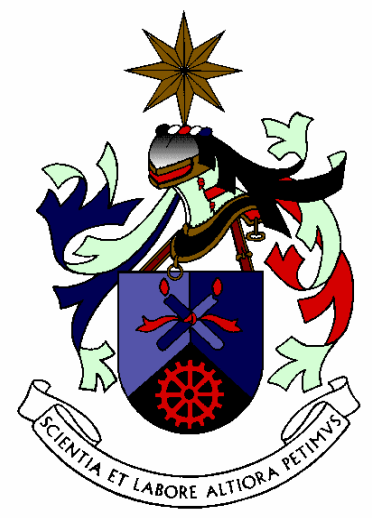
University of Beira Interior



Androgen receptor alternative spliced transcripts: tissue expression and evolutionary analysis

Master thesis in Biochemistry

Joana Filipa Melfe Tomás

Covilhã, September of 2008

Androgen receptor alternative spliced transcripts: tissue expression and evolutionary analysis

Transcritos alternativos do receptor de androgénio: expressão em tecidos e análise evolutiva.

Master thesis in Biochemistry

Joana Filipa Melfe Tomás

Supervisor: Prof. Dr. Sílvia Socorro (University of Beira Interior)

Co-Supervisor: Prof. Dr. J.E. Cavaco (University of Beira Interior)

The practical part of this work was conducted in the Health Sciences Research Centre of the Medical Sciences Faculty from the University of Beira Interior.

Centro de Investigação em Ciências da Saúde
Health Sciences Research Centre

Faculdade de Ciências da Saúde

All contents in this thesis are of exclusive responsibility of the author.

(Joana Filipa Melfe Tomás)

Contents

Acknowledgements

Um grande obrigado aos meus pais por todo o apoio prestado e que me proporcionaram a oportunidade de realizar mais este projecto académico.

Gostaria ainda de agradecer à Sandra por tudo o que me ensinou e pela disponibilidade que sempre mostrou para me ajudar.

Aos meus orientadores Prof. Doutora Sílvia Socorro e Prof. Doutor J.E. Cavaco um especial agradecimento pela oportunidade que me deram em realizar a minha dissertação de mestrado e por toda a ajuda e orientação prestada.

A todos os colegas de laboratório um especial obrigado por terem transformado estes meses numa experiência magnífica tanto a nível profissional como pessoal.

Por último agradeço a todos aqueles que de alguma forma me apoiaram, ao CICS e à Faculdade de Ciências da Saúde da Universidade da Beira Interior.

My bigger acknowledgement go to my parents without whom this project would have been impossible to reach, thanks once more for all the support you gave me all these years.

I would like to say thanks to Sandra for all the time she lose with me teaching me and for all the availability she had.

To my supervisors a special acknowledgement for giving me the opportunity to work with them and for all the support given during this months. I have still to tank to all my lab colleges for the sociability and for the mutual support.

In last I thanks to everyone that help me at any ocasion, to CICS and to Medical Sciences Faculty from University of Beira Interior.

Abstract

The vast majority of human genes express multiple mRNAs by alternative splicing of their pre-mRNAs. This mechanism allows the expression of many different mRNAs from the same gene, and greatly increases the complexity of “cell-specific” protein function. It can alter the function of proteins by removing or adding specific domains (nuclear localization signals, transcription activation domains, DNA or RNA binding domains, trans-membrane domains), post-translation modification sites, or by causing substantial changes in protein structure. The existence of several variant RNA transcripts has been described for many steroid receptors. However, the information on alternative spliced mRNAs of androgen receptor (AR) gene is scarce. A previous work in our research group (Laurentino *et al.* 2006, results not published) allowed the identification for the first time of AR alternative transcripts in human testis. Two transcripts result from exon 2 (AR∆2) and exon 3 (AR∆3) skipping. Transcripts lacking exon 2 but retaining part of intron 2 of the AR gene (AR55), and lacking part of exon 4 (AR94) were also detected. The aims of the present study are: 1) characterize the distribution of AR transcripts in different human tissues (heart, kidney, liver, and lung); 2) perform an evolutionary analysis by characterizing the expression of AR transcripts in testis of chicken (*Gallus gallus*), dog (*Cannis lupus*), rat (*Rattus norvegicus*) and sea bream (*Sparus aurata*). For this purpose the following experimental approach was established: 1) exon spanning primers suitable for the detection of alternative spliced ARs were designed; 2) Total RNA was extracted and used for cDNA synthesis; 3) RT-PCR was carried out and the amplified products were cloned and sequenced; 4) the obtained sequences were analysed using bioinformatics tools. In human heart, kidney, liver and lung was observed the presence of AR∆2 and AR∆3 transcripts previously identified in testis. In seabream, rat and dog testis were detected several AR transcripts some of which are homologous to those identified in human testis.

Resumo

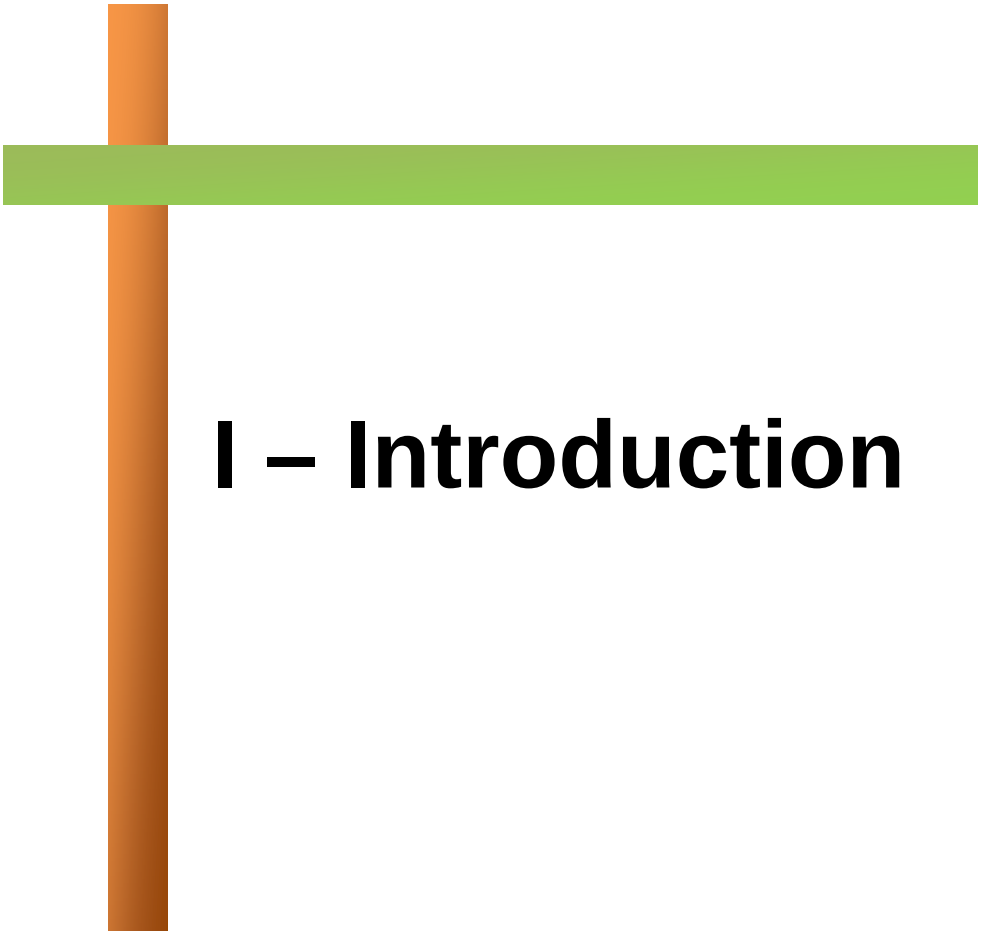
A grande maioria dos genes humanos expressa múltiplos mRNAs devido aos seus pré-mRNA sofrerem splice alternativo. Este mecanismo permite a expressão de diferentes mRNAs para um mesmo gene, aumentando significativamente a complexidade da função proteica. Ele pode alterar a função das proteínas por remover ou adicionar domínios específicos (sinal de localização nuclear, domínios de activação da transcrição, domínios de ligação do DNA ou RNA, domínios transmenbranares), por modificar os sítios de pós-translação, ou por causar mudanças significativas na estrutura da proteína. A existência de diversas variantes de transcritos de RNA tem vindo a ser descrita para vários receptores nucleares. No entanto, a informação sobre transcritos alternativos do receptor de androgénio (AR) é escassa. Um trabalho anterior do nosso grupo de investigação (Laurentino *et al.* 2006, resultados não publicados) permitiu pela primeira vez a identificação de transcritos alternativos do AR em testículos humanos. Transcritos com delecção do exão 2 (AR∆2) e do exão 3 (AR∆3). Transcritos sem o exão 2 mas com parte do intrão 2 do gene do AR (AR55) e sem parte do exão 4 (AR94) foram também detectados. Os objectivos deste trabalho são: 1) caracterizar a distribuição de transcritos do AR em diferentes tecidos humanos (coração, fígado, rim e pulmão); 2) realizar uma análise evolutiva através da caracterização da expressão dos transcritos do AR em testículos de galinha (*Gallus gallus*), cão (*Cannis lupus*), rato (*Rattus norvegicus*) e dourada (*Sparus aurata*). Com este objectivo foi estabelecida a seguinte abordagem experimental: 1) foram desenhados primers flanqueando exões indicados para a detecção de transcritos alternativos do AR; 2) foi extraído RNA total e foi utilizado para síntese de cDNA; 3) foi realizado RT-PCR e os produtos amplificados foram clonados e sequenciados; 4) as sequência obtidas foram analisadas com programas bioinformáticos. Nos tecidos humanos, coração, fígado, rim e pulmão foi observada a presenças dos transcritos AR∆2 e AR∆3 previamente identificados nos testículos. Nos testículos de dourada, cão e rato foram detectados vários transcritos do AR, alguns dos quais homólogos aos identificados em testículos humanos.

Figures list

Table list

Abbreviations list

A – Absorvance
AF1 – N-terminal activation function 1
AF2 – C-terminal activation function 2
AMP – Ampicilin
AR – Androgen receptor
AREs– Androgen response elements
AS – Alternative splicing
BLAST - Basical Local Alignment Search Tool
BSA - Bovine serum albumin
bp – Base pare
CE – capillary electrophoresis
cDNA – Complementary deoxyribonucleic acid
Cl – *Cannis lupus*
Conc. – Concentration
dATP – 2'-deoxyadenosine 5'-triphosphate
dCTP – 2'-deoxycytidine 5'-triphosphate
DBD – DNA -binding domain
DEPC – Diethyl pyrocarbonate
DHT – 5α-dihydrotestosterone
DNA – Deoxyribonucleic acid
dGTP – 2'-deoxyguanosine 5'-triphosphate
dNTP – Deoxyribonucleotide triphosphate
DTT – Dithiothreitol
dTTP – 2'-deoxythymidine 5'-triphosphate)
ER – Estrogen receptor
Fwd - Forward
Gg – *Gallus gallus*
GR – Glucocorticoid receptor
H – Heart
h - Human
IPTG – Isopropyl-beta-D-thiogalactopyranoside
K – Kidney
L – Liver
LB – Luria-Bertani
LBD – Ligand binding domain
Lg – Lung
MAPK – Mitogen-activated protein kinase
Min – Minutes
Mm – Mus musculus
mRNA – Messenger RNA
NCBI – National Center for Biotechnology Information
NLS – Nuclear localization signal
NTD – Amino-terminal domain
OD – Optical density
PCR – Polymerase chain reaction
PKA – Protein hinase A
PR – Progesterone receptor
pre-mRNA – mRNA precursor
PROM2 – Protamine 2
RT-PCR – Reverse transcript polymerase chain reaction
Rn – *Rattus norvegicus*
RNA – Ribonucleic acid
Rv – Reverse
Sa – *Sparus aurata*
SHBG – Sex Hormone-Binding Globulin
Sec – Seconds
snRNPs – Small nuclear ribonucleoproteins
ss – Splice sites
SSHRs – Sex steroid hormone receptors
T – Testosterone
Ta – Annealing temperature
UK – United Kingdom
UV – Ultra-violet
V - Volume
W – Weigth
WT – Wild Type
X-GAL – 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside



I – Introduction

1. Alternative splicing

Alternative splicing (AS) is a mechanism, by which more than one mRNA transcript are generated from the same mRNA precursor (pre-mRNA) due to variations in the incorporation of coding regions giving rise to functional proteins (Kim *et al.* 2007; Marden 2006; Srebrow and Kornblihtt 2006; Venables 2006). AS of untranslated regions (UTR) can also determine mRNA localization and stability, as well as efficiency of translation (Kim *et al.* 2007). AS allows a tremendous amplification of protein diversity (Marden 2006) and 30% to 60% of human genes have alternatively spliced forms, suggesting that alternative splicing is one of the most significant components of the functional complexity of the human genome (Huang *et al.* 2005).

Alternative splicing events are classified into five main subgroups: (1) exon skipping (cassette exons), where the exon can be spliced out of the transcript together with its flanking introns, (2) alternative 5' splice sites (ss) and (3) 3' spliced sites selection, which are the result of the recognition of two or more splice sites at one end of an exon, (4) intron retention, in which an intron can remain in the mature mRNA molecule and (5) mutually exclusive exons which involves the selection of only one from an array of two or more exon variants (Figure 1.1; Kim *et al.* 2007; Matlin *et al.* 2005; Srebrow and Kornblihtt 2006; Stamm *et al.* 2005).

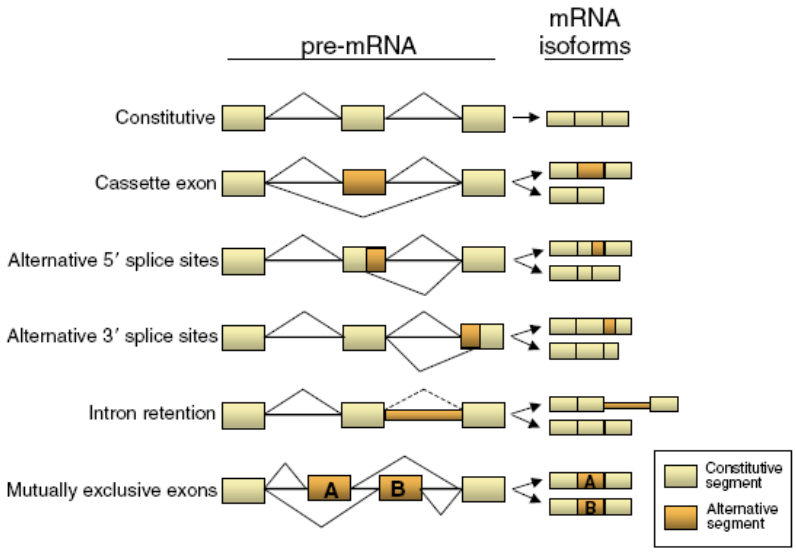


Figure 1.1: Types of alternative splicing (Srebrow and Kornblihtt 2006).

1.1. Splice sites, enhances and silencers

Splicing occurs when a group of small nuclear ribonucleoproteins (snRNPs) recognize specific RNA sequences conserved at the boundaries of introns. The classical spliceosome, recognizes a 5' donor ss beginning with a GU dinucleotide and a 3' acceptor ss ending with an AG dinucleotide (the GU-AG rule). A second type of spliceosome acts on a minor class of Pol-II-transcribed introns obeying the AU-AC rule (Srebrow and Kornblihtt 2006).

When the sequence of a splice site deviates from the consensus shown in the site can still be used, although less efficiently. These weak sites have less affinity for their spliceosomal protein or snRNP partners and are the main cause of alternative splicing. A second class of cis-acting sequence can influence the recognition and use of weak sites by the splicing apparatus: the enhancers and silencers of splicing. These are short (~10 nucleotides) conserved sequences located in exons or introns, either isolated or in clusters, that stimulate or inhibit the use of weak splice sites. An exonic sequence is defined as an exonic splicing enhancer if its mutation reduces inclusion of the corresponding exon into the mature mRNA. Conversely, mutation of an exonic splicing silencer increases inclusion of the exon (Srebrow and Kornblihtt 2006).

1.2. Mechanism of splicing

The positions and sequences of the consensus *cis*-acting elements help to define the splice sites of a typical intron (Matlin *et al.* 2005; Stamm *et al.* 2005). The removal of introns is catalysed by the spliceosome, an assembly of five snRNP particles (U1, U2, U4, U5 and U6) that are associated with a large number of additional proteins and occur in the nucleus (Elliott and Grellscheid 2006; Matlin *et al.* 2005). The spliceosome assembles onto the pre-mRNA through a series of complexes (Figure 1.2; Matlin *et al.* 2005).

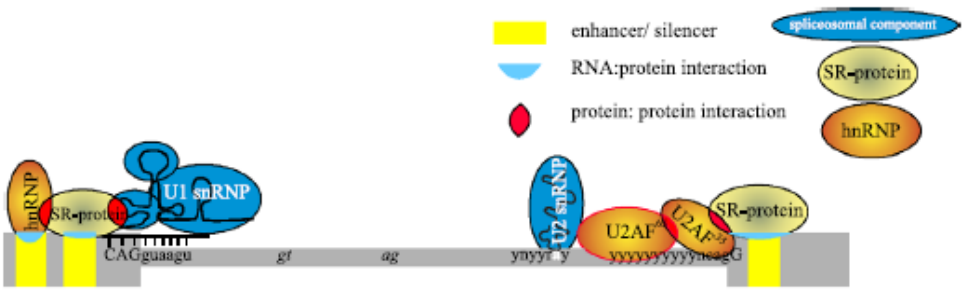


Figure 1.2: Splice-site selection. Exons are indicated as boxes, the intron as a thick line. Splicing regulator elements (enhancers or silencers) are shown as yellow boxes in exons or as thin boxes in introns. The 5' splice site (CAGguaagu) and 3' splice site (y) 10ncagG, as well as the branch point (ynyyray), are indicated (y=c or u, n=a, g, c or u). Two major groups of proteins, hnRNPs (orange) and SR proteins (green), bind to splicing regulator elements; the protein–RNA interaction is shown in blue. This protein complex assembles around an exon enhancer, stabilizing the binding of the U1 snRNP close to the 5' splice site.. Factors at the 3' splice site include U2AF, which facilitates binding to U2 snRNP to the branch point sequence. In exons with weak polypyrimidine tracts, the binding of U2AF is facilitated by the SR proteins binding to exonic enhancers. Green: SR and SR-related proteins; orange: hnRNPs; blue: protein–RNA interaction; red: protein–protein interaction; thick black line with stripes: RNA–RNA hybridization. (Stamm *et al.* 2005)

1.3. Functional significance

The prevalence of AS shows that it may play important roles in many biological processes (Lareau *et al.* 2004). Gene regulation through AS is more versatile than is regulation through promoter activity and it lead to proteome expansion (Lareau *et al.* 2004; Stamm *et al.* 2005). Variant transcripts generated through AS, similar to those initiated from distinct promoters, are often tissue and/or developmental specific, resulting in effects seen only in certain cells or developmental stages. The structural changes fall into three categories: (1) introduction of stop codons, (2) changes of the protein structure and (3) changes in the 5' or 3' UTRs. The effects caused by AS range from a complete loss of function to subtle effects that are difficult to detect (Stamm *et al.* 2005). Evolutionary conservation of AS variants suggests functional importance and provides insight into the evolutionary history of AS (Lareau *et al.* 2004).

AS is likely to be important in many if not all developmental pathways, *e.g.*, in transcriptional re-programing after meiosis AS converts the transcription factor cAMP responsive element modulator (CREM) from an antagonist to a potent activator required for the transcription from an array of promoters in round spermatids (Elliott and Grellscheid 2006) .

2. The androgen receptor

Androgen receptor (*AR*) is a member of the nuclear receptor superfamily which is composed of over 100 members and continues to grow (Gao *et al.* 2005). The term superfamily is used to encompass all of the known nuclear hormone receptors. The superfamily is often further divided into the steroid receptor family and the thyroidlretinoidlvitamin D (or nonsteroid) receptor family. Nuclear receptors can be

grouped into four classes according to their ligand binding, DNA binding, and dimerization properties: steroid receptors, RXR heterodimers, homodimeric orphan receptors, and monomeric orphan receptors (Figure 1.3; Mangelsdorf *et al.* 1995; Stunnenberg 1993).

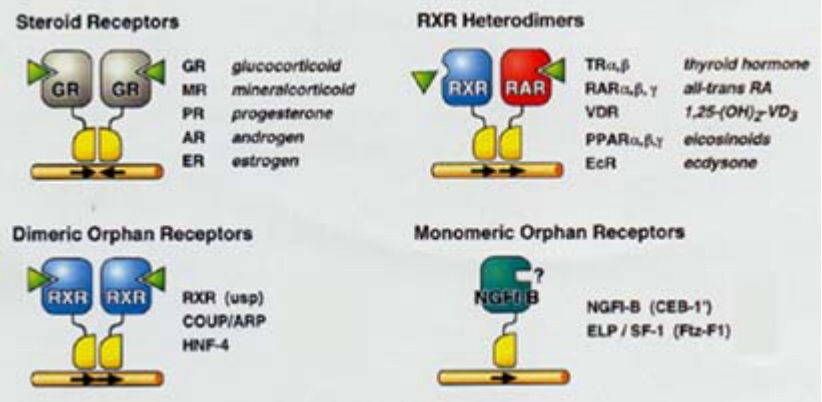


Figure 1.3: Nuclear receptor super-family (*Adapted from* Mangelsdorf *et al.* 1995).

AR belongs to the steroid receptors class and like other steroid receptors is a soluble protein that functions as an intracellular transcriptional factor (Gao *et al.* 2005). One of the known features of this family is the existence of several mRNA variants and in contrast with the members of this family *AR* variants have been relatively poorly studied (Hirata *et al.* 2003).

AR function is regulated by the binding of androgens, which induces an allosteric change that affect receptor protein interactions and receptor-DNA interactions (Gao *et al.* 2005; Mangelsdorf *et al.* 1995). *AR* can be activated by the binding of endogenous androgens, including testosterone (T) and 5α-dihydrotestosterone (DHT). DHT binds more specifically to *AR* than T, since the later binds not only to *AR* but also to other steroid receptors, such as estrogen receptors and progesterone receptor, although with lower affinity (Gao *et al*., 2005). The known *AR* ligands can be classified as steroidal or nonsteroidal based on their structure or as agonist or antagonist based on their ability to activate or inhibit transcription of *AR* target genes (Gao *et al.* 2005; Singh *et al.* 2000). After binding the androgen, *AR* goes through a conformational change and is translocated to the nucleus where it binds to target DNA sequence and initiates the formation of the pre-initiation complex at the promoter regions of androgen responsive genes (Thompson 2006).

Androgens action need a functional *AR* which is responsible for the development of the phenotype during embryogenesis and for male sexual maturation at puberty (Heemers and Tindall 2007). In adult males, androgen is mainly responsible

for maintaining libido, spermatogenesis, muscle mass and strength, bone mineral density, and erythropoisis (Gao *et al.* 2005; Johansen 2004). The actions of androgen in the reproductive tissues, including prostate, seminal vesicle, testis, and accessory structures, are known as the androgenic effects, while the nitrogen retaining effects of androgen in muscle and bone are known as the anabolic effects (Gao *et al.* 2005).

Recently, hormone replacement therapy in aging males has also been proposed to improve body composition, bone and cartilage metabolism, and certain domains of brain function and even decrease cardiovascular risk (Oettel *et al.* 2003).

2.1. Gene organization

The genomic organization of the *AR* gene is conserved throughout mammalian evolution from rodent to man (Gelmann 2002). Human *AR* is encoded by a single copy gene on the long arm of chromosome X at Xq11-12 (Brown *et al.* 1989; Lubahn *et al.* 1988). The mRNA transcript is 10.6 kb long and has an open reading frame of 2757 bp, which codes for the eight exons of *AR* termed A-H or 1-8 (Figure 1.4; Thompson 2006).

