# CHARACTERIZATION OF THE

# ANDROGEN RECEPTOR TRANSCRIPTION UNIT

Karakterisatie van de

androgeenreceptor transcriptie-eenheid

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof. Dr. C.J. Rijnvos
en volgens besluit van het College van dekanen.

De openbare verdediging zal plaatsvinden op
woensdag 26 mei om 11.45 uur

door

Pieter Wiebren Faber
geboren in de Noordoostelijke Polder

Promotiecommissie

| | |
|---|---|
| Promotor: | Prof. Dr. F.T. Bosman |
| Overige leden: | Prof. Dr. D. Bootsma<br>Prof. Dr. J.A. Grootegoed<br>Prof. Dr. P.C. van der Vliet |
| Co-Promotor: | Dr. Ir. J. Trapman |

Gedrukt door: Drukkerij Haveka B.V., Alblasserdam

Dit proefschrift werd bewerkt binnen de afdeling Pathologie van de Faculteit der Geneeskunde en Gezondsheidswetenschappen, Erasmus Universiteit te Rotterdam.

Aan: mijn ouders
& Saskia

CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| Ad-MLP | :adenovirus major late promoter |
| AIS | :androgen insensitivity syndrome |
| AR | :androgen receptor |
| ARE | :androgen responsive element |
| AR-TIS I | :androgen receptor-transcription initiation site I |
| AR-TIS II | :androgen receptor transcription initiation site II |
| bp | :basepair |
| DBD | :DNA-binding domain |
| DHT | :5$\alpha$-dihydrotestosterone |
| CAT | :chloramphenicol acetyltransferase |
| cDNA | :complementary deoxyribonucleic acid |
| dhfr | :dihydrofolate reductase |
| EcR | :ecdysone receptor |
| ER | :estrogen receptor |
| G | :glycine |
| GR | :glucocorticoid receptor |
| GRE | :glucocorticoid responsive element |
| GTF | :general transcription factor |
| GUS | :glucuronidase |
| HRE | :hormone responsive element |
| Inr | :initiator |
| kb | :kilobase |
| LBD | :ligand-binding domain |
| LNCaP | :lymph node carcinome of the prostate (human) |
| MMTV-LTR | :mouse mammary tumor virus-long terminal repeat |
| MR | :mineralocorticoid receptor |
| mRNA | :messenger ribonucleic acid |
| NLS | :nuclear localization signal |
| PCR | :polymerase chain reaction |
| PR | :progesterone receptor |
| Q | :glutamine |
| RAR | :retinoic acid receptor |
| T | :testosterone |
| tau | :transactivation unit |
| TBP | :TATA binding protein |
| TdT | :terminal deoxynucleotidyltransferase |
| THR | :thyroid hormone receptor |
| VDR | :vitamin D3 receptor |
| UTR | :untranslated region |

# CHAPTER I

## STEROID HORMONE RECEPTORS

## &

## TRANSCRIPTION INITIATION BY RNA POLYMERASE II

## 1.1 STEROID HORMONE RECEPTORS

### 1.1.1 Introduction

Steroid hormones (androgens, estrogens, glucocorticoids, mineralocorticoids and progestins) are small hydrophobic, cholesterol derived molecules which exert profound effects in cellular development, differentiation and homeostasis. The physiological effect of steroid hormones is mediated by soluble intra-cellular receptor proteins that function as ligand-dependent transcription factors (Evans, 1988; Green & Chambon, 1988). Steroid hormone receptors are organized in and characterized by three relatively discrete functional domains (Fig. 1). These domains are a ligand-binding domain (LBD) in the C-terminal part of the protein, a DNA-binding domain (DBD) and an N-terminal domain which plays a role in transcription activation (Evans, 1988; Green & Chambon, 1988; Carson-Jurica et al., 1990; see also § 1.1.3). Steroid hormone receptors play key roles in the signal transduction pathways of their respective ligands. The receptor structure is changed upon ligand binding in a way which enables it to interact with high-affinity chromatin binding sites, termed hormone responsive elements (HREs), located in the transcriptional control regions of target genes. The activated DNA-bound receptor subsequently can modulate the transcription from this target gene (Beato, 1989, 1991; Wahli & Martinez, 1991; Gronemeyer, 1992).


FIGURE 1: Schematical representation of a steroid hormone receptor. Indicated are the three main domains: the N-terminal domain (broken line), the DBD (black box) and the LBD (hatched box).

Steroid hormone receptors form a subfamily of a larger superfamily of structurally related proteins, known as the nuclear receptor family. The members of this family contain a well conserved DBD, a moderateraly conserved LBD and show no conservation in the N-terminal domain. The homologies in the corresponding domains of the steroid hormone receptors (57-80% in the DBD, 20-55% in the LBD, <15% in the N-terminal domain which is also the least conserved in size) are schematically depicted in Fig. 2. The reported number of members of the family presently exceeds thirty and includes the receptors for the thyroid hormones, the hormonal forms of vitamin A (retinoids) and vitamin D, the *Drosophila* steroid ecdysone as well as a variety of receptor-like proteins for which no ligand has been identified as yet, and which are

collectively referred to as "orphan receptors" (Evans, 1988; Green & Chambon, 1988; Carson-Jurica et al., 1990; Fuller, 1991; Koelle et al., 1991; for a recent review on "orphan receptors" see O'Malley & Conneely, 1992).

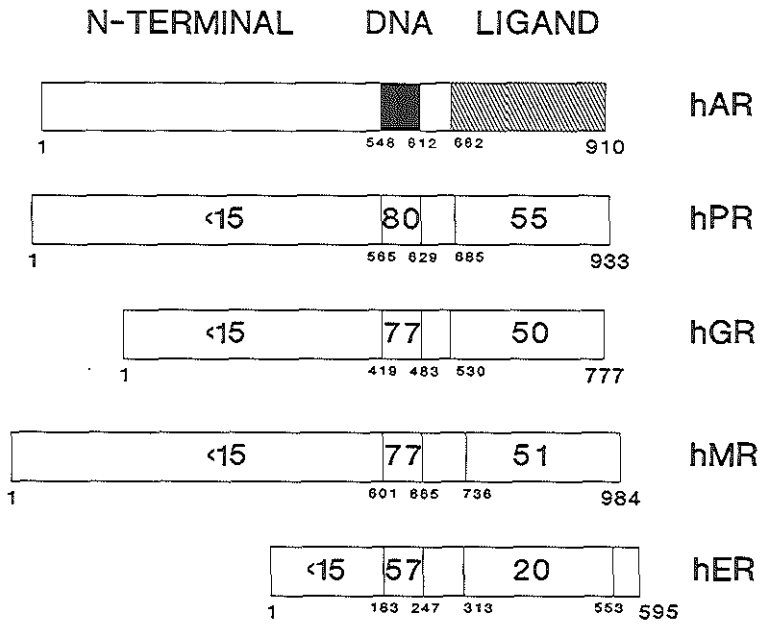


FIGURE 2: Structural homology in the N-terminal domain (open box), DBD (black box) and LBD (hatched box) of the human androgen receptor, progesterone receptor, glucocorticoid receptor, mineralocorticoid receptor and estrogen receptor (Trapman et al., 1988; Faber et al., 1989 [Chapter II]; Misrahi et al., 1987; Hollenberg et al., 1985; Arriza et al., 1987; Green et al., 1986)

Members of the nuclear receptor family originate from a single precursor protein as shown through the generation of phylogenetic trees for the two conserved domains, the DBD and the LBD (Laudet et al., 1992; Amero et al., 1992). Based on the phylogenetic tree for the DBD three subfamilies were defined that could have arisen from simple gene duplications of the common ancestor as schematically presented in Fig. 3. Subfamily I contains the receptors for the thyroid hormones (THR$\alpha$ and THR$\beta$) and retinoic acid (RAR$\alpha$, RAR$\beta$ and RAR$\gamma$). The chromosomal localization of THR$\alpha$ and RAR$\gamma$ (17q21) and THR$\beta$ and RAR$\beta$ (3p24) present an example of gene and chromosome duplications. RAR$\gamma$, which is located at 12q13 must be the result from a prior duplication event. Subfamily II contains the majority of the orphan receptors and represents the least characterized subfamily, whereas the steroid hormone receptors, the vitamin D3 receptor (VDR) and the *Drosophila* ecdysone receptor (EcR) are found in

subfamily III. The phylogenetic tree of the LBD generally confirmed these subdivisions, but also provides evidence that in addition to gene duplications, other events (recombination, translocation and exon shuffling) have taken place. As an example the subfamily III members VDR and EcR are shown which have DBDs from subfamily III but contain LBDs from subfamily I.

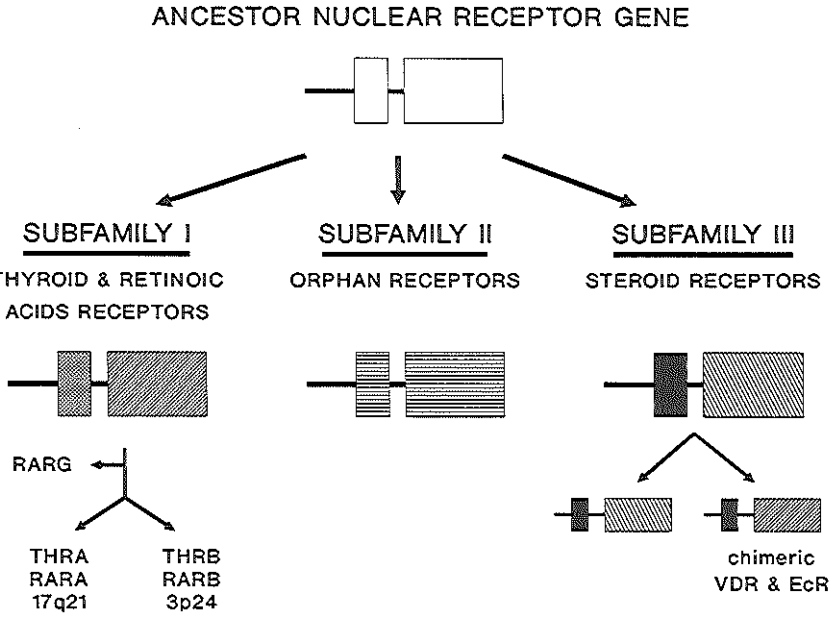


FIGURE 3: An evolutionary scenario for nuclear receptor genes, indicating the three nuclear receptor subfamilies. Subfamily I contains the THR and RAR, subfamily II contains the majority of the "orphan receptors", whereas the steroid hormone receptors are found in subfamily III. Adapted with permission from Laudet et al., 1992; EMBO J. 11, 1003-1013.

### 1.1.2 Genomic organization of steroid hormone receptor genes

The organization of steroid receptor genes has been elucidated for the human AR (hAR) and mouse AR (mAR) genes (Kuiper et al., 1989; Faber et al., 1991b [Chapters II and V]), the hER gene (Ponglikitmongkol et al., 1988), the hGR and mGR genes (Encío & Detera-Wadleigh, 1991; Strähle et al., 1992) and the chicken PR (cPR) gene (Huckaby et al., 1987; Jeltsch et al., 1990). In accordance with the observation that the receptors have a common ancestor and show relatively high conservation in their primary amino acid sequence, the organization is strikingly similar with respect to the exons containing protein coding information (Fig. 4). The N-terminal domain is encoded in exon 1, the DBD is encoded in the exons 2 and 3 and the information for the LBD is divided over the exons 4 to 8 with the positions of the individual splice sites being

conserved between the various receptors.

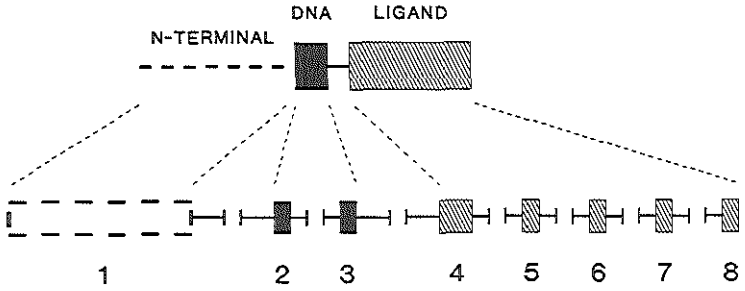


FIGURE 4: Schematical representation of the structural organization of steroid hormone receptor genes. Indicated is the distribution of the information for the three domains, N-terminal (broken line), DBD (black box) and LBD (hatched box) over the eight coding exons of the genes. For refs of the individual receptor genes, see text.

Human steroid hormone receptors are encoded by large genes which are localized on different chromosomes (summarized in Table I). In contrast to the similarity in the organization of the protein coding region, there is a marked heterogeneity in the organization of the genes with regard to the number of (identified) promoters and upstream untranslated exons (Fig. 5). The simplest organization is observed in case of the hAR, mAR and rAR genes which lack 5'-untranslated exons and have a single promoter region (Tilley et al., 1990; Faber et al., 1991a, 1991b [Chapters IV and V]; Baarends et al., 1990). One promoter region, upstream from exon 1, has been described for the hER gene (Green et al., 1986). However, cDNA clones, containing the 3'-part of an untranslated exon (exon 0) which splices to a position just upstream from the ATG translation initiation codon in exon 1 have also been characterized (Keaveney et al., 1991). In the mER gene this exon 0 has been characterized and a functional promoter was identified (White et al., 1987). No promoter upstream from exon 1, similar to the one found by Green et al. for the hER gene, has been described for the mER gene. One untranslated exon ($0^C$) and one promoter region have been described for the hGR gene (Zong et al., 1990; Encío & Detera-Wadleigh, 1991). However, for the corresponding mGR gene three untranslated exons ($0^{A,B,C}$), all of which splice to a position 13 bp upstream from the ATG translation initiation codon in exon 1, and three corresponding promoter regions were identified (Strähle et al., 1992). The hPR gene lacks 5'-untranslated exons but two functional promoter regions were identified as indicated in Fig. 5. The use of the upstream promoter generates a mRNA which encodes the full-length 933 amino acid hPR, whereas the use of the downstream promoter results in an mRNA which cannot encode this 933 amino acid hPR. Instead, translation starts at an in-frame ATG ($ATG_2$) resulting in a functional hPR lacking 165

amino acids at the N-terminal end (Kastner et al., 1990). The same promoter regions are probably functional in the cPR gene, which also generates two different functional PRs (Jeltsch et al., 1990). In contrast, in the rabbit PR gene only one promoter region is active which is in accordance with a single rabbit PR (Misrahi et al., 1988). So far, no information regarding the genomic organization and promoter regions of the MR gene has been presented.

TABLE 1

OVERVIEW OF THE CHROMOSOMAL LOCALIZATION AND LENGTH OF HUMAN STEROID HORMONE RECEPTOR GENES

| | CHROMOSOME | LENGTH |
|---|---|---|
| hAR | X | > 80 kb |
| hER | 6 | > 140 kb |
| hGR | 5 | > 80 kb |
| hMR | 4 | nd |
| hPR | 11 | nd |

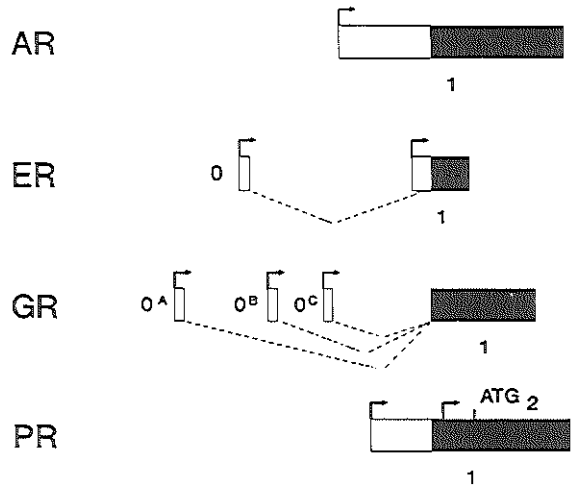


FIGURE 5: Overview of the 5' organization of steroid hormone receptor genes. Indicated are the (identified) upstream untranslated exons and promoters in the androgen receptor-, estrogen receptor-, glucocorticoid receptor-, and progesterone receptor genes (for refs, see text). Open boxes represent 5'-untranslated regions; black boxes represent open reading frames.

In summary, the use of alternative promoters and upstream exons in steroid hormone receptor genes can result in the generation of distinct mRNAs which encode functional full-length receptors, except in case of the hPR and cPR, where the use of an alternative promoter results in the synthesis of an N-terminally truncated receptor. The use of multiple promoters in a single gene provides transcriptional flexibility and is usually associated with tissue- and/or differentiation stage specific gene expression (Schibler & Sierra, 1987). This is in agreement with the differential expression patterns of the individual receptors. As studies regarding the molecular mechanisms of receptor mRNA expression have been scarce so far it is to be expected that the scheme depicted in Fig. 5 will not represent the full promoter potential of steroid hormone receptor genes.

### 1.1.3 Functional domains of steroid hormone receptors

Steroid hormone receptors are characterized by a unique modular structure which consists of three main functional domains and includes several functional subregions (Fig. 6). Structural comparisons between steroid hormone receptors (as presented in Fig. 2) followed by structure/function analysis of the hER and hGR provided the basis for the classical receptor domain structure: the moderately conserved C-terminal domain contains the ligand-binding function, the highly conserved cysteine-rich central domain mediates DNA-binding, whereas the non-conserved N-terminal domain is important to invoke the maximal transcriptional response from hormone-inducible promoters (Kumar et al., 1986, 1987; Giguère et al., 1986; reviewed in Evans, 1988; Green & Chambon, 1988; Carson-Jurica et al., 1990).

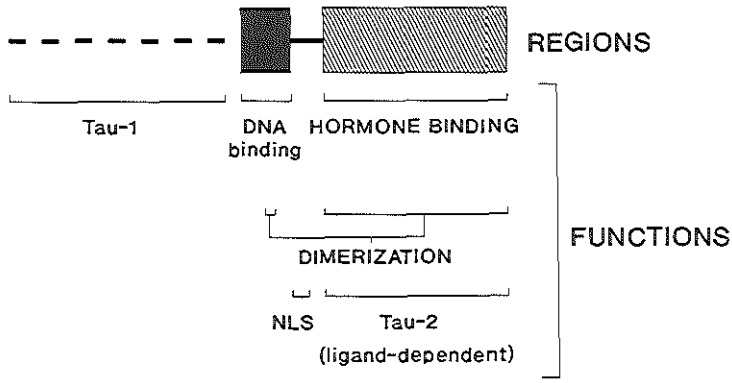


FIGURE 6: Schematic illustration of the structure/function organization of steroid hormone receptors. The N-terminal domain is shown as a broken line, the DBD and LBD as black and hatched boxes, respectively. Domain functions are depicted under the receptor scheme. tau, transactivation unit; NLS, nuclear localization signal.

Steroid hormone receptors interact as dimers with their cognate response elements, as determined for the hER and hPR (Kumar & Chambon, 1988; Tsai et al., 1988; Guiochon-Mantel et al., 1989) and subregions involved in protein-protein interactions leading to receptor dimerization have been identified in the DBD (Umesono & Evans, 1989) and in the LBD (Kumar & Chambon, 1988; Guiochon-Mantel et al., 1989; Fawell et al., 1990). A transactivation domain (tau-1) has been identified as part of the N-terminal domain for the hER, hGR, and hPR (Tora et al., 1989; Hollenberg et al., 1988; Meyer et al., 1990) (note that a tau domain is defined as a region which can function as a transcription activation domain when attached to a heterologous DBD). However, receptors deleted for their N-terminal domains were still able to activate transcription in a ligand-dependent way albeit at reduced revels compared to the full-length receptor. This observation led to the identification of a hormone-inducible transactivation function (tau-2) in the LBD (Webster et al., 1988; Meyer et al., 1990). Steroid hormone receptors require a nuclear localization signal (NLS) for nuclear import as their size exceeds the apparent exclusion limit for free diffusion through nuclear pores (60 kD, Peters, 1986). The subregion of steroid hormone receptors, involved in the translocation of the receptor to the nucleus, is located between the DBD and LBD as determined for the hGR, hPR, hAR and hER (Picard et al., 1987; Guiochon-Mantel et al., 1989; Simental et al., 1991; Jenster et al., 1991; Ylikomi et al., 1992). A conserved basic lysine/arginine rich subregion is present in all receptors. This region resembles the bipartite NLS, which consists of two stretches of basic amino acids with a 10 amino acid spacer, originally identified in the nucleoplasmin protein (Robbins et al., 1991; Dingwall & Laskey, 1991).

### 1.1.3.1 The DNA-binding domain

The distinguishing feature of steroid hormone receptors is the highly conserved DBD which mediates the sequence specific interaction of the recep- tor with the DNA (reviewed in Schwabe & Roades, 1991; Freedman, 1992). This domain contains several invariant residues (31 of 66 for the human receptors, see Fig. 7), which include eight cysteine residues (indicated with an asterisk in Fig. 7). Originally, the DBD of steroid hormone receptors was proposed to fold into two $Zn^{2+}$-coordinated "fingers", similar to those proposed for the *Xenopus laevis* transcription factor TFIIIA (Miller et al., 1985). In TFIIIA each $Zn^{2+}$-ion is coordinated by two cysteine and two histidine residues and defines a single structural unit which can interact with the DNA. However, through recent mutation/ deletion analysis and NMR/crystallographic analysis of the GR and ER DBD it was shown that the steroid hormone receptor DBD is folded differently (reviewed in Schwabe & Roades, 1991; Freedman, 1992).

Mutation analysis of the GR and ER DBD focussed on the finding that the consensus response elements for these receptors are related but different. Both receptors bind to a short fifteen bp palindromic sequence with a six bp halfsite and a

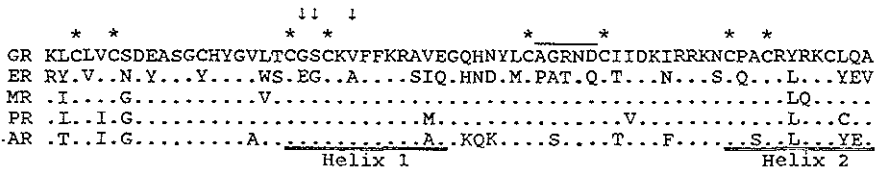


FIGURE 7: Overview of the DBD of the steroid hormone receptors. Indicated are the conserved cysteine residues (asterisk), the three amino acids that allow the receptor to discriminate between a GRE and an ERE (arrows), the D-box which is involved in protein-protein interactions (overlined) and the amino acids that form the the two $\alpha$-helices (doubly underlined). For refs. of individual receptors, see FIGURE 2.

spacing of three nucleotides, but differ at two positions per halfsite (Fig. 8; see Beato, 1991). In an effort to determine the amino acids responsible for this difference, receptors were constructed containing chimeric DBDs, with both ER and GR sequences. This approach initially led to the observation that the N-terminal "finger" contains the amino acids involved in sequence specific interaction of the receptor with the DNA (Green et al., 1988). Subsequently, three amino acids were identified in this region of the GR and ER DBD that are essential to the ability of the receptor to interact with its cognate response element ([GSxxV vs EGxxA], indicated with arrows in Fig. 7; Umesono & Evans., 1989; Mader et al., 1989). The elucidation of the three-dimensional structure of the GR and ER DBD by NMR spectroscopy and the crystallographic analysis of a GR DBD-DNA complex showed the steroid hormone receptor DBD to be folded into a novel $Zn^{2+}$ -coordinated structure (Härd et al., 1990; Schwabe et al., 1990; Luisi et al., 1991). The DBD consists of two $\alpha$-helices, perpendicular to each other (double underlined in Fig. 7). Each helix is anchored and oriented by a $Zn^{2+}$-ion at its N-terminus. The amino acids involved in receptor specific recognition of the HRE are located in helix 1. This recognition $\alpha$-helix contacts the major groove in the DNA and facilitates binding of the receptor to one halfsite of the HRE. A second receptor then binds cooperatively to the remaining halfsite. Although the second "finger" is not involved in specific DNA recognition, it contains a sequence, referred to as the D-box (overlined in Fig. 7) which facilitates cooperate binding of the DBD of two receptors to the DNA by protein-protein interactions leading to receptor dimerization (Umesono et al., 1989; Luisi et al., 1991).

GRE AGAACA nnn TGTTCT

ERE AGGTCA nnn TGACCT

FIGURE 8: Comparison of a glucocorticoid and estrogen response element. Shown are the two 6 bp halfsites and the 3 bp spacer. The two nucleotide difference between the elements is indicated (enlarged).

### 1.1.3.2 The ligand-binding domain

The second, relatively conserved domain is the LBD which is located in the C-terminal part of the receptor. This region has a length of approximately 250 amino acids and binds the ligand by the forming of a hydrophobic pocket. The integrity of the complete domain is important for steroid binding, as most deletions or point mutations in this region abolish ligand binding (Kumar et al., 1986; Giguère et al., 1986; Gronemeyer et al., 1987; Jenster et al., 1991). Deletion of the complete LBD leads to a truncated receptor which is constitutively transcriptionally active, suggesting that one of the functions of the LBD is to repress the activity of transactivating domains in other parts of the protein (Carson et al., 1987; Hollenberg et al., 1988; Picard et al., 1988; Tora et al., 1989; Jenster et al., 1991). The LBD contains a ligand-inducible transactivation function (tau-2) as demonstrated with chimeric activators, consisting of the GAL4 DBD and the either the hER, hGR or hPR LBD (Webster et al., 1988; Meyer et al., 1990). To investigate whether tau-2 is encoded in one of the five exons that encode the LBD (see Fig. 4), fusion proteins consisting of the GAL4 DBD and the individual hER exons were expressed. However, none of the fusion proteins could function as an activator (Webster et al., 1989). More likely, tau-2 corresponds to a protein surface created upon ligand-binding from several dispersed elements in the LBD. Mutational analysis of the mER LBD has identified a subregion near the C-terminal end, involved in receptor dimerization (Fawell et al., 1990), but the general importance of this region in steroid hormone receptor dimerization remains to be confirmed for other receptors.

### 1.1.3.3 The N-terminal domain

The third major structural region is the N-terminal domain, which is the least conserved, both in size and in sequence (see Fig. 2). Originally referred to as the immuno-reactive domain, as most antibodies generated against receptors recognized epitopes in this domain, a tau-region (tau-1) has now been identified in the N-terminal domains of the hGR, hER and hPR (Hollenberg et al., 1988; Tora et al., 1989; Meyer et al., 1990). Deletion analysis of the N-terminal domains has established that the tau-1 regions are located at different positions and structurally differ (Fig. 9) (Hollenberg et al., 1988; Tora et al., 1989; Meyer et al., 1992). The N-terminal domain of the hGR contains an acidic tau-1 region although subsequent experiments indicated that additional non-acidic transactivation functions might exist outside this defined tau-1 region. The tau-1 region of the hPR is located just upstream from the DBD in a proline-rich amino acid region. A region, located between amino acid 1 and 165, which corresponds to the difference of the two functional hPR isoforms, cannot function independently from the remainder of the N-terminal domain but has a modulatory function as the absence of this region results in a partial loss of the transactivation potential of tau-1. No subregions have been defined in the short N-terminal domain of the hER but as this region lacks acidic stretches and proline-clusters it must be

structurally different from the hGR and hPR tau-1 region.

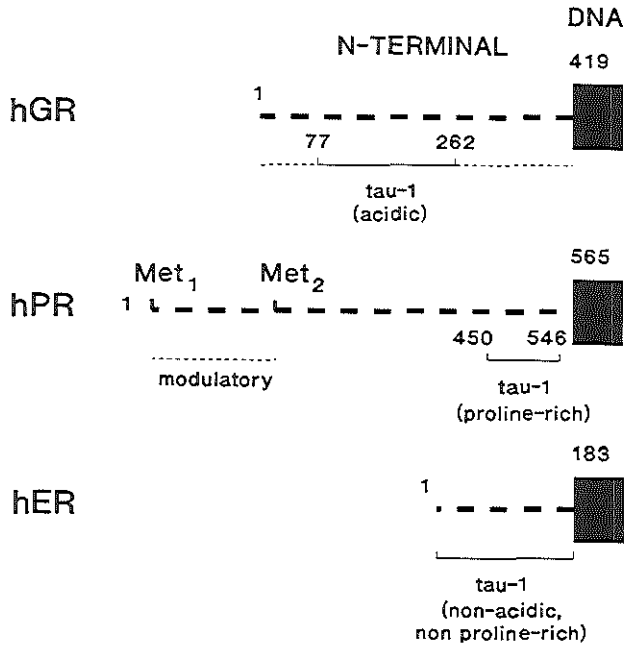


FIGURE 9: Overview of the identified tau-regions in the N-terminal domains of the human glucocorticoid receptor, progesterone receptor and estrogen receptor. Indicated are the positions and structural features of the the tau-regions.

**1.1.4.1** Isolation and characterization of cDNA clones encoding the androgen receptor

The androgen receptor (AR) is the steroid hormone receptor which mediates the intra-cellular response to androgens, the male sex steroids. The principal androgen, testosterone (T), is produced by the testis. In some tissues T is reduced to its more potent metabolite 5$\alpha$-dihydrotestosterone (DHT). Androgen action affects a wide variety of tissues but exerts its most pronounced effect on the development and maintenance of the male phenotype (reviewed in Mooradian et al., 1987).

Because no anti-AR antibodies or structural data on the AR protein were available, cDNA clones encoding the AR could not be obtained by conventional isolation techniques. To circumvent this problem it was assumed that the putative AR DBD would be structurally related to the DBDs of the hGR, hER and hPR, the sequences of which were available and highly similar (see Figs 2, 7 and 10). Oligonucleotides, corresponding to the most conserved region of these DBDs were synthesized and used to screen

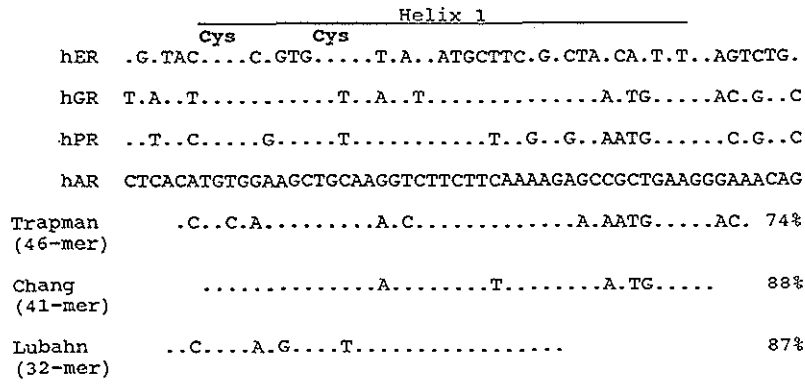


FIGURE 10: Overview of the oligonucleotides, used to isolate human and rat androgen receptor cDNA clones (Trapman et al., 1988; Chang et al., 1988a; Lubahm et al., 1988a). The oligonucleotides are compared with the sequences of the human estrogen receptor, glucocorticoid receptor and progesterone receptor and the actual sequence of the human androgen receptor.

cDNA libraries. The various oligonucleotides are depicted in Fig. 10 and compared to the actual AR nucleotide sequence (note that this conserved region is the recognition $\alpha$-helix in the DBD [overlined sequence]). This approach was successfully and independently applied to isolate partial hAR and rat AR (rAR) cDNA clones (Chang et al., 1988a; Lubahn et al., 1988a; Trapman et al., 1988). Next, the complete hAR and rAR cDNA sequences were reported (Chang et al., 1988b; Lubahn et al., 1988b; Tan et al., 1988; Faber et al., 1989 [Chapter II]; Tilley et al., 1989). Using the reported hAR and rAR cDNA sequences, the mouse AR (mAR) cDNA was isolated by a combination of conventional screening techniques and polymerase chain reaction (PCR) methods (He et al., 1990; Gaspar et al., 1990; Faber et al., 1991b [Chapter V]; Charest et al., 1991).

The primary amino acid sequence of the hAR as deduced from the nucleotide sequence of the isolated cDNA clones contains the three structural domains, characteristic of the nuclear receptors (see Fig. 2) with high homology between the AR, GR and PR (80% in the DBD, 50 % in the LBD) and moderate homology between the AR and ER (57% in the DBD, 20% in the LBD). A homology comparison of the hAR, mAR and rAR shows high conservation (100%) in the DBD and LBD and over 75% conservation in the large N-terminal domain (Fig. 11, taken from Chapter V). The N-terminal domain of the AR is remarkable as it contains a variety of homopolymeric stretches, including very long Q- and G-stretches. The position and the presence of these stretches, however, is species dependent as the Q-stretch in the rAR is located at a more C-terminal position than the hAR Q-stretch. The mAR Q-stretch, which is at the rAR position, is intermingled with three histidine residues. A long G-stretch is absent in the rAR and mAR, as they harbor at the identical position of the hAR G-stretch, only five

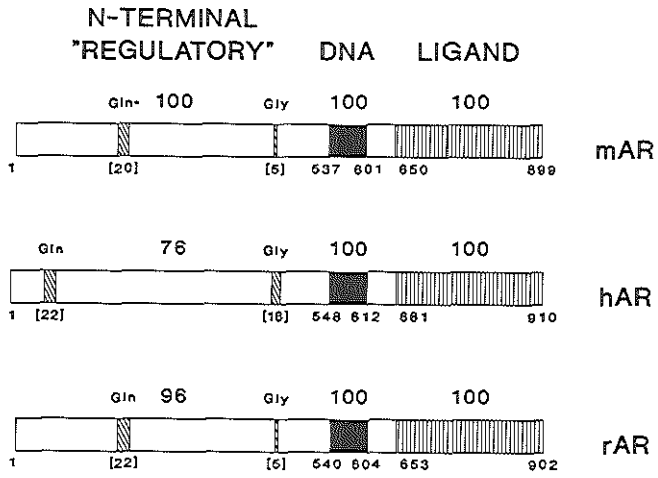


FIGURE 11: Schematic comparison of the mouse, human and rat androgen receptor. Indicated are the sequence similarities in the N-terminal domain, the DBD and the LBD. Also indicated are the position and composition of the Q- and G-stretches in the N-terminal domain. Gln* indicates the intermingling of the seventeen Q-residues with three histidine residues in the mAR Q-stretch.

glycine residues. This situation is reminiscent of a Q-stretch in the GR, which has a length of 19 amino acids in the rGR, 8 amino acids in the mGR and is absent from the hGR (Miesfeld et al., 1986; Danielsen et al., 1986; Hollenberg et al., 1985). The differences in the length of these stretches account for most of the variation in length between the hAR, mAR and rAR proteins which is 910, 899 and 902 amino acids, respectively. In addition there is an intra-species specific variation in receptor length which is most pronounced for the hAR for which protein lengths of 910, 918, 919, and 917 amino acids have been reported (Faber et al., 1989 [Chapter II]; Chang et al., 1988b; Lubahn et al., 1988b; Tilley et al., 1989). This difference is due to a natural variation in the length of the Q- and G-stretches (Sleddens et al., 1992; Edwards et al., 1992; Sleddens, 1993). The variation in the reported length is less in case of the rAR where 902 and 901 amino acids have been reported, resulting from the absence of one glutamine residue in the Q-stretch (Chang et al., 1988b; Tan et al., 1988). For the mAR only one size (899 amino acids) has been observed (He et al., 1990; Caspar et al., 1990; Faber et al., 1991b [Chapter V]; Charest et al., 1991).

### 1.1.4.2 Functional analysis of recombinant AR

The properties of the recombinant hAR were investigated using transient

transfections of cultured mammalian cells with eukaryotic hAR expression vectors. This approach showed that the recombinant hAR was undistinguishable from the native AR with respect to ligand-binding and electrophoretic mobility on SDS-PAGE gels (Trapman et al., 1988; Brinkmann et al., 1989; Tilley et al., 1989; Quarmby et al., 1990). The ability of the hAR to function as a transcription factor was investigated in co-transfection experiments of the AR expression plasmid with reporter plasmids containing the androgen-responsive MMTV-LTR as well as a synthetic promoter (Schüle et al., 1988) which contains two copies of a consensus GRE upstream from the TK promoter linked to the chloramphenicol acetyl transferase (CAT) reporter gene. These experiments showed that recombinant hAR could function as a ligand-dependent transcription factor on these promoters (Brinkmann et al., 1989; Quarmby et al., 1990; Rundlett et al., 1990). The use of eukaryotic rAR and mAR expression vectors indicated little difference between the hAR, rAR and mAR in these kinds of experiments (Yarbrough et al., 1990; Faber et al., 1991b [Chapter V]; He et al., 1991). In addition to the functional activity in mammalian cells, the hAR has also been shown to function as a transcription factor in yeast cells in a ligand-dependent manner (Purvis et al., 1991) which is consistent with the functional activity of other members of the nuclear receptor family in yeast (see Srinivan, 1992 for refs).

Subsequent structure/function analysis confirmed the domain structure of the hAR which originally was based only on homology comparisons to other receptors. The C-terminal domain has the ligand-binding capacity, the cysteine-rich "zinc-finger" domain mediates DNA binding and regions important for transactivation are located in the N-terminal domain and the LBD, although it remains to be established if these regions are bona-fide tau-domains and can function when linked to a heterologous DBD (Rundlett et al., 1990; Jenster et al., 1991; Simental et al., 1991). Although eukaryotic expression systems are useful, they do not provide sufficient amounts of protein for in vitro experiments. Therefore bacterial and insect expression systems are now being used to express full-length and partial AR sequences for additional studies regarding AR struction/function (Young et al., 1990; Chang et al., 1992; Xie et al., 1992; Roehrborn et al., 1992).

#### 1.1.4.3 Androgen-regulated gene expression

The three amino acids in the first "zinc finger" of the steroid receptor DBD that are essential for the recognition of the cognate response element are similar for the AR, GR, PR and MR (see Fig. 7), indicating that these receptors could activate transcription through similar response elements. This had prior to the analysis of the receptor DBD already been inferred from the findings that endogenous AR in the human mammary tumor cell line T47D had the ability to activate transcription on the Mouse Mammary Tumor Virus (MMTV) Long Terminal Repeat (LTR) through elements defined as GREs and PREs. This was established by mutation analysis of the GREs as well as the use of

a synthetic promoter, consisting of two GREs, linked to the TK promoter (Darbre et al., 1986; Cato et al., 1987; Ham et al., 1988; Gowland et al., 1989) (note that the MMTV-LTR and the synthetic promoter were later used to establish that recombinant AR could function as a ligand-dependent transcription factor in §1.1.4.2).

So far only a small number of cellular genes have been characterized which are regulated by androgens. Among the best characterized examples are the rat Prostatic Binding Protein (PBP) gene and the human Prostate Specific Antigen (PSA) gene, where nuclear run-on experiments have demonstrated a direct effect of androgens on the transcription rates of these genes (Page & Parker, 1982: Riegman et al., 1991; Wolf et al., 1992). Analysis of the transcriptional control regions identified AREs which are similar to the GRE (Fig. 12). Additional examples of androgen-regulated genes are the murine $\beta$-Glucuronidase (GUS) gene (Lund et al., 1990; reviewed in Paigen, 1989), the murine Sex-limited protein (Slp) gene (Adler et al., 1991, 1992) and the human blood clotting Factor IX gene (Crossley et al., 1992). Sequences, similar to the GRE were identified in all these genes (summarized in Fig. 12). Evidence that these sequences constitute the ARE for the GUS and Slp genes comes from the finding that in a mouse strain deleted for the ARE, GUS mRNA levels are not influenced by androgens and that the Slp ARE can mediate transcriptional regulation by the AR in transient transfection experiments. In accordance with the ability of the AR to activate transcription through GREs it has been shown that the PBP, Slp and PSA ARE can mediate activation by the GR (Claessens et al., 1988; Adler et al., 1992; Cleutjens, unpublished).

| | |
|---|---|
| "consensus GRE" | AGAACA nnn TGTTCT |
| PBP ARE | AGTACG tga TGTTCT |
| PSA ARE | AGCACT gca TGTTCT |
| Slp ARE | AGAACA gcc TGTTTC |
| Factor IX ARE | AGCTCA gct TGTATC |
| GUS ARE | AGTACT tgt TGTTCT |
| "consensus ARE" | GGAACA nnn TGTTCT<br>T |

FIGURE 12: The ARE is similar to a GRE. Overview of identified ARE sequences in comparison with the consensus GRE and ARE (see text for refs).

To determine if the limited number of AREs identified so far are a true representation of the actual ARE, a DNA-binding site selection assay was performed (Roche et al.,

1992). From the results a consensus ARE (GGA/TACAnnnTGTTCT) was compiled, which again is similar to the consensus GRE (AGAACAnnnTGTTCT) (Fig. 12). The observation that the ARE and GRE, and in fact also the PRE and MRE are similar in nucleotide sequence, combined with the differential gene networks that are activated by the various receptors, suggests that additional levels of regulation, such as tissue-specific receptor expression, protein-protein interactions and promoter context determine the precise pattern of steroid induced gene expression.

