Available online at www.sciencedirect.com
ScienceDirect
Gene 399 (2007) 105–111
GENE
www.elsevier.com/locate/gene

Phylogenetic conservation of the androgen receptor AR45 variant form in placental mammals

Bertram Weiss [a], Hortensia Faus [b,c], Bernard Haendler [b,*]

[a] *Target Discovery, Bayer Schering Pharma AG, D-13342 Berlin, Germany*
[b] *Therapeutic Research Group Oncology, Bayer Schering Pharma AG, D-13342 Berlin, Germany*
[c] *Institut für Chemie/Biochemie, Freie Universität Berlin, D-14195 Berlin, Germany*

Received 6 December 2006; received in revised form 13 April 2007; accepted 29 April 2007
Available online 22 May 2007
Received by M. Di Giulio

Abstract

A cDNA coding for a tissue-specific AR45 variant form of the androgen receptor (AR) has recently been identified in humans, with highest expression levels found in heart. The deduced protein comprises the DNA-binding domain, hinge region and ligand-binding domain of the AR, but not the N-terminal domain which is replaced by a unique, short, seven amino-acid-long stretch. This sequence is encoded by the mutually exclusive exon 1B, located between exons 1 and 2 of the human AR gene. As transcript variants of the steroid receptor family have been shown to have important implications for hormone function, we set out to analyse the genomes of different organisms for potential AR45 expression. We found exon 1B to be conserved in the syntenic chromosomal region of non-human primates such as the chimpanzee *Pan troglodytes*, the orang-utan *Pongo pygmaeus*, the macaque *Macaca mulatta* and the marmoset *Callithrix jacchus*, and of the elephant *Loxondonta africana*, the pig *Sus scrofa* and the dog *Canis familiaris*. Quantification of AR45 transcript levels in heart, skeletal muscle and lung of *Macaca fascicularis* showed the heart to be the main organ of expression. A complete AR45 cDNA was furthermore isolated from the heart of this species. Comparative analysis of the identified AR45 exon 1B regions and of the deduced amino acids revealed a high conservation among species. The four N-terminal residues were identical in all eight species, whereas a few changes were seen in the other three residues in the marmoset, elephant and pig. In contrast, we observed more divergence in the mouse *Mus musculus* and rat *Rattus norvegicus* syntenic regions. Here a stop codon was found downstream of the potential start codon in the putatively deduced protein sequence and it can be inferred that no protein corresponding to AR45 exists in these two species. The existence of AR45 in different placental mammals with the exception of mouse and rat suggests a disappearance in rodents late in evolution, before the separation of the mouse and rat lineages, about 16 million years ago. In view of the potential function of AR45 as a regulator of AR function, and considering the multiple roles of androgens in normal physiology and in several diseases, these findings have important implications with regard to subtle differences in the action of the male sexual hormone in various organisms.

Keywords: Androgen receptor; Steroid receptors; Phylogeny; Exon; Alternative splicing

1. Introduction

Steroid receptors belong to a subgroup of the nuclear receptor superfamily that acquired ligand-binding properties through evolution and bind as homodimers to DNA response elements present in gene regulatory regions (Escriva et al., 2004). They are modular proteins composed of an N-terminal domain, a central DNA-binding domain (DBD), a hinge region and a C-terminal ligand-binding domain (LBD).

Variant forms of steroid receptors resulting from alternating use of transcription start sites, from exon skipping or duplication, or from use of cryptic exons have been described in recent years, adding to the complexity of hormone signaling (Hirata et al., 2003). Among the best characterized examples are the two forms of the progesterone receptor (PR), PR-A and PR-B, which differ by the length of their N-terminal domains (Mulac-Jericevic and Conneely, 2005). They are expressed at varying ratios in several tissues and possess different *in vivo* functions, as demonstrated by

Abbreviations: AR, androgen receptor; DBD, DNA-binding domain; ER, estrogen receptor; GR, glucocorticoid receptor; LBD, ligand-binding domain; MR, mineralocorticoid receptor; PR, progesterone receptor.

* Corresponding author. Tel.: +49 30 468 12669; fax: +49 30 468 18069.
E-mail address: bernard.haendler@bayerhealthcare.com (B. Haendler).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.04.037

the analysis of mice engineered to express only one or the other form. Multiple splice variants of PR transcripts lacking exon 4, exon 6 or parts thereof have been detected in human endometrium, with no evidence however for differences in their levels during the menstrual cycle (Marshburn et al., 2005).

Multiple variants of estrogen receptor (ER) α transcripts generated through differential promoter usage or alternative splicing have been identified, mostly in human tumors and in cell lines (Hopp and Fuqua, 1998; Reid et al., 2002). They give rise to smaller ER protein forms such as ERα46, which lacks the N-terminal domain and ERα36, which additionally lacks part of the LBD (Flouriot et al., 2000; Wang et al., 2005). A regulatory role of the smaller forms on ERα function has been shown following overexpression in different cell lines. Additional splice variants where single or multiple exons are skipped have been found in breast tumors and also in rat uterus (Varayoud et al., 2005). A splice variant of ERβ has been described in mouse under the name ERβ2 (Zhao et al., 2005). It differs from the main form by a region encoding an 18 amino-acid-long stretch inserted in the LBD and leading to altered responses of the corresponding protein to several estrogens. Another variant, ERβcx, which results from alternative splicing of exon 8, exists in humans and non-human primates (Sierens et al., 2004). Interestingly, a correlation has been found between the expression levels of ERβ splice variants and disease outcome in breast cancer (Davies et al., 2004).