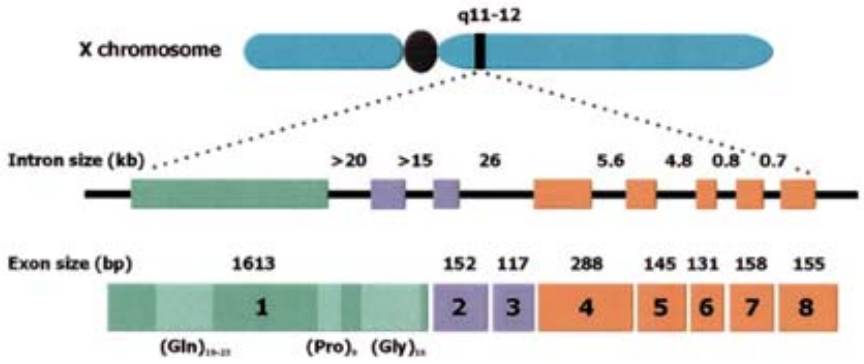


Figure 1.4: Structural organization of the human *AR* gene (*Adapted from* Gelmann 2002).

2.2. Protein structure

There are two possible transcription start sites for the *AR* gene located 1.1kbp upstream of the translation start codon in the 5' untranslated region (UTR). The two transcription start sites are only 10bp apart and therefore code for the same protein (Faber *et al.* 1991). The *AR* gene codes for a protein of 919 amino acids that has three major functional domains (Gao *et al.* 2005). The N-terminal domain (NTD), which serves a modulatory function, encoded by exon 1, the DNA-binding domain (DBD)

encoded by exons 2 and 3 (McEwan 2004) and the ligand binding domain (LBD) encoded by exons 4, 5, 6, 7 and 8. There is also a small hinge region between the DNA-binding domain and ligand-binding domain (Figure 1.5; Gao *et al.* 2005).

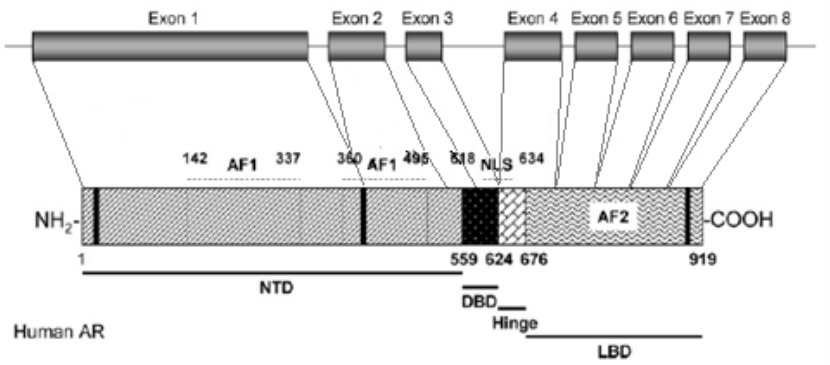


Figure 1.5: Structural organization of the human *AR* gene and protein (*Adapted from* Gao *et al.* 2005).

Two transactivation functions have been identified. The N-terminal activation function 1 (AF1) is constitutively active in truncated receptor that does not contain the ligand-binding domain and is not conserved in sequence compared to other steroid receptors (Figure 3), whereas the C-terminal activation function 2 (AF2) functions in a ligand-dependent manner and is relatively more conserved in sequence as compared to other steroid hormone receptors, particularly with regard to the charge-clamp residues (Chawbshang *et al*., 2002) A nuclear localization signal (NLS) spans the region between the DNA-binding domain and the hinge region (Gao *et al.* 2005), and bind to importin α playing a crucial role in nuclear translocation (Claessens *et al.* 2008).

The *AR* NTD is relatively long and displays the most sequence variability among nuclear receptors. NTD contains the major transactivation function of AR named AF-1. Deletion of the LBD from the *AR* results in a residual N-terminal fragment with constitutive activity almost equal to the transcriptional activity of the ligand-bound, full- length protein (Gelmann 2002; Heemers and Tindall 2007).

The central located DBD is the most conserved region within the nuclear receptor family (Heemers and Tindall 2007). This region includes 9 cystein residues, which 8 are involved in forming two zinc fingers the contained four cysteins and a Zn^{2+} ion, and a C-terminal extention (Figure 1.6; Gelmann 2002; Heemers and Tindall 2007) which produces a helix–loop–helix structural domain that interacts with specific DNA sequences, termed androgen response elements (AREs) (Hiipakka and Liao 1998), forming the structure that binds to the major groove of DNA (Gelmann 2002).

The first zinc finger, located in exon 2, determines the specificity of DNA recognition (Heemers and Tindall 2007) while the second zinc finger, located in exon 3, stabilizes the binding complex by hydrophobic interaction with the fist finger, contributes to specificity of receptor DNA binding and plays a role in receptor dimerization that occurs during DNA binding (Gelmann 2002). The C-terminal extension is important for the overall three-dimensional structure of the DBD and plays a role in mediating the *AR* selectivity of DNA interaction (Figure 1.6; Heemers and Tindall 2007).

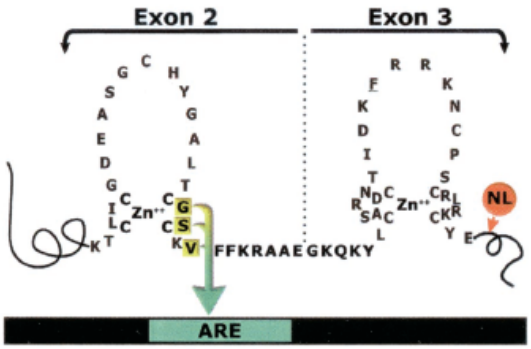


Figure 1.6: Amino acid sequence of the *AR* DBD showing the formation of two zinc fingers. Three residues in boxes, G, S, and V, are highly conserved among steroid hormone receptors such as progesterone and glucocorticoid receptor and are the sites of contact with the androgen responsive elements (ARE). The nuclear localization signal sequence (NL in circle) is C-terminal to the DBD (Gelmann 2002).

The hinge region is a flexible linker between the DBD and the LBD and it is involved in DNA biding and in *AR* dimerization (Heemers and Tindall 2007).

The LBD in nuclear receptors consist of twelve discrete α-helices. Insertion of the agonist into the *AR* ligand-binding pocket changes the conformation of the LBD in such a way that helix 12 is stabilized (Heemers and Tindall 2007). It contains the second transactivation region AF-2. And domain structure and ARs features as for example zinc fingers motifs are absolutely conserved in all vertebrates (Figure 1.7; Gao *et al.* 2005). The human *AR* protein is 88% identical to dog and 85% to rat AR protein. Between human and a frog *AR* protein the identity is 54%, and with a piscine species the percentage of identity is 36%.

The *AR* protein is highly conserved throughout vertebrate evolutive line (Figure 1.7; Gao *et al.* 2005). Conservation of segments of the *AR* protein throughout evolution implicates these regions as being critical for the activity of the molecule. The DBD and the LBD are highly conserved from *Xenopus* to human. Other regions of the gene that retain a striking concentration of sequence conservation include much of the hinge region and the LBD. A large number of LBD sites that are targets for mutation resulting in androgen insensitivity syndrome are conserved from frog to man, also the hinge

region, particularly its N-terminal domain, is highly conserved from frog to rodent (Thornton and Kelley 1998).

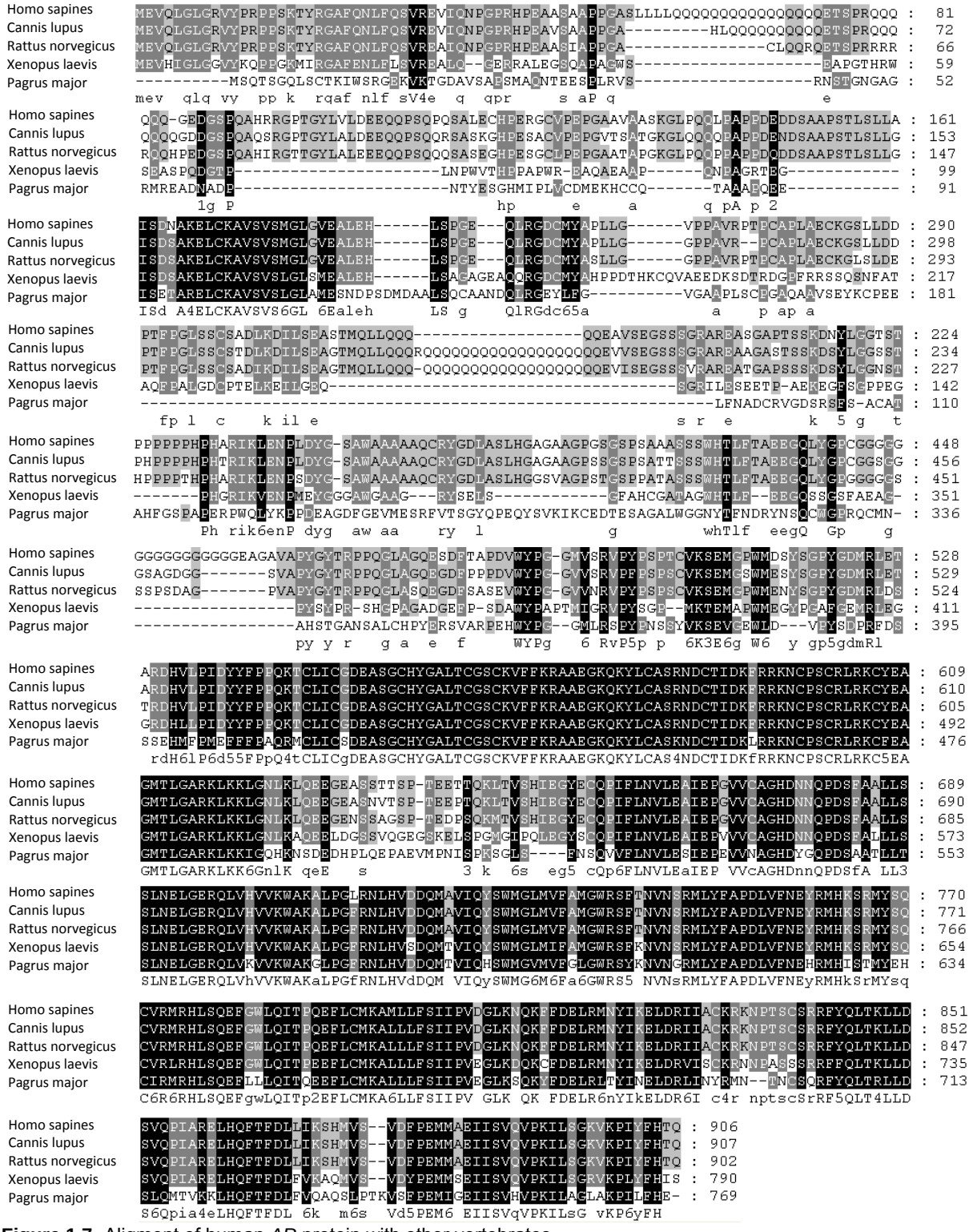


Figure 1.7: Aligment of human *AR* protein with other vertebrates.

2.3. Mecanism of action

Each *AR* domain plays an important role in AR function and signaling. This is either via intra-receptor interactions or via functional interactions with AREs and/or coregulatory proteins. The domains do not function independently, but synergize or antagonize with each other to produce a receptor function that exquisitely regulates the genomic actions of androgens in target tissues (Thompson 2006).

In the absence of androgens, *AR* resides in the cell cytoplasm as a heteroprotein complex with heat-shock proteins (HSP). After T entry into the cell and the possible cell-specific conversion of T to DHT, *AR* binds the androgen. This induces a conformational change in which the HSPs are released and allows *AR* to be translocated to the nucleus (Thompson 2006). Once inside the nucleus, *AR* binds to target ARE which are characterized by the six-nucleotide half-site consensus sequence 5'-TGTTCT-3' spaced by three random nucleotides and is located in the promoter or enhancer region of *AR* target genes (Claessens and Gewirth 2004). Recruitment of other transcription co-regulators (including co-activators and co-repressors) and transcriptional machinery to form the pre-initiation complex are necessary to ensure the transactivation of *AR* regulated gene expression (Figure 1.8; Lee and Chang 2003).

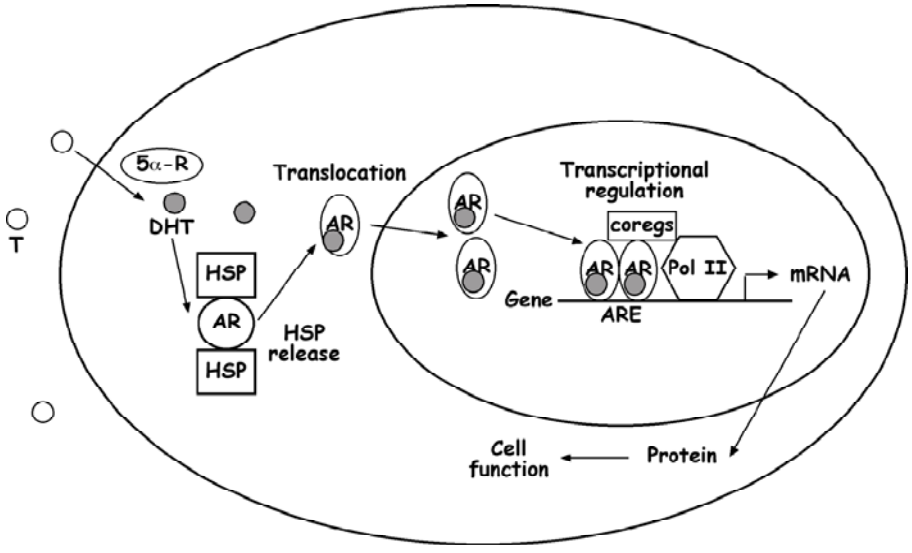


Figure 1.8: *AR* nuclear translocation and transcriptional regulation. *AR*, androgen receptor; ARE, androgen response element; DHT, 5α-dihydrotestosterone; 5α-R, 5α-reductase; T, testosterone; HSP, heat shock protein; Pol II, RNA polymerase II. Testosterone enters the cell and dependening on cell type can be converted to DHT (Thompson 2006).

The general mechanism of nuclear receptor activation by the binding of ligand involves a repositioning of helix 12 in such a way that the ligand-binding pocket is closed, and a hydrophobic cleft is formed on the surface of the LBD. This restructured part of the LBD surface is synonymous to AF-2 and serves as a docking site for coactivators (Claessens *et al.* 2008). A functional AF-2 region is believed to be crucial for co-activator recruitment, because the nuclear receptor box "LxxLL" motif from the nuclear-receptor-interacting domain of coactivators specifically binds to this surface. On the other hand, similar motifs from the *AR* N-terminal domain can also interact with the AF-2 region (He *et al.* 2002).

DNA binding is also required for *AR*-regulated gene expression. The androgen response element half-site sequence can be arranged as either inverted repeats or direct repeats and *AR* recognizes and binds to the ARE site through two zinc fingers located in the DNA-binding domain. Like other steroid receptors, ligand-bound AR forms homodimers and appears to form "head-to-head' dimers even when it is bound to the direct repeats of androgen response element. Selective recognition of specific androgen response element sequences could be regulated by ligand binding or the presence of other transcriptional factors, which bind to their own DNA binding sites as well (Gao *et al.* 2005).

According to the classical model of steroid hormone action, described before, steroid hormones diffuse into the cell, bind to their cognate receptors and induce transcription of the target genes. The length of time between steroid hormone entry into the cell and the accumulation of significant amounts of protein to affect cell function can range from 15-30 minutes to several hours. Within the last decade, an increasing amount of information has been produced on a phenomenon termed the nongenomic actions of steroid hormones (Losel and Wehling 2003). The nongenomic actions of steroid hormones are defined as rapid cellular responses to steroid hormones that occur in the second to minute range of time, although they can occur over longer time periods (Thompson 2006). The effects of nongenomic signaling are not mediated by alterations in transcription and protein synthesis, but rather the immediate activation/repression of cytoplasmic kinase-signaling cascades and intracellular calcium levels (Losel and Wehling 2003).

Androgens may stimulate second messenger cascades in a nongenomic manner through more than one mechanism. Androgens may stimulate the MAPK pathway through stimulation of the nonreceptor tyrosine kinase c-Src (Src). Androgen binding by SHBG can stimulate cAMP and PKA. In addition to these receptors, a plasma membrane, G protein-coupled receptor may also directly bind androgens or

indirectly influence the activity of a membrane androgen-binding protein. One of the effects mediated by this putative receptor is to increase intracellular calcium levels. The elevation of intracellular calcium activates signal transduction cascades, including PKA, PKC, and MAPK, and can modulate the activity of transcription factors. Both mechanisms potentially influence the transcriptional activation of the nuclear *AR* (Figure 1.9; Heinlein and Chang 2002).

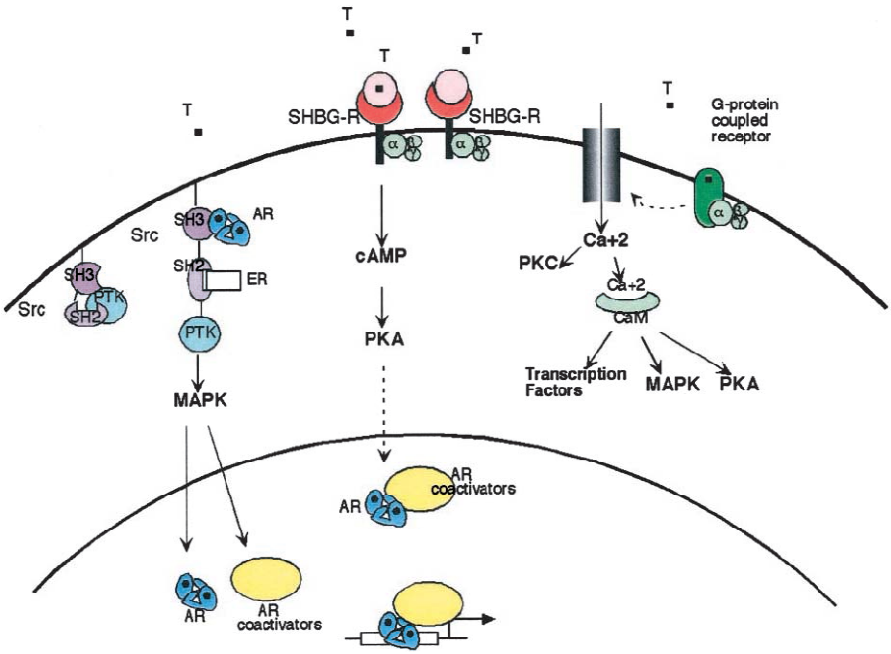


Figure 1.9: Nongenomic action of androgens through multiple receptors (Heinlein and Chang 2002).

3. AR mRNA variants/isoforms

Non-wild-type mRNAs in normal and abnormal tissues are usually called 'isoform' and 'variant', respectively, the terms 'isoform' and 'variant' are not strictly defined. In the present study, it was adopted the more widely used terms.

The open reading frames of human sex steroid hormone receptors (hSSHRs) are composed of eight exons. In addition, the presence of various exons – including 5'-untranslated exons, alternative coding exons and novel 'intronic' exons – has been demonstrated in the genes encoding hSSHRs. The isoform/variant hSSHR mRNAs generated from these exons can be tentatively classified into seven types described in table 1.1 (Hirata *et al.* 2003).

Table 1.1: Type of variant/isoform hSSHR mRNAs (Hirata *et al.* 2003).

Type	Mechanism
Type 1	Alternative use of transcription start sites
Type 2	Deletion of one or more exons
Type 3	Duplication of one or more exons
Type 4	Alternative use of 50-untranslated exons
Type 5	Alternative use of coding exons
Type 6	Use of 'intronic' exon and exons 4 or 5 and 8
Type 7	Insertion of 'intronic' exons

The existence of several variant/isoform RNA transcripts has been described for many steroid receptors. Examples include the estrogen receptor (ER) the ERa46, an ERα form lacking the region encoded by the first exon, which modulates the activity of ERα in MCF7 cells (Flouriot *et al.* 2000), and ERβcx, a C-terminally truncated form containing 26 specific amino acids, which forms heterodimers with ERα to inhibit its function (Ogawa *et al.* 1998). The progesterone receptor (PR) exists as two isoforms, PRA and PRB, differing by the length of the N-terminal end and originating from translation reinitiation at an internal methionine codon. Studies with transgenic mice show that the ratio of both forms is essential for proper development of the mammary gland (Mulac-Jericevic *et al.* 2003). Concerning the glucocorticoid receptor (GR), a splice variant named GRβ has been described which results from an alternative splicing replacement of the final 50 amino acids of hGRα with 15 unique residues. Overexpression of GRβ leads to glucocorticoid resistance in a number of pathological conditions (Yudt *et al.* 2003). In the case of the mineralocorticoid receptor a variant lacking the hinge and ligand-binding regions but able to increase the activity of the full-length receptor has been described (Zennaro *et al.* 2001). However, the information on *AR* mRNA variants is scarce, there are only 5 different variants describe: ∆3 (Zhu *et al.* 1997), AR45 (Ahrens-Fath *et al.* 2005), AR23 (Jagla *et al.* 2007), $AR^{1727372b}$ and AR^{17272b} (Dehm *et al.* 2008). Very recently, our research group described four *AR* variants named AR∆2, AR∆3, AR55 and AR94 (Laurentino *et al*. 2006, results not published).

3.1. ARΔ3

Zhu *et al.* (1997) described for the first time an *AR* variant laking exon 3 in some breast cancer samples and cell lines (Figure 1.10). Expression levels of the variant relative to the WT message varied considerably between samples.

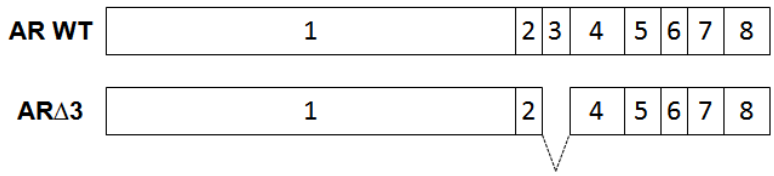


Figure 1.10: Diagram of WT *AR* and exon 3 deleted variant (Δ3) present in breast cancer cells.

The probable functional activity of ARΔ3 variant protein were summarised by considerating the known functional domains of the WT AR. Exon 3 is the coding exon for the second zinc finger which is important in orientating the receptor for DNA binding, in stabilising DNA–protein interaction and in providing the interface for receptor dimerisation. Partial deletion of the DNA-binding domain *in vitro* completely inactivates the ability of receptor protein to stimulate transcription, though normal steroid-binding properties are retained (Hollenberg *et al.* 1987). The presence of ARΔ3 in breast cancer cells may reduce or abolish the DNA-binding capacity of *AR*. Thus, the growth-inhibitory role of androgens through AR may be reduced or lost (Zhu *et al.* 1997).

3.2. AR45

Ahrens-Fath, together with his research collaborators (2005), identified and characterized AR45, a human *AR* isoform composed of a unique N-terminal extension linked to the DBD, hinge region and LBD of the *AR*. It lacked the entire region encoded by exon 1 of the AR which was replaced by a short, unique seven amino acid-long N-terminal extension (Figure 1.11). The deduced molecular mass was about 45 kDa, hence the name AR45. This *AR* variant after homology search revealed that the AR45-specific coding region was located on chromosome Xq11, between exons 1 and 2 of the *AR* gene. A survey of human tissues revealed that AR45 was expressed in most tissues, but it was mainly expressed in heart and skeletal muscle. AR45 may act as a dominant-negative inhibitor of *AR* function suggesting an important biological role in modulating androgen action (Ahrens-Fath *et al.* 2005).