1.1.4.4 Androgen receptor structural defects

The gene encoding the AR is located on the X-chromosome. Thus the effect of mutations in the AR will become manifest in 46,XY male individuals. The elucidation of the genomic organization hAR gene (Kuiper et al., 1988 [Chapter III]) provided the opportunity to detect structural defects in the AR through amplification of separate exons in PCR fragments. Strong evidence linking structural defects in the AR to the observed phenotype so far have been presented for the X-linked androgen insensitivity syndrome (AIS) and Kennedy's disease (Fig. 13).

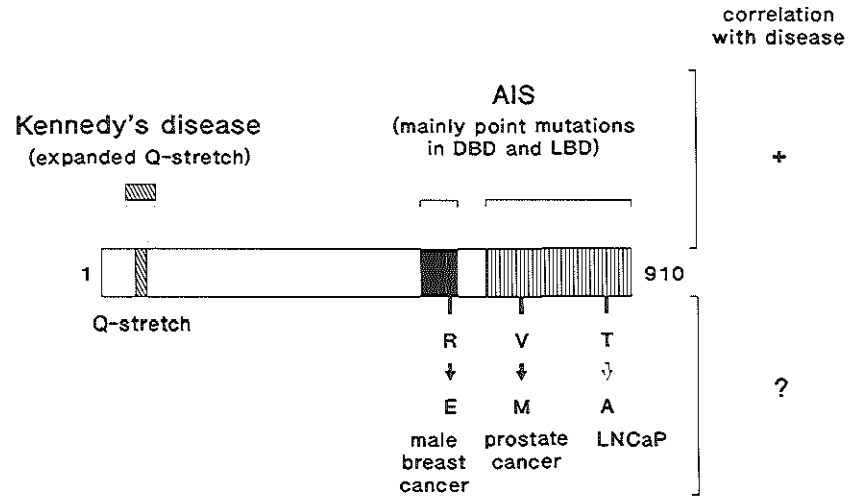


FIGURE 13: Overview of mutations in the hAR. Shown is a schematical representation of the 910 amino acid hAR with the N-terminal domain (open box), DBD (black box) and the LBD (hatched box). Over the hAR are indicated the mutations which correlate with disease. The mutations under the hAR might correlate with disease, but this remains to be established.

AIS results in abnormal male development and differentiation (Griffin & Wilson, 1987; Pinsky & Kaufman, 1987) and can vary from a complete female phenotype in a 46,XY individual (complete AIS) to partial disorders of male sexual differentiation

(partial AIS). Analysis of the hAR primary structure from mainly complete AIS patients by our group (Ris-Stalpers et al., 1990, 1991) as well as by others has identified a plethora of mutations in the hAR gene (reviewed in McPhaul et al., 1991; Brinkmann & Trapman, 1992; Pinsky et al., 1992). Although in some cases chromosomal deletions in the hAR gene have been identified, the majority of defects consists of point mutations in the AR LBD and DBD. Functional analysis of the mutations showed that in all cases correct AR function was impaired. In addition, both an AIS rat and mouse strain have been shown to contain inactivating mutations in their respective AR genes (Yarbrough et al., 1990; Gaspar et al., 1991; He et al., 1991).

In Kennedy's disease, a motor neuron disease resulting in bulbar and spinal muscular atrophy, abnormalities in the length of the CAG-repeat, encoding the Q-stretch in the AR N-terminal domain, correlate with the presence of the disease. Whereas the normal repeat number is between 19 and 30, affected patients showed numbers between 40 and 52 in every case that was examined (La Spada et al., 1991).

An additional mutation in the hAR has been identified in the human prostate tumor cell line LNCaP, where a point mutation is present in exon eight that results in a threonine to alanine change (Fig. 13, Veldscholte et al., 1990a). Functional analysis of this mutation shows that it accounts for the previously observed altered binding specificity for steroids of the LNCaP AR which has an increased preference for oestradiol and progestins compared to the native AR (Veldscholte et al., 1990b). In addition, in the LNCaP cell line, androgen regulated gene expression can now also be stimulated at low anti-androgen concentrations. However, it remains unclear if the mutation was present in the original tumor material from which the LNCaP originated. In addition to the mutation in LNCaP cells, recently a mutation in a primary prostate tumor has been described (Fig. 13, Newmark et al., 1992). Twenty six independent samples were examined and one point mutation in exon four was found, resulting in a valine to methionine change. This mutation, however, remains to be functionally characterized.

In addition to the systematic studies of patient material described above, a case report describing a germ-line point mutation in exon 3 of the AR in two brothers with partial AIS has been described (Fig. 13, Wooster et al., 1992). This mutation results in an arginine to glutamic acid change. Interestingly, both brothers also appeared with a breast tumor, suggesting a correlation between male breast cance and AR mutations.

## 1.2: TRANSCRIPTION INITIATION BY RNA POLYMERASE II

### 1.2.1 Introduction

Transcription initiation by RNA polymerase II is a major control point in the regulation of gene expression. The purified RNA polymerase II enzyme has no intrinsic affinity for (promoter) DNA and transcribes double-stranded DNA poorly, initiating only at single-stranded regions [see Sawadogo & Sentenac (1990) for a review]. The development of cellfree transcription systems, allowing the analysis of fractionated cellular extracts, led to the identification of two major classes of transcription factors that function in concert with RNA polymerase II to regulate proper initiation (schematically presented in Fig. 14) (Matsui et al., 1980; Davidson et al., 1983; Samuels et al., 1982; Dynan & Tjian, 1983).

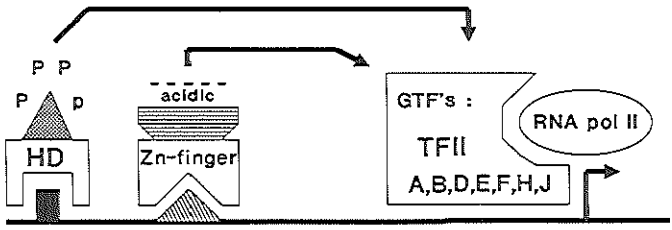


FIGURE 14: Two classes of transcription factors are involved in regulating transcription initiation. Indicated are the GTF's and RNA polymerase II, which form the basal transcription complex in the vicinity of the transcription start site. Also indicated are gene-specific factors, which interact through their DBD (examplified with the HD and "zinc-finger" DBDs) with the DNA in a sequence-specific manner and stimulate the transcription frequency through their activation domains (examplified with the proline-rich and acidic activation domains).

The first class consists of proteins which form, together with RNA polymerase II, the transcription initiation complex on the promoter. These factors (TFIIA,-B,-D,-E,-F,-H,-J) are active on most, if not all promoters and are referred to as the general transcription factors (GTFs) (Reinberg et al., 1987; Reinberg & Roeder, 1987; Maldonado et al., 1990; Inostroza et al., 1991; Cortes et al., 1992; reviewed in Mermelstein et al., 1989; Saltzman & Weinmann, 1989; Sawadogo & Sentenac, 1990; Zawel & Reinberg, 1992). The second class consists of sequence specific DNA binding factors which function in a more restricted manner. They are not required for basal levels of transcription but instead stimulate the transcription frequency through enhancement of the rate of initiation and/or the formation of a more stable transcription initiation complex. These transcription factors are referred to as gene-specific transcription factors. They are characterized by a modular structure and contain distinct domains involved in DNA binding and transactivation (reviewed in Ptashne, 1986; Maniatis et al., 1987; Mitchel

& Tjian, 1989, Johnson & McKnight, 1989). In addition, several different structural DBDs and transactivation domains have been characterized, of which the "zinc-finger" and homeodomain (HD) DBDs and the proline-rich and acidic transactivation domains are shown as examples in Fig. 14. The DBD positions these factors on the DNA whereas their transactivation domains allows them to "communicate" directly or indirectly with each other and the GTFs to modulate levels of transcription initiation.

The general transcription machinery interacts with promoter DNA and positions RNA polymerase II for proper start site selection. Two distinct DNA-elements can be defined, which constitute minimal promoters and are capable of directing the formation of a competent transcription initiation complex independently of additional promoter sequences (Fig. 15). The first element consists of an A/T rich sequence, the TATA-box (consensus TATAAA) which directs transcription initiation from an approximately 25-30 bp downstream position (Breathnach & Chambon, 1981). The second element overlaps with the site of initiation, lacks a well defined consensus sequence as yet, and is referred to as an initiator (Inr) (Smale & Baltimore, 1989; Weis & Reinberg, 1992). An Inr seems to be present in all TATA-less promoters, including those originally referred to as GC-rich promoters, which are found in housekeeping genes and for which the transcription factor Sp1 (consensus binding sequence GGGGCGGGGC) was thought to be responsible for the scattered patterns of initiation observed on such promoters (Fig. 16) (Dynan, 1986). Analysis of a canonical GC-rich promoter, the dihydrofolate reductase (dhfr) promoter, has shown that although the Sp1 binding sequences are important parameters of the levels of initiation the site of initiation is determined by an Inr (Blake et al., 1990; Means & Farnham, 1990; Weis & Reinberg, 1992).

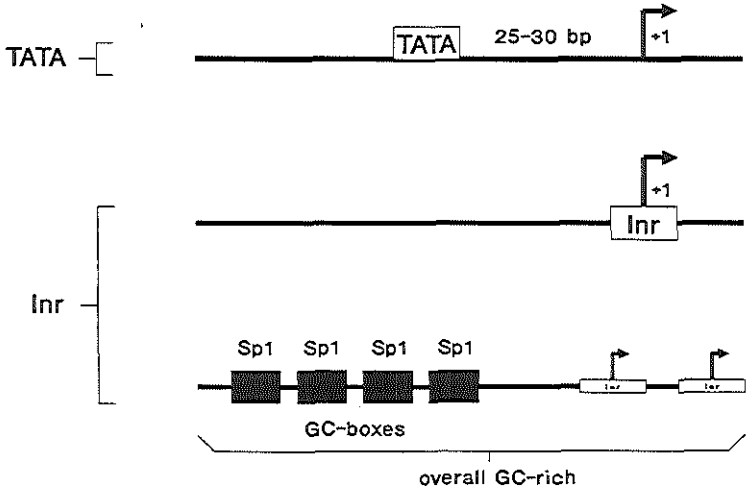


FIGURE 15: Two classes of DNA elements can constitute a minimal promoter. The TATA box directs initiation from a 25-30 bp downstream position whereas the Inr directs initiation from a position within itself.

### 1.2.2 Mechanisms of transcription initiation

#### 1.2.2.1 Transcription initiation on TATA-containing promoters

RNA polymerase II needs auxilliary proteins to catalyze accurate transcription. Todate seven such factors (TFIIA, -B,-D,-E,-F,-H,-J) have been identified, primarily through analysis of the adenovirus major late promoter (Ad-MLP) (for a recent review see Zawel & Reinberg, 1992). Only one of these factors, TFIID, can specifically interact with DNA and recognizes the TATA-box (Parker & Topol, 1984). The additional factors, including RNA polymerase II, enter the complex through protein-protein interactions. The assembly of the GTFs into the initiation complex proceeds in a highly ordered fashion. The time of entry of a specific factor into the complex can be elegantly visualized by DNA binding assays, as originally performed by Buratowski et al., 1989 (see also Zawel & Reinberg, 1992). Native gels are used to resolve complexes between the GTFs and labelled promoter DNA. TFIID is the first factor to associate with the DNA through a sequence specific interaction with the TATA-box. The TFIID-DNA complex provides the entry point for TFIIA and TFIIB, which results in a more stable DNA-bound TFII-ABD complex. Subsequently, RNA polymerase II is directed into this complex by TFIIF. Next TFIIE and TFIIH enter the complex through an interaction with RNA polymerase II, finally TFII-J enters.

The recent isolation and characterization of cDNA clones encoding several of the GTFs (TFIID, TFIIB, TFIIE, TFIIF), in combination with the use of highly purified GTF-protein fractions, has enabled studies addressing the mechanism of transcription initiation by RNA polymerase II. The major breakthrough, however, was the isolation of cDNA clones encoding the TATA-binding component (TBP) of TFIID (Greenblatt, 1991 for refs.). Functional studies with recombinant TBP showed that it could interact with the TATA-box in a sequence specific manner and could functionally substitute for native TFIID in the process of basal transcription initiation (Peterson et al., 1990; Pugh & Tjian, 1990, 1991; Dynlacht et al., 1990; Kelleher et al., 1992). Recombinant TBP, however, was not able to mediate the response to any of the gene-specific transcription factors tested todate, in contrast to a highly purified cellular TFIID fraction (Peterson et al., 1990; Pugh & Tjian, 1990, 1991; Dynlacht et al., 1990). This finding, combined with the observation that the molecular weight of TBP (38 kD) is much smaller than the estimated size of TFIID (> 100 kD) led to the identification of TFIID as a multiprotein complex consisting of TBP and several tightly bound TBP-associated factors (TAFs) (Dynlacht et al., 1991; Tanese et al., 1991; Pugh & Tjian, 1992). Whereas TBP can sustain basal levels of transcription, the TAFs seem essential to mediate the response to gene-specific transcription factors.

In addition to cDNA clones encoding TBP, TFIIB- (Ha et al., 1992), TFIIE- (Peterson et al., 1991) and TFIIF encoding cDNA clones (Sopta et al., 1989; Aso et al., 1992; Finkelstein et al., 1992) have been isolated. In contrast to TFIID, recombinant TFIIB,

TFIIE and TFIIF can functionally substitute for the corresponding purified cellular fractions in gel mobility shifts, basal transcription and activated transcription. The mechanism of action of each of these GTFs, however, remains to be established.

### 1.2.2.2 Transcription initiation on TATA-less promoters

The formation of a functional transcription initiation complex is nucleated by the assocation of TFIID with the TATA-box. However, a great variety of promoters do not contain a TATA-box, but yet maintain accurate levels of transcription initiation. The breakthrough in the analysis of transcription initiation on TATA-less promoters came with the study of the TATA-less terminal deoxynucleotidyltransferase (TdT) gene promoter, where a 17 basepair fragment, overlapping the +1 initiation site was identified as being the element, which could direct accurate transcription initiation from the +1 site, independently from additional promoter sequence (Smale & Baltimore, 1989). This element, which was later constricted to an eight bp (-3/+5) fragment, was referred to as the initiator (Inr, see Fig. 16).

| | |
|---|---|
| TdT | $_{-5}$CCCTCA$^{+1}$TTCT$_{+5}$ |
| dhfr | $_{-11}$ATTTCGCGCCAA+1ACTT$_{+5}$ |
| PGBD | $_{-1}$CA$^{+1}$ $_{+5}$TCCTGGTTAC$_{+14}$ |
| Ad-P5 | $_{-4}$CTCCA$^{+1}$TTTT$_{+5}$ |
| rp | $_{-4}$CTTCC$^{+1}$CTTTTCC$_{+8}$ |
| Ad-MLP | $_{-5}$TCCTCA$^{+1}$CTCT$_{+5}$ |
| Ad-IVa2 | $_{-5}$GTCTCA$^{+1}$GAGT$_{+5}$ |
| AR-TIS I | $_{-5}$CTCCCA$^{+1}$GCGC$_{+5}$ |
| AR-TIS II | $_{-5}$CCCTCC$^{+1}$GAGA$_{+5}$ |

FIGURE 16: AR-TIS I and AR-TIS II are structurally related to an Inr. Shown is a comparison of identified Inr sequences with the two transcription initiation sites of the hAR promoter, AR-TIS I and AR-TIS II.

Examination of additional TATA-less promoters identified functional Inr elements in the dhfr-gene (Blake et al., 1989; Means & Farnham, 1990), the porphobilinogen deaminase (PBGD) gene (Beaupain et al., 1990), the adeno-associated virus type 2 P5 (Ad-P5) promoter (Shi et al., 1991) and a mammalian ribosomal protein (rp) gene (Hariharan et al., 1990). Interestingly, the presence of an Inr sequence is not restricted to TATA-less promoters as functional Inr sequences have been identified on the TATA-containing Ad-ML promoter (Roy et al., 1991) and the Ad-IVa2 promoter (Carcamo et

al., 1990). Although the identified elements share a preponderance of pyrimidine-residues, no overall consensus sequence could be constructed. This observation leads to the suggestion that the Inr is a heterogenous element and based on their nucleotide sequence several Inr subfamilies were postulated (reviewed in Weis & Reinberg, 1992).

Regarding the mechanism of Inr-mediated transcription initiation, efforts have been focussed on the identification of protein factors, essential to this process. Using in vitro transcription experiments, it was shown that the GTFs, as defined for transcription initiation on TATA-containing promoters and including TFIID, are necessary to sustain basal and stimulated levels of initiation on synthetic promoters containing the TdT Inr (Smale et al., 1990; Pugh & Tjian, 1991). The assumption that protein-Inr interactions would provide the basis for the formation of the transcription initiation complex, led to the search for proteins capable of sequence specific interaction with the respective Inr sequences. So far, proteins capable of interacting with the Inr of the dhfr (Blake et al., 1989; Means & Farnham, 1990), the Ad-P5 (Shi et al., 1991), and the Ad-ML promoter (Roy et al., 1991) have been identified. Characterization of these proteins confirmed the heterogeneity of the Inr as it resulted in the identification of the transcription factors E2F (also called HIP1) and YY1 as the factors interacting with the dhfr- and Ad-P5 Inr, respectively (Blake et al., 1989; Means & Farnham, 1990; Shi et al., 1991), whereas an as yet uncharacterized protein, TFII-I, interacts with the AD-MLP Inr (Roy et al., 1991). So far, however, the link between these proteins and the GTFs, resulting in Inr-mediated transcription initiation remains unclear.

#### 1.2.2.4 Transcription initiation on the hAR promoter

The promoter of the AR gene is located approximately 1 kb upstream from the translation initiation codon in exon 1 and has been identified and characterized in the human, mouse and rat species (Tilley et al., 1990; Baarends et al., 1990; Faber et al., 1991a, 1991b [Chapters IV and V]]). Two major transcription initiation sites (AR-TIS I and AR-TIS II) are located in a 13 bp pyrimidine-rich region. No direct sequence homology exists between AR-TIS I, AR-TIS II and any of the previously described Inr sequences, but both AR-TIS I and AR-TIS II are structurally related to those Inr sequences (see Fig. 16). Additional structural elements with regulatory potential consist of a short GC-box containing an Sp1 binding sequence and a long homopurine stretch which is variable in length between the hAR, mAR and rAR promoter based on the presence of 4, 6 and 8 (GGGGA)-repeats, respectively (see Fig. 17). As discussed previously (§1.1.2), no alternative promoter utilization has been demonstrated for the AR gene and, in addition, the relative ratio of the AR-TIS I and AR-TIS II transcripts is similar in the small group of cell lines and tissues tested so far (Tilley et al., 1990, 1991; Takane et al., 1991).

Analysis of the sequences involved in AR-TIS I and AR-TIS II selection on the TATA-less hAR gene promoter identifies two different initiation pathways that are active on

overlapping sequences (Fig. 18, taken from Chapter VI).

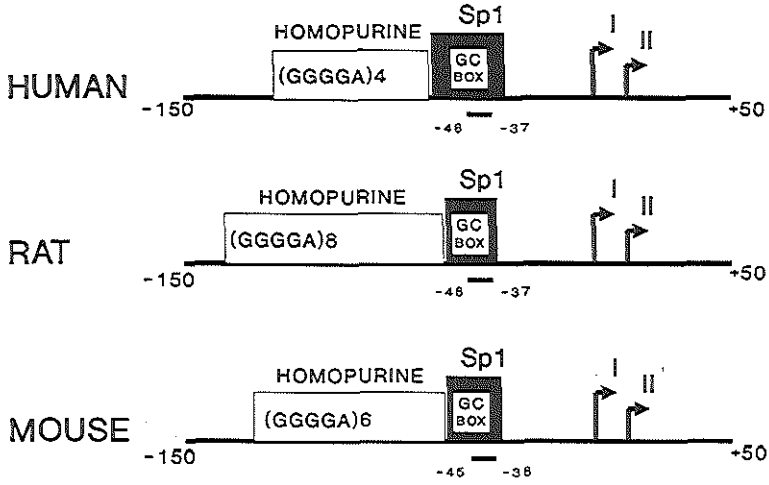


FIGURE 17: Schematical comparison of the proximal promoter regions of the hAR, rAR and mAR genes. Shown are the transcription initiation sites, the GC-box with the Sp1 binding site and the homopurine stretch. The number of GGGGA repeats in the 5'-segment of the homopurine stretch is given.

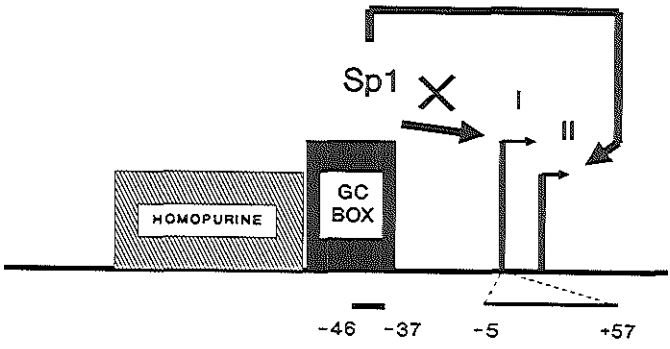


FIGURE 18: Overview of the transcription initiation events on the hAR promoter. Shown is a representation of the hAR proximal promoter. Indicated are the homopurine stretch (hatched area), the GC-box (black area) and the transcription initiation sites, I and II, respectively. The Sp1 binding sequence at -46/-37 is responsible for AR-TIS II utilization but does not affect AR-TIS I selection which is dependent on sequences between position -5 and +57.

AR-TIS I utilization is determined by sequences which are located between positions +5 and +57, whereas the short GC-box and more precisely the Sp1 binding sequence at -46/-37 is essential to AR-TIS II utilization. As the -5/+57 sequence does not contain sequences homologous to the binding sites of known regulatory proteins, the only protein so far identified in start site selection on the hAR promoter is Sp1. Although it might be possible that Sp1 is the sole determinant of AR-TIS II utilization, a comparable promoter, either a native one or a synthetic one, in which Sp1 performs such a function has not been described. Instead Sp1 binding sequences are often found in the proximal promoter region and Sp1 has been referred to as a protein, which increases the relative initiation frequency from a large variety of genes (see Mermelstein et al., 1989; Seipel et al., 1992). As the AR-TIS II sequence is reminiscent to an Inr, this sequence could represent the actual determinant of initiation, although this remains to experimentally proven. With regard to AR-TIS I utilization, the precise sequences and factors involved remain to be established.

## 1.3: SCOPE OF THE THESIS

Androgens, the male sex steroid hormones, are essential for the development of the male phenotype during embryogenesis as well as male virilization during postnatal life. The physiological response to androgens is mediated by a soluble intra-cellular receptor protein, the androgen receptor, which functions as a ligand-dependent transcription factor. Defects in the androgen-androgen receptor signal transduction pathway, including structural abnormalities in the androgen receptor protein have been implicated in several sex-linked disorders, notably the androgen insensitivity syndrome and Kennedy's disease. As part of our ongoing work on the role of the androgen receptor protein in normal and abnormal physiology, this thesis describes the molecular cloning and the characterization of the androgen receptor transcription unit.

In chapters II-IV the molecular cloning and characterization of human androgen receptor complementary DNA (cDNA) and genomic clones is described. Chapter II presents the amino acid sequence of the large N-terminal domain of the androgen receptor. Chapter III deals with the genomic organization of the human androgen receptor gene with respect to the protein coding exons. Chapter IV extends the work of chapters II and III with the isolation and characterization of cDNA clones covering the full-lenght 11 kb human androgen receptor mRNA. It includes the identification of a functional promoter and polyadenylation signals as well as the characterization of an alternative splice which takes place in the 3'-untranslated region (3'-UTR) of the human androgen receptor mRNA. Chapter V describes the cloning and characterization of the mouse androgen receptor cDNA and gene. The experiments described in chapter VI concern the process of transcription initiation on the TATA-less androgen receptor promoter.

1.4 References


Adler, A.J., Scheller, A., Hoffman, Y., & Robins, D.M. (1991) Multiple components of a complex androgen dependent enhancer. Mol. Endocrinol. 5, 1587-1596.

Adler, A.J., Danielsen, M., & Robins, D.M. (1992) Androgen-specific gene activation via a consensus glucocorticoid response element is determined by interaction with nonreceptor factors. Proc. Natl. Acad. Sci. USA 89, 11660-11663.

Amero, S.A., Kretsinger, R.H., Moncrief, N.D., Yamamoto, K.R., & Pearson, W.R. (1992) The origin of nuclear receptor proteins: a single precursor distinct from other transcription factors. Mol. Endocrinol. 4, 3-7.

Arriza, J.F., Weinberger, C., Cerelli, G., Glaser, T.M., Handelin, B.L., Housman, D.E., & Evans, R.M. (1987) Cloning of human mineralocorticoid receptor complementary DNA: structural and functional kinship with the glucocorticoid receptor. Science 237, 268-275.

Aso, T., Vasavada, H.A., Kawaguchi, T., Germino, F.J., Ganguly, S., Kitajima, S., Weissman, S.M., & Yasukochi, Y. (1992) Characterization of cDNA for the large subunit of the transcription factor TFIIF. Nature 355, 461-464.

Baarends, W.M., Themmen, A.P.N., Blok, L.J., Mackenbach, P., Brinkmann, A,O., Meijer, D., Faber, P.W., Trapman, J., & Grootegoed, J.A. (1990) The rat androgen receptor promoter. Mol. Cell. Endocrinol. 74, 75-84.

Beaupain, D., Eléouët, J.F., & Roméo, P.H. (1990) Initiation of transcription of the erythroid promoter of the porphobilinogen deaminase gene is regulated by a cis-acting sequence around the cap site. Nucl. Acids Res. 18, 6509-6515.

Beato, M. (1989) Gene regulation by steroid hormones. Cell 56, 335-344.

Beato, M. (1991) Transcriptional control by nuclear receptors. FASEB J. 5, 2044-2051.

Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W., & Azizkhan, J.C. (1990) Transcriptional initiation is controlled by upstream GC-box interactions in a TATA-less promoter. Mol. Cell. Biol. 10, 6632-6641.

Breathnach, R., & Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. 50, 349-383.

Brinkmann, A.O., Faber, P.W., van Rooij, H.C.J., Kuiper, G.G.J.M., Klaassen, P., van der Korput, J.A.G.M., Voorhorst, M.M., van Laar, J.H., Mulder, E., & Trapman, J. (1989) The human androgen receptor: domain structure, genomic organization and regulation of expression. J. Steroid Biochem. 34, 307-310.

Brinkmann, A.O., & Trapman, J. (1992) Androgen receptor mutants that affect normal growth and development. Cancer Surveys 14, 95-111.

Buratowski, S., Hahn, S., Guarente, L., & Sharp, P.A. (1989) Five inermediate complexes in transcription initiation by RNA polymerase II. Cell 56, 549-561.

Carcamo, J., Maldonado, E., Ahn, M., Ha, I., Kasai, Y., Flint, J., & Reinberg, D. (1990) A TATA-like sequence located downstream of the transcription initiation site is required for expression of an RNA polymerase II transcribed gene. Genes Dev. 4, 1611-1622.

Carson, M.A., Tsai, M-J., Conneely, O.M., Maxwell, B.L., Clark, J.H., Dobson, A.D.W., Elbrecht, A., Toft, D.O., Schrader, W.T., & O'Malley, B.W. (1987) Structure-function properties of the chicken progesterone receptor A synthesized from complementary deoxyribonucleic acid. Mol. Endocrinol. 1, 791-801.

Carson-Jurica, M.A., Schrader, W.T., & O'Malley, B.W. (1990) Steroid receptor family: structure and functions. Endocr. Rev. 11, 201-220.

Cato, A.C.B., Henderson, D., & Ponta, H. (1987) The hormone responsive element of the mouse mammary tumour virus DNA mediates progestin and androgen induction of transcription in the proviral long terminal repeat region. EMBO J. 6, 363-368.

Chang, C., Kokontis, J., & Liao, S. (1988a) Molecular cloning of human and rat complementary DNA encoding androgen receptors. Science 240, 324-326.

Chang, C., Kokontis, J., & Liao, S. (1988b) Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors. Proc. Natl. Acad. Sci. USA 85, 7211-7215.

Chang, C., Wang, C., DeLuca, H.F., Ross, T.K., & Shih, C. C-Y. (1992) Characterization of human androgen receptor overexpressed in the baculovirus system. Proc. Natl. Acad. Sci. USA 89, 5946-5950.

Charest, N.J., Zhou, Z-X, Lubahn, D.B., Olsen, K.L., Wilson, E.M., French, F.S. (1991) A frameshift mutation destabilizes androgen receptor messenger RNA in the Tfm mouse. Mol. Endocrinol. 5, 573-581.

Claessens, F., Celis, L., Peeters, B., Heyns, W., Verhoeven, G., & Rombauts, W. (1989) Functional characterization of an androgen response element in the first intron of the C3(1) gene of prostatic binding protein. Biochem. Biophys. Res. Comm. 164, 833-840.

Cortes, P., Flores, O., & Reinberg, D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II: Purification and analysis of transcription factor TFIIA and identification of transcription factor TFIIJ. Mol. Cell. Biol. 12, 413-421.

Crossley, M., Ludwig, M., Stowell, K.M., De Vos, P., Olek, K., Brownlee, G.G. (1992) Recovery from hemophilia B Leyden: an androgen-responsive element in the factor IX promoter. Science 257, 377-379.

Danielsen, M., Northrop, J.P., & Ringold, G.M. (1986) The mouse glucocorticoid receptor: mapping of functional domains by cloning, sequencing and expression of wild-type and mutant receptor proteins. EMBO J. 5, 2513-2522.
Darbre, P., Page, M., & King, R.J.B. (1986) Androgen regulation by the long terminal repeat of mouse mammary tumor virus. Mol. Cell. Biol. 6, 2847-2854.
Davidson, B.L., Egly, J.M., Mulvihill, E.R., & Chambon, P. (1983) Formation of stable preinitiation complexes between eucaryotic class B transcription factors and promoter sequences. Nature, 301, 680-686.
Dingwall, C., & Laskey, R.A. (1991) Nuclear targeting sequences- a consensus? Trends Biochem. Sci. 16, 478-481.
Dynan, W.S., & Tjian, R. (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35, 79-87.
Dynan, W.S. (1986) Promoters for housekeeping genes. Trends Genet. 196-197.
Dynlacht, B.D., Hoey, T., & Tjian, R. (1991) Isolation of coactivators associated with the TATA-binding protein that mediate transcriptional activation. Cell 66, 563-576.
Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T., & Chakraborty, R. (1992) Genetic variation in five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12, 241-253.
Enclo, I.J., & Detera-Wadleigh, S.D. (1991) The genomic structure of the human glucocorticoid receptor. J. Biol. Chem. 266, 7182-7188.
Evans, R.M. (1988) The steroid and thyroid hormone receptor superfamily. Science 240, 889-895.
Faber, P.W., Kuiper, G.G.J.M., van Rooij, H.C.J., van der Korput, J.A.G.M., Brinkmann, A.O., & Trapman, J. (1989) The N-terminal domain of the human androgen receptor is encoded by one, large exon. Mol. Cell. Endocrinol. 61, 257-262.
Faber, P.W., van Rooij, H.C.J., van der Korput, J.A.G.M., Baarends, W.M., Brinkmann, A.O., Grootegoed, J.A., & Trapman, J. (1991a) Characterization of the human androgen receptor transcription unit. J. Biol. Chem. 266, 10743-10749.
Faber, P.W., King, A., van Rooij, H.C.J., Brinkmann, A.O., de Both, N.J., & Trapman, J. (1991b) The mouse androgen receptor: functional analysis of the protein and characterization of the gene. Biochem. J. 278, 269-278.
Fawell, S.E., Lees, J.A., White, R., & Parker, M.G. (1990) Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor. Cell 60, 953-962.
Finkelstein, A., Kostrub, C.F., Li, J., Chavez, D.P., Wang, B.Q., Fang, S.M., Greenblatt, J., & burton, Z.F. (1992) A cDNA encoding RAP74, a general initiation factor for transcription by RNA polymerase II. Nature 355, 464-467.
Freedman, L.P. (1992) Anatomy of the steroid receptor zinc finger region. Endocrin. Rev. 13, 129-145.
Fuller, P.J. (1991) The steroid receptor superfamily: mechanisms of diversity. FASEB J. 5, 3092-3099.
Gaspar, M.L., Meo, T., & Tosi, M. (1990) Structure and size distribution of the androgen receptor messenger RNA in wild-type and Tfm/y mutant mice. Mol. Endocrinol. 4, 1600-1610.
Giguère, V., Hollenberg, S.M., Rosenfeld, M.G., & Evans, R.M. (1986) Functional domains of the human glucocorticoid receptor. Cell 46, 645-652.
Gowland, P.L., & Buetti, E. (1989) Mutations in the hormone regulatory element of mouse mammary tumor virus differentially affect the response to progestins, androgens, and glucocorticoids. Mol. Cell. Biol. 9, 3999-4008.
Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J-M., Argos, P., & Chambon, P. (1986) Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature 320, 134-139.
Green, S., & Chambon, P. (1988) Nuclear receptors enhance our understanding of transcription regulation. Trends Genet. 4, 309-314.
Green, S., Kumar, V., Theulaz, I., Wahli, W., & Chambon, P. (1988) The N-terminal DNA-binding "zinc finger" of the oestrogen and glucocorticoid receptors determines target gene specificity. EMBO J. 7, 3037-3044.
Greenblatt, J. (1991) Roles of TFIID in transcriptional activation by RNA polymerase II. Cell 66, 1067-1070.
Griffin, J.E., & Wilson, J.D. (1989) The androgen resistance syndromes: 5-alpha-reductase deficiency, testicular feminization, and related syndromes. In: Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D. (Eds) The Metabolic Basis of Inherited Disease, pp 1919-1944 (New York: McGraw-Hill).
Gronemeyer, H., Turcotte, B., Quirin-Stricker, C., Bocquel, M.T., Meyer, M.E., Krozowski, Z., Jeltsch, J.M., Lerouge, T., Garnier, J.M., & Chambon, P. (1987) The chicken progesterone receptor: sequence, expression and functional analysis. EMBO J. 6, 3985-3994.
Gronemeyer, H. (1992) Control of transcription activation by steroid hormone receptors. FASEB J. 6, 2524-2529.
Guiochon-Mantel, A., Loosfelt, H., Lescop, P., Sar, S., Atger, M., Perrot-Applanat, M., & Milgrom, E. (1989) Mechanisms of nuclear localization of the progesterone receptor: evidence for interaction between monomers. Cell 57, 1147-1154.
Ha , I., Lane, W.S., Reinberg, D. (1992) Cloning of a human gene encoding the general transcription initiation factor IIB. Nature 352, 689-695.

Ham, J., Thomson, A., Needham, M., Webb, P., & Parker, M. (1988) Characterization of response elements for androgens, glucocorticoids and progestins in mouse mammary tumor virus. Nucl. Acids Res. 16, 5263-5276.
Härd, T., Kellenbach, E., Boelens, R., Maler, B.A., Dahlman, K., Freedman, L.P., Carlstedt-Duke, J., Yamamoto, K.R., Gustaffson, J-A., & Kaptein, R. (1990) Solution structure of the glucocorticoid receptor DNA binding domain. Science 249, 157-160.
Hariharan, N., & Perry, R.P. (1990) Functional dissection of a ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region. Proc. Natl. Acad. Sci. USA 87, 1526-1530.
He, W.W., Fischer, L.M., Sun, S., Bilhartz, D.L., Zhu, X., Young, C. Y-F., Kelly, D.B., & Tindall, D.J. (1990) Molecular cloning of androgen receptor from divergent species with a polymerase chain reaction technique: complete cDNA sequence of the mouse androgen receptor and isolation of androgen receptor cDNA probes from dog, guinea pig and clawed frog. Biochem. Biophys. Res. Comm. 171, 697-704.
He, W.H., Kumar, M.J., & Tindall, D.J. (1991) A frame-shift mutation in the androgen receptor gene causes complete androgen insensitivity in the testicular-feminized mouse. Nucl. Acids Res. 19, 2373-2378.
Hollenberg, S.M., Weinberger, C., Ong, E.S., Cerelli, C.E., Oro, A., Lebo, R., Thompson, E.B., Rosenfeld, M.G., & Evans, R.M. (1985) Primary structure and expression of a functional human glucocorticoid receptor cDNA. Nature 318, 635-641.
Hollenberg, S.M., & Evans, R.M. (1988) Multiple and cooperative trans-activation domains of the human glucocorticoid receptor. Cell 55, 899-906.
Huckaby, C.S., Conneely, O.M., Beattie, W.G., Dobson, A.D.W., Tsai, M-J., & O'Malley, B.W. (1987) Structure of the chicken progesterone receptor gene. Proc. Natl. Acad. Sci USA 84, 8380-8384.
Inostroza, J., Flores, O., & Reinberg, D. (1991) Factors involved in specific transcription by mammalian polymerase II: purification and functional analysis of general transcription factor IIE. J. Biol. Chem. 266, 9304-9308.
Jenster, G., van der Korput, J.A.G.M., van Vroonhoven, C., van der Kwast, T.H., Trapman, J., & Brinkmann, A.O. (1991) Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. Mol. Endocrinol. 5, 1396-1404.
Jeltsch, J-M., Turcotte, B., Garnier, J-M., Lerouge, T., Krozowski, Z., Gronemeijer, H., & Chambon, P. (1990) Characterization of multiple mRNAs from the chicken progesterone receptor gene. J. Biol. Chem. 265, 3967-3974.
Johnson, P.F., & McKnight, S.L. (1989) Eucaryotic transcriptional regulatory proteins. Annu. Rev. Biochem. 58, 799-839.
Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H., & Chambon, P. (1990) Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. EMBO J. 9, 1603-1614.
Keaveney, M., Klug, J., Dawson, M.T., Nestor, P.V., Neilan, J.G., Forde, R.C., & Gannon, F. (1991) Evidence for a previously unidentified upstream exon in the human oestrogen receptor gene. J. Mol. Endocrinol. 6, 111-115.
Kelleher, R.J. III., Flanagan, P.M., Chasman, D.I., Ponticelli, A.S., Struhl, K., & Kornberg, R.D. (1992) Yeast and human TFIIDs are interchangeable for the response to acidic transcriptional activators in vitro. Genes Dev. 6, 296-303.
Koelle, M.R., Talbot, W.S., Segraves, W.A., Bender, M.T., Cherbas, P., & Hogness, D.S. (1991) The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell 67, 59-77.
Kuiper, G.G.J.M., Faber, P.W., van Rooij, H.C.J., van der Korput, J.A.G.M., Ris-Stalpers, C., Klaassen, P., Trapman, J., & Brinkmann, A.O. (1989) Structural organization of the human androgen receptor gene. J. Mol. Endocinol. 2, R1-R4.
Kumar, V., Green, S., Staub, A., & Chambon, P. (1986) Localisation of the oestradiol-binding and putative DNA-binding domains of the human oestrogen receptor. EMBO J. 5, 2231-2236.
Kumar, V., Green, S., Stack, G., Berry, M., Jin, J-R., & Chambon, P. (1987) Functional domains of the human estrogen receptor. Cell 51, 941-951.
Kumar, V., & Chambon, P. (1988) The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell 55, 145-156.
La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., & Fischbeck, K.H. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352, 77-79.
Laudet, V., Hänni, C., Coll, J., Catzeflis, F., & Stéhelin, D. (1992) Evolution of the nuclear receptor gene superfamily. EMBO J. 11, 1003-1013.
Lubahn, D.B., Joseph, D.R., Sullivan, P.M., Willard, H.F., French, F.S., & Wilson, E.M. (1988) Cloning of human androgen receptor complementary DNA and localization to the X-chromosome. Science 240, 327-330.
Lubahn, D.B., Joseph, D.R., Sar, M., Tan, J-A., Higgs, H.N., Larson, R.E., French, F.S., & Wilson, E.M. (1988b) The human androgen receptor: Complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in the prostate. Mol. Endocrinol. 2, 1265-1275.
Lund, S.D., Gallagher, P.M., Wang, B., Porter, S.C., & Ganschow, R.E. (1991) Androgen responsiveness of the murine β-Glucuronidase gene is associated with nuclease hypersensitivity protein binding, and haplotype-specific sequence diversity within intron 9. Mol. Cell. Biol. 11, 5426-5434.

Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., & Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. Nature 352, 497-505.
Mader, S., Kumar, V., de Verneuil, H., & Chambon, P. (1989) Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid responsive element. Nature 338, 271-274.
Maldonado, E., Ha, I., Cortes, P., Weis, L., & Reinberg, D. (1990) Factors involved in specific transcription by RNA polymerase II: role of transcription factors IIA, IID, and IIB during formation of a transcription-competent complex. Mol. Cell. Biol. 10, 6335-6347.
Maniatis, T., Goodbourn, S., & Fischer, J.A. (1987) Regulation of inducible and tissue-specific gene expression. Science 236, 1237-1245.
Matsui, T., Segall, J., Weil, P.A., & Roeder, R.G. (1980) Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. J. Biol. Chem. 255, 11992-11996.
McPhaul, M.J., Marcelli, M., Tilley, W.D., Griffin, J.E., & Wilson, J.D. (1991) Androgen resistance caused by mutations in the androgen receptor gene. FASEB J. 5, 2910-2915.
Means, A.L., & Farnham, P.J. (1990) Transcription initiation from the dihydrofolate reductase promoter is positioned by HIP1 binding at the initiation site. Mol. Cell. Biol. 10, 653-661.
Mermelstein, F.H., Flores, O., & Reinberg, D. (1989) Initiation of transcription by RNA polymerase II. Biochim. Biophys. Acta 1009, 1-10.
Meyer, M-E., Pornon, A., Ji, J., Bocquel, M-T., Chambon, P., & Gronemeyer, H. (1990) Agonistic and antagonistic activities of RU486 on the functions of the human progesterone receptor. EMBO J. 9, 3923-3932.
Meyer, M-E., Quirin-Stricker, C., Lerouge, T., Bocquel, M-T., & Gronemeyer, H. (1992) A limiting factor mediates the differential activation of promoters by the human progesterone receptor isoforms. J. Biol. Chem. 267, 10882-10887.
Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wikstrom, A-C., Gustafsson, J-A., & Yamamoto, K.R. (1986) Genetic complementation of a glucocorticoid receptor deficiency by expression of cloned receptor DNA. Cell 46, 389-399.
Miller, J., McLachlan, A.D., and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIa from Xenopus oocytes. EMBO J. 4, 1609-1614.
Misrahi, M., Atger, M., d'Auriol, L., Loosfelt, H., Meriel, C., Fridlanski, F., Guiochon-Mantel, A., Galibert, F., & Milgrom, E. (1987) Complete amino acid sequence of the human progesterone receptor deduced from cloned cDNA. Biochem. Biophys. Res. Comm. 143, 740-748.
Misrahi, M, Loosfelt, H., Atger, M., Meriel, C., Zerah, V., Dessen, P., & Milgrom, E. (1988) Organization of the entire rabbit progesterone receptor mRNA and the promoter and 5'-flanking region of the gene. Nucl. Acids Res. 16, 5459-5472.
Mitchell, P.J., & Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245, 371-378.
Mooradian, A.D., Morley, J.E., & Korenman, S.G. (1987) Biological actions of androgens. Endocrin. Rev. 8, 1-26.
Newmark, J.R., Hardy, D.O., Tonb, D.C., Carter, B.S., Epstein, J.I., Isaacs, W.B., Brown, T.B., & Barrack, E.R. (1992) Androgen receptor gene mutations in prostate cancer. Proc. Natl. Acad. Sci. USA 89, 6319-6323.
O'Malley, B.W., & Conneely, O.M. (1992) Orphan receptors: in search of a unifying hypothesis for activation. Mol. Endocrinol. 6, 1359-1361.
Page, M.J., & Parker, M.G. (1982) Effect of androgen on the transcription of rat prostate binding protein genes. Mol. Cell. Endocrinol. 27, 343-355.
Paigen, K. (1989) Mammalian $\beta$-glucuronidase: genetics, molecular biology, and cell biology. Progr. Nucl. Acid Res. Mol. Biol. 37, 155-205.
Parker, C.S., & Topol, J. (1984) A Drosophila RNA polymerase II transcription factor contains a promoter-region-specific DNA-binding activity. Cell 36, 357-369.
Peters, R. (1986) Fluorescence microphotolysis to measure nucleocytoplasmic transport and intracellular mobility. Biochim. Biophys. Acta. 864, 305-359.
Peterson, M.G., Tanese, N., Pugh, B.F., & Tjian, R. (1990) Functional domains and upstream activation properties of cloned human TATA binding protein. Science 248, 1625-1630.
Peterson, M.G., Inostroza, J., Maxon, M.E., Flores, O., Admon, A., Reinberg, D., & Tjian, R. (1991) Structure and functional properties of human general transcription factor IIE. Nature 354, 369-373.
Picard, D., & Yamamoto, K.R. (1987) Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. EMBO J. 6, 3333-3340.
Picard, D., Salser, S.J., & Yamamoto, K.R. (1988) A movable and regulable inactivation function within the steroid binding domain of the glucocorticoid receptor. Cell 54, 1073-1080.
Pinsky, L., & Kaufman, M. (1987) Genetics of steroid receptors and their disorders. Adv. Human Genet. 16, 299-472.
Pinsky, L., Trifiro, M., Kaufman, M., Beitel, L.K., Mhatre, A., Kazemi-Esfarjani, P., Sabbaghian, N., Lumbroso, R., Alvarado, C., Vasiliou, M., & Gottlieb, B. (1992) Androgen resistance due to mutation of the androgen receptor. Clin. Invest. Med. 15, 456-472.
Ponglikitmongkol, M., Green, S., & Chambon, P. (1988) Genomic organization of the human oestrogen receptor gene. EMBO J. 7, 3385-3388.
Ptashne, M. (1986) Gene regulation by proteins acting nearby and at a distance. Nature 322, 697-701.

Pugh, B.F., & Tjian, R. (1990) Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell 61, 1187-1197.
Pugh, B.F., & Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev. 5, 1935-1945.
Pugh, B.F., & Tjian, R. (1992) Diverse transcriptional functions of the multisubunit eucaryotic TFIID complex. J. Biol. Chem. 267, 679-682.
Purvis, I. J., Chotai, D., Dykes, C.W., Lubahn, D.B., French, F.S., Wilson, E.M., & Hobden, A.N. (1991) An androgen inducible expression system for Saccharomyces cerevisiae. Gene 106, 35-42.
Quarmby, V.E., Kemppainen, J.A., Sar, M., Lubahn, D.B., French, F.S., & Wilson, E.M. (1990) Expression of recombinant androgen receptor in cultured mammalian cells. Mol. Endocrinol. 4, 1399-1407.
Reinberg, D., & Roeder, R.G. (1987) Factors involved in specific transcription by mammalian RNA polymerase II: Purification and functional analysis of initiation factors IIB and IIE. J. Biol. Chem. 262, 3310-3321.
Reinberg, D., Horikoshi, M., & Roeder, R.G. (1987) Factors involved in specific transcription in mammalian RNA polymerase II: Functional analysis of initiation factors IIA and IID and identification of a new factor operating at sequences downstream from the initiation site. J. Biol. Chem. 262, 3322-3330.
Riegman, P.H.J., Vlietstra, R.J., van der Korput, J.A.G.M., Brinkmann, A.O., & Trapman, J. (1991) The promoter of the prostate specific antigen gene contains a functional androgen response element. Mol. Endocrinol. 5, 1921-1930.
Ris-Stalpers, C., Kuiper, G.G.J.M., Faber, P.W., Schweigert, H.U., van Rooij, H.C.J., Zegers, N.D., Hodgins, M.B., Degenhart, H.J., Trapman, J., & Brinkmann, A.O. (1990) Aberrant splicing of androgen receptor mRNA results in the synthesis of a non-functional receptor protein in a patient with androgen insensitivity. Proc. Natl. Acad. Sci. USA 87, 7966-7970.
Ris-Stalpers, C., Trifiro, M.A., Kuiper, G.G.J.M., Jenster, G., Romalo, G., Sai, T., van Rooij, H.C.J., Kaufman, M., Rosenfeld, R.L., Liao, S., Schweikert, H.U., Trapman, J., Pinsky, L., & Brinkmann, A.O. (1991) Substitution of aspartic acid-686 by histidine or asparagine in the human androgen receptor leads to a functionally inactive protein with altered hormone-binding characteristics. Mol. Endocrinol. 5, 1562-1569.
Robbins, J., Dilworth, S.M., Laskey, R.A., & Dingwall, C. (1991) Two independent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. Cell 64, 615-623.
Roche, P.J., Hoare, S.A., & Parker, M.G. (1992) A consensus DNA-binding site for the androgen receptor. Mol. Endocrinol. 6, 2229-2235.
Roehrborn, C.G., Zoppi, S., Gruber, J.A., Wilson, C.M., & McPhaul, M.J. (1992) Expression and characterization of full-lenght and partial human androgen receptor fusion proteins. Implications for the production and applications of soluble steroid receptors in Escherichia coli. Mol. Cell. Endocrinol. 84, 1-14.
Roy, A.L., Meisterernst, M., Pognonec, P., & Roeder, R.G. (1991) Cooperative interaction of an initiator-binding transcription factor and the helix-loop-helix activator USF. Nature, 354, 245-248.
Rundlett, S.E., Wu, X-P., & Miesfeld, R.L. (1990) Functional characterizations of the androgen receptor confirm that the molecular basis of androgen action is transcriptional regulation. Mol. Endocrinol. 4, 708-714.
Saltzman, A.G., & Weinmann, R. (1989) Promoter specificity and modulation of RNA polymerase II transcription. FASEB J. 3, 1723-1733.
Samuels, M., Fire, A., & Sharp, P. (1982) Separation and characterization of factors mediating accurate transcription by RNA polymerase II. J. Biol. Chem. 261, 2003-2013.
Sawadogo, M., & Sentenac, A. (1990) RNA polymerase B (II) and general transcription factors. Annu. Rev. Biochem. 59, 711-754.
Schibler, U., & Sierra, F. (1987) Alternative promoters in developmental gene expression. Ann. Rev. Genet. 21, 237-257.
Schwabe, J.W.R., Neuhaus, D., & Rhodes, D. (1990) Solution structure of the DNA-binding domain of the oestrogen receptor. Nature 348, 458-461.
Schwabe, J.W.R., & Rhodes, D. (1991) Beyond zinc fingers: steroid hormone receptors have a novel structural motif for DNA recognition. Trends. Biochem. Sci. 16, 291-296.
Schüle, R., Muller, M., Kaltschmidt, C., & Renkawitz, R. (1988) Many transcription factors interact synergistically with steroid receptors. Science 242, 1418-1420.
Seipel, K., Georgiev, O., & Schaffner, W. (1992) Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J. 11, 4961-4968.
Shi, Y., Seto, E., Chang, L-S., & Shenk, T. (1991) Transcriptional repression by YY1, a human GLI-kruppel related protein, and relief of repression by adenovirus E1A protein. Cell 67, 377-388.
Simental, J.A., Sar, M., Lane, M.V., French, F.S., & Wilson, E.M. (1991) Transcriptional activation and nuclear targeting signals of the human androgen receptor. J. Biol. Chem. 266, 510-518.
Sleddens, H.F.B.M., Oostra, B.A., Brinkmann, A.O., & Trapman, J. (1992) Trinucleotide repeat polymorphism in the androgen receptor gene (AR) Nucl. Acids Res. 20, 1427.
Sleddens, H.F.B.M., Oostra, B.A., Brinkmann, A.O., & Trapman, J. (1993) Trinucleotide (GGN) repeat polymorphism in the human androgen receptor gene. Human Mol. Genet. in press
Smale, S.T., & Baltimore, D. (1989) The "initiator" as a transcription control element. Cell 57, 103-113.

Smale, S.T., Schmidt, M.C., Berk, A.J., & Baltimore, D. (1990) Transcriptional activation by Sp1 directed through TATA or initiator: requirement for transcription factor IID. Proc. Natl. Acad. Sci. USA 87, 4509-4513.
Sopta, M., Burton, Z., & Greenblatt, J. (1989) Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. Nature 341, 410-414.
Srinivasan, G. (1992) Overexpression of receptors of the steroid/thyroid family. Mol. Endocrinol. 6, 857-860.
Strähle, U., Schmidt, A., Kelsey, G., Stewart, A.F., Cole, T.J., Schmid, W., & Schutz, G. (1992) At least three promoters direct expression of the mouse glucocorticoid receptor gene. Proc. Natl. Acad. Sci. USA 89, 6731-6735.
Takane, K.K., Wilson, J.D., McPhaul, M.J. (1991) Decreased levels of the androgen receptor in the mature phallus are associated with decreased levels of androgen receptor messenger ribonucleic acid. Endocrinol. 129, 1093-1100.
Tan, J-A., Joseph, D.R., Quarmby, V.E., Lubahn, D.B., Sar, M., French, F.S., & Wilson, E.M. (1988) The rat androgen receptor: Primary structure, autoregulation of its messenger ribonucleic acid, and immunocytochemical localization of the receptor protein. Mol. Endocrinol. 2, 1276-1285.
Tanese, N., Pugh, B.F., & Tjian, R. (1992) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. Genes Dev. 5, 2212-2224.
Tasset, D., Tora, L., Fromental, C., Scheer, E., & Chambon, P. (1990) Distinct classes of transcriptional activating domains function by different mechanisms. Cell 62, 1177-1187.
Tilley, W.D., Marcelli, M., Wilson, J.D., & McPhaul, M.J. (1989) Characterization and expression of a cDNA encoding the human androgen receptor. Proc. Natl. Acad. Sci. USA 86, 327-331.
Tilley, W.D., Marcelli, M., & McPhaul, M.J. (1990a) Expression of the human androgen receptor gene uses a common promoter in diverse human tissues and cell lines. J. Biol. Chem. 265, 13776-13781.
Trapman, J., Klaassen, P., Kuiper, G.G.J.M., van der Korput, J.A.G.M., Faber, P.W., van Rooij, H.C.J., Geurts van Kessel, A., Voorhorst, M.M., Mulder, E., & Brinkmann, A.O. (1988) Cloning, structure and expression of a cDNA encoding the human androgen receptor. Biochem. Biophys. Res. Comm. 153, 241-248.
Tora, L., White, J., Brou, C., Tasset, D., Webster, N., Scheer, E., & Chambon, P. (1989) The human estrogen receptor has two independent transcriptional activation functions. Cell 59, 477-487.
Tsai, S.Y., Carlstedt-Duke, J., Weigel, N.L., Dahlman, K., Gustafsson, J-A., Tsai, M-J., & O'Malley, B.W. (1988) Molecular interactions of steroid hormone receptor with its enhancer elements: evidence for dimer formation. Cell 55, 361-369.
Umesono, K., & Evans, R.M. (1989) Determinants of target gene specificity for steroid/thyroid hormone receptors. Cell 57, 1139-1146.
Veldscholte, J., Ris-Stalpers, C., Kuiper, G.G.J.M., Jenster, G., Berrevoets, C.A., Claasen E., van Rooij, H.C.J., Trapman, J., Brinkmann, A.O., & Mulder, E. (1990a) A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affect steroid binding characteristics and response to anti-androgens. Biochem. Biophys. Res. Comm. 173, 534-540.
Veldscholte, J., Voorhorst-Ogink, M.M., Bolt-de Vries, J., van Rooij, H.C.J., Trapman, J., & Mulder, E. (1990b) Unusual specificity of the androgen receptor in the human prostate tumor cell line LNCaP: high affinity for progestogenic and estrogenic steroids. Biochim. Biophys. Acta 1052, 187-194.
Wahli, W., & Martinez, E. (1991) Superfamily of steroid nuclear receptors: positive and negative regulators of gene expression. FASEB J. 5, 2243-2249.
Webster, N.J.G., Green, S., Jin, J.R., & Chambon, P. (1988) The hormone-binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function. Cell 54, 199-207.
Webster, N.J.G., Green, S., Tasset, D., Ponglikitmongkol, M., & Chambon, P. (1989) The transcriptional activation function located in the hormone binding domain of the human oestrogen receptor is not encoded in a single exon. EMBO J. 8, 1441-1446.
Weis, L., & Reinberg, D. (1992) Transcription by RNA polymerase II: Initiator-directed formation of transcription complexes. FASEB J. 6, 3300-3309.
White, R., Lees, J.A., Needham, M., Ham, J., & Parker, M. (1987) Structural organization and expression of the mouse estrogen receptor. Mol. Endocrinol. 1, 735-744.
Wolf, D.A., Schultz, P., & Fittler, F. (1992) Transcriptional regulation of prostate kallikrein-like genes by androgen. Mol. Endocrinol. 6, 753-762.
Wooster, R., Mangion, J., Eeles, R., Smith, S., Dowsett, M., Averill, D., Barret-Lee, P., Easton, D.F., Ponder, B.A.J., & Stratton, M.R. (1992) A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome. Nature Genetics, 2, 132-134.
Yarbrough, W.G., Quarmby, V.E., Simenthal, J.A., Joseph, D.R., Sar, M., Lubahn, D.B., Olsen, K.L., French, F.S., & Wilson, E.M. (1990) A single base mutation in the androgen receptor gene causes androgen insensitivity in the testicular feminized rat. J. Biol. Chem. 265, 8893-8900.
Young, C. Y.-F., Qiu, S, Prescott, J.L., & Tindall, D.J. (1990) Overexpression of a partial human androgen receptor in E. coli: characterization of steroid binding, DNA binding and immunological properties.
Xie, Y-B., Sui, Y-P., Shan, L-X., Palvimo, J.J., Phillips, D.M., & Jänne, O.A. (1992) Expression of androgen receptor in insect cells. J. Biol. Chem. 267, 4939-4948.

Ylikomi, T., Bocquel, M.T., Berry, M., Gronemeyer, H., & Chambon, P. (1992) Cooperation of proto-signals for nuclear accumulation of estrogen and progesterone receptors. EMBO J. 11, 3681-3694.
Zawel, L., & Reinberg, D. (1992) Initiation of transcription by RNA polymerase II: a multi-step process. Progr. Nucl. Acid Res. Mol. Biol. 44, 67-108.
Zong, J., Ashraf, J., & Thompson, E.B. (1990) The promoter and first, untranslated exon of the human glucocorticoid receptor gene are GC rich but lack consensus glucocorticoid receptor element sites. Mol. Cell. Biol. 10, 5580-5585.

## CHAPTER II

---

# THE N-TERMINAL DOMAIN OF THE HUMAN ANDROGEN RECEPTOR IS ENCODED BY ONE, LARGE EXON

FABER, P.W., KUIPER, G.G.J.M., van ROOIJ, H.C.J., VAN DER KORPUT, J.A.G.M., BRINKMANN, A.O., & TRAPMAN, J.

MOLECULAR AND CELLULAR ENDOCRINOLOGY 61, 257-262 (1989).

*Molecular and Cellular Endocrinology*, 61 (1989) 257–262
Elsevier Scientific Publishers Ireland, Ltd.



MCE 02027

Rapid Communication

# The N-terminal domain of the human androgen receptor is encoded by one, large exon

P.W. Faber [1], G.G.J.M. Kuiper [2], H.C.J. van Rooij [1], J.A.G.M. van der Korput [1], A.O. Brinkmann [2] and J. Trapman [1]

[1] *Department of Pathology and* [2] *Department of Biochemistry II, Erasmus University, 3000 DR Rotterdam, The Netherlands*

(Received 23 November 1988; accepted 24 November 1988)

*Key words:* Androgen, receptor; Gene, structure; (Human)

**Summary**

Using specific cDNA hybridization probes, the first coding exon of the human androgen receptor gene was isolated from a genomic library. The exon contained an open reading frame of 1586 bp, encoding an androgen receptor amino-terminal region of 529 amino acids. The deduced amino acid sequence was characterized by the presence of several poly-amino acid stretches of which the long poly-glycine stretch (16 residues) and the poly-glutamine stretch (20 residues) were most prominent. Androgen receptor cDNAs from different sources contained information for poly-glycine stretches of variable size (23 and 27 residues, respectively). The androgen receptor amino-terminal domain was found to be hydrophilic and have a net negative charge. Combined with the previously described, partially overlapping cDNA clone 7A2M27 (Trapman et al. (1988) Biochem. Biophys. Res. Commun. 153, 241–248), the complete human androgen receptor was deduced to have a size of 910 amino acids.

## Introduction

Androgens exert their physiological function by an interaction with the intracellular androgen receptor (AR). This interaction 'activates' AR and enables it to modulate specific gene expression by binding to androgen response elements located in the flanking regions of target genes (Yamamoto, 1985). AR is a member of the superfamily of ligand-responsive transcriptional modifiers (Green and Chambon, 1986; Evans, 1988) which includes steroid hormone receptors, the vitamin D3 receptor, thyroid hormone receptors and retinoic acid receptors. All members of this family are characterized by a similar domain structure: a C-terminal domain which is essential for ligand binding is preceded by a DNA-binding domain composed of two DNA-binding Zn fingers and an N-terminal domain of variable size. The involvement of the N-terminal domain in the fine tuning of transcriptional regulation has been established for some receptors (Danielsen et al., 1987; Kumar et al., 1987; Tora et al., 1988).

Dysfunctioning of the system of AR-modulated gene expression is presumed to be involved in several forms of the androgen insensitivity syndrome (Wilson et al., 1983; Pinsky and Kaufman, 1987) and in the (progressive) growth of human prostate cancer (Coffey and Pienta, 1987). There-

Address for correspondence: Dr. P.W. Faber, Department of Pathology, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

fore, a detailed knowledge of human AR (hAR) structure and expression is of great clinical importance.

Recently, the isolation of partial hAR cDNA clones has led to the elucidation of the primary structure of the putative DNA-binding domain (Chang et al., 1988a; Lubahn et al., 1988; Trapman et al., 1988) and the putative ligand-binding domain (Trapman et al., 1988). Here we describe the cloning and characterization of a genomic DNA fragment containing the information for the complete N-terminal region (hAR-N).

## Materials and methods

The human genomic library (Mbo 1 partial digest in EMBL3) was a gift of Dr. G. Grosveld (Rotterdam, The Netherlands). T47D cDNA libraries in γgt10 were kindly provided by Dr. E. Milgrom (Paris, France) and the human testis cDNA library (Clontech, Palo Alto, CA, U.S.A.) was made available to us by Dr. d'Azzo (Rotterdam).

Phages were propagated in *Escherichia coli* C600HΠ or Y1090. Screening of the libraries was performed according to standard high stringency hybridization procedures. Probes were labelled as described (Feinberg and Vogelstein, 1983). Positive phages were isolated by three rounds of purification. Phages were grown, DNA purified and characterized by restriction mapping and Southern blotting using standard methods (Maniatis et al., 1982; Davis et al., 1986).

Insert fragments were subcloned in pUC9 or pTZ19 and mapped in detail. Appropriate fragments were subcloned in M13mp18/19 and sequenced by the dideoxy chain termination method (Sanger et al., 1977) using sequenase (USB, Cleveland, OH, U.S.A.) and GTP or ITP. In addition, GC-rich regions were sequenced by the chemical cleavage method (Maxam and Gilbert, 1977).

## Results

Previously we reported the isolation of a partial hAR cDNA clone (7A2M27) from a T47D (mammary tumor cell line) cDNA library (Trapman et al., 1988). Clone 7A2M27 contained the information for the putative DNA-binding and ligand-binding domain of hAR but lacked most of hAR-N encoding sequences. Using an appropriate 7A2M27 5′-fragment a human genomic library and several cDNA libraries were screened. This resulted in the isolation of a genomic clone (G17.1) and the cDNA clones 3.3 (T47D) and T4.1-A1 (testis) (Fig. 1).

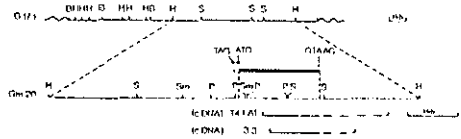

Fig. 1. Partial restriction map of G17.1 A 7 kb *Hin*dIII/*Hin*dIII fragment (Gm 20) containing the first coding exon of the hAR gene is shown in more detail. The black bar over Gm 20 indicates the position of the 1586 bp ORF, starting with the indicated ATG start codon and ending with the splice donor consensus sequence 5′-GTAAG-3′. The indicated TAG stop codon is located in-frame at position −105 to −103. Two independent cDNAs (T4.1-A1 and 3.3) are shown under Gm 20 with the straight line indicating corresponding sequence and the broken line indicating deviating sequence. B = *Bam*HI, H = *Hin*dIII, P = *Pst*I, S = *Sac*I, Sm = *Sma*I.

Sequence analysis of the hybridizing fragment of G17.1 and comparison with cDNA sequences revealed the presence of an open reading frame (ORF) of 1586 bp (Fig. 2). The first ATG in this ORF was preceded by an in-frame stop codon (TAG) at position −105 to −103 (Figs. 1 and 2). S1-nuclease mapping showed that G17.1 lacked intervening sequences in the coding region presented (data not shown). Starting at position 1587, G17.1 deviated from the cDNA sequences and a splice donor consensus sequence (5′-GTAAG-3′) (compare Figs. 1 and 2) was found at this position. Combination of the hAR-N ORF with the 7A2M27 sequence (Trapman et al., 1988) resulted in the construction of a 2730 bp fragment encoding the complete hAR. This ORF can encode a 910 amino acid protein with a calculated MW of 99 kDa. This value is somewhat below the value of 110 kDa obtained by SDS-PAGE analysis of the photolabelled receptor (Brinkmann et al., 1988). Whether this difference is due to post-translational modifications and/or an aberrant mobility is at present unknown. hAR contains ample sites which could be subject to modifications such as serine, threonine and tyrosine phosphorylation

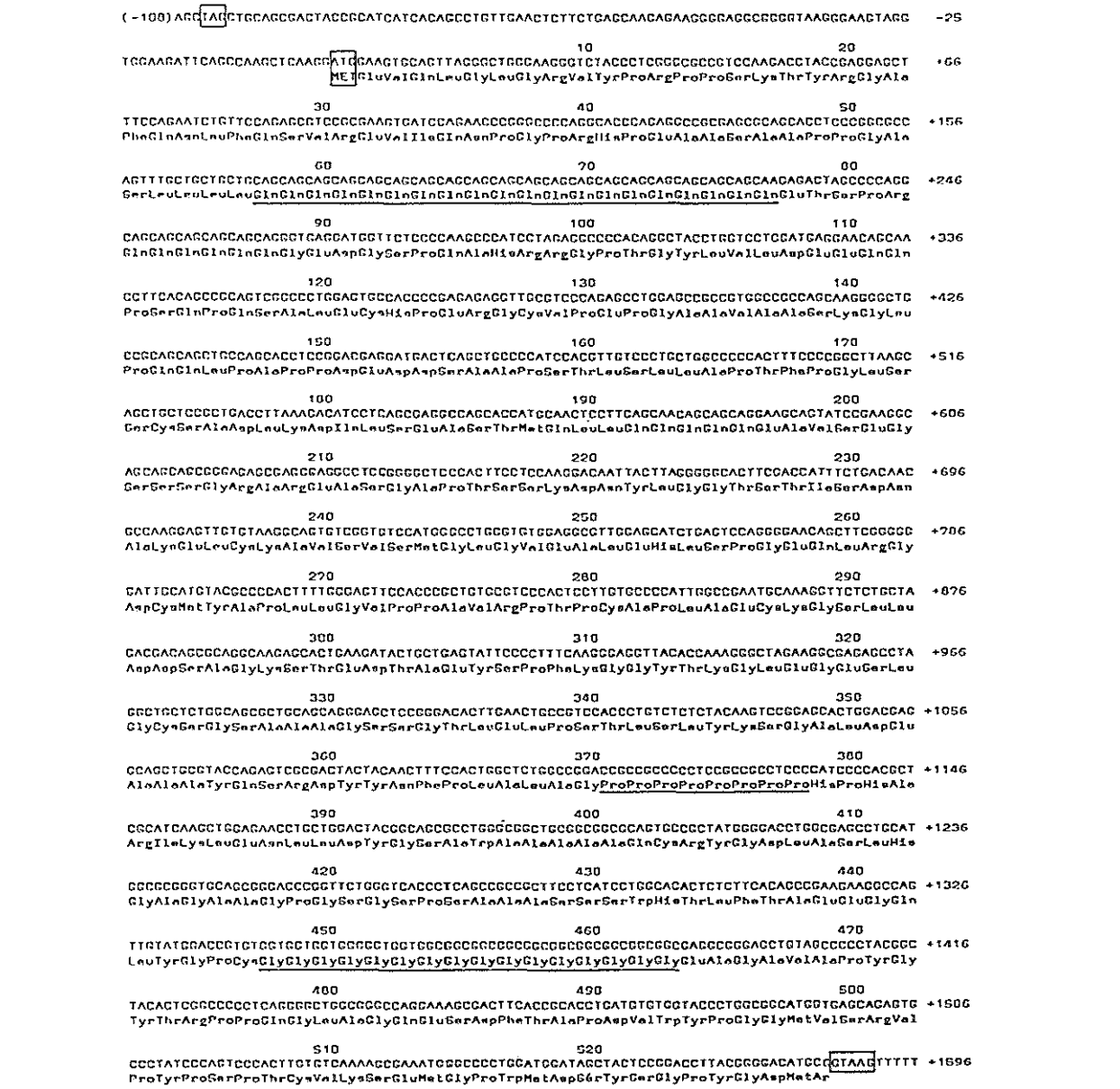

Fig. 2. Genomic nucleotide sequences and deduced amino acid sequences of the N-terminal region of hAR. Numbers at the right indicate nucleotide sequences and numbers over amino acids indicate amino acid sequences. The TAG stop codon, the ATG start codon and the splice donor consensus sequence 5′-GTAAG-3′ are boxed. The poly-Gln, poly-Pro and poly-Gly stretches are underlined.

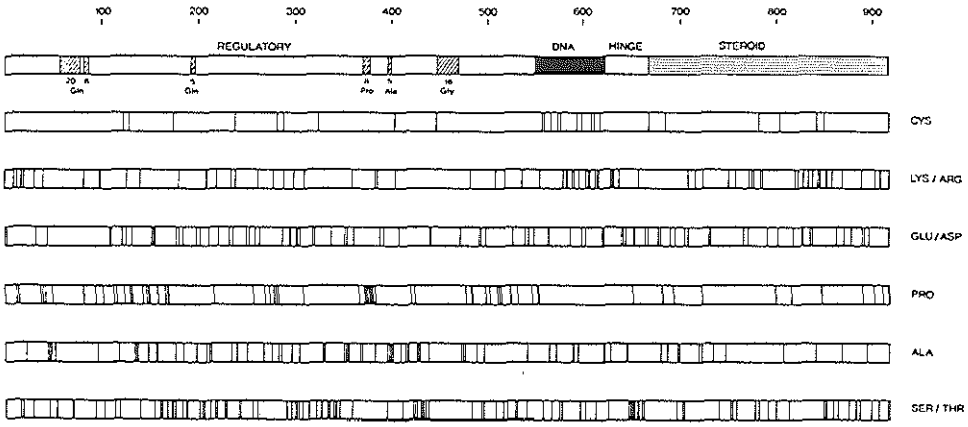


Fig. 3. Schematic presentation of the distribution of several amino acid residues over hAR, showing the abundance of Cys, Lys and Arg residues in the DNA-binding domain, the presence of an acidic domain between 100 and 325 and the bias for proline, alanine and serine/threonine residues in the N-terminal region of hAR. The amino acid composition of the DNA- and steroid-binding domains is from Trapman et al. (1988).

and/or *O*-glycosylation (see Jackson and Tjian, 1988).

The hAR N-terminal domain comprises the first 529 amino acids. A considerable bias for certain amino acids, notably proline (10%), alanine (11%), glycine (9%), glutamine (9%) and serine/threonine (12%) can be observed in this region (see Fig. 3). In addition, several remarkable stretches of homopolymeric amino acids are present (Figs. 2 and 3) of which the long poly-Gln (58–77; CAG/CAA triplet) and poly-Gly (448–463; GGN triplet) regions are most prominent, but which also include two shorter poly-Gln (83–88; 192–196) stretches, a poly-Pro (371–377) stretch and a poly-Ala (397–401) stretch. The hydrophobicity plot of hAR revealed the N-terminal region to be very hydrophilic (Fig. 4). Calculation of the net charge of hAR-N resulted in a negative value of −22, mainly because of high amounts of acidic amino acids in the region 100–325 (see Fig. 3).

Although the protein coding sequences of the genomic DNA clone and those of the two cDNA clones were in excellent agreement with each other, one striking difference was found. The GGN repeat, encoding the poly-Gly stretch was composed of 16 GGN triplets in the genomic clone whereas 23 GGN triplets and 27 GGN triplets were found in the T47D clone 3.3 and testis clone T4.1-A1, respectively. The deviations started at an identical position: one GGT triplet followed by a series of GGC triplets was lacking in both cDNAs. Sequence analysis of several different clones isolated from the same library gave rise to consistent results. The differences found could be the result of unequal recombination during the preparation or propagation of the libraries, which is sometimes

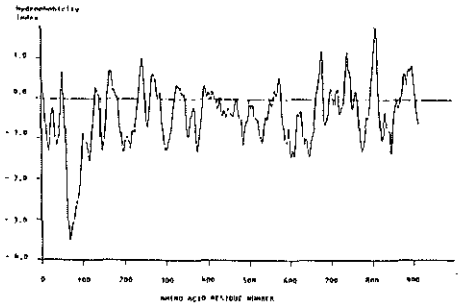

Fig. 4. Hydrophobicity plot of hAR. Note the hydrophilicity of the N-terminal region (1–529), the second DNA-binding finger and the hinge region (570–663) and the hydrophobicity of the steroid-binding domain (664–910).

observed with repeat sequences. An alternative explanation is the occurrence of a hypervariable GGN repeat in the original DNA samples. Which of the possibilities is correct remains to be established. It will be obvious, however, that the definite size of the GGN repeat still is a matter of dispute.

## Discussion

The deduced primary structure of hAR-N has revealed it to encompass the first 529 amino acids of the 910 amino acid hAR. hAR-N is encoded by the first, large coding exon of the hAR gene. A similar organization has been found in the case of the chicken progesterone receptor (cPR) gene where the N-terminal domain is also encoded by the first, large exon (Huckaby et al., 1987). Preliminary work on the organization of the remainder of the hAR gene indicates that this is also highly similar to that of the cPR gene.

Although hAR-N shows no obvious homology to the corresponding region of other steroid hormone receptors, analysis of the amino acid composition revealed some relationship with the human PR (hPR) (Misrahi et al., 1987). Like hAR, the N-terminal domain of hPR is characterized by high amounts of proline (15%), alanine (16%), glycine (9%) and serine/threonine (13%) and it shows the same hydrophilicity at the N-terminus (data not shown) suggesting a close evolutionary kinship between the hAR and hPR. Interestingly, high amounts of the aforementioned amino acids have also been reported in the putative regulatory domains of several recently characterized mammalian transcriptional regulatory factors: SP-1 (Kadonaga et al., 1987), C/EBP (Landschulz et al., 1988) and AP-1 (jun) (Bohman et al., 1987).

An acidic domain of variable size seems to be common to many transcriptional regulatory proteins and in the case of two yeast factors (GCN4 and GAL4) it is indispensable for proper function (Ptashne, 1988). Whether the acidic domain of the hAR N-terminal region exerts a similar function remains to be established.

The most conspicuous feature of hAR-N is the presence of two transcribed repetitive elements, the pen repeat (GGN; Haynes et al., 1987) and the opa repeat (CAG/CAA; Wharton et al., 1985). Interestingly, most proteins in which similar repeats were identified (Drosophila and mouse homeotic gene products (see Haynes et al., 1987 for pen references and Burglin, 1988 for opa references) and the rat and mouse glucocorticoid receptor (GR) (Danielsen et al., 1987; Miesfeld et al., 1986)) are involved in developmental control and/or regulation of gene expression. The Gln repeat is absent in hGR (Hollenberg et al., 1985). A specific function has not yet been attributed to these repeats and there is evidence suggesting that they do not contribute to any function at all (Danielsen et al., 1987). Still it is remarkable that the repeats have been identified exclusively in proteins with specific properties and it might well be that the exact conditions under which they exert their function have not been examined.

The pen repeat seems to be hypervariable in size as 16 GGN triplets were found in the genomic clone and 23 and 27 GGN triplets in two independent cDNA clones. Therefore, the actual size of hAR can be different from the 910 amino acids reported here. Studies are in progress to find out whether the variability observed is a reflection of the in vivo situation. If this turns out to be the case, the GGN repeat could be a hot spot of naturally occurring mutations which could result in hAR deficiencies. Very recently Chang et al. (1988b) also reported the characterization of hAR. Differences with our data are in the two long homopolymeric stretches (17 glutamines; 27 glycines) and at positions 389 (Leu/Pro) and 467 (Gly/Glu).

The cloning of the complete hAR coding region has now provided the tools which are necessary to investigate in detail the system of hAR-modulated gene expression. Specific antibodies against various parts of hAR can be raised and presumed defects in hAR structure can be investigated at protein and DNA levels. Hopefully, this will lead to a better understanding of the role of hAR in human disease.

## Acknowledgements

We are indebted to Mrs. P. Delfos for artwork and photography and to Ing. P. Clotscher for computer analyses. This study was supported by a grant from the Netherlands Cancer Foundation KWF.

## References


Bohman, D., Bos, T.J., Nishimura, T., Vogt, P. and Tjian, R. (1987) Science 238, 1386–1392.

Brinkmann, A.O., Kuiper, G.G.J.M., Bolt-de Vries, J. and Mulder, E. (1988) J. Steroid Biochem. 30, 257–261.

Burglin, T.R. (1988) Cell 53, 339–340.

Chang, C., Kokontis, J. and Liao, S. (1988a) Science 240, 324–326.

Chang, C., Kokontis, J. and Liao, S. (1988b) Proc. Natl. Acad. Sci. U.S.A. 85, 7211–7215.

Coffey, D.S. and Pienta, K.J. (1987) Prog. Clin. Biol. Res. 239, 1–73.

Danielsen, M., Northrop, J.P., Jonklaas, J. and Ringold, G.M. (1987) Mol. Endocrinol. 1, 816–822.

Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) Basic Methods in Molecular Biology, Elsevier, New York.

Evans, R.M. (1988) Science 240, 889–895.

Feinberg, A.P. and Vogelstein, P. (1983) Anal. Biochem. 132, 6–13.

Green, S. and Chambon, P. (1986) Nature 324, 615–617.

Haynes, S.R., Rebbert, M.L., Mozer, B.A., Forquignon, F. and Dawid, I.B. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 1819–1823.

Hollenberg, S.M., Weinberger, C., Ong, E.S., Cerelli, G., Oro, A., Lebo, R., Thompson, E.B., Rosenfeld, M.G. and Evans, R.M. (1985) Nature 318, 635–641.

Huckaby, C.S., Conneely, O.M., Beattie, W.G., Dobson, A.D.W., Tsai, M.-J. and O'Malley, B.W. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 8380–8384.

Jackson, S.P. and Tjian, R. (1988) Cell 55, 125–133.

Kadonaga, J.T., Carner, K.R., Masiarz, R. and Tjian, R. (1987) Cell 51, 1079–1090.

Kumar, V., Green, S., Stack, G., Berry, M., Jin, J.-R. and Chambon, P. (1987) Cell 51, 941–951.

Landschulz, W.H., Johnson, P.F., Adashi, E.Y., Graves, B.J. and McKnight, S.L. (1988) Genes Dev. 2, 786–800.

Lubahn, D.B., Joseph, D.R., Sullivan, P.M., Huntington, F.W., French, F.S. and Wilson, E.M. (1988) Science 240, 327–330.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560–563.

Miesfeld, R., Rusconi, S., Godowski, P.J., Maler, B.A., Okret, S., Wikstrom, A.C., Gustafsson, J.-A. and Yamamoto, K.R. (1986) Cell 46, 389–399.

Misrahi, M., Atger, M., d'Auriol, L., Loosfelt, H., Meriel, C., Fridlansky, F., Guiochon-Mantel, A., Galibert, F. and Milgrom, E. (1987) Biochem. Biophys. Res. Commun. 143, 740–748.

Pinsky, L. and Kaufman, M. (1987) Adv. Hum. Genet. 16, 299–472.

Ptashne, M. (1988) Nature 335, 683–689.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467.

Tora, L., Gronemeyer, H., Turcotte, B., Gaub, M.-P. and Chambon, P. (1988) Nature 333, 185–188.

Trapman, J., Klaassen, P., Kuiper, G.G.J.M., van der Korput, J.A.G.M., Faber, P.W., van Rooij, H.C.J., Geurts van Kessel, A., Voorhorst, M.M., Mulder, E. and Brinkmann, A.O. (1988) Biochem. Biophys. Res. Commun. 153, 241–248.

Wharton, K.A., Yedvobnick, B., Finnerty, V.G., and Artvanis-Tsakonas, S. (1985) Cell 40, 55–62.

Wilson, J.D., Griffin, J.E., Leshin, M. and MacDonald, P.C. (1983) in The Metabolic Basis of Inherited Disease, pp. 1001–1026.

Yamamoto, K.R. (1985) Annu. Rev. Genet. 19, 209–252.

## CHAPTER III

---

# STRUCTURAL ORGANIZATION OF THE HUMAN ANDROGEN RECEPTOR GENE

KUIPER, G.G.J.M., FABER, P.W., van ROOIJ, H.C.J., van der KORPUT,
J.A.G.M., RIS-STALPERS, C., KLAASSEN, P.,
TRAPMAN, J., & BRINKMANN, A.O.

JOURNAL OF MOLECULAR ENDOCRINOLOGY 2, R1-R4 (1989).

# STRUCTURAL ORGANIZATION OF THE HUMAN ANDROGEN RECEPTOR GENE

G.G.J.M.Kuiper(*), P.W.Faber, H.C.J.van Rooij, J.A.G.M.van der Korput, C.Ris-Stalpers (*), P.Klaassen, J.Trapman and A.O.Brinkmann (*).

Department of Biochemistry II (*) and Department of Pathology, Erasmus University Rotterdam, P.O.Box 1738, 3000 DR Rotterdam, The Netherlands.

