Production of Recombinant Androgen Receptor in a Heterologous Expression System

Olli A. Jäanne,1,2 Jorma J. Palvimo,1 Pekka Kallio,1 Merja Mehto,1 Yan-Bo Xie,2 and Ya-Ping Sui2

To facilitate detailed studies of androgen receptor, we have produced a full-length receptor protein and some of its deletion mutants in Spodoptera frugiperda (Sf9) insect cells, using the baculovirus expression system. Recombinant baculovirus DNA-infected Sf9 cells expressed these proteins in very high quantities, which represented as much as 30–40% of total insect cell protein at 72 h after infection. Only <1% of the recombinant protein was soluble in low-salt buffers; the majority formed electron-dense cytoplasmic aggregates 30–40 nm in diameter. These aggregates could be solubilized in 6 mol/L guanidine HCl, and biologically active receptor was generated by diluting the guanidine HCl preparation 20- to 50-fold. The full-length receptor, expressed either in a soluble or aggregated form, had characteristics typical of a native receptor: it bound steroids with high affinity and specificity, interacted with DNA in a sequence-specific fashion, and was recognized by domain-specific receptor antibodies. Androgen-receptor protein purified to homogeneity in guanidine HCl required the presence of Zn²⁺ ions during the refolding to reconstitute its DNA-binding form; ZnCl₂ was not, however, needed to restore the receptor's steroid-binding activity.

Indexing Terms: steroid hormones · gene expression · baculovirus · receptor protein · recombinant DNA

Androgens play an essential role in the differentiation, development, and maintenance of normal male reproductive functions. In addition, many sexually dimorphic responses in gonadal tissues are controlled by androgens through mechanisms that appear to be identical to those in the organs of the reproductive tract; therefore, classification of biological responses to male sex steroids as either androgenic or anabolic is not very informative (1). Physiological androgens such as testosterone and 5α-dihydrotestosterone elicit their actions by binding to intracellular receptor proteins from a family of ligand-responsive transcription factors. This large protein family includes receptors for different classes of steroid hormones, multiple receptor species for thyroid hormones and retinoic acids, vitamin D receptor, and several so-called orphan receptors, whose physiological ligands are currently unknown (2–6). Receptor–androgen complexes interact with regulatory regions of specific genes and thereby modulate their activity, which eventually results in the expression of an androgen-induced phenotype. Acquisition of this phenotype occurs in a cell- and tissue-specific fashion; it involves both hyperplastic and hypertrophic growth responses in the accessory sex organs but only a hypertrophic response in nonreproductive organs (1).

Availability of purified androgen-receptor protein is essential for many structural and functional studies; therefore, the protein should be available in large quantities. However, this receptor is present in androgen-target tissues in very low concentrations (7), which, along with its poor stability, has made it impossible to purify the receptor in amounts needed for detailed structure–function experiments. Truncated forms of the androgen-receptor protein have been produced in both prokaryotic and eukaryotic expression vectors (8–12). Our laboratory has adopted the baculovirus expression system (13, 14) to generate a full-length form of this protein in large quantities (15). Here, we will review our experience with the baculovirus expression system in a large-scale production of the full-length androgen receptor and some of its variant forms, and compare our data with those obtained in other laboratories, who have used a similar approach for a heterologous production of other members of the steroid/thyroid hormone receptor superfamily.

Structural Features of the Androgen-Receptor Protein

Typical structural features of a protein belonging to the steroid/thyroid hormone receptor superfamily are depicted in Figure 1. The androgen receptor is one of the largest proteins in this family; its deduced amino acid sequence contains 917–919 residues, from which one can predict a molecular size of ~96,000 Da for the human receptor (16–19). The reported sequence for the rat androgen receptor is slightly shorter, 902 amino acids (16, 18). The overall molecular size of an individual member in the receptor superfamily is largely determined by the length of its modulator (or transactivation) domain, also termed the A/B region, which extends from the N-terminus to the beginning of the DNA-binding domain (Figure 1). The length of this sequence ranges from ~25 amino acid residues (vitamin D receptor) to >600 residues (mineralocorticoid receptor). The function of the A/B domain has not been delineated in sufficient detail for any one of the members, but this region contains sequence information for at least two key features: (a) optimization of transactivation capability of the receptor, and (b) gene recognition specificity for the transcriptional regulation. An example of the latter feature is the presence of two forms of progesterone receptor (A and B forms), which differ in the length of their A/B domains and which have been shown to

---

1 Department of Physiology, University of Helsinki, Siltaamanner Penger 20 J, SF-00170 Helsinki, Finland (address for correspondence).
2 The Population Council, New York, NY 10021.

