Molecular mechanisms of androgen receptor-mediated gene regulation: structure–function analysis of the AF-1 domain

I J McEwan

School of Medical Sciences, College of Life Sciences and Medicine, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, Scotland, UK

(Requests for offprints should be addressed to I J McEwan; Email: lain.mcewan@abdn.ac.uk)

Abstract

The androgen receptor is a ligand-activated transcription factor that binds DNA response elements as a homodimer. Binding sites for the receptor have been identified both upstream and downstream of the transcription start site. Once bound to DNA, the receptor contacts chromatin remodelling complexes, coactivator proteins and components of the general transcription machinery in order to regulate target gene expression. The main transactivation domain, termed AF1, is located within the structurally distinct amino-terminal domain. This region is structurally flexible but adopts a more folded conformation in the presence of the binding partner TFIIF, and this in turn enhances subsequent protein–protein interactions. Thus, there is likely to be a dynamic interplay between protein–protein interactions and protein folding, involving AF1, that is proposed to lead to the assembly and/or disassembly of receptor-dependent transcription complexes.

Introduction

In 1849, Arnold A Berthold reported that the testes were responsible for producing a substance, acting via the bloodstream, that restored normal characteristics of cockerels to castrated birds. Some 86 years later, the ‘active principle’, testosterone, was crystallised by Ernst Laqueur and co-workers (reviewed in Freeman et al. 2001). Testosterone is a C19 steroid; together with the more potent 5α-reduced metabolite dihydrotestosterone (DHT), it is responsible for male development in utero, and secondary sexual characteristics and male reproductive function and fertility at puberty and in adult life. Both testosterone and DHT, acting through a single intracellular receptor protein termed the ‘androgen receptor’ (AR), alter patterns of gene expression in target tissues. In addition, androgens can elicit rapid responses that have been termed ‘non-genomic’ to distinguish them from the direct effects of the AR on gene expression (reviewed in Heinlein & Chang 2002).

Prostate cancer

The normal growth and function of the prostate gland are under the control of androgens, and it is this regulation that is targeted by ‘hormone-based’ therapies for prostate cancer. Prostate cancer is a leading cause of cancer-related deaths in men in the West (for recent reviews, see Taplin & Ho 2001, Debes & Tindall 2002, Gelmann 2002). Therapeutic strategies involve androgen ablation by orchietomy or through the use of gonadotropin-releasing hormone analogues in the first instance. Such treatments can be combined with or followed by the use of an antiandrogen to block AR function and are initially successful in providing disease management and controlling tumour growth. However, by mechanisms that remain unclear, the disease eventually progresses to a more aggressive, hormone-refractory or androgen-independent state. This may be accompanied by amplification of the AR gene or the acquisition of point mutations with a gain of function (for example, promiscuous ligand
binding), which have been selected for under conditions of low androgens (reviewed in Taplin & Ho 2001, Debes & Tindal 2002, Gelman 2002). In addition, there is now significant evidence showing that the AR can be activated by growth factors and related molecules in a hormone-dependent or -independent manner (see Gnanapragasam et al. 2000 and references therein). Thus, given the central role of the AR in prostate biology, and current therapeutic interventions in prostate cancer, considerable research effort is now focused on understanding the structure and function of the receptor protein and the identification and characterisation of androgen-regulated genes. This review will focus on recent findings that shed light on the molecular details of receptor-dependent gene regulation.

The AR gene and protein structure

The gene for the AR is located on the long arm of the X chromosome (q11-12) and consists of eight exons spanning a region of 170-180 kbp of genomic DNA in the rat and human (www.ensembl.org). Figure 1 illustrates the organisation of the AR gene and the domain structure of the receptor protein (reviewed in Quigley et al. 1995, MacLean et al. 1997, Hiipakka & Liao 1998). The C-terminal part of the protein consists of a signal transduction domain involved in ligand binding and receptor dimerisation. The ligand-binding domain (LBD) is encoded by exons 4-8. The central 60–100 amino acids, encoded by exons 2 and 3, represent the DNA-binding domain (DBD), and are responsible for targeting the receptor to specific sequences in the genome associated with target genes. A hinge domain, containing a bipartite nuclear localisation signal (NLS), amino acids 608–625, links the LBD and DBD. The structurally distinct N-terminal domain (NTD) is involved in contacting the transcriptional machinery and transcription regulation, and is encoded by a single large exon (exon 1). Thus, the receptor protein consists of four structurally and functionally distinct domains, which together mediate the genomic actions of testosterone in androgen target tissues.

