Structural and functional consequences of glutamine tract variation in the androgen receptor

Grant Buchanan1, Miao Yang1, Albert Cheong1, Jonathan M. Harris2, Ryan A. Irvine3, Paul F. Lambert1, Nicole L. Moore1, Michael Raynor1, Petra J. Neufing1,§, Gerhard A. Coetzee3 and Wayne D. Tilley1,*

1Dame Roma Mitchell Cancer Research Laboratories, Adelaide University/Hanson Institute, Adelaide, SA, Australia, 2Centre for Molecular Biotechnology, School of Life Sciences, Queensland University of Technology, Brisbane, QLD, Australia and 3Departments of Molecular Microbiology and Immunology, Urology and Preventative Medicine, USC/Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Received February 6, 2004; Revised April 26, 2004; Accepted June 1, 2004

The androgen receptor (AR) gene contains a polymorphic trinucleotide repeat region, (CAG)n, in its N-terminal transactivation domain (NTD) that encodes a polyglutamine (polyQ) tract in the receptor protein. Whereas the length of the CAG repeat ranges from 6 to 39 in healthy individuals, the variations in repeat length both within and outside the normal range are associated with disease, including impaired spermatogenesis and Kennedy’s disease, and with the risk of developing breast and prostate cancer. Whereas it has been proposed that the inverse relationship between polyQ tract length within the normal range and AR transactivation potential may be responsible for altered risk of disease, the molecular mechanisms underlying polyQ length modulation of AR function have not been elucidated. In this study, we provide detailed characterization of a somatic AR gene mutation detected in a human prostate tumor that results in interruption of the polyQ tract by two non-consecutive leucine residues (AR-polyQ2L). Compared with wtAR, AR-polyQ2L exhibits disrupted inter-domain communication (N/C interaction) and a lower protein level, but paradoxically has markedly increased transactivation activity. Molecular modeling and the response to cofactors indicate that the increased activity of AR-polyQ2L results from the presentation of a more stable platform for the recruitment of accessory proteins than wild-type AR. Analysis of the relationship between polyQ tract length and AR function revealed a critical size (Q16–Q29) for maintenance of N/C interaction. That between 91 and 99% of AR alleles in different racial-ethnic groups encode a polyQ tract in the range of Q16–Q29 suggests that N/C interaction has been preserved as an essential component of androgen-induced AR signaling.

INTRODUCTION

Trinucleotide repeat sequences occur throughout the human genome, where they expand and contract during replication in a dynamic process that gives rise to increased or decreased length (length polymorphisms) in subsequent generations (1,2). Whereas the majority of trinucleotide repeats are located in non-coding sequences of the genome and have no discernable effects, those located in coding regions can influence gene expression, modulate the structure and stability of human molecular genetics, vol. 13, no. 16 # oxford university press 2004; all rights reserved
intermediate mRNAs, and can also function the content of the result-
tant proteins (3). The longer the trinucleotide repeat the
more predisposed it is to expansion, a phenomenon known
as anticipation. Expansion of trinucleotide repeats beyond
a critical number in a given coding region almost invariably
results in disease (4–8). The most common trinucleotide
repeat sequence found in human genes is (CAG)n, which
predominantly gives rise to polyglutamine (polyQ) tracts rather
than polyserine or polyalanine when located in the coding
sequence (9). There are at least 10 hereditary neurological
diseases related to expansion of a polyQ region in a causative
protein, including Huntington’s disease, dentatorubral-
pallidoluysian atrophy, spinocerebellar ataxia types 1, 2, 3,
6, 7 and 17 and spinal and bulbar muscular atrophy/Kenny’s
disease (4,5,10,11).

The human androgen receptor (AR) gene contains a poly-
orphic CAG repeat within exon 1 that encodes an uninter-
rupted polyQ tract in the N-terminal transactivation domain
(NTD) of the receptor (Fig. 1A). The normal distribution
of the AR-CAG microsatellite has been reported as 6–39
repeats, with a median of 19–20 in African American, 21–22
in White Caucasian, 22–23 in Asian and 23 in Hispanic popu-
lations (12–19). Expansion of the AR-CAG microsatellite to 40
or more repeats causes a rare, X-linked, adult onset, neurode-
generative disorder known as spinal and bulbar muscular
atrophy/Kenny’s disease (20,21). In addition to progressive
muscle weakness and atrophy due to loss of brain stem and
spinal cord motor neurons, men with this disorder frequently
present with symptoms of partial androgen insensitivity (i.e.
gynecomastia and testicular atrophy), indicative of aberrant
AR function (22,23). Conversely, several studies have dem-
strated that AR-CAG repeat length is inversely correlated with
the risk of developing prostate cancer, age of onset and/or
increased risk of advanced disease at diagnosis (15,18,24,25,
reviewed in 26). In addition, AR-CAG repeat length has been
inherited in the inherited form of androgen insensitivity
(AIS), hereditary hearing impairment, schizophrenia, benign
prostatic hyperplasia and the risk of developing breast and
endometrial cancers (27–33).