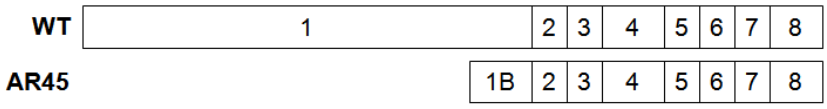


Figure 1.11: Diagram of WT AR and AR45 variant present in human tissues.

3.3. AR23

The AR23 variant results from an aberrant splicing of intron 2 that conduct to the insertion of 23-amino acids between the two zinc fingers in the DNA-binding domain (Figure 1.12) which correspond to the last 69 nucleotides of intron 2. This variant was identified in a hormone refractory metastic prostate cancer cell line (Jagla *et al.* 2007).

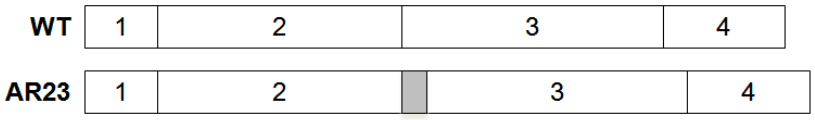


Figure 1.12: Diagram of WT *AR* and AR23 variant in metastic prostate cancer cell line. Shadow box corresponds to a 23 aminoacid insertion.

The 23-amino acids insertion occurs in a key region of the DBD, between the carboxyl-terminal part of the first zinc finger, which contains the P-box responsible for specific DNA-binding and the nuclear export signal, and the second zinc finger that contains the dimerization domain and a nuclear localization signal what turn AR23 unable to be translocated into the nucleus (Jagla *et al.* 2007). However, AR23 showed to be functional once it affects transcriptional activities increasing NF-κB and decreasing AP-1 activity. The NF-κB family of transcription factors is known to regulate the expression of adhesion molecules that promote cell migration and cell-cell interactions and to induce expression of genes that contribute to tumor progression such as antiapoptotic genes and genes that regulate cell growth (Suh and Rabson 2004). AP-1 transcription factor has been implicated in cell growth, differentiation, and development by mediating gene regulation in response to a various range of physiological and pathological stimuli (Shaulian and Karin 2002). These findings suggest that AR23 may contribute to tumor progression by protecting prostate cancer cells from apoptosis (Jagla *et al.* 2007).

3.4. $AR^{1/2/2b}$ and $AR^{1/2/3/2b}$

Very recently two new variants of *AR* were identified in 22Rv1 prostate cancer cells by Dhem *et al*. (2008). The AR NTD/DBD isoforms expressed in 22Rv1 cells arise from splicing of a novel *AR* exon 2b after either AR exon 2 or exon 3. AR exon 2b spliced after exon 2 or exon 3 gives rise to protein species possessing the entire AR NTD fused to the first zinc finger of the *AR* DBD (termed $AR^{Ex1/2/2b}$) or the entire AR NTD fused to the complete AR DBD (termed $AR^{Ex1/2/3/2b)}$, respectively (Figure 1.13; Dehm *et al.* 2008).

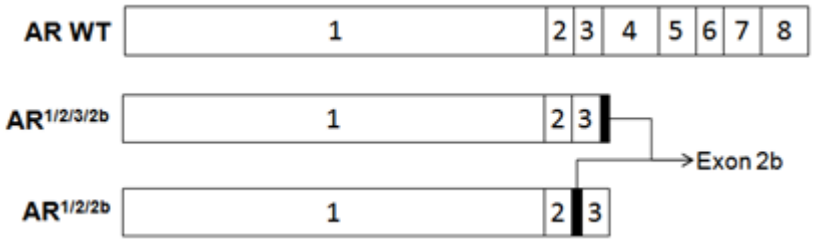


Figure 1.13: Diagram of WT *AR* and $AR^{1/2/2b}$ and $AR^{1/2/3/2b}$ isoforms present in prostate cancer cells.

The two isoforms with exon 2b insertion generate a truncated *AR* consisting in a complete NTD, the entire zinc fingers and the exon 2b-derived sequence. Both isoforms can independently function as constitutively active and transcription factors (Dehm *et al.* 2008). The first zinc finger, encoded by exon 2, harbors the recognition helix that directly engages with one hexameric half-site in an ARE. The second zinc finger, encoded by exon 3, mediates dimerization with an AR molecule engaged with the neighboring ARE half-site (Shaffer *et al.* 2004). Dehm *et al.* finding that $AR^{Ex1/2/2b}$ is able to constitutively activate *AR* responsive promoters shows that the first zinc finger is sufficient for the *AR* to engage with AREs, which agrees with this structural data. However, because the second exon 3 encoded zinc finger is responsible for AR dimerization and subsequent stabilization of ARE-bound ARs, the repertoire of genes that are activated by $AR^{Ex1/2/2b}$ may be diminished compared with the repertoire of genes that are activated by full-length, liganded AR. In their study Dehm and co-authors showed that short AR isoforms mediate androgen-independent proliferation of 22Rv1 cells (Dehm *et al.* 2008).

3.5. AR∆2, AR∆3, AR55 and AR94

In figure 1.14 are represented the four *AR* mRNA variants detected in human testis tissues (Laurentino *et al.* 2006, results not published). AR∆2 is an *AR* mRNA variant with an exon 2 delection which gives rise to premature stop codon. AR∆3 variant was previously described by Zhu *et al.*, 1997 in breast cancer cells and this variant results in a diminut DNA-binding capacity of AR. AR55 variant correspond to a 55 aminoacid insertion that correspond to part of intron 2 resultin in the insertion of a premature STOP codon. AR94 lacks the last 94 aminoacids coded by exon 4 and it will result in a protein with an incomplete LBD.

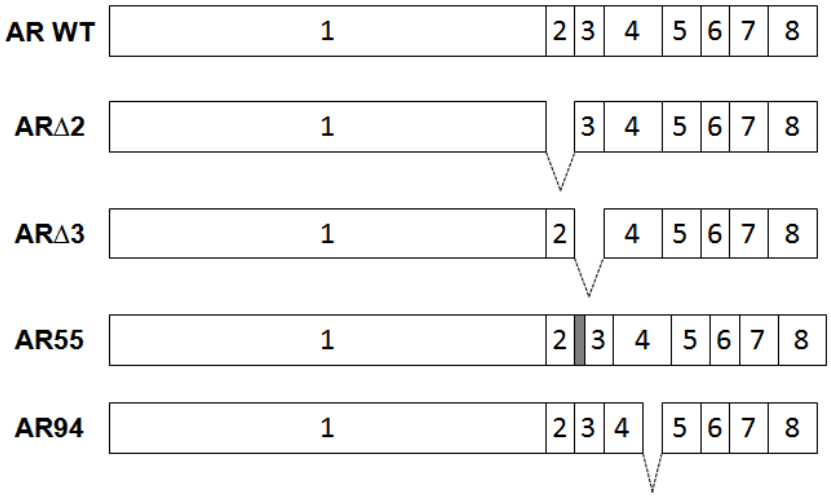
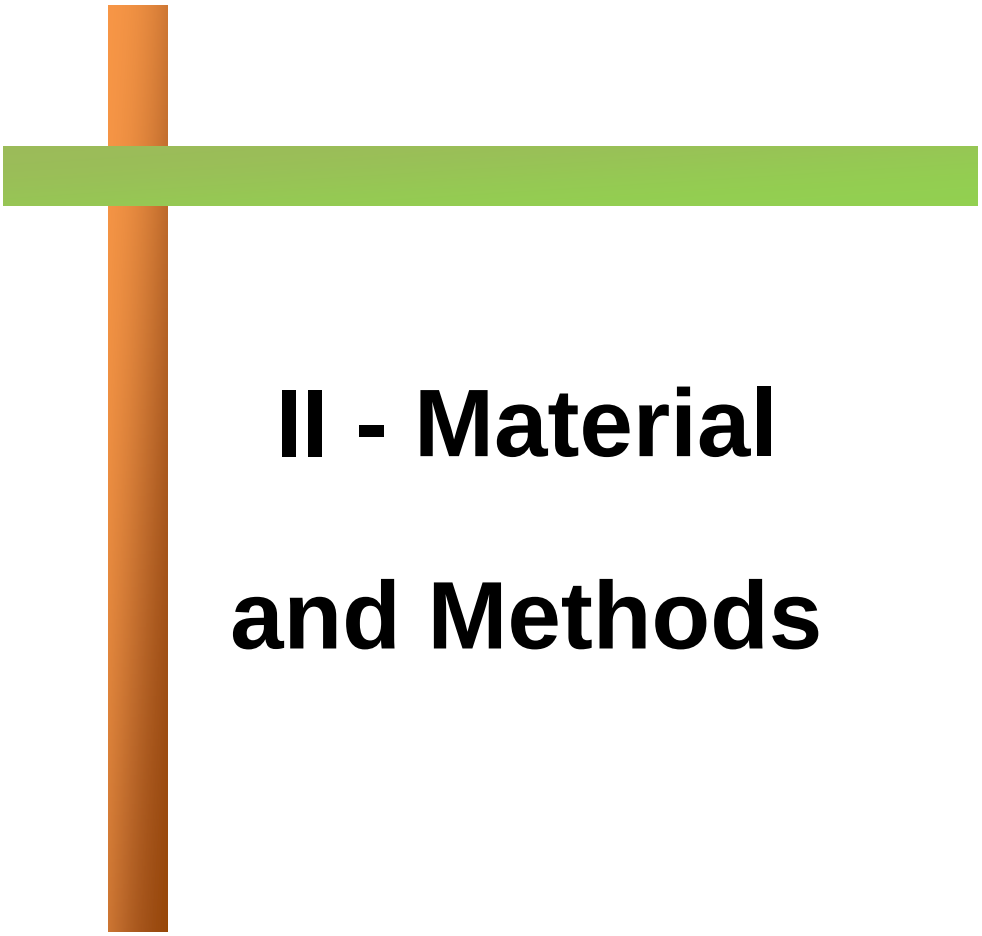


Figure 1.14: Diagram of WT *AR* and ∆2, ∆3, AR55 and AR94 variants detected in testis.

4. Aim of the study

The existence of several variant RNA transcripts has been described for many steroid receptors. However, the information on alternative spliced mRNAs of AR gene is scarce. A previous work in our research group (Laurentino, S., data not published) allowed the identification for the first time of AR alternative transcripts in human testis. The aims of the present study was to characterize the distribution of AR transcripts found in testis samples in different human tissues (heart, kidney, liver, and lung) and to perform an evolutionary analysis by characterizing the expression of AR transcripts in testis of chicken (*Gallus gallus*), dog (*Cannis lupus*), rat (*Rattus norvegicus*) and sea bream (*Sparus aurata*).



II - Material and Methods

1. Samples

Cannis lupus and *Galus gallus* testis were obtained from surgery at a veterinary Clinic (Clínica Veterinária da Covilhã, Dr. Hugo Brancal), or collected by Prof.Dr. J.E.Cavaco (Health Sciences Research Centre, University of Beira Interior) at a local farm. After collection, tissues were immediately frozen in liquid nitrogen and stored at -80ºC until RNA extraction. *Sparus aurata* polyA$^+$ RNA and *Rattus norvegicus* total RNA testis samples were provide by Prof.Dr. Sílvia Socorro (Health Sciences Research Centre, University of Beira Interior). Comercial total RNA samples from human heart, kidney, liver and lung were obtained from Clontech (Appendix A).

2. RNA extraction

Total RNA was extracted from *C.luppus* and *G.gallus* testis using TRI Reagent® (Sigma-Aldrich, St. Louis, USA) according to manufacturer's instructions (Appendix B).

The quality of total RNA was determinated by running 10µl of a 1:10 dilution in a 1% agarose gel stained with ethidium bromide and photographed under UV illumination. The integrity of RNA was also assessed by calculating the ratio between 260nm (A260) and 280nm (A280) absorvances.

Concentration of total RNA from each sample was photometrically determinated using a NanoPhotometer (IMPLEN, Reutlingen, Germany). Total RNA samples were stored at -80°C until use

3. Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

3.1. cDNA synthesis

Initially ≈1µg of total RNA was pre-incubated at 65°C for 5 min with 50ng of Random Hexamers Primers (Invitrogen, Paisley, UK), 0,5mM dNTP Mix (10mM each dATP, dGTP, dCTP and dTTP at neutral pH; GE Healthcare, Buckinghamshire, UK) and sterile water to a final volume up to 12µl. After chilling on ice and a brief centrifugation, 1x first strand buffer (Invitrogen, Paisley, UK), 0.01mM DTT (Invitrogen, Paisley, UK) and 20 units of RNaseOUT™ (Recombinant Ribonuclease Inhibitor 40 units/µl; Invitrogen, Paisley, UK) were added to each sample and the tubes were incubated at 37ºC for 2 min. Two hundred units of M-MLV reverse transcriptase (Invitrogen, Paisley, UK) were added to each tube and incubated for 10 min at 25°C followed by 50 min at 37ºC. Reverse transcriptase activity was inactivated by incubation at 70°C for 15min. cDNA samples were stored at -20°C until use.

3.2. Primer design

Using the *AR* sequences retrieved from the GenBanck database (Table 2.1), species specific primers spanning exons 2 to 4 of *AR* (Table 2.2) were designed using Primer3 V.0.4.0 software (http://frodo.wi.mit.edu). These primers set were used to amplify the *AR* transcripts.

Table 2.1: *AR* sequences of species used in present study.

Species	Accession number
Cannis lupus	NM_001003053
Gallus gallus	NM_001040090
Homo sapiens	NM_000044
Rattus norvegicus	NM_012502
Sparus aurata	Prof. Sílvia Socorro (not published)
Xenopus laevis	NM_001090884

In appendix C are provided all *AR* nucleotide and aminoacid sequences with indication of exon boundaries, protein domains and primer sequeences.

Specific primers for 18S ribossomal RNA (18S) were also designed for amplification of 18S in order to control the success of cDNA synthesis

Table 2.2: Details of primers used for amplification of *AR* and 18S cDNA

Species	Primer set	Exon location	Sequence	Amplicon size
Cannis lupus	ClAR1/5	Ex1 Ex5	5′-CAG AGT GCC CTT TCC AAG TC-3′ 5′-TGA ATG ACT GCC ATC TGG TC-3′	687bp
Gallus gallus	GgAR1/5	Ex1 Ex5	5′-CTG CAT CAA GAG CGA GCT G-3′ 5′-ATC TGG TCA TCC ACA TGC AA-3′	651bp
Homo sapiens	18ShrCAV[1]		5′-AAG ACG AAC CAG AGC GAA AG-3′ 5′-GGC GGG TCA TGG GAA TAA-3′	152bp
	hAR1/4	Ex1 Ex4	5′-GTC AAA AGC GAA ATG GGC CCC-3′ 5′-CTT CTG GGT TGT CTC CTC AGT-3′	420bp
	hAR1/5	Ex1 Ex5	5′-GTC AAA AGC GAA ATG GGC CCC-3′ 5′-GTC GTC CAC GTG TAA GTT GCG-3′	642bp
Mus musculus	MmAR1/5[2]	Ex1 Ex5	5′-GTG AAA TGG GAC CTT GGA TG-3′ 5′-TAC TGA ATG ACC GCC ATC TG-3′	655bp
Spaurus aurata	Sa18S		5′-TCA AGA ACG AAA CTC GGA GG-3′ 5′-GGA CAT CTA AGG GCA TCA CA -3′	500bp
	SaAR1/5	Ex1 Ex5	5′-CGA TGT CCC CTA CAA TGA CC-3′ 5′-GGT CGT CCA CAT GGA GAT TT-3′	599bp

[1] – These primers were 100% identical to *Rattus norvegicus* 18S sequence and were used to amplify *AR* in this specie.
[2] – These primers were 100% identical to *Rattus norvegicus AR* sequence and were used to amplify *AR* in this specie.

3.3. PCR reaction

PCR reactions were performed with 1U of GoTaq DNA Polymerase (Promega, Madison, USA), 1X PCR Buffer, 0,2mM of dNTP's, 1,5mM of $MgCl_2$, 0,5µM of each primer and 1µL of cDNA. A negative control was run in parallel with sterile H_2O instead of cDNA. Reactions were carried out using a Px2 ThermoHybaid thermal cycler. RT-PCR products were separated on a 1,5% (w/v) ethidium bromide-stained agarose gel and photographed under UV illumination. Primes were optimized in terms of annealing temperature and number of cycles (data not shown). Optimized PCR reactions were performed as showed in table 2.3 in duplicate for each sample.

Table 2.3: Conditions used in PCR reactions for amplification of *AR* and 18S cDNA.

Primer set	Conditions	Cycles
ClAR1/5	94ºC – 5'; 94ºC – 1', 55ºC – 1', 72 ºC – 30''; 72ºC – 5'.	35
GgAR1/5	94ºC – 5'; 94ºC – 1', 58ºC – 1', 72 ºC – 30''; 72ºC – 5'.	45
18ShrCAV	94ºC – 3'; 94ºC – 30'', 56ºC – 30'', 72ºC – 30''; 72ºC – 5'.	25
hAR1/4	94ºC – 2'; 94ºC – 1', 55ºC – 1', 72ºC – 1''; 72ºC – 5'.	35
hAR1/5	94ºC – 2'; 94ºC – 1', 55ºC – 1', 72ºC – 1''; 72ºC – 5'.	35
MmAR1/5	94ºC – 5'; 94ºC – 30', 55ºC – 30', 72ºC – 1''; 72ºC – 5'.	38
Sa18S	94ºC – 5'; 94ºC – 45'', 55ºC – 30'', 72ºC – 45''; 72ºC – 5'.	15
SaAR1/5	94ºC – 5'; 94ºC – 1', 55ºC – 1', 72 ºC – 30''; 72ºC – 5'.	40

4. Cloning and sequencing

PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). Briefly, equal volume of Membrane Wash Solutution was added to each PCR product. The mixtures were transferred into SV Minicolumns, inserted into collection tubes and centrifuged, flowthrough was rejected and the membrane was washed. SV Minicolumns were transferred into a sterile 1,5mL tube and the DNA was eluted with 50µL of Nuclease Free Water. The eluted DNA was ligated to pGEM®-T Easy Vector (Promega). The ligation reactions were performed with 50ng of pGEM-T Easy Vector (Promega), 1x rapid ligation buffer (Promega), 3µL of DNA sample and 100 units of T_4 DNA ligase (Promega) and incubated overnight at 4ºC. *E. coli* competent cells were transformed with 10µL of the ligation products, by an incubation on ice for at least 35 minutes, followed by an heat shock at 42ºC for 2 minutes. Immediately, bacterias were plated in LB agar (USB, USA) plates with ampicilin (50µg/mL, USB, USA), IPTG (100µg/mL, Sigma, USA) and X-Gal (100µg/mL,

Eppendorf, Germany) and were incubated overnight at 37ºC in an air incubator (JP Selecta, Spain). White colonies were collected to 2mL of LB broth containing ampicillin (100µg/mL) and incubated overnight at 37ºC in an orbital shaking incubator (200 rpm; ARABAL Agitorb 200E, Portugal). Plasmid DNA was purified from liquid cultures using Wizard® *Plus* SV Minipreps DNA Purification System (Promega). Briefly, cells were pelleted by centrifugation, the pellet was ressuspend, cells were lysed, proteins were precipitated and the obtained lysates were transferred into Spin Columns. After centrifugation DNA was retained on the spin column and washed with a solution containing ethanol. Plasmid DNA was eluted with 100µL of Nuclease Free Water. To confirm the presence of inserts restriction digestions were performed using: 1XH Buffer, 0,01% Triton X-100, 0,01% BSA, 0,5 units of NotI (Takara), 2,5µL plasmid DNA and water up to 10µL. Digestions were carried out at 37ºC for 1 hour and the digestion products were run on a 1% agarose gel stained with ethidium bromide and photographed under UV illumination. Plasmids containing the insert were sequenced by StabVida (Oeiras, Portugal) using T7 fwd pGEM primer (5'- GTA TAC GAC TCA CTA TAG GGC -3'). To determine its identity and structural osganization, obtained sequences were aligned with GenBank sequences using Clustal W software (Thompson *et al.* 1994).

III – Results

1. Confirmation of total RNA integrity

1.1. Electrophoresis and spectrophotometric analysis

Total RNA extracted from *C. lupus* and *G. gallus* testis was analysed by agarose gel electrophoreisis (Figure 3.1) and spectrophotometrically at 260 and 280 nm (results listed in table 3.1).

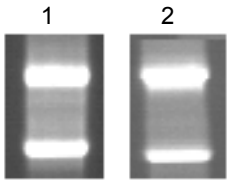


Figure 3.1: Electrophoresis of total RNA in 1% agarose gel. Each lane was loaded with 1µL of total RNA, 19µL of DEPC-treated water and 2µL of loading buffer. 1 – *C. Lupus*; 2 – *G. Gallus*.

Table 3.1: A260/A280 Ratio and concentration of each RNA sample.

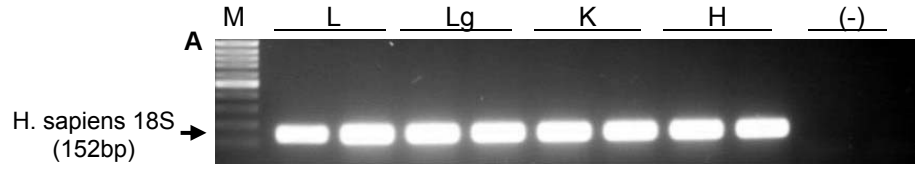
RNA sample	A260nm	A280nm	A260/A280	Conc. (µg/µL)
C.lupus	0,938	0,503	1.865	1,87
G.gallus	1,870	0,930	2,011	18,7

For other samples the integrity of total RNA was assessed by the expression analysis of 18S.

1.2. Expression of 18S ribosomal RNA

To confirm the success of cDNA synthesis, the expression of 18S was determinate in human, *R. Norvegicus* and *S. aurata* samples.

Figure 3.2 shows the electrophoresis of the 18S amplified products. All bands correspond to the expected size, which demonstrate the integrity of RNA samples and the quality of the synthesized cDNA.



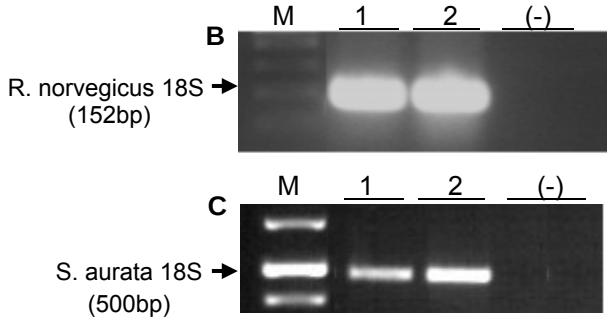


Figure 3.2: Electrophoresis of PCR products using 18ShrCAV (A), Sa18S (B) and Rn18S (C) primer sets primers. M - GeneRuler™ 1kb Plus DNA Ladder (Fermentas); L – liver, Lg – lung, K – hidney, H – heart; 1,2 – two individual samples of *R. Norvegicus* and *S. aurata* testis cDNA; (-) – PCR reaction without cDNA.

2. *AR* alternative transcripts are expresses in several human tissues

RT-PCR analysis showed wild type (WT) *AR* expression in all tissues used namely liver, lung, kidney and heart (Figure 3.3). Under each WT band visible in electrophoresis it is possible to observe the presence of smaller bands corresponding to putative alternative spliced transcripts.

PCR products corresponding to putative alternative spliced transcripts amplified in lung samples were selected to be cloned in order to confirm their identity. The presence of inserts was confirmed by Not1 digestion (Figure 3.4).