Androgen-regulated genes

Androgens have long been known to be essential for the development and function of male reproductive tissues.

Figure 1 AR gene organisation and domain structure of the protein. The AR gene (top) consists of eight exons that give rise to the characteristic domain structure of the receptor protein (bottom). LBD, ligand-binding domain; DBD, DNA-binding domain; TAU, transactivation unit. Q and G represent polyglutamine and polyglycine repeats respectively. The receptor AF1 transactivation domain, amino acids 142–485, is highlighted together with tryptophan residues (W) 396 and 432.
Recent results from DNA microarrays (Nelson et al. 2002, Eder et al. 2003, Jiang & Wang 2003) and proteomic analyses (Umar et al. 2003a,b) have led to the identification of a large number of genes coding for proteins involved in protein folding, trafficking and secretion, metabolism, the cytoskeleton, cell-cycle regulation and signal transduction regulated by androgens. Figure 2 shows the organisation of transcription factor-binding sites within the regulatory regions of a selection of well-characterised androgen-regulated genes. The genes for prostatic specific antigen (PSA), probasin, keratinocyte growth factor (KGF) and antiapoptotic factor p21 are induced by androgens, while the gene for the tumour-suppressing serpin, maspin, is negatively regulated. A key feature of these genes is the presence of one or more AR-binding sequences (ARE, ARR or HRE), together with binding sites for housekeeping (Sp1, CCAAT and NF-1), inducible (NF-κB, Ets and AP-1) and tissue-specific transcription factors. Thus, androgen-regulated gene expression is likely to involve the coordinated interactions of the receptor protein and other transcription factors. The AR gene is autoregulated by androgens, and both upregulation and downregulation have been reported in different cell lines: the interesting features of the receptor gene are the absence of a TATA element and the location of the ‘androgen-regulated region’ over 2 kb downstream of the start site (see Grad et al. 1999 and references therein). The transcription factors Sp1 (Chen et al. 1997), cAMP-response element binding (CREB) protein (Mizokami et al. 1994) and c-myc (Grad et al. 1999) have all been associated with upregulation of the AR gene. In contrast, negative regulation involved the inducible transcription factor NF-κB (SUPukar et al. 1995) and the constitutively active factor NF-1 (Song et al. 1999).

**AR mechanism of action: DNA recognition and binding**

The regulation of gene expression requires the specific tethering of the AR homodimer to the promoter and/or enhancer sequences of target genes (Fig. 2). Upon binding the hormone testosterone, the LBD undergoes a rearrangement of a three-layer, α-helical, sandwich structure.
conformation, causing a realignment of the most C-terminal helix (helix 12) with the core of the domain, and resulting in a more compact structure (Matias et al. 2000). Concomitant with this conformational change is the release of a number of molecular chaperones (hsp90 and hsp70) and the translocation of the receptor to the nucleus. Recognition and binding of DNA is achieved by the DBD (amino acids 550–624\(^1\)), which comprises two Zn-binding modules, with four cysteine residues coordinating each Zn ion. Nuclear magnetic resonance spectroscopy and X-ray crystallography studies have revealed a common structural fold for the DBD of both steroid and non-steroid receptors (reviewed in Zilliacus et al. 1995, Verrijdt et al. 2003). The main feature of this fold comprises two \(\alpha\)-helices positioned perpendicular to each other: the N-terminal helix is the ‘recognition helix’ and is positioned within the major groove of the DNA. Within this helix are the amino acids glycine 568, serine 569 and valine 572, collectively termed the ‘P-box residues’, that are involved in part in DNA sequence recognition. The P-box residues found within the AR are identical to those at the corresponding positions within the glucocorticoid, mineralocorticoid and progesterone receptors (jointly referred to as class I steroid receptors, with the oestrogen receptor being the sole member of class II). Clearly, for the class I steroid receptors, determinants other than the P-box residues must also play a role in determining DNA recognition and binding specificity in cells where more than one receptor is present (for review, see Verrijdt et al. 2003). Indeed, mutational analysis has highlighted the 12 amino acids immediately C-terminal of the core DBD as being important for AR-specific DNA binding (Schoenmakers et al. 1999). Further mutational analysis by Claessens and co-workers identified glycine 618 and leucine 625 within the ‘C-terminal extension’ sequence and threonine 593 within the second Zn-module as crucial residues in AR selective binding (Schoenmakers et al. 2000).