## ABSTRACT

The complete coding region of the human androgen receptor gene has been isolated from a genomic library. The information for the androgen receptor was found to be divided over eight exons and the total length of the gene exceeded 90 kb. The sequence encoding the N-terminal region is present in one large exon. The two putative DNA-binding fingers are encoded separately by two small exons.The information for the hormone-binding domain is split over five exons. Positions of introns are identical to those reported for the chicken progesterone receptor and the human oestrogen receptor genes. Southern blot analysis of genomic DNA with various specific probes reveal that the human androgen receptor is encoded by a single-copy gene.

## INTRODUCTION

Regulation of cell growth and differentiation are complex processes which require the cooperation of many factors involved in modulation of gene expression. Among these is the superfamily of steroid/thyroid hormone/retinoic acid receptors, whose trans-acting activity is controlled by the tight and specific binding of the cognate ligand [Yamamoto, 1985; Evans, 1988; Green & Chambon, 1988]. The molecular cloning and structural characterization of cDNA's encoding many of these ligand responsive factors, has helped to define the structural requirements for their proper functioning [Giguère, Hollenberg, Rosenfeld et al. 1986; Kumar, Green, Stack et al. 1987; Guiochon-Mantel, Loosfelt, Ragot et al. 1988]. In general three functional domains can be distinguished. Firstly the largely hydrophobic C-terminal domain, which is involved in ligand binding and probably also in transcriptional activation [Webster, Green, Jin & Chambon, 1988; Guiochon-Mantel et al. 1988]. Secondly a DNA-binding domain which contains the highly conserved Zn-finger motif and thirdly the acidic N-terminal domain which is supposed to play a role in transcriptional activation [Evans, 1988; Green & Chambon, 1988].

Androgenic hormones play an essential role in male sexual differentiation and are believed to exert their function at the transcriptional level via the nuclear androgen receptor. The recently published primary structure of the human androgen receptor reveals a domain structure identical to that of the other members of the superfamily of ligand responsive factors [Trapman, Klaassen, Kuiper et al. 1988; Chang, Kokontis & Liao, 1988; Lubahn, Joseph, Sar et al. 1988]. Three mRNA species sizes (11, 8.5 and 4.7 kb) have been detected for the human androgen receptor and the localization of the corresponding gene has been confined to the X-chromosome [Trapman et al. 1988]. Defects in the human androgen receptor have been directly correlated to several forms of the X-linked androgen insensitivity syndrome [Pinsky & Kaufman, 1987; Brown, Lubahn, Wilson et al. 1988].

In order to define more precisely the regions of the human androgen receptor, which are involved in normal male phenotypic expression as well as in abnormal male sexual differentiation, a detailed knowledge of the organization of the human androgen receptor gene is of high importance. In the present investigation the molecular cloning and partial characterization of the human androgen receptor gene is described.

## MATERIALS AND METHODS

An EMBL3 human genomic library [Mbo I partial digest] was kindly provided by Dr. G.Grosveld (Rotterdam). This library was screened several times with appropriate, 32P-labeled [Feinberg & Vogelstein, 1983] human androgen receptor cDNA fragments [ see Trapman et al. 1988; Faber, Kuiper, van Rooij et al. 1989; see also Figure 2A (probes A-E)]. Duplicate nitrocellulose filters were hybridized overnight in 6xSSC, 10x Denhardt, 0.1% SDS, and 100 µg/ml salmon sperm DNA at 65°C . Filters were washed twice in 3xSSC for 20 min at 65°C, twice in 1xSSC for 20 min at 65°C and once in 0.3xSSC for 10 min at 65°C. Filters were exposed to Kodak X-ARS film for 18 to 64 h at -70°C using intensifier screens. Positive plaques were purified by three isolation cycles.

Phages were propagated, DNA isolated and restriction maps prepared using standard procedures [Maniatis, Fritsch & Sambrook, 1982; Davis, Dibner & Battey, 1986]. Appropriate fragments, containing exon sequences, were subcloned in pUC9 or pTZ19 for detailed restriction mapping. For sequencing fragments were subcloned in M13mp18/19. Sequencing was performed by the dideoxy chain termination method using Sequenase (USB, Cleveland) and dGTP or dITP [Sanger, Nicklen & Coulson, 1977].

Southern blots of human genomic DNA were prepared by standard methods (Maniatis et al. 1982]. Hybridizations under stringent conditions were performed as described above.

## RESULTS

Using human androgen receptor cDNA fragments as hybridization probes a total number of 9 different, partially overlapping human genomic fragments was isolated (Figure 1). After restriction mapping with EcoRI, BamHI, SstI and HindIII and subsequent hybridization with cDNA probes three groups of overlapping clones could be identified. The phages 17.1 and 5.1 span about 21 kb and contain sequences encoding the human androgen receptor N-terminal region. Phages 2.1, 4.2 and 8.2 overlap and contain a fragment hybridizing with the DNA-binding domain. These three phages span a range of approximately 22 kb. Phage 9.2 also hybridized with a human androgen receptor cDNA probe specific for the DNA-binding domain, but did not overlap with 2.1 or 8.2. Phages 8.1, 7.2 and 18.1 were overlapping and contained sequences encoding the steroid binding domain.

J. Mol. Endocr. (1989) 2, R1–R4  Printed in Great Britain
0952-5041/89/002-0R1 $02.00/0 Accepted 17 February 1989

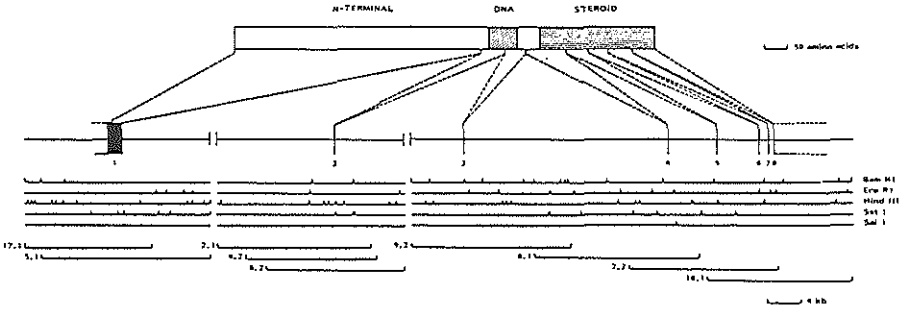


Figure 1
Structural organization of the human androgen receptor gene. The eight exons of the human androgen receptor gene are numbered 1-8 and their corresponding position in the human androgen receptor cDNA is indicated together with the location of the N-terminal-, DNA-binding and steroid-binding domains respectively. The position of the 9 individual clones (17.1; 5.1; 2.1; 9.2; 8.2; 9.2; 8.1; 7.2; 18.1) is shown at the bottom of the figure. In addition a restriction map of the genomic clones after digestion with BamHI, EcoRI, HindIII, SstI and SalI is shown.

Restriction mapping and hybridization experiments with terminal fragments of the various inserts revealed that 9.2 and 8.1 contained a common region. The phages 9.2, 8.1, 7.2 and 18.1 span a region with a size of about 52 kb (Figure 1).

Appropriate hybridizing fragments were subcloned in a plasmid vector and detailed restriction maps were made (data not shown). Hybridization with human androgen receptor cDNA probes revealed the position of the various exons. Restriction sites known to be present in the cDNA sequence and sites found to be located in the direct vicinity of the exons were used for subcloning of fragments in M13mp18/19. Subsequently the nucleotide sequences of these regions were determined. A comparison of the genomic sequence with the cDNA-coding sequence allowed the identification of 8 exons. Detailed information on the sequences at the exon/intron boundaries is shown in Table I. At all boundaries splice consensus sequences are found.

TABLE I

HUMAN ANDROGEN RECEPTOR GENE EXON/INTRON JUNCTIONS

| EXON | INTRON | EXON |
|---|---|---|
| GAC ATG CG<br>Asp Met Ar<br>527 528 52 | GTAAGTTTTTCCT-- 1 --TGTGTCTTTCCAG<br>>24 kb | T TTG GAG<br>g Leu Glu<br>9 530 531 |
| GCT GAA G<br>Ala Glu G<br>578 579 5 | GTAAAGGGTCTTG-- 2 --TTTGTTCTCCCAG<br>>15 kb | GG AAA CAG<br>ly Lys Gln<br>80 581 582 |
| CTG GGA G<br>Leu Gly A<br>617 618 6 | GTAAGATACTTTT-- 3 --TCCTTCCCAATAG<br>26 kb | CC CGG AAG<br>la Arg Lys<br>19 620 621 |
| TTG CCT G<br>Leu Pro G<br>713 714 7 | GTAAGGAAAAGGG-- 4 --TCTTCTTCTCCAG<br>5.6 kb | GC TTC CGC<br>ly Phe Arg<br>15 716 717 |
| TTC AAT GA<br>Phe Asn Gl<br>761 762 76 | GTAAGTGCTCCTG-- 5 --ATCTCCTTCCCAG<br>4.8 kb | G TAC CGC<br>u Tyr Arg<br>3 764 765 |
| AGC ATT A<br>Ser Ile I<br>805 806 8 | GTAAGTGCCTAGA-- 6 --CATCCCACATCAG<br>0.8 kb | TT CCA GTG<br>le Pro Val<br>07 808 809 |
| CAG CCT<br>Gln Pro<br>858 859 | GTAAGCAAACGAT-- 7 --TTGTTCCCTACAG<br>0.7 kb | ATT GCG<br>Ile Ala<br>860 861 |

The splice position in the amino acid sequence is relative to +1 as determined in Faber et al. 1989. The position in the amino acid sequence in Trapman et al. 1988 can be obtained by subtracting 458 from each number.

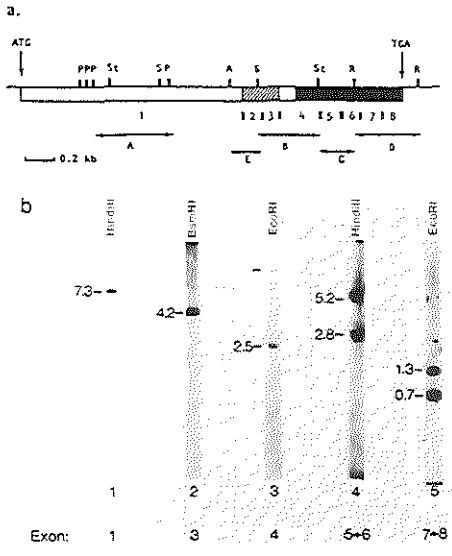


Figure 2

a. Human androgen receptor cDNA probes (A-D) used for Southern blot analysis of human genomic DNA. Probes A-E were used for screening of genomic libraries. Numbers indicate different exons. For detection of exon 3 a specific XbaI-XbaI genomic fragment containing the complete exon was used. A: Asp718; P: PvuII; R: EcoRI; S: SstI; St: StuI. Hatched region represents sequences encoding the DNA-binding domain; black region represents the steroid binding domain.

b. Southern blot analysis of genomic DNA. DNA was digested either with BamHI, EcoRI or HindIII (as indicated at the top of each lane). The following hybridization probes were used (see figure 2a): lane 1: probe A (exon 1); lane 2: genomic XbaI-XbaI probe encompassing exon 3; lane 3: probe B (exon 3 + 4); lane 4: probe C (exon 5 + 6); lane 5: probe D (exon 7 + 8). The high molecular weight band in lane 3 probably represents exon 3.

The first exon contains the ATG translation start codon and the information for the complete N-terminal domain of the human androgen receptor (1586 bp). The size of the first exon is over 3 kb, because of the very long 5'-noncoding sequence.The DNA-binding domain is represented by two small exons. The sequence encoding the first Zn-finger is in exon 2 (152 bp), whereas exon 3 (117 bp) encodes the second Zn-finger. The steroid binding domain is divided over 5 exons (288, 145, 131, 158 and 153 bp, respectively). Exon 8 contains the last part of the open reading frame (153 bp) and presumably all of the corresponding large 3'-noncoding region of the human androgen receptor mRNA (over 6 kb). Introns between exons 1 and 2, exons 2 and 3, and exons 3 and 4 are very large ( >24 kb, >15 kb and approximately 25 kb, respectively). The other introns are much smaller: 5.6 kb, 4.8 kb, 0.8 kb and 0.7 kb respectively. The total length of the human androgen receptor gene will be more than 90 kb.

High stringency hybridization of Southern blots of human genomic DNA was performed with a panel of different human androgen receptor specific probes (Figure 2a). Examples of these experiments are shown in Figure 2b. Only hybridizing bands were identified which correspond exactly in size and number to those predicted from the restriction maps of the phage inserts after HindIII, BamHI and EcoRI digestion (Figures 1 and 2). This result strongly indicates that the human androgen receptor is encoded by a single-copy gene.

## DISCUSSION

The organization of the human androgen receptor gene was elucidated from overlapping fragments isolated from a genomic library. The human androgen receptor gene has a total length of more than 90 kb and the protein coding part was found to be separated over eight exons. All exon-intron junctions were sequenced and the exact size of the various exons was determined. Analysis of genomic DNA showed that the human androgen receptor is encoded by a single-copy gene. This is in concordance with the genes of the other steroid hormone receptors, which are also single-copy genes and are located on separate chromosomes [Green & Chambon, 1988]. In contrast for both the retinoic acid receptor and the thyroid hormone receptor at least two genes have been reported, one of each on chromosomes 3 and 17 [Evans, 1988; Green & Chambon, 1988]. Because only one gene for the androgen receptor is found, the previously described three mRNA species of the human androgen receptor [Trapman et al., 1988] may be evolved from alternative processing of one primary transcript. Alternatively, one or both smaller mRNA's are specific degradation products of the large (approx. 11 kb) transcript. Detailed analysis of the extremely long 5'- and 3'- noncoding regions of the cDNA and the corresponding parts of the gene is in progress.

The N-terminal domain of the human androgen receptor is encoded by one large exon [Faber et al., 1989]. The putative DNA-binding domain is divided over two small exons. This finding suggests that each part of the DNA-binding domain may represent a separate functional sub-domain. Supportive evidence for this speculative suggestion has been reported recently [Green, Kumar, Theulaz et al., 1988]. In that study the first 'zinc finger' in the oestrogen and the glucocorticoid receptor is found to be involved in specific recognition of hormone responsive elements, whereas the second finger is supposed to stabilize the DNA-protein interaction through non-specific DNA-binding and dimer formation. Whether this mechanism can be generally applied for all the members of the steroid/thyroid hormone /retinoic acid receptor family awaits further conclusive experiments.

Recently the structure of the chicken progesterone receptor and human oestrogen receptor genes have been reported [Huckaby, Conneely, Beattie et al., 1987; Ponglikitmongkol, Green & Chambon, 1988] and a comparison of the genomic organization of these genes with that of the human androgen receptor gene revealed striking conservation. Although intron sizes

varied considerably, all exon-intron boundaries were at exactly the same positions. This finding substantiates earlier observations showing that steroid receptors are members of the same subgroup of ligand responsive transcription activating factors.

In the thyroid hormone receptor and retinoic acid receptor genes the position of the intron separating the exons encoding the two zinc fingers is different from that found in steroid hormone receptors [Zachraoui & Cuny, 1987; Green & Chambon, 1988].

In the human oestrogen and glucocorticoid receptor as well as in the human progesterone receptor a hormone-inducible transcription activating region is present close to or within the hormone binding domain [Webster et al., 1988; Guiochon-Mantel et al., 1988]. In this regard it would be of interest to investigate whether one of the exons encoding the androgen binding domain provides specific information for transcription activation. Exon deletion as well as exon swapping experiments will have to be performed to substantiate this point further.

The information in the present investigation on the organization of the human androgen receptor gene is of extreme importance for the detailed analysis of the androgen receptor gene in patients with androgen resistance. Knowledge of the human androgen receptor gene structure allows direct screening for the presence of each individual exon by Southern blotting and hybridization with specific androgen receptor probes. Furthermore each individual exon can now be amplified using the DNA-polymerase chain reaction [Saiki, Gelfand, Stoffel et al., 1988] and can be subsequently sequenced. This approach allows the detection of point mutations and other small aberrations in the androgen receptor structure, which may be the cause of androgen insensitivity. A similar methodology has been successfully applied very recently for the detection of point mutations in the DNA-binding domain of the vitamin D receptor in patients with hypocalcemic vitamin D-resistant rickets [Hughes, Malloy, Kieback et al., 1988].

ACKNOWLEDGEMENTS

The present investigation was supported by grants from the Dutch Organization for Scientific Research (Medigon-NWO) and from the Netherlands Cancer Foundation KWF.

REFERENCES

Brown, T. R., Lubahn, D. B., Wilson, E. M., Joseph, D. R., French, F. S. & Migeon, C. J. (1988). Proceedings of the National Academy of Sciences of the USA 85, 8151-8155.

Chang, C., Kokontis, J. & Liao, S. (1988). Proceedings of the National Academy of Sciences of the USA 85, 7211-7215.

Davis, L. G., Dibner, M. D. & Battey, J. F. (1986). Basic Methods in Molecular Biology. Elsevier, New York.

Evans, R. M. (1988). Science 240, 889-895.

Faber, P. W., Kuiper, G. G. J. M., van Rooij, H. C. J., van der Korput, J. A. G. M., Brinkmann, A. O. & Trapman, J. (1989). Molecular and Cellular Endocrinology 61, 257-262.

Feinberg, A. P. & Vogelstein, P. (1983). Analytical Biochemistry 132, 6-13.

Giguère, V., Hollenberg, S. M., Rosenfeld, M. G. & Evans, R. M. (1988). Cell 46, 645-652.

Green, S., Kumar, V., Theulaz, J., Wahli, W. & Chambon, P. (1988). EMBO Journal 7, 3037-3044.

Green, S. & Chambon, P. (1988). Trends in Genetics 4, 309-314.

Guiochon-Mantel, A., Loosfelt, H., Ragot, T., Bailly, A., Atger, M., Misrahi, M., Perricaudet, M. & Milgrom, E. (1988). Nature 336, 695-698.

Huckaby, C. S., Conneely, O. M., Beattie, W. G., Dobson, A. D. W., Tsai, M. J. & O'Malley, B. W. (1987). Proceedings of the National Academy of Sciences of the USA 84, 8380-8384.

Hughes, M. R., Malloy, P. J., Kieback, D. G., Kesterson, R. A., Pike, J. W., Feldman, D., and O'Malley, B. W. (1988). Science 242, 1702-1705.

Kumar, V., Green, S., Stack, G., Berry, M., Jin, J. R. & Chambon, P. (1987). Cell 51, 941-951.

Lubahn, D. B., Joseph, D. R., Sar, M., Tan, J. Higgs, H. N., Larson, R. E., French, F. S. & Wilson, E. M. (1988). Molecular Endocrinology 2, 1265-1275.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Pinsky, L. & Kaufman, M. (1987). Advances in Human Genetics 16, 299-472.

Ponglikitmongkol, M., Green, S. &'Chambon, P. (1988). EMBO Journal 7, 3385-3388.

Saiki, R. K., Gelfand, D. H., Stoffel, G., Scharf, G. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Ehrlich, H. A. (1988). Science 239, 487-494.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). Proceedings of the National Academy of Sciences of the USA 74, 5463-5467.

Trapman, J., Klaassen, P., Kuiper, G. G. J. M., van der Korput, J. A. G. M., Faber, P. W., van Rooij, H. C. J., Geurts van Kessel, A., Voorhorst, M. M., Mulder, E. & Brinkmann, A.O. (1988). Biochemical Biophysical Research Communications 153, 241-248.

Webster, N. J. G., Green, S., Jin, J. R. & Chambon, P. (1988). Cell 54, 199-207.

Yamamoto, K. R. (1985). Annual Review of Genetics 19, 209-252.

Zahraoui, A. & Cuny, G. (1987). European Journal of Biochemistry 166, 63-69.

## CHAPTER IV

# CHARACTERIZATION OF THE HUMAN ANDROGEN RECEPTOR TRANSCRIPTION UNIT

FABER, P.W., VAN ROOIJ, H.C.J., VAN DER KORPUT, J.A.G.M., BAARENDS, W.M., GROOTEGOED, J.A., BRINKMANN, A.O., & TRAPMAN, J.

JOURNAL OF BIOLOGICAL CHEMISTRY 266, 10743-10749 (1991).

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1991 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 266, No. 17, Issue of June 15, pp. 10743-10749, 1991
Printed in U.S.A.

# Characterization of the Human Androgen Receptor Transcription Unit*

(Received for publication, May 29, 1990)

Peter W. Faber‡§, Henri C. J. van Rooij‡, Hetty A. G. M. van der Korput‡, Willy M. Baarends¶, Albert O. Brinkmann¶, J. Anton Grootegoed¶, and Jan Trapman‡

*From the Departments of ‡Pathology and ¶Biochemistry II, Erasmus University, P. O. Box 1738, 3000 DR Rotterdam, The Netherlands*

A full length human androgen receptor (hAR) cDNA was constructed from cDNA and genomic clones. Structurally the 10.6-kilobase (kb) hAR cDNA consists of a long 5′-untranslated region (5′-UTR, 1.1 kb), a previously described open reading frame (ORF, 2.7 kb) (Trapman, J., Klaassen, P., Kuiper, G. G. J. M., van der Korput, J. A. G. M., Faber, P. W., van Rooij, H. C. J., Geurts van Kessel, A., Voorhorst, M. M., Mulder, E., and Brinkmann, A. O. (1988) *Biochem. Biophys. Res. Commun.* 153, 241–248; Faber, P. W., Kuiper, G. G. J. M., van Rooij, H. C. J., van der Korput, J. A. G. M., Brinkmann, A. O., and Trapman, J. (1989) *Mol. Cell. Endocrinol.* 61, 257–262), and a very long 3′-untranslated region (3′-UTR, 6.8 kb). The complete 5′- and 3′-UTRs were found to be encoded by the previously reported first and eight protein coding exons of the hAR gene, respectively (Kuiper, G. G. J. M., Faber, P. W., van Rooij, H. C. J., van der Korput, J. A. G. M., Ris-Stalpers, C., Klaassen, P., Trapman, J., and Brinkmann, A. O. (1989) *J. Mol. Endocrinol.* 2, R1–R4). Two major sites of transcription initiation were identified in a 13-base pair region. DNA fragments spanning these transcription initiation sites conferred promoter activity upon a promoterless chloramphenicol acetyltransferase reporter gene construct. Two equally effective, functional polyadenylation signals (ATTAAA and CATAAA) at a mutual distance of 221 base pairs were detected. The ATTAAA hexamer sequence gave rise to multiple sites of poly(A) addition, whereas only one position was used following the CATAAA hexamer. In LNCaP prostatic carcinoma cells an alternatively spliced hAR mRNA species was identified which lacks 3 kb of the 3′-UTR.

The process of cell growth and differentiation requires the intricate interplay of many factors involved in the regulation of gene transcription. Among these are the members of the recently identified steroid/thyroid hormone/retinoic acid receptor family, which function in response to binding of the corresponding ligand as transcription factors of specific target genes (reviewed in Refs. 1–4).

The androgen receptor (AR)[1] is a member of this receptor family and mediates the physiological response to testosterone and dihydrotestosterone. It plays a major role in the process of male sexual development and maintenance of male sex characteristics (see, for reviews, Refs. 5–7). Little is known about the regulation of AR expression, apart from the observation that moderate or low levels of AR expression are found in many tissues and high expression is observed in cells of the male urogenital system (5). In addition, it has been reported that androgens are able to down-regulate the expression of AR mRNA (8). Defects in human AR (hAR) expression and/or function are known to be involved in the androgen insensitivity syndrome (7, 9–11). They may also play a role in the progressive growth of human prostatic tumors (12). Therefore, identification of the hAR transcription unit and of elements which regulate hAR transcription, hAR mRNA processing, and translation are of high interest.

The primary structure of the hAR has been deduced from cDNA sequences (13–19). The hAR contains the domain structure (trans-activating, DNA-binding, and ligand-binding) characteristic of the members of the steroid receptor family. The hAR is encoded by a single copy gene which is localized on the X-chromosome (14, 15). The information for the protein-coding part of the gene is separated over eight exons (11, 20). The large N-terminal domain is encoded by exon 1, the two zinc fingers of the DNA-binding domain are encoded separately by exons 2 and 3 and the information for the ligand-binding domain is split over exons 4–8. In the LNCaP prostatic carcinoma cell line two major hAR mRNA species of 8.5 and 11 kb have been identified (15).

In the present study we characterize the complete hAR cDNA. Two functional polyadenylation signals and the origin of the 8.5-kb mRNA species are revealed. In addition, we identify the transcription initiation sites of the hAR gene.

## MATERIALS AND METHODS AND RESULTS[2]

*Identification of Transcription Initiation Sites of the hAR Gene*—In order to find out the presumed localization of the transcription initiation site(s) of the hAR gene in the genomic fragment as presented in Fig. 2, several experimental approaches were applied. A S1 nuclease protection experiment was performed using an anti-sense single-stranded DNA

* This work was supported by a grant from the Dutch Cancer Society (to K. W. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

*The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M63504 and M63505.*

§ To whom correspondence should be addressed. Tel.: 3110-4087956; Fax: 3110-4088152.

[1] The abbreviations used are: hAR, human androgen receptor; kb, kilobase; CAT, chloramphenicol acetyltransferase; pENH, pCAT-enhancer expression plasmid; PCR, polymerase chain reaction.

[2] Portions of this paper (including "Materials and Methods," part of "Results," "Discussion," and Figs. 1, 2, 4, and 6) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are included in the microfilm edition of the Journal that is available from Waverly Press.

probe (G-*Bss*HII, Fig. 3*A*) and LNCaP total RNA. RNA from foreskin fibroblasts from an androgen insensitivity syndrome individual lacking the complete hAR gene served as a control. The results of the experiment are shown in Fig. 3*B* and summarized in Fig. 2. Several protected fragments were identified, with the two most prominent signals mapping in a 13-base pair region between 1114 and 1126 nucleotides in front of the ATG translation initiation codon. Less intense signals were found in the more upstream region. One of the major protected fragments mapped to a splice acceptor consensus sequence. However, other fragments did not, and no branch point sequence could be identified (Fig. 2). Primer extension experiments failed to provide proper signals which could be due to the high G + C content of the 5′-untranslated region.

Subsequently, a functional assay was performed. A 0.75-kb *Sst*I-*Sma*I (−737 to +21) and a 1.1-kb *Sst*I-*Pvu*II (−737 to +378) fragment (Fig. 3*A*), which span the potential transcription initiation sites, were cloned in sense and anti-sense orientation in front of the CAT reporter gene in the promoterless pCAT-enhancer expression plasmid (pENH). This gives rise to the constructs pENH1 and pENH rev1 for the *Sst*I-*Sma*I fragment and pENH2 and pENH rev2 for the *Sst*I-*Pvu*II fragment (Fig. 4). All constructs were transiently expressed in LNCaP cells. Controls included transfections with pENH and pSV2CAT. The results of the CAT assays are shown in Fig. 4. Extracts from cells transfected with pENH gave a low background activity. This activity was considerably increased by both pENH1 and pENH2, whereby pENH2 elevates the CAT activity more than pENH1. This suggests that, although pENH1 contains all the sequences necessary for promoter activity, additional sequences present in pENH2 can augment expression of the CAT reporter gene. Extracts from cells transfected with the reverse constructs pENH rev1 and pENH rev2 were devoid of CAT activity. Importantly, a S1 nuclease protection experiment using RNA isolated from COS cells transfected with pENH2 revealed the same tran-

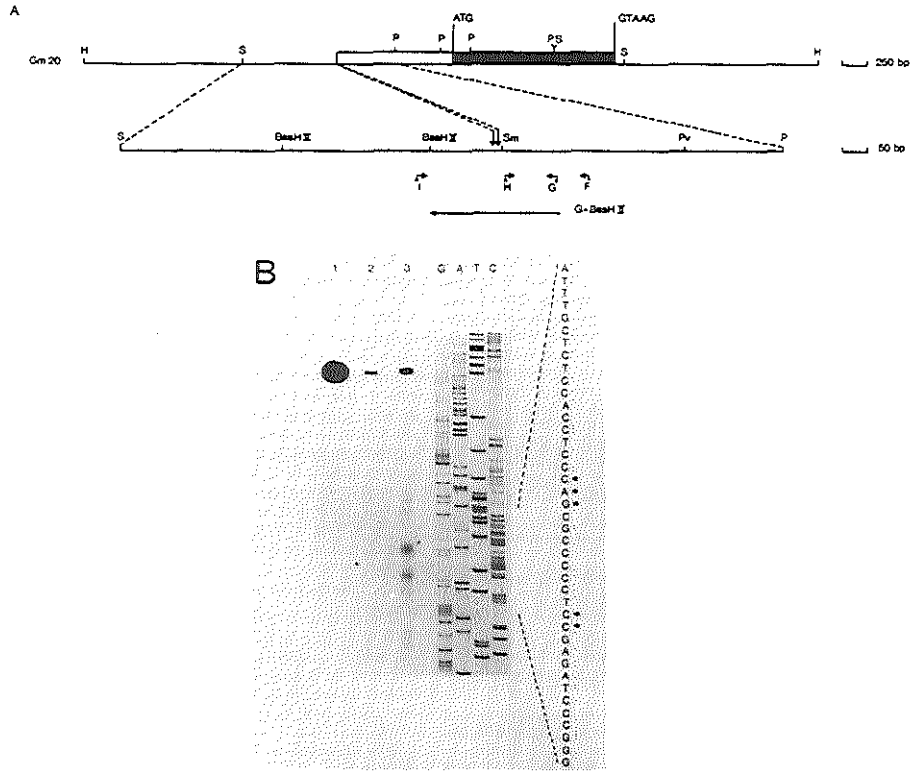


FIG. 3. **Determination of the transcription initiation sites of the hAR gene by S1 nuclease mapping.** *A*, partial restriction map of Gm20 from which an 1.3-kb *Sst*I-*Pst*I fragment is shown in more detail. The position of primers F, G, H, and I and the single stranded DNA probe G-*Bss*HII is indicated. The *black triangle* marks the start of the rAR cDNA. *H*, *Hin*dIII, *P*, *Pst*I, *Pv*, *Pvu*II, *S*, *Sst*I, *Sm*, *Sma*I. *B*, S1 nuclease mapping of hAR mRNA with the single stranded probe G-*Bss*HII. *Lane 1*, probe; *lane 2*, S1 analysis of control RNA; *lane 3*, S1 analysis of LNCaP RNA. A sequence ladder prepared with primer G was run alongside the S1 nuclease mapping products. The end points of the major protected fragments are indicated with an *asterisk* alongside this sequence ladder.

scription initiation sites as identified for LNCaP hAR mRNA (data not shown).

*Localization of the Polyadenylation Signals of the hAR mRNA*—Because the screening of oligo(dT)-primed cDNA libraries had failed to identify hAR cDNA clones primed on the hAR mRNA poly(A) tail, an anchorage-PCR experiment was performed (33). Poly(A)$^+$ RNA, which was isolated from LNCaP cells, was transcribed into first strand cDNA with the primer T12-X, from which the T12 sequence should prime on poly(A) stretches and from which the X sequence, that contains several restriction sites, serves as a PCR adaptor. The first strand cDNA preparation was then amplified using the primers X and A (Fig. 5A). Primer A was chosen just downstream from the AAAAAAAAGAAAAAAAAAATTAAAAA sequence (on which the various clones isolated from an oligo(dT)-primed cDNA library were primed; see Fig. 1) to prevent amplification of cDNAs originating from this sequence. Controls included PCR reactions with first strand cDNA preparations from which reverse transcriptase had been omitted. Analysis of Southern blots of PCR products, which were hybridized to probe 7 (Fig. 5*A*), resulted in the identification of two DNA fragments of 400 base pairs (ARp1) and 650 base pairs (ARp2), respectively (Fig. 5, *A* and *B*, summarized in Fig. 1). ARp1 and ARp2 were sequenced and compared to the genomic sequence of clone Gm100 (Fig. 6). Both fragments ended in long poly(A) tails which were preceded by the polyadenylation signals ATTAAA (ARp1) and CATAAA (ARp2). Interestingly, the three ARp1 clones that were sequenced showed poly(A) addition at three different positions (14, 15, and 24 nucleotides following the polyadenylation signal), whereas the two ARp2 clones sequenced showed poly(A) addition at a single position (16 nucleotides following the polyadenylation signal). Both sites of poly(A) addition are followed by G + T-rich segments in the genomic sequence (Fig. 6). Summarizing, the data obtained demonstrate that the complete hAR 3′-untranslated region has a length of 6.8 kb and is contained in exon 8 of the hAR gene, which also encodes the last segment of the open reading frame (Fig. 1 and Ref. 20).

*Identification of an Alternatively Spliced hAR mRNA*—hAR cDNA and genomic probes were used in Northern blot analysis to identify the nature of the 4.7-, 8.5-, and 11-kb hAR mRNAs (15). Fig. 7*B* shows the results of a representative experiment in which fragments 2, 3, 5, and 6 (Fig. 7*A*) were used as hybridization probes. In contrast to previous studies, the minor 4.7-kb mRNA species was not detected in the RNA preparations, indicating that it may represent a degradation product of the other two mRNAs. The two mRNAs of 8.5 and 11 kb were clearly visualized by probes 2, 3, and 6, whereas probe 5 only detected the 11-kb mRNA. These findings suggest that the 8.5-kb mRNA differs from the 11-kb mRNA by the absence of part of the hAR mRNA 3′-untranslated region. The precise origin of the 8.5-kb mRNA was determined in a reverse transcriptase-PCR experiment. LNCaP poly(A)$^+$ RNA was transcribed into first strand cDNA using primers B and C (Fig. 7*A*). The first strand cDNA preparations were subsequently amplified using the primer combinations B/C and B/E or the combinations C/D and C/E. Controls included

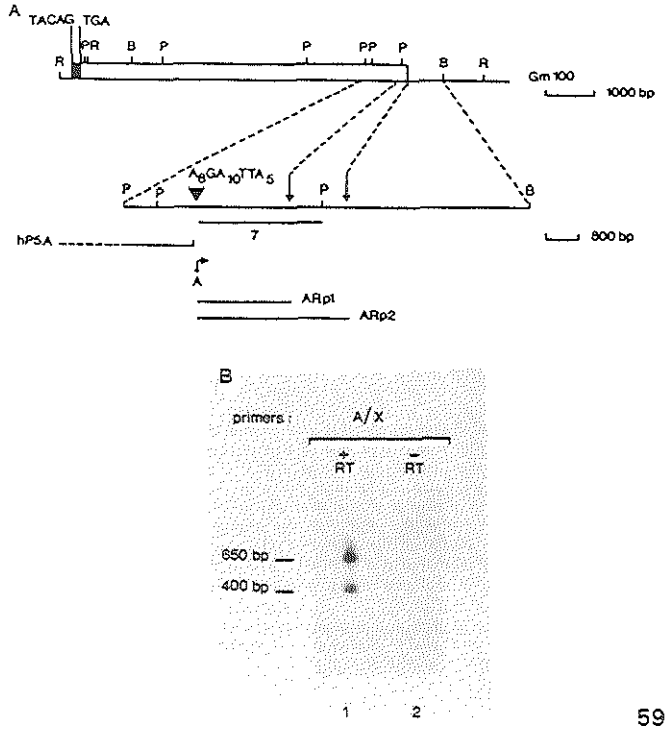


FIG. 5. Determination of the hAR mRNA polyadenylation sites by anchorage-PCR. *A*, restriction map of Gm100, from which an 1.5-kb *Pst*I-*Bam*HI fragment is shown in more detail. The sequence $A_8GA_{10}TTA_5$ marks the end of clone hP-5A (Fig. 1). The position of primer A and hybridization probe 7 are indicated. *B*, *Bam*HI; *P*, *Pst*I; *R*, *Eco*RI. *B*, autoradiograph of Southern blot analysis of the PCR products, hybridized with probe 7. *Lane 1*, first strand cDNA preparation with reverse transcriptase (*RT*); *lane 2*, first strand cDNA preparation without reverse transcriptase.

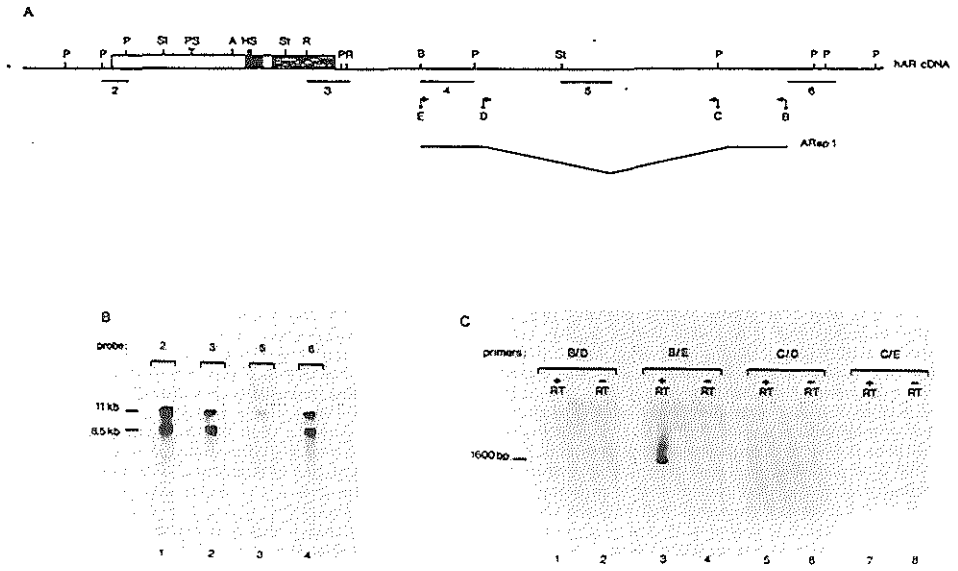


FIG. 7. Northern blot analysis of hAR mRNAs and determination of the origin of the 8.5-kb hAR mRNA species. *A*, partial restriction map of the complete hAR cDNA. The positions of probes 2, 3, 4, 5, and 6 are indicated by *horizontal bars*. The positions of the primers used in the PCR experiment to identify the nature of the 8.5-kb hAR mRNA are indicated by *arrows*. *A*, Asp-718; *B*, *Bam*HI; *H*, *Hin*dIII; *R*, *Eco*RI; *S*, *Sst*I; *St*, *Stu*I. *B*, autoradiograph of a Northern blot of LNCaP poly(A)$^+$ RNA hybridized with probes 2, 3, 5, and 6. *C*, DNA fragments produced after amplification of a first strand cDNA preparation synthesized with either primer B or C in the absence or presence of reverse transcriptase (*RT*) were separated on agarose gels and blotted onto a nitrocellulose filter. An autoradiograph of the blot hybridized to probe 4 is shown. The primer combinations and the presence or absence of reverse transcriptase are indicated at the top of the lane.

amplifications with first strand cDNA preparations from which reverse transcriptase had been omitted. Hybridization of Southern blots of the PCR products with probe 4 (Fig. 7*A*) resulted in the identification of one amplified fragment of 1600 bp (ARsp1) generated by the primer combination B/E (Fig. 7*C*). All other primer combinations were negative. Hybridization of the blot to the 1.2-kb *Pst*I-*Pst*I fragment in which primer B is situated gave an identical result (data not shown). The sequence of clone ARsp1 was determined and compared to the sequence of the full length hAR cDNA.[3] The comparison showed clone ARsp1 to be identical to the full length hAR cDNA up to nucleotide +5650 after which clone ARap1 continues with nucleotide +8685. As the sequence surrounding position +5650 (CTG(+5650)/GTGAG) and +8685 (CTTTAACTTTCTCTGCATCTTTATATTTTGG-TTCCAG/A(+8685)TCA) conforms to the consensus splice donor and acceptor sites, the deletion of 3-kb of hAR mRNA 3′-untranslated region must be caused by an alternative splice event (in the splice acceptor sequence the potential branch point sequence is underlined).

[3] The complete sequence of the human androgen receptor cDNA together with portions of the 5′- and 3′-flanking genomic sequences will be presented elsewhere.

*Acknowledgments*—We thank Dr. L. Pinsky for the hAR fibroblasts, Drs. Milgrom and Grosveld for libraries, and Drs. Kuiper and Themmen for stimulating discussions. We are grateful to Paula Delfos for artwork and photography.