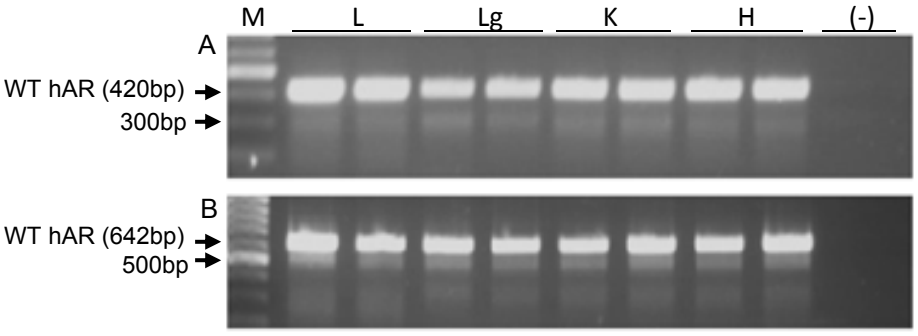


Figure 3.3: Electrophoresis of PCR products using hAR1/4 (A) and hAR1/5 (B) primer sets. M - GeneRuler™ DNA 100bp Ladder Mix, K - kidney, L - liver, Lg - lung, H - hear, (-) – PCR reaction without cDNA.

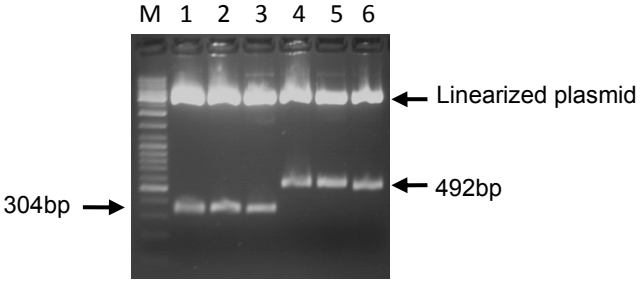


Figure 3.4.: Electrophoresis of plasmid digestions confirming the cloning of putative alternative transcripts amplified in human lung samples using hAR1/4 (1, 2 and 3) and hAR1/5 (4, 5 and 6) primers sets.

Clustal W alignments of PCR products sequence with *AR* sequence retrieved from GenBank (accession no. NM_000044) showed that the fragment with approximate 300bp correspond to an exon 3 deleted *AR* variant (AR∆3, 304bp, Figure 3.5), which does not contain the second Zn finger coding region. The amplificated product with approximately 500bp correspond to an exon 2 deleted *AR* variant (AR∆2, 492bp, Figure 3.6), which correspond to the insertion of a premature STOP codon.

```
Homosapiens   2640 TGAGCAGAGTGCCCTATCCCAGTCCCACTTGTGTCAAAAGCGAAATGGGCCCCTGGATGG 2700
PCRproduct         ------------------------------GTCAAAAGCGAAATGGGCCCCTGGATGG
                                                  ****************************

Homosapiens   2701 ATAGCTACTCCGGACCTTACGGGGACATGCGTTTGGAGACTGCCAGGGACCATGTTTTGC 2760
PCRproduct         ATAGCTACTCCGGACCTTACGGGGACATGCGTTTGGAGACTGCCAGGGACCATGTTTTGC
                   ************************************************************

Homosapiens   2761 CCATTGACTATTACTTTCCACCCCAGAAGACCTGCCTGATCTGTGGAGATGAAGCTTCTG 2820
PCRproduct         CCATTGACTATTACTTTCCACCCCAGAAGACCTGCCTGATCTGTGGAGATGAAGCTTCTG
                   ************************************************************

Homosapiens   2821 GGTGTCACTATGGAGCTCTCACATGTGGAAGCTGCAAGGTCTTCTTCAAAAGAGCCGCTG 2880
PCRproduct         GGTGTCACTATGGAGCTCTCACATGTGGAAGCTGCAAGGTCTTCTTCAAAAGAGCCGCTG
                   ************************************************************

Homosapiens   2881 AAGGGAAACAGAAGTACCTGTGCGCCAGCAGAAATGATTGCACTATTGATAAATTCCGAA 2940
PCRproduct         AAG-------------------------------------------------------
                   ***

Homosapiens   2941 GGAAAAATTGTCCATCTTGTCGTCTTCGGAAATGTTATGAAGCAGGGATGACTCTGGGAG 3000
PCRproduct         ------------------------------------------------------------

Homosapiens   3001 CCCGGAAGCTGAAGAAACTTGGTAATCTGAAACTACAGGAGGAAGGAGAGGCTTCCAGCA 3060
PCRproduct         CCCGGAAGCTGAAGAAACTTGGTAATCTGAAACTACAGGAGGAAGGAGAGGCTTCCAGCA
                   ************************************************************

Homosapiens   3061 CCACCAGCCCCACTGAGGAGACAACCCAGAAGCTGACAGTGTCACACATTGAAGGCTATG 3120
PCRproduct         CCACCAGCCCCACTGAGGAGACAACCCAGAAGA-------------------------
                   ********************************
```

Figure 3.5: Clustal alignment of sequenced PCR product obtained with hAR1/4 primer set with human *AR* sequence available at GenBank (accession no NM_000044). Underlined sequences correspond to forward and reverse primers. In grey is highlighted the sequence deleted in the PCR product which correspond to exon 3, as could be confirmed by nucleotide position at appendix C.

```
Homosapiens   2640 TGAGCAGAGTGCCCTATCCCAGTCCCACTTGTGTCAAAAGCGAAATGGGCCCCTGGATGG 2700
PCRproduct         ---------------------------------TCAAAAGCGAAATGGGCCCCTGGATGG
                                                     **************************

Homosapiens   2701 ATAGCTACTCCGGACCTTACGGGGACATGCGTTTGGAGACTGCCAGGGACCATGTTTTGC 2760
PCRproduct         ATAGCTACTCCGGACCTTACGGGGACATGCG----------------------------
                   *****************************

Homosapiens   2761 CCATTGACTATTACTTTCCACCCCAGAAGACCTGCCTGATCTGTGGAGATGAAGCTTCTG 2820
PCRproduct         ------------------------------------------------------------

Homosapiens   2821 GGTGTCACTATGGAGCTCTCACATGTGGAAGCTGCAAGGTCTTCTTCAAAAGAGCCGCTG 2880
PCRproduct         ------------------------------------------------------------

Homosapiens   2881 AAGGGAAACAGAAGTACCTGTGCGCCAGCAGAAATGATTGCACTATTGATAAATTCCGAA 2940
PCRproduct         ---GGAAACAGAAGTACCTGTGCGCCAGCAGAAATGATTGCACTATTGATAAATTCCGAA
                      ************************************************************

Homosapiens   2941 GGAAAAATTGTCCATCTTGTCGTCTTCGGAAATGTTATGAAGCAGGGATGACTCTGGGAG 3000
PCRproduct         GGAAAAATTGTCCATCTTGTCGTCTTCGGAAATGTTATGAAGCAGGGATGACTCTGGGAG
                   ************************************************************

Homosapiens   3001 CCCGGAAGCTGAAGAAACTTGGTAATCTGAAACTACAGGAGGAAGGAGAGGCTTCCAGCA 3060
PCRproduct         CCCGGAAGCTGAAGAAACTTGGTAATCTGAAACTACAGGAGGAAGGAGAGGCTTCCAGCA
                   ************************************************************

Homosapiens   3061 CCACCAGCCCCACTGAGGAGACAACCCAGAAGCTGACAGTGTCACACATTGAAGGCTATG 3120
PCRproduct         CCACCAGCCCCACTGAGGAGACAACCCAGAAGCTGACAGTGTCACACATTGAAGGCTATG
                   ************************************************************

Homosapiens   3121 AATGTCAGCCCATCTTTCTGAATGTCCTGGAAGCCATTGAGCCAGGTGTAGTGTGTGCTG 3180
PCRproduct         AATGTCAGCCCATCTTTCTGAATGTCCTGGAAGCCATTGAGCCAGGTGTAGTGTGTGCTG
                   ************************************************************

Homosapiens   3181 GACACGACAACAACCAGCCCGACTCCTTTGCAGCCTTGCTCTCTAGCCTCAATGAACTGG 4240
PCRproduct         GACACGACAACAACCAGCCCGACTCCTTTGCAGCCTTGCTCTCTAGCCTCAATGAACTGG
                   ************************************************************

Homosapiens   4241 GAGAGAGACAGCTTGTACACGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGCAACT 4300
PCRproduct         GAGAGAGACAGCTTGTACACGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGCAACT
                   ************************************************************

Homosapiens   4301 TACACGTGGACGACCAGATGGCTGTCATTCAGTACTCCTGGATGGGGCTCATGGTGTTTG 4360
PCRproduct         TACACGTGGACGACA----------------------------------------------
                   **************
```

Figure 3.6: Clustal alignment of sequenced PCR product obtained with hAR1/5 primers and human *AR* sequence available at GenBank (accession no NM_000044). Underlined sequences correspond to forward and reverse primers. In grey is highlighted the sequence deleted in the PCR product which correspond to exon 2, as could be confirmed by nucleotide position at appendix C.

3. *AR* transcripts are expressed in vertebrate evolutive line

3.1. An exon 2 deleted *AR* mRNA variant is expressed in *Spaurus aurata*

Using the SaAR1/5 primer set, was possible to amplify the WT *AR* and also a smaller band corresponding to putative alternative spliced *AR* transcripts (Figure 3.7).

PCR products corresponding to putative alternative spliced transcripts were cloned. Presence of inserts was confirmed by Not1 digestion (Figure 3.8)

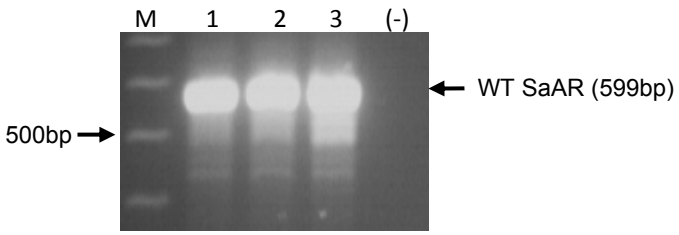


Figure 3.7: Electrophoresis of PCR products using SaAR1/5 primer set . M - GeneRuler™ DNA 100bp Ladder Mix, (-) – PCR reaction without cDNA.

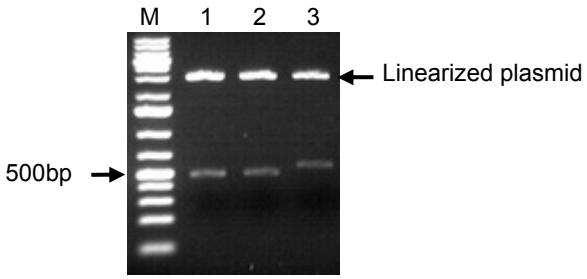


Figure 3.8: Electrophoresis of plasmid digestions confirming the cloning of putative alternative transcripts amplified S. aurata testis samples using SaAR1/5 (1, 2 and 3) primers sets.

Clustal W alignments of PCR products sequence with *AR* sequence provided by Prof. Dr.ª S. Socorro showed that the fragment with approximate 500bp correspond to an exon 2 deleted *AR* (AR∆2, 448, Figure 3.9), which give rise to the insertion of a premature STOP codon.

```
Sparusaurata     361 GAGTGGCTCGATGTCCCCTACAATGACCCCAGGTTCGACGCCAGCAGCGAGCACATGTTC 420
PCRproduct           --------CGATGTCCCCTACAATGACCCCAGG----------------------------
                             *************************

Sparusaurata     421 CCAATGGAGTTCTTCTTTCCAGCTCAGAGGATGTGTATGATCTGTTCAGACGAGGCGTCT 480
PCRproduct           ------------------------------------------------------------

Sparusaurata     481 GGCTGCCATTACGGTGCACTCACCTGTGGCAGCTGCAAGGTTTTCTTCAAAAGAGCTGCA 540
PCRproduct           ------------------------------------------------------------

Sparusaurata     541 GAAGGCAAACAGAAATACCTATGCGCAAGCAAAAATGACTGCACTATTGATAAGCTAAGA 600
PCRproduct           -----CAAACAGAAATACCTATGCGCAAGCAAAAATGACTGCACTATTGATAAGCTAAGA
                          ************************************************

Sparusaurata     601 AGAAAGAACTGTCCGTCCTGTCGGCTGAGGAAGTGCTTCGAAGCTGGAATGACTCTGGGA 660
PCRproduct           AGAAAGAACTGTCCGTCCTGTCGGCTGAGGAAGTGCTTCGAAGCTGGAATGACTCTGGGA
                     ************************************************************

Sparusaurata     661 GCACGTAAACTGAAAAAGATTGGGCAACAGAAGAACCCGGACGAAGATCATCCTCTCCAG 720
PCRproduct           GCACGTAAACTGAAGAAGATTGGGCAACAGAAGAACCCGGACGAAGATCATCCTCCTCAG
                     ************** ****************************************  ***
```

```
Sparusaurata     721 GAGCCTGCAGAGGTTATGCCCAATATCTCTCCTAAAATGGGCCTGAGCTTCAACTCTCAA 780
PCRproduct           GAGCCTGCAGAGGTTATGCCCAATATCTCTCCTAAAATGGGCCTGAGCTTCAACTCTCAA
                     ************************************************************

Sparusaurata     781 GTGGTCTTCTTGAACATCCTGGAGTCCATTGAGCCCGAGGTGGCGTACGCAGGACACGAC 840
PCRproduct           GTGGTCTTCCTGAACATCCTGGAGTCCATTGAGCCCGAGGTGGCGTACGCAGGACACGAC
                     ********* ************************************************

Sparusaurata     841 TATGGCCAACCGGATTCAGCCCCCACCCTGTTCACTAGCCTCAACGAGCTTGGGGAAAAA 900
PCRproduct           TACGGCCAACCGGACTCAGCCGCCACCCTGCTCACTAGCCTCAACGAGCTTGGGGAGAAA
                     ** *********** ****** ******** ************************ ***

Sparusaurata     901 CAGCTGGTGAAAGTGGTCAAATGGGCAAAAGGATTCCCAGGTTTTAGAAATCTCCATGTG 960
PCRproduct           CAGCTGGTGAAAGTGGTCAAATGGGCAAAAGGATTGCCAGGTTTTAGAAATCTCCATGTG
                     ********************************* ************************

Sparusaurata     961 GACGACCAAATGACTGTCATCCAGTATTCATGGATGGGGGTGATGGTGTTCGGCCTCGGG 1020
PCRproduct           GACGACCA----------------------------------------------------
                     ********
```

Figure 3.9: Clustal alignment of PCR product obtained with SaAR1/5 primers with *S.aurata AR*. Underlined sequences correspond to forward and reverse primers. In grey is highlighted the sequence deleted in the PCR product which correspond to exon 2, as could be confirmed by nucleotide position at appendix C.

3.2. An exon 4 deleted *AR* mRNA variant is expressed in *Cannis* lupus and *Rattus norvegicus*

Using the ClAR1/5 and the MmAR1/5 primer sets, was possible to amplify the WT *AR* and also a smaller band corresponding to putative alternative spliced *AR* transcripts (Figure 3.10).

PCR products corresponding to putative alternative spliced transcripts were cloned. Presence of inserts was confirmed by NotI digestion (Figure 3.11).

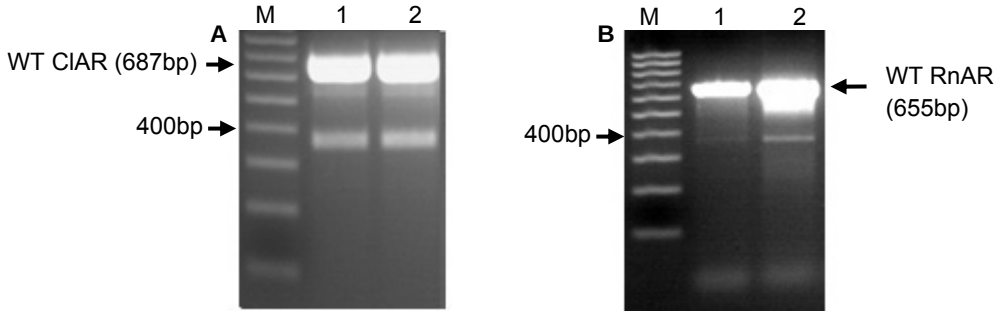


Figure 3.10: Electrophoresis of PCR products using ClAR1/5 (A) and GgAR1/5 (B) primer sets used for amplifications of WT and alternative spliced transcribed *AR*. M - GeneRuler™ DNA 100bp. 1, 2 – two indivual samples of *C. lupus* ans *R. norvegicus* testis cDNA.

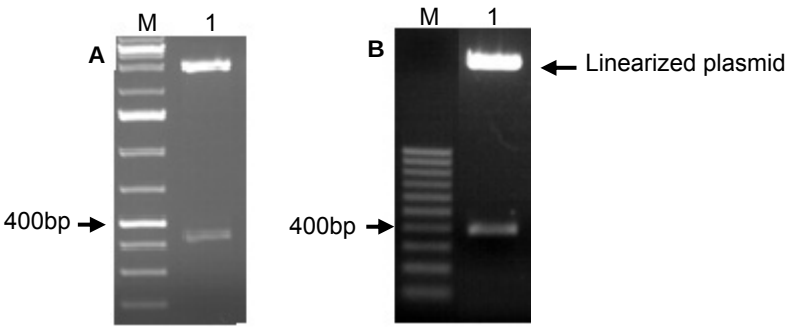


Figure 3.11: Electrophoresis of plasmid digestions confirming the cloning of putative alternative transcripts amplified in *C. lupus* and *R. norvegicus* testis samples using ClAR1/5 (A) and MmAR1/5 (B) primers sets.

Clustal W alignments of PCR products sequence with AR sequence retrieved from GenBank (accession no. NM_012502) showed that both fragments with approximate 400bp correspond to an exon 4 deleted AR variant (ARΔ4, 360bp for *C. lupus*, Figure 3.12 and 360bp for *R. norvegicus*, Figure 3.13), which does not contains the hinge coding region and misses the initial part of the NTD coding region, in both species.

```
Cannislupus   1441 GAAGGTGACTTCCCTCCACCTGATGTGTGGTATCCGGGCGGTGTGGTGAGCAGAGTGCCC 1500
PCRproduct         --------------------------------------------------CAGAGTGCCC
                                                                      **********

Cannislupus   1501 TTTCCAAGTCCTAGTTGTGTCAAAAGCGAGATGGGCTCTTGGATGGAGAGCTACTCCGGA 1560
PCRproduct         TTTCCAAGTCCTAGTTGTGTCAAAAGCGAGATGGGCTCTTGGATGGAGAGCTACTCCGGA
                   ************************************************************

Cannislupus   1561 CCCTATGGGGACATGCGTTTGGAGACTGCCAGGGACCATGTTCTACCCATTGACTATTAC 1620
PCRproduct         CCCTATGGGGACATGCGTTTGGAGACTGCCAGGGACCATGTTCTACCCATTGACTATTAC
                   ************************************************************

Cannislupus   1621 TTTCCACCTCAGAAGACCTGTCTGATCTGCGGTGATGAAGCTTCTGGCTGTCACTATGGA 1680
PCRproduct         TTTCCACCTCAGAAGACCTGTCTGATCTGCGGTGATGAAGCTTCTGGCTGTCACTATGGA
                   ************************************************************

Cannislupus   1681 GCTCTCACTTGTGGAAGCTGCAAAGTCTTCTTTAAAAGAGCCGCTGAAGGGAAACAGAAG 1740
PCRproduct         GCTCTCACTTGTGGAAGCTGCAAAGTCTTCTTTAAAAGAGCCGCTGAAGGGAAACAGAAG
                   ************************************************************

Cannislupus   1741 TACCTGTGTGCCAGCAGAAATGATTGTACCATCGATAAATTCCGAAGG-AAAAATTGTCC 1800
PCRproduct         TACCTGTGTGCCAGCAGAAATGATTGTACCATCGATAAATTCCGAAGGAAAAAATTGTCC
                   ********************************************** ***********

Cannislupus   1801 ATCTTGTCGCCTCCGGAAATGCTATGAAGCAGGGATGACTCTGGGAGCCCGGAAGCTAAA 1860
PCRproduct         ATCTTGTCGCCTCCGGAAATGCTATGAAGCAGGGATGACTCTGGGAG------------
                   *******************************************

Cannislupus   1861 GAAACTGGGGAATCTGAAACTGCAAGAGGAAGGAGAGGCTTCCAATGTCACCAGCCCCAC 1920
PCRproduct         ------------------------------------------------------------

Cannislupus   1921 TGAGGAGCCAACCCAGAAGCTGACGGTGTCACACATTGAAGGCTATGAGTGTCAGCCCAT 1980
PCRproduct         ------------------------------------------------------------

Cannislupus   1981 CTTTCTGAATGTCCTTGAAGCCATCGAGCCAGGCGTGGTGTGTGCTGGACATGACAACAA 2040
PCRproduct         ------------------------------------------------------------

Cannislupus   2041 CCAGCCCGACTCCTTTGCAGCCTTGCTCTCTAGCCTTAATGAATTGGGTGAAAGGCAGCT 2100
PCRproduct         ------------------------------------------------------------
```

```
Cannislupus    2101 TGTACATGTGGTCAAGTGGGCCAAGGCCTTGCCGGGCTTCCGCAACCTGCACGTGGATGA 2160
PCRproduct          ----------------------------------GCTTCCGCAACCTGCACGTGGATGA
                                                       ************************

Cannislupus    2161 CCAGATGGCAGTCATTCAGTACTCCTGGATGGGGCTCATGGTGTTTGCCATGGGCTGGCG 2220
PCRproduct          CCAGATGGCAGTCATTCAA----------------------------------------
                    ******************
```

Figure 3.12: Clustal alignment of PCR product obtained with ClAR2/4 primers with *C.lupus* AR sequence available at GenBank (accession no NM_001003053). Underlined sequences correspond to forward and reverse primers. In grey is highlighted the sequence deleted in the PCR product which correspond to exon 4, as could be confirmed by nucleotide position at appendix C.

```
Rattusnovegicus    2469 GTTGTGAACAGAGTCCCCTATCCCAGTCCCAGTTGTGTTAAAAGTGAAATGGGACCTTGG 2520
PCRproduct              ----------------------------------------GTGAAATGGGACCTTGG
                                                                ******************

Rattusnovegicus    2521 ATGGAGAACTACTCCGGACCTTATGGGGACATGCGTTTGGACAGTACCAGGGACCACGTT 2580
PCRproduct              ATGGAGAACTACTCCGGACCTTATGGGGACATGCGTTTGGACAGTACCAGGGACCACGTT
                        ************************************************************

Rattusnovegicus    2581 TTACCCATCGACTATTACTTCCCACCCCAGAAGACCTGCCTGATCTGTGGAGATGAAGCT 2640
PCRproduct              TTACCCATCGACTATTACTTCCCACCCCAGAAGACCTGCCTGATCTGTGGAGATGAAGCT
                        ************************************************************

Rattusnovegicus    2641 TCTGGTTGTCACTACGGAGCTCTCACTTGTGGCAGCTGCAAGGTCTTCTTCAAAAGAGCT 2700
PCRproduct              TCTGGTTGTCACTACGGAGCTCTCACTTGTGGCAGCTGCAAGGTCTTCTTCAAAAGAGCT
                        ************************************************************

Rattusnovegicus    2701 GCGGAAGGGAAACAGAAGTATCTATGTGCCAGCAGAAATGATTGCACCATTGATAAATTT 2760
PCRproduct              GCGGAAGGGAAACAGAAGTATCTATGTGCCAGCAGAAATGATTGCACCATTGATAAATTT
                        ************************************************************

Rattusnovegicus    2761 CGGAGGAAAAATTGTCCATCGTGTCGTCTCCGGAAATGTTATGAAGCAGGGATGACTCTG 2820
PCRproduct              CGGAGGAAAAATTGTCCATCGTGTCGTCTCCGGAAATGTTATGAAGCAGGGATGACTCTG
                        ************************************************************

Rattusnovegicus    2821 GGAGCTCGTAAGCTGAAGAAACTTGGAAATCTCAAACTACAGGAAGAAGGAGAAAACTCC 2880
PCRproduct              GGAG------------------------------------------------------
                        ****

Rattusnovegicus    2881 AGTGCTGGTAGCCCCACTGAGGACCCATCCCAGAAGATGACTGTATCACACATTGAAGGC 2940
PCRproduct              ------------------------------------------------------------

Rattusnovegicus    2941 TATGAATGTCAACCTATCTTTCTTAATGTCCTGGAAGCCATTGAGCCAGGAGTGGTGTGT 3000
PCRproduct              ------------------------------------------------------------

Rattusnovegicus    3001 GCCGGACATGACAACAACCAGCCTGATTCCTTTGCTGCCTTGTTATCTAGTCTCAACGAG 3060
PCRproduct              ------------------------------------------------------------

Rattusnovegicus    3061 CTTGGCGAGAGACAGCTTGTACATGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGC 3120
PCRproduct              ------------------------------------------------GCTTCCGC
                                                                        ********

Rattusnovegicus    3121 AACTTGCATGTGGATGACCAGATGGCAGTCATTCAGTATTCCTGGATGGGACTGATGGTA 3180
PCRproduct              AACTTGCATGTGGATGACCAGATGGCGGTC----------------------------
                        ************************* ***
```

Figure 3.13: Clustal alignment of PCR product obtained with MmAR1/5 primers with *R.norvegicus* AR sequence available at GenBank (accession no NM_012503). Underlined sequences correspond to forward and reverse primers. In grey is highlighted the sequence deleted in the PCR product which correspond to exon 4, as could be confirmed by nucleotide position at appendix C.