**RESULTS**

AR-polyQ2L has increased activity but lower protein level compared with wtAR

The AR-polyQ2L variant was reconstructed in an AR expression vector and assayed for transactivation function by transient transfection of the AR negative human prostate cancer cell line, PC-3. Compared with wtAR, AR-polyQ2L

![Figure 1](image-url)
exhibited a ~2-fold greater activity on the androgen responsive probasin promoter (ARR3-tk-luc) over a range of DHT concentrations (0.01–10 nM) (Fig. 1B). Addition of a 1000-fold excess of the AR antagonist, hydroxyflutamide, blocked the DHT-induced activities of wtAR and AR-polyQ2L with similar efficiency (data not shown). Immunoblot analysis of PC-3 cells transfected under similar conditions used for the transactivation studies consistently demonstrated a ~40–50% lower level of AR-polyQ2L steady-state protein compared with wtAR (Fig. 1C). When AR transactivation activity was corrected for this difference in protein level, AR-polyQ2L was up to 4-fold more active than wtAR (Fig. 1D). A similar augmented transactivation capacity was observed with the androgen responsive prostate specific antigen (PSA) luciferase reporter (PSA540-luc; Fig. 1E). In this latter experiment, immunoblot analysis was performed directly on lysates used for analysis of transactivation function, confirming a lower steady-state level of AR-polyQ2L protein in transfected cells. Independent analysis using the androgen responsive mouse mammary tumor virus chloramphenicol acetyl transferase reporter construct (MMTV-CAT) confirmed both the increased activity and the lower protein levels of the AR-polyQ2L variant compared with wtAR (Fig. 1F and G).

Transfection of increasing amounts of wtAR expression vector with either the ARR3-tk-luc or the PSA540-luc reporter constructs resulted in a near linear increase in AR transactivation activity (Fig. 2A) and protein (Fig. 2B). In addition, the affinity of AR-polyQ2L for DHT (0.29 ± 0.07 nM) was not significantly different ($P > 0.05$; *t*-test) from wtAR (0.23 ± 0.04 nM) in transiently transfected COS-1 cells. Together, these data indicate that the higher activity of AR-polyQ2L compared with wtAR cannot be attributed to: (i) non-specific inhibition of wtAR activity (squelching) due to higher receptor levels compared with AR-polyQ2L; (ii) limiting quantities of the reporter construct; or (iii) altered affinity of AR-polyQ2L for DHT.

Steady-state AR RNA and protein levels

No significant difference in the relative level of wtAR or AR-polyQ2L RNA determined by quantitative real-time polymerase chain reaction (PCR) analysis was observed at either 4 or 24 h after transfection of PC-3 cells (Fig. 3A). This observation is consistent with previous studies demonstrating that AR CAG repeat length within the normal range does not affect AR expression or mRNA stability (25,38–41,47). In contrast, immunoblot analysis performed on lysates of the same transiently transfected cells used for quantitative PCR analysis demonstrated significantly less (~40%) AR-polyQ2L protein than wtAR 24 h post-transfection ($P = 0.041$; Fig. 3B). These findings suggest that the decreased level of AR-polyQ2L protein compared with wtAR is most likely due to reduced AR protein stability or turnover rather than altered transcription rates or RNA stability.