REFERENCES

1. Green, S., and Chambon, P. (1988) *Trends Genet.* 4, 309–314
2. Evans, R. M. (1988) Science 240, 889–895
3. Beato, M. (1989) *Cell* 56, 335–344
4. O'Malley, B. (1990) *Mol. Endocrinol.* 4, 363–369
5. Mooradian, A. D., Morley, J. E., and Korenman, S. G. (1987) *Endocr. Rev.* 8, 1–27
6. Cunha, G. R., Donjacour, A. A., Cooke, P. S., Mee, S., Bigsby, R. M., Higgins, S. J., and Sugimura, Y. (1987) *Endocr. Rev.* 8, 338–362
7. Griffin, J. E., and Wilson, J. D. (1989) in *The Metabolic Basis of*

*Inherited Disease* (Scriver, S. C., ed) pp. 1919–1944, McGraw-Hill ISC, New York
8. Quarmby, V. E., Yarbrough, W. G., Lubahn, D. B., French, F. S., and Wilson, E. M. (1990) *Mol. Endocrinol.* **4**, 22–28
9. Pinsky, L., and Kaufman, M. (1987) *Adv. Hum. Genet.* **16**, 299–472
10. Brown, T. R., Lubahn, D. B., Wilson, E. M., Joseph, D. R., French, F. S., and Migeon, C. J. (1988) Proc. Natl. Acad. Sci. U. S. A. **85**, 8151–8155
11. Lubahn, D. B., Brown, T. R., Simental, J. A., Higgs, H. N., Migeon, C. J., Wilson, E. M., and French, F. S. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 9534–9538
12. Coffey, D. S., and Pientu, K. J. (1987) *Prog. Clin. Biol. Res.* **239**, 1–73
13. Chang, C., Kokontis, J., and Liao, S. (1988) *Science* **240**, 324–326
14. Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S., and Wilson, E. M. (1988) *Science* **240**, 327–330
15. Trapman, J., Klaassen, P., Kuiper, G. G. J. M., van der Korput, J. A. G. M., Faber, P. W., van Rooij, H. C. J., Geurts van Kessel, A., Voorhorst, M. M., Mulder, E., and Brinkmann, A. O. (1988) *Biochem. Biophys. Res. Commun.* **153**, 241–248
16. Chang, C., Kokontis, J., and Liao, S. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 7211–7215
17. Lubahn, D. B., Joseph, D. R., Sar, M., Tan, J., Higgs, H. N., Larson, R. E., French, F. S., and Wilson, E. M. (1988) *Mol. Endocrinol.* **2**, 1265–1275
18. Faber, P. W., Kuiper, G. G. J. M., van Rooij, H. C. J., van der Korput, J. A. G. M., Brinkmann, A. O., and Trapman, J. (1989) *Mol. Cell. Endocrinol.* **61**, 257–262
19. Tilley, W. D., Marcelli, M., Wilson, J. D., and McPhaul, M. J. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 327–331
20. Kuiper, G. G. J. M., Faber, P. W., van Rooij, H. C. J., van der Korput, J. A. G. M., Ris-Stalpers, C., Klaassen, P., Trapman, J., and Brinkmann, A. O. (1989) *J. Mol. Endocrinol.* **2**, R1–R4
21. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
22. Davis, L. G., Dibner, M. D., and Battey, J. F. (1986) *Basic Methods in Molecular Biology*, Elsevier Service Publishing Co., Inc., New York
23. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
24. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
25. Maxam, A., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 560–563
26. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
27. Favoloro, J., Treisman, R., and Kamen, R. (1980) *Methods Enzymol.* **65**, 718–749
28. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239**, 487–491
29. Chen, C., and Okayama, H. (1987) *Mol. Cell. Biol.* **7**, 2745–2752
30. Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051
31. Seed, B., and Sheen, J-Y. (1988) *Gene (Amst.)* **67**, 271–277
32. Tan, J., Joseph, D. R., Quarmby, V. E., Lubahn, D. B., Sar, M., French, F. S., and Wilson, E. M. (1988) *Mol. Endocrinol.* **2**, 1276–1285
33. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8998–9002
34. Ponglikitmongol, M., Green, S., and Chambon, P. (1988) *EMBO J.* **7**, 3385–3388
35. Huckaby, C. S., Connely, O. M., Beattie, W. C., Dobson, A. D. W., Tsai, M-J., and O'Malley, B. W. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 8380–8384
36. Jeltsch, J.-M., Turcotte, B., Garnier, J.-M., Lerouge, T., Krozowski, Z., Gronemeyer, H., and Chambon, P. (1990) *J. Biol. Chem.* **265**, 3967–3974
37. Zahraoui, A., and Cuny, G. (1987) *Eur. J. Biochem.* **166**, 63–69
38. Hughes, M. R., Malloy, P. J., Kieback, D. G., Kesterson, R. A., Pike, J. W., Feldman, D., and O'Malley, B. W. (1988) *Science* **242**, 1702–1705
39. Smale, S. T., and Baltimore, D. (1989) *Cell* **57**, 103–113
40. Means, A. L., and Farnham, P. J. (1990) *Mol. Cell. Biol.* **10**, 653–661
41. Baarends, W. M., Themmen, A. P. N., Blok, L. J., Mackenbach, P., Brinkmann, A. O., Meijer, D., Faber, P. W., Trapman, J., and Grootegoed, J. A. (1990) *Mol. Cell. Endocrinol.* **74**, 75–84
42. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J-M., Argos, P., and Chambon, P. (1986) *Nature* **320**, 134–139
43. Misrahi, M., Loosfelt, H., Atger, M., Meriel, C., Zerah, V., Dessen, P., and Milgrom, E. (1988) *Nucleic Acids Res.* **16**, 5459–5472
44. Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H., and Chambon, P. (1990) *EMBO J.* **9**, 1603–1614
45. Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G., and Evans, R. M. (1985) *Nature* **318**, 635–641
46. Arriza, J. L., Weinberger, C., Cerelli, G., Glaser, T. M., Handelin, B. L., Housman, D. E., and Evans, R. M. (1987) *Science* **237**, 268–275
47. Sassevage, N. L., Smith, M., Gillam, S., Woychick, R. P., and Rottmann, F. M. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 223–227
48. Wiedemann, L. M., and Perry, R. P. (1984) *Mol. Cell. Biol.* **4**, 2518–2528
49. Simonsen, C. C., and Levinson, A. D. (1983) *Mol. Cell. Biol.* **3**, 2250–2258
50. Blok, L. J., Mackenbach, P., Trapman, J., Themmen, A. P. N., Brinkmann, A. O., and Grootegoed, J. A. (1989) *Mol. Cell. Endocrinol.* **63**, 267–271

Supplementary material to

CHARACTERIZATION OF THE TRANSCRIPTION UNIT OF THE HUMAN ANDROGEN RECEPTOR GENE

Peter W. Faber, Henri C.J. van Rooij, Hetty A.G.M. van der Korput, Willy M. Baarends, Albert O. Brinkmann, J. Anton Grootegoed and Jan Trapman.

## MATERIALS AND METHODS

### cDNA and genomic DNA libraries

Random primed T47D mammary tumor cDNA libraries were kind gifts of Dr. Milgrom (Paris). The oligo(dT) primed prostate and testis cDNA libraries were purchased from Clontech (Palo Alto, CA). The human genomic DNA library was provided by Dr. Grosveld (Rotterdam). The libraries were plated, transferred to nitrocellulose filters (Schleicher & Schuell, Dassell, FRG) and hybridized to $^{32}$P-labelled DNA probes according to standard procedures (21-23).

### Isolation and characterisation of cDNA and genomic DNA clones

Positive phages were plaque-purified by three rounds of purification. All experiments concerning the isolation, subcloning and characterization of phage DNA inserts by restriction mapping and Southern blot hybridization were performed by standard procedures (21-23). Sequence analysis was done by the dideoxy chain termination method (24) on both strands of fragments, which were subcloned in M13mp18/19, using T7 polymerase (Pharmacia, Uppsala, Sweden). When necessary, (G+C)-rich regions were sequenced by the chemical cleavage method (25).

### RNA preparation and Northern blot analysis

Total cellular RNA was isolated by the guanidinium thiocyanate method (26). Poly(A)$^+$ RNA was prepared by oligo(dT) affinity chromatography. For Northern blot analysis, glyoxal denatured poly(A)$^+$ RNA was separated on agarose gels and blotted onto nylon membranes (GeneScreen, NEN, Boston). Hybridizations with $^{32}$P-labelled probes were carried out at 42 °C in 50% formamide under standard conditions.

### S1-nuclease protection experiments

A single stranded DNA probe was prepared using primer G (see Fig. 3A). This primer was end-labelled, using $^{32}$P-ATP and T4 polynucleotide kinase (Gibco BRL, Grand Island, NY) and annealed to a single stranded M13 phage containing the 1.1 kb SstI-PvuII fragment (Fig. 3A). After elongation by Klenow DNA polymerase (Pharmacia) and digestion with BssHII, the resulting single stranded probe was isolated from a 6% polyacrylamide gel. Approximately 1x10$^5$ cpm of probe was annealed to 20-30 ug total RNA (from LNCaP cells) in 80% formamide, 40 mM Pipes (pH=6.5), 0.4 M NaCl and 1 mM EDTA overnight at 55 °C. S1-nuclease (Boehringer, Mannheim, FRG) digestions were carried out for 1 h at 37 °C and the resulting protected fragments were analysed on a 6% polyacrylamide gel (27).

### DNA amplification

Oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesizer. Amplification by the Polymerase Chain Reaction (PCR) (28) was performed in 100 ul reaction mixtures containing 1 ug of genomic DNA or 2% of the first strand cDNA preparation, using 2U TAQ polymerase (Promega, Madison, WI) under conditions as described by the manufacturer. For first strand cDNA preparations, 5 ug total RNA or 1 ug poly(A)$^+$ RNA was annealed to 100 ng of the appropriate primer and cDNA synthesis was carried out using AMV reverse transcriptase according to a standard protocol (Promega). Amplification was performed in a Bioexcellence DNA incubator during 24 cycles (genomic DNA) or 30 cycles (cDNA). Standard conditions were: denaturation for 1 min at 95 °C, annealing for 2 min at 60 °C and extension for 1-5 min at 70 °C. The amplification products were recovered after chloroform/isoamylalcohol extraction, electrophoresed on 1-1.5% agarose gels and blotted onto nitrocellulose filters for hybridization. For isolation and subcloning the products were extracted with phenol/chloroform and ethanol precipitated. The oligonucleotides used were:

A: 5'-GCTTTCCTCTAGACTGGAAC-3' (sense, +10046 to +10065),
B: 5'-CTGCCTTCACCTAGAAATG-3' (anti-sense, +9363 to +9345),
C: 5'-TCTGCCTTCAACTGCAGATAAC-3' (anti-sense, +8536 to +8515),
D: 5'-CCATCTGGTGAGTTTACTC-3' (sense, +5644 to +5662),
E: 5'-CCATTTGGATCCAGGTCTGCT-3' (sense, +4901 to +4921),
F: 5'-TGCAAGAGGCGTTGGCTGT-3' (anti-sense +171 to +189),

G: 5'-TGTACAGCACTGGAGCGGCTA-3' (anti-sense, +106 to +125),
H: 5'- GAGCCAGCTTGCTGGGAGA-3' (sense +25 to +43),
I: 5'- TCTCCAAAGCCACTAGGCAG-3' (sense -133 to -152),
T12-X: 5'-TTTTTTTTTTTTGGATCCGAATTCGCATGC-3',
X: 5'-GGATCCGAATTCGCATGC-3'.

CAT reporter gene constructs

The 1.1 kb SstI-PvuII (-737 to +378) and the 0.75 kb SstI-SmaI (-737 to +21) fragment were isolated from Gm 20 (Fig. 3A). The SstI site was blunt-ended using T4 polymerase (Promega). The fragments were cloned in the blunt-ended HindIII site of the promoterless pCAT-Enhancer plasmid (Promega).

Transfection assay

Four days prior to transfection, 2-3x$10^5$ LNCaP cells were seeded into 6 cm dishes. Culture medium was Dulbecco's-MEM supplemented with 5% fetal calf serum and antibiotics. Approximately 40% confluent cell cultures were transfected by the calcium phosphate precipitation method (29). Per culture dish, 10 ug of DNA (5 ug CAT reporter gene construct and 5 ug carrier DNA [pTZ, Pharmacia]) was used. All experiments were performed in duplicate with two different plasmid isolations. In part of the experiments, the pCH110 beta-galactosidase expression plasmid (Pharmacia) was co-transfected as an internal control. Cell extracts were prepared 48 h following the transfection and the CAT assays were done essentially as described (30). For quantification butyryl CoA was used, which allows a direct counting of the butyrylated chloramphenicol (31).

## RESULTS

Characterization of hAR cDNA clones

Previously we have described the isolation of a cDNA clone (7A2M27) (15) and a genomic DNA fragment (Gm 20) (18) from which the hAR open reading frame (ORF) of 2730 nucleotides was deduced (Fig.1). The cDNA clone 7A2M27 contains the information for the DNA-binding and ligand-binding domains, and the genomic DNA clone Gm 20 contains hAR exon 1, which encodes the complete N-terminal domain of the receptor and partly overlaps with 7A2M27 (Fig. 1 and Refs 15,18). A genomic PstI-PstI fragment (probe 2), spanning the ATG translation initiation codon, and an EcoRI-EcoRI cDNA fragment (probe 3) were used to screen random primed T47D cDNA libraries for overlapping cDNA clones. Whereas no additional clones could be identified using probe 2, several were found by screening the libraries with probe 3. The longest clone isolated (2-4-1) with a size of approximately 4kb is depicted in Fig.1. Rescreening the libraries with a 3' situated fragment of clone 2-4-1 (StuI-EcoRI, probe 5) resulted in the isolation of clone 0.3A, which also has a size of about 4kb. Together clones 2-4-1 and 0.3A span over 6kb of the 3'-untranslated region (UTR) of the hAR cDNA. A comparison of the cDNA restriction map to that of a genomic fragment (Gm 100), which contains the last protein coding exon (exon 8), showed the maps to be completely identical (see Fig.1 and Ref. 20).

Subsequently, oligo(dT) primed cDNA libraries were screened using probes 3, 5 and 6, the latter is a 3' situated fragment (PvuII-EcoRI) of clone 0.3A. Hybridization with probes 3 and 6 resulted in the identification of several cDNA clones, whereas no clones could be detected using probe 5. The set of clones detected by probe 3 originated from a position just downstream from the ORF, at which point an A-rich region has been reported in the rat AR (rAR) mRNA (32). Sequence analysis of clone 2-4-1 confirmed the presence of a similar stretch of A-residues in the hAR mRNA (data not shown). Subsequent structural analysis of the cDNA clones obtained by hybridization with probe 3 showed all of them to be primed in this A-rich region. The cDNA clones detected by probe 6, of which clone hP-5A is given as an example (Fig. 1), originated from a point just downstream of the 3' end of clone 0.3A. Sequence analysis showed all of these clones to have a short stretch of A residues at their 3' end at the same position, however, no polyadenylation signal could be identified. Structural analysis of the

corresponding genomic DNA region from clone Gm 100 demonstrated the presence of the sequence AAAAAAAAGAAAAAAAAAATTAAAAA at the precise point of the short poly(A) stretches in the cDNAs. Therefore, all cDNA clones isolated from oligo(dT) primed cDNA libraries were found to be primed on poly(A) stretches present in the hAR gene. Library screening with a genomic DNA fragment from clone Gm 100 (XbaI-PstI, probe 7) failed to identify any additional cDNA clones.

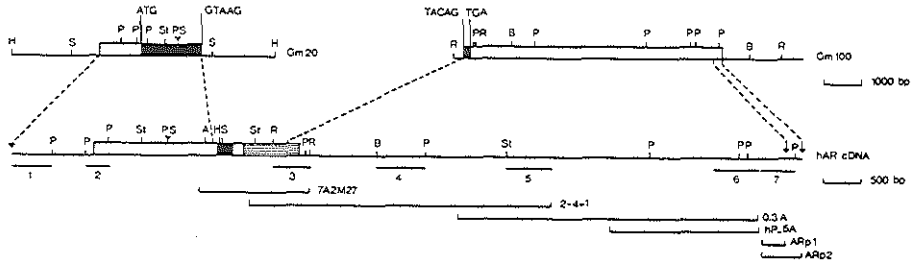


FIGURE 1: Partial restriction map of the complete hAR cDNA. The 2730 bp ORF is shown as a box in which the black and hatched areas represent the DNA-binding and ligand-binding domain, respectively. Above the cDNA a 7 kb HindIII-HindIII genomic DNA fragment (Gm 20) and a 9 kb EcoRI-(SalI) genomic DNA fragment (Gm 100) (20), which contain the complete first and last exons of the hAR transcription unit, respectively, are shown. The exons are boxed and the positions of the translation initiation (ATG) and termination (TGA) codons as well as the splice donor (GTAAG) and acceptor (TACAG) sequences at the end of exon 1 and the beginning of exon 8 are indicated (see Refs 15,18,20). Arrows represent transcription initiation sites in case of Gm 20 and sites of polyadenylation in case of Gm 100 (see text). Under the cDNA the positions of several hybridization probes are given (1 to 7) as well as the cDNA clones representing the 3'-UTR of the hAR mRNA. A = Asp718, B = BamHI, H = HindIII, P = PstI, R = EcoRI, S = SstI, St = StuI.

Sequence of the region spanning the transcription initiation sites of the hAR gene

Both hAR and rAR cDNA clones containing part of the 5'-UTR, with lengths of 531 and 993 nucleotides respectively, have been reported (16, 32). The genomic DNA sequence of clone Gm20 (Fig. 1) upstream from the ATG initiation codon was determined and compared to the rAR cDNA sequence. A strong homology (>70%) was observed up to position 1112 in the human genomic sequence. Part of this sequence surrounding the begin point of the rAR cDNA is given in Fig. 2. The difference in length is due to several gaps in the rat sequence. The presence of this sequence in the hAR mRNA was confirmed in a PCR experiment, in which the region between primers G and H (Figs 2 and 3A) was sequenced after amplification from a first strand cDNA preparation, made with primer F and LNCaP total RNA (data not shown). No PCR product could be produced from a first strand cDNA preparation from which reverse transcriptase (RT) had been omitted and in the same experiment no PCR product could be obtained using the primers G and I (Figs 2 and 3A), suggesting the presence of either an intron/exon boundary or the hAR gene promoter in this region. However, the region lacked well known promoter elements (TATA and CAAT-boxes or G+C-rich) and splice acceptor/branchpoint sequences.

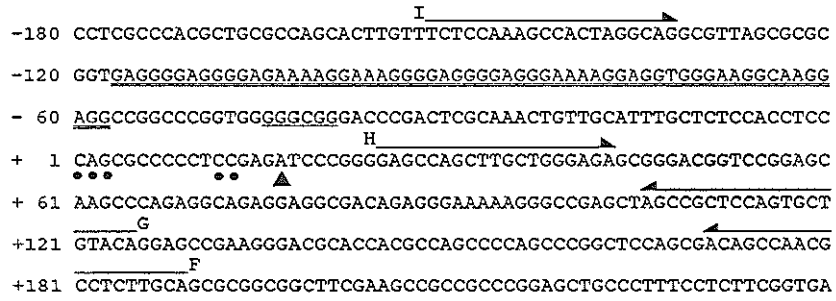


FIGURE 2: Sequence of the region spanning the hAR gene transcription initiation sites. Shown is the sequence from position +240 to -180 (see text). The major transcription initiation sites (see text) are indicated with dots. Arrows represent primers used in RT-PCR experiments. The position of the closed triangle marks the start of the rAR cDNA sequence. Two putative promoter elements, the consensus Sp1 binding site (-40 to -45) and the homopurine stretch (-58 to -117) are underlined and double underlined, respectively.

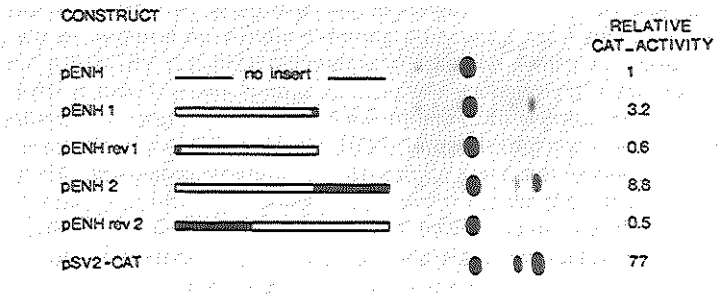


FIGURE 4: Functional activity of the hAR gene promoter. On the left hand side the constructs used in the transfection are shown. pENH is the pCAT-Enhancer plasmid. pENH 1 (SstI-PvuII), pENH rev1 (PvuII-SstI), pENH 2 (SstI-SmaI) and pENH rev2 (SmaI-SstI) are the four hAR gene promoter constructs of which the inserts are shown. The open part of the horizontal bar represents 5' flanking sequence and the closed part of the bar represents 5'-UTR sequence. On the right hand side the autoradiograph of the products of the CAT assay, separated by thin layer chromatography, are shown. Direct counting of the butyrylated chloramphenicol resulted in the relative CAT activities presented.

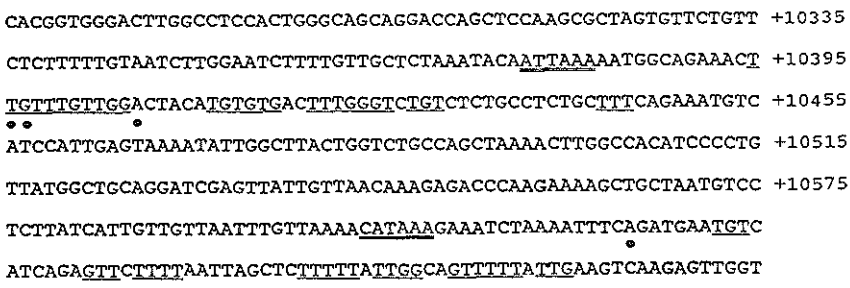


FIGURE 7: Sequence of the genomic region surrounding the two polyadenylation signals of the hAR mRNA. Numbering of the nucleotides is relative to the transcription initiation sites. The two polyadenylation signals at positions +10377 (ATTAAA) and +10604 (CATAAA) are double underlined. The actual sites of poly(A) addition are indicated with dots and the TG-boxes following the polyadenylation signals are underlined.

## DISCUSSION

The present study concerns the isolation and characterization of the complete transcription unit and the promoter region of the hAR gene. A full lenght hAR cDNA of 10.6 kb was constructed from cDNA and genomic clones. In addition to the previously reported 2.7 kb ORF this cDNA consists of a long 5'-UTR (1.1 kb) and a very long 3'-UTRs (6.8 kb) that were encoded by the protein coding exons 1 and 8 (this study and Ref. 20). This fixes the number of exons in the hAR gene to eight. A comparison of the hAR gene to the human estrogen receptor (hER) (34) and chicken progesterone receptor (cPR) (35,36) gene shows the three genes to be similarly organized. All these genes contain eight exons with conserved intron/exon boundaries; the complete 5'-UTR is in exon 1 and the 3'-UTR is in exon 8. The intron/exon junctions are less conserved when compared to genes encoding more distant members of the same receptor family, such as the chicken c-erb-A/thyroid hormone receptor (37) and the human vitamin D3 receptor (38). This indicates an early divergence in evolution between the steroid hormone receptors and other receptors of the same family.

The promoter region of the hAR gene does not follow the most common rules of promoter structure. Two major transcription initiation sites were identified in a 13 bp region. The sequence upstream from these transcription initiation sites does not contain the classical TATA and CCAAT box elements. Neither does the sequence surrounding the transcription initiation sites conform to the recently identified "initiator" sequence (39) nor to the "HIP1" sequence (40), found around the transcription initiation sites of several (G+C)-rich promoters. The most notable features of the hAR gene promoter are a short stretch of G and C residues containing one consensus Sp1 binding site and a homopurine stretch of 60 bp containing several GGGGA and GGGA sequence motifs. These elements are also present in the recently identified rAR gene promoter (41).

A comparison of the promoter region of the hAR gene to that of the hER and cPR genes shows these three promoters to be completely different (35,36,42). The promoter of the hER gene contains TATA and CCAAT box elements, whereas the

promoter of the cPR has a large (G+C)-rich region, surrounding the transcription initiation sites. Similar to the hAR promoter, in the human PR B and rabbit PR promoter regions canonical TATA and CCAAT boxes are absent (43,44). However, excluding a Sp1 binding site, there is no obvious homology with the hAR promoter. These comparisons indicate that if the steroid hormone receptor genes originate from one primordial gene, the receptors must have adapted to different expression and functioning by a rapid divergence of the first exons and the attached promoter regions. Alternatively there was no primordial gene and the individual genes originated from a fusion event of one set of exons encoding the DNA-binding and ligand-binding domains to a number of different exons, supplying the information for the both $NH_2$-terminal domains of the receptors and the promoter regions.

The untranslated regions of the hAR mRNA (5'-UTR= 1.1 kb and 3'-UTR= 6.8 kb) are remarkably long. The 5'-UTR even is one of the longest of any gene published so far. Untranslated regions have been implicated both in control of mRNA transcription, translation and processing. So far our experiments indicate the involvement of the 5'-UTR in either control of transcription or translation. The precise nature of this effect is being investigated at present. The 3'-UTR is the longest of all steroid hormone receptor mRNAs although they all are characterized by long 3'-UTRs (34,43,45,46). As yet no studies regarding their role have been performed. Two equally effective functional polyadenylation signals ATTAAA (+10377) and CATAAA (+10604) direct poly(A) addition to the hAR mRNA. Although polyadenylation usually takes place at a single nucleotide following the polyadenylation signal, three different sites were used follwing the ATTAAA hexamer sequence. A similar heterogeneity in poly(A) addition has previously only been reported for the bovine prolactin (47) and mouse ribosomal L30 (48) RNAs, where an AATAAA polyadenylation signal was followed by poly(A) addition at several sites in a 12-14 bp region. In case of the hepatitis B surface antigen (49) a TATAAA polyadenylation signal directed polyadenylation at positions 13, 14 and 20 following the polyadenylation signal. Although the precise structural requirements which are responsible for the heterogeneity of poly(A) addition sites are unknown, a possible explanation could stem from the inadequacy of the ATTAAA which then results in both the observed scattering of poly(A) addition sites as well as the use of a second immediately downstream located polyadenylation signal with the same frequency. However, for some reason this region must have a selective advantage to direct poly(A) addition to the hAR mRNA as similar signals were also identified in other parts of the 3'-UTR but were not functional.

The alternative splice event which deletes approximately 3 kb of the hAR mRNA 3'-UTR explains the nature of the 8.5 kb hAR mRNA that was reported in RNA preparations from LNCaP prostatic carcinoma cells (15). A hAR mRNA of 7 kb has also been reported from human protatic tissue in conjunction with a 10 kb full length hAR mRNA (14). We consider it to be quite well possible that the 7/8.5 kb hAR mRNAs are identical. All reports dealing with AR mRNAs from other than human scources only mention a 10 kb mRNA [rat (14,33,50), mouse (14)]. Whether or not this alternatively spliced hAR mRNA is present in all tissues in which the hAR is expressed, and whether the splicing event influences the stability or the translation efficiency of the transcript remains to be investigated.

In summary, we have defined the complete transcription unit of the hAR. Several potentially regulatory elements/events which could influence hAR expression, hAR mRNA translation and processing were identified. Further detailed analysis of the regulation of hAR expression will contribute to the elucidation of the role of the AR in normal and abnormal male sexual development and differentiation.

## CHAPTER V

# THE MOUSE ANDROGEN RECEPTOR: FUNCTIONAL ANALYSIS OF THE PROTEIN AND CHARACTERIZATION OF THE GENE

FABER, P.W., KING, A, van ROOIJ, H.C.J., de BOTH, N.J., BRINKMANN, A.O., & TRAPMAN. J.

THE BIOCHEMICAL JOURNAL 278, 269-278 (1991).

Biochem. J. (1991) 278, 269-278 (Printed in Great Britain)



# The mouse androgen receptor

## Functional analysis of the protein and characterization of the gene

Peter W. FABER,* Audrey KING,* Henri C. J. VAN ROOIJ,* Albert O. BRINKMANN,† Nico J. DE BOTH* and Jan TRAPMAN*

Departments of *Pathology and †Biochemistry II, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Screening a mouse genomic DNA library with human androgen-receptor (hAR) cDNA probes resulted in the isolation and characterization of eight genomic fragments that contain the eight exons of the mouse androgen-receptor (mAR) gene. On the basis of similarity to the hAR gene, the nucleotide sequences of the protein-coding parts of the exons as well as the sequences of the intron/exon boundaries were determined. An open reading frame (ORF) of 2697 nucleotides, which can encode an 899-amino-acid protein, could be predicted. The structure of the mAR ORF was confirmed by sequence analysis of mAR cDNA fragments, which were obtained by PCR amplification of mouse testis cDNA, using mAR specific primers. A eukaryotic mAR expression vector was constructed and mAR was transiently expressed in COS-1 cells. The expressed protein was shown by Western blotting to be identical in size with the native mAR. Co-transfection of HeLa cells with the mAR expression plasmid and an androgen-responsive chloramphenicol acetyltransferase (CAT) reporter-gene construct showed mAR to be able to *trans*-activate the androgen-responsive promoter in a ligand-dependent manner. Transcription-initiation sites of the mAR gene were identified by S1-nuclease protection experiments, and the functional activity of the promoter region was determined by transient expression of mAR promoter–CAT-reporter-gene constructs in HeLa cells. Structural analysis revealed the promoter of the mAR gene to be devoid of TATA/CCAAT elements. In addition, the promoter region is not remarkably (G+C)-rich. Potential promoter elements consist of a consensus Sp1 binding sequence and a homopurine stretch. The polyadenylation sites of mAR mRNA were identified by sequence similarity to the corresponding sites in the hAR mRNA.

## INTRODUCTION

The androgen receptor (AR) is a transcription-regulating protein that plays a pivotal role in the programming of male sexual differentiation and development. Absence or mutation of the X-chromosome-located AR gene can lead to complete androgen-insensitivity, and the affected 46,*XY* individual displays the external phenotype of a female [1–7]. Mutations in AR are also thought to be involved in less severe forms of aberrant male sexual development.

Structurally and functionally AR belongs to the superfamily of ligand-responsive transcription modifiers which encompasses the receptors for the steroid and thyroid hormones, retinoic acid, vitamin $D_3$ and several 'orphan' receptors for which a ligand has not as yet been identified [8–11]. The structural regions that these receptors have in common, namely a *C*-terminal ligand-binding domain, an internal DNA-binding domain, consisting of two Cys–Cys zinc finger motifs and an *N*-terminal hypervariable 'regulatory' domain, enable them to react to a hormonal stimulus by modulation of gene transcription through recognition of, and binding to, hormone-responsive elements (HREs) located in the control regions of target genes.

The isolation of human AR (hAR) and rat AR (rAR) cDNAs [12–17] and the elucidation of the structural organization of the hAR gene [18–20a] has provided important tools to address questions regarding AR function in the androgen signal-transduction pathway. In common with the glucocorticoid receptor and the progesterone receptor, the androgen receptor is able to regulate expression of a mouse-mammary-tumour-virus long-terminal-repeat (MMTV LTR)-driven promoter, although perhaps less efficiently [21–23]. This indicates that the DNA motif recognized by the AR is identical with, or closely related to, the GRE and PRE consensus sequence GGTACAnnnTGTTCT.

In order to extend findings to different species, it is necessary to isolate species-specific tools. Especially for detailed studies *in vivo* concerning molecular and genetic mechanisms of AR synthesis and function, including transgenic animals, and study of the physiological effects of manipulation of the AR system, knowledge of the mouse AR (mAR) system is of high importance.

In the present study we describe the characterization of the mAR gene and cDNA. A mAR expression vector (pmAR$^a$) was constructed and applied for the functional characterization of the mAR protein. In addition, the promoter of the mAR gene is structurally and functionally characterized.

## MATERIALS AND METHODS

### Isolation and characterization of mAR genomic DNA clones

A mouse genomic DNA library in lambda EMBL3 SP6/T7 was purchased from Clontech (Palo Alto, CA, U.S.A.). The library was plated, transferred to nitrocellulose filters (Schleicher and Schüll, Dassel, Germany) and hybridized to $^{32}$P-labelled hAR cDNA probes (see Fig. 2*a* below), according to standard procedures [24,25]. Positive phages were isolated by three rounds of purification. All experiments concerning the isolation, subcloning and characterization of phage DNA inserts by restriction mapping and Southern-blot hybridization were performed

Abbreviations used: hAR, mAR and rAR: human, mouse and rat androgen receptor; ORF, open reading frame; CAT, chloramphenicol acetyltransferase; HRE, hormone-responsive element; MMTV, mouse mammary-tumour virus; MCS, multiple cloning site; MEM, minimal essential medium; GRE/PRE, glucocorticoid/progesterone-responsive element; cPR, chicken progesterone receptor; hER, human oestrogen receptor; LTR, long terminal repeat; UTR, untranslated region; SV40, simian virus 40; PMSF, phenylmethanesulphonyl fluoride; DTT, dithiothreitol.

These sequence data have been submitted to the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession nos. X59590 [mouse (1A) androgen-receptor cDNA], X59591 [mouse (5′-MAR) androgen-receptor cDNA and X59592 [mouse (5′-MAR, 1SA, 1SB, 1BA, 19A, 6B, 3A, 1A) androgen-receptor cDNA].

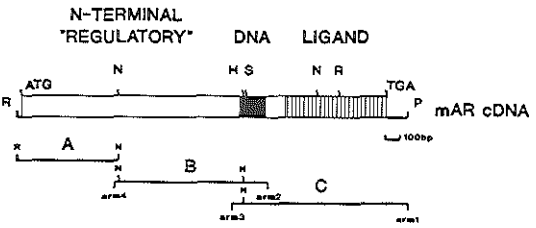


**Fig. 1. Partial restriction map of mAR cDNA**

Indicated are the AR domain structure as well as the three DNA fragments from which the cDNA was constructed. Key to restriction endonucleases: H, *Hin*dIII; N, *Nco*I; P, *Pst*I; R, *Eco*RI; S, *Sst*I. Arm1–Arm4 are mAR-specific primers (see the text and Fig. 3).

according to standard procedures [24,25]. Sequence analysis was done by the dideoxy-chain-termination method on both DNA strands using single-stranded M13mp18/19 DNA preparations [26] and T7 polymerase (Pharmacia, Uppsala, Sweden).

### RNA preparation and S1-nuclease-protection experiments

Total cellular RNA was isolated by the guanidinium thiocyanate method [27]. For the S1-nuclease-protection assay a 0.6 kb *Eco*RI–*Hin*dIII double-stranded DNA fragment (see Fig. 5*a* below) was used as a probe. The fragment was end-labelled using T4 polynucleotide kinase (Gibco BRL, Grand Island, NY, U.S.A.), and approx. $1 \times 10^6$ c.p.m. of probe were annealed to 40 $\mu$g of RNA in 80% formamide/40 mM-Pipes (pH 6.5)/0.4 M-NaCl/1 mM-EDTA overnight at 55 °C [28]. S1-nuclease (Boehringer, Mannheim, Germany) digestions were carried out for 1 h at 37 °C, and the resulting protected fragments were analysed on a denaturing 6% (w/v) polyacrylamide gel.

### DNA amplification

Oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesizer. Amplification by the PCR [29] was performed in 100 $\mu$l reaction mixtures containing 2% of a first-strand cDNA preparation, using 2 units of *Taq* DNA polymerase (Promega, Madison, WI, U.S.A.) under the conditions described by the manufacturer. For first-strand cDNA preparations, 100 ng of the appropriate primer was annealed to 5 $\mu$g of total RNA, and cDNA synthesis was performed with avian-myeloblastosis-virus reverse transcriptase according to a standard protocol (Promega). Amplification was performed in a Bioexcellence DNA incubator for 30 cycles. Standard conditions were: denaturation for 1 min at 95 °C, annealing for 2 min at 60 °C and extension for 1–5 min at 70 °C. The amplification products were recovered after chloroform/3-methylbutan-1-ol extraction, electrophoresed through 1–1.5%-(w/v)-agarose gels and blotted on nitrocellulose filters for hybridization. For isolation and subcloning, the fragments were extracted with phenol/chloroform and ethanol-precipitated. The oligonucleotides used (see Fig. 3) were:

arm1: 5′-CAGAGAAGTAGTGCAGAGTT-3′
(anti-sense, 3309–3328)

arm2: 5′-CAGAGTCATCCCTGCTTC-3′
(anti-sense, 2062–2079)

arm3: 5′-TTTGGACAGTACCAGGGACC-3′
(sense, 1747–1766)

arm4: 5′-AGTGCCAAGGAGTTGTGTAA-3′
(sense, 802–821)

### Chloramphenicol acetyltransferase (CAT) reporter gene constructs for analysis of promoter activity

The 0.5 kb *Pst*I–*Hin*dIII and the 1.3 kb *Pst*I–*Pst*I genomic DNA fragments were isolated from a subcloned 1.5 kb *Eco*RI–*Eco*RI fragment (see Fig. 5*a* below). The *Pst*I *Pst*I fragment was cloned in both orientations in the *Pst*I site of the multiple cloning site (MCS) of the promoterless pCAT-Enhancer plasmid (Promega; referred to as 'pmAR-CAT-0'). This resulted in the construction of pmAR-CAT-1 and pmAR-CAT-1 rev (reverse) respectively. The 0.5 kb *Pst*I–*Hin*dIII (blunt-ended) fragment was cloned in a *Pst*I–*Acc*I (blunt-ended) pCAT-Enhancer vector, and in pCAT-Enhancer (*Pst*I–*Hin*dIII), resulting in the construction of pmAR-CAT-2 and pmAR-CAT-2 rev, which contain the fragment in the normal and reverse orientation respectively.

### Construction of the mAR expression plasmid pmAR°

Using genomic DNA fragments, and cDNA fragments obtained by PCR amplification a mAR expression plasmid, containing the complete mAR ORF, was constructed. A partial restriction map of the mAR cDNA, together with the three fragments from which it was derived, is presented in Fig. 1. Fragment A (*Eco*RI–*Nco*I) was derived from the 2 kb *Eco*RI–*Eco*RI genomic DNA fragment that contains the protein-coding region of exon 1 (see Fig. 5*a* below), whereas the fragments B (*Nco*I–*Hin*dIII) and C [*Hin*dIII–3′-untranslated region (UTR)] originated from reverse-transcriptase PCR using a first-strand cDNA preparation made with primer arm1 (Fig. 3 below) and mouse testis RNA. The fragments were produced by PCR with the primer combinations arm1 and arm3 (C) and arm2 and arm4 (B) respectively. The cDNA construct was prepared in a pGEM vector and completely sequenced. A mAR expression vector was constructed using the cDNA containing the complete mAR ORF, the simian-virus-40 (SV40) early promoter and the rabbit $\beta$-globin polyadenylation signal [21]. The resulting vector is referred to as 'pmAR°'.

### Transfection assay

A day before transfection $3 \times 10^5$ HeLa cells were seeded into 6 cm-diameter dishes. Culture medium was Dulbecco's MEM supplemented with 5% (v/v) fetal-calf serum and antibiotics. The approx. 40%-confluent cell cultures were transfected by the calcium phosphate method [30]. For *trans*-activating studies 2.5 $\mu$g of pmAR° was used in combination with 2.5 $\mu$g of reporter plasmid (pG 29 G-TK-CAT/pBL2-CAT) [31,32], 2.5 $\mu$g of pCH110 ($\beta$-galactosidase expression plasmid; Pharmacia) and 2.5 $\mu$g of carrier DNA (pTZ; Pharmacia). For promoter studies

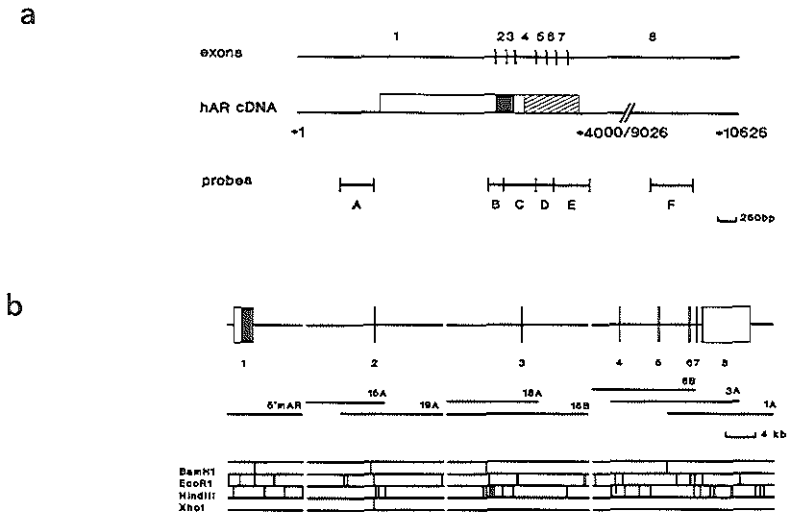


**Fig. 2. Molecular cloning and characterization of the mAR gene**

(*a*) Probes used in the screening of the mouse genomic DNA library. Depicted is a schematic representation of the full-length hAR cDNA (10625 nucleotides). The ORF is shown as a box with the black and hatched areas representing the DNA- and ligand-binding domains respectively. The distribution of the hAR cDNA sequence over the eight exons of the hAR gene is indicated above the hAR cDNA. The hybridization probes used (A–F) are indicated under the cDNA. (*b*) Structural organization of the mAR gene. The eight exons are numbered 1–8. The positions of the eight individual clones (5'mAR, 15A, 19A, 18A, 15B, 6B, 3A and 1A) are shown as horizontal bars. A restriction map of the genomic clones after digestion with *Bam*HI, *Eco*RI, *Hin*dIII and *Xho*I is presented at the bottom.