A consensus DNA-binding site, 5'-GGA/TACAnnnTGTTCCT-3', for the AR was proposed over 10 years ago by Roche et al. (1992) using a DNA-binding site selection assay. This sequence is remarkably similar to the consensus derived from a comparison of over 20 AR-binding sequences: 5'-AGA/TACA/Tgca/gT/AGTTCT-3'. Both these sequences are palindromic in nature, with the receptors binding in a ‘head-to-head’ configuration, and can be thought of as hormone or steroid response elements (HRE or SRE) to reflect the non-selective nature of binding by class I steroid receptors. In contrast, a comparison of five selective androgen response elements (ARE) — DNA binding and/or activation being observed only with the AR, not the GR or PR—identified in the genes for PEM, probasin, secretory component and sex-limited protein, gives a slightly modified consensus, 5'-A/GGCTCnnnA/TGTTCCT/C-3', with the main difference being the central two base pairs of the 5' half-site (underlined). Interestingly, this sequence can be read as an imperfect palindrome or as a direct repeat. The binding of the AR to direct repeat sequences has previously been suggested on the basis of a 17 bp sequence derived from a pool of degenerate oligonucleotides (Zhou et al. 1997). Claessens and co-workers have argued compellingly that the AR has two modes of DNA recognition and binding, with specific ARE sequences recognised as direct repeats and bound in a ‘head to-tail’ configuration by the receptor dimer (see Schoenmakers et al. 2000 and references therein). The latter is intriguing and, if proven, may have implications for subsequent receptor–target protein interactions and the control of gene expression (see below). In addition to half-site recognition, there is evidence from DNA selection studies that nucleotides in the flanking and spacer sequences may also play a role in DNA binding and response element selection (Nelson et al. 1999). Thus, it will be important to solve the three-dimensional structure of the AR–DBD dimer bound to a typical HRE and to one or more selective ARE sequences to determine the binding configuration of the receptor and the specific amino acid–DNA contacts.

Robins and co-workers have argued that the selective response of the AR depends not only on the DNA architecture of the response sequence but also on the presence of non-receptor proteins and possible communication between receptor domains (see Gonzalez & Robins 2001). Thus, Grad et al. (2001) showed that the AR–NTD, not the DBD, is responsible for the AR-specific response from the exonic enhancer located within the AR gene (see Fig. 2). The question of whether this is a special case, restricted to autoregulation of the AR gene, or a more general mechanism awaits further analysis of AR selective response elements. The conclusion from these studies is that multiple mechanisms, alone or in combination, are likely to be involved in ensuring specificity for AR-dependent gene regulation.

**Transcriptional activation: structure of the AF1 domain**

The AR–NTD is characterised by a number of amino acid repeat sequences, including poly-glutamine (Q), poly-glycine (G) and poly-proline (P). The largest poly-Q repeat and a poly-G repeat have been found to be polymorphic, alteration in length being correlated with

---

\(^1\)The numbering for amino acids in the AR protein used in this paper is based on poly-Q and poly-G repeats of 20 and 16 residues respectively.
different disease conditions (see McEwan 2001 and references therein). Expansion of this poly-Q repeat to greater than 40 residues results in spinal bulbar muscular atrophy, a neurodegenerative condition associated with selective neuronal cell death in brainstem and spinal cord (reviewed in McEwan 2001). This change leads to a reduction in the ability of the full-length AR or the isolated NTD to activate transcription. More modest alterations in the length of the poly-Q has, controversially, been associated with an increased risk of prostate cancer (shortened repeats), infertility (longer repeats) and, more recently, male patterned baldness (shortened repeats) (Correa-Cerro et al. 1999, Edwards et al. 1999, Ellis et al. 2001, McEwan 2001 and references therein).

