 250 μl passive lysis buffer (Promega) for protein-based assays and in 300 μl nucleic acid purification lysis solution (Applied Biosystems, Darmstadt, Germany) for real-time PCR.

**Luciferase assay**

Dual-luciferase Reporter Assay System (Promega) was used to quantify AR-driven reporter gene activity (firefly luciferase) and transfection control (R. luciferase) using 20 μl cell lysate, 50 μl luciferase assay reagent II, and 50 μl
Stop&Glo reagent. Luciferase activities were quantified in 96-well plates in a GeniosPro plate reader (TECAN, Grailsheim, Germany). Firefly luciferase activities were corrected for variation in transfection efficiency by the respective R. luciferase activities. Relative luciferase units of polyG23 + DHT were set as 100%.

**ELISA**

Fifty microliters of transfected HeLa cell extracts (as used for luciferase assays) per well were assayed by ELISA using the NR Sandwich AR kit (Active Motif, Rixensart, Belgium) according to the manufacturer’s recommendations. Colorimetric quantification was performed in a DYNEX MRX II microplate photometer (DYNEX Technologies, Berlin, Germany) in 96-well plates. For AR quantification in activity assays, optical densities were normalized by R. luciferase luminescence and polyG23 + DHT was set as 100%. For AR quantification in stability assays, optical densities of each AR polyG allele were normalized to their respective optical densities at $t=0$. Each AR polyG allele at $t=0$ was set as 100%.

**Real-time PCR**

Total RNA was extracted from transfected HeLa cell samples in nucleic acid purification lysis solution (Applied Biosystems) through the use of an ABI PRISM 6100 Nucleic Acid PrepStation (Applied Biosystems). The quality and quantity of the RNA was verified on a NanoDrop ND-1000 spectrophotometer (Peqlab Biotechnologie, Erlangen, Germany). RT-PCR was performed with the SuperScriptIII First-Strand Synthesis SuperMix (Invitrogen) with oligo dT primers. Relative quantifications in real-time experiments were performed in 384-well plates with 100 ng cDNA per well in a total volume of 20 μl on the ABI Prism 7900HT Fast Real-Time PCR System (Applied Biosystems) using the TaqMan Gene Expression Assay (Applied Biosystems) for AR (Hs00171172_m1) and the pre-developed TaqMan Endogenous Controls Human Cyc (Cyclophilin, 4326316E) and Human B2M (β-2-microglobulin, 4326319E; Applied Biosystems). Each sample was assayed in triplicate, and normalized by the house-keeping genes cyclophilin and B2M, and CT values were transformed by $100 000 000/2^{CT}$. Values were calculated relative to polyG23 + DHT which was set as 100%.

**SDS–PAGE and western blot**

SDS–PAGE was performed according to standard protocol (Sambrook & Russell 2001). In brief, 40 μl pooled cell extract (triplicates and duplicates were pooled) were mixed with 20 μl 3X sample buffer. The volume of 40 μl were loaded on 10% polyacrylamide gels and separated by electrophoresis. Proteins were transferred on polyvinylidifluoride membranes by the semi-dry blotter system (Biometra, Göttingen, Germany). Membranes were stained with ponceau for 2 min and blocked with 5% nonfat milk powder in PBST (phosphate buffered saline containing 0.2% Tween) for 1 h. Rabbit anti-AR antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) and mouse anti-β-actin antibody (Sigma–Aldrich) were used for incubation in 1:500 and 1:15 000 dilutions in 5% nonfat milk PBST for 1 h. Membranes were washed thrice for 5 min with PBST and were incubated with 1:30 000 horseradish peroxidase-conjugated anti-rabbit and anti-mouse antibodies (Jackson ImmunoResearch Laboratories, Suffolk, UK) respectively in 5% nonfat milk PBST for 1 h. Membranes were washed thrice for 20 min in PBST and assayed for peroxidase activity using SuperSignal West Dura (Pierce) and autoradiography.

**Statistical analysis**

The statistical analysis consisted of the non-parametric Kruskal–Wallis ANOVA on ranks. $P$ values were judged significant at $P<0.05$.

**Acknowledgements**

We thank Dr R B Lanz for plasmid GRE2-TATA-Luc and Dr J Isaacs for plasmid SG5-AR. This work was supported by the European Research Grant for Research on Alopecia funded by Pfizer Consumer Healthcare. Conflict of interest statement: no conflicts declared.

**References**


Zeegers MP, Kiemeney LA, Nieder AM & Ostrer H 2004 How strong is the association between CAG and GGN repeat length polymorphisms in the androgen receptor gene and prostate cancer risk? *Cancer Epidemiology, Biomarkers and Prevention* **13** 1765–1771.

Received in final form 18 April 2007

Accepted 18 April 2007

Made available online as an Accepted Preprint 2 May 2007