5 $\mu$g of the appropriate mAR promoter construct was used together with 2.5 $\mu$g of pCH110 and 2.5 $\mu$g of pTZ. pSV2-CAT (5 $\mu$g) was used in control experiments. All experiments were carried out in duplicate with at least two different plasmid preparations. Cell extracts were prepared at 48 h after transfection, and CAT assays were performed essentially as described in [33]. For quantification, butyryl-CoA was used, which allows the direct liquid-scintillation counting of the butyrylated chloramphenicol [34]. For immunoprecipitation and Western blotting, $2 \times 10^6$ COS-1 cells were transfected using 10 $\mu$g of pmAR° and 10 $\mu$g of pTZ.

### Immunoprecipitation and Western-blot analysis

COS-1 cells, transfected with the mAR and hAR expression plasmids pmAR° and pAR° [21], mock-transfected COS-1 cells and mouse testicular tissue were lysed in 40 mM-Tris (pH 7.4)/1 mM-EDTA/10 % (v/v) glycerol/10 mM-dithiothreitol (DTT)/50 mM-NaF/0.6 mM-phenylmethanesulphonyl fluoride (PMSF)/0.1 mM-bacitracin/0.5 mM-leupeptin/1 % (v/v) Triton/0.5 % (w/v) sodium deoxycholate/0.08 % (w/v) SDS. For the immunoprecipitation the anti-hAR monoclonal antibody F39.4.1 [35] was used. The non-specific control monoclonal antibody was ER-Pr 27, which recognizes prostate-specific antigen [36]. Western blotting and immunostaining with the anti-hAR polyclonal antibody Sp061 (diluted 1:1000) were done as described in [37].

## RESULTS

### Molecular cloning and characterization of the mAR gene

To isolate the mAR gene, a mouse genomic DNA library was screened using hAR cDNA probes. Previously we elucidated the organization of the hAR gene and showed that the hAR gene is composed of eight exons [18, 20a]. In Fig. 2(*a*) the hAR cDNA is schematically depicted, together with the cDNA probes used and the distribution of the hAR cDNA sequence over the exons of the hAR gene. The probes cover the complete protein-coding regions of the hAR gene exons 1–8. Screening of the mouse genomic DNA library resulted in the isolation of eight clones (Fig. 2*b*). Probes A and B hybridized to clone 5'mAR; clones 15A and 19A were both recognized by probes B and C; probe C hybridized to clones 15B and 18A; clones 6B and 3A could be identified by probes C, D and E; probes D, E and F recognized clone 1A. A physical map of the mAR gene was constructed on the basis of restriction-enzyme-digestion patterns combined with sequence analysis of the individual exons (see Fig. 2*b*). For this sequence analysis, mouse genomic DNA fragments hybridizing with hAR cDNA probes were subcloned in plasmid vectors and mapped in more detail. Small fragments containing the exons were subsequently cloned in M13 vectors and sequenced. Fig. 2(*b*) shows the positions of the eight mAR exons and the restriction map for *Bam*HI, *Eco*RI, *Hin*dIII and *Xho*I. The sequence obtained is presented in Fig. 3 and is discussed in detail below. Additional hybridization experiments using probes

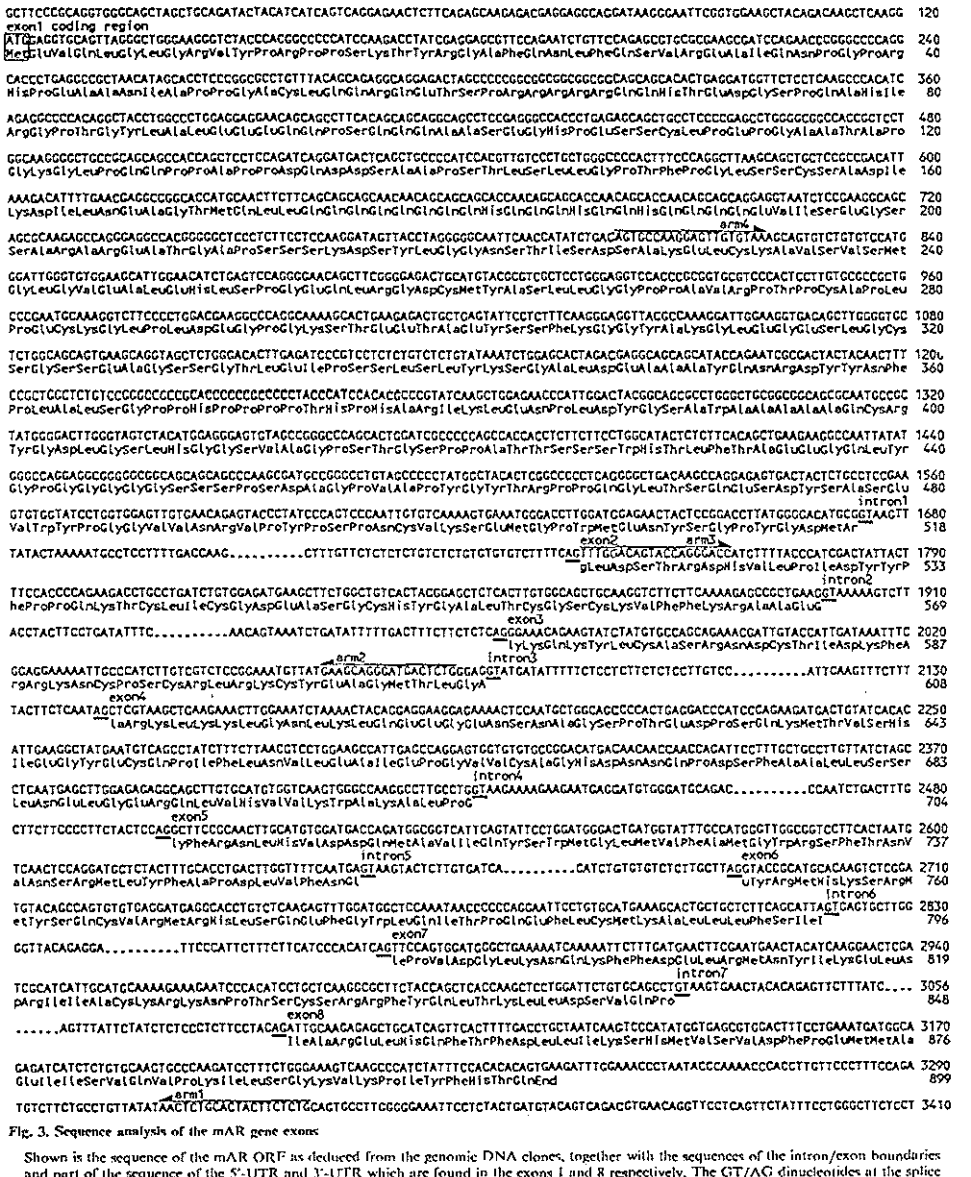


Fig. 3. Sequence analysis of the mAR gene exons

Shown is the sequence of the mAR ORF as deduced from the genomic DNA clones, together with the sequences of the intron/exon boundaries and part of the sequence of the 5'-UTR and 3'-UTR which are found in the exons 1 and 8 respectively. The GT/AG dinucleotides at the splice donor and acceptor sites are doubly underlined. The ATG translation-initiation codon is boxed, and the TGA translation-termination codon is marked 'End'. The number of each individual exon is given after the AG splice acceptor sequence; the number of the individual intron is given above the GT splice donor site. The positions of the four primers used in the construction of the full-length mAR cDNA (arm1–arm4) are indicated by horizontal arrows.

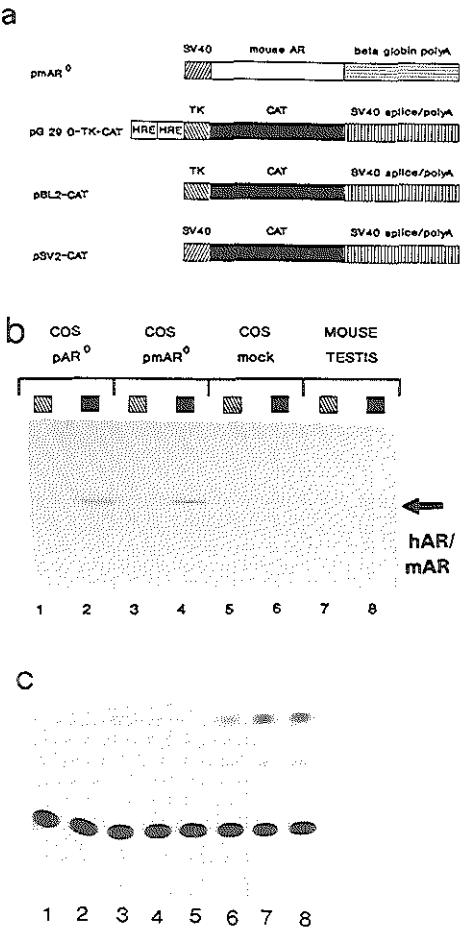


**Fig. 4. Functional analysis of the mAR**

(*a*) Constructs used for the functional assay of mAR activity. Shown are the mAR expression plasmid pmAR°, the androgen-responsive CAT reporter-gene construct pG 29 G-TK-CAT and the two control constructs, pBL2-CAT and pSV2-CAT. (*b*) Determination of the size of mAR. Shown is the Western-blot analysis of the mAR and hAR as obtained by immunoprecipitation from COS-1 cells and mouse testicular tissue. Immunoprecipitations were performed with specific (■) and non-specific monoclonal antibodies (□) using whole-cell extracts from pmAR°-transfected COS-1 cells (lanes 3 and 4), from pAR°-transfected COS-1 cells (hAR) (lanes 1 and 2), from mock-transfected COS-1 cells (lanes 5 and 6) and mouse testicular tissue (lanes 7 and 8). AR was revealed with an anti-hAR polyclonal antibody. The arrow indicates the position of AR. Molecular-mass markers were $\beta$-galactosidase (116 kDa) and phosphorylase *b* (97 kDa). (*c*) Determination of the *trans*-activating activity of mAR by co-transfection of HeLa cells with the androgen-responsive CAT reporter-gene construct pG 29 G-TK-CAT and pmAR° (lanes 5 and 6). Controls include pG 29 G-TK-CAT (lanes 1 and 2), a co-transfection of pmAR° and pBL2-CAT (lanes 3 and 4) and pSV2-CAT (lanes 7 and 8). Cells were grown in the presence (lanes 2, 4, 6, 8) or absence (lanes 1, 3, 5, 7) of the synthetic steroid R1881. The autoradiograph displays the conversion of [$^{14}$C]chloramphenicol into acetylated products.

specific for the promoter region of the hAR gene and the polyadenylation sites of the hAR mRNA indicated their presence in clones 5'mAR and 1A, respectively, and they are depicted as such in Fig. 2(*b*).

The eight overlapping clones cover a 60 kb region of genomic DNA. Although all the exons of the mAR gene are present in the clones isolated, part of the information for the introns 1, 2 and 3 is lacking. The lengths of the introns as deduced from these clones are >14.5 kb (intron 1), >19.1 kb (2), >12.4 kb (3), 5.2 kb (4), 4.1 kb (5), 1.0 kb (6) and 0.7 kb (7) respectively. The physical map of the mAR gene was checked by Southern-blot analysis of mouse genomic DNA digested with the restriction enzymes *Bam*HI, *Eco*RI and *Hin*dIII. Blots were probed with genomic probes spanning all exons of the mAR gene. These experiments confirmed the restriction map presented in Fig. 2(*b*) (results not shown). In addition, the sizes of the various hybridizing fragments detected in these genomic blots indicated that the length of intron 3 exceeds 20 kb. This finding extends the minimum length of the mAR gene to at least 68 kb.

### Sequence analysis of the mAR open reading frame (ORF)

As discussed above, small fragments containing exon information were sequenced to predict the structure of the mAR cDNA ORF. Protein-coding sequences as well as intron/exon boundaries were deduced on the basis of similarity to the hAR ORF and splice consensus sequences respectively. This resulted in a predicted ORF for the mAR cDNA shown in Fig. 3, together with the sequences of the intron/exon boundaries, which all conform to the GT/AG rule (doubly underlined in Fig. 3). The mAR ORF starts at the boxed ATG translation initiation codon in exon 1 and continues in exons 2–7 until encountering a termination codon in exon 8. The length of the ORF is 2697 nucleotides, which can encode a protein of 899 amino acids. All domains that characterize the steroid-receptor family are present in this ORF. A large *N*-terminal domain is followed by the DNA-binding domain consisting of the two Cys-Cys zinc fingers and the ligand-binding domain at the C-terminus of the protein. The *N*-terminal domain is encoded by exon 1, the information for the first and second zinc finger motif is present in the exons 2 and 3 respectively, whereas the ligand-binding domain is encoded by the exons 4–8. The 2697 nucleotides of the mAR ORF are distributed over the exons as follows: exon 1, 1553; exon 2, 152; exon 3, 117; exon 4, 288; exon 5, 145; exon 6, 131; exon 7, 158; and exon 8, 153 nucleotides respectively.

### Functional analysis of mAR expressed from the molecular cloned cDNA

A mAR cDNA containing the complete ORF was constructed using conventional cloning methods combined with the PCR techniques described in the Materials and methods section. After insertion of the mAR cDNA in a pGEM vector, the 2.8 kb insert was completely sequenced and the sequence was matched with that of the deduced mAR genomic DNA structure. The sequence showed the mAR ORF to be identical with the predicted ORF as already presented in Fig. 3, and thereby confirmed the positions of the splice sites within the mAR gene (clones containing PCR artefacts were omitted). To test the functional properties of the cloned mAR, the cDNA was inserted in an expression vector

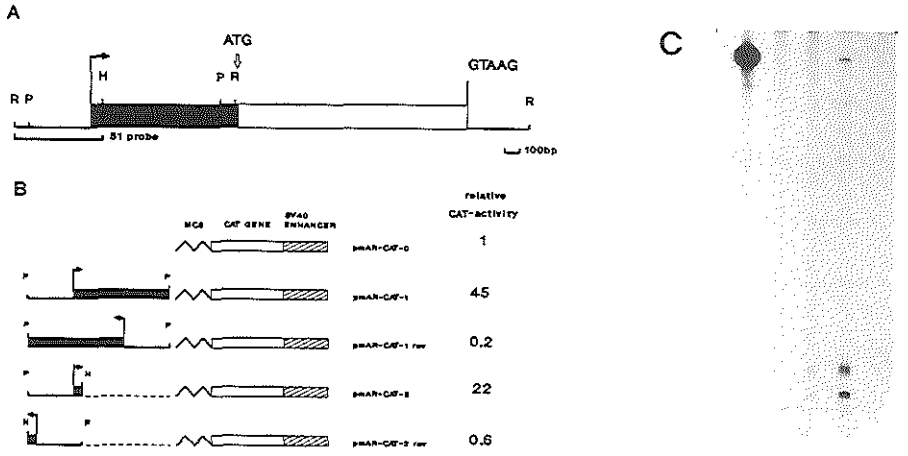


**Fig. 5. Structural and functional analysis of the promoter region of the mAR gene**

(*a*) Partial restriction map of the genomic region surrounding exon 1 of the mAR gene. Indicated are the ATG translation-initiation codon and the splice donor site GTAAG. The position of the double-stranded DNA probe *Eco*R1 *Hin*dIII that was used in the S1-nuclease-protection experiment is indicated by a horizontal bar. (*b*) Compositions of the four mAR gene promoter-CAT constructs, as well as of the plasmid pmAR-CAT-0, are shown. The four mAR promoter CAT constructs were transiently expressed in HeLa cells and subsequently cell lysates were assayed for CAT activity. Direct counting of butyrylated chloramphenicol resulted in the relative CAT activities presented. Abbreviations: H, *Hin*dIII; P, *Pst*I; R, *Eco*R1; MCS, multiple cloning site. (*c*) S1-nuclease mapping of the mAR-gene transcription start sites. The S1 probe is indicated in (*a*). Lane 1, probe; lane 2, RNA from mock-transfected COS-1 cells; lane 3, RNA from mouse prostatic tissue; lane 4, RNA from pmAR-CAT-1-transfected COS-1 cells.

(pmAR°; Fig. 4*a*). mAR was transiently expressed in COS-1 cells. To determine the size of the protein produced, Western-blot analysis was performed after immunoprecipitation of mAR from whole-cell extracts using an AR-specific monoclonal antibody. The blot was probed with an anti-AR polyclonal antibody. Controls included cell extracts prepared from mock-transfected COS-1 cells, COS-1 cells transfected with the hAR expression plasmid pAR°, cell extracts prepared from mouse testicular tissue and immunoprecipitations with a non-specific monoclonal antibody. The results of these experiments are shown in Fig. 4(*b*). A 110 kDa protein is specifically detected in pmAR°-transfected cells (compare lanes 3 and 4). Proteins of similar size are precipitated from pAR°-transfected cells (compare lanes 2 and 4) and mouse testicular tissue (compare lanes 4 and 8). The 110 kDa protein could not be precipitated from mock-transfected cells (lanes 5 and 6). These findings clearly show the pmAR° expression vector to give rise to a full-length mAR protein. Scatchard-plot analysis revealed a dissociation constant ($K_d$) of 0.12 nM for methyltrienolone (R1881) binding to the expressed protein (results not shown). To test the *trans*-activating function of the cloned mAR, co-transfections with a CAT reporter-gene construct were performed in Hela cells. In Fig. 4(*a*) the constructs used in these experiments are shown. pmAR° is the mAR expression vector. The androgen-responsive CAT reporter gene construct is pG 29 G-TK-CAT [31], which consists of two synthetic copies of an HRE (GTTACAaacTGTTCT) upstream of the TK promoter, which is linked to the CAT reporter gene. PBL2-CAT [32] is a control construct which is comparable with pG 29 G-TK-CAT, but lacks the two HREs. pSV2-CAT is a control construct used for monitoring the transfection efficiency. The autoradiograph of the t.l.c. analyses of the CAT assays is depicted in Fig. 4(*c*). The control experiments show that the synthetic androgen R1881 has no effect on CAT expression from the pG 29 G-TK-CAT reporter gene in the absence of the mAR (compare lanes 1 and 2). Similarly, in the presence of ligand, mAR has no stimulating effect on the TK promoter (compare lanes 3 and 4). In lanes 7 and 8 the activity of pSV2-CAT is illustrated, which, as expected, is not influenced by R1881. Co-transfection of pmAR° and pG 29 G-TK-CAT shows a ligand-dependent increase in CAT-activity (compare lanes 5 and 6). These experiments prove that mAR is able to *trans*-activate an androgen-responsive promoter through the appropriate response elements. Quantification of the CAT assay resulted in the calculation of an induction factor of 30 (results not shown).

### Determination of the transcriptional start sites and polyadenylation sites of the mAR gene

Hybridization studies using hAR probes spanning the transcription initiation sites of the hAR gene identified a region of sequence similarity in clone 5'mAR (Fig. 2*b*). In Fig. 5(*a*) a restriction map of the region surrounding the first exon of the mAR gene is shown. The ATG translation-initiation codon as well as the splice donor site at the exon 1/intron 1 boundary (GTAAG) are indicated. To determine whether the region upstream from the ATG codon contained promoter activity, a CAT reporter-gene construct containing SV40 enhancer sequences, but lacking transcription-initiation sites was used. Four CAT expression plasmids with mAR gene-promoter fragments were constructed (Fig. 5*b*). The original plasmid [pCAT-Enhancer (Promega)] will be referred to as 'pmAR-CAT-0'. Two

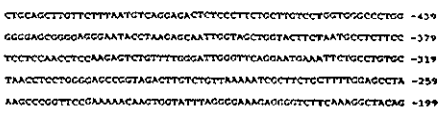

Fig. 6. Sequence analysis of the promoter region of the mAR gene

The position of the transcription-initiation sites as determined from the S1-nuclease-protection experiments are indicated by dots. Two possible control elements, the consensus Sp1 binding site and the homopurine stretch, are underlined and doubly underlined respectively.

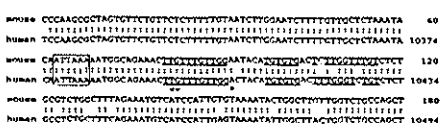

Fig. 7. Sequence analysis of the putative polyadenylation signals of mAR mRNA

The sequence as obtained from mAR genomic DNA clone 1A (Fig. 2) is compared with the sequence surrounding the well-defined polyadenylation signals of the hAR mRNA. The polyadenylation signals ATTAAA and CATAAA are boxed. The actual sites of poly(A) addition as determined in the hAR mRNA are indicated by dots, whereas the TG boxes after the polyadenylation signals are underlined in both sequences.

fragments, a 0.5 kb *Pst*I *Hin*dIII fragment and a 1.3 kb *Pst*I *Pst*I fragment were cloned in front of the CAT gene in both sense and anti-sense orientation, giving rise to the constructs pmAR-CAT-1 and pmAR-CAT-1 rev in the case of the 1.3 kb *Pst*I–*Pst*I fragment and pmAR-CAT-2 and pmAR-CAT-2 rev in the case of the 0.5 kb *Pst*I–*Hin*dIII fragment. All constructs were transfected into HeLa cells, together with the internal control plasmid pCH110, and 48 h after transfection cell extracts were assayed for CAT activity. Values obtained were normalized for $\beta$-galactosidase activity and the activity of the pmAR-CAT-0 plasmid was arbitrarily set to 1. The results of the experiments are given as relative CAT activities in Fig. 5(*b*) and show that the insertion of the two mAR gene fragments in the proper orientation produces a marked increase in CAT activity, whereas the insertion of the same fragments in the reverse orientation slightly reduces basal activity. Because pmAR-CAT-2 is active, the promoter of the mAR gene has to be located in this region. The precise positions of the transcription-initiation sites were subsequently mapped by S1-nuclease-protection experiments. RNA from mouse prostatic tissue as well as RNA from COS-1 cells transfected with the pmAR-CAT-1 promoter construct was isolated. RNA from mock-transfected COS-1 cells served as a control. The end-labelled 0.6 kb *Eco*RI–*Hin*dIII fragment (Fig. 5*a*) was used as a probe. The results are shown in Fig. 5(*c*). Specific protected fragments were absent in the control experiment (lane 2), whereas, in the case of mouse prostatic RNA (lane 3) and pmAR-CAT-1 transfected COS-1 RNA (lane 4), two sets of protected fragments were observed. The sequence of the region surrounding the transcription-initiation sites was determined (Fig. 6). The positions of the transcription-initiation sites is indicated with dots, and the most 5′ nucleotide of the two sets of transcription initiation sites is numbered +1. The promoter region of the mAR gene lacks the canonical TATA/CCAAT elements which are found at 20–30 bp and 70–100 bp upstream from the transcription start sites of many genes. Neither is the promoter of the mAR gene exceptionally (G+C)-rich, as is often observed in TATA/CCAAT minus promoters. Potential promoter elements which are present in the mAR gene promoter are a consensus Sp1 binding site at positions −36 to −45 (underlined) and a homopurine stretch at positions −55 to −125 (doubly underlined).

Because clone 1A (Fig. 2*b*) hybridized to a hAR cDNA probe located at the very 3′ end of the cDNA, additional hAR cDNA probes spanning the complete 6.8 kb 3′-UTR were hybridized to subcloned fragments from clone 1A. All probes gave rise to hybridization signals, but the probe spanning the polyadenylation signals of hAR mRNA produced the most intense signal (results not shown). These results suggested a moderate conservation for most of the 3′-UTR sequences and a more pronounced conservation for the sequences surrounding the polyadenylation sites. The fragment from clone 1A, which hybridizes to the 3′-terminus of hAR cDNA, was sequenced, and this sequence was matched with the corresponding hAR sequence (Fig. 7). The hAR DNA sequence extends from 10315 to 10625 from the transcription start site (cDNA), followed by a small stretch of 3′-flanking genomic sequences [20a]; the corresponding mouse sequence is numbered 1–359. The polyadenylation signals that were functionally identified in the hAR mRNA (ATTAAA and CATAAA) are boxed, the actual sites of poly(A) addition in the hAR mRNA are indicated by dots, and the TG boxes, usually found in the immediate downstream region of polyadenylation sites, are doubly underlined for both the human and mouse sequence.

## DISCUSSION

In the present study the structural organization of the mAR gene is described. In addition, the mAR protein is functionally characterized. The mAR gene was shown to consist of eight exons. From the genomic clones an ORF of 2697 nucleotides was deduced that encodes a 899-amino-acid protein. This sequence was verified by the isolation of mAR cDNA using conventional cloning techniques coupled with PCR amplification.

The 899-amino-acid protein contains the domains (*N*-terminal 'regulatory', DNA-binding and ligand-binding) characteristic of the receptor family. Fig. 8 shows a schematic comparison of the sequence similarity between the mAR, the hAR and the rAR [12–17]. Total (100%) identity is observed in the DNA-binding and ligand-binding domains. In the *N*-terminal 'regulatory' domain, some (minor) differences were observed, resulting in 76 and 96% similarity for the hAR and rAR in these domains respectively. The overall similarity was 85% for the hAR and 98% for the rAR. Also indicated in Fig. 8 are the homopolymeric stretches of glutamine and glycine residues. As to the position within the protein, the mAR stretches resemble those of the rAR, but the composition of the glutamine stretch is different, because it is shorter by two amino acids and intermingled with three histidine residues ($Gln_4$-1His-$Gln_3$-1His-$Gln_2$-1His-$Gln_4$).

Co-transfections of pmAR" with the androgen-responsive CAT reporter gene construct pG 29 G-TK-CAT showed the

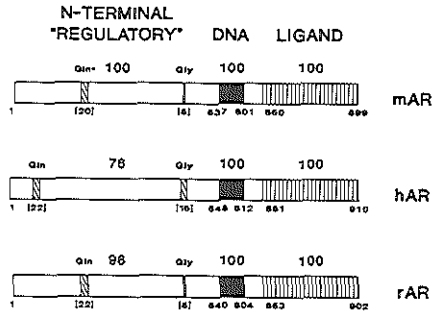


Fig. 8. Schematic comparison of the mAR, hAR and rAR

Indicated are the sequence similarities in the *N*-terminal 'regulatory' domain, the DNA-binding domain and the ligand-binding domain. Also indicated are the position and composition of the polyglutamine (Gln) and polyglycine (Gly) stretches in the *N*-terminal 'regulatory' domains. Gln* indicates the intermingling of the 17 glutamine residues with three histidine residues in the mAR polyglutamine stretch.

ability of mAR to *trans*-activate through a HRE in a ligand-dependent manner. The induction factor observed (30-fold) was comparable with that observed when the hAR expression plasmid pAR° was used in a parallel experiment (P. W. Faber, unpublished work). Similar results were obtained in experiments in which the transcriptional activating properties of mAR and hAR were compared using the complete MMTV promoter in the reporter-gene construct (P. W. Faber, unpublished work). This indicates that the polyglycine stretch, which is long in hAR (16–27 residues), but virtually absent in mAR and rAR (five residues) is not essential for AR *trans*-activating function, at least not under the experimental conditions applied. The presence of a polyglutamine stretch of similar length and composition in the three receptors, although at a different position, would argue in favour of an important role of this stretch in AR function. A link between glutamine-rich protein domains and transcriptional

Table 1. Comparison of the intron lengths of mAR and hAR genes

| Intron | Intron length (kb) mAR | hAR |
|---|---|---|
| 1 | > 14.5 | > 24.0 |
| 2 | > 19.1 | > 15.0 |
| 3 | > 20.0 | 26.0 |
| 4 | 5.2 | 5.6 |
| 5 | 4.1 | 4.8 |
| 6 | 1.0 | 0.8 |
| 7 | 0.7 | 0.7 |

activation has been made in the transcription factor SP1 [38]. Interestingly, the recently cloned human TFIID transcription factor also contains a long polyglutamine stretch [39–41]. During the preparation of this manuscript the sequence of the mAR cDNA was reported by others [42,43]. Comparison of these cDNA sequences with the mAR sequence presented here showed the three sequences to be completely identical.

The information of the mAR cDNA is separated over exons 1–8 of the mAR gene in a manner identical with that found in the hAR gene [18–20a]. When compared with the hAR gene the lengths of the individual introns are roughly conserved between the two species (Table 1). The introns 1, 2 and 3 are large (part of the information for the introns 1 and 2 is missing for both genes, whereas intron 3 is 26 kb in the human gene and at least 20 kb in the mouse gene). The lengths of the introns 4–7 could be determined for both genes and were found to be highly comparable.

A similar genomic organization with conserved intron/exon boundaries has been reported for the chicken progesterone receptor (cPR) [44,45] and human oestrogen receptor (hER) [46] genes. The positions of the intron/exon boundaries are less well conserved in genes encoding more distant members of the receptor family, such as the human vitamin $D_3$ receptor [47] and the chicken thyroid-hormone receptor [48], indicating an early divergence during evolution between the steroid receptors and other receptors of the same family. The lengths of the introns 4–7

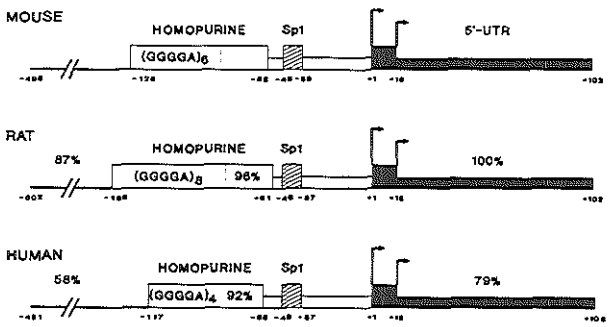


Fig. 9. Schematic comparison of the promoter regions of the mAR, rAR and hAR genes

Shown are the transcriptional start sites, the Sp1-binding site and the homopurine stretch, which was separated in a 5'- and 3'-segment, as well as part of the 5'-UTR sequences and part of the sequences upstream from the homopurine stretch. The sequence similarities are given for the 5'-UTR sequences, the 3' segment of the homopurine stretch and the sequences upstream from the homopurine stretch. In the 5'-segment of the homopurine stretch the number of GGGGA sequence motifs is given.

of the cPR and hER genes are unrelated to those observed in the mAR and hAR gene, with those of the cPR being substantially smaller and those of the hER gene being substantially larger.

Two major sites of transcription initiation were identified in a 13 bp region. These transcription-initiation sites could be defined using RNA preparations originating either from mouse prostatic tissue or from COS-1 cells transfected with mAR promoter–CAT constructs. Both promoter constructs (pmAR-CAT-1 and pmAR-CAT-2), which differed only by the length of the 5'-UTR, conferred promoter activity to a promoterless CAT reporter gene. The activity of pmAR-CAT-1, which contains an approx. 0.8 kb longer 5'-UTR than pmAR-CAT-2, was slightly higher. A similar effect has been noted using hAR promoter constructs [20a], suggesting a potentiating role of 5'-UTR sequences in transcriptional activation, an increased stability of the RNA produced or an increase in translation efficiency caused by the presence of the longer 5'-UTR.

Putative promoter elements are the consensus Sp1 binding site (*GGGGCGGGAC*, −36 to −45) and the homopurine stretch (−55 to −125). Sp1 is an ubiquitous transcription factor with DNA binding sites usually clustered in the promoter region of (G+C)-rich promoters, whereas homopurine stretches with various compositions have been identified in the promoter region of several genes [49]. These stretches are usually associated with a high sensitivity *in vitro* to S1-nuclease degradation and an unusual DNA structure [50]. Whether in the cell nucleus polypurine stretches are correlated with unfolding of the DNA is not known. Deletion of a polypurine/pyrimidine stretch in the promoter region of the epidermal-growth-factor-receptor gene resulted in a 3-fold decrease in promoter activity [51]. Analysis of this fragment by protein–DNA interaction experiments showed specific binding of two proteins, one of which turned out to be Sp1. A similar analysis of the promoter of the mouse K-*ras* gene also indicated the homopurine stretch to be involved in transcriptional activation and specific protein–DNA interactions [52].

Recently, we and others identified the promoter region of the hAR and rAR gene [20,20a,53]. A schematic comparison of the promoter regions of the three genes is given in Fig. 9. The positions of the transcription-initiation sites are identical in the three genes. The putative Sp1-binding site and the homopurine stretch are also conserved. The homopurine stretch can be divided in a 5'-fragment, consisting mainly of a multimer of the sequence motif GGGGA, and a 3'-fragment, without repeat elements. The 5'-segments of the homopurine stretch contains six, eight and four (GGGGA) blocks in the case of the mAR, rAR and hAR gene promoters respectively. More-upstream regions are moderately well conserved; the 5'-UTR shows high sequence similarity (see Fig. 9). A more detailed functional analysis will have to be conducted to evaluate the importance of the various structural elements in the AR promoter region.

The 3'-UTR of the mAR mRNA has a length of approx. 6 kb. This length is inferred from the identification of two potential polyadenylation signals in the mAR gene sequence (ATTAAA and CATAAA) at exactly the same position, where the functional polyadenylation signals of the hAR are situated [20a]. The two signal sequences differ from the canonical AATAAA hexamer sequence, but have previously been implicated in polyadenylation processes [54]. Although not strictly proven, the remarkable conservation between the hAR and mAR sequences in this region indicates that the same signals will be used for polyadenylation in the hAR and mAR transcripts. In addition, the strong sequence similarity suggests that the complete region is of high importance for polyadenylation or other, so-far-unidentified, functions of the AR gene or mRNA. The mRNA which would be produced from the complete mAR transcription unit would have a length of approx. 10 kb (1 kb [5'-UTR]+2.7 kb [ORF]+6 kb [3'-UTR]+0.2 kb [poly(A)-tail]), which closely corresponds to the reported length of the mAR mRNA [42,43,55].

In summary, we have defined the transcription unit of the mAR gene and functionally characterized the mAR protein product that originates from this gene. The information obtained provides the opportunity for a detailed analysis of regulation of expression of the mAR gene and mAR function in "*in vivo*" model systems. The sequence similarity between the mAR and hAR proteins and the promoter region of the mAR and hAR genes suggests a functional interchangeability for the individual protein domains as well as the regulatory regions of the promoters. A particularly interesting possibility would be either the correction of the naturally occurring mutation in the AR of the Tfm mouse [56] or the introduction of mutations, observed in the hAR in individuals with an aberrant male sexual development, in the mAR gene by homologous recombination techniques [57] to prove that these indeed correlate with the observed phenotypes. The information obtained from such experiments would be of great value for understanding of the role of AR in normal and abnormal male sexual development.

We are indebted to C. A. Berrevoets, J. A. G. M. van der Korput and J. Veldscholte for carrying out the Western and Southern blots and R1881 binding experiments and to Mrs. Paula Delfos for photography. This study was supported by a grant from the Dutch Cancer Society (KWF).

## REFERENCES

1. Pinsky, L. & Kaufman, M. (1987) Adv. Hum. Genet. 16, 299–472
2. Griffin, J. E. & Wilson, J. D. (1989) in The Metabolic Basis of Inherited Disease (Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D., eds.), pp. 1919–1944, McGraw-Hill, New York
3. Brown, T. R., Lubahn, D. B., Wilson, E. M., Joseph, D. R., French, F. S. & Migeon, C. J. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 8151–8155
4. Sai, T., Seino, S., Chang, C., Trifiro, M., Pinsky, L., Mhatre, A., Kaufman, M., Lambert, B., Trapman, J., Brinkmann, A. O., Rosenfield, R. L. & Liao, S. (1990) Am. J. Hum. Genet. 46, 1095–1100
5. Marcelli, M., Tilley, W. D., Wilson, C. M., Wilson, J. E., Griffin, J. E. & McPhaul, M. J. (1990) J. Clin. Invest. 85, 1522–1528
6. Marcelli, M., Tilley, W. D., Wilson, C. M., Griffin, J. E., Wilson, J. D. & McPhaul, M. J. (1990) Mol. Endocrinol. 4, 1105–1116
7. Ris-Stalpers, C., Kuiper, G. G. J. M., Faber, P. W., Schweigert, H. U., van Rooij, H. C. J., Zegers, N. D., Hodgins, M. B., Degenhart, H. J., Trapman, J. & Brinkmann, A. O. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 7866–7870
8. Green, S. & Chambon, P. (1988) Trends Genet. 4, 309–314
9. Evans, R. M. (1988) Science 240, 889–895
10. Beato, M. (1989) Cell (Cambridge, Mass.) 56, 335–344
11. O'Malley, B. (1990) Mol. Endocrinol. 4, 363–369
12. Trapman, J., Klaassen, P., Kuiper, G. G. J. M., van der Korput, J. A. G. M., Faber, P. W., van Rooij, H. C. J., Geurts van Kessel, A., Voorhorst, M. M., Mulder, E. & Brinkmann, A. O. (1988) Biochem. Biophys. Res. Commun. 153, 241–248
13. Chang, C., Kokontis, J. & Liao, S. (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 7211–7215
14. Lubahn, D. B., Joseph, D. R., Sar, M., Tan, J.-A., Higgs, H. N., Larson, R. E., French, F. S. & Wilson, E. M. (1988) Mol. Endocrinol. 2, 1265–1275
15. Tan, J., Joseph, D. R., Quarmby, V. E., Lubahn, D. B., Sar, M., French, F. S. & Wilson, E. M. (1989) Mol. Endocrinol. 2, 1276–1285
16. Faber, P. W., Kuiper, G. G. J. M., van Rooij, H. C. J., van der Korput, J. A. G. M., Brinkmann, A. O. & Trapman, J. (1989) Mol. Cell. Endocrinol. 61, 257–262
17. Tilley, W. D., Marcelli, M., Wilson, J. D. & McPhaul, M. J. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 327–331
18. Kuiper, G. G. J. M., Faber, P. W., van Rooij, H. C. J., van der Korput, J. A. G. M., Ris-Stalpers, C., Klaassen, P., Trapman, J. & Brinkmann, A. O. (1989) J. Mol. Endocrinol. 2, R1–R4

19. Lubahn, D. B., Brown, T. R., Simental, J. A., Higgs, H. N., Migeon, C. J., Wilson, E. M. & French, F. S. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 8151–8155
20. Tilley, W. D., Marcelli, M. & McPhaul, M. J. (1990) J. Biol. Chem. 265, 13776–13781
20a. Faber, P. W., van Rooij, H. C. J., van der Korput, J. A. G. M., Baarends, W. M., Brinkmann, A. O., Grootegoed, J. A. & Trapman, J. (1991) J. Biol. Chem., in the press
21. Brinkmann, A. O., Faber, P. W., van Rooij, H. C. J., Kuiper, G. G. J. M., Ris, C., Klaassen, P., van der Korput, J. A. G. M., Voorhorst, M. M., van Laar, J. H., Mulder, E. & Trapman, J. (1989) J. Steroid Biochem. 34, 307–310
22. Quarmby, V. E., Kemppainen, J. A., Sar, M., Lubahn, D. B., French, F. S. & Wilson, E. M. (1990) Mol. Endocrinol. 4, 1399–1407
23. Rundlett, S. E., Wu, X.-P. & Miesfeld, R. L. (1990) Mol. Endocrinol. 4, 708–714
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) Basic Methods in Biology, Elsevier, New York
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. (1977) Biochemistry 18, 5294–5299
28. Favoloro, J., Treisman, R. & Kamen, R. (1980) Methods Enzymol. 65, 718–749
29. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Ehrlich, H. A. (1988) Science 239, 487–491
30. Chen, C. & Okayama, H. (1987) Mol. Cell. Biol. 7, 2745–2752
31. Schule, R., Muller, M., Kaltschmidt, C. & Renkawitz, R. (1988) Science 242, 1418–1420
32. Luckow, B. & Schutz, G. (1987) Nucleic Acids Res. 15, 5490
33. Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982) Mol. Cell. Biol. 2, 1044–1052
34. Seed, B. & Sheen, J. Y. (1988) Gene 67, 271–277
35. Zegers, N. D., Claassen, E., Neelen, C., Mulder, E., van Laar, J. H., Voorhorst, M. M., Berrevoets, C. A., Brinkmann, A. O., van der Kwast, T. H., Ruizeveld de Winter, J. A., Trapman, J. & Boersma, W. J. A. (1991) Biochim. Biophys. Acta 1073, 23–32
36. Gallee, M. P. W., Visser-de Jong, E., van der Korput, J. A. G. M., van der Kwast, Th. H., ten Kate, F. J. W., Schroeder, F. H. & Trapman, J. (1990) Urol. Res. 18, 181–187
37. van Laar, J. H., Voorhorst-Ogink, M. M., Zegers, N. D., Boersma, W. J. A., Claassen, E., van der Korput, J. A. G. M., Ruizeveld-de Winter, J. A., van der Kwast, Th. H., Mulder, E., Trapman, J. & Brinkmann, A. O. (1989) Mol. Cell. Endocrinol. 67, 29–38
38. Courey, A., Holtzman, D. A., Jackson, S. P. & Tjian, R. (1989) Cell (Cambridge, Mass.) 59, 827–836
39. Peterson, M. G., Tanese, N., Pugh, B. F. & Tjian, R. (1990) Science 248, 1625–1630
40. Kao, C. C., Lieberman, P. M., Schmidt, M. C., Zhou, Q., Pei, R. & Berk, A. J. (1990) Science 248, 1646–1650
41. Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. & Roeder, R. G. (1990) Nature (London) 346, 387–390
42. He, W. W., Fischer, L. M., Sun, S., Bilhartz, D. L., Zhu, X., Young, C. Y.-F., Kelley, D. B. & Tindall, D. J. (1990) Biochem. Biophys. Res. Commun. 171, 697–704
43. Gaspar, M. L., Meo, T. & Tosi, M. (1990) Mol. Endocrinol. 4, 1600–1610
44. Huckaby, C. S., Conneely, O. M., Beattie, W. G., Dobson, A. D. W., Tsai, M.-J. & O'Malley, B. W. (1987) Proc. Natl. Acad. Sci. U.S.A. 84, 8380–8384
45. Jeltsch, J.-M., Turcotte, B., Garnier, J.-M., Lerouge, T., Krozowski, Z., Gronemeyer, H. & Chambon, P. (1990) J. Biol. Chem. 265, 3967–3974
46. Ponglikitmongol, M., Green, S. & Chambon, P. (1988) EMBO J. 7, 3385–3388
47. Zahraoui, A. & Cuny, G. (1987) Eur. J. Biochem. 166, 63–69
48. Hughes, M. R., Malloy, P. J., Kieback, D. G., Kesterson, R. A., Pike, J. W., Feldman, D. & O'Malley, B. W. (1988) Science 242, 1702–1705
49. Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M. & Wohlrab, F. (1988) FASEB J. 2, 2939–2949
50. Wells, R. D. (1988) J. Biol. Chem. 263, 1095–1098
51. Johnson, A. C., Jinno, Y. & Merlino, G. T. (1988) Mol. Cell. Biol. 8, 4174–4184
52. Hoffman, E. K., Trusko, S. P., Murphy, M. & George, D. L. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 2705–2709
53. Baarends, W. M., Themmen, A. P. N., Blok, L. J., Mackenbach, P., Brinkmann, A. O., Meijer, D., Faber, P. W., Trapman, J. & Grootegoed, J. A. (1990) Mol. Cell. Endocrinol. 74, 75–84
54. Sheets, M. D., Ogg, S. C. & Wickens, M. P. (1990) Nucleic Acids Res. 18, 5799–5805
55. Lubahn, D. B., Joseph, D. R., Sullivan, P. M., Willard, H. F., French, F. S. & Wilson, E. M. (1988) Science 240, 327–330
56. He, W. W., Young, C. Y. & Tindall, D. J. (1990) Ann. Meet. Endocrine Soc. 72nd, Atlanta, GA, abstr. 862
57. Capecchi, M. R. (1989) Science 244, 1288–1292

Received 14 January 1991/26 February 1991; accepted 4 March 1991

CHAPTER VI

---

# TWO DIFFERENT, OVERLAPPING PATHWAYS OF TRANSCRIPTION INITIATION ARE ACTIVE ON THE TATA-LESS HUMAN ANDROGEN RECEPTOR GENE PROMOTER: THE ROLE OF Sp1

FABER, P.W., van ROOIJ, H.C.J., SCHIPPER, H.J.,
BRINKMANN, A.O., & TRAPMAN, J.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
1993, in press

# TWO DIFFERENT, OVERLAPPING PATHWAYS OF TRANSCRIPTION INITIATION ARE ACTIVE ON THE TATA-LESS HUMAN ANDROGEN RECEPTOR PROMOTER: THE ROLE OF Sp1.