Using the GgAR1/5 primer sets was possible to amplify the WT *AR* and also a smaller band corresponding to putative alternative spliced transcripts (figure 3.14).

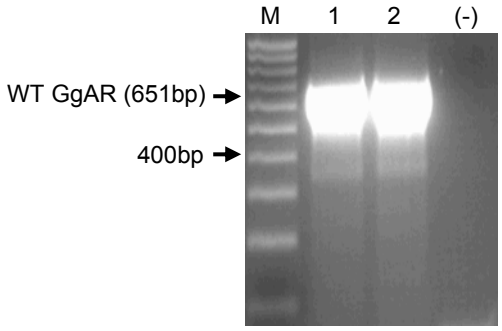
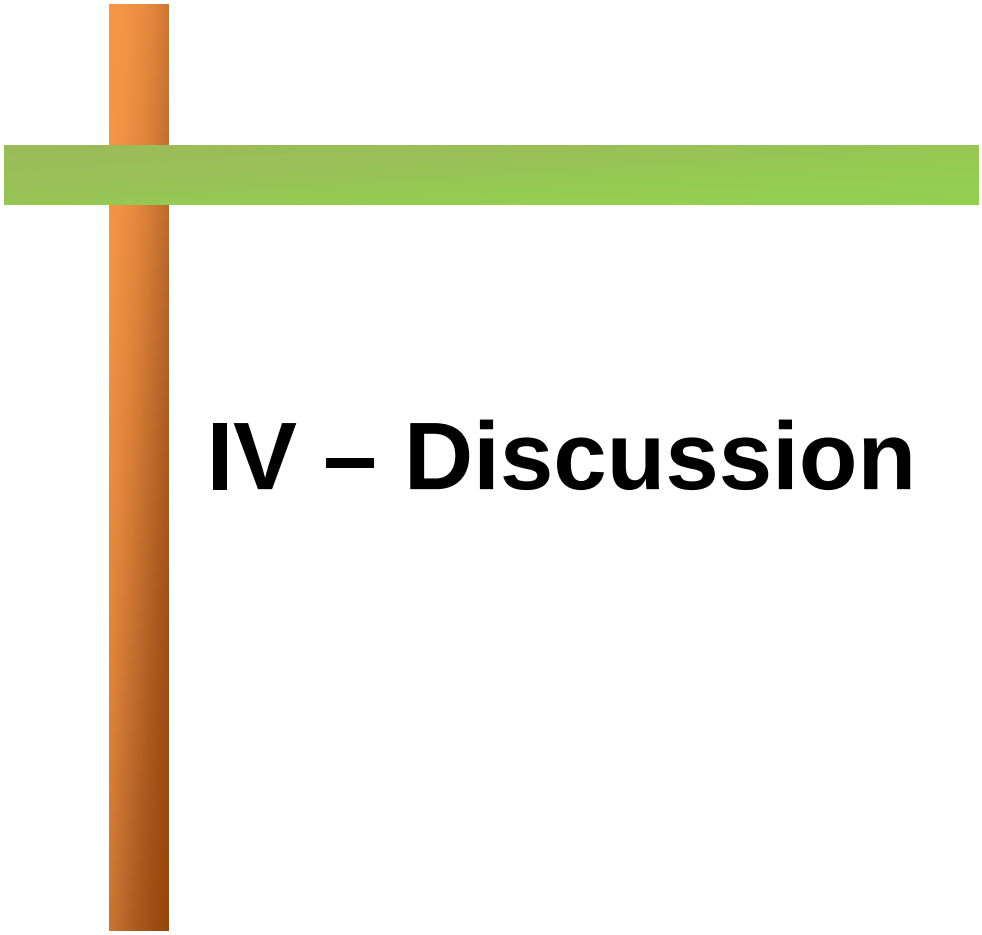


Figure 3.14: Electrophoresis of products from PCR using GlAR1/5 primer set used for amplifications of WT and alternative spliced transcribed *AR*. M - GeneRuler™ DNA 100bp; 1,2 – two individual samples of *G.gallus* testis cDNA; (-) – PCR reaction without cDNA.

PCR products corresponding to putative alternative spliced transcripts were ligated to pGEMT. However, it was not possible to achieve a successful transformation being impossible to determine the identity of proximally 400bp band.



IV – Discussion

The purpose of the present study was to characterize the expression of *AR* variants, namely ARΔ2, ARΔ3, AR55 and AR94, previously identified in human testis (Laurentino *et al.* 2006, results not published). For this purpose, their expression was analysed in a variety of human tissues (heart, kidney, liver, and lung), and an evolutionary analysis was performed by characterizing the expression of *AR* transcripts in *C. lupus*, *G. gallus*, *R*. *norvegicus* and *S. aurata* testis.

With the primer set used for amplification of *AR* transcripts in human testis, was possible to amplify ARΔ2 and ARΔ3 variants in heart, kidney, liver and lung. Since these variants are expressed ubiquitously among human tissues it is likely that they will be translated to proteins (Elliott and Grellscheid 2006). The ARΔ2 and ARΔ3 variants result from exon 2 and 3 spkipping, respectively, a common process in alternative spliced variants. Exon 2 and 3 encode the first and second Zn^{2+} finger, respectively, which constitute the DBD (Gao *et al.* 2005). The ARΔ2 variant will result in the presence of a premature STOP codon that will lead to an incomplete *AR* protein only with the first domain, the NTD. Considerating the role of Zn^{2+} fingers in functional *AR* DBD, it is espected that ARΔ3 variant will be unable to bind DNA. ARΔ3 variant may be unable to orientate the receptor for DNA binding, to stabilize DNA-protein interaction and to provide the interface receptor dimerization (Zhu *et al.* 1997). The other *AR* variants identified in testis, AR55 and AR94, were not detected in other human tissues suggesting that it may be implicated on specific testicular functions.

The precise functions of *AR* transcripts identified in testis, and other human tissues, remain to be clarified.

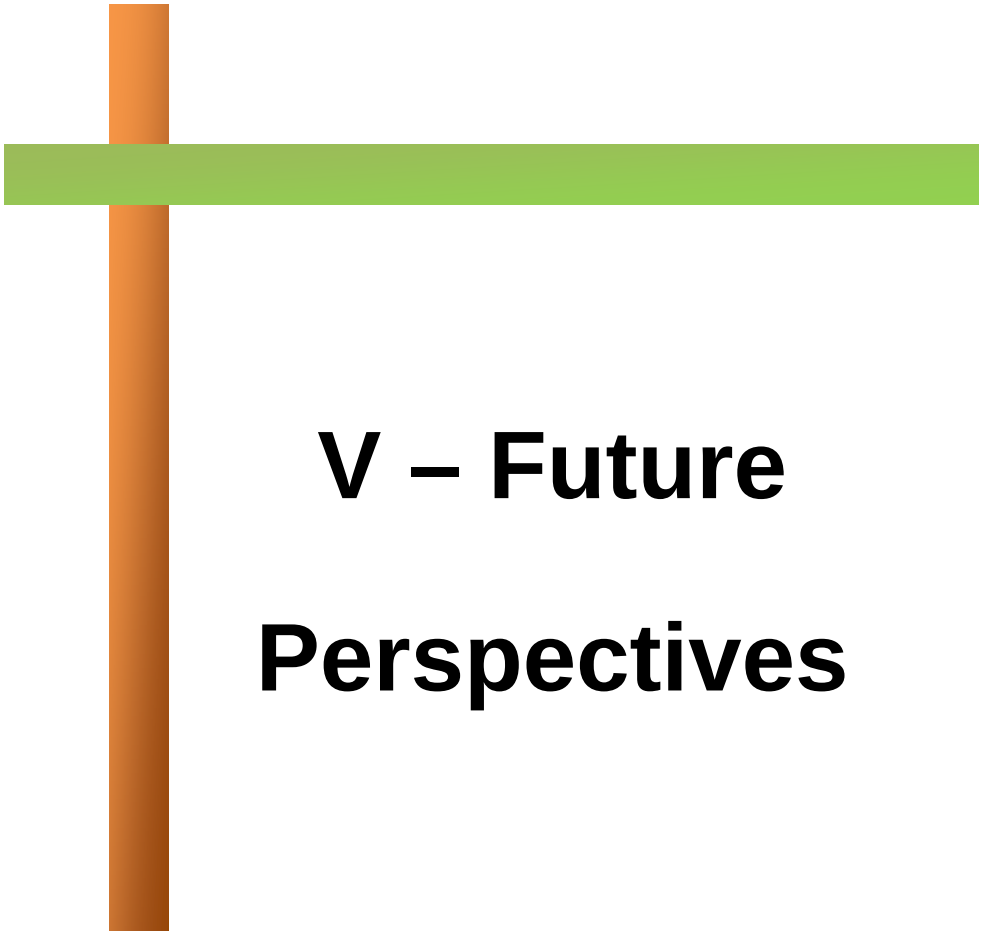
The evolutionary analysis of *AR* transcripts in *C. lupus*, *G. gallus*, *R, norvegicus* and *S. aurata* testis was performed with primers spanning exon 2 to 4, in order to amplify AR variants equivalent to those identified in human testis (Laurentino *et al.* 2006, results not published). Exons 2, 3 and 4 encode the highly conserved region of DBD (Thornton and Kelley 1998) and the hinge region.

In *S. aurata* testis, a piscine species, which is evolutionary the most distant from *H. sapiens* used in this study, was identified an ARΔ2 variant. This variant is equivalent to the human ARΔ2 showing the insertion of a premature STOP codon which will give rise to an incomplete *AR* protein. The majority of alternative spliced cassette exons are not evolutionary conserved, they may result from aberrant splicing and likely do not create functional protein. Diverged alternative splicing is enriched in genes expressed

in testis and cancerous cell lines where increased rate of aberrant splicing may result from abnormal cellular conditions, rapid cell proliferation or faulty surveillance mechanism (Kan *et al.* 2005). However, evolutionary conservation is an established indicator of functional importance (Elliott and Grellscheid 2006; Kan *et al.* 2005; Matlin *et al.* 2005). Despite the fact that AR∆2 variant generates an incomplete protein its conservation between man and sea bream highlight for a relevant function and suggest that these transcripts may generate functional proteins.

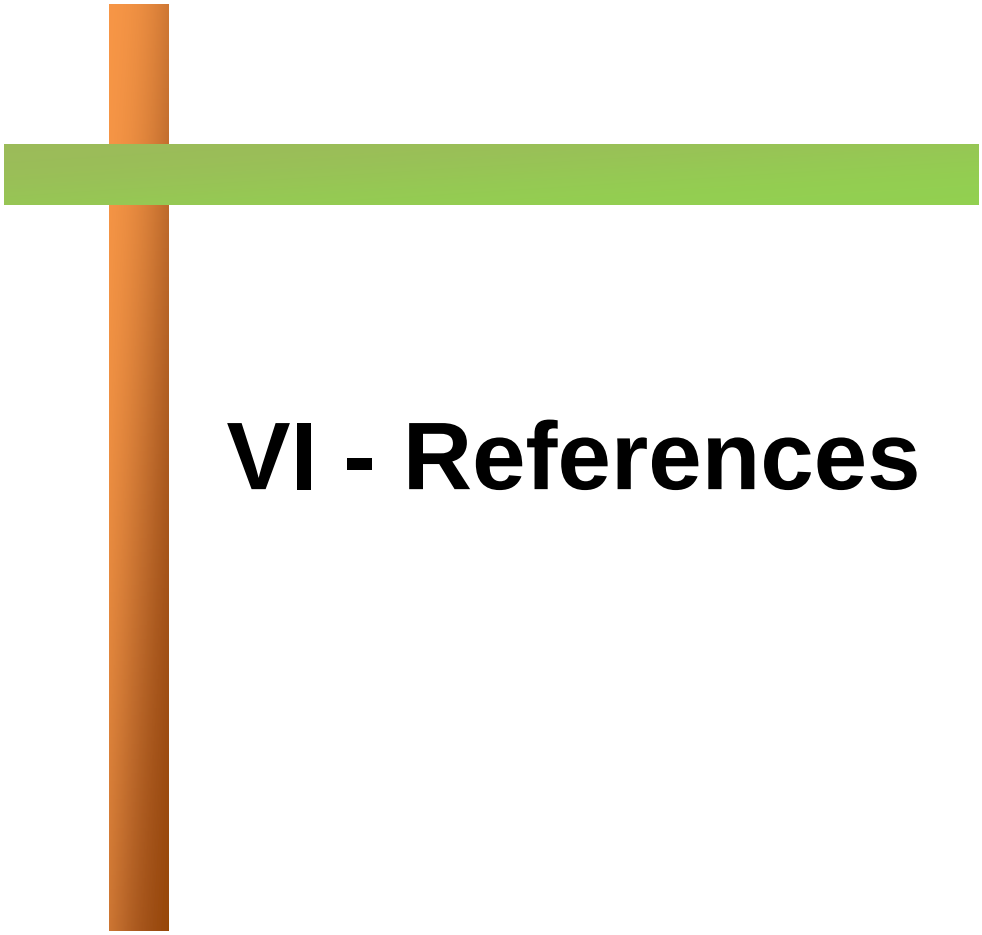
In *C. lupus* and *R. Norvegicus* the expression of a new *AR* variant with deletion of exon 4 (AR∆4), not detected in other species, was described. According to genomic and protein organization, for both species an AR∆4 variant lacks almost the entire hinge region and the first 47 aminoacids of the LBD. The hinge region comprises the C-terminal end of the DBD and the first helix of the LBD, and contains the bipartite NLS that enables targeting of the activated *AR* to the nucleus (Thompson 2006). Thus, the nuclear translocation of the AR∆4 may be compromised and it would correspond to a cytoplasmatic protein. Furthermore, the hinge region has a role in regulating transcriptional activity by modulating the NTD/LBD interaction (Zhou *et al.* 1994). So this *AR* variant will be unable to have NTD/LBD interactions what may lead to an unstable structure (Claessens *et al.* 2008). AR∆4 variant also lacks part of the LBD. This domain has a total of 18 amino acids that directly interact with the androgen bound in the ligand-binding pocket (Matias *et al.* 2000). Due to the high conservation of LBD (99% identity between *H. sapiens*, *C. lupus* and *R. Norvegicus*) it is possible to infer that 5 of these 18 aminoacids are directly involved in ligand binding and encoded by exon 4. Thus, due to the lost of the first 47 aminoacids of the LBD the AR∆4 variant may not be able to bind to androgen or if so it may be a week and/or instable interaction. If we add the fact that an exon 4 deleted variant implicate the lost of the hinge region and an incomplete LBD we can hipotetize that if translated this variant will differ from the wild type *AR* not only in cell localization but also in function. Although, as discussed before alternative spliced cassette exons are not evolutionary conserved and they probably will not create functional protein (Kan *et al.* 2005). However, in the case of AR∆4 variant it is conserved between *C. lupus* and *R. Norvegicus* and if evolutionary conservation at sequence level is an established indicator of functional importance (Elliott and Grellscheid 2006; Kan *et al.* 2005; Matlin *et al.* 2005), this *AR* variant will probably generate a functional protein.

In summary with the present work we showed that the AR∆2 and AR∆3 variants identified in testis (Laurentino *et al.* 2006, results not published) are expressed in a variety of human tissues, namely, heart, kidney, liver and lung. In *C. lupus*, *G. gallus*, *R. norvegicus* and *S. aurata* testis were detected several *AR* transcripts some of which are homologous to those identified in human testis. Altogether present results suggest that AR variants may have a relevant physiological function and showed the increasing complexity of *AR* signaling.



V – Future Perspectives

This study demonstrated the presence of *AR* variants in a variety of human tissues and troughout vertebrate evolutive line. This raised important questions about the function of these variants in *AR* signalling and function. Further studies will be necessary to confirm weather the identified *AR* are translated and start deciphering the biological function of the corresponding proteins. It would be also of interest to characterize the ligand-binding and trasactivation properties of these proteins.



VI - References

Ahrens-Fath I, Politz O, Geserick C, Haendler B (2005) Androgen receptor function is modulated by the tissue-specific AR45 variant. *Febs J* 272, 74-84.

Brown CJ, Goss SJ, Lubahn DB, Joseph DR, Wilson EM, French FS, Willard HF (1989) Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism. *Am J Hum Genet* 44, 264-269.

Claessens F, Denayer S, Van Tilborgh N, Kerkhofs S, Helsen C, Haelens A (2008) Diverse roles of androgen receptor (AR) domains in AR-mediated signaling. *Nucl Recept Signal* 6, e008.

Claessens F, Gewirth DT (2004) DNA recognition by nuclear receptors. *Essays Biochem* 40, 59-72.

Dehm SM, Schmidt LJ, Heemers HV, Vessella RL, Tindall DJ (2008) Splicing of a novel androgen receptor exon generates a constitutively active androgen receptor that mediates prostate cancer therapy resistance. *Cancer Res* 68, 5469-5477.

Elliott DJ, Grellscheid SN (2006) Alternative RNA splicing regulation in the testis. *Reproduction* 132, 811-819.

Faber PW, King A, van Rooij HC, Brinkmann AO, de Both NJ, Trapman J (1991) The mouse androgen receptor. Functional analysis of the protein and characterization of the gene. *Biochem J* 278 (Pt 1), 269-278.

Flouriot G, Brand H, Denger S, Metivier R, Kos M, Reid G, Sonntag-Buck V, Gannon F (2000) Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1. *Embo J* 19, 4688-4700.

Gao W, Bohl CE, Dalton JT (2005) Chemistry and structural biology of androgen receptor. *Chem Rev* 105, 3352-3370.

Gelmann EP (2002) Molecular biology of the androgen receptor. *J Clin Oncol* 20, 3001-3015.

He B, Lee LW, Minges JT, Wilson EM (2002) Dependence of selective gene activation on the androgen receptor NH2- and COOH-terminal interaction. *J Biol Chem* 277, 25631-25639.

Heemers HV, Tindall DJ (2007) Androgen receptor (AR) coregulators: a diversity of functions converging on and regulating the AR transcriptional complex. *Endocr Rev* 28, 778-808.

Heinlein CA, Chang C (2002) The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions. *Mol Endocrinol* 16, 2181-2187.

Hiipakka RA, Liao S (1998) Molecular mechanism of androgen action. *Trends Endocrinol Metab* 9, 317-324.

Hirata S, Shoda T, Kato J, Hoshi K (2003) Isoform/variant mRNAs for sex steroid hormone receptors in humans. *Trends Endocrinol Metab* 14, 124-129.

Hollenberg SM, Giguere V, Segui P, Evans RM (1987) Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. *Cell* 49, 39-46.

Huang X, Li J, Lu L, Xu M, Xiao J, Yin L, Zhu H, Zhou Z, Sha J (2005) Novel development-related alternative splices in human testis identified by cDNA microarrays. *J Androl* 26, 189-196.

Jagla M, Feve M, Kessler P, Lapouge G, Erdmann E, Serra S, Bergerat JP, Ceraline J (2007) A splicing variant of the androgen receptor detected in a metastatic prostate cancer exhibits exclusively cytoplasmic actions. *Endocrinology* 148, 4334-4343.

Johansen KL (2004) Testosterone metabolism and replacement therapy in patients with end-stage renal disease. *Semin Dial* 17, 202-208.

Kan Z, Garrett-Engele PW, Johnson JM, Castle JC (2005) Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. *Nucleic Acids Res* 33, 5659-5666.

Kim E, Magen A, Ast G (2007) Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* 35, 125-131.

Lareau LF, Green RE, Bhatnagar RS, Brenner SE (2004) The evolving roles of alternative splicing. *Curr Opin Struct Biol* 14, 273-282.

Laurentino S, Socorro S, Cavaco J (2006, results not published) Study of the Expression of Androgen Receptor, Estrogen Receptors α and β and α-Inhibin and Localization of Androgen Receptor in Testicular Tissue of Infertile Men. *not published results*.

Lee DK, Chang C (2003) Molecular communication between androgen receptor and general transcription machinery. *J Steroid Biochem Mol Biol* 84, 41-49.

Losel R, Wehling M (2003) Nongenomic actions of steroid hormones. *Nat Rev Mol Cell Biol* 4, 46-56.

Lubahn DB, Joseph DR, Sullivan PM, Willard HF, French FS, Wilson EM (1988) Cloning of human androgen receptor complementary DNA and localization to the X chromosome. *Science* 240, 327-330.

Mangelsdorf DJ, Thummel C*, et al.* (1995) The nuclear receptor superfamily: the second decade. *Cell* 83, 835-839.

Marden JH (2006) Quantitative and evolutionary biology of alternative splicing: how changing the mix of alternative transcripts affects phenotypic plasticity and reaction norms. *Heredity* 100, 111-120.

Matias PM, Donner P*, et al.* (2000) Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J Biol Chem* 275, 26164-26171.

Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6, 386-398.

McEwan IJ (2004) Molecular mechanisms of androgen receptor-mediated gene regulation: structure-function analysis of the AF-1 domain. *Endocr Relat Cancer* 11, 281-293.

Mulac-Jericevic B, Lydon JP, DeMayo FJ, Conneely OM (2003) Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. *Proc Natl Acad Sci U S A* 100, 9744-9749.

Oettel M, Hubler D, Patchev V (2003) Selected aspects of endocrine pharmacology of the aging male. *Exp Gerontol* 38, 189-198.

Ogawa S, Inoue S, Watanabe T, Orimo A, Hosoi T, Ouchi Y, Muramatsu M (1998) Molecular cloning and characterization of human estrogen receptor betacx: a potential inhibitor ofestrogen action in human. *Nucleic Acids Res* 26, 3505-3512.

Sarropoulou E, Nousdili D, Magoulas A, Kotoulas G (2008) Linking the genomes of nonmodel teleosts through comparative genomics. *Mar Biotechnol (NY)* 10, 227-233.

Shaffer PL, Jivan A, Dollins DE, Claessens F, Gewirth DT (2004) Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A* 101, 4758-4763.

Shaulian E, Karin M (2002) AP-1 as a regulator of cell life and death. *Nat Cell Biol* 4, E131-136.

Singh SM, Gauthier S, Labrie F (2000) Androgen receptor antagonists (antiandrogens): structure-activity relationships. *Curr Med Chem* 7, 211-247.

Srebrow A, Kornblihtt AR (2006) The connection between splicing and cancer. *J Cell Sci* 119, 2635-2641.

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. *Gene* 344, 1-20.

Stunnenberg HG (1993) Mechanisms of transactivation by retinoic acid receptors. *Bioessays* 15, 309-315.

Suh J, Rabson AB (2004) NF-kappaB activation in human prostate cancer: important mediator or epiphenomenon? *J Cell Biochem* 91, 100-117.

