A bioinformatics-based functional analysis shows that the specifically androgen-regulated gene SARG contains an active direct repeat androgen response element in the first intron

Karine Steketee, Angelique C J Ziel-van der Made, Hetty A G M van der Korput, Adriaan B Houtsmuller and Jan Trapman

Department of Pathology, Josephine Nefkens Institute, Erasmus Medical Center, PO Box 1738, 3000 DR Rotterdam, The Netherlands

(Requests for offprints should be addressed to J Trapman; Email: j.trapman@erasusmc.nl)

Abstract

We characterized the specifically androgen-regulated gene (SARG), which is expressed in the androgen receptor (AR) and glucocorticoid receptor (GR) positive cell line lymph node carcinoma of the prostate-1F5 (LNCaP-1F5). SARG mRNA expression can be up-regulated by androgens, but not by glucocorticoids. SARG mRNA expression is high in prostate tissue. SARG is composed of four exons and spans a region of 14·5 kbp on chromosome 1q32·2. Transcripts of 5·5, 3·3 and 2·3 kb are the result of alternative polyadenylation. SARG mRNA splice variants lack exon 2 and vary in length of exon 1. The SARG protein has a length of 601 amino acids and is located in the cytoplasm. By screening the 18 kbp genomic sequence flanking the transcription start site we identified the imperfect direct repeat 5’TGTGCTaacTGTTCT-3’ in intron 1 as an active androgen response element (ARE-SARG+4·6). A 569 bp genomic DNA fragment containing this element functioned as an androgen-specific enhancer in transiently transfected LNCaP-1F5 cells. ARE-SARG+4·6 cooperated with flanking sequences for optimal activity. Inactivation of ARE-SARG+4·6 completely abolished the androgen response of the enhancer. Chromatin immunoprecipitation (ChIP) experiments showed chromatin structural changes of the enhancer in the presence of R1881. ARE-SARG+4·6 was able to bind to the androgen receptor, but not to the glucocorticoid receptor, correlating with its androgen-specific activity in transfections.

Journal of Molecular Endocrinology (2004) 33, 477–491

Introduction

Androgens are essential in the development and maintenance of the male phenotype. They mediate their function by activation of the androgen receptor (AR), which is a member of the nuclear receptor family of ligand-activated transcription factors. Nuclear receptors have a modular structure composed of a moderately conserved carboxyl-terminal ligand-binding domain (LBD), a highly conserved central DNA-binding domain (DBD) and a non-conserved amino-terminal domain (NTD). Most ligand-activated nuclear receptors bind as homodimers or heterodimers to hormone response elements (HREs) in the regulatory regions of their target genes. HREs are composed of an inverted or direct repeat of two 6 bp half-sites separated by a spacer of variable size (Khorasanizadeh & Rastinejad 2001). Together with coactivators, chromatin remodelling complexes, general transcription factors and RNA polymerase II, nuclear receptors initiate the transcription of target genes in a tightly controlled fashion (Glass & Rosenfeld 2000, Lee & Lee Kraus 2001, McKenna & O’Malley 2002).

An important class of nuclear receptors is the family of steroid hormone receptors, which is composed of AR, glucocorticoid receptor (GR), mineralocorticoid receptor (MR), progesterone receptor (PR) and estrogen receptors alpha and beta (ERα and ERβ) (Thornton 2001). Steroid hormone receptors display distinct physiological functions, reflected in their tissue-specific expression pattern and to some extent in their spectrum...
of target genes. However, AR, GR, MR and PR all bind with high affinity to the same inverted repeat consensus sequence 5′-AGAACAnnnTGACCT-3′ (Nordeen et al. 1990, Roche et al. 1992, Lieberman et al. 1993, Lombes et al. 1993). As a result, the activity of several promoters can be regulated by more than one of these receptors. Examples are the MMTV promoter, and the promoters of the C3, the cystatin-related protein (CRP) and the prostate specific antigen (PSA) gene (Ham et al. 1988, Claessens et al. 1989, De Vos et al. 1994, Cleutjens et al. 1997, Devos et al. 1997). The consensus high affinity binding site of ERα and ERβ is slightly different, 5′-AGGTCAnnnTGACCT-3′ (Klein-Weiβ et al. 1989), therefore ERα and ERβ direct the expression of a different panel of target genes. Because GR, MR, PR and AR recognize the same DNA sequence, it has been postulated that additional mechanisms are necessary to explain their specificity. These include differences in expression levels of the various receptors in specific cell types (Strahle et al. 1989), selective interaction with specific transcription factors, coactivators and corepressors and ligand availability (Glass & Chang 2002).

In spite of the identical high-affinity recognition sequence for AR, PR, GR and MR, steroid response elements can also direct receptor specificity. In natural promoters steroid receptor binding sites can deviate considerably from the consensus high-affinity binding site. These sequences might have a different affinity to the various receptors. Additionally, sequences directly flanking the response element can contribute to receptor affinity and preference (Nelson et al. 1999, Haelens et al. 2003). On top of this, the AR seems to have adopted an exclusive mechanism of specificity. A few genes are known to be preferentially regulated by AR (Claessens et al. 2001). The structures of the androgen response elements (AREs) that direct androgen-specificity to these genes resemble more direct repeats of the sequence 5′-TGTTCT-3′ than classic inverted repeats of this sequence.