AR-polyQ2L exhibits reduced N/C interaction

Conformational changes induced by agonist binding to the AR facilitate a strong inter- or intra-molecular interaction between the ligand-dependent AF-2 surface in the ligand-binding domain (LBD) and an $\alpha$-helix forming peptide, $^{23}$FQNLF$^{27}$ in the NTD (48). This is referred to as the N/C interaction (49,50). Numerous AR coregulators, which contain similar LxxLL-like $\alpha$-helix peptides, compete with the $^{23}$FQNLF$^{27}$ peptide for the AF-2 surface following ligand binding (51,52). The N/C interaction is dependent on agonist ligand binding to the receptor LBD, and has been suggested to be a necessary requirement for AR transactivation in vivo (53). As the polyQ region is located only 21 residues C-terminal to the $^{23}$FQNLF$^{27}$ peptide, we sought to determine whether
The AR-polyQ2L variant is distinct from other AR variants with N/C defects, in that it has markedly enhanced transactivation function compared with wtAR (Fig. 1). As it has recently been demonstrated that structural order within the AR-NTD is correlated with recruitment of the transcription machinery and receptor activity (57,58), we sought to assess whether the polyQ2L mutation alters AR-NTD structure. Secondary structure prediction algorithms (59) suggest that the disruption of the polyQ tract by two leucine residues results in an extended alpha-helix along the repeat sequence (Fig. 5A). In a more rigorous approach, we utilized simulated annealing and molecular modeling to generate theoretical structures of a pure polyQ tract of 23 residues and of an equal length glutamine tract containing the dual leucine interruption (Fig. 5B). These structures indicate that the polyQ2L peptide adopts a markedly different structure compared with a wild-type polyQ tract. The calculated energy for the polyQ2L peptide (−822.7 kcal mol$^{-1}$) was considerably lower than that of the pure polyQ peptide (−626.5 kcal mol$^{-1}$), indicating that insertion of the leucine residues also has a substantial impact on the flexibility of the AR-polyQ tract, and may result in an increase in structural order.

**Coactivation of AR-polyQ2L by GRIP1 and RAN**

Although the interaction between the AR and p160 coactivators is mediated by a region of the AR-NTD that does not include the polyQ tract (60), the transactivation response of the AR to the p160 proteins is inversely related to the length of the glutamine repeat (42). In this study, we did not detect any difference in the ability of wtAR and AR-polyQ2L to recruit p160 proteins (data not shown), but coexpression of the p160 protein, AIB1, resulted in a greater increase in activity of AR-polyQ2L compared with wtAR (Fig. 5C). Coexpression of RAN, an AR coregulator whose interaction with the AR-NTD is altered by polyQ tract length (61), also resulted in a greater increase in activity of AR-polyQ2L compared with wtAR (Fig. 5D). These results suggest that the structure of the polyQ tract, in addition to length, plays a key role in the recruitment and/or response to exogenous coregulators, and that the increased transcriptional activity of AR-polyQ2L derives from an altered interplay between components of the transactivation complex recruitment to the NTD compared with wtAR.
Although it is well accepted that AR polyQ length is inversely related to the transactivation capacity of the receptor (25,39,42,47,61–63), only one other group has compared more than three polyQ alleles within the accepted normal range of 6–39 repeats (64). In that study, only a modest effect of polyQ length was observed. Here, we conclusively demonstrate using multiple transfection experiments a significant inverse relationship between transactivation and polyQ length within the normal range ($P < 0.001$; Fig. 6A). There was on average a 1.7% decrease in activity for each additional
glutamine repeat. Significantly, however, the linear relationship did not apply to an AR with nine glutamine residues, which had a transactivation activity lower than an AR with 16 repeats ($P = 0.05$; Fig. 6A). Alterations in polyQ length within the range Q9–Q35 did not significantly affect the steady-state level of AR protein in transfected cells (Fig. 6B).