Thompson J (2006) Molecular Mechanisms of Androgen Receptor Interactions. *Institute of Biomedicine/Physiology, University of Helsinki, Finland*.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680.

Thornton JW, Kelley DB (1998) Evolution of the androgen receptor: structure-function implications. *Bioessays* 20, 860-869.

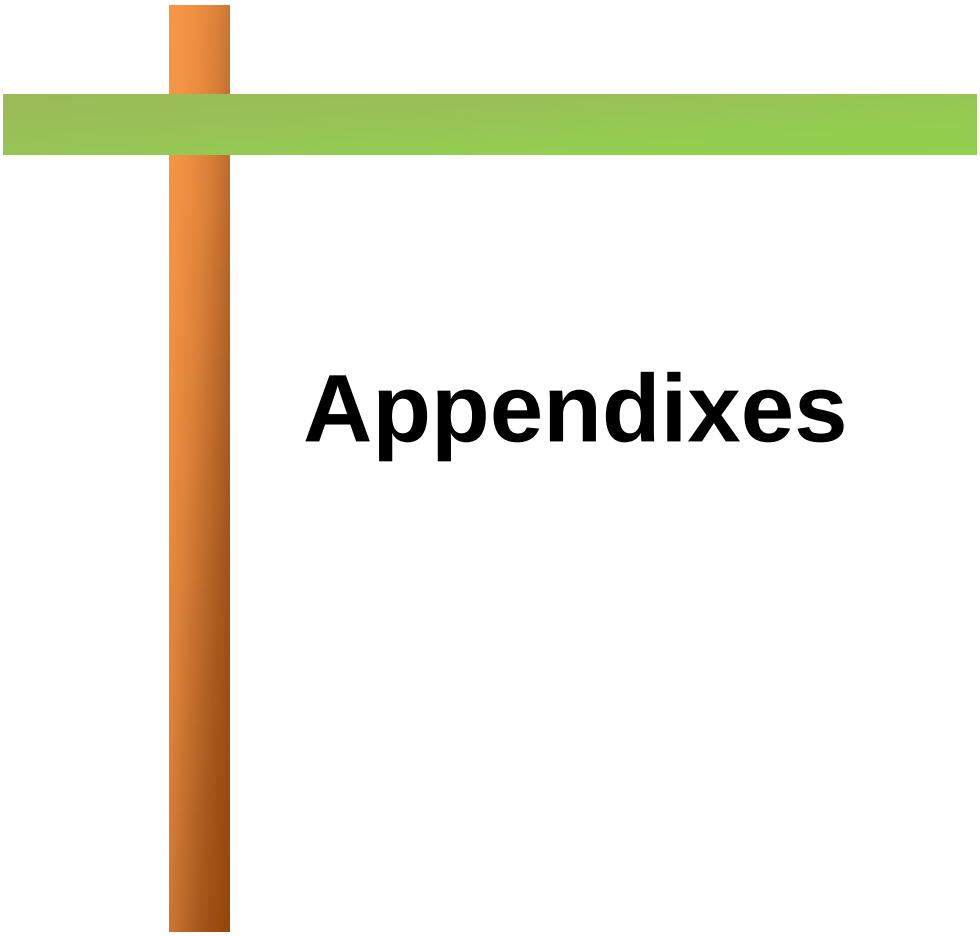
Venables JP (2006) Unbalanced alternative splicing and its significance in cancer. *Bioessays* 28, 378-386.

Yudt MR, Jewell CM, Bienstock RJ, Cidlowski JA (2003) Molecular origins for the dominant negative function of human glucocorticoid receptor beta. *Mol Cell Biol* **23**, 4319-4330.

Zennaro MC, Souque A, Viengchareun S, Poisson E, Lombes M (2001) A new human MR splice variant is a ligand-independent transactivator modulating corticosteroid action. *Mol Endocrinol* 15, 1586-1598.

Zhou ZX, Sar M, Simental JA, Lane MV, Wilson EM (1994) A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor. Requirement for the DNA-binding domain and modulation by NH2-terminal and carboxyl-terminal sequences. *J Biol Chem* 269, 13115-13123.

Zhu X, Daffada AA, Chan CM, Dowsett M (1997) Identification of an exon 3 deletion splice variant androgen receptor mRNA in human breast cancer. *Int J Cancer* 72, 574-580.



Appendixes

APPENDIX A – Comercial RNA samples from Clontech

A.1

PRODUCT: Human Heart Total RNA
CATALOG No. 636532
LOT NUMBER: 7030456
CONCENTRATION: 1 µg/µl
FORM: Suspension of total RNA in DEPC-treated water

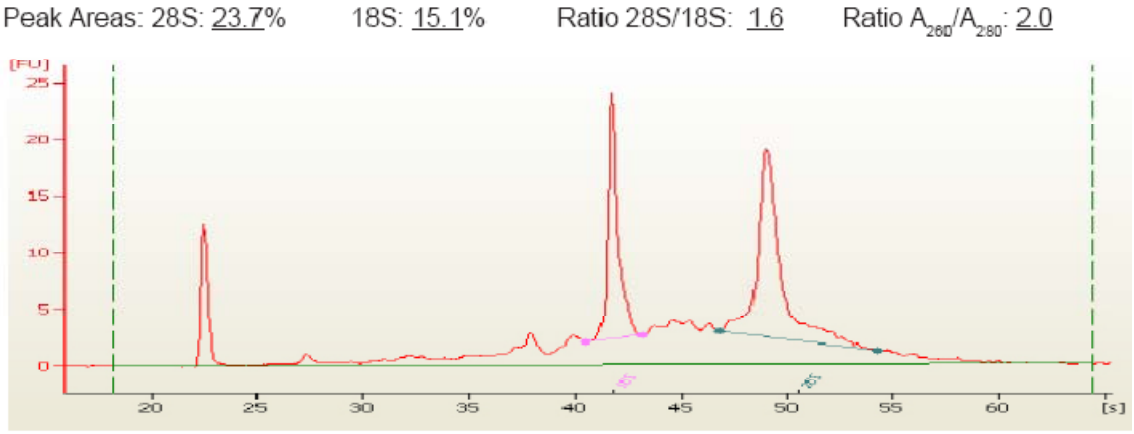


Figure A.1 – Total RNA analyzed by capillary electrophoresis (CE) using an Agilent 2100 Bioanalyzer.

A.2

PRODUCT: Human Kidney Total RNA
CATALOG No. 636529
LOT NUMBER: 7030175
CONCENTRATION: 1 µg/µl
FORM: Suspension of total RNA in DEPC-treated water

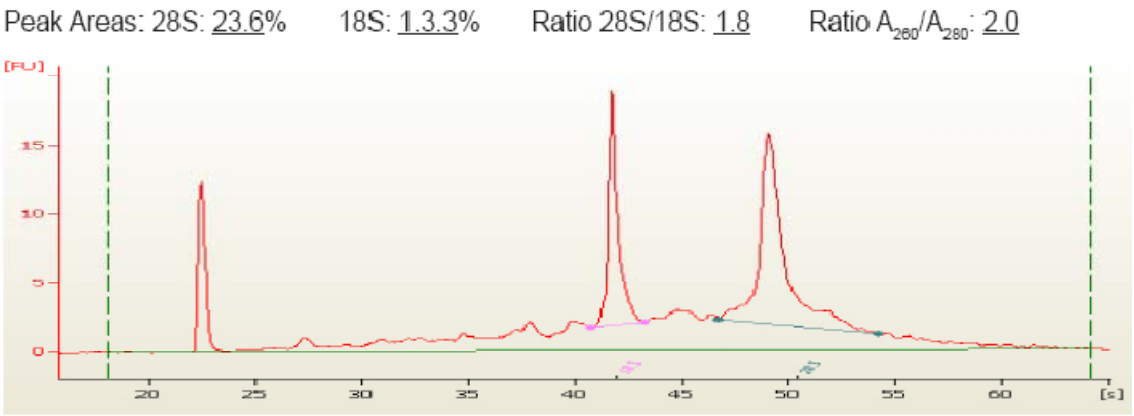


Figure A.2 – Total RNA analyzed by capillary electrophoresis (CE) using an Agilent 2100 Bioanalyzer.

A.3

PRODUCT: Human Lung Total RNA
CATALOG No. 636524
LOT NUMBER: 6120233
CONCENTRATION: 1 µg/µl
FORM: Suspension of total RNA in DEPC-treatedwater

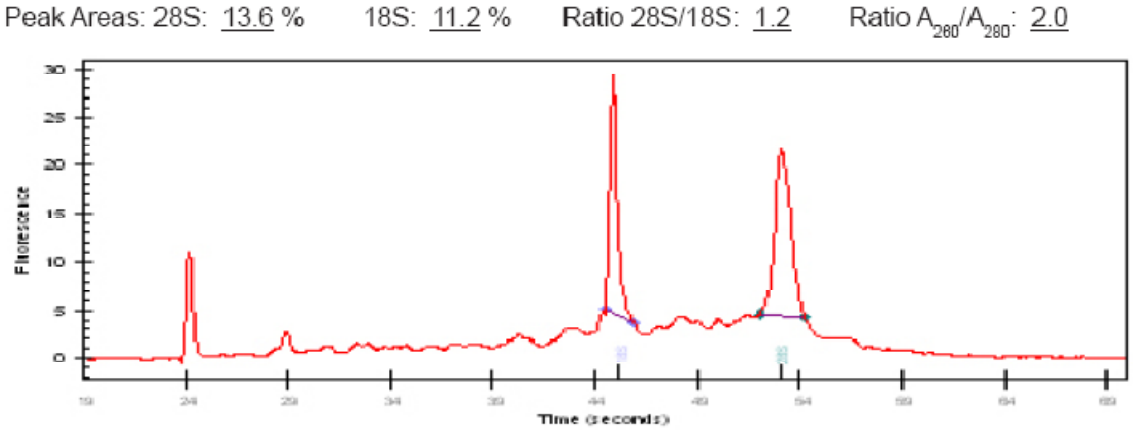


Figure A.3 – Total RNA analyzed by capillary electrophoresis (CE) using an Agilent 2100 Bioanalyzer.

A.4

PRODUCT: Human Liver Total RNA
CATALOG No. 636531
LOT NUMBER: 7070189
CONCENTRATION: 1 µg/µl
FORM: Suspension of total RNA in DEPC-treatedwater

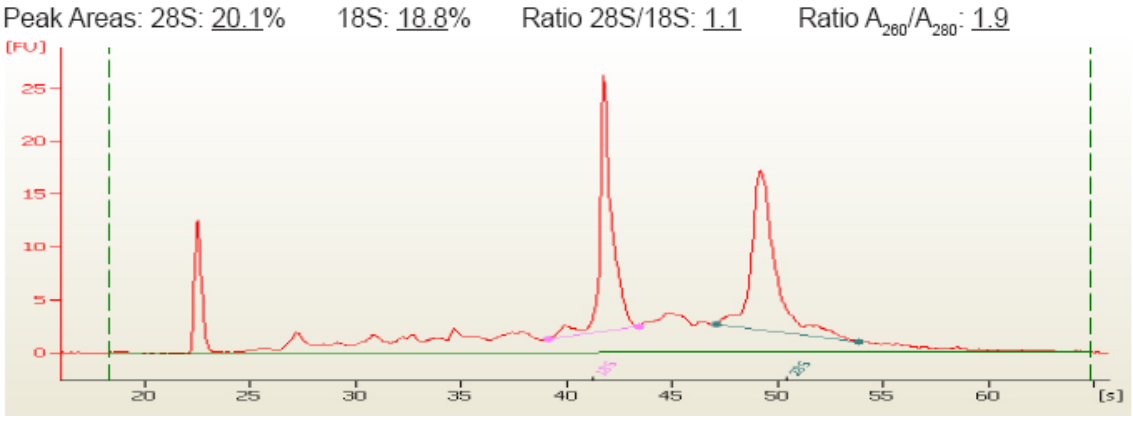


Figure A.4 – Total RNA analyzed by capillary electrophoresis (CE) using an Agilent 2100 Bioanalyzer.

APPENDIX B – RNA extraction with TRI Reagent® RNA Isolation Reagent, Sigma-Aldrich Protocol

1. Verify if the centrifuge is at 4ºC
2. Weight 50-100mg of tissue in a 2mL eppendorf and put on ice
3. Add 1mL of TRI reagent
4. Homogenise the tissue in Ultra-Turrax T25 Basic, IKA – Werke
5. Incubate 5min at room temperature
6. Add 200µL of chloroform and invert the tubes until everything is mixed
7. Incubate 510in at room temperatute
8. Spin 15min at 4°C at 12000xg
9. Transfer supernatant to a new tube
10. Add 500µL of isopropanol (500 µL of isopropanol/1mL of TRI), invert the tubes to mix
11. Incubate 10min at room temperature
12. Spin 10min at 4°C at 12000xg
13. Pipette the supernatant and recover the pellet – CAUTION
14. Wash the pellet whit 500µL of 75% EtOH at -20ºC
15. Spin 5min at 4°C at 7500xg
16. Repeat step 13, 14 1nd 15
17. Remove the supernatant, remove the last drop ph EtOH – CAREFULLY
18. Leave to dry (it is dry when the pellet lose its white colour)
19. Reconstitute the pellet with DEPC water (10, 20, 50, 80 or 100µL)
20. Store RNA at -80ºC

APPENDIX C – cDNA AR sequences, exons, domains and primers

C.1. *Cannis lupus*

```
1     atg gag gta cag tta ggg cta ggg agg gtc tac ccc cgg ccg ccg tcc aag acc tat cga gga gct ttc    69
1      M   E   V   Q   L   G   L   G   R   V   Y   P   R   P   P   S   K   T   Y   R   G   A   F    23

70    cag aac ctg ttc cag agt gtg cgc gaa gtg atc cag aac ccg ggc ccc agg cac cct gag gcc gtg agc   138
24     Q   N   L   F   Q   S   V   R   E   V   I   Q   N   P   G   P   R   H   P   E   A   V   S    46

136   gca gca cct ccc ggt gcc cat ttg cag cag cag cag cag cag cag cag cag cag cag gag acc agt cct cgg   207
47     A   A   P   P   G   A   H   L   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   E   T   S   P   R    69

208   cag cag cag cag caa cag cag ggt gac gat ggc tct ccc caa gcg cag agc aga ggc ccc aca ggc tac   276
70     Q   Q   Q   Q   Q   Q   Q   G   D   D   G   S   P   Q   A   Q   S   R   G   P   T   G   Y    92

275   ctg gct ctg gat gag gaa cag cag cct tcc caa cag cgg tca gcc tcc aag ggc cat ccg gag agt gcc   345
93     L   A   L   D   E   E   Q   Q   P   S   Q   Q   R   S   A   S   K   G   H   P   E   S   A   115

346   tgc gtt cca gag cct gga gtg act tcg gcc acc ggc aag ggg ctg cag cag cag cag cca gca cca ccg   414
116    C   V   P   E   P   G   V   T   S   A   T   G   K   G   L   Q   Q   Q   Q   P   A   P   P   138

415   gac gag aat gac tca gct gcc cca tcc aca ttg tca ctg ctg ggc ccc act ttc ccg ggc tta agt agc   483
139    D   E   N   D   S   A   A   P   S   T   L   S   L   L   G   P   T   F   P   G   L   S   S   161

484   tgt tcc acc gat ctt aaa gac atc ctg agc gag gct gga acc atg caa ctc ctt cag cag cag cgg cag   552
162    C   S   T   D   L   K   D   I   L   S   E   A   G   T   M   Q   L   L   Q   Q   Q   R   Q   184

553   cag cag cag cag cag cag caa cag caa cag cag cag cag cag cag cag caa cag cag gag gta gta tca gaa   621
185    Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   E   V   V   S   E   207

622   ggt agc agc agc ggg aga gca agg gag gcc gcc ggt gct tcc acc tcc tcc aag gac agt tac cta ggt   690
208    G   S   S   S   G   R   A   R   E   A   A   G   A   S   T   S   S   K   D   S   Y   L   G   230

691   ggc agt tcg acc atc tcg gac agc gcc aag gag ttg tgt aag gca gtg tcg gtg tcc atg ggt ttg ggt   759
231    G   S   S   T   I   S   D   S   A   K   E   L   C   K   A   V   S   V   S   M   G   L   G   253

760   gtg gag gca ttg gaa cat ctg agc cct ggg gaa cag ctt cgg ggg gat tgt atg tac gcc ccg ctc ctg   828
254    V   E   A   L   E   H   L   S   P   G   E   Q   L   R   G   D   C   M   Y   A   P   L   L   276

829   gga ggt cca ccc gcc gta cgt cct tgc gct ccg ctg gcc gaa tgc aaa ggt tct ctg ctg gat gac ggc   897
277    G   G   P   P   A   V   R   P   C   A   P   L   A   E   C   K   G   S   L   L   D   D   G   299

898   ccg ggc aag ggc acc gaa gaa act gct gaa tat tcc cct ttc aag gca ggt tat gcg aaa ggg ttg gat   966
300    P   G   K   G   T   E   E   T   A   E   Y   S   P   F   K   A   G   Y   A   K   G   L   D   322

967   ggg gac agc ctg ggc tgt tcg agc agc agt gaa gca ggg ggc tcc gga aca ctt gag atg cca tcc acc  1035
323    G   D   S   L   G   C   S   S   S   S   E   A   G   G   S   G   T   L   E   M   P   S   T   345

1036  ctg tct ctt tac aag tct gga gca cta gat gaa gcg gca gct tat caa agt cga gac tac tac aac ttt  1104
346    L   S   L   Y   K   S   G   A   L   D   E   A   A   A   Y   Q   S   R   D   Y   Y   N   F   368

1105  cct ctc tcc cta ggc ggg ccg cct ccc cat cca cca cct ccc cat cct cac acc cgc atc aag ctg gaa  1173
369    P   L   S   L   G   G   P   P   P   H   P   P   P   P   H   P   H   T   R   I   K   L   E   391

1174  aac cct ctg gac tat ggc agc gcc tgg gcg gct gca gcg gca caa tgc cgc tac ggg gat ctg gcg agc  1242
392    N   P   L   D   Y   G   S   A   W   A   A   A   A   A   Q   C   R   Y   G   D   L   A   S   414

1243  ctg cac gga gcg ggt gca gca gga ccc agc tcg ggc tca cct tcg gcc acc acc tcc tct tcc tgg cac  1311
415    L   H   G   A   G   A   A   G   P   S   S   G   S   P   S   A   T   T   S   S   S   W   H   437

1312  act ctc ttc aca gca gaa gaa ggc cag ctg tat ggg ccc tgc ggc ggg agt gga ggc ggc agt gca ggc  1380
461    T   L   F   T   A   E   E   G   Q   L   Y   G   P   C   G   G   S   G   G   G   S   A   G   460

1381  gac ggg gga tct gta gcc ccc tat ggc tac act cgg cca cct cag gga ttg gcg ggt cag gaa ggt gac  1449
461    D   G   G   S   V   A   P   Y   G   Y   T   R   P   P   Q   G   L   A   G   Q   E   G   D   483

1450  ttc cct cca cct gat gtg tgg tat ccg ggc ggt gtg gtg agc aga gtg ccc ttt cca agt cct agt tgt  1518
484    F   P   P   P   D   V   W   Y   P   G   G   V   V   S   R   V   P   F   P   S   P   S   C   506
```

```
1519  gtc aaa agc gag atg ggc tct tgg atg gag agc tac tcc gga ccc tat ggg gac atg*cgt ttg gag act  1587
507    V   K   S   E   M   M   G   S   W   M   E   S   Y   S   G   P   Y   G   D   R   L   E   T   529

1588  gcc agg gac cat gtt cta ccc att gac tat tac ttt cca cct cag aag acc tgt ctg atc tgc ggt gat  1656
530    A   R   D   H   V   L   P   I   D   Y   Y   F   P   P   Q   K   T   C   L   I   C   G   D   552

1657  gaa gct tct ggc tgt cac tat gga gct ctc act tgt gga agc tgc aaa gtc ttc ttt aaa aga gcc gct  1725
553    E   A   S   G   C   H   Y   G   A   L   T   C   G   S   C   K   V   F   F   K   R   A   A   575

1726 *gaa ggg aaa cag aag tac ctg tgt gcc agc aga aat gat tgt acc atc gat aaa ttc cga agg aaa aat  1794
576    E   G   K   Q   K   Y   L   C   A   S   R   N   D   C   T   I   D   K   F   R   R   K   N   598

1795  tgt cca tct tgt cgc ctc cgg aaa tgc tat gaa gca ggg atg act ctg gga* gcc cgg aag cta aag aaa  1863
599    C   P   S   C   R   L   R   K   C   Y   E   A   G   M   T   L   G   A   R   K   L   K   K   621

1864  ctg ggg aat ctg aaa ctg caa gag gaa gga gag gct tcc aat gtc acc agc ccc act gag gag cca acc  1932
622    L   G   N   L   K   L   Q   E   E   G   E   A   S   N   V   T   S   P   T   E   E   P   T   644

1933  cag aag ctg acg gtg tca cac att gaa*ggc tat gag tgt cag ccc atc ttt ctg aat gtc ctt gaa gcc  2001
645    Q   K   L   T   V   S   H   I   E   G   Y   E   C   Q   P   I   F   L   N   V   L   E   A   667

2002  atc gag cca ggc gtg gtg tgt gct gga cat gac aac aac cag ccc gac tcc ttt gca gcc ttg ctc tct  2070
668    I   E   P   G   V   V   C   A   G   H   D   N   N   Q   P   D   S   F   A   A   L   L   S   690

2071  agc ctt aat gaa ttg ggt gaa agg cag ctt gta cat gtg gtc aag tgg gcc aag gcc ttg ccg ggc ttc  2139
691    S   L   N   E   L   G   E   R   Q   L   V   H   V   V   K   W   A   K   A   L   P   G   F   713

2116  cgc aac ctg cac gtg gat gac cag atg gca gtc att cag tac tcc tgg atg ggg ctc atg gtg ttt gcc  2208
714    R   N   L   H   V   D   D   Q   M   A   V   I   Q   Y   S   W   M   G   L   M   V   F   A   736

2206  atg ggc tgg cga tcc ttc acc aat gtc aac tcc agg atg ctc tac ttc gcc cct gac ctg* gtt ttc aat  2277
737    M   G   W   R   S   F   T   N   V   N   S   R   M   L   F   A   P   D   L   V   F   N   E   759

2251  gag tac cgc atg cac aag tcc cgg atg tac agc cag tgt gtc cga atg agg cac ctc tct caa gaa ttt  2346
760    Y   Y   R   M   H   K   S   R   M   Y   S   Q   C   V   R   M   R   H   L   S   Q   E   F   782

2341  gga tgg ctc caa atc acc ccg cag gaa ttt ttg tgc atg aag gcg ctg ctg cta ttc agc att* att cca  2415
783    G   W   L   Q   I   T   P   Q   E   F   L   C   M   K   A   L   L   L   F   S   I   I   P   805

2386  gtg gat ggg ctg aaa aat caa aaa ttc ttt gat gaa ctt cga atg aac tac atc aag gaa ctt gat cgt  2484
806    V   D   G   L   K   N   Q   K   F   F   D   E   L   R   M   N   Y   I   K   E   L   D   R   828

2476  atc att gct tgc aag aga aaa aat ccc aca tcc tgc tca agg cgc ttc tac cag ctc acc aag ctc ctg  2553
829    I   I   A   C   K   R   K   N   P   T   S   C   S   R   R   F   Y   Q   L   T   K   L   L   851

2521  gac tct gtg caa cct* att gct cga gag ctg cat cag ttc act ttt gac ctg cta atc aag tcc cac atg  2622
852    D   S   V   Q   P   I   A   R   E   L   H   Q   F   T   F   D   L   L   I   K   S   H   M   874

2611  gtg agc gtg gac ttt cca gaa atg atg gca gaa atc atc tcc gtg caa gtg ccc aag att ctt tct ggg  2691
875    V   S   V   D   F   P   E   M   M   A   E   I   I   S   V   Q   V   P   K   I   L   S   G   897

2656  aaa gtc aag ccc atc tat ttc cac acg cag tga agc ttt aga agc ccc tgt ctt                      2745
898    K   V   K   P   I   Y   F   H   T   Q                                                    907
```

Figure C.1: *Canis lupus* androgen receptor cDNA sequence (GenBank accession no. NM_001003053). Protein specific domains, namely, NTD, DBD, Hinge region and LBD, are indicated alternated white and shadowed areas. Exon boundaries are indicated by *. The sequence amplified by the primer set ClAR1/5 is in bold. Primer sequences are underlined.