The androgen-sensitive human prostate carcinoma cell line, lymph node carcinoma of the prostate (LNCaP) expresses AR, but lacks GR and PR (Horoszewicz et al. 1983, Berns et al. 1986). It was previously shown that growth of LNCaP cells and PSA mRNA expression in these cells can be stimulated by androgens (Schuurmans et al. 1988, Riegman et al. 1991, Young et al. 1991). In order to directly compare the molecular and biological function of AR and GR, the LNCaP-1F5 subline, containing a stably integrated GR expression vector, was generated (Cleutjens et al. 1997). PSA mRNA expression in LNCaP-1F5 can be induced by both androgens and glucocorticoids but cell growth is selectively induced by androgens. We identified in LNCaP-1F5 cells a novel gene that is specifically regulated by androgens (Cleutjens et al. 1997). In the present study an integrated experimental and bioinformatics-based approach was applied to characterize the gene, designated specifically androgen-regulated gene (SARG), and to decipher the molecular mechanism of androgen-specific regulation of the gene.

Materials and methods

Materials

Methyltrienolone (R1881) was purchased from NEN (Boston, MA, USA), dexamethasone (Dex) was obtained from Steraloids (Wilton, NH, USA). Cell culture media were from Bio Whittaker (Verviers, Belgium), fetal calf serum (FCS) was from Roche Diagnostics (Almere, The Netherlands).

Plasmid construction

pLUC and pPSA-4-LUC have been described previously (Cleutjens et al. 1996). pHisXpress-cSARG, expressing (His)6-Xpress-SARG protein, contains the SARG cDNA fragment 209–2579 (SARG ORF is from 251 to 253) inserted in the eukaryotic expression vector pcDNA3·1 His (Invitrogen, Carlsbad, CA, USA).

The SARG genomic fragments SARG-8·5, SARG-7·3 and SARG+4·6, with sizes of 510 bp, 476 bp and 569 bp respectively, were obtained by PCR on PAC90 L18 DNA (GenomeSystems, St Louis, MO, USA) as template with the primer sets: -8·5F: 5′-GATCCAGCTGGATCCAGGGA CATGGATGAAGCTG-3′; -8·5R: 5′-GATCATGGATCGATCCTGGAGAGGCAGTCTAGTCAG-3′; -7·3F: 5′-GATCATGGATCCTGGATGTCATAAT-3′; -7·3R: 5′-GATCAGCCTGGATGTCATAAT-3′; -4·6F: 5′-GATCATGGATCCTGGATGTCATAAT-3′; +4·6R: 5′-GATCAGCCTGGATGTCATAAT-3′.
The resulting amplified fragments were inserted in pGEM-T Easy (Promega, Madison, WI, USA), sequenced and subsequently inserted as BamHI/BamHI fragments in pPSA-4-LUC, yielding pSARG - 8‘-3-PSA-LUC, pSARG - 7‘-3-PSA-LUC and pSARG + 4‘-6-PSA-LUC respectively.

The -55 to +168 genomic fragment SARG-S was obtained by PCR on PAC90 L18 DNA (Genome Systems), utilizing the primers SARG-A: 5’-GCTAAGAGGGAACAGCACCAC-3’ and SARG-B: 5’-CCCAGGAGATCTACTATGCTCACTGGGT-3’. The PCR product was inserted in pGEM-T Easy (Promega), verified by sequencing and inserted as a 240 bp PvuII/BglII fragment in pLUC, yielding pSARG-S-LUC. To generate pSARG-L-LUC, the 3012 to 1559 HindIII/PvuII genomic fragment was isolated from PAC90 L18 and inserted in the corresponding sites of pSARG-S-LUC, yielding pSARG-L-LUC. Subsequently, the 1559 to 55 PvuII/PvuII genomic SARG fragment was inserted into the PvuII site of pSARG-L-LUC. The resulting construct pSARG-L-LUC contains SARG bp -3012 to +168.

SARG+4‘-6 was inserted as a BamHI/BamHI fragment upstream of the SARG-S promoter or SARG-L promoter in pSARG-S-LUC and pSARG-L-LUC respectively, yielding pSARG + 4‘-6-SARG-S-LUC and pSARG+4‘-6-SARG-L-LUC. pSARG+4‘-6 m-SARG-S-LUC was generated by mutagenesis utilizing the QuikChange Site Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA) using pSARG + 4‘-6-SARG-S-LUC as template, primer mut–4603S: 5’-CAACTAAACTATGATAACTATTATCTCATTTAATC-3’ and its complementary strand, mut–4603AS.

SARG+4‘-6-(+4447/+4659)-S-LUC and SARG + 4‘-6-(+4548/+4659)-S-LUC were constructed by PCR utilizing the primers S+4‘-6-A (5’-GACGTGGATCCCTTCTTTTCTGAGATCC TG-3’) and S+4‘-6-B (5’-GATCAGCTGGATCCCTTCTTTTTCTGAGATCC TG-3’) and S+4‘-6-S (5’-GATCAGCTGGATCCCTTCTTTTTCTGAGATCC TG-3’) as respective forward primers, and S+4‘-6-C (5’-GATCAG ATGGATCCGGCAAATTTACCTCTGAGTCTG-3’) as reverse primer. Amplified cDNA fragments were sequenced prior to insertion into pSARG-S-LUC as BamHI/BamHI fragments.

pRIT2 TAR, encoding rat AR DBD, was described previously (De Vos et al. 1991). prGR-DBD-PRT2T, encoding rat GR DBD, was constructed by BamHI/SalI insertion of a PCR fragment, synthesized with primers rGR-DBD-1: 5’-CAGCGGATTCGCCAGGGACCCAC-3’ and SARG-B: 5’-CCCAGGAGATCTACTATGCTCACTGGGT-3’. The PCR product was inserted in pGEM-T Easy (Promega), verified by sequencing and inserted as a 240 bp PvuII/BglII fragment in pLUC, yielding pSARG-S-LUC. To generate pSARG-L-LUC, the 3012 to 1559 HindIII/PvuII genomic fragment was isolated from PAC90 L18 and inserted in the corresponding sites of pSARG-S-LUC, yielding pSARG-L-LUC.