Analysis of the amino acid sequence of the AR–NTD from a diverse range of organisms has revealed the presence of three areas of sequence conservation: amino acids 1–30, 224–258 and 500–541. The first 30 amino acids of the NTD have been shown to be important for the interaction with the AR–LBD. In particular, the sequence 23FqnLF27 and flanking residues are critical for the N/C-repeat (reviewed in McEwan 2001). This change leads to a conformational change in the presence of the binding partner TATA-binding protein (TBP). The AR–AF1 domain (amino acids 142–485) (Simental et al. 1999, 1997, 1998), PR-AF3 (transactivation function unique to the B-isoform of the PR) (Tung et al. 2001) and AR-AF1 (Betney & McEwan 2003). The ligand-dependent AF2 in the LBD is at best weak in the AR, compared with other members of the nuclear receptor superfamily, and the main determinant(s) for receptor-dependent transcriptional activation reside(s) within the AF1 domain (amino acids 142–485) (Simental et al. 1991, Jenster et al. 1995). The AR–TAD has a modular structure, as revealed by deletion and point mutations and the use of NTD-fusion proteins (Jenster et al. 1991, 1995, Simental et al. 1991, Chamberlain et al. 1996). These studies have defined amino acids 142–485 as the AR–AF1 domain, which comprises both the TAU-1 (amino acids 101–360) and TAU-5 (amino acids 360–485) regions originally identified by Brinkman and co-workers (Jenster et al. 1995). When fused to the LexA–DBD, this region retained at least 70% of the activity of the full-length AR–NTD; importantly, a double point mutation that significantly reduced the activity of the full-length AR had an identical effect on the isolated AR–AF1 domain (Reid et al. 2002a).

Although this region is clearly important for receptor-dependent transactivation, there is, as yet, no atomic resolution structure available. We have therefore used a combination of optical spectroscopy approaches (circular dichroism (CD) and fluorescence spectroscopy), together with secondary structure predictions and sensitivity to partial protease digestion, to investigate the folding and conformation of the AR–AF1 domain. These studies generally revealed that AR–AF1 lacked stable secondary structure in aqueous solution, with a measured α-helix content of 13% and β-sheet of 20% (Reid et al. 2002a). Folding of the transactivation domain polypeptide was observed in the presence of the structure-stabilising solute trifluoroethanol (TFE), with a transition to a more α-helical structure at the expense of β-turn and the ‘other’ structure, which was interpreted as being mainly random coil (Reid et al. 2002a). Local folding of the AF1 domain was observed by measuring the intrinsic fluorescence from tryptophans 396 and 432 and the adoption of a protease-resistant conformation in the presence of TFE or the natural osmolyte trimethylamine N-oxide (TMAO). Significantly, a similar protease-resistant conformation was observed upon binding of the target protein TFIIIF, supporting an induced-protein folding model for the AR–AF1 domain structure (Reid et al. 2002a).

While the NTD of nuclear receptors share little, if any, amino acid sequence homology, they do share some structural characteristics. Thus, hydrophobicity has been shown to be important for transcriptional activation of GR-AF1 (Almåsf et al. 1997, 1998), PR-AF3 (transactivation function unique to the B-isoform of the PR) (Tung et al. 2001) and AR-AF1 (Betney & McEwan 2003). Like the AR, the AF1 domains of the glucocorticoid receptor (Dahlman-Wright et al. 1995, Baskakov et al. 1999) and the peroxisome proliferator-activated receptor α (Hi et al. 1999) lack stable secondary structure in aqueous solution but adopt a more α-helical conformation in the presence of the structure-stabilising solutes TFE and TMAO. The NTD of the oestrogen receptor (ER) α and β isoforms are similarly unstructured in an aqueous environment, but, interestingly, the ERα-NTD undergoes a conformational change in the presence of the binding partner TATA-binding protein (Wärnmark et al. 2000). Horwitz and co-workers have extensively studied the biophysical and functional properties of the N-terminal domains of the A and B isoforms of the progesterone receptor (PR) (reviewed in Takimoto et al. 2003). The PR-NTD is largely unstructured in the absence of the DBD. However, intramolecular interactions with the DBD induce a more ordered structure, which is further stabilised by DNA
binding. It can be concluded from these studies that an N-terminal transactivation domain of nuclear receptors is structurally flexible and is likely to adopt a more structured conformation upon intra- and/or intermolecular protein–protein interactions.