C.2. *Gallus gallus*

```
601   ATG GAG GTG CAG CTG GGG ATC GGG CGC GTG TAC CCA CGG CCG CCG GGA CGG ACC TTC CGC GGC GTC TTC   669
1      M   E   V   Q   L   G   I   G   R   V   Y   P   R   P   P   G   R   T   F   R   G   V   F    23

670   CAG ACG TTC TTC CAG AGC GTG TGC GAA GCG TTC CAG GCG CCG CGG GAC GAG CCG GGC CCC GGC CTG CCC   738
24     Q   T   F   F   Q   S   V   C   E   A   F   Q   A   P   R   D   E   P   G   P   G   L   P    46

739   GCA CCG GGA GCC CCG TGC CCG CAG AGT CCT CGC CCG CCG CCC GTC GCC TCG CCC GCG TTC CTG CCG CTG   807
47     A   P   G   A   P   C   P   Q   S   P   R   P   P   P   V   A   S   P   A   F   L   P   L    69
```

```
808   CCC GAG CCC CGC GCC GCC GCC CGC CCG GCC ATG GGC TCG CCC TTC CCG TGC GCC GGG GAC CTG AAG GAG   876
70     P   E   P   R   A   A   A   R   P   A   M   G   S   P   F   P   C   A   G   D   L   K   E    92

877   CTG CTG GGC GAG CCG GGC GTG CTG CCG CTG CTG CCG CCC GAG GCC GAG CCG GGG GCC GGG CGA GCC GAG   945
93     L   L   G   E   P   G   V   L   P   L   L   P   P   E   A   E   P   G   A   G   R   A   E   115

946   CCG GCG CTC AAG GAA GAC TTT CTG GGC GAC AGC GCC AAG GAG CTC TGC AAA GCC GTG TCC GCC TCC ATG  1014
116    P   A   L   K   E   D   F   L   G   D   S   A   K   E   L   C   K   A   V   S   A   S   M   138

1015  GGG CTG GCG GTC GAG ACT CTG GAG GCG CCC AGA GAG CCG CCG CCC CGC GAG GAC TGC ATG TTC GCG CTG  1083
139    G   L   A   V   E   T   L   E   A   P   R   E   P   P   P   R   E   D   C   M   F   A   L   161

1084  CCC GGA GGG CCG CCC CGC GCC CCG CGC CCC GAC GCC GCC GAA CCA CCG GAG CCG CCC ACA CCG GCC GCC  1152
162    P   G   G   P   P   R   A   P   R   P   D   A   A   E   P   P   E   P   P   T   P   A   A   184

1153  TTC AAG GGC AGC GGA GCC GAG GCC GCC CTG GCC GTC GAG GTG CCC GCG GGG CTG CCG CTG TAC CGC GTC  1221
185    F   K   G   S   G   A   E   A   A   L   A   V   E   V   P   A   G   L   P   L   Y   R   V   207

1222  CCG TCC CCG CCC GAG GAG CCG CCC GGC CGC GAC TGC TTC GTT CTC CCG CCG CCC GCC CGC ATC AAG CTG  1290
208    S   P   P   P   E   E   P   P   G   R   D   C   F   V   L   P   P   P   A   R   I   K   L   230

1291  GAG AGC CCC CCC GAG CCC GCG GCA GTC GGC GCC TGG GGC TCG CCG GTC CCC GCG CCG CCG TGG CCC TCC  1359
231    E   S   P   P   E   P   A   A   V   G   A   W   G   S   P   V   P   A   P   P   W   P   S   253

1360  TTC TTC GCC GAC GAG GGG CAG CTG TAC GGA CCC TGC GCC GAG CCG CCC CCC GGC GCC TTC GGC TGC GGC  1428
254    F   F   A   D   E   G   Q   L   Y   G   P   C   A   E   P   P   P   G   A   F   G   C   G   276

1429  CGC CCG GAG AAC GCC GAC TTC GCC GCC GAC GCG TGG CAC CCG ATG GCG CGG GCG CCC TAC GCC GCG CCC  1497
277    R   P   E   N   A   D   F   A   A   D   A   W   H   P   M   A   R   A   P   Y   A   A   P   299

1498  GGC TCC TGC ATC AAG AGC GAG CTG GGA CCC TGG GCA GAG GGC TAC GCC GGA GCC TAC GGC GAC GTC CGG  1566
300    G   S   C   I   K   S   E   L   G   P   W   A   E   G   Y   A   G   A   Y   G   D   V   R   322

1567  CTG GAG GCT GGC CGG GAG CAC ATC CTG CCC ATT GAC TAT TAC TTT CCA CCT CAG AAG ACG TGT CTG ATC  1635
323    L   E   A   G   R   E   H   I   L   P   I   D   Y   Y   F   P   P   Q   K   T   C   L   I   345

1636  TGC GGT GAT GAG GCA TCC GGG TGC CAC TAT GGA GCC CTC ACC TGC GGG AGC TGC AAA GTG TTC TTC AAG  1704
346    C   G   D   E   A   S   G   C   H   Y   G   A   L   T   C   G   S   C   K   V   F   F   K   368

1705  CGG GCG GCT GAA GGT AAG CAG AAG TAC CTG TGT GCC AGC CGA AAC GAC TGC ACC ATC GAC AAG TTC AGG  1773
369    R   A   A   E   G   K   Q   K   Y   L   C   A   S   R   N   D   C   T   I   D   K   F   R   391

1774  AGG AAA AAC TGC CCG TCC TGC CGC CTG CGG AAG TGC TAC GAG GCG GGG ATG ACT CTT GGA GCC CGC AAG  1842
392    R   K   N   C   P   S   C   R   L   R   K   C   Y   E   A   G   M   T   L   G   A   R   K   414

1843  CTG AAG AAG CTG GGC AGT CTG AAG ACA CAA GAT GAA GCT GAG GCA GCC AGC TCA TCC AGC CCC ACG GAG  1911
415    L   K   K   L   G   S   L   K   T   Q   D   E   A   E   A   A   S   S   S   S   P   T   E   437

1912  GAG CAA GCT CCT AAG ATG GTG ATG ACA CAC GTC AAT GGC TTT GAG TGC CAG CCC ATC TTT CTC AAT GTG  1980
438    E   Q   A   P   K   M   V   M   T   H   V   N   G   F   E   C   Q   P   I   F   L   N   V   460

1981  CTG GAA GCC ATC GAG CCC GCT GTG GTG TGT GCT GGC CAT GAC AAC AGC CAG CCA GAC TCC TTC TCC AAC  2049
461    L   E   A   I   E   P   A   V   V   C   A   G   H   D   N   S   Q   P   D   S   F   S   N   483

2050  CTG CTA AGC AGC CTG AAT GAA CTT GGG GAG AGA CAG CTG GTC TAC GTC GTC AAG TGG GCA AAG GCT TTG  2118
484    L   L   S   S   L   N   E   L   G   E   R   Q   L   V   Y   V   V   K   W   A   K   A   L   506

2119  CCA GGA TTT CGC AAT TTG CAT GTG GAT GAC CAG ATG TCA ATA ATC CAG TAT TCT TGG ATG GGT CTC ATG  2187
507    P   G   F   R   N   L   H   V   D   D   Q   M   S   I   I   Q   Y   S   W   M   G   L   M   529

2188  ATT TTT GCC ATG GGT TGG AGA TCT TTC ACT AAC GTC AAT TCC AGG ATG CTT TAC TTT GCC CCA GAT TTG  2256
530    I   F   A   M   G   W   R   S   F   T   N   V   N   S   R   M   L   Y   F   A   P   D   L   552

2257  GTC TTC AAC GAG TAC CGG ATG CAC AAA TCC AGA ATG TAC AGC CAA TGC GTC AGG ATG CGG CAG CTC TCC  2325
553    V   F   N   E   Y   R   M   H   K   S   R   M   Y   S   Q   C   V   R   M   R   Q   L   S   575

2326  CAG GAG TTT GGG TGG CTT CAG ATC ACG CCC CAG GAG TTT CTC TGC ATG AAG GCT CTG CTC TTC TTC AGT  2394
576    Q   E   F   G   W   L   Q   I   T   P   Q   E   F   L   C   M   K   A   L   L   F   F   S   598

2395  ATT ATT CCA GTG GAT GGC CTG AAG AAC CAG AAG CTC TTT GAT GAA CTT CGC ATG AAC TAC ATT AAG GAA  2463
599    I   I   P   V   D   G   L   K   N   Q   K   L   F   D   E   L   R   M   N   Y   I   K   E   621

2464  CTC GAT CGG ATC ATT GCT TGC AAG AGG AAG AAC CCC ACC TCC TGC TCT AGG CGG TTT TAC CAG CTC ACC  2532
622    L   D   R   I   I   A   C   K   R   K   N   P   T   S   C   S   R   R   F   Y   Q   L   T   644

2533  AAG GTC CTG GAC TCC GTG CAT CCT ATT GCC AAG GAT CTG CAC CAG TTT ACA TTT GAT CTT TTA ATT AAG  2601
645    K   V   L   D   S   V   H   P   I   A   K   D   L   H   Q   F   T   F   D   L   L   I   K   667
```

```
2602  GCC CAC ATG GTG AGC GTG GAC TAC CCG GAA ATG ATG GCC GAG ATC ATC TCT GTG CAG GTT CCC AAG ATC  2670
668    A   H   M   V   S   V   D   Y   P   E   M   M   A   E   I   I   S   V   Q   V   P   K   I    690

2671 CTG TCT GGA AAA GTG AAA CCC ATT TAC TTC CAC GCG GAG TGA   2684
691    L   S   G   K   V   K   P   I   Y   F   H   A   E   *
```

Figure C.2: *Gallus gallus* androgen receptor cDNA sequence (GenBank accession no. NM_001040090). Protein specific domains, namely, NTD, DBD, Hinge region and LBD, are indicated alternated white and shadowed areas. Exon boundaries are indicated by *. The sequence amplified by the primer set GgAR1/5 is in bold. Primer sequences are underlined..

C.3. *Homo sapiens*

```
1116  atg gaa gtg cag tta ggg ctg gga agg gtc tac cct cgg ccg ccg tcc aag acc tac cga gga gct ttc  1184
1      M   E   V   Q   L   G   L   G   R   V   Y   P   R   P   P   S   K   T   Y   R   G   A   F     23

1185  cag aat ctg ttc cag agc gtg cgc gaa gtg atc cag aac ccg ggc ccc agg cac cca gag gcc gcg agc  1253
24     Q   N   L   F   Q   S   V   R   E   V   I   Q   N   P   G   P   R   H   P   E   A   A   S     46

1254  gca gca cct ccc ggc gcc agt ttg ctg ctg ctg cag cag cag cag cag cag cag cag cag cag cag cag  1322
47     A   A   P   P   G   A   S   L   L   L   L   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q     69

1323  cag cag cag cag cag cag cag cag cag cag cag caa gag act agc ccc agg cag cag cag cag cag cag ggt  1391
70     Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   E   T   S   P   R   Q   Q   Q   Q   Q   Q   G     92

1392  gag gat ggt tct ccc caa gcc cat cgt aga ggc ccc aca ggc tac ctg gtc ctg gat gag gaa cag caa  1460
93     E   D   G   S   P   Q   A   H   R   R   G   P   T   G   Y   L   V   L   D   E   E   Q   Q    115

1461  cct tca cag ccg cag tcg gcc ctg gag tgc cac ccc gag aga ggt tgc gtc cca gag cct gga gcc gcc  1529
116    P   S   Q   P   Q   S   A   L   E   C   H   P   E   R   G   C   V   P   E   P   G   A   A    138

1530  gtg gcc gcc agc aag ggg ctg ccg cag cag ctg cca gca cct ccg gac gag gat gac tca gct gcc cca  1598
139    V   A   A   S   K   G   L   P   Q   Q   L   P   A   P   P   D   E   D   D   S   A   A   P    161

1599  tcc acg ttg tcc ctg ctg ggc ccc act ttc ccc ggc tta agc agc tgc tcc gct gac ctt aaa gac atc  1667
162    S   T   L   S   L   L   G   P   T   F   P   G   L   S   S   C   S   A   D   L   K   D   I    184

1668  ctg agc gag gcc agc acc atg caa ctc ctt cag caa cag cag cag gaa gca gta tcc gaa ggc agc agc  1736
185    L   S   E   A   S   T   M   Q   L   L   Q   Q   Q   Q   Q   E   A   V   S   E   G   S   S    207

1737  agc ggg aga gcg agg gag gcc tcg ggg gct ccc act tcc tcc aag gac aat tac tta ggg ggc act tcg  1805
208    S   G   R   A   R   E   A   S   G   A   P   T   S   S   K   D   N   Y   L   G   G   T   S    230

1806  acc att tct gac aac gcc aag gag ttg tgt aag gca gtg tcg gtg tcc atg ggc ctg ggt gtg gag gcg  1874
231    T   I   S   D   N   A   K   E   L   C   K   A   V   S   V   S   M   G   L   G   V   E   A    253

1875  ttg gag cat ctg agt cca ggg gaa cag ctt cgg ggg gat tgc atg tac gcc cca ctt ttg gga gtt cca  1943
254    L   E   H   L   S   P   G   E   Q   L   R   G   D   C   M   Y   A   P   L   L   G   V   P    276

1944  ccc gct gtg cgt ccc act cct tgt gcc cca ttg gcc gaa tgc aaa ggt tct ctg cta gac gac agc gca  2012
277    P   A   V   R   P   T   P   C   A   P   L   A   E   C   K   G   S   L   L   D   D   S   A    299

2013  ggc aag agc act gaa gat act gct gag tat tcc cct ttc aag gga ggt tac acc aaa ggg cta gaa ggc  2081
300    G   K   S   T   E   D   T   A   E   Y   S   P   F   K   G   G   Y   T   K   G   L   E   G    322

2082  gag agc cta ggc tgc tct ggc agc gct gca gca ggg agc tcc ggg aca ctt gaa ctg ccg tct acc ctg  2150
323    E   S   L   G   C   S   G   S   A   A   A   G   S   S   G   T   L   E   L   P   S   T   L    345

2151  tct ctc tac aag tcc gga gca ctg gac gag gca gct gcg tac cag agt cgc gac tac tac aac ttt cca  2219
346    S   L   Y   K   S   G   A   L   D   E   A   A   A   Y   Q   S   R   D   Y   Y   N   F   P    368

2220  ctg gct ctg gcc gga ccg ccg ccc cct ccg ccg cct ccc cat ccc cac gct cgc atc aag ctg gag aac  2288
369    L   A   L   A   G   P   P   P   P   P   P   P   P   H   P   H   A   R   I   K   L   E   N    391

2289  ccg ctg gac tac ggc agc gcc tgg gcg gct gcg gcg gcg cag tgc cgc tat ggg gac ctg gcg agc ctg  2357
390    P   L   D   Y   G   S   A   W   A   A   A   A   A   Q   C   R   Y   G   D   L   A   S   L    414

2358  cat ggc gcg ggt gca gcg gga ccc ggt tct ggg tca ccc tca gcc gcc gct tcc tca tcc tgg cac act  2426
415    H   G   A   G   A   A   G   P   G   S   G   S   P   S   A   A   A   S   S   S   W   H   T    437
```

```
2427  ctc ttc aca gcc gaa gaa ggc cag ttg tat gga ccg tgt ggt ggt ggt ggg ggt ggt ggc ggc ggc ggc  2495
438    L   F   T   A   E   E   G   Q   L   Y   G   P   C   G   G   G   G   G   G   G   G   G    460

2496  ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc gag gcg gga gct gta gcc ccc tac ggc tac  2564
461    G   G   G   G   G   G   G   G   G   G   G   G   G   E   A   G   A   V   A   P   Y   G   Y    483

2565  act cgg ccc cct cag ggg ctg gcg ggc cag gaa agc gac ttc acc gca cct gat gtg tgg tac cct ggc  2633
484    T   R   P   P   Q   G   L   A   G   Q   E   S   D   F   T   A   P   D   V   W   Y   P   G    506

2634 ggc atg gtg agc aga gtg ccc tat ccc agt ccc act tgt gtc aaa agc gaa atg ggc ccc tgg atg gat   2702
507   G   M   V   S   R   V   P   Y   P   S   P   T   C   V   K   S   E   M   G   P   W   M   D    529

2703  agc tac tcc gga cct tac ggg gac atg* cgt ttg gag act gcc agg gac cat gtt ttg ccc att gac tat  2771
530    S   Y   S   G   P   Y   G   D   M   R   L   E   T   A   R   D   H   V   L   P   I   D   Y    552

2772  tac ttt cca ccc cag aag acc tgc ctg atc tgt gga gat gaa gct tct ggg tgt cac tat gga gct ctc  2840
553    Y   F   P   P   Q   K   T   C   L   I   C   G   D   E   A   S   G   C   H   Y   G   A   L    575

2841  aca tgt gga agc tgc aag gtc ttc ttc aaa aga gcc gct gaa* ggg aaa cag aag tac ctg tgc gcc agc  2909
576    T   C   G   S   C   K   V   F   F   K   R   A   A   E   G   K   Q   K   Y   L   C   A   S    598

2910  aga aat gat tgc act att gat aaa ttc cga agg aaa aat tgt cca tct tgt cgt ctt cgg aaa tgt tat  2978
599    R   N   D   C   T   I   D   K   F   R   R   K   N   C   P   S   C   R   L   R   K   C   Y    621

2979  gaa gca ggg atg act ctg gga* gcc cgg aag ctg aag aaa ctt ggt aat ctg aaa cta cag gag gaa gga  3047
622    E   A   G   M   T   L   G   A   R   K   L   K   K   L   G   N   L   K   L   Q   E   E   G    644

3048  gag gct tcc agc acc acc agc ccc act gag gag aca acc cag aag ctg aca gtg tca cac att gaa ggc  3116
645    E   A   S   S   T   T   S   P   T   E   E   T   T   Q   K   L   T   V   S   H   I   E   G    667

3117  tat gaa tgt cag ccc atc ttt ctg aat gtc ctg gaa gcc att gag cca ggt gta gtg* tgt gct gga cac  3185
668    Y   E   C   Q   P   I   F   L   N   V   L   E   A   I   E   P   G   V   V   C   A   G   H    690

3186  gac aac aac cag ccc gac tcc ttt gca gcc ttg ctc tct agc ctc aat gaa ctg gga gag aga cag ctt  3254
691    D   N   N   Q   P   D   S   F   A   A   L   L   S   S   L   N   E   L   G   E   R   Q   L    713

3255  gta cac gtg gtc aag tgg gcc aag gcc ttg cct* ggc ttc cgc aac tta cac gtg gac gac cag atg gct  3323
714    V   H   V   V   K   W   A   K   A   L   P   G   F   R   N   L   H   V   D   D   Q   M   A    736

3324  gtc att cag tac tcc tgg atg ggg ctc atg gtg ttt gcc atg ggc tgg cga tcc ttc acc aat gtc aac  3392
737    V   I   Q   Y   S   W   M   G   L   M   V   F   A   M   G   W   R   S   F   T   N   V   N    759

3393  tcc agg atg ctc tac ttc gcc cct gat ctg gtt ttc aat* gag tac cgc atg cac aag tcc cgg atg tac  3461
760    S   R   M   L   Y   F   A   P   D   L   V   F   N   E   Y   R   M   H   K   S   R   M   Y    782

3462  agc cag tgt gtc cga atg agg cac ctc tct caa gag ttt gga tgg ctc caa atc acc ccc cag gaa ttc  3530
783    S   Q   C   V   R   M   R   H   L   S   Q   E   F   G   W   L   Q   I   T   P   Q   E   F    805

3531  ctg tgc atg aaa gca ctg cta ctc ttc agc att* att cca gtg gat ggg ctg aaa aat caa aaa ttc ttt  3599
806    L   C   M   K   A   L   L   L   F   S   I   I   P   V   D   G   L   K   N   Q   K   F   F    828

3591  gat gaa ctt cga atg aac tac atc aag gaa ctc gat cgt atc att gca tgc aaa aga aaa aat ccc aca  3668
829    D   E   L   R   M   N   Y   I   K   E   L   D   R   I   I   A   C   K   R   K   N   P   T    851

3669  tcc tgc tca aga cgc ttc tac cag ctc acc aag ctc ctg gac tcc gtg cag* cct att gcg aga gag ctg  3737
851    S   C   S   R   R   F   Y   Q   L   T   K   L   L   D   S   V   Q   P   I   A   R   E   L    874

3738  cat cag ttc act ttt gac ctg cta atc aag tca cac atg gtg agc gtg gac ttt ccg gaa atg atg gca  3806
875    H   Q   F   T   F   D   L   L   I   K   S   H   M   V   S   V   D   F   P   E   M   M   A    897

3807  gag atc atc tct gtg caa gtg ccc aag atc ctt tct ggg aaa gtc aag ccc atc tat ttc cac acc cag  3875
898    E   I   I   S   V   Q   V   P   K   I   L   S   G   K   V   K   P   I   Y   F   H   T   Q    920

3861  tga   3878
921    *
```

Figure C.3: *Homo sapiens* androgen receptor cDNA sequence (GenBank accession no. NM_000044). Protein specific domains, namely, NTD, DBD, Hinge region and LBD, are indicated alternated white and shadowed areas. Exon boundaries are indicated by *. The sequence amplified by the primer set hAR1/4 and hAR1/5 is in bold. Primer sequences are underlined.