Screening of a prostate cDNA library

Screening of a λgt10 human prostate cDNA library (BD Biosciences Clontech, Palo Alto, CA, USA) was performed according to the manufacturer’s protocol. Hybridization probes were the SARG differential display PCR (DD-PCR) fragment (GenBank accession Number AF007835) and SARG cDNA fragment 855–1957 (see GenBank accession Number AY352640).

RACE-PCR

For RACE-PCR we applied the Marathon-Ready prostate cDNA cloning kit (BD Biosciences Clontech), primer SARG-RACE: 5’-CCGGTCTTTCTCTGGAATGTG-3’ and the standard AP1 primer of the kit. Amplified cDNA fragments were inserted in pGEM-T Easy and sequenced.

Analysis of alternative splicing of mRNA by RT-PCR

cDNA was synthesized from 1 µg total RNA isolated from LNCaP cells incubated for 24 h in RPMI 1640 supplemented with 5% dextran-coated charcoal treated FCS (FCS-DCC), antibiotics and 1 nM R1881. cDNA synthesis was performed at 55°C utilizing M-MuLV reverse transcriptase and an oligo-dT primer. Subsequently, PCR was carried out under standard conditions on the LNCaP cDNA template utilizing either forward primer SARG-F1A: 5’-GATCAGCTGGATCCCTTCTTTTCTGAGATCC TG-3’ and the standard AP1 primer of the kit. Amplified cDNA fragments were sequenced prior to insertion into pSARG-S-LUC as BamHI/BamHI fragments.
RNA isolation and Northern analysis

Total RNA from LNCaP and LNCaP-1F5 cells was isolated by the guanidinium thiocyanate method (Sambrook & Russell 2001). RNA (10 µg per lane) was separated by electrophoresis on a 1% (w/v) agarose formaldehyde gel in TBE. Following electrophoresis RNA was transferred to a Hybond-N* membrane (Amersham Biosciences). The blot was hybridized under standard conditions at 65 °C utilizing the 32P-labelled HindIII/HindIII SARG cDNA fragment (854–1957) as a probe (Sambrook & Russell 2001). Actin cDNA was utilized as a hybridization control. Blots were exposed to X-ray film with intensifying screens at −80 °C.

Analysis of tissue specific expression of mRNA by PCR

Tissue specificity of SARG mRNA was assayed by semi-quantitative PCR on Human MTC Panel II cDNA (BD Biosciences Clontech), containing cDNAs from spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood lymphocytes, essentially according to the procedure described in the user manual. G3PDH primers from the cDNA kit were used as a control (30 amplification cycles). SARG primers utilized were 5'-AGTCTGAGCCAGCCACA ACT-3' (F-ex3) and 5'-TGTGGATATTCGGGAGG-3' (R-ex4) (30 amplification cycles, primer annealing was at 55 °C).

Immunocytochemistry

LNCaP cells were seeded at a density of 3 x 10^5 cells per well on sterile micro-slides in four-well tissue culture plates (Heraeus Instruments, Hanau, Germany), cultured until 50% confluence in RPMI 1640, supplemented with 5% (v/v) FCS and antibiotics and subsequently transfected with 5 µg pHisXpress-cSARG. After overnight incubation, cells were washed twice in PBS and fixed in acetone for 10 min. Next, slides were rinsed twice in PBS, followed by overnight incubation in mouse anti-Xpress antibody solution (Invitrogen) diluted 1:500 in PBS at 4 °C. Incubation was stopped by four PBS washes. Next, slides were incubated for 30 min at room temperature in goat anti-mouse peroxidase conjugate antibody (DAKO, Glostrup, Denmark) solution (1:100 dilution in PBS). After four PBS washes, immunoreactivity was visualized by dianaminobenzidine (DAB) staining. The reaction was stopped in water. Cells were counterstained with Mayers Hematoxylin.

Isolation of genomic DNA fragments

The SARG DD-PCR fragment was randomly 32P-labelled and utilized to screen a genomic human PAC library on gridded filters (Genome-Systems, St Louis, MO, USA) according to the manufacturer’s protocol. DNA was isolated from positive PACs by standard procedure (Sambrook & Russell 2001). For Southern blot analysis 10 µg PAC90 L18 DNA was HindIII, PstI or EcoRI digested, electrophoresed on a 0.8% TAE-agarose gel and subsequently transferred to a Hybond-N* membrane. Filters were hybridized at high stringency with randomly 32P-labelled SARG probes under standard conditions (Sambrook & Russell 2001). HindIII, PstI or EcoRI digested PAC DNA was shot-gun cloned in the corresponding sites of pBSKS+ (- Stratagene). Clones were utilized for isolation of genomic fragments by screening with randomly 32P-labelled SARG cDNA fragments and for SARG gene walking with overlapping HindIII, PstI and EcoRI fragments. Hybridizing inserts were sequenced.