Received August 6, 1992; accepted October 19, 1992.
exhibit gene-specific regulatory functions (20). The A/B domains among the different family members have the least-conserved amino acid sequences; consequently, they represent most immunogenic parts of the proteins. It is also probable that this lack of identity of amino acid sequences will enable interaction of a given A/B domain with other cellular proteins in a receptor-specific fashion.

The human androgen receptor contains long homopolymeric amino acid segments that comprise motifs of 21 glutamine residues, 8 proline residues, and 24 glycines in the transactivation domain (6). The lengths of these segments vary among healthy individuals, indicating that the androgen-receptor protein is polymorphic. The rat receptor contains similar poly(amino acid) motifs, although their positions in the A/B domain differ from those in the human receptor (16, 18). Interestingly, X-linked spinal and bulbar muscular atrophy (Kennedy disease), an adult-onset form of motor neuron disease, has recently been found to contain an increased size of the polyglutamine region in the transactivation domain (CAG repeats in the receptor gene) (21, 22). The amplified CAG repeats were absolutely associated with the disease and were about twice as long in the affected patients as in healthy control subjects (21, 22). The biological function of the polyglutamine region or the other poly(amino acid) motifs of the A/B domain is not known; however, similar repeats have been identified, not only in other steroid receptors, but also in several transcription-regulating proteins, particularly in the homeotic gene products (23). Enlargement of the polyglutamine repeat appears to impair the function of the androgen receptor, at least to some extent, because the patients with Kennedy disease show late-onset signs of defective androgen action. The affected patients suffer from degeneration of motor neurons, the growth of which may be controlled in a cell-specific fashion by the transactivation domain of the androgen receptor.

The DNA-binding domain of steroid receptors consists of 66–68 amino acid residues. This region is cysteine-rich and appears to fold into two so-called zinc finger structures, with one zinc atom being tetrahedrally coordinated with four cysteines in each case (2–6). Human and rat androgen receptors have identical amino acid sequences in this domain. The similarity between the androgen receptor and other steroid hormone receptors is also greatest in the DNA-binding domain. For example, this region has an 82% amino acid sequence identity with the human progesterone receptor, 79% with human glucocorticoid and mineralocorticoid receptors, 59% with the human estrogen receptor, and ~40–50% with other members of the steroid/thyroid hormone receptor superfamily (16–19). The DNA-binding domain interacts with cis-acting elements of the regulated genes, known as the hormone-responsive elements, which are usually located in the 5′-flanking region of the gene. A unique androgen-responsive element (ARE) has not yet been defined; in some instances, such as the mouse mammary tumor virus long terminal repeat, the same element that mediates induction by glucocorticoids and progestins also confers androgen-responsiveness on a heterologous promoter (24, 25). Plausibly, there are indeed no distinct AREs, or if there are, they are very similar to the responsive elements for glucocorticoids and progestins. As mentioned above, the DNA-binding domain of the androgen receptor also displays a remarkable amino acid sequence identity with the glucocorticoid, progesterone, and mineralocorticoid receptors. This amino acid sequence similarity in the DNA-binding domains and the nucleotide sequence identity of hormone-responsive elements in genes regulated by four different steroids poses an intriguing biological enigma: how are the unique actions of these four steroids actually specified under physiological conditions? In some cases, this may be exerted through the presence or absence of a given receptor; however, a simple abundance of receptor may not suffice in most cases. It is, therefore, logical to assume that an important additional determinant for the hormone specificity lies in the interaction of the A/B domain with cell- and tissue-specific proteins, along with general transcription factors.