To better define how the polyQ tract influences AR function, we created a series of constructs expressing the AR-NTD with different polyQ lengths as either a fusion with the DNA-binding domain (DBD) of GAL4 or the transactivation domain of VP16. The length of the polyQ tract did not significantly alter the steady-state level of GAL4-AR-NTD (Fig. 6C) or VP16-AR-NTD (Fig. 6E) fusion proteins. In contrast to full-length AR, the size of the polyQ did not affect transactivation capacity of the AR-NTD ($P = 0.17$; Fig. 6D), or interaction of the NTD with the p160 coactivator, GRIP1 ($P = 0.59$; Fig. 6F). Whereas there was no change in N/C interaction for AR-NTD with glutamine residues in the range Q16–Q29, a polyQ length of 9 or 35 residues resulted in a significant decrease in the capacity of the AR to undergo N/C interaction ($P = 0.05$; Fig. 6G). Together, the above results suggest that polyQ length alters AR transactivation function directed through AF-1/AF-2, and that short and long repeats can disrupt the N/C interaction. These observations are consistent with analysis of the AR-polyQ2L mutation where disruption of the repeat did not affect AF-5 mediated transactivation of the AR or the NTD interaction with p160 coactivators, but markedly reduced N/C interaction (Figs 4 and 5).
Figure 6. Effect of polyQ length on AR transactivation, and on p160 and N/C interactions. (A) Transactivation capacity of wtAR with different polyQ tract lengths. PC-3 cells in 96-well plates (10,000 cells/well) were transfected with 0.5 ng of AR expression plasmid containing different CAG repeat lengths (pCDNA-AR) and 100 ng of the MMTV-luc reporter gene. Following transfection, cells were incubated for 30 h in the presence of 1 nM DHT as indicated. Results represent the amalgamated data (+SEM) of six independent experiments each containing seven individually transfected wells (i.e., 42 datapoints). The decrease in activity for increasing polyQ tract length was highly significant \( P < 0.001 \), analysis of variance (ANOVA). (B) Immunoblot analysis for AR on lysates of PC-3 cells transfected in 6-well dishes with pCDNA-AR expression clones containing different CAG repeat lengths as indicated. (C) Immunoblot analysis for AR and β-actin on lysates of COS-1 cells transfected in 6-well dishes with pM-AR expression clones containing different CAG repeat lengths. (D) Transactivation capacity of the AR-NTD with different polyQ tract lengths determined in COS-1 cells transfected with pM-AR expression clones and the pGK1 reporter corrected for protein level of each AR construct. There was no significant trend in AR-NTD activity with increasing polyQ tract length \( P = 0.17 \), ANOVA). (E) Immunoblot analysis for AR and β-actin on lysates of COS-1 cells transfected in 6-well dishes with pVP16-AR expression clones containing different CAG repeat lengths. (F) Mammalian two-hybrid analysis for interaction of AR containing different polyQ lengths and the p160 coactivator, GRIP1, was performed as detailed in Figure 6. Data are shown as RLU and represents the mean (+SEM) of six independently transfected wells corrected for protein level of each AR construct. There was no significant trend in AR-NTD/p160 interaction with increasing polyQ tract length \( P = 0.59 \), ANOVA). (G) Mammalian two-hybrid analysis for N/C interaction with AR containing different polyQ lengths performed as detailed in Figure 6. Data are shown as RLU and represents the mean (+SEM) of six independently transfected wells corrected for protein level of each AR construct. Whereas there was a highly significant effect of polyQ length on N/C interaction for the series \( P < 0.001 \), ANOVA), a significant change in N/C interaction was only observed between 9–16 and 29–35 repeats \( * P < 0.05 \), t-test).
The role of N/C interaction

To further analyze the role of N/C interaction in AR function, we utilized two previously characterized AR LBD mutations; the T875A variant identified from the LNCaP prostate cancer cell line that results in promiscuous activation of the receptor by non-androgenic ligands (65), and the E895Q substitution in the core of the AF-2 surface, previously reported to inhibit the N/C interaction without altering ligand binding characteristics of the receptor (66). Whereas DHT effectively induced an N/C interaction with the LBD of the wtAR, DHT, progesterone and hydroxyflutamide all induced a strong N/C interaction with the LNCaP AR variant LBD (Fig. 7A). These results parallel the agonist ability of these ligands on wild-type and LNCaP ARs as previously reported (53). In contrast, the E895Q mutation eliminated the capacity of the AR to form an N/C interaction in the presence of all ligands examined (Fig. 7A). However, transactivation analysis clearly demonstrated that the E895Q AR variant retains up to 60% of wtAR activity in response to DHT on both ARR3-tk and PSA540 reporters (Fig. 7B). These characteristics have been previously demonstrated for AR AF-2 core sequence mutations (50,67), suggesting that N/C interaction is not an absolute requirement of AR transactivation activity per se. Immunoblot analysis demonstrated a marked reduction in AR-E895Q protein level compared with wtAR (Fig. 7C), consistent with: (i) a role of N/C in preventing receptor degradation (49,56); and (ii) the concomitant decrease in N/C interaction and steady-state protein levels observed for the polyQ2L AR variant.

AR-polyQ length distribution

In order to better understand the implications of the above results, we re-examined available AR polyQ length distribution data from control populations in the published literature (14,24,68,69) (Fig. 8A). As noted by Edwards et al. in 1992, but little appreciated in the vast majority of subsequent studies, polyQ length is not distributed in a Gaussian manner, but exhibits a more complex pattern (14) (Fig. 8A). Whereas this is most apparent for African Americans, distinct peaks and troughs are observed in allele distributions for all ethnic groups in each of the individual reports from which this data were compiled. We then assessed polyQ length distribution in relation to maintenance of maximal N/C interaction determined above (i.e. 16–29 repeats). Across the four ethnic groups examined, 91–99% of all polyQ alleles fall within the range Q16–Q29 (Fig. 8B).