Search for candidate androgen response elements

The MatInspector professional program (www. genomatix.de/mat_fam) (Quandt et al. 1995) was utilized for detection of candidate AREs with queries for the inverted repeat 5'-RGWACAN NNTTTCT-3' (R=A/G, W=G/T) and the direct repeat 5'-TGTTCTNNNTGTTCT-3'. The threshold for candidate AREs was set at nine out of 12 matches. The MatInspector program searched both the sense and anti-sense strand. Identified sequences were further selected manually according to additional criteria. Candidate inverted repeat AREs should contain G and C at the double-underlined positions in the sequence above, and at least one of either single underlined C or G. Candidate direct repeat AREs should contain three out of four single-underlined C and G residues.
Cell culture, transfection and luciferase assay

LNCaP and LNCaP-1F5 cells were cultured in RPMI 1640 supplemented with 5% (v/v) FCS and antibiotics. Four hours prior to transfection the medium was substituted by Dulbecco’s Modification of Eagle’s Medium (DMEM) supplemented with 5% (v/v) FCS-DCC. Transient transfections were performed following the calcium phosphate precipitation method (Sambrook & Russell 2001) utilizing 1 x 10⁶ cells per 25 cm² flask and 5 µg of one of the pLUC-constructs. After 4 h the medium was removed and cells were incubated for 90 s at room temperature in PBS containing 15% (v/v) glycerol. Next, transfected cells were cultured in DMEM-FCS-DCC medium for 24 h in the absence or presence of 1 nM R1881 or 10 nM Dex. Transfected cells were washed in PBS, and subsequently incubated in 300 µl lysis buffer (25 mM Tris–phosphate, pH 7.8; 8 mM MgCl₂; 1 mM DTT; 1% (v/v) Triton X-100; 15% (v/v) glycerol). Next, 100 µl 0.25 mM luciferin (Sigma, St Louis, MO, USA) and 0.25 mM ATP in lysis buffer was added to 150 µl lysate and luciferase activity was measured in a LUMAC 2500 M Biocounter (LUMAC, Landgraaf, The Netherlands).

Electrophoretic mobility shift assay

AR DBD and GR DBD were produced in *Escherichia coli* and purified as described previously (De Vos et al. 1991). AR DBD and GR DBD were expressed from pRIT2 TAR and pRIT2 TrGR-DBD respectively. The following oligonucleotide sequences were filled in by standard M-MuLV-RT reaction in the presence of α-³²P-dATP and subsequently purified on a non-denaturing polyacrylamide gel. For EMSA, 50 x 10³ c.p.m. probe was added to 20 µl reaction mixture, containing 2 µg poly dIdC, 2 µg BSA, 10 µM ZnCl₂, 1 mM DTT and 2 µl 10x binding buffer (100 mM Hepes, pH 7.6; 300 mM KCl; 62.5 mM MgCl₂; 4% (v/v) ficoll 400) and 5 pmol AR DBD or GR DBD. Incubation was for 30 min on ice. Samples were electrophoresed on a 4% (w/v) polyacrylamide gel (19:1 mono/bis acrylamide ratio) gel in a 25 mM Tris.HCl, 41.5 mM boric acid, 0.5 mM EDTA buffer for 2 h at 150 V at room temperature. Subsequently, gels were fixed, dried and exposed to X-ray film.

Chromatin immunoprecipitation (ChIP)

ChIPs were done essentially according to the method described in the Acetyl-Histone H3 ChIP assay kit (Upstate Biotechnology, Chicago, IL, USA). In short, LNCaP cells were grown for at least 3 days in 5% FCS-DCC supplemented RPMI 1640 medium. To half of the cultures R1881 was added to a final concentration of 10 nM. After 1 h cells were cross-linked with formaldehyde (1% final concentration) at 22°C for 10 min. Cross-linking was stopped by addition of glycine to a final concentration of 125 mM. Next, cells were washed in ice-cold PBS and harvested in PBS supplemented with protease inhibitors (Roche Diagnostics). Cell pellets were resuspended in SDS lysis buffer and sonicated to shear the DNA. Sonicated samples were centrifuged, diluted in Chip Dilution Buffer (Upstate Biotechnology, Chicago, IL, USA). In short, LNCaP cells were grown for at least 3 days in 5% FCS-DCC supplemented RPMI 1640 medium. To half of the cultures R1881 was added to a final concentration of 10 nM. After 1 h cells were cross-linked with formaldehyde (1% final concentration) at 22°C for 10 min. Cross-linking was stopped by addition of glycine to a final concentration of 125 mM. Next, cells were washed in ice-cold PBS and harvested in PBS supplemented with protease inhibitors (Roche Diagnostics). Cell pellets were resuspended in SDS lysis buffer and sonicated to shear the DNA. Sonicated samples were centrifuged, diluted in Chip Dilution Buffer (Upstate Biotechnology, Chicago, IL, USA). In short, LNCaP cells were grown for at least 3 days in 5% FCS-DCC supplemented RPMI 1640 medium. To half of the cultures R1881 was added to a final concentration of 10 nM. After 1 h cells were cross-linked with formaldehyde (1% final concentration) at 22°C for 10 min. Cross-linking was stopped by addition of glycine to a final concentration of 125 mM. Next, cells were washed in ice-cold PBS and harvested in PBS supplemented with protease inhibitors (Roche Diagnostics). Cell pellets were resuspended in SDS lysis buffer and sonicated to shear the DNA. Sonicated samples were centrifuged, diluted in Chip Dilution Buffer (Upstate Biotechnology, Chicago, IL, USA). In short, LNCaP cells were grown for at least 3 days in 5% FCS-DCC supplemented RPMI 1640 medium. To half of the cultures R1881 was added to a final concentration of 10 nM. After 1 h cells were cross-linked with formaldehyde (1% final concentration) at 22°C for 10 min. Cross-linking was stopped by addition of glycine to a final concentration of 125 mM. Next, cells were washed in ice-cold PBS and harvested in PBS supplemented with protease inhibitors (Roche Diagnostics). Cell pellets were resuspended in SDS lysis buffer and sonicated to shear the DNA. Sonicated samples were centrifuged, diluted in Chip Dilution Buffer (Upstate Biotechnology, Chicago, IL, USA). In short, LNCaP cells were grown for at least 3 days in 5% FCS-DCC supplemented RPMI 1640 medium. To half of the cultures R1881 was added to a final concentration of 10 nM. After 1 h cells were cross-linked with formaldehyde (1% final concentration) at 22°C for 10 min. Cross-linking was stopped by addition of glycine to a final concentration of 125 mM. Next, cells were washed in ice-cold PBS and harvested in PBS supplemented with protease inhibitors (Roche Diagnostics). Cell pellets were resuspended in SDS lysis buffer and sonicated to shear the DNA. Sonicated samples were centrifuged, diluted in Chip Dilution Buffer (Upstate Biotechnology, Chicago, IL, USA). In short, LNCaP cells were grown for at least 3 days in 5% FCS-DCC supplemented RPMI 1640 medium. To half of the cultures R1881 was added to a final concentration of 10 nM. After 1 h cells were cross-linked with formaldehyde (1% final concentration) at 22°C for 10 min. Cross-linking was stopped by addition of glycine to a final concentration of 125 mM. Next, cells were washed in ice-cold PBS and harvested in PBS supplemented with protease inhibitors (Roche Diagnostics).
AGCTAGTC-3'; –7·3R: 5'–CTGAGATGCTGA GAGGCTGA-3'; +4·6F: 5’–CAAGTCTACAGT CTCCCATC-3' and +4·6R: 5’–CTCAAATCCC AGTTTAGCCA-3’. PCR fragments were separated over an agarose gel.