The C-terminal domain of the receptor proteins, which encompasses ~250 amino acid residues, is involved in the ligand-binding and in the interaction with heat-shock proteins such as hsp90 (Figure 1). The similarity of this sequence between the androgen receptor and other steroid receptors agrees with previous data from in vitro steroid-binding studies and biological experiments in vivo (26), in that androgens, progestins, and glucocorticoids can, at least to some extent, interact with more than

---

**Fig. 1.** Schematic representation of the structural domains of a protein belonging to the steroid/thyroid hormone receptor superfamily. The numbers in brackets depict the approximate number of amino acid residues in each of the four regions. Some of the putative functional characteristics of the domains are also listed. Adapted from references 2–6.

---

3 Nonstandard abbreviations: ARE, androgen-responsive element; nt, nucleotides; Sc, Spodoptera frugiperda; and Gdn, guanidine.
one receptor type. This is particularly true for synthetic steroids, which may have higher affinities for multiple receptors than does the corresponding physiological hormone for its cognate receptor (27). Although both of the two physiological androgens, testosterone and 5α-dihydrotestosterone, interact directly with the androgen receptor and mediate hormonal responses, in certain tissues conversion of testosterone to the more potent agonist 5α-dihydrotestosterone is required for the steroid action to occur. This requirement is particularly clear during male sexual development, when formation of 5α-dihydrotestosterone is mandatory for virilization of external genitalia and development of the prostate gland. Virilization of the Wolffian duct structures is, however, thought to be mediated by testosterone (28). Because the mammalian genome contains only one androgen-receptor gene, these effects are most probably mediated by a single receptor protein. Furthermore, immunoblotting studies with receptor antibodies have failed to uncover the presence of multiple, tissue-specific androgen-receptor species (29, 30).

Production of Androgen Receptors in a Heterologous System

Construction of the Recombinant Baculovirus Strains

A full-length rat androgen receptor cDNA was inserted into the baculovirus transfer plasmid pVL1393 by use of the EcoRI and Pst I sites on the multiple cloning sequence as described in detail previously (15). The transfer plasmid, termed pVXRAR, contained the entire protein-coding region and 32 nucleotides (nt) and 87 nt of the 5'- and 3'-untranslated regions of the mRNA, respectively. Because the initiation codon for the polyhedrin protein translation is mutated in pVL1393, translation of the RNA encoded by the recombinant baculovirus initiates at the authentic AUG of the coding sequence for androgen receptor mRNA. Recombinant baculovirus DNA (designated AcrAR) was generated in Spodoptera frugiperda (Sf9) cells by homologous recombination with wild-type viral (AcNPV) DNA and pVXRAR DNA by standard techniques (13–15).

Sf9 insect cells infected with AcrAR DNA for 3–4 days produced extremely high quantities of androgen-receptor protein that migrated on polyacrylamide gel electrophoresis at \( M_r = 110,000 \), indicating that a full-length protein was indeed formed in the recipient cells. At the time of peak protein production, the amount of receptor in infected Sf9 cells ranged from 30% to 50% of total insect cell protein (15). However, only a fraction of the recombinant receptor was soluble upon homogenization of infected Sf9 cells in buffers containing nonionic detergents or moderate salt concentrations, with the soluble fraction representing usually <1% of the total recombinant receptor content (15). The rest of the receptor population remained as aggregated cytoplasmic particles, 30–40 nm in diameter, and exhibited a relatively regular shape in transmission electron microscopy. These two receptor populations had identical characteristics in all physical-chemical and functional studies performed so far, including molecular size, steroid-bind-

ing affinity and specificity, antibody recognition, and binding to different AREs (15, and below). The reason for this very high level of androgen-receptor expression is currently unknown to us, but it may relate to the structure or the length of the 5'-untranslated region in the construct used to generate the AcrAR baculovirus strain. In this context, we note that the length of the 5'-untranslated region reportedly influences the expression of recombinant 1,25-dihydroxyvitamin D\(_3\) receptor in insect cells (31).