Peter W. Faber[1], Henri C.J. van Rooij[1], Helma J. Schipper[1], Albert O. Brinkmann[2], and Jan Trapman[1].

Departments of Pathology[1] and Endocrinology & Reproduction[2], Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

## SUMMARY

In this study the minimal promoter requirements of the TATA-less human Androgen Receptor (hAR) gene promoter are described. The hAR promoter is characterized by a short GC-box (-59/-32) and a long homopurine stretch (-117/-60). Two major transcription initiation sites, AR-transcription initiation site I (AR-TIS I, [+1/2/3]) and AR-transcription initiation site II (AR-TIS II, [+12/13]) are located in a 13 basepair (bp) region (Faber, P.W., van Rooij, H.C.J., van der Korput, J.A.G.M., Baarends, W.M., Brinkmann, A.O., Grootegoed, J.A., and Trapman, J. (1991) J. Biol. Chem. 266, 10743-10749). Transient transfection of COS cells with hAR promoter deletion and mutant constructs, followed by RNA isolation and S1-nuclease protection analysis showed that the process of transcription initiation through AR-TIS I and AR-TIS II is regulated by different promoter sequences. The GC-box directed initiation from AR-TIS II but did not affect AR-TIS I utilization, which is dependent upon sequences between positions -5 and +57. Bandshift analysis identified the transcription factor Sp1 as the protein interacting with the GC-box. A single Sp1 binding sequence was found to be present in the GC-box. Footprint analysis confirmed the interaction of Sp1 with this sequence. The differential initiation through AR-TIS I and AR-TIS II was substantiated by the introduction of point mutations in the Sp1 binding sequence: only mutations that specifically abolished Sp1 binding interfered with AR-TIS II utilization, but all mutations left AR-TIS I initiation intact.

## INTRODUCTION

Initiation of transcription is a major control point in the process of gene expression directed by RNA polymerase II. It is mediated by a set of general transcription factors (1,2), via the minimal promoter, and regulated by gene-specific factors, via more distal control elements (3,4). In higher eucaryotes, the use of alternative (tissue-specific) promoters and first exons in a transcription unit results in additional mechanisms of regulation (5). The minimal promoter contains element(s), involved in transcription start site selection. Promoters with only one or a few clustered sites of initiation contain as minimal promoter elements either a TATA-box at -25/-30 (6,7), a pyrimidine rich

initiator (Inr) element which overlaps with the site of initiation (8,9), or a combination of both. The molecular cloning of the TATA-binding component of the basal transcription factor TFIID (TBP) has helped to address the mechanism of initiation on TATA-containing promoters (10). Much less is known about TATA-less promoters, although proteins capable of interacting with Inr sequences have recently been characterized (9,11).

The human androgen receptor (hAR) is a ligand-dependent transcriptional modifier that belongs to the steroid/thyroid hormone/retinoic acid receptor family (see 12,13 for recent reviews). The hAR mediates the physiological response to testosterone and dihydrotestosterone and plays a major role in the process of male sexual development and the maintenance of male sex characteristics (14,15). Expression levels of hAR are low in most tissues but elevated in specific cells of the male urogenital system (14). Defects in hAR expression and/or function are involved in the androgen insensitivity syndrome (16,17) and may play a role in the progressive growth of prostatic tumors (18). Yet, although the identification of factors involved in regulation of hAR transcription and translation is of considerable interest, little is known about molecular mechanisms underlying hAR expression.

The hAR is encoded by a single copy gene, located on the X-chromosome. Two major mRNAs of 11 and 7.5 kb are transcribed from this gene. The 7.5 kb mRNA results from an additional splice event in the 3'-untranslated region (3'-UTR) of the hAR messenger RNA (mRNA) and does not affect the hAR open reading frame (ORF) (19). The hAR gene promoter lacks TATA/CCAAT boxes (19,20). Two transcription initiation sites, AR-TIS I (+1/2/3) and AR-TIS II (+12/13) are located in a pyrimidine rich 13 bp region. Potentially regulatory elements consist of a short GC-box (-59/-32) and a long homopurine stretch (-117/-60).

In the present study we have identified sequences involved in minimal hAR promoter activity. From deletion and mutation analysis it is concluded that two unrelated mechanisms of transcription initiation are active on overlapping sequences: initiation from AR-TIS I depends on sequences located between positions -5 and +57, whereas initiation from AR-TIS II is regulated by the single Sp1 binding sequence at -46/-37.

## MATERIALS & METHODS

### hAR promoter constructs

hAR promoter constructs were generated by standard cloning procedures (21,22), either using restriction fragments or DNA fragments obtained by the polymerase chain reaction (PCR) (23). All fragments were inserted in the multiple cloning site of the pCAT-basic and pCAT-enhancer plasmids (Promega, Madison WI, USA). Restriction fragments were isolated from Gm20 (19) and used to generate the following constructs: phAR-CAT-(E)1, SstI-AccIII (-737/+57); phAR-CAT-(E)2, HindIII-AccIII (-2.5kb/+57); phAR-CAT-(E)3, SstI-PstI (-737/+575); phAR-CAT-(E)5, BssHII-PstI (-125/+575); phAR-CAT-(E)7, DraII-PstI (-39/+575). The constructs generated with PCR fragments are: phAR-CAT-(E)6 [primer AR(-63)], phAR-CAT-(E)4 [Primer AR(-

5)] and phAR-CAT-3-Sp1-mut1/2/3 [Primers Sp1-mut1/2/3]. The sequences of the primers are given below. An SstI restriction site (underlined) was added to the primers AR(-63) and AR(-5) to facilitate cloning procedures. Primers Sp1-mut1/2/3 contain a NaeI site (underlined) which is present at position -55 in the hAR promoter. The mutation introduced in the Sp1 binding sequence (GG to TT) is indicated.

Primer AR(-63): 5'-GCAGAGCTCA$^{-63}$GGAGGCCGGCCCGGTG-3'

Primer AR(-5) : 5'-GCAGAGCTC$^{-5}$CTCCCAGCGCCCCCTCCGA-3'

Primer Sp1-mut1: 5'-AGGAGGCCGGCCCGGTttGGG-3'

Primer Sp1-mut2: 5'-AGGAGGCCGGCCCGGTGGGttCGG-3'

Primer Sp1-mut3: 5'-AGGAGGCCGGCCCGGTGGGGGCGttACCC-3'

All primers were used in PCR experiments in combination with primer G (+105/+125) (19). The resulting amplified DNA fragments were digested either with SstI and Sma I (position +21) in case of AR(-63) and AR(-5) or with NaeI and SmaI in case of Sp1-mut1/2/3. The resulting SstI-SmaI and NaeI-SmaI fragments were inserted in the SstI-SmaI and NaeI-SmaI sites of phAR-CAT-(E)3, respectively. All PCR generated fragments were completely sequenced. A typical PCR reaction used 100 pg template plasmid DNA (phAR-CAT-3), 300 ng of each primer, 0.1U Supertaq (HT Biotechnology, UK) in a 100 $\mu$l volume. Standard amplification conditions were 1' 94°C, 2' 50°C and 2' 72°C.

## Transfection-, CAT- and S1-nuclease protection assays.

One day prior to transfection, 4 x $10^5$ T47D cells were seeded in 6 cm dishes. Culture medium was Dulbecco's MEM supplemented with 5% fetal calf serum and antibiotics. Approximately 40% confluent cells were transfected by the calcium phosphate precipitation method (24) as described previously (19). Cellular extracts were prepared 48 hrs following transfection and CAT activity was measured as described (19). COS cells were transfected similarly but 48 hrs following transfection total cellular RNA was prepared by the guanidine thiocyanate method (25). S1-nuclease protection assays were performed as described (19). For S1-probes doublestranded DNA fragments were used that were generated by PCR. For RNA from phAR-CAT-(E)3/4/5/6/7 transfected COS cells and control LNCaP RNA the S1-probe was generated by the primer combination G and I (19). For RNA from phAR-CAT-(E)1 transfected COS cells primer I was used in combination with a primer within the CAT ORF, corresponding to position 2361-2385 (5'-GGCATCGTAAAGAACATTTTGAGGC-3') of the pCAT-enhancer construct (Promega) (Fig.6A). The RSV-LacZ plasmid (which contains the RSV-LTR linked to the lacZ reporter gene) was kindly provided by Dr. Meijer, Rotterdam. To identify RSV-lacZ transcripts a probe was generated using a primer within the RSV-LTR (5'-AAGCACCGTGCATGCCGATT-3') in combination with a primer in the lacZ gene (5'-AAGCCGTGGCGGTCTGGTAC-3').

## Bandshift and footprint analysis.

Oligonucleotides for bandshift analysis were end-labelled using gamma$^{32}$P-ATP and T4 polynucleotide kinase (Gibco BRL, Grand Island, NY). The Sp1 consensus oligonucleotide was purchased from Promega. Nuclear protein extracts from HeLa cells were prepared as described (26). Approximately 1x$10^4$ cpm of DNA were used per bandshift with 5-10 $\mu$g nuclear protein (27). The footprint analysis with purified Sp1 protein was performed according to a protocol supplied by the supplier (Promega). The -194/+103 NheI(blunt-ended) fragment of the hAR promoter was inserted in the sense orientation in the SmaI site of pTZ 19. The EcoRI/HindIII fragment from this clone was end-labelled on the EcoRI site, resulting in the specific labelling of the non-coding DNA strand. Approximately 1 x $10^4$ cpm of DNA were used per footprint experiment. A Maxam and Gilbert G+A sequence reaction was performed and used as a position marker.

## RESULTS

Functional activity of the hAR promoter in T47D cells

The large first exon of the hAR gene contains the information for the N-terminal domain of the receptor as well as the complete 1.1 kilobase (kb) 5'-untranslated region (5'-UTR) of the hAR mRNA (Fig.1A)(19). To measure hAR promoter activity, T47D mammary tumor cells, which endogenously express hAR mRNA, were transiently transfected with constructs, containing hAR promoter fragments linked to the chloramphenicol acetyltransferase (CAT) reporter gene (Fig.1B, phAR-CAT-1/2/3/4).

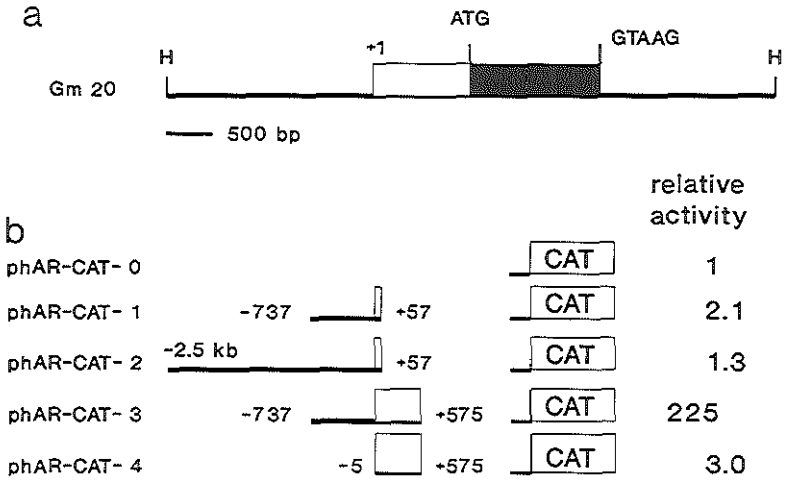


FIGURE 1: Functional activity of the hAR promoter in T47D cells. A: Overview of Gm 20, a 7.3 kb HindIII-HindIII fragment, containing the first exon of the hAR gene. The open box represents the 5'-UTR, the closed box represents the protein coding region. B: Large genomic hAR promoter fragments were linked to the CAT reporter gene and tested for promoter activity in transient transfection assays of T47D cells. The size of the fragments corresponds to the scale drawn in Fig. 1A. Activity is presented relative to the activity of the phAR-CAT-0 construct, which was arbitrarily set to 1.

Hardly any reporter gene activity was measured with phAR-CAT-1 which contains the -737/+57 promoter fragment (Fig. 1B). This low activity could not be increased by additional upstream sequences in phAR-CAT-2 (-2.5kb/+57), but was substantially increased by downstream sequences in phAR-CAT-3 (-737/+575). The activity measured with phAR-CAT-3 was dependent on sequences both upstream and downstream from the transcription start sites as a construct which lacked all upstream sequences except the transcription initiation sites, phAR-CAT-4 (-5/+575), displayed a low activity, comparable to phAR-CAT-1 and phAR-CAT-2. Summarizing, the data

indicate that the -5/+57 fragment, encompassing the two transcription start sites, contains the elements needed for minimal promoter activity. Additional segments, both upstream (untill -737) and downstream (till +575) of the start sites are able to enhance the signal above background level.

Identification of AR-TIS I and AR-TIS II transcripts in transfected COS cells

The -737/+575 hAR promoter fragment was used as a basis for subsequent experiments. First it was tested if the two transcription start sites (19) could be identified by S1-nuclease analysis in cells, transfected with phAR-CAT-3 (Fig.2A). The inclusion of the hAR 5'-UTR sequences in the construct, however, excluded the use of T47D cells because an endogenous hAR mRNA signal would also be detected. Since we were interested in the minimal hAR promoter which, by definition, would be active in all cell types, transfection experiments were done in COS cells (which do not express AR mRNA).

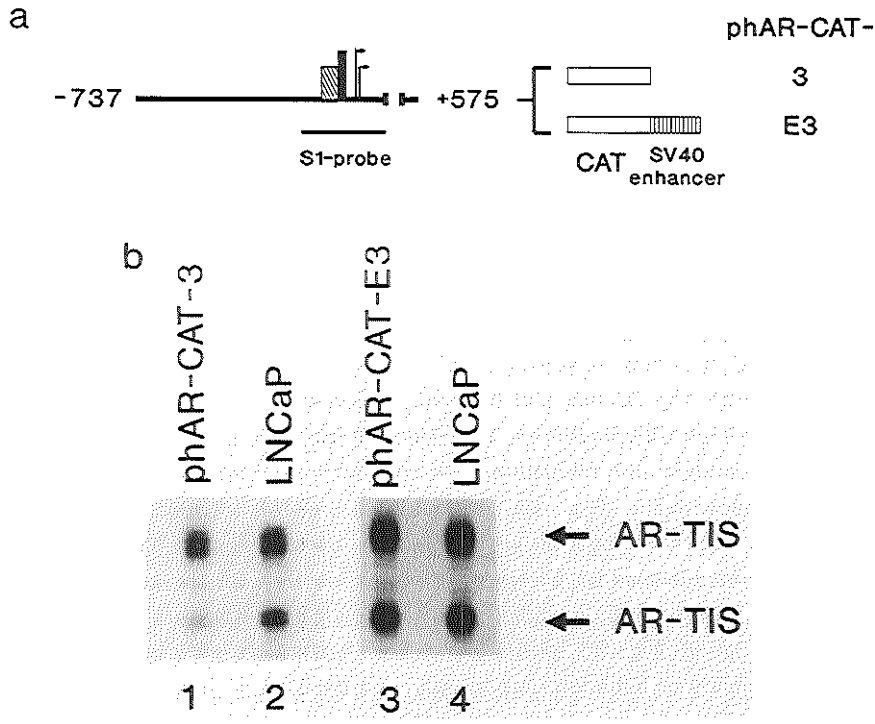


FIGURE 2: Identification of AR-TIS I and AR-TIS II transcripts in phAR-CAT-(E)3 transfected COS cells. A: Schematic representation of phAR-CAT-(E)3. Indicated are the homopurine stretch (hatched area), the GC-box (black area), the transcription initiation sites (arrows) and the probe used in the S1-assays. B: phAR-CAT-3 and phAR-CAT-E3 were transfected to COS cells, RNA was isolated and used in S1-nuclease mapping experiments. RNA from the prostatic tumor cell line LNCaP served as a control. The origin of the RNA is indicated above each lane.

Transcripts originating from both AR-TIS I and AR-TIS II could be identified by S1-nuclease protection mapping in phAR-CAT-3 transfected COS cells (Fig. 2B, lane 1). However, the relative ratio of the AR-TIS I to AR-TIS II transcripts markedly differed from the ratio in control RNA of T47D cells (data not shown) and of the prostate tumor cell line LNCaP, which expresses high levels of hAR mRNA and is preferable over T47D as a control. The AR-TIS I transcript, which is the slightly more abundant one in LNCaP RNA is the major transcript in phAR-CAT-3 transfected COS cells (Fig. 2B, compare lane 1 to lane 2). In previous studies we had used SV40 enhancer containing constructs to determine the location of the transcription initiation sites of the hAR and rat AR (rAR) promoter and had observed no difference in the relative ratio between RNA from transfected cells and control RNA (19, 28). We therefore generated and tested phAR-CAT-E3 (Fig. 2A), which is identical to phAR-CAT-3 but contains the SV40 enhancer sequence downstream from the CAT gene. This resulted in the identification of AR-TIS I and AR-TIS II transcripts in a ratio comparable to that of the control RNA (Fig. 2B, compare lane 3 to lane 4). The finding that the inclusion of the SV40 enhancer sequence in phAR-CAT-3 can restore the relative ratio of the two transcripts indicates that it is caused by a transcriptional control mechanism and does not result from a different stability of the two transcripts. As a consequence this implies that transcription initiation from AR-TIS I and AR-TIS II is regulated by distinct mechanisms and/or promoter sequences. This notion was further explored by analysis of promoter constructs containing deletion and mutant hAR promoter fragments in combination with bandshift and footprint experiments.

Different hAR promoter sequences regulate AR-TIS I and AR-TIS II utilization

In Fig. 3A the structural organization of the hAR proximal promoter, including the homopurine stretch, the GC-box and the two transcription initiation sites, AR-TIS I and AR-TIS II, is schematically presented. The effect of upstream sequences in transcription initiation was examined with reporter constructs, starting at positions -737, -125, -63, -39 and -5, respectively (Fig. 3B, phAR-CAT-3/5/6/7/4). These constructs were transfected into COS cells together with an internal control plasmid (RSV-LacZ) and the resulting RNA was analyzed by S1-nuclease protection (Fig. 3C, lanes 1-5). Sequences upstream from position -63, including the homopurine stretch, are not essential for transcription initiation (Fig. 3C, phAR-CAT-3/5/6, lanes 1-3), although they do affect the overall efficiency of initiation. Importantly, the deletion of additional sequences up to -39, which include the major part of the GC-box, abolished initiation from AR-TIS II but left initiation from AR-TIS I essentially intact (Fig. 3C, phAR-CAT-6/7, compare lanes 3 and 4). Initiation from AR-TIS I was below the detection level if sequences up to -5 were deleted (Fig. 3C, phAR-CAT-4, lane 5). To boost the signals, experiments were repeated with the comparable SV40 enhancer containing reporter constructs (Fig. 3B, phAR-CAT-E3/E5/E6/E7/E4), a selection of which is presented (Fig. 3C, lanes 6-8).

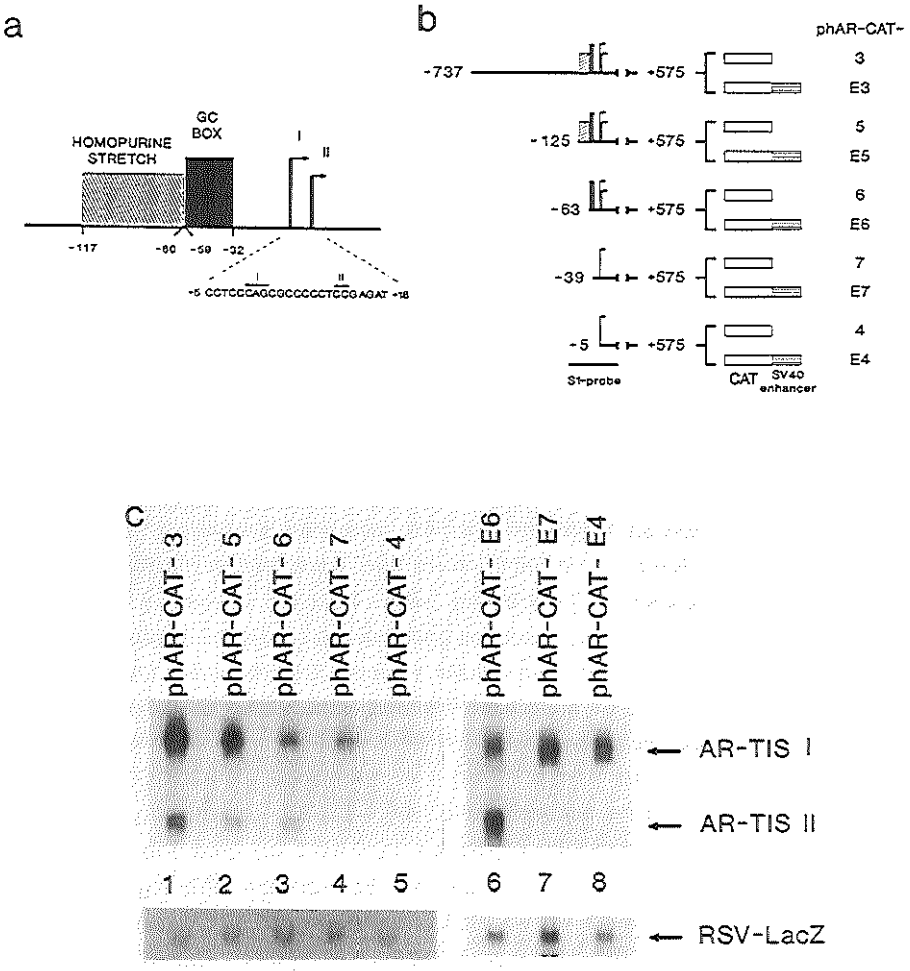


FIGURE 3: Different hAR promoter sequences gouvern AR-TIS I and AR-TIS II utilization. A: Overview of the hAR proximal promoter. Indicated are the homopurine stretch (hatched area), the GC-box (black area), AR-TIS I and AR-TIS II (arrows) as well as the sequence surrounding the two transcription initiation sites. B: hAR promoter deletion constructs, either lacking or containing the SV40 enhancer sequence, used to identify the sequences essential to transcription initiation. The 5'-startpoints of the constructs were chosen to match the boundaries of structural elements in the hAR promoter. C: phAR-CAT-3/4/5/6/7 and phAR-CAT-E4/E6/E7 were transfected together with the internal control plasmid RSV-lacZ into COS cells, RNA was isolated and used in S1-protection assays. The origin of the RNA is indicated above each lane.

With regard to AR-TIS II initiation identical results were obtained. If sequences from -63 to -39 are deleted, even in the presence of the strong SV40 enhancer, initiation from AR-TIS II no longer takes place (phAR-CAT-E6/E7, compare lanes 6 and 7). With regard to AR-TIS I initiation in the -5/+575 fragment a different result was obtained. In the presence of the SV40 enhancer initiation from AR-TIS I is clearly observed (phAR-CAT-E4, lane 8). This shows that sequences downstream from -5 have the intrinsic ability to direct initiation from the correct AR-TIS I position. Detection of AR-TIS I utilization with S1-nuclease mapping, however, depends on the presence of additional sequences which enhance the initiation frequency. Whereas the SV40 enhancer serves as this additional sequence in phAR-CAT-E4, hAR promoter sequences upstream from -5 fulfill this role in the hAR promoter constructs without the SV40 enhancer.

AR-TIS II utilization: physical evidence for Sp1 involvement

The experiments described above show that AR-TIS I and AR-TIS II utilization is governed by different sequences in the hAR promoter and identify the GC-box region as the region responsible for AR-TIS II initiation. To define proteins capable of interacting with this region, the -63/-32 fragment (Fig. 4A, overlined sequence), was used in bandshift experiments. Although in addition to the transcription factor Sp1 other factors, which are capable of interacting with GC-rich sequences have been documented, the -63/-32 hAR probe produces under several different binding conditions and with nuclear extracts from different cellular origins (COS, HeLa and LNCaP, shown is HeLa) an identical pattern of shifted bands and shows the same sensitivity to competitor oligonucleotides as a consensus Sp1 binding oligonucleotide (Promega) (Fig. 4B, compare lanes 1-5 to lanes 6-10). A sequence, 5'-GGGGCGGGAC-3', corresponding to a Sp1 binding sequence (Fig. 4A, -46/-37, underlined sequence) is present in this region (29). Specific interaction between Sp1 and this region was demonstrated by footprint analysis using purified Sp1 protein (Promega) (Fig. 4C, compare lanes 1 and 2 to lanes 3 and 4). Sp1 protects the -52/-32 region in the hAR promoter which encompasses the sequence 5'-GGGGCGGGAC-3' (-46/-37) in the hAR promoter.

a

```
        -63                          -39
         ▼                            ▼
-65 CAAGGAGGCCGGCCCGGTGGGGGCGGGACCCGACTCGCAAACT -23

                    -5
                    ▼
-22 GTTGCATTTGCTCTCCACCTCCCAGCGCCCCCTCCGAGATCCC +21
                          ***        **
                           I         II
```

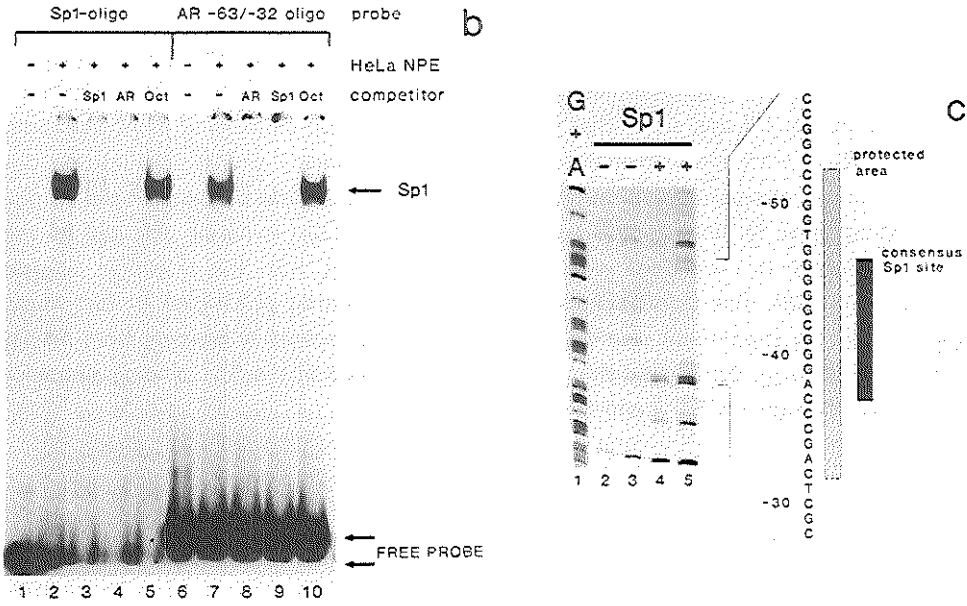


FIGURE 4: Sp1 interacts with the GC-box in the hAR promoter. A: Shown is a sequence of 86 bp (-65/+21) surrounding AR-TIS I and AR-TIS II. Endpoints of phAR-CAT-6, phAR-CAT-7 and phAR-CAT-4 are indicated with arrowheads. The positions of AR-TIS I and AR-TIS II are shown by an asterisk and indicated by I and II, respectively. The Sp1 binding sequence is underlined and the region used as a probe in the bandshift experiments is overlined. B: HeLa nuclear protein extract was used in bandshift experiments with a consensus Sp1 binding oligo (lanes 1-5) and the AR -63/-32 probe (lanes 6-10). Competions with 100 fold excess non-labelled oligo's were performed as indicated above each lane.

The upper-strand sequence of the double-stranded oligonucleotides used in the bandshift assays and for competitions are:

Sp1: 5'-ATTCGATCGGGGCGGGGCGAGC-3'

AR: 5'-AGGAGGCCGGCCCGGTGGGGGCGGGACCCGAC-3'

Oct: 5'-TGTCGAATGCAAATCACTAGAA-3'

C: Purified Sp1 protein (Promega) was used in a footprint experiment with the -194/+103 NheI hAR promoter fragment. Shown is the protection over the GC-box. A Maxam and Gilbert G+A reaction was performed and used as a size marker (lane 1). As the non-coding strand was labelled however this lane should be read as a C+T sequence. Duplicate experiments were performed with no Sp1 added to the reaction mixture (lanes 2 and 3) and with Sp1 added to the reaction mixture (lanes 4 and 5).

## AR-TIS II utilization: functional evidence for Sp1 involvement

By deletion mapping and DNA-protein interaction experiments the transcription factor Sp1 and its binding sequence at -46/-37 in the hAR promoter have been linked to initiation from AR-TIS II but not, or to a much lesser extend, from AR-TIS I. To test the functionality of the Sp1 binding sequence in the context of the -737/+575 hAR promoter fragment, a series of clustered point mutations with known effects on Sp1 binding was introduced in the Sp1 binding sequence (Fig. 5A, Sp-mut1/2/3) (27,28). One mutation (Sp1-mut2) is non-permissive to Sp1 binding, whereas two similar point

mutations (Sp1-mut1 and Sp1-mut3) are permissive to Sp1 binding to its recognition sequence (28). COS cells were transfected with the mutant constructs and the resulting RNA analyzed. phAR-CAT-3 transfected COS cells served as a control. The results are presented in Fig. 5B and show that the introduction of the permissive point mutations does not affect start site utilization (compare lane 1 to lanes 2 and 4). The introduction however of the non-permissive mutation results in the complete absence of the AR-TIS II transcript (lane 3). Importantly, initiation through AR-TIS I is hardly affected by this mutation, which is consistent with the results obtained with the promoter deletion mutants (Fig. 3B). The SV40 enhancer containing construct phAR-CAT-E1-Sp1-mut2, used in a similar experiment also did not give rise to transcripts originating from AR-TIS II (data not shown), further substantiating the importance of the Sp1 binding sequence in initiation through AR-TIS II and not AR-TIS I.

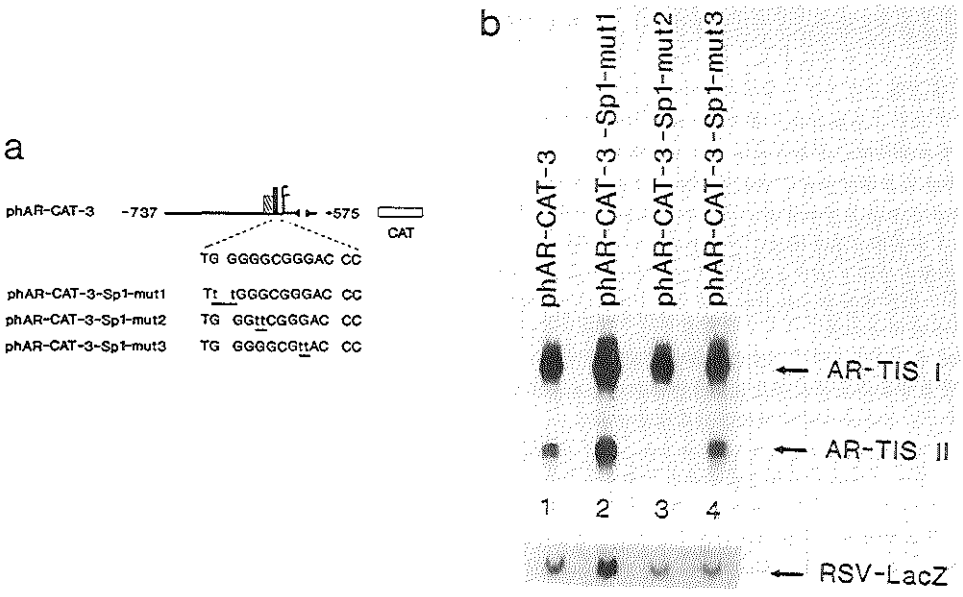


FIGURE 5: An intact Sp1 binding sequence is essential for AR-TIS II, but not for AR-TIS I, utilization. A: Schematic representation of phAR-CAT-3 with highlighted the sequence of the Sp1 binding sequence. Three mutated Sp1 binding sequences that were introduced in phAR-CAT-3, resulting in phAR-CAT-3-Sp1-mut1/2/3 are shown below. Mutations (GG to tt) are underlined. B: S1-protection mapping analysis of RNA isolated from transfected COS cells. The origin of the RNA is indicated above each lane.

The hAR 5'-UTR downstream of +57 is not involved in start site selection

The upstream boundary for AR-TIS I utilization is located at position -5 (Fig.3). A downstream boundary was established by transfection of COS cells with constructs starting at position -737 and containing progressive deletions between position +575 and +57 in SV40 enhancer containing hAR promoter constructs. Because all longer constructs used gave rise to identical results, only the S1-nuclease protection assay with RNA isolated from COS cells transfected with phAR-CAT-E1, which contains the smallest (-737/+57) promoter fragment (Fig.6A) is shown as an example (Fig. 6B, lane 1 [see also Fig. 2B, lane 3 for the -737/+575 promoter fragment]).

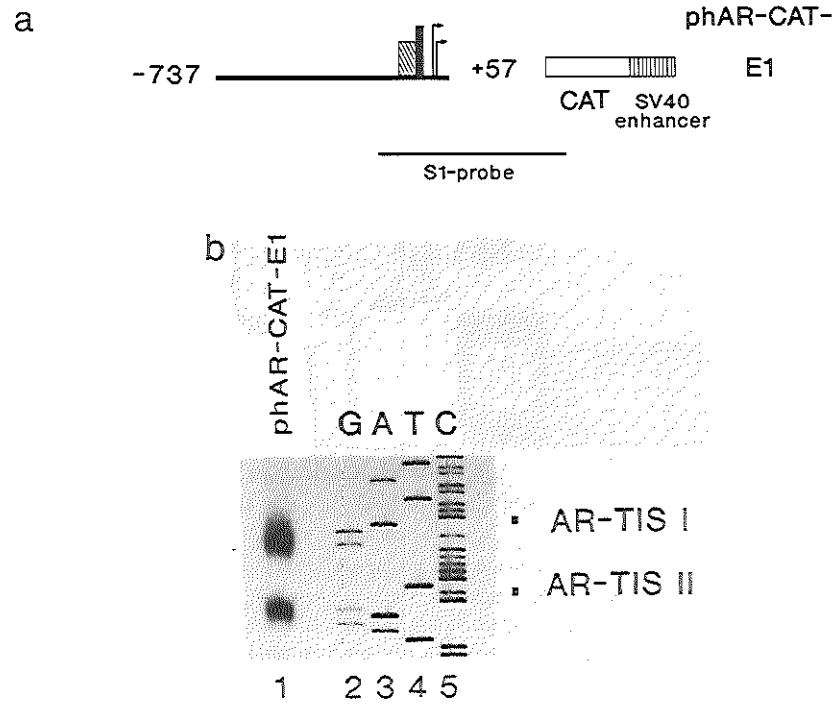


FIGURE 6: Sequences between +57 and +575 do not affect start site selection. A: Schematic representation of phAR-CAT-E1. Indicated are the homopurine stretch (hatched area), the GC-box (black area), the transcription initiation sites (arrows) and the probe used in the S1-analysis. B: S1-protection analysis of RNA isolated from phAR-CAT-E1 transfected COS cells. A sequence ladder, prepared with the same primer as the S1-probe was run alongside.

Both transcripts from AR-TIS I and AR-TIS II could be identified. This experiment, in combination with the experiments illustrated in Fig. 3 (phAR-CAT-(E)4), indicates that the minimal promoter requirements for AR-TIS I utilization are contained in the -5/+57 hAR promoter fragment.

## DISCUSSION

The present work is part of a study focussed on the identification of elements involved in the regulation of expression of hAR mRNA and concerns the characterization of the hAR minimal promoter. Using transient transfections of COS cells with hAR promoter constructs followed by S1-analysis of the resulting mRNA it was established that the utilization of the two transcription initiation sites of the hAR promoter, AR-TIS I and AR-TIS II, is governed by different promoter sequences. A single Sp1 binding sequence at -46/-37 is responsible for AR-TIS II utilization as determined by analysis of promoter deletion constructs (Fig.3) and constructs containing specific mutations in the Sp1 binding sequence (Fig.5). Sp1 specifically interacted with this sequence as determined by bandshift and footprint analysis (Fig. 4). This Sp1 binding sequence however is not involved in AR-TIS I utilization (Fig.3, Fig.5) which is dependent upon sequences located between position -5 and +57 (Fig.3, Fig.6). These results are summarized in the model presented in Fig.7.

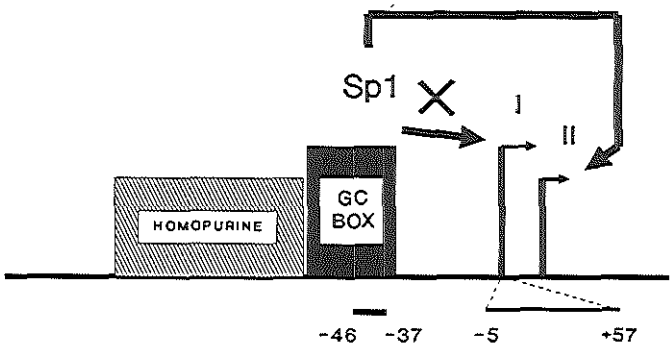


FIGURE 7: Overview of the transcription initiation events on the hAR promoter. Shown is a schematic representation of the hAR proximal promoter. Indicated are the homopurine stretch (hatched area), the GC-box (black area) and the transcription initiation sites, indicated as I and II. The Sp1 binding sequence at -46/-37 is responsible for AR-TIS II utilization but does not affect AR-TIS I selection which is dependent on sequences between position -5 and +57.

Although the differential use of overlapping promoter regions, including the use of closely linked transcription initiation sites, is well established in bacteria and yeast (31,32), the hAR gene represents the first gene described in a higher eucaryote with such a promoter organization.