**Results**

**SARG mRNA expression in LNCaP-1F5 cells**

Utilizing DD-PCR technology we previously identified in LNCaP-1F5 prostate cancer cells, which express both AR and GR, a novel androgen-specific regulated gene, denoted 21·1 in the initial study and \(SARG\) in the present study (Cleutjens et al. 1997). SARG mRNA expression was found to be up-regulated by the synthetic androgen R1881, but not by the synthetic glucocorticoid Dex. Utilizing the DD-PCR fragment as hybridization probe, a 5·5 kb transcript was identified in R1881-incubated LNCaP-1F5 cells.

For further characterization of SARG we first isolated full-length SARG mRNA. Overlapping SARG cDNA fragments were obtained by repeated screening of a human prostate cDNA library. In the first screen, the DD-PCR fragment was utilized as a hybridization probe (Fig. 1a). The longest cDNA, containing a polyadenylation signal and a polyA tail, was 3·6 kbp. Screening of the cDNA library with a 5’ fragment of this cDNA as a probe resulted in the detection of an overlapping 2·7 kbp cDNA with a second polyadenylation signal and a polyA tail. This cDNA fragment extended the cDNA sequence to approximately 4·9 kbp. Further, 5’ SARG cDNA sequence was obtained by RACE-PCR, utilizing a primer in the 2·7 kbp cDNA. Two related 5’ cDNA fragments of 665 bp and 537 bp, were found. The shorter fragment (RACE1) lacked nucleotides 42–169 of the longer fragment (RACE2) (Fig. 1a). The longest SARG cDNA sequence of 5487 bp was deposited in GenBank under Accession Number AY352640.

A BLAST search of the EST database (www.ncbi.nlm.nih.gov/) identified two EST clusters overlapping the SARG cDNA sequence. The first group represented the 3’ parts of the 5·5 kb and 3·3 kb transcripts, as detected in the cDNA library (Unigene cluster Hs.32417), the second cluster represented a 2·3 kb transcript, which contained SARG 5’ cDNA sequences and a third polyadenylation sequence and polyA tail (Unigene cluster Hs.223394). We confirmed the presence of three polyadenylation signals in the 5·5 kbp SARG cDNA sequence (Fig. 1a). A 1·1 kbp SARG cDNA fragment (nucleotides 854 to 1957) hybridized with all three predicted SARG mRNAs in a Northern blot of LNCaP-1F5 RNA (Fig. 1b). The 3·3 kb transcript showed the highest expression. All SARG transcripts were up-regulated by R1881 but their expression could not be induced by Dex. Semi-quantitative RT-PCR indicated high SARG mRNA expression in prostate tissue (Fig.1c, lane 3) as compared with spleen, thymus, testis, ovary, small intestine, colon and lymphocytes.

**SARG protein**

The SARG open reading frame (ORF) encodes a 601 amino acid protein (Fig. 2a). This ORF is identical to that of the hypothetical protein MGC2742 (Genbank). Because Unigene cluster Hs.23417 encompasses the 3’ part of the two longest SARG transcripts, a predicted protein from this EST cluster (MGC4309) is unlikely.

To determine its cellular localization, SARG protein was Xpress-tagged and transiently expressed in LNCaP cells. Immunocytochemical staining with anti-Xpress antibody showed that SARG protein was exclusively present in the cytoplasm (Fig. 2b).

**SARG gene structure and splice variants**

The complete \(SARG\) gene was isolated in one PAC (90 L18) by screening of a human genomic PAC library with the SARG DD-PCR fragment (see Fig. 1a) as probe. To characterize \(SARG\), subcloned overlapping HindIII, PstI and EcoRI fragments of PAC 90 L18 were hybridized with appropriate cDNA fragments. Comparison with the cDNA sequence revealed that \(SARG\) was composed of four exons, and spanned 14·5 kbp (Fig 3a).

A BLAST search of the EST database (www.ncbi.nlm.nih.gov/) identified two EST clusters overlapping the SARG cDNA sequence. The first group represented the 3’ parts of the 5·5 kb and 3·3 kb transcripts, as detected in the cDNA library (Unigene cluster Hs.32417), the second cluster represented a 2·3 kb transcript, which contained SARG 5’ cDNA sequences and a third polyadenylation sequence and polyA tail (Unigene cluster Hs.223394). We confirmed the presence of three polyadenylation signals in the 5·5 kbp SARG cDNA sequence (Fig. 1a). A 1·1 kbp SARG cDNA fragment (nucleotides 854 to 1957) hybridized with all three predicted SARG mRNAs in a Northern blot of LNCaP-1F5 RNA (Fig. 1b). The 3·3 kb transcript showed the highest expression. All SARG transcripts were up-regulated by R1881 but their expression could not be induced by Dex. Semi-quantitative RT-PCR indicated high SARG mRNA expression in prostate tissue (Fig.1c, lane 3) as compared with spleen, thymus, testis, ovary, small intestine, colon and lymphocytes.