In addition to the full-length receptor, we also produced several deletion mutants for the rat androgen receptor cDNA, and generated respective recombinant baculovirus strains for expression of mutated proteins in insect cells. Schematic representation of some of the mutated proteins is shown in Figure 2. In these experiments, the basic transfer plasmid construct was the same, i.e., the pVL1393 series, and the 5'-untranslated sequence of androgen receptor mRNA was kept unchanged. In more recent studies, we used linearized baculovirus DNA in lieu of the circular wild-type AcNPV DNA to produce recombinant virus strains by homologous recombination and found that this improves the yield of recombinant plaques. As was the case with the production of full-length receptor protein, the mutated proteins were expressed to very high quantities and represented ~30% of total insect cell protein 72 h after infection (Palvimo et al., unpublished results). Distribution of the expressed recombinant protein between the soluble and aggregated forms was not significantly influenced by the above mutations: only a minor portion of each recombinant protein could be released from Sf9 cells by homogenization in low-salt buffers.

![Fig. 2. Structures of the full-length androgen receptor and its mutants produced in insect cells by use of the baculovirus expression system](image_url)

The numbers underneath the full-length protein depict the beginning and the end of the various domains in the rat androgen receptor. Deletion mutants were constructed as follows: \( \Delta 46-408 \), a fragment between two Nhe I sites was deleted; \( \Delta 38-296 \), a fragment between two Sma I sites has been deleted; \( \Delta 40-147 \), the sequence between two Apa I sites was deleted; and \( \Delta 787-902 \), a fragment between an EcoRI site and the end of the coding region has been removed.
Properties of the Recombinant Receptor Proteins

As mentioned previously, no significant functional differences were observed between the two forms in which the recombinant receptors were expressed; therefore, the experimental data reviewed below for one form can be extended to include the other form, too. The aggregated receptor was routinely solubilized in 6 mol/L guanidine HCl (GdnHCl) at room temperature for 1 h, after which it was either purified to virtual homogeneity by gel-exclusion chromatography (15) or used without further enrichment for subsequent studies. Recombinant receptor was, of course, biologically inactive in 6 mol/L GdnHCl, but it retained most of its functional characteristics upon dilution of the GdnHCl by a factor of 20–50. This renaturation applied to all activities tested, including sequence-specific interaction with DNA, steroid-specific hormone binding, and recognition by domain-specific antisera (15).

The presence of Zn\(^{2+}\) ions in the dilution buffer used for reconstituting the recombinant receptor from 6 mol/L GdnHCl was required for the receptor's acquisition of specific DNA binding. The optimal Zn\(^{2+}\) ion concentration was \(\sim 50\) μmol/L, higher concentrations being inhibitory. Of the several other divalent cations tested, Cd\(^{2+}\) ions were as active as Zn\(^{2+}\) ions but at \(\sim 10\)-fold lower concentration; and, similar to ZnCl\(_2\), higher CdCl\(_2\) concentrations prevented receptor binding to ARE sequences. As mentioned previously, the DNA-binding domain of all steroid receptors contains two cysteine-rich zinc "fingers," motifs that are required for their binding to the hormone-responsive elements (2–6). A recombinant glucocorticoid receptor fragment encompassing the DNA-binding domain (total length: 150 amino acids) purified to homogeneity has been previously shown to ligate reversibly two Zn\(^{2+}\) ions (32).

Release of the zinc ions from the protein by chelating agents yielded an apoprotein that failed to bind to its target DNA element; the binding was restored by a preincubation with Zn\(^{2+}\) or Cd\(^{2+}\) ions. Our studies with the full-length recombinant androgen receptor purified to homogeneity in the presence of 6 mol/L GdnHCl by gel filtration thus agreed completely with the data on the glucocorticoid receptor and indicated that Zn\(^{2+}\) ions are essential for folding the receptor from GdnHCl in such a way that it acquires the specific DNA-binding function. In our experiments (15), the optimal ZnCl\(_2\) concentration of \(\sim 50\) μmol/L was one-fifth the optimum (250 μmol/L) reported by Freedman et al. (32) for the DNA-binding domain of the recombinant glucocorticoid receptor. The reason for this difference is not known. However, the optimal Cd\(^{2+}\) ion concentration for the full-length androgen receptor was the same as that for the DNA-binding domain of the glucocorticoid receptor (5 μmol/L in both cases) (15, 32).