Concerning the glucocorticoid receptor (GR), two human forms originating from differential splicing of their 3′-terminal exon exist (DeRijk et al., 2002; Schaaf and Cidlowski, 2002). The main form, GRα, is bound by glucocorticoids and acts as a transcriptional activator whereas GRβ is not stimulated by ligands, due to its shortened LBD. An important role of GRβ in immunomodulation has been reported (DeRijk et al., 2002). The situation is made more complex by the fact that these splice variants may be translated from different initiation sites, thus increasing the number of possible forms (Chrousos and Kino, 2005). Interestingly, GRβ could not be evidenced in mice, but no data are available about other non-human species (Otto et al., 1997).

Several splice variants also exist for the mineralocorticoid receptor (MR). In one case, MR+4, this leads to the synthesis of a protein with a four amino-acid-long insert that enlarges the space between the two zinc fingers of the DBD (Wickert et al., 1998). The ratio of MR+4 to MR transcripts varies in different human tissues. Another variant MR form, depleted of the hinge and LBD domains due to skipping of exons 5 and 6, has been identified in several human tissues (Zennaro et al., 2001). It acts as a ligand-independent transcriptional activator.

Two different AR genes have been identified in tetraploid species such as *Xenopus laevis* and several fish (Fischer et al., 1993; Ikeuchi et al., 1999; Takeo and Yamashita, 1999). In diploid organisms only one AR gene giving rise to a single major product exists. A few variant AR forms have nonetheless been described. They include AR-B, which was originally found in human tissues as a shortened form of the AR originating from an internal ATG codon (Wilson and McPhaul, 1994), but more recent data strongly suggest it to be a degradation product (Gregory et al., 2001). Alternative splicing events in untranslated regions have been reported (Faber et al., 1991). Aberrant alternative splicing due to a mutation in a splice donor site that leads to skipping of exon 2 and premature stop has been reported in a patient suffering from complete androgen insensitivity (Hellwinkel et al., 1999). AR transcripts lacking the exon 3 region coding for part of the DBD have been identified in patients with androgen insensitivity syndrome (Quigley et al., 1992; Ris-Stalpers et al., 1994) and in human breast cancer samples (Zhu et al., 1997). AR45 is the only AR variant form that has so far been described in normal human tissue and the corresponding transcripts are predominantly detected in the heart (Ahrens-Fath et al., 2005). The deduced protein possesses a short, seven amino-acid-long N-terminal stretch instead of the long N-terminal domain found in the AR. DBD, hinge and LBD regions are maintained between AR45 and AR. The AR45-specific stretch is encoded by a DNA sequence located between exons 1 and 2 of the human AR gene which we named exon 1B. Functional *in vitro* studies indicate that AR45 represses AR activity, a role which requires intact DNA- and ligand-binding properties. AR45 can also activate target promoters when the cofactor TIF-2 is overexpressed.

Here we report that the AR45-specific exon 1B is conserved in the genome of different Old and New World monkeys, of dogs, pigs and elephants. It is more divergent in rodents where it is probably not translated. Expression of macaque AR45 was furthermore confirmed by the amplification of the corresponding cDNA. These results suggest important differences in the mode of action of androgens in various organisms.

2. Materials and methods

2.1. Identification of genomic regions containing AR sequences

Starting with the sequence of human AR45 (NM_001011645), we identified the orthologous AR loci in various species through the blastn and megablast web interfaces of Genbank (Benson et al., 2006), Ensembl (Hubbard et al., 2005) or the NCBI Trace Archive (www.ncbi.nlm.nih.gov/Traces/trace.cgi?). The identified sequences were downloaded and aligned to the human AR45 exon 1B sequence using ssearch3 (Pearson, 1990). The following database entries were used: *Homo sapiens* AL049564.11.1.139033 Ensembl NCBI35:X:66444022:66582954; *Pan troglodytes* AADA01164817 Ensembl and NCBI trace ids 233228415, 187294708; *Pongo pygmaeus* NCBI Trace ids 1221376572, 760105968, 820575451, 840802775, 850260406, 855963445, 855969627, 894731170, 911859047; for *Macaca mulatta*, 33 NCBI trace ids were used for assembly (wph29g02.g1, RTCM762TR, 1091119488110, 92590827, MQAA-ahe09g08.b1, MQAA-aly56e02.g1, 124533172, 127810861, 132130991, 132789224, 132798838, PPAC-aly31h11.g1, PPAC-aqz98f06.g1, PPAC-bjs35c08.g1, PPAC-bfm68a12.b1, 170983688, S215P61276RA8.T0, 17000178552096, 130707973, 136990342, 17000182912617, XSNP51152-108m04.p1k, XSNP49605-220c01.q1k, XSNP49605-297a08.q1k, XChr_X_11321_2-225j16.q1k, XSNP50242-141g23.p1k, XSNP49605-182f18.p1k, 137826797, 17000196044675, 17000181700581, 17000110148470, PPAC-awc69c03.g1, PPAC-bkz24e08.b1, 17000037504609, XSNP50242-28c03.q1k,