DISCUSSION

The AR-LBD is highly ordered, consisting of 11 alpha helices and four beta sheets arranged in a highly conserved tertiary structure (70,71). Conformational changes in the LBD brought about by high-affinity ligand binding result in the formation of the conserved hydrophobic AF-2 surface (70), which is the docking site contended by AR coregulators and the $^{23}$FQNLF$^{27}$ peptide in the AR-NTD (66,72). Chemical analysis has revealed that the AR-NTD has only a small number of predicted secondary structure elements, and adopts a disordered but flexible structure typical of transactivation domains (57,58). Nonetheless, key structural elements in the AR-NTD, including the $^{23}$FQNLF$^{27}$ peptide (72) and the core sequences of AF-1a and AF-1b (73) (Fig. 1), have been implicated in AR function, and appear to be critical for interaction of the NTD with coregulators and members of the basal transcription machinery (57,58). In this study, enhanced transactivation capacity of AR-polyQ2L in response to ligand and exogenous coregulators was associated with a predicted increase in secondary structure compared with wtAR, providing additional evidence for a relationship between structural order of the AR-NTD and transcriptional competence. Critically, our data support a model whereby each additional glutamine residue in the wtAR polyQ tract causes a reduction in the overall NTD order, thereby resulting in a reduced ability to recruit coregulators and components of the transcription machinery. This provides a mechanism to explain how polyQ length within the normal range is inversely related to AR activity and the response to coregulators. Moreover, the increased transcriptional activity of AR-polyQ2L is likely to be the result of a more effective recruitment of the transcription machinery due to increased structural order in the NTD compared with wtAR.

The reduction in N/C interaction and steady-state AR protein levels for AR-polyQ2L and E895Q AR variants compared with wtAR is analogous to observations for AR mutations identified in AIS and derived in vitro in the $^{23}$FQNLF$^{27}$ peptide (49,72,74). However, AR-polyQ2L is distinct from those mutations in that it retains both the structure of the $^{23}$FQNLF$^{27}$ peptide and the integrity of the LBD. Interactions of coregulators with the intact AF-2 surface of AR-polyQ2L can therefore occur with less competition from the N/C interaction compared with wtAR (51), and in the absence of NTD mutations that destabilize structural elements. This provides an additional mechanism, which may act in concert with effects associated with increased NTD structural order discussed above, to enhance the transactivation capacity of AR-polyQ2L compared with wtAR. In contrast to the AR-polyQ2L and AR-E895Q variants, a reduction in N/C interaction for ARs with short (≤16) or long (≥35) glutamine repeats was not associated with lower AR protein levels. This finding suggests that the effect of polyQ2L and E895Q mutations on AR protein level are distinct from changes that result in a reduced N/C interaction. Like AIS and $^{23}$FQNLF$^{27}$ mutations discussed above, these variants may exert a more dramatic effect on AR structure and function, or alter the recruitment of accessory factors that promote or inhibit receptor degradation compared with wtAR having different polyQ tract lengths.

Whereas only a small number of studies have examined the AR-NTD for mutations in prostate cancer, we recently reported that they occur with higher frequency following androgen ablation, and predominantly collocate to either the polyQ tract or to a small region of the transactivation function, AF-5 (26,75,76). However, in contrast to missense mutations in the LBD that result in increased promiscuity for activation by androgens and non-classical ligands (reviewed in 26,77–79), the functional consequence of the majority of mutations in the NTD is unknown. In this study, we have utilized the AR-polyQ2L variant to demonstrate a new mechanism...
whereby missense mutations identified in the AR-NTD in prostate cancer can exhibit a gain of function phenotype, thereby potentially contributing to disease progression.

Whereas the results of this and previous studies demonstrate that the N/C interaction is not essential for activity of the AR per se, it is highly indicative of a normal response of the receptor to physiological androgens and acts to stabilize the receptor following ligand binding (74). In this study, the observation that N/C interaction is only sustained for polyQ lengths of between 16 and 29 residues, and can be disrupted within this range by mutation, provides the first clear evidence that the polyQ tract plays a mandatory role in ensuring proper androgen-induced function of the human AR. Maximizing N/C interaction would serve to exclude coregulators from the promiscuous AR AF-2 surface (51), and therefore could be an important means of maintaining the specificity of AR function. That 91–99% of polyQ alleles across different ethnic groups are between 16 and 29 residues, a range that

Figure 7. The effect of LBD mutations on AR transactivation capacity and N/C interaction. (A) AR N/C interaction for the LBD AR variants E895Q and T875A (LNCaP AR variant) was compared with the wild-type AR LBD using the mammalian two-hybrid system as detailed in Figure 6. Data are shown as RLU and represent the mean (+ SEM) of four independently transfected wells. (B) Transactivation capacity of wtAR and the AR-E895Q variant on two promoters. PC-3 cells in 96-well plates (10,000 cells/well) were transfected with 0.5 ng AR expression plasmid (pCDNA-AR) and 100 ng of either the AR responsive probasin (ARR3-tk-luc) or PSA (pGL3-PSA540-luc) reporter. Following transfection, cells were incubated for 30 h in the presence or absence of 1 nM DHT as indicated. Results represent the mean (+ SEM) of six individually transfected wells. (C) Duplicate [(i) and (ii)] immunoblot analysis of wtAR and the E895Q AR variant in transfected PC-3 cells.
would maintain maximum N/C interaction, strongly supports this hypothesis. As the effect of polyQ repeat on transactivation is cell specific, presumably due to distinct profiles of coregulator proteins (25), determining in which cell types the relationship between N/C interaction and polyQ length is maintained will provide important additional information.