**SARG protein**

The SARG open reading frame (ORF) encodes a 601 amino acid protein (Fig. 2a). This ORF is identical to that of the hypothetical protein MGC2742 (Genbank). Because Unigene cluster Hs.23417 encompasses the 3’ part of the two longest SARG transcripts, a predicted protein from this EST cluster (MGC4309) is unlikely.

To determine its cellular localization, SARG protein was Xpress-tagged and transiently expressed in LNCaP cells. Immunocytochemical staining with anti-Xpress antibody showed that SARG protein was exclusively present in the cytoplasm (Fig. 2b).

**SARG gene structure and splice variants**

The complete \(SARG\) gene was isolated in one PAC (90 L18) by screening of a human genomic PAC library with the SARG DD-PCR fragment (see Fig. 1a) as probe. To characterize \(SARG\), subcloned overlapping HindIII, PstI and EcoRI fragments of PAC 90 L18 were hybridized with appropriate cDNA fragments. Comparison with the cDNA sequence revealed that \(SARG\) was composed of four exons, and spanned 14·5 kbp (Fig 3a). The two cDNA fragments obtained by RACE-PCR represented two forms of exon 1, the short exon 1A, and the extended exon 1AB. All splice junctions were consistent with the GT/AG rule (Fig. 3b). The SARG ORF started in exon 2 and ended in the large exon 4 (Fig. 3a). \(SARG\) is part of BAC RP11–564A8 (Genbank Accession Number AC098935-2). The transcription start site is at
Figure 1. SARG cDNA, and androgen-regulated and prostate specific SARG mRNA expression. (a) Schematic overview of the isolation of the complete SARG cDNA. The start and stop codons are indicated. An asterisk indicates a polyadenylation signal. DD–PCR indicates the original differential display fragment (Cleutjens et al. 1997). RACE1 and RACE2 are two related 5′ ends of SARG cDNA. (b) Northern blot of SARG mRNA expression in LNCaP-1F5 cells incubated for 24 h without hormone (−), with 1 nM R1881 or 10 nM Dex. Transcript sizes are indicated. Actin was utilized as a loading control. (c) RT-PCR analysis of SARG mRNA expression in (1) spleen, (2) thymus, (3) prostate, (4) testis, (5) ovary, (6) small intestine, (7) colon and (8) lymphocytes. cDNAs were normalized for G3PDH expression as loading control.
Figure 2 The SARG open reading frame and the cellular localization of the SARG protein. (a) The SARG ORF encodes a protein of 601 amino acids. Methionine 247 (in bold) is the first amino acid residue of the shorter SARG protein translated from mRNA lacking exon 1B and/or exon 2 sequences. (b) The SARG protein is located in the cytoplasm. LNCaP cells were transfected with pcHisXpress-cSARG encoding Xpress-tagged SARG protein. Xpress-SARG was visualized by immunostaining with anti-Xpress antibody.
position 184 776 in this clone. **SARG** maps at chromosome band 1q32·2.

**RACE-PCR and RT-PCR** revealed four different **SARG** splice variants (Fig. 3c). The largest variant contained all exons, smaller variants lacked either part B of exon 1AB, exon 2 or both. The splice variants lacking exon 2, which formed a minority, are predicted to encode a protein of 355 amino acids, starting at methionine 247 (Fig. 2a). The corresponding ATG codon is in exon 4, in frame with the long **SARG** ORF (Fig 3c).

### Functional and bioinformatics-based selection of candidate androgen response elements

To establish androgen response of **SARG**, the promoter fragments **SARG-L** (−3012 to +168) and **SARG-S** (−55 to +168) were inserted in front of the luciferase reporter gene in the constructs **pSARG-L-LUC** and **pSARG-S-LUC** respectively. Transient transfection of these constructs to LNCaP cells showed in both cases a very weak androgen response, indicative of the absence of strong **AR** binding sites (see Fig. 4d). This prompted us to screen for candidate AREs in a region of approximately 18 kbp, from 9 kbp upstream to 9 kbp downstream of the transcription start site using a bioinformatics-based approach. This sequence is present in BACs AC098935·2 and AC023534. The MatInspector program was applied to search both DNA strands for sequences homologous to the direct repeat 5′-TGTTCCT nnTGTCCTTCT-3′ or to the inverted repeat consensus ARE sequence 5′-A/GGA/TACAnnnTGT TCT-3′ (see Materials and methods). We selected sequences that showed at least nine out of 12 matches in the two half sites. Out of the sequences obtained, candidate AREs were further selected manually, based on the criteria of presence of at least three out of four underlined C or G residues in the direct repeat or presence of the double-underlined C and G residue and at least one of the two single-underlined C and G residues in the inverted repeat. Utilizing this approach, we identified 34 candidate AREs in the 18 kbp region, 12 inverted repeats and 22 direct repeats. None of these was completely identical to the consensus inverted repeat. One was a perfect direct repeat and two sequences deviated at one position from a perfect direct repeat. Four sequences matched the inverted or direct repeat at ten out of 12 positions. Several candidate AREs clustered in the genome. At approximately −8·5 kbp a cluster of four candidate AREs was present, including the imperfect direct repeat 5′-TGAACAatgAGA ACA-3′ (11/12 matches). At +4·6 kbp a cluster of five candidate AREs was detected, including the imperfect direct repeat 5′-TGTTCTctcTGT TCT-3′ (11/12 matches). The latter cluster is located in **SARG** intron 1. The perfect direct repeat 5′-TGTTCTctcTGT TCT-3′ mapped at −7·3 kbp, one ARE-like sequence was close to this repeat (Fig. 4a,b).