The Sp1 binding sequence at -46/-37 in the hAR promoter plays a central role in transcription initiation. First because it determines AR-TIS II initiation and secondly because it does not influence AR-TIS I utilization. The finding that Sp1 is involved in transcription initiation is not unique. Although initially identified as a factor involved in specific expression of SV40 early RNA (33), binding sequences for Sp1 have been identified in the (proximal) promoters of many cellular genes (29). In model systems of

TATA-lacking promoters, Sp1 has been shown to be able to mediate transcription initiation by influencing the formation of a competent transcription initiation complex (34). However, no natural or synthetic promoter has been described where a single Sp1 binding sequence functions as a minimal promoter. Sp1 always functions with a second element, which can consist of additional Sp1 binding sequences, a TATA box or an Initiator (Inr) sequence (35,36, 37). Either the hAR promoter presents an exception to this rule or a second element is present that was not detected in our experiments but should be present in all constructs showing AR-TIS II initiation and thus is located downstream from position -63. Although no Initiator consensus is available as yet the AR-TIS II sequence (+12/13, CCCCTC$^{+12}$C$^{+13}$GAGA) bears some homology to the described Inr sequence from the terminal deoxynucleotidyl transferase (TdT) gene (CCCTCA$^{+1}$TTCT) (8), to the Inr sequence of the adenovirus major late gene (TCCTCA$^{+1}$CTCT) (11) and to the adenovirus IVa2 gene Inr sequence (GTCTCA$^{+1}$GAGT) (38) and could represent the second element. The observation that the Sp1 binding sequences in the distal SV40 enhancer cannot functionally substitute for the proximal Sp1 binding site in the SV40 enhancer containing constructs would be in agreement with previous studies on the TdT Inr were it was demonstrated that the productive interaction between an Inr and an Sp1 binding sequence is distance dependent (8). In addition, the strong stimulation of AR-TIS II initiation by the SV40 enhancer in phAR-CAT-E3 (Fig. 2b) can readily be explained by a synergism between the Sp1 binding sequence at -46/-37 and the distant Sp1 binding sequences (39). Although the SV40 enhancer stimulates AR-TIS I initiation (Fig.2b, compare phAR-CAT-4 and phAR-CAT-E4) the proximal Sp1 binding site is not involved in AR-TIS I selection. At present the reason for this remains unclear. A more detailed analysis will have to elucidate the elements involved in AR-TIS I selection and the role of Sp1.

Regarding the physiological role of the differential regulation of AR-TIS I and AR-TIS II initiation no experimental data are available as yet. Although the promoter of the AR gene has also been characterized for the mouse and rat genes (40,28) and the use of the AR promoter has been investigated in several human cell lines and rat tissues no evidence for a differential use of AR-TIS I and AR-TIS II has been presented so far (18, 41). However the AR is one of the key factors involved in the process of male sexual differentiation and development (14,15). It can be speculated that two pathways of transcription initiation that differ in their response to the action of upstream activators and one of which involves the ubiquitously expressed transcription factor Sp1 facilitates the fine-tuning of AR expression.

Acknowledgements.
This study were supported by the Dutch Organization for Scientific Research (NWO). The authors would like to thank Mr. F.L. van der Panne for skillful photography.

## REFERENCES


1. Mermelstein, F.H., Flores, O., and Reinberg, D. (1989) Biochim. Biophys. Acta 1009, 1-10
2. Saltzman, A.G., and Weinman, R. (1989) FASEB. J. 3, 1723-1733
3. Ptashne, M. (1988) Nature 335, 683-689 .
4. Mitchell, P.J., and Tjian, R. (1989) Science 245, 371-378
5. Leff, S.E, Rosenfeld, M.G., and Evans, R.M. (1986) Ann. Rev. Biochem 55, 1091-1117
6. Breathnach, R., and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383
7. Nakajima, N., Horikoshi, M., and Roeder, R.G. (1988) Mol. Cell. Biol. 8, 4028-4040
8. Smale, S.T., and Baltimore, D. (1989) Cell 57, 103-113
9. Shi, Y., Seto, E., Chang, L-S., and Shenk, T. (1991) Cell 67, 377-388
10. Greenblatt, J. (1991) Cell 66, 1067-1070.
11. Roy, A.L., Meisterernst, M., Pognonec, P., and Roeder, R.G. (1991) Nature 354, 245-248
12. Wahli, W., and Martinez, E. (1991) FASEB J. 5, 2243-2249
13. Gronemeyer, H. (1991) FASEB J. 6, 2524-2529.
14. Mooradian, A.D., Morley, J.E., and Korenman, S.G. (1987) Endocrine Rev. 8, 1-27
15. Cunha, G.R., Donjacour, A.A., Cooke, P.S., Mee, S., Bigsby, R.M., Higgins, S.J., and Sugimur, Y. (1987) Endocrine Rev. 8, 338-362
16. McPhaul, M.J., Marcelli, M., Tilley, W.D., Griffin, J.E., and Wilson, J.D. (1991) FASEB J. 5, 2910-2915
17. Brinkmann, A.O., and Trapman, J. (1992) Cancer Surveys 14, 95-111
18. Coffey, D.S., and Pinta, K.J. (1987) Prog. Clin. Biol. Res. 239, 1-73
19. Faber, P.W., van Rooij, H.C.J., van der Korput, J.A.G.M., Baarends, W.M., Brinkmann, A.O., Grootegoed, J.A., and Trapman, J. (1991) J. Biol. Chem. 266, 10743-10749
20. Tilley, W.D., Marcelli, M., and McPhaul, M.J. (1990) J. Biol. Chem. 265, 13776-13781
21. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
22. Davis, L.G., Dibner, M.D., and Battey, J.F. (1986) Basic Methods in Molecular Biology, Elsevier Publishing Co., Inc, New York
23. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Ehrlich, H.A. (1988) Science 239, 487-491
24. Chen, C., and Okayama, H. (1987) Mol. Cell Biol. 7, 2745-2752
25. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) Biochemistry 18, 5294-5299
26. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983) Nucl. Acids Res. 11, 1475-1489
27. Singh, H., Sen, R., Baltimore, D., and Sharp, P.A. (1986) Nature 319, 154-158
28. Baarends, W.M., Themmen, A.P.N., Blok, L.J., Mackenbach, P., Brinkmann, A.O., Meijer, D., Faber, P.W., Trapman, J., and Grootegoed, J.A. (1990) Mol. Cell. Endocrinol. 74, 75-84
29. Kadonaga, J.T., Jones, K.A., and Tjian, R. Trends biochem. Sci. (1986) 11, 20-23
30. Jones, K.A., Kadonaga, J.T., Luciw, P.A., and Tjian, R. Science (1986) 232, 755-759
31. Goodrich, J.A., and McClure, W.R. (1991) Trends biochem. Sci 16, 394-402
32. Pellman, D., McLaughlin, M.E., and Fink, G.R. (1990) Nature 348, 82-85
33. Dynan, W.S., and Tjian, R. (1983) Cell 32, 669-680
34. Pugh, B.F., and Tjian, R. (1991) Genes Dev. 5, 1935-1945
35. Means, A.L., and Farnham, P.J. (1990) Mol. Cell. Biol. 10, 653-661
36. Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J.W., and Clifford Azizkhan, J. (1990) Mol. Cell. Biol. 10, 6632-6641
37. Travis, A., Hagman, J., and Grosschedl, R. (1991) Mol. Cell. Biol. 11, 5756-5766
38. Kasai, Y., Chen, H., and Flint, S.J. (1992) Mol. Cell. Biol.12, 2884-2897
39. Su, W., Jackson, S., Tjian, R., and Echols, H. (1991) Genes Dev. 5, 820-826

40. Faber, P.W., King, A., van Rooij, H.C.J., Brinkmann, A.O., de Both, N.J., and Trapman, J. (1991) Biochem. J. 278, 269-278
41. Takane, K.K., Wilson, J.D., and McPhaul, M.J. (1991) Endocrinol. 129, 1093-1100

**FOOTNOTES**

Abbreviations used: hAR, human androgen receptor; AR-TIS I, androgen receptor-transcription start site I; AR-TIS II, androgen receptor-transcription start site II; bp, basepair; kb, kilobase; UTR, untranslated region; mRNA, messenger RNA; CAT, chloramphenicol acetyltransferase; Inr, initiator.

## CHAPTER VII: CONCLUDING REMARKS

### 7.1 Introduction

The gene encoding the AR is located on the X-chromosome. The hAR and mAR genes have been characterized. Both genes are composed of eight exons and seven introns. The exact size of the genes is not known as part of the intron information is lacking. Both the hAR and mAR genes, however, are large, as their size exceeds 90 and 60 kb, respectively. A major AR mRNA species of 10/11 kb is transcribed from these genes. The AR gene promoter lacks TATA/CCAAT boxes and is not particularly GC-rich. Two major sites of transcription initiation [referred to as AR-TIS I (+1/2/3) and AR-TIS II (+12/13)] are located in a 13 bp region (Chapters II to V). Analysis of the sequences involved in start site selection on the TATA-less hAR promoter identified two different, overlapping pathways of transcription initiation. AR-TIS I utilization is determined by sequences between -5/+57, whereas AR-TIS II utilization is regulated by an Sp1 binding sequence at -47/-36 (Chapter VI).

### 7.2.1 Functional activity of the hAR promoter

Transient transfection experiments of cultured mammalian cells with promoter constructs have been performed to investigate the ability of the AR promoter to induce reporter gene activity. In this way it was established that the hAR promoter is functionally active in two human cell lines, which endogenously express hAR mRNA, notably T47D mammary tumor cells and LNCaP prostatic tumor cells (Faber et al., 1993 [Chapter VI], and unpublished results). Similarly, the rAR promoter is functionally active in cultured rat Sertoli cells, which express rAR mRNA, as well as in LNCaP cells (Blok et al., 1992). In both studies large promoter fragments, which include major parts of the 1.1 and 1.0 kb 5'-UTR of the hAR and rAR mRNA, respectively, were used (-737/+575 and -435/+978, respectively). The addition of more upstream sequences (up to -2.5 and -7 kb for the hAR and rAR promoter, respectively) in the constructs had little influence on promoter activity. Further, the hAR promoter was found to be functionally active both in COS monkey kidney cells and in HeLa cervix carcinoma cells, which do not express hAR mRNA (Faber et al., 1993 [Chapter VI], and unpublished results). As it is likely that a transcriptional control mechanism underlies the difference in expression levels between the various cell-types, these results suggest that elements, which are involved in tissue-specific expression of the AR gene, are not located in the constructs used so far (also see §7.4). Alternatively, the difference is such that it cannot be identified in transient transfection assays as it depends on local chromatin structure (the homopurine stretch?). Therefore, additional studies, including the identification of novel transcriptional control regions and/or the generation of transgenic mice, carrying various hAR and/or mAR promoter fragments, linked to a lacZ reporter gene, might be necessary to clarify the in vivo cell type specific AR expression.

Functional regions of the -737/+575 hAR promoter fragment

The contribution of sequences within the -737/+575 hAR promoter fragment to the promoter activity was investigated (Faber et al., 1993 [Chapter VI], and unpublished results). Both upstream and downstream sequences were needed for high promoter activity as two deletion constructs lacking either the downstream sequences (-737/+57) or the upstream sequences (-5/+575) had low activity. However, because the -5/+575 promoter fragment only can induce initiation from AR-TIS I, it is more appropriate to compare the activity of the -63/+575 (which can induce initiation from AR-TIS I and AR-TIS II), and the -737/+575 fragments to establish the contribution of upstream sequences to promoter activity. In T47D cells the -737/+575 fragment is approximately five-fold more active than the -63/+575 fragment. In comparison, the -737/+575 fragment is over one hundred fold more active than the -737/+57 fragment, suggesting that the major contribution to AR mRNA expression is provided by sequences between +57 and +575. These downstream sequences, which will ultimately be present in the mRNA, can function at several levels. They can harbor an intra-genic enhancer to affect the transcription rate of the hAR promoter, they can affect the stability of the mRNA transcribed from the reporter gene construct or they can influence the translational competence of this mRNA. At what level the sequences between +57/+575 in the hAR promoter function remains to established. Preliminary Northern blotting experiments, however, using RNA from COS cells transfected with hAR promoter constructs, indicate that the amount of mRNA correlates with the relative level of reporter gene activity. This would exclude a major effect of the 5'-UTR sequences on translation, but doesnot discriminate between a transcriptional or a stability effect. In addition, the -737/+575 fragment contains approximately half of the 1126 nucleotide 5'-UTR of the hAR mRNA. So far, a function of sequences located between +575 and +1126, in which a small ORF (+688/+711, eight amino acids) is present, is completely unknown. However, the exeriments with the rAR promoter constructs, in which most of these sequences, including the small ORF, are present, show that the inclusion of these sequences in the constructs doesnot abolish reporter gene activity (Blok et al., 1992).

In summary, deletion analysis of the -737/+575 hAR promoter fragment has shown that downstream sequences make a major contribution to the ability of hAR promoter fragments to induce reporter gene activity. A minor contribution is made by upstream sequences.

### 7.3.1 Transcription initiation on the hAR promoter

The sequence requirements for initiation from AR-TIS I and AR-TIS II were determined using transient transfections of COS cells with hAR promoter constructs, followed by isolation of RNA from the transfected cells and S1-nuclease protection assays with probes spanning the initiation sites (Faber et al., 1993). Although COS cells do not express AR mRNA, this approach is feasible as the mechanism of basal

transcription initiation, in contrast to the process of regulated gene expression by gene-specific factors, is similar in all cell types. In comparison, the archetypal Inr of the TdT gene, a gene which is normally only expressed in precursor B and T lymphocytes, was identified this way (Smale & Baltimore, 1989).

### 7.3.2 Transcription initiation from AR-TIS I

Utilization of AR-TIS I is dependent on hAR promoter sequences, located between -5/+57, as deduced from the observation that both a -737/+57 and a -5/+575 fragment can direct initiation from AR-TIS I (Faber et al., 1993 [Chapter VI]). The -737/+21 fragment has also been used in similar experiments; it can direct initiation from AR-TIS I, narrowing down the sequences involved in AR-TIS I utilization to the -5/+21 fragment (unpublished results). These findings are in line with the observation that the AR-TIS I sequence is related to an Inr (Chapter I, Fig. 16). Recent experiments indicate that the AR-TIS I sequence can specifically interact with protein factors, present in crude nuclear extracts, as determined in bandshift experiments (unpublished results). It remains to be established if the protein factor, interacting with AR-TIS I, is similar to any other factor, previously identified as interacting with Inr sequences and whether or not this factor is actually involved in the process of transcription initiation from AR-TIS I. These findings, however, open the possibility for a thorough mutational analysis of the AR-TIS I sequence in both transcription assays in combination with DNA-protein interaction experiments.

### 7.3.3 Transcription initiation from AR-TIS II

Utilization of AR-TIS II is utterly dependent on the presence of an intact Sp1 binding sequence at -46/-37 (Faber et al., 1993 [Chapter VI]). This leads to two possibilities, either the Sp1 binding sequence is the sole determinant of AR-TIS II initiation, or, a second element is involved but the assay systems used are not sensitive enough to detect this element. As sequences upstream from the Sp1 binding sequence are not involved in transcription start site selection, this putative second element is located downstream from the Sp1 binding sequence. A potential candidate is the AR-TIS II sequence (CCCTCC$^{+1}$GAGA), which, like the AR-TIS I sequence, is related to an Inr, especially the Ad-IVa2 Inr (GTCTCA$^{+1}$GAGT) (Chapter I, Fig. 16). This Inr belongs to the TdT Inr family for which a four nucleotide sequence ($^{+3}$CTCA$^{+1}$) is the common denominator (Weis & Reinberg, 1992). A similar motif ($^{+3}$CTCC$^{+1}$) is present in the AR-TIS II sequence. Additional experiments might include mutational analysis of the AR-TIS II sequence, p.e. a C → A mutation at position +1 should strengthen the Inr potential of the AR-TIS II sequence. Also the distance between the Sp1 binding sequence and the AR-TIS II sequence can be manipulated. If AR-TIS II is an Inr, this would not affect the actual position of AR-TIS II initiation, although it could affect the transcription rate from AR-TIS II. Similar experiments with the TdT Inr showed that correct initiation took place if the proximal Sp1 binding sequence was between 30 and 66 nucleotides upstream

from the Inr (Smale & Baltimore, 1989). The Sp1/AR-TIS II distance of 49 nucleotides is well within this optimum range.

As discussed above, the AR-TIS II sequence ($CCCTC^{+1}CGAGA$) is related to, but not identical to identified Inr sequences. In in vitro reconstitution experiments RNA polymerase II, in combination with the basal transcription factors TFIID, TFIIB and TFIIF, specifically can recognize the Ad-IVa2 Inr through the $^{+3}CTCA^{+1}$ motif (Carcamo et al., 1991; Weis & Reinberg, 1992). So far, it has not been possible to demonstrate protein-DNA interactions between the AR-TIS II sequence and proteins in crude nuclear extracts in bandshift assays (unpublished results). However, whether or not the AR-TIS II sequence is similar to the Ad-IVa2 type of Inr in this respect remains to be established by additional protein-DNA binding studies using purified components of the basal transcription machinery in combination with in vitro transcription assays using wild-type and mutant AR-TIS II sequences.

### 7.4 Does the isolated human androgen receptor gene represent the complete human androgen receptor transcription unit?

The isolated hAR gene contains a single promoter region. As discussed in §1.1.2 this makes the hAR gene the "simplest" steroid hormone receptor gene as the hER, hGR and hPR genes have additional (5'-untranslated) exons and/or promoter regions. So far, no evidence for such regions in the hAR gene exists, although many cDNA libraries (prostate, epididymis, testis, T47D) have been screened with hAR probes and LNCaP hAR mRNA has been specifically analyzed by RACE-PCR (Frohman et al., 1988) with primers located in the 5'-UTR of the hAR mRNA (unpublished experiments). However, in this kind of experiments it is never certain if the appropriate cDNA libraries or the correct mRNA (with respect to the tissue-specificity or the time point during development) has been chosen. Therefore, the possibility of additional exons/promoters in the hAR gene remains.

One promoter region has been identified approximately 1 kb upstream from the ATG translation initiation codon in exon 1 and functional analysis of more upstream sequences in the hAR and rAR gene has failed to identify additional regulatory elements. Some evidence, however, for an additional enhancer sequence comes from the experiments presented in Chapter VI. The relative ratio of the AR-TIS I to AR-TIS II transcript in LNCaP and T47D RNA is approximately 1:1, but if phAR-CAT-3, containing the -737/+575 hAR promoter fragment, is transfected in COS cells, AR-TIS I initiation is favoured over AR-TIS II initiation. The use of an external SV40 enhancer restores the 1:1 ratio, suggesting that the -737/+575 fragment lacks an important segment, involved in AR-TIS II utilization. This finding would agree with the recent observation that Sp1 only functions in proximal promoter regions and is highly sensitive to the action of upstream enhancers (Seipel et al., 1992). Alternatively, the -737/ +575 fragment contains an enhancer which does not function in COS cells. At the moment, however, we favor the former possibility as preliminary deletion analysis of the -

737/+575 fragment reveals identical promoter segments to be active in LNCaP, T47D and COS cells (unpublished results). The appropriate experiment therefore would be to transiently transfect LNCaP and/or T47D cells with phAR-CAT-3 and establish if the relative ratio of the AR-TIS I to AR-TIS II transcript is different from the ratio observed with endogenous hAR mRNA. If so, this might be taken as evidence for the existence of an enhancer which is located outside the -737/+575 fragment. Subsequently, the location of this enhancer could be determined by mapping DNAse I hypersensitive regions in the hAR gene locus using nuclei from either LNCaP or T47D cells and/or testing the effect of other regions in the hAR transcription unit on hAR promoter activity in stable and transient transfection assays.

### 7.5 Final remarks

Two different, overlapping pathways of transcription initiation are active on the AR promoter. So far, the functionality of this promoter organization remains obscure as no cell lines or tissues, which endogenously express AR RNA, have been observed in which the relative ratio of the AR-TIS I to AR-TIS II transcript is different from the 1:1 ratio, observed for LNCaP RNA. As discussed in §7.2.1 and §7.4, additional transcriptional regulatory elements might be present in the hAR transcription unit. The identification of such elements and the elucidation of their effect on AR-TIS I and/or AR-TIS II initiation might shed light on this question. Alternatively, the mAR gene promoter might be used to generate transgenic "promoter knockout" mice, which are unable to initiate from either AR-TIS I or AR-TIS II. For AR-TIS II initiation, a mutation in the Sp1 binding sequence might be sufficient. The advantage of such "knockout" mice would be to split AR-TIS I from AR-TIS II initiation and vice versa. It provides an in vivo system the establish if their is a functional difference between AR-TIS I and AR-TIS II initiation at some stage and/or tissue during male development.

#### References


Blok, L.J., Themmen, A.P.N., Peters, A.H.F.M., Trapman, J., Baarends, W.M., Hoogerbrugge, J.W., & Grootegoed, J.A. (1992) Transcriptional regulation of androgen receptor gene expression in Sertoli cells and other cell types. Mol. Cell. Endocrinol. 88, 153-164.

Carcamo, J., Buckbinder, L., & Reinberg, D. (1991) The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proc. Natl. Acad Sci. USA 88, 8052-8056.

Faber, P.W., van Rooij, H.C.J., Schipper, H.J., Brinkmann, A.O., & Trapman, J. (1993) Two different, overlapping pathways of transcription initiation are active on the TATA-less human androgen receptor promoter: the role of Sp1. J. Biol. Chem., in press.

Frohman, M.A., Dush, M.K., & Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA 85, 8998-9002.

Seipel, Georgiev, O., & Schaffner, W. (1992) Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. EMBO J. 11, 4961-4968.

Smale, S.T., & Baltimore, D. (1989) The "initiator" as a transcription control element. Cell 57, 103-113.

Weis, L., & Reinberg, D. (1992) Transcription by RNA polymerase II: initiator-directed formation of transcription-competent complexes. FASEB. J. 6, 3300-3309.

## SUMMARY

In this study the androgen receptor (AR) transcription unit is presented. Chapter II describes the isolation and characterization of one genomic clone, from which the amino acid sequence of the N-terminal domain of the hAR was deduced. This amino acid sequence was characterized by the presence of several homopolymeric stretches, of which a long Q-stretch (20 residues) and a long G-stretch (16 residues) are most conspicuous. In combination with a previously described cDNA clone, which encoded the hAR DBD and LBD, an open reading frame of 2730 bp was deduced, encoding a protein of 910 amino acids. Next, the complete coding region of the hAR gene was isolated from a genomic library, as described in Chapter III. The information for the hAR was found to be separated over eight exons. The total lenght of the single copy hAR gene, which is located on the X-chromosome, exceeds 90 kb. As described in Chapter II, the N-terminal domain is present in one single exon. The DBD is encoded in the exons 2 and 3, with each exon containing the information for one "zinc-finger". The LBD is encoded by the exons 4 to 8. Interestingly, the positions of the introns were found to be conserved between the hAR gene and the cPR and hER genes. The experiments described in Chapter IV extend the work of the Chapters II and III and deal with the characterization of the complete hAR cDNA and gene. Northern blot analysis of hAR mRNA identified two hAR mRNA species of 11 and 8 kb, respectively, in the human prostatic tumor cell line LNCaP. A full-length hAR cDNA was constructed from cDNA and genomic clones. Structurally, the 11 kb cDNA consists of a long 5'-UTR (1.1 kb), a 2.7 kb ORF, and a very long 3'-UTR (6.8 kb). The complete 5'-UTR and 3'-UTR were found to be encoded in the exons 1 and 8 of the hAR gene, respectively, fixing the number of exons in this gene at 8. The promoter of the hAR gene, located 1.1 kb upstream from the ATG translation initiation codon in exon 1, was structurally and functionally characterized. Two major sites of transcription initiation in a 13 bp region were identified by S1-nuclease protection experiments. DNA fragments, spanning these sites of initiation, conferred promoter activity upon a promoterless reporter gene construct. S1-nuclease protection experiments with RNA from COS cells, transiently transfected with these constructs, showed usage of the correct initiation sites. Structurally, the promoter lacks TATA/CCAAT boxes and potential regulatory elements consist of a short GC-box (-59/-32), which includes an Sp1 binding sequence (-46/-37), and a long homopurine stretch (-60/-117). The 3'-UTR contains two equally effective polyadenylation signals at a mutual distance of 221 bp. In addition, it was shown that the 8 kb hAR mRNA results from an alternative splice in the 3'-UTR, which does not affect the hAR ORF.

In order to extend findings regarding the AR to different species, it is necessary to isolate species-specific tools. The experiments, described in Chapter V, deal with the characterization of the mouse AR protein and gene. Using hAR cDNA probes, the mAR gene was isolated from a genomic library and found to be similarly organized as the hAR gene. Sequence analysis of the exons resulted in the prediction of a 2697 bp ORF,

which can encode a 899 amino acid protein. The structure of this ORF was confirmed by sequence analysis of mAR cDNA fragments, which were obtained by RT-PCR with mouse testis cDNA and mAR specific primers. An eukaryotic mAR expression vector was constructed and the mAR was transiently expressed in COS cells. The expressed protein was similar in ligand-binding affinity and size to the native mAR. In cotransfection experiments in HeLa cells with an androgen-responsive gene construct, the mAR was able to transactivate this promoter in a ligand-dependent manner. The transcription initiation sites of the mAR gene were identified by S1-nuclease protection experiments with mouse prostatic RNA and RNA from COS cells, transiently transfected with mAR promoter constructs. The mAR gene promoter was found to be similar to the hAR gene promoter. Two major sites of transcription initiation are located in a 13 bp region, no TATA/CCAAT boxes are present, and potentially regulatory elements consist of a short GC-box with an Sp1 binding sequence and a long homopurine stretch. In the 3'-UTR, the polyadenylation sites were identified through sequence homo- logy with the corresponding hAR signals. The size of the mAR cDNA, identitified this way is in accordance with the reported lenght of 10 kb of the mAR mRNA.

Although the hAR and mAR gene promoter lack a TATA-box, transcription initiation takes place at two well defined positions. The experiments of Chapter VI deal with the regulation of transcription initiation on the hAR gene promoter. Transient transfections of COS cells with reporter gene constructs, containing large fragments spanning the hAR promoter, followed by RNA isolation and S1-nuclease protection assays, showed correct usage of the two initiation sites of the hAR promoter, AR-TIS I and AR-TIS II. hAR promoter deletion and point mutant constructs were generated and tested in a similar manner. Surprisingly, transcription initiation through AR-TIS I and AR-TIS II is regulated by different promoter sequences. The GC-box, and more precisely the Sp1 binding sequence (-46/-37), is essential for AR-TIS II utilization, but hardly affects AR-TIS I utilization. Sequences, essential for AR-TIS I utilization, were found to be located between position -5 and +57, but the protein factors involved remain to characterized.

## SAMENVATTING

In deze studie wordt de androgeenreceptor (AR) transcriptie-eenheid gepresenteerd. Hoofdstuk II beschrijft de isolatie en karakterisering van een genomische kloon, waaruit de aminozuurvolgorde van het N-terminale domein van de hAR afgeleid werd. Deze aminozuurvolgorde wordt gekarakteriseerd door de aanwezigheid van een aantal homopolymere reeksen, waarvan een lange Q-reeks (20 residuen) en een lange G-reeks (16 residuen) het meest in het oog springen. In combinatie met een eerder beschreven cDNA kloon, die kodeerde voor het DNA-bindende en het hormoon-bindende gebied van de hAR, werd een open leesraam van 2730 baseparen (bp) afgeleid, dat kodeert voor een eiwit van 910 aminozuren. Vervolgens werd het komplete koderende gebied van het hARgen geisoleerd uit een genomische DNA bank, zoals beschreven in Hoofdstuk III. De informatie voor de hAR is verdeeld over 8 exonen. De totale lengte van het hARgen, dat gelegen is op het X-chromosoom, is groter dan 90 kilobasen (kb). Zoals beschreven in Hoofdstuk II, is de informatie voor het N-terminale gebied aanwezig in een enkel exon. Het DNA-bindende gebied wordt gekodeerd door de exonen 2 en 3, waarbij ieder exon de informatie voor èèn "zink-vinger" bevat. Het hormoon-bindende gebied wordt gekodeerd door de exonen 4 tot 8. Een interessante bevinding was, dat de posities van de intronen gekonserveerd waren tussen het hARgen en zowel het kippe progesteron-receptorgen als het humane oestrogeenreceptorgen. De experimenten die worden beschreven in Hoofdstuk IV verbreden het werk uit de Hoofdstukken II en III en behandelen de karakterisering van het komplete hAR cDNA en gen. Via Northern blot analyse van hAR mRNA, geisoleerd uit de humane prostaat tumor cellijn LNCaP, werden twee hAR mRNAs van respectievelijk 11 en 8 kb geïdentificeerd. Het komplete hAR cDNA werd geconstrueerd uit cDNA en genomische fragmenten. Dit 11 kb cDNA bestaat structureel uit een lang 5'-niet-vertaald gebied (1.1 kb), een open leesframe van 2.7 kb en een erg lang 3'-niet-vertaald gebied (6.8 kb). Het komplete 5'-niet-vertaalde gebied en 3'-niet-vertaalde gebied worden gekodeerd door respectievelijk exon 1 en 8 van het hARgen, hetgeen het aantal exonen in dit gen op 8 bepaalt. De promotor van het hARgen, die 1.1 kb voor het ATG translatie start codon in exon 1 gelegen is, werd structureel en functioneel gekarakteriseerd. S1-nuclease experimenten toonden het voorkomen van twee preferentiële initiatieplaatsen in een gebied van 13 bp aan. DNA fragmenten, gelegen rond deze initiatieplaatsen, waren in staat activiteit te verlenen aan een promotorloos reportergen construct. S1-nuclease experimenten, die werden uitgevoerd met RNA van COS cellen, die kortdurend getransfecteerd waren met deze constructen, toonden aan dat de correcte initiatieplaatsen gebruikt werden. Structureel gezien mist de promotor TATA/CCAAT boxen en potentiële regulerende elementen bestaan uit een korte GC-box (-59/-32), die een Sp1 bindingsplaats (-46/-37) bevat, en een lange homopurine reeks (-60/-117). Aan het einde van het 3'-niet-vertaalde gebied werd de aanwezigheid van twee in gelijke mate werkzame polyadenylerings- signalen, die 221 bp van elkaar lagen, aangetoond. Bovendien werd gedemonstreerd dat het 8 kb

hAR mRNA wordt gegenereerd door een alternatieve "splice" in het 3'-niet-vertaalde gebied, die echter geen invloed heeft op het open leesraam van de hAR.

Om bevindingen met betrekking tot de AR te kunnen extrapoleren naar verschillende species is het nodig om species-specifieke probes te isoleren. De experimenten, zoals beschreven in Hoofdstuk V, behandelen de karakterisering van het muis AR (mAR) eiwit en gen. Het mARgen, dat geisoleerd werd uit een genomische DNA bank, is net zoals het hARgen georganiseerd. Sequentie-analyse van de exonen resulteerde in het voorspellen van een 2697 bp mAR open leesraam, dat een eiwit van 899 aminozuren kan koderen. De structuur van dit open leesraam werd bevestigd door sequentie-analyse van mAR cDNA fragmenten, die met behulp van RT-PCR met mAR specifieke primers gegenereerd werden uit muis testis RNA. Een eukaryote expressievector voor de mAR werd geconstrueerd en de mAR werd kortdurend tot expressie gebracht in COS cellen. Het tot expressie gebrachte eiwit had een identieke affiniteit voor hormonen en grootte als de natieve mAR. De mAR was in staat om in cotransfectie experimenten met een androgeen-gereguleerd reportergenconstruct deze promotor hormoon-afhankelijk te activeren. De transcriptie initiatieplaatsen van het mAR gen werden geidentificeerd via S1-nuclease experimenten met muis prostaat RNA en RNA, dat geisoleerd was uit COS cellen, die met mAR promoter constructen getransfecteerd waren. De promotor van het mARgen is identiek aan de promotor van het hARgen. Twee preferentiële initiatieplaatsen zijn gelegen in een gebied van 13 bp, er zijn geen TATA/CCAAT boxen en potentiële regulerende elementen bestaan uit een korte GC-box met een Sp1 bindingsplaats en een lange homopurine reeks. De polyadenyleringsignalen in het 3'-niet-vertaalde gebied werden geidentificeerd via sequentiehomologie met de korresponderende hAR signalen. De grootte van het mAR cDNA dat afgeleid kan worden uit deze experimenten is 10 kb, hetgeen overeenkomt met de gerapporteerde lengte van het mAR mRNA.

Alhoewel de promotoren van de hAR en mARgenen geen TATA/CCAAT boxen bevatten, treedt er transcriptie-initiatie op via twee goed gedefinieerde posities. De experimenten uit Hoofdstuk VI behandelen de regulering van transcriptie-initiatie op de promotor van het hARgen. Wanneer COS cellen worden getransfecteerd met reportergen constructen, die grote fragmenten rond de hAR promoter bevatten, kan na isolatie van RNA, via S1-nuclease experimenten worden aangetoond dat de correcte initiatieplaatsen, AR-TIS I en AR-TIS II, worden gebruikt. In deze studie zijn constructen met hAR promotor deletiefragmenten en fragmenten met puntmutaties gekonstrueerd en op dezelfde manier getest. Een interessante bevinding was dat transcriptie-initiatie via AR-TIS I en AR-TIS II door verschillende promotorsequenties gereguleerd worden. De GC-box, en precies gezegd de Sp1 bindingsplaats (-46/-37), is essentieel voor het gebruik van AR-TIS II, maar heeft weinig invloed op het gebruik van AR-TIS I. De sequenties, die verantwoordelijk zijn voor het gebruik van AR-TIS I, liggen tussen positie -5 en +57, maar de eiwitfactoren die hierbij betrokken zijn moeten nog gekarakteriseerd worden.

## LIST OF PUBLICATIONS

J.Trapman, P. Klaassen, G.G.J.M. Kuiper, J.A.G.M. van der Korput, P.W. Faber, H.C.J. van Rooij, A. Geurts van Kessel, M.M. Voorhorst, E. Mulder and A.O. Brinkmann. (1988). Cloning, structure and expression of a cDNA encoding the human androgen receptor. Biochem. Biophys. Res. Comm. 153, 241-248.

P.W. Faber, G.G.J.M. Kuiper, H.C.J. van Rooij, J.A.G.M. van der Korput, A.O. Brinkmann and J. Trapman. (1989) The N-terminal domain of the human androgen receptor is encoded by one large exon. Mol. Cell. Endocrinol. 61, 257-262.

P.W. Faber, G.G.J.M. Kuiper, H.C.J. van Rooij, J.A.G.M. van der Korput, A.O. Brinkmann and J. Trapman. (1989) Characterisation of the human androgen receptor. In: The steroid/thyroid hormone receptor family and gene regulation. (J. Carlstedt-Duke, H. Eriksson and J-A. Gustafsson eds) Birkhauser Verlag, Basel, pp169-182.

A.O. Brinkmann, P.Klaassen, G.G.J.M. Kuiper, J.A.G.M. van der Korput, J. Bolt, W. de Boer, A. Smit, P.W. Faber, H.C.J. van Rooij, A. Geurts van Kessel, M.M. Voorhorst, E. Mulder and J. Trapman. (1989) Structure and function of the androgen receptor. Urol. Res. 17, 87-89.

G.G.J.M. Kuiper, P.W. Faber, H.C.J. van Rooij, J.A.G.M. van der Korput, C. Ris-Stalpers, P. Klaassen, J. Trapman and A.O. Brinkmann. (1989) Structural organisation of the human androgen receptor gene. J. Mol. Endocrinol. 2, R1-R4.

A.O Brinkmann, P.W. Faber, H.C.J. van Rooij, G.G.J.M. Kuiper, C. Ris, P. Klaassen, J.A.G.M. van der Korput, M.M. Voorhorst, J.H. van Laar, E. Mulder and J. Trapman. (1990) The human androgen receptor: domain structure, genomic organisation and regulation of expression. J. Steroid Biochem. 34, 307-310.

J.Trapman, C. Ris-Stalpers, J.A.G.M. van der Korput, G.G.J.M. Kuiper, P.W. Faber, J.C. Romeijn, E. Mulder and A.O. Brinkmann. (1990) The androgen receptor: Functional structure and expression in transplanted human prostate tumors and prostate tumor cell lines. J. Steroid Biochem. Mol. Biol. 37, 837-842.

C. Ris-Stalpers, G.G.J.M. Kuiper, P.W.Faber, H.U. Schweikert, H.C.J. van Rooij, M.D. Hodgins, H.J. Degenhart, J. Trapman and A.O. Brinkmann. (1990) Aberrant splicing of androgen receptor messenger RNA in a patient with androgen insensitivity. Proc. Natl. Acad. Sci. USA. 87, 7866-7870.

W.M. Baarends, A.P.N. Themmen, L.J. Blok, P. Mackenbach, A.O. Brinkmann, D. Meyer, P.W. Faber, J. Trapman and J.A. Grootegoed. (1990) The rat androgen receptor promoter. Mol. Cell. Endocrinol. 74, 75-84.

P.W. Faber, H.C.J. van Rooij, J.A.G.M. van der Korput, W.M. Baarends, A.O. Brinkmann, J.A. Grootegoed and J. Trapman. (1991) Characterization of the human androgen receptor transcription unit. J. Biol. Chem. 266, 10743-10749.

P.W. Faber, A. King, H.C.J. van Rooij, A.O. Brinkmann, N.J. de Both and J. Trapman. (1991) The mouse androgen receptor: Functional analysis of the protein and characterization of the gene. Biochem. J. 278, 269-278.

J.A Ruizeveld de Winter, W.M. van Weerden, P.W. Faber, G.J. van Steenbrugge, J. Trapman, A.O. Brinkmann, Th. H. and van der Kwast. (1992) Regulation of androgen receptor expression in the human heterotransplantable prostate carcinoma PC-82. Endocrinol. 131, 3045-3050.

P.W. Faber, H.C.J. van Rooij, H.J. Schipper, A.O. Brinkmann, and J. Trapman. (1993) Two different, overlapping pathways of transcription initiation are active on the TATA-less human androgen receptor promoter: the role of Sp1. J. Biol. Chem. in press.

## CURRICULUM VITAE

| | |
|---|---|
| 22 maart 1962 | Geboren in de Noordoostelijke Polder. |
| 1980 | Eindexamen Gymnasium-$\beta$ aan het Prof. ter Veen Lyceum te Emmeloord. |
| 1980-1986 | Rijksuniversiteit Groningen,<br>Faculteit Wiskunde en Natuurwetenshappen<br>Hoofdvak: Subfaculteit Scheikunde,<br>Afdeling Biochemie,<br>Dr. G. AB, Prof. Dr. M. Gruber.<br>Bijvak: Faculteit Geneeskunde,<br>Afdeling Fysiologische Chemie,<br>Dr. D. Hoekstra. |
| 1987-1992 | Promotie onderzoek op de afdeling Pathologie van de Erasmus Universiteit Rotterdam.<br>Promotor: Prof. Dr. F.T. Bosman.<br>Co-Promotor: Dr. Ir. J. Trapman. |
| 1992-heden | Onderzoeksmedewerker op de afdeling Pathologie van de Erasmus Universiteit Rotterdam. |

## NAWOORD

De totstandkoming van dit proefschrift is niet alleen mijn verdienste maar is tevens te danken aan de steun en toewijding van anderen in de voorbije jaren. Allen die met mij hebben meegewerkt dank ik daarom hierbij terwijl ik sommigen met naam wil noemen.

In de eerste plaats wil ik mijn ouders en mijn vrouw Saskia bedanken voor de steun en het vertrouwen die er altijd is en vooral Saskia voor het moeiteloos accepteren van de werktijden die dit vak soms met zich meebrengt.

Mijn co-promoter, Jan Trapman, wil ik bedanken voor de steun, het vertrouwen en vooral de plezierige samenwerking in de afgelopen jaren. Mijn promoter, prof. Bosman wil bedanken voor zijn bijdrage gedurende de laatste fase van het promotie-onderzoek en het vlot doorwerken van het uiteindelijke manuscript. Ook de overige leden van de promotie-commissie, prof. Bootsma, prof. Grootegoed, en prof. van der Vliet wil ik bedanken voor het zeer vlot doornemen van het proefschrift.

Naast mensen die vooral achter de schermen hun werk hebben gedaan heb ik ook bij het praktische werk veel steun mogen ondervinden waarbij de bijdrage van eèn persoon een speciale vermelding verdiend. Henri, bedankt voor de enorme bijdrage aan het verrichtte werk en vooral de plezierige en stimulerende samenwerking gedurende pakweg de afgelopen vijf jaar. Verder verdienen alle anderen die aan het praktische werk bijgedragen hebben, Paul, George, Hetty, Audrey en gedurende de afgelopen 1,5 jaar Helma een pluim op de hoed.

Niet alleen medewerkers van de afdeling pathologie hebben hun bijdrage geleverd maar ook medewerkers van andere afdelingen, met name Endocrinologie & Voortplanting, Genetica en Urologie hebben hun steentje bijgedragen. In dit kader verdient voornamelijk de al jaren durende positieve samenwerking met de groep van Albert Brinkmann een vermelding.

Het werk is gefinancierd in de afgelopen jaren door achtereenvolgens de EUR, het KWF en het NWO, hiervoor mijn dank.