Protein–protein interactions

In order to regulate transcription, two general steps must be accomplished: (1) the remodelling of chromatin structure to open up regulatory regions and the promoter and (2) the recruitment of the general transcription machinery to the promoter to enhance transcription initiation and/or elongation. The AR can potentially regulate both these steps, leading to an increase in the levels of target gene mRNA. Recently, evidence for the hormone-dependent remodelling of the chromatin at the MMTV promoter by the AR has been presented (List et al. 1999, Huang et al. 2003). Huang et al. (2003) further demonstrated the importance of BRG-1, the ATPase subunit of the SWI/SNF chromatin remodelling complex, for AR-dependent alterations in DNA topology and gene activation. Recruitment of the SWI/SNF complex was thought to be indirect via interactions with the coactivators SRC and CREB-binding protein (CBP), with the histone acetyl transferase (HAT) activity of the latter important for receptor-dependent activation and the stabilisation of the interaction with SWI/SNF (Huang et al. 2003). Although this study showed an increase in histone acetylation, it is worth noting that three lysine residues, within the NLS, are acetylated by the coactivator proteins p300 and p300/CBP-associated factor (P/CAF) (Fu et al. 2000) and Tip60 (Gaughan et al. 2002). Consequently, acetylation may act to modulate AR activity directly as well as through opening the chromatin structure at receptor-target genes. Indeed, Fu et al. (2003) recently reported that acetylation of the AR enhanced binding of the p300 coactivator protein. Figure 3 shows sites of post-translational modification within the AR protein, including phosphorylation (van Laar et al. 1991, Gioeli et al. 2002), acetylation (Fu et al. 2000, Gaughan et al. 2002) and sumoylation (Poukka et al. 2000). It is striking that the majority of identified phosphorylation sites map to the AR–NTD and the AF-1 region (Gioeli et al. 2002), suggesting that this modification may modulate receptor-dependent transactivation directly. Possible mechanisms could involve altering protein–protein interactions and/or alterations in protein structure and stability.

Since the first report of a direct interaction of the AR–AF1 domain with TFIIF, a component of the general transcription machinery (McEwan & Gustafsson 1997), a number of protein targets have been found to bind to the AR–NTD (Table 1). TFIIF is a heterotetramer of RAP30 and RAP74 subunits, and binding sites for the receptor have been mapped to the N- and C-terminal globular domains of the large subunit (Reid et al. 2002b). TFIIF is involved at several distinct steps of the transcription cycle: preinitiation complex formation, promoter escape and transcription elongation by RNA polymerase II. We have therefore argued that the AR may regulate target genes at multiple steps during transcription initiation and elongation, in part, by binding to TFIIF (McEwan & Gustafsson 1997, Reid et al. 2002b). Results from Chang and co-workers (Lee et al. 2000, 2001) demonstrating the interaction of the receptor with subunits of TFIIH and the elongation factor P-TEFb, factors involved in the early stages of transcription initiation and elongation, strongly suggest that the AR can indeed act at multiple steps to enhance transcription.

TFIIF, TFIIH and P-TEFb are components of the general transcription machinery and are unlikely therefore to explain tissue and/or gene specific regulation by the AR. In addition to TFIIF, the binding sites for p160 coactivators (Bevan et al. 1999, Reid et al. 2002b), cell-cycle regulatory protein cyclin E1 (Yamamoto et al.