The time course of appearance of androgen binding (\(K_d = 5\) nmol/L for \(^{3}H\)mibolerone, a synthetic androgen) in the soluble cytosol of SF9 cells revealed that, as measured by the steroid-binding capacity, \(\sim 150\) 000 receptors/cell were present by 72 h after infection (15). Even though the soluble receptor fraction represented \(<\)1% of the total expressed receptor population in SF9 cells, its concentration was at least fivefold greater than that reported for rat prostate, which is supposed to be the richest tissue source for the receptor. An identical time course of accumulation was found by using immunoblotting to estimate receptor protein content in both soluble and aggregated forms. The steroid-binding specificity of the soluble or reconstituted androgen receptor expressed in insect cells was very similar to that of the native receptor in its physiological target tissues (33), in that the rank order of competition was androgen > estradiol > progesterone > glucocorticoid. The equilibrium dissociation constant for \(^{3}H\)mibolerone binding (5 nmol/L) was somewhat higher for the recombinant protein purified to homogeneity than that reported for the physiologically occurring receptor form—which may indicate that some auxiliary proteins are required for this event. However, in another study, no such protein was needed for a truncated androgen receptor expressed in prokaryotic cells to bind its cognate ligand with high affinity (10).

An additional criterion to evaluate functional properties of the recombinant receptor is its ability to interact with specific DNA motifs. We have used for this purpose a glucocorticoid-responsive element (GRE) of the tyrosine aminotransferase gene, which has been previously shown to confer androgen responsiveness upon a reporter gene (34), and an ARE present in the first intron of the rat C3(1) gene (35). Soluble extracts from AcAr-infected cells, or samples reconstituted from 6 mol/L GdnHCl in the presence of ZnCl\(_2\) interacted in a nucleotide sequence-specific fashion with the two DNA elements, as revealed by band-shift experiments (15). Samples that were prepared and incubated in the presence of 50 mmol/L 5α-dihydrotestosterone yielded results identical with those studied in the absence of the steroid. The mutant receptor with a C-terminal deletion (Δ787–902, Figure 2) bound to ARE sequences as well as the full-length receptor. DNA-binding experiments conducted with the deletion mutants Δ46–408 and Δ38–296 revealed that a partial removal of the A/B domain did not abolish the ability of these variant receptors to interact with specific DNA sequences; however, the apparent affinity of the two mutants for ARE sequences was not as high as that of the full-length receptor (Palvimo et al., unpublished observations). Properties of the other variant receptors depicted in Figure 2 have not yet been studied in this respect.

Several antisera that we have raised (15, 36) in rabbits against defined sequences corresponding to various regions of the androgen receptor have proven very useful reagents. The locations of these sequences in the human receptor are illustrated in Figure 3. Three of the antisera were particularly valuable: antibodies against androgen receptor peptide 3 (N-terminal), peptide 4 (mid-region of the A/B domain), and peptide 5 (hinge region). The first antiserum, corresponding to residues 14–32 of the human receptor, reacted well with all recombinant receptors (Figure 2) so far produced in
insect cells, indicating that the synthesis of these proteins indeed commences at the authentic AUG codon of the receptor mRNA. This antibody along with that raised against androgen receptor peptide 4 served as excellent specificity controls in band-shift experiments, because they both supershifted and stabilized ARE receptor complexes (Figure 4). The fact that responsive element–receptor complexes were significantly stabilized by these antibodies may be interpreted to indicate that accessory proteins are involved in the formation of this complex under in vivo conditions. The stabilization effect appeared to be more dramatic with the deletion mutants than with the full-length recombinant receptor. By contrast, the hinge-region antiserum (against androgen receptor peptide 5) inhibited the interaction of both soluble and renatured receptor preparations with specific DNA elements (Figure 4). Two possible mechanisms may account for this effect: (a) direct masking by the antibody of the zinc finger region in the receptor, and (b) inhibition of receptor dimerization, which may be important for the formation of specific DNA-receptor complexes (2-6). It remains to be elucidated whether this particular antiserum also blunts androgen receptor-dependent transcription in vitro, in a fashion similar to an antiserum directed against the DNA-binding domain of the estrogen receptor, which prevented both estrogen response element–receptor interaction and estrogen receptor-mediated transcription in a cell-free system (37).