Two important lines of evidence support the hypothesis that polyQ tracts evolved as regulatory elements involved in fine-tuning the function of transcription factors (3,80–82): (i) the positional context of the polyQ tract in transcription factors is a major determinant of transcriptional competence (9) and (ii) the length of the polyQ tract in some cellular regulators (e.g. CBP) is absolutely conserved, implying critical reasons for maintenance of a particular repeat length in some proteins (83). For the AR, the distribution of polyQ alleles in all ethnic populations examined was positively skewed (i.e. towards longer repeats), which may reflect the bias of CAG repeats to expansion, a greater tolerance for longer alleles in maintaining AR function, or repeat length constraints (2,84). On the basis of the current study, we propose that polyQ tracts could regulate protein function by serving as flexible spacers to separate regions of biological activity. In the context of the AR, maintenance of a polyQ of 16–29 residues would allow sufficient flexibility for the upstream FQNLF motif.
to fold into the LBD (N/C interaction) while maintaining the capacity of the proximally located activation function (AF-1) to interact with coregulators and the transcription machinery. This provides a mechanism to explain how both increased and decreased polyQ allele length, and altered structure as observed for the polyQ2L mutation, can influence N/C interaction and receptor function. We further propose that the moderate reduction in AR activity with increasing length of the polyQ tract within the normal range is a tolerated consequence of maintaining greater receptor control. This mechanism may be distinct from how polyQ tracts expanded beyond 39 repeats cause Kennedy’s disease, as recent studies suggest that the fundamental properties of polyQ tracts change beyond this length (37,85). However, with a reduced capacity to undergo N/C interaction, ARs with expanded polyQ tracts may be more vulnerable to aggregation. The AR also contains a polymorphic GGN repeat in the AR-NTD that has been weekly associated with prostate cancer risk (18,86,87). Whereas the encoded polyglycine tract is located in a region of activation function AF-5 that binds coregulators and appears to be required for maximum N/C interaction (60,72), it is distinct from the polyQ in that it does not display as broad a spectrum of length variation and there is a bias for specific repeat numbers (18,69,86,87). As polyglycine peptides are considerably more hydrophobic and flexible than polyQ peptides (J.M. Harris, unpublished data), this may indicate that polyglycine tract length is constrained by specific requirements for folding within the NTD tertiary structure where it is less likely to have a dramatic influence on AR function.

The observation that normal AR function is sustained over a critical, but limited, range of polyQ lengths, could in part explain why analysis of AR CAG repeat length and risk of prostate cancer and other diseases has produced conflicting results. Our data suggest that changes in functionality of a receptor with a polyQ length either shorter or longer than the critical range of 16–29 residues could be a more important mediator of disease phenotype than a stepwise reduction in receptor activity with increasing length across the entire range. In this model, ARs with a polyQ tract within the critical range can be considered as functionally equivalent, whereas those with shorter or longer repeats as distinct. A consequence of this model is that analysis of samples by median AR-CAG repeat data and defining the downstream effects of polyQ alterations are warranted to better understand the endocrine basis of AR related diseases and the assessment of risk in pre-symptomatic individuals.

**MATERIALS AND METHODS**

**Cell culture**

Cell lines were obtained from the American Type Culture Collection (ATCC; Rockville, MD) and maintained in RPMI 1640 medium (GIBCO, BRL, Melbourne, Australia) supplemented with 5% fetal bovine serum (FBS). Antibiotics were not used in the culture medium. The PC-3 cells used in this study were negative for AR message and protein (PC-3AR−) as described previously (88).