C.4. *Rattus norvegicus*

```
994   atg gag gtg cag tta ggg ctg gga agg gtc tac cca cgg ccc ccg tcc aag acc tat cga gga gcg ttc  1062
1      M   E   V   Q   L   G   L   G   R   V   Y   P   R   P   P   S   K   T   Y   R   G   A   F     23

1063  cag aat ctg ttc cag agc gtg cgc gaa gcg atc cag aac ccg ggc ccc agg cac cct gag gcc gct agc  1131
24     Q   N   L   F   Q   S   V   R   E   A   I   Q   N   P   G   P   R   H   P   E   A   A   S     46

1132  ata gca cct ccc ggt gcc tgt tta cag cag cgg cag gag act agc ccc cgg cgg cgg cgg cgg cag cag  1200
47     I   A   P   P   G   A   C   L   Q   Q   R   Q   E   T   S   P   R   R   R   R   R   Q   Q     69

1201  cac cct gag gat ggc tct cct caa gcc cac atc aga ggc acc aca ggc tac ctg gcc ctg gag gag gaa  1269
70     H   P   E   D   G   S   P   Q   A   H   I   R   G   T   T   G   Y   L   A   L   E   E   E     92

1270  cag cag cct tca cag cag cag tca gcc tcc gag ggc cac cct gag agc ggc tgc ctc ccg gag cct gga  1338
93     Q   Q   P   S   Q   Q   Q   S   A   S   E   G   H   P   E   S   G   C   L   P   E   P   G    115

1339  gct gcc acg gct cct ggc aag ggg ctg ccg cag cag cca cca gct cct cca gat cag gat gac tca gct  1407
116    A   A   T   A   P   G   K   G   L   P   Q   Q   P   P   A   P   P   D   Q   D   D   S   A    138

1408  gcc cca tcc acg ttg tcc cta ctg ggc ccc act ttc cca ggc tta agc agc tgc tcc gca gac att aaa  1476
139    A   P   S   T   L   S   L   L   G   P   T   F   P   G   L   S   S   C   S   A   D   I   K    161

1477  gac atc ctg agc gag gcc ggc acc atg caa ctt ctt cag cag cag cag caa cag caa cag cag cag cag  1545
162    D   I   L   S   E   A   G   T   M   Q   L   L   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q    184

1546  cag cag cag cag cag cag cag caa cag cag cag gag gta ata tcc gaa ggc agc agc agc gtg aga gca  1614
185    Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   Q   E   V   I   S   E   G   S   S   S   V   R   A    207

1615  agg gag gcc act ggg gct ccc tct tcc tcc aag gat agt tac cta ggg ggc aat tcg acc ata tct gac  1683
208    R   E   A   T   G   A   P   S   S   S   K   D   S   Y   L   G   G   N   S   T   I   S   D    230

1664  agt gcc aag gag ttg tgt aaa gca gtg tct gtg tcc atg ggg ttg ggt gtg gaa gca ctg gaa cat ctg  1752
231    S   A   K   E   L   C   K   A   V   S   V   S   M   G   L   G   V   E   A   L   E   H   L    253

1753  agt cca ggg gag cag ctt cgg ggc gac tgc atg tac gcg tcg ctc ctg gga ggt cca ccc gcc gtg cgt  1821
254    S   P   G   E   Q   L   R   G   D   C   M   Y   A   S   L   L   G   G   P   P   A   V   R    276

1822  ccc act cct tgt gcg cct ctg gcc gaa tgc aaa ggt ctt tcc ctg gac gaa ggc ccg ggc aaa ggc act  1890
277    P   T   P   C   A   P   L   A   E   C   K   G   L   S   L   D   E   G   P   G   K   G   T    299

1891  gaa gag act gct gag tat tcc tct ttc aag gga ggt tac gcc aaa ggg ttg gaa ggt gag agt ctg ggc  1959
300    E   E   T   A   E   Y   S   S   F   K   G   G   Y   A   K   G   L   E   G   E   S   L   G    322

1960  tgc tct ggc agc agt gaa gca ggt agc tct ggg aca ctt gag atc ccg tcc tca ctg tct ctg tat aag  2028
323    C   S   G   S   S   E   A   G   S   S   G   T   L   E   I   P   S   S   L   S   L   Y   K    345

2029  tct gga gca gta gac gag gca gca gca tac cag aat cgc gac tac tac aac ttt ccg ctc gct ctg tcc  2097
346    S   G   A   V   D   E   A   A   A   Y   Q   N   R   D   Y   Y   N   F   P   L   A   L   S    368

2098  ggg ccg ccg cac ccc ccg ccc cct acc cat cca cac gcc cgc atc aag ctg gag aac ccg tcg gac tac  2166
369    G   P   P   H   P   P   P   P   T   H   P   H   A   R   I   K   L   E   N   P   S   D   Y    391

2167  ggc agc gcc tgg gct gcg gcg gca gcg caa tgc cgc tat ggg gac ttg gct agc cta cat gga ggg agt  2235
392    G   S   A   W   A   A   A   A   A   Q   C   R   Y   G   D   L   A   S   L   H   G   G   S    314

2236  gta gcc gga ccc agc act gga tcg ccc cca gcc acc gcc tct tct tcc tgg cat act ctc ttc aca gct  2304
415    V   A   G   P   S   T   G   S   P   P   A   T   A   S   S   S   W   H   T   L   F   T   A    437

2305  gaa gaa ggc caa tta tat ggg cca gga ggc ggg ggc ggc agc agt agc cca agc gat gct ggg cct gta  2373
438    E   E   G   Q   L   Y   G   P   G   G   G   G   G   S   S   S   P   S   D   A   G   P   V    460

2374  gcc ccc tat ggc tac act cgg ccc cct cag ggg ctg gca agc cag gag ggt gac ttc tct gcc tct gaa  2442
461    A   P   Y   G   Y   T   R   P   P   Q   G   L   A   S   Q   E   G   D   F   S   A   S   E    483

2443  gtg tgg tat cct ggt gga gtt gtg aac aga gtc ccc tat ccc agt ccc agt tgt gtt aaa agt gaa atg  2511
484    V   W   Y   P   G   G   V   V   N   R   V   P   Y   P   S   P   S   C   V   K   S   E   M    506

2512  gga cct tgg atg gag aac tac tcc gga cct tat ggg gac atg* cgt ttg gac agt acc agg gac cac gtt  2580
507    G   P   W   M   E   N   Y   S   G   P   Y   G   D   M   R   L   D   S   T   R   D   H   V    529

2581  tta ccc atc gac tat tac ttc cca ccc cag aag acc tgc ctg atc tgt gga gat gaa gct tct ggt tgt  2649
530    L   P   I   D   Y   Y   F   P   P   Q   K   T   C   L   I   C   G   D   E   A   S   G   C    552
```

```
2650  cac tac gga gct ctc act tgt ggc agc tgc aag gtc ttc ttc aaa aga gct gcg gaa* ggg aaa cag aag  2718
553     H   Y   G   A   L   T   C   G   S   C   K   V   F   F   K   R   A   A   E   G   K   Q   K   575

2719  tat cta tgt gcc agc aga aat gat tgc acc att gat aaa ttt cgg agg aaa aat tgt cca tcg tgt cgt  2787
576     Y   L   C   A   S   R   N   D   C   T   I   D   K   F   R   R   K   N   C   P   S   C   R   598

2788  ctc cgg aaa tgt tat gaa gca ggg atg act ctg gga* gct cgt aag ctg aag aaa ctt gga aat ctc aaa  2856
599     L   R   K   C   Y   E   A   G   M   T   L   G   A   R   K   L   K   K   L   G   N   L   K   621

2857  cta cag gaa gaa gga gaa aac tcc agt gct ggt agc ccc act gag gac cca tcc cag aag atg act gta  2925
622     L   Q   E   E   G   E   N   S   S   A   G   S   P   T   E   D   P   S   Q   K   M   T   V   644

2926  tca cac att gaa ggc tat gaa tgt caa cct atc ttt ctt aat gtc ctg gaa gcc att gag cca gga gtg  2994
645     S   H   I   E   G   Y   E   C   Q   P   I   F   L   N   V   L   E   A   I   E   P   G   V   667

2995  gtg tgt gcc gga cat gac aac aac cag cct gat tcc ttt gct gcc ttg tta tct agt ctc aac gag ctt  3063
668     V   C   A   G   H   D   N   N   Q   P   D   S   F   A   A   L   L   S   S   L   N   E   L   690

3064  ggc gag aga cag ctt gta cat gtg gtc aag tgg gcc aag gcc ttg cct* ggc ttc cgc aac ttg cat gtg  3132
691     G   E   R   Q   L   V   H   V   V   K   W   A   K   A   L   P   G   F   R   N   L   H   V   713

3133  gat gac cag atg gca gtc att cag tat cc tgg atg gga ctg atg gta ttt gcc atg ggt tgg cgg tcc  3201
714     D   D   Q   M   A   V   I   Q   Y   S   W   M   G   L   M   V   F   A   M   G   W   R   S   736

3202  ttc act aat gtc aac tct agg atg ctc tac ttt gca cct gac ctg gtt ttc aat* gag tat cgc atg cac  3270
737     F   T   N   V   N   S   R   M   L   Y   F   A   P   D   L   V   F   N   E   Y   R   M   H   759

3271  aag tct cga atg tac agc cag tgc gtg agg atg agg cac ctt tct caa gag ttt gga tgg ctc cag ata  3339
760     K   S   R   M   Y   S   Q   C   V   R   M   R   H   L   S   Q   E   F   G   W   L   Q   I   782

3330  acc ccc cag gaa ttc ctg tgc atg aaa gca ctg cta ctc ttc agc att* att cca gtg gat ggg ctg aaa  3408
783     T   P   Q   E   F   L   C   M   K   A   L   L   L   F   S   I   I   P   V   D   G   L   K   805

3409  aat caa aaa ttc ttt gat gaa ctt cga atg aac tac atc aag gaa ctt gat cgc atc att gca tgc aaa  3477
806     N   Q   K   F   F   D   E   L   R   M   N   Y   I   K   E   L   D   R   I   I   A   C   K   828

3478  aga aaa aat ccc aca tcc tgc tca agg cgc ttc tac cag ctc acc aag ctc ctg gat tct gtg cag cct  3546
829     R   K   N   P   T   S   C   S   R   R   F   Y   Q   L   T   K   L   L   D   S   V   Q   P   851

3547* att gca aga gag ctg cat caa ttc act ttt gac ctg cta atc aag tcc cat atg gtg agc gtg gac ttt  3615
852     I   A   R   E   L   H   Q   F   T   F   D   L   L   I   K   S   H   M   V   S   V   D   F   874

3616  cct gaa atg atg gca gag atc atc tct gtg caa ccc aag atc ctt tct ggg aaa gtc aag ccc atc gtg  3684
875     P   E   M   M   A   E   I   I   S   V   Q   V   P   K   I   L   S   G   K   V   K   P   I   897

3685  tat ttc cac aca cag tga  3702
898     Y   F   H   T   Q   *
```

Figure C.4: *Rattus norvegicus* androgen receptor cDNA sequence (GenBank accession no. NM_012502). Protein specific domains, namely, NTD, DBD, Hinge region and LBD, are indicated alternated white and shadowed areas. Exon boundaries are indicated by *. The sequence amplified by the primer set MmAR1/5 is in bold. Primer sequences are underlined.

C.5. *Sparus aurata*

```
64    ATG GAG AGC AGT TTC GCA ACA ACA AGT GGT TAT CAG TCG GAG CAG TAC AGC GTG AAA ATC AAA TGC GAG   132
1      M   E   S   S   F   A   T   T   S   G   Y   Q   S   E   Q   Y   S   V   K   I   K   C   E    23

133   GGC ACC GAA TCT GCC GGA GCG TTG TGG GGC GGT AAT CAC AGC TTT AAT GAC AGG TAC AAC TCC CAG TGT   201
24     G   T   E   S   A   G   A   L   W   G   G   N   H   S   F   N   D   R   Y   N   S   Q   C    46

202   TGG GGT CCG AGG CAG TGC GTG AGC GCA CAC GGA GCA GGA GGC AAC AGC GCG TTA TGT AAT CCA TAC GAG   270
47     W   G   P   R   Q   C   V   S   A   H   G   A   G   G   N   S   A   L   C   N   P   Y   E    69

271   AGG AGC ATG GCG CGG CCG GAA CAA TGG TAC CCA GGC GGG ATG CTG AGG TCG CCA TAT CCC AAC TCC AGC   339
70     R   S   M   A   R   P   E   Q   W   Y   P   G   G   M   L   R   S   P   Y   P   N   S   S    92

340   TAC GTG AAG AGT GAA GTC AGC GAG TGG CTC GAT GTC CCC TAC AAT GAC CCC* AGG TTC GAC GCC AGC AGC   408
93     Y   V   K   S   E   V   S   E   W   L   D   V   P   Y   N   D   P   R   F   D   A   S   S   115
```

```
409  GAG CAC ATG TTC CCA ATG GAG TTC TTC TTT CCA GCT CAG AGG ATG TGT ATG ATC TGT TCA GAC GAG GCG  477
116   E   H   M   F   P   M   E   F   F   F   P   A   Q   R   M   C   M   I   C   S   D   E   A  138

478  TCT GGC TGC CAT TAC GGT GCA CTC ACC TGT GGC AGC TGC AAG GTT TTC TTC AAA AGA GCT GCA* GAA GGC  546
139   S   G   C   H   Y   G   A   L   T   C   G   S   C   K   V   F   F   K   R   A   A   E   G  161

547  AAA CAG AAA TAC CTA TGC GCA AGC AAA AAT GAC TGC ACT ATT GAT AAG CTA AGA AGA AAG AAC TGT CCG  615
162   K   Q   K   Y   L   C   A   S   K   N   D   C   T   I   D   K   L   R   R   K   N   C   P  184

616  TCC TGT CGG CTG AGG AAG TGC TTC GAA GCT GGA ATG ACT CTG GGA* GCA CGT AAA CTG AAA AAG ATT GGG  684
185   S   C   R   L   R   K   C   F   E   A   G   M   T   L   G   A   R   K   L   K   K   I   G  207

685  CAA CAG AAG AAC CCG GAC GAA GAT CAT CCT CTC CAG GAG CCT GCA GAG GTT ATG CCC AAT ATC TCT CCT  753
208   Q   Q   K   N   P   D   E   D   H   P   L   Q   E   P   A   E   V   M   P   N   I   S   P  230

754  AAA ATG GGC CTG AGC TTC AAC TCT CAA GTG GTC TTC TTG AAC ATC CTG GAG TCC ATT GAG CCC GAG GTG  822
231   K   M   G   L   S   F   N   S   Q   V   V   F   L   N   I   L   E   S   I   E   P   E   V  253

823  GCG TAC GCA GGA CAC GAC TAT GGC CAA CCG GAT TCA GCC CCC ACC CTG TTC ACT AGC CTC AAC GAG CTT  891
254   A   Y   A   G   H   D   Y   G   Q   P   D   S   A   P   T   L   F   T   S   L   N   E   L  276

892  GGG GAA AAA CAG CTG GTG AAA GTG GTC AAA TGG GCA AAA GGA TTC CCA* GGT TTT AGA AAT CTC CAT GTG  960
277   G   E   K   Q   L   V   K   V   V   K   W   A   K   G   F   P   G   F   R   N   L   H   V  299

961  GAC GAC CAA ATG ACT GTC ATC CAG TAT TCA TGG ATG GGG GTG ATG GTG TTC GGC CTC GGG TGG CGG TCC 1029
300   D   D   Q   M   T   V   I   Q   Y   S   W   M   G   V   M   V   F   G   L   G   W   R   S  322

1030 TAT AAG AAC GTC AAC GGC AGA ATG CTG TAC TTC GCT CCA GAT CTT GTG TTC AAT* GAA CAC CGG ATG CAC 1098
323   Y   K   N   V   N   G   R   M   L   Y   F   A   P   D   L   V   F   N   E   H   R   M   H  345

1099 ATC TCC ACC ATG TAC GAG CAC TGC ATA CGG ATG AGA CAT CTT TCA CAG GAG TTC CTG CTG CTG CAG ATC 1167
346   I   S   T   M   Y   E   H   C   I   R   M   R   H   L   S   Q   E   F   L   L   L   Q   I  368

1168 ACT CAG GAG GAG TTC CTC TGC ATG AAG GCC CTG CTT CTC TTC AGC ATT* ATT CCA GTT GAA GGT CTG AAG 1236
369   T   Q   E   E   F   L   C   M   K   A   L   L   L   F   S   I   I   P   V   E   G   L   K  391

1237 AGT CAG AAG TAC TTT GAC GAG TTG CGT CTC ACC TAC ATC AAC GAG CTC GAT CGC CTC ATT AAC TAT CGG 1305
392   S   Q   K   Y   F   D   E   L   R   L   T   Y   I   N   E   L   D   R   L   I   N   Y   R  414

1306 ATG AGC GCT AAT TGT TCT CAG AGG TTC TAC CAG CTC ACC CGA CTA CTG GAC TCT CTG CAG ATG* ACG GTG 1374
415   M   S   A   N   C   S   Q   R   F   Y   Q   L   T   R   L   L   D   S   L   Q   M   T   V  437

1375 AAG AAG CTC CAT CAG TTT ACA TTT GAC CTT TTC GTC CAG GCT CAA TCA CTC CCC ACG AAA GTC AGC TTT 1443
438   K   K   L   H   Q   F   T   F   D   L   F   V   Q   A   Q   S   L   P   T   K   V   S   F  460

1444 CCG GAG ATG ATC GGA GAG ATA ATC TCT GTG CAC GTA CCA AAG ATC CTG GCA GGT TTG GCT AAA CCA ATC 1512
461   P   E   M   I   G   E   I   I   S   V   H   V   P   K   I   L   A   G   L   A   K   P   I  483

1512 TTG TTT CAC GAG TAG   1527
484   L   F   H   E   *     487
```

Figure C.5: *Sparus aurata* androgen receptor cDNA sequence. Protein specific domains, namely, NTD, DBD, Hinge region and LBD, are indicated alternated white and shadowed areas. Exon boundaries are indicated by *. The sequence amplified by the primer set SaAR1/5 is in bold. Primer sequences are underlined. *S. aurata* genomic organization was performed by comparative analysis with *Gasterosteu aculateus* AR gene. This is a piscine species highly related to S. aurata (Sarropoulou *et al.* 2008). Putative exon bounderires were detected with the RNASPL program and with *G. aculateus* genome with BLAT program.

APPENDIX D – 18S rRNA sequences and primers

D.1. *Homo sapiens*

```
   1 tacctggttg atcctgccag tagcatatgc ttgtctcaaa gattaagcca tgcatgtcta
  61 agtacgcacg gccggtacag tgaaactgcg aatggctcat taaatcagtt atggttcctt
 121 tggtcgctcg ctcctctcct acttggataa ctgtggtaat tctagagcta atacatgccg
 181 acgggcgctg acccccttcg cgggggggat gcgtgcattt atcagatcaa aaccaacccg
 241 gtcagcccct ctccggcccc ggccgggggg cgggcgccgg cggctttggt gactctagat
 301 aacctcgggc cgatcgcacg ccccccgtgg cggcgacgac ccattcgaac gtctgcccta
 361 tcaactttcg atggtagtcg ccgtgcctac catggtgacc acgggtgacg gggaatcagg
 421 gttcgattcc ggagagggag cctgagaaac ggctaccaca tccaaggaag gcagcaggcg
 481 cgcaaattac ccactcccga cccggggagg tagtgacgaa aaataacaat acaggactct
 541 ttcgaggccc tgtaattgga atgagtccac tttaaatcct ttaacgagga tccattggag
 601 ggcaagtctg gtgccagcag ccgcggtaat tccagctcca atagcgtata ttaaagttgc
 661 tgcagttaaa aagctcgtag ttggatcttg ggagcgggcg ggcggtccgc cgcgaggcga
 721 gccaccgccc gtccccgccc cttgcctctc ggcgcccccc cgatgctctt agctgagtgt
 781 cccgcggggc ccgaagcgtt tactttgaaa aaattagagt gttcaaagca ggcccgagcc
 841 gcctggatac cgcagctagg aataatggaa taggaccgcg gttctatttt gttggttttc
 901 ggaactgagg ccatgattaa gagggacggc cgggggcatt cgtattgcgc cgctagaggt
 961 gaaattcttg gaccggcgca agacggacca gagcgaaagc atttgccaag aatgttttca
1021 ttaatcaaga acgaaagtcg gaggttcgaa gacgatcaga taccgtcgta gttccgacca
1081 taaacgatgc cgaccggcga tgcggcggcg ttattcccat gacccgccgg gcagcttccg
1141 ggaaaccaaa gtctttgggt tccgggggga gtatggttgc aaagctgaaa cttaaaggaa
1201 ttgacggaag ggcaccacca ggagtggagc ctgcggctta atttgactca acacgggaaa
1261 cctcacccgg cccggacacg gacaggattg acagattgat agctctttct cgattccgtg
1321 ggtggtggtg catggccgtt cttagttggt ggagcgattt gtctggttaa ttccgataac
1381 gaacgagact ctggcatgct aactagttac gcgacccccg agcggtcggc gtcccccaac
1441 ttcttagagg gacaagtggc gttcagccac ccgagattga gcaataacag gtctgtgatg
1501 cccttagatg tccggggctg cacgcgcgct acactgactg gctcagcgtg tgcctaccct
1561 acgccggcag gcgcgggtaa cccgttgaac cccattcgtg atggggatcg gggattgcaa
1621 ttattcccca tgaacgaggg aattcccgag taagtgcggg tcataagctt gcgttgatta
1681 agtccctgcc ctttgtacac accgcccgtc gctactaccg attggatggt ttagtgaggc
1741 cctcggatcg gccccgccgg ggtcggccca cggccctggc ggagcgctga gaagacggtc
1801 gaacttgact atctagagga agtaaaagtc gtaacaaggt ttccgtaggt gaacctgcgg
1861 aaggatcatt a
```

Figure D.1: *Homo sapiens* 18S ribosomal RNA sequence (GenBank accession no. NR_003286). Primer sequences are underlined. The sequence amplified by the primer set 18ShrCAV is in bold.

D.2. *Rattus norvegicus*

```
   1 tacctggttg atcctgccag tagcatatgc ttgtctcaaa gattaagcca tgcatgtcta
  61 agtacgcacg gccggtacag tgaaactgcg aatggctcat taaatcagtt atggttcctt
 121 tggtcgctcg ctcctctcct acttggataa ctgtggtaat tctagagcta atacatgccg
 181 acgggcgctg acccccttc ccgtgggggg aacgcgtgca tttatcagat caaaaccaac
 241 ccggtcagcc ccctcccggc tccggccggg ggtcgggcgc cggcggcttt ggtgactcta
 301 gataacctcg ggccgatcgc acgtccccgt ggcggcgacg acccattcga acgtctgccc
 361 tatcaacttt cgatggtagt cgccgtgcct accatggtga ccacgggtga cggggaatca
 421 gggttcgatt ccggagaggg agcctgagaa acggctacca catccaagga aggcagcagg
 481 cgcgcaaatt acccactccc gacccgggga ggtagtgacg aaaaataaca atacaggact
 541 ctttcgaggc cctgtaattg gaatgagtcc actttaaatc ctttaacgag gatccattgg
 601 agggcaagtc tggtgccagc agccgcggta attccagctc caatagcgta tattaaagtt
 661 gctgcagtta aaaagctcgt agttggatct tgggagcggg cgggcggtcc gccgcgaggc
 721 gagctcaccg ccctgtcccc agcccctgcc tctcggcgcc ccctcgatgc tcttagctga
```

```
 781 gtgtcccgcg gggcccgaag cgtttacttt gaaaaaatta gagtgttcaa agcaggcccg
 841 agccgcctgg ataccgcagc taggaataat ggaataggac cgcggttcta ttttgttggt
 901 tttcggaact gaggccatga ttaagaggga cggccggggg cattcgtatt gcgccgctag
 961 aggtgaaatt cttggaccgg cgcaagacga accagagcga aagcatttgc caagaatgtt
1021 ttcattaatc aagaacgaaa gtcggaggtt cgaagacgat cagataccgt cgtagttccg
1081 accataaacg atgccgactg gcgatgcggc ggcgttattc ccatgacccg ccgggcagct
1141 tccgggaaac caaagtcttt gggttccggg gggagtatgg ttgcaaagct gaaacttaaa
1201 ggaattgacg gaagggcacc accaggagtg gagcctgcgg cttaatttga ctcaacacgg
1261 gaaacctcac ccggcccgga cacggacagg attgacagat tgatagctct ttctcgattc
1321 cgtgggtggt ggtgcatggc cgttcttagt tggtggagcg atttgtctgg ttaattccga
1381 taacgaacga gactctcggc atgctaacta gttacgcgac ccccgagcgg tcggcgtccc
1441 ccaacttctt agagggacaa gtggcgttca gccaccgaga ttgagcaata acaggtctgt
1501 gatgccctta gatgtccggg gctgcacgcg cgctacactg aactggctca gcgtgtgcct
1561 accctacgcc ggcaggcgcg ggtaacccgt tgaaccccat tcgtgatggg gatcggggat
1621 tgcaattatt ccccatgaac gaggaattcc cagtaagtgc gggtcataag cttgcgttga
1681 ttaagtccct gccctttgta cacaccgccc gtcgctacta ccgattggat ggtttagtga
1741 ggccctcgga tcggccccgc cggggtcggc ccacggcctt ggcggaggcc tgagaagacg
1801 gtcgaacttg actatctaga ggaagtaaaa gtcgtaacaa ggtttccgta ggtgaacctg
1861 cggaaggatc atta
```

Figure D.2: *Rattus norvegicus* 18S ribosomal RNA sequence (GenBank accession no. NR_003286). Primer sequences are underlined. The sequence amplified by the primer set 18ShrCAV is in bold.