Next, it was investigated whether three genomic fragments containing the indicated direct repeats and flanking candidate AREs could function as enhancer regions. Genomic fragments with a size of approximately 500 bp, **SARG−8·5**,
SARG−7·3 and SARG+4·6, were coupled to PSA4-LUC, containing the 600 bp promoter of the PSA gene. This promoter was weakly responsive to androgens but combination with an upstream PSA enhancer fragment resulted in a strong androgen-inducible promoter (Cleutjens et al. 1997). In transfected LNCaP cells, SARG−8·5 and SARG−7·3 had no significant effect on the weak R1881 induction of the PSA4 promoter. In contrast, SARG+4·6 clearly increased R1881 induced PSA4 activity (Fig. 4c). Next, SARG+4·6 was linked to SARG-S (−55 to +168) and SARG-L (−3012 to +168), which both showed, as mentioned above, a very low androgen induction. Similar to the PSA promoter experiment, linkage of SARG+4·6 to both SARG-L and SARG-S showed androgen response in transfected LNCaP cells (Fig. 4d).

**SARG intron 1 contains a functional direct repeat androgen response element**

To determine whether the imperfect direct repeat 5′-TGTGCTaacTGT TCT-3′ in SARG+4·6 was responsible for androgen induction it was mutated to 5′-TATGATaacTAT TAT-3′. The mutated fragment was coupled to SARG-S and tested in LNCaP cells for its response to R1881. As shown in Fig. 5a, the androgen induction of SARG+4·6 was completely abolished by the mutations.

Next, the direct repeat 5′-TGTGCTaacTGT TCT-3′, designated ARE-SARG+4·6, was tested in an EMSA for its ability to bind to AR DBD (Fig. 5b). Control AREs were PSA ARE I (5′-AGAACAgcaAGTGCT-3′), which has been shown to bind strongly to the AR DBD, and rat probasin ARE II (5′-GGTTCTtggAGT ACT-3′), which is considered as a direct repeat, strongly interacting with AR DBD (Riegman et al. 1991, Rennie et al. 1993, Claessens et al. 1996, Cleutjens et al. 1996). ARE-SARG+4·6 bound to AR DBD, albeit more weakly than the PSA and probasin AREs. The change of ARE-SARG+4·6 into the perfect DR 5′-TGTGCTaacTGT TCT-3′ did not affect its capacity to bind AR DBD. AR DBD was unable to bind inactive mutant ARE-mSARG+4·6 (5′-TATGATaacTAT TAT-3′).

**Characterization of enhancer SARG+4·6**

To further decipher the role of the 569 bp enhancer SARG+4·6 (+4297/+4865) in androgen...
regulation, two deletion constructs of SARG+4·6-S-LUC were generated. Construct SARG+4·6-(+4447/4659) lacked all four candidate weak ARE
sequences present in the SARG+4·6 enhancer, but contained the imperfect direct repeat ARE (ARE-SARG+4·6). The 112 bp enhancer fragment SARG+4·6(+4548/4659) lacked even more up-
stream sequences, but also still contained ARE-
SARG+4·6. In transfection experiments the shortened enhancer SARG+4·6-(4447/4659) was less active than the 569 bp fragment, suggesting that clustering of ARE-SARG+4·6 with weak ARE-like sequences is important for full enhancer activity (Fig. 6a). Interestingly, further shortening of the enhancer completely abolished its activity, although ARE-SARG+4·6 was still present. However, in the deleted region we could not detect an obvious ARE-like sequence.

We carried out ChIP assays in order to investigate the in vivo function of enhancer SARG+4·6. Utilizing an antibody directed against acetylated histone H3, we observed a difference in H3 acetylation over SARG+4·6 between LNCaP cells grown in the presence and in the absence of R1881, showing a difference in chromatin structure on this part of the gene (Fig. 6b). The higher signal with AcH3 antibody in the presence of R1881 indicated an active structure of the enhancer region. Such a difference was not detected for the genomic fragments SARG−8·5 and SARG−7·3. These findings were in accordance with the transient transfection studies, as shown in Fig. 4c. Unfortunately, ChIP assays with a large series of different antibodies against the AR were not successful, probably due to the low affinity of AR for ARE-SARG+4·6.

ARE-SARG+4·6 is androgen receptor specific

To address the question of whether ARE-
SARG+4·6 is involved in androgen specificity, SARG+4·6 coupled to both SARG-S-LUC and PSA4-LUC was tested for activation by Dex. The constructs were transfected to LNCaP-1F5 cells cultured in the presence of 1 nM R1881 or 10 nM Dex or in the absence of hormone (Figure 7a,b). SARG+4·6 did not significantly stimulate Dex induced activity of SARG-S and PSA4. In contrast, R1881 induced activity of these two promoters was clearly increased by SARG+4·6.

ARE-SARG+4·6 was also tested in an EMSA for its ability to bind to GR DBD (Fig. 7c). Control PSA ARE I did bind to GR DBD, but rat probasin ARE II did not. Importantly, ARE-SARG+4·6

Figure 5 The direct repeat in SARG+4·6 is a functional androgen response element. (a) Effect of mutation of ARE-SARG+4·6 on the transcriptional activity of SARG+4·6. LNCaP cells were transfected with pSARG-S-LUC, pSARG+4·6-S-LUC or pSARG+4·6 m-SARG-S-LUC and incubated for 24 h with or without 1 nM R1881. Transcriptional activity was measured in a luciferase assay. Values +/− S.E.M. are from two experiments carried out in duplicate. Fold induction is the ratio of luciferase activity measured in the presence and absence of R1881. (b) EMSA of indicated AREs and AR DBD. ARE sequences are shown below the Figure. The arrowhead indicates the position of the AR DBD-ARE complex.
was also not able to bind to GR DBD, which correlated with the R1881 specificity of SARG+4·6 in the transfection assay.