**Plasmid vectors**

The probasin (ARR3-tk-luc) and MMTV (MMTV-CAT, MMTV-LUC) reporter constructs have been used previously in our studies (42). pGL3-PSA540 consists of the androgen responsive regions of the PSA enhancer (nucleotides 502–1951) and promoter (nucleotides 4924–5784; GenBank accession no. U37672) driving expression of the firefly luciferase gene. The pGK1 reporter used consists of a minimal adenovirus E1b promoter and five tandem Gal4 response elements that drive luciferase gene expression (89). AR expression vectors encoding different polyQ lengths [pcDNA-AR(CAG)n] were generated as described previously (42), pCMV-AR-E895Q expression vector was generated in the parental AR expression clone, pCMV-AR (90) using the PCR based megaprimer method previously described (91). The product, generated with the specific mutagenic (5'-ATGATGGCACAGATCATCTCTCTGTG-3') and flanking sense (5'-TGGAGATGAAAGCTTGGGTGT-3') and antisense (5'-CCCTCTAGATGAGCTAGATGAGCTAGATGAG-3') oligonucleotides, was cloned in place of the homologous HindIII fragment of pCMV-AR. pCMV-AR-polyQ2L was created using a 220 bp BglII fragment of the mutant polyQ region (amplified by PCR from human tumor p332 and digested with BglII) as a megaprimer with the antisense oligonucleotide, 5'-TGACACCCCGAGTCTATCTCTCA-3'. The resulting mutant PCR product was cloned in place of the homologous Smal to HindIII fragment of pCMV-AR. pCMV-MV-trunc707 contains the coding sequence for the first 707 amino acids of the wtAR and was generated by replacing the HindIII–XbaI fragment of pCMV-AR with a PCR product generated using a specific antisense oligonucleotide (5'-TGCATCTAGATGACACAGGGATTCTACCTTGAGCTAGTAATG-3') that incorporates a Stop codon. AR-polyQ2L-trunc707 was derived by replacing the EagI–BstEII fragment of pCMV-AR with the homologous region of CMV-AR-polyQ2L. The mammalian two-hybrid AR expression vector, pMAR-LBD (60,72) contains AR residues 644–917 cloned in-frame with the GAL4 DBD of pM (BD Biosciences Clontech, Palo Alto, CA) and is described elsewhere (60). pVP16-AR(NTD), which encompasses AR amino-acid residues 1–538, was created by cloning a PCR product generated using appropriate sense (5’-AGCTGATATT CATGGAAGTGCAGTTAGGGCTG-3') and antisense (5’-TGCTCTAGATCCGGCTGAGAGCTGAGTCGA-3') oligonucleotides in the EcoRI and XbaI sites of pVP16 (Clontech). Mutations and variations in (CAG)n length were cloned into these vectors using appropriate restriction fragments from the parental vectors listed above. To create pCMV-RAN, a PCR fragment containing the RAN coding sequence was amplified by PCR from T47D cDNA using specific sense (5’-CTAGAATTCCTCGCATGTTAGGGCTG-3') and antisense (5’-GTAGAATTCCTCGAGCATGTTAGGGCTG-3') oligonucleotides, and cloned into the EcoRI site of pCMV. pCDNA-AIB1 is described elsewhere (42). The integrity of all clones was determined by sequence analysis in both directions.
**Transactivation assays**

AR transactivation assays in 96-well plates was performed as previously described (91) using 10 000 PC-3 cells/well with 0.1–10 ng of AR expression vector and 100 ng of reporter. The effect of RAN was determined by adding 1–5 ng of pCMV-RAN or empty vector control. In all cases, transfection mixes were balanced with respect to the molar ratio of all expression vectors (using the appropriate empty vector) and total plasmid [using the promoter-less vector, pCAT-basic or the prokaryotic vector, pBS-sk(–)]. Transfection was performed using LipofectAMINE 2000™ (GIBCO-BRL) according to the manufacturer’s instructions. Luciferase activity was determined in cell lysates using the Luciferase™ Reporter Gene Assay Kit (Promega) and a plate reading luminometer (Top Count™, Packard, Mount Waverley, ACT, Australia). CAT assays for AR activity and coactivation by AIB1 were performed as previously described (42). Mammalian two-hybrid assays were performed in COS-1 cells (15 000 cells/well) in 96-well plates. Each well was cotransfected with an equal molar amount of the appropriate pM and pVP16 vectors (maximum 15 ng of each vector), and 25 ng of the pGK1 reporter. DNA was balanced by adding an appropriate amount of pBS-sk(–). Following transfection, cells were treated for 30 h in phenol-red free medium supplemented with the appropriate steroids, and luciferase activity was determined in cell lysates as detailed above.