Expression of Other Members of the Steroid Receptor Superfamily in the Baculovirus System

During the last few years, several members of the steroid/thyroid hormone receptor superfamily have been overexpressed by use of the baculovirus system. These include at least the following proteins: mouse and human estrogen receptor (37, 38), human and chicken progesterone receptors (39, 40), human and rat glucocorticoid receptor (41, 42), human mineralocorticoid receptor (43), rat and human vitamin D receptor (31, 44), and human thyroid hormone receptor β1 (45, 46). The results from these studies indicate very clearly that the baculovirus expression system is well suited for a large-scale production of the ligand-induced transcription factors, such as those belonging to the steroid/thyroid hormone receptor superfamily. A relatively high level of expression was achieved in all cases and, even more importantly, biologically active full-length receptors were produced in every instance. The level of expression in our studies with the androgen receptor appeared to be higher than the expression of other receptors cited above, although accurate comparisons cannot be made because different methods were used to extract receptors from infected cells. Furthermore, we are not aware of other reports on steroid/thyroid hormone receptor overproduction with the baculovirus sys-

Fig. 3. Location of the peptide sequences used to raise androgen receptor antibodies
Schematic representation of the human receptor structure is based on the sequence published by Luhahn et al. (17). Properties of these antisera have been described in detail in previous publications from this laboratory (15, 36). The number of + signs in immunoblotting indicates how well an antiserum performed in Western blotting and (or) immunohistochemistry. In the DNA-binding experiments, + refers to stabilization and up-shifting of ARE–receptor complexes, − refers to inhibition of ARE-receptor interaction, and ? indicates that this function has not yet been tested by gel-retardation experiments.

Fig. 4. Influence of domain-specific androgen receptor antibodies on the interaction between a full-length recombinant receptor protein and an androgen-responsive element (ARE) of the rat tyrosine aminotransferase gene, as demonstrated by an electrophoretic mobility shift assay
Receptor preparations originated from soluble insect cell extracts and were prepared as described by Xie et al. (16). F, free [32P]ARE; AR, androgen receptor–[32P]ARE complex; Ab-AR, antibody–androgen receptor–[32P]ARE complex; NRS, normal rabbit serum; Ab, antiserum; ARPS, ARP4, and ARPS, antisera raised against AR peptides 3, 4, and 5, respectively (see Fig. 3). The two bands between free [32P]ARE and AR–ARE complexes represent non-specific DNA–protein interaction of unknown nature.
system, in which much aggregated form of the receptor was expressed. However, aggregation of the expressed protein similar to that in our work has been reported for other proteins produced in insect cells, such as protein phosphatase 1 (47) and urate oxidase (48). Aggregation did not necessarily present a problem in any of these cases; for instance, the initial insolubility permitted purification of the androgen receptor to virtual homogeneity in a single gel-filtration step (15).

Except for the human and chicken progesterone receptors (39, 40), all other intracellular receptors expressed in insect cells were able to interact with specific DNA elements in a hormone-independent fashion. The reason for this dissimilarity remains to be elucidated. Possibly there are inherent differences among the members of this superfamily in their activation process, i.e., the mechanisms by which their DNA-binding activity is generated. However, it is equally likely that this disparity relates to the different experimental conditions used in different studies.

Expression of recombinant proteins belonging to the steroid/thyroid hormone receptor superfamily by use of a baculovirus vector has become very popular during the last 2 years among the investigators in this field. The insect cell system offers many advantages, such as expression of full-length proteins, absence of fusion protein formation, low proteolytic activity, and apparently proper posttranslational modifications of expressed proteins in these cells. Furthermore, the system is not overly demanding to set up with currently available reagents. We anticipate that in the near future this overexpression system will enable production of native receptor proteins in amounts sufficient for their crystallization.

The work performed in the authors' laboratory was supported by grants DK37692 and HD13541 from the National Institutes of Health, the Medical Research Council of the Academy of Finland, the Finnish Life and Pension Insurance Companies, the Finnish Foundation for Cancer Research, and the Sigrid Jusélius Foundation.

References
2. Evans RM. The steroid and thyroid hormone receptor superfamily [Review]. Science 1988;240:889-95.