**Discussion**

The androgen-specific regulated SARG gene was identified in the LNCaP-1F5 subline that expresses endogenous AR and GR from a stable-integrated cDNA expression vector (Cleutjens et al. 1997). We showed that SARG mRNA expression could be up-regulated by androgens, but not by glucocorticoids. SARG is a four-T exon gene of 14·5 kbp, mapping to chromosome band 1q32·2. Exon 1 can
appear in a short or long form, 1A and 1AB respectively. The 2-3, 3-3 and 5-5 kb transcripts result from alternative polyadenylation. Splice variants might lack either exon 2, part B of exon 1, or both. The predicted genes MGC2742 and MGC4309 are both part of SARG. SARG is preferentially expressed in the prostate (Fig. 1C). Our findings are substantiated by data in the expression profile database GeneNote (http://bioinformatics/weizmann.ac.il/cards/). In this database high SARG expression was only documented for prostate and lung.

The SARG ORF encodes a protein of 601 amino acids; splice variants lacking exon 2 are expected to code for a carboxyl-terminal fragment of 355 amino acids of the full-length protein. Transient transfection experiments showed that the SARG protein is located in the cytoplasm. No homology to other proteins was found. Unfortunately, the amino acid composition of SARG does not indicate motifs that could predict its function. The SARG mouse ortholog has 605 amino acids, with a homology to human SARG of 65%. Highest homology is in the amino-terminal and carboxyl-terminal regions of the proteins (data not shown).

To explain androgen specificity of SARG expression we first studied a 3 kb promoter region in transfection assays. Because these experiments were unsuccessful, we decided to carry out an in silico search for candidate AR binding sites in the 18 kb region. Without clustering as a selection criterion, the bioinformatics approach would not have been selective because of the high density of candidate AREs (one per 500 bp). We realize that functional AREs in enhancers and promoters might also cluster with binding sites for other transcription factors. Moreover, although less likely, it cannot be completely excluded that some candidate ARE sequences did not pass the selection criteria. One such ARE should be in SARG-S (−55 to +168), which is weakly androgen inducible. A candidate is the sequence 5'-GGGCCAggCAGCACA-3' (+5 to +17) in exon 1, which deviates at four positions from a perfect direct repeat.

A complicating factor in direct repeat ARE search is the lack of a consensus sequence for high-affinity, high-specificity AR binding due to the limited number of this type of AREs identified so far. Rat probasin ARE II (5'-GGTTCTtggAGT ACT-3'), which deviates from a perfect direct repeat at three positions, seems at present the specific ARE with highest AR affinity (Rennie et al. 1993, Kasper et al. 1994, Claessens et al. 1996, Kasper et al. 1999, Claessens et al. 2001). However, a search of the 18 kb SARG sequence did not detect candidate AREs closely resembling this sequence (data not shown). Our data provide the first evidence that an almost perfect direct 5'-TGT TCT-3' repeat, as present in SARG+4-6, can function as an AR-specific element in a natural enhancer. Other direct repeat-like functional AREs with variable AR specificity have been found in the
SC (secretory component) gene, the mouse Slp (sex limited protein) gene and the PEM (placenta and embryo) homeobox gene, which all deviate at least at three positions from a perfect direct repeat (Verrijdt et al. 1999, Verrijdt et al. 2000, Barbulescu et al. 2001).

Comparison of functional AREs of a large series of preferentially androgen regulated genes should reveal the sequence of a consensus high-affinity, high-specificity AR binding site in a natural context. Such genes might be identified by expression profiling of the AR and GR positive LNCaP-1F5 cell line followed by an unbiased functional study of a large series of overlapping fragments flanking the transcription start sites of these genes. This should also give a better insight in selection criteria for a bioinformatics-based search for functional AR binding sites in novel genes. Such an approach might also include comparison with data from other species.

It may be possible that the composition of an ARE can influence receptor activity by transduction of a particular conformation via the DNA-bound DBD to other AR domains. Recent evidence indicates that binding to different AREs indeed induces different conformational changes (Geserick et al. 2003). A different DBD conformation might directly or indirectly affect the association with coactivators, as shown for ER (Wood et al. 2001, Hall et al. 2002). In addition, it remains to be elucidated whether the ARE sequence is the major molecular determinant of AR specificity, or whether AR protein–protein interactions, including interactions with other specific transcription factors can contribute significantly to receptor specificity (Karvonen et al. 1997, Scheller et al. 1998).

Acknowledgement

This work was supported in part by a grant from the Dutch Cancer Society KWF. We thank Karin Hermans for technical support and art work.

References

Borns EM, de Boer W & Mulder E 1986 Androgen-dependent growth regulation of and release of specific protein(s) by the androgen receptor containing human prostate tumor cell line LNCaP. Prostate 9 247–259.
Cleutjens CB, Steketee K, van Eekelen CC, van der Korput JA, Brinkmann AO & Trapman J 1997 Both androgen receptor and glucocorticoid receptor are able to induce prostate-specific antigen expression, but differ in their growth-stimulating properties of LNCaP cells. Endocrinology 138 5293–5300.


Thornton JW 2001 Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. PNAS 98 5671–5676.


Received 4 June 2004
Accepted 5 July 2004