**Immunoblot analysis**

Cells transfected in 60–100 mm dishes with the appropriate vectors were harvested into cytosol or RIPA buffers containing protease inhibitors and soluble cytosol fractions prepared by centrifugation. Total cellular protein (20 mg) was electophoresed on 6% sodium dodecyl sulfate–polyacrylamide gels, transferred to Hybond-C or Hybond-P membranes (Amerham-Pharmacia Biotech) and immunostained using AR antibodies U402 (92) or N20 (Santa Cruz Biotechnology, Santa Cruz, CA) and control cytokeratin 8 antibody (Sigma, St Louis, MO). Immunoreactivity was detected using appropriate horseradish peroxidase conjugated IgG and visualized using ECL western blotting reagents (Amerham). Protein levels were determined from immunoblots by quantitative scanning densitometry using a Bio-Rad Model GS-710 Imaging Densitometer. Immunoblot analysis was also performed on lysates pooled from replica samples in 96-well plates following transactivation analysis or mammalian two-hybrid assays.

**Affinity**

Affinity analysis of AR for DHT was performed as previously described (90,93,94) in COS-1 cells (10⁶ cells per 100 mm dish) transiently transfected with 20 μg of wtAR or AR-polyQ2L expression vectors and cultured in medium containing 5% charcoal stripped FBS for 48 h. Soluble cytosol fractions were incubated with 0.1–6.0 nM [³H]DHT for 16 h at 4°C before excess steroid was removed using dextran coated charcoal and centrifugation. Specific activity in fractions was determined by liquid scintillation and the affinity calculated using Scatchard plot analysis and linear regression.

**Quantitative real-time PCR for RNA and plasmid DNA in transfected cells**

RNA was extracted from PC-3 cells (7.5 × 10⁵ cells per well in 6-well culture dishes) 4 and 24 h after transfection with 250 ng AR expression vectors using the RNA easy Mini Kit (Qiagen) according to the manufacturer’s instructions. A portion of the RNA was treated with DNase, extracted using phenol:chloroform:isoamylalcohol, precipitated with ethanol/sodium acetate and reverse transcribed using SuperscriptII™ reverse transcriptase (GIBCO-BRL, Melbourne, VIC, Australia) according to manufacturer’s instructions. A second aliquot of the primary RNA preparation was treated with 2 U DNase free RNase A (Boehringer Manheim) and similarly extracted to yield AR vector DNA from transfected cells. Relative levels of AR and β-actin RNA were determined in each sample by quantitative real-time PCR on a Rotor-Gene 2000 (Corbett Research, Mortlake NSW) using 2 × SYBR Green PCR Mastermix (Applied Biosystems) and specific oligonucleotide primers (AR, 5′-AGCCATTGAGCCAGGTGTAGTGTG-3′ and 5′-GT GAAGGATCGCCGCCCAT-3′; β-actin, 5′-GCAAAACA CATGTCTGTCTGG-3′ and 5′-TACTCTGTTGCTGATCACA-3′). Relative AR expression was determined by normalization of cycle threshold values to those of β-actin for each sample using Rotor-Gene software (Version 4.4 Build 1). Values were corrected for the level of AR vector DNA determined by analogous reactions from the DNA preparation. The AR primers span ~13 kb of intron 4 of the AR gene and can therefore be regarded as specific for the vector DNA in these samples. β-Actin primers provide a similar fragment for DNA as for RNA above.

**Modeling of the AR-polyQ tract**

Molecular modeling of the AR polyQ tract was performed by constructing an alpha-helical polyQ stretch as a free peptide and subjected it to simulated annealing with molecular dynamics using the program Biomer (http://www.scripps.edu/~nwhite/B/). This simulation was repeated using a polyQ peptide containing the two leucine substitutions. Both simulations treated the polyQ tract as a free peptide rather than a constrained peptide within a larger assembly. The energy of each peptide was calculated using the AMBER implementation in ChemsitePro. Secondary structure predictions were performed on sequences of at least 50 residues in length using the NNpredict program of Kneller (1990) available online at http://www.cmparm.ucsf.edu/~nomi/nnpredict.html.

**ACKNOWLEDGEMENTS**

The authors would like to thank Dr Lisa M. Butler for assistance, Dr Mike Stallcup (University of Southern California, Los Angeles, CA, USA) for providing mammalian two-hybrid AR and GRIP1 expression vectors and Dr Marco Marcelli (Bristol Myers Squibb, Princeton, NJ, USA) for pGL3-PSA540. Ms Marie A. Pickering, Ms Kathleen M. Saint and...
Mr George Hatzinikolas provided technical assistance. This work was supported by the National Health and Medical Research Council of Australia (ID#299048), The Cancer Council of South Australia (RG 53/99), B1 funding from the Faculty of Health Sciences at the University of Adelaide, the Prostate Cancer Foundation of Australia, the NIH (CA84890) and United States Department of Defense (W81XWH-04-1-0017, DAMD17-00-1-0102).

REFERENCES