Figure 3 Post-translational modifications of the AR protein. Sites of phosphorylation (P), sumoylation (SUMO-1) and acetylation (Ac) are shown. See text for details.
<table>
<thead>
<tr>
<th>Protein</th>
<th>Receptor binding site</th>
<th>Comments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General transcription machinery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFIIF</td>
<td>141 to 485</td>
<td>TFIIF is a heterotetramer of RAP30 and RAP74 subunits involved in transcription initiation and elongation. AR binding maps to the CTD of the large subunit RAP74</td>
<td>1, 2</td>
</tr>
<tr>
<td>TFIIF</td>
<td>38 to 561</td>
<td>Binding mapped to Cdk7/cyclin H subunits of CAK; involved in phosphorylating the CTD of the large subunit of RNA PolII</td>
<td>3</td>
</tr>
<tr>
<td>P-TEFb</td>
<td>38 to 634 (incl. DBD)</td>
<td>Involved in the transition of the RNA PolII to elongating form: phosphorylates the CTD of the large subunit of RNA PolII enzyme</td>
<td>4</td>
</tr>
<tr>
<td><strong>Co-activators</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARA160</td>
<td>38 to 566</td>
<td>1093 amino acid protein; alternative name “TMF”, TATA-element modulating factor</td>
<td>5</td>
</tr>
<tr>
<td>ART27</td>
<td>153 to 336</td>
<td>Novel coactivator of 157 amino acids, identified in two-hybrid screen</td>
<td>6</td>
</tr>
<tr>
<td>CBP</td>
<td>1 to 566, 297 to 640 (incl. DBD)</td>
<td>CREB-binding protein; AR binding site mapped to amino acids 271 to 452</td>
<td>7, 8</td>
</tr>
<tr>
<td>p160 family</td>
<td>1 to 556 (360 to 494?)</td>
<td>Family of co-activator proteins (SRC-1/TIF2/ACTR) originally identified binding to LBD of nuclear receptor via LxxLL motifs</td>
<td>9–11</td>
</tr>
<tr>
<td>AES</td>
<td>1 to 559</td>
<td>‘Amino terminal enhancer of split’, member of the Groucho/TLE family of co-repressors</td>
<td>12</td>
</tr>
<tr>
<td><strong>Co-repressors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMRT</td>
<td>171 to 505 (171 to 328)</td>
<td>270 kDa protein originally identified binding to TR and RAR in absence of ligand and acting as a co-repressor</td>
<td>13</td>
</tr>
<tr>
<td>SMAD3</td>
<td>330 to 563 (DBD–LBD)</td>
<td>Transcription factor, involved in TGFβ signalling</td>
<td>14</td>
</tr>
<tr>
<td><strong>Transcription factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>234 to 558</td>
<td>Transcription factor mediating signalling by the IL-6 family of cytokines</td>
<td>15</td>
</tr>
<tr>
<td>ANT-1</td>
<td>1 to 532</td>
<td>p102 US small nuclear ribonucleoprotein particle-binding protein; human homologue of yeast splicing factor Prp6</td>
<td>16</td>
</tr>
<tr>
<td><strong>Co-regulatory-proteins</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARNIP</td>
<td>11 to 172</td>
<td>30 kDa RING domain (C3H2C3 Zn finger) protein with E3 ubiquitin ligase activity; inhibits AR-N/C interaction</td>
<td>17</td>
</tr>
<tr>
<td>BRCA-1</td>
<td>1 to 555</td>
<td>Breast cancer susceptibility gene 1, a tumour suppressor gene; also interacts with p160 family member GRIP1 (TIF2)</td>
<td>18, 19</td>
</tr>
<tr>
<td>Caveolin-1</td>
<td>1 to 501</td>
<td>Transient interaction; co-sedimentation with the AR in lipid-rafts; may serve as a scaffolding protein for signalling molecules</td>
<td>20</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>1 to 502</td>
<td>Role in AR signalling appears to be independent of its role with Cdk4 in cell-cycle regulation</td>
<td>21</td>
</tr>
<tr>
<td>Cyclin E</td>
<td>419 to 556</td>
<td>Complex with Cdk2 not required for enhancement of AR signalling</td>
<td>22</td>
</tr>
<tr>
<td>pRb</td>
<td>502 to 565</td>
<td>Product of the retinoblastoma tumour suppressor gene; involved in cell-cycle regulation. Putative pRb binding motif: LxCxD</td>
<td>23</td>
</tr>
</tbody>
</table>

McEwan: AR transactivation domain

2000), the transcription factors SMAD3 (Hayes et al. 2001) and STAT3 (Ueda et al. 2002), the novel coactivator ART-27 (Markus et al. 2002) and the corepressor SMRT (Dotzlaw et al. 2002) have been mapped to the AF1 domain. The binding site for p160 steroid receptor coactivators includes amino acids 360-494, although additional regions of the AR-NTD may contribute to this interaction (Bevan et al. 1999). The binding site for the AR-AF1 was mapped to the C-terminal region of SRC-1a (Reid et al. 2002b), in good agreement with studies using the AR-NTD (amino acids 1-556) that have highlighted the Q-rich domain of p160 coactivators (Bevan et al. 1999, Christiaens et al. 2002). A number of point mutations have been introduced into the AR-AF1 that disrupt function and/or TFIIF binding (Reid et al. 2002b, Betney & McEwan 2003). None of the mutations described disrupted the binding of SRC-1a, indicating that the binding site was likely to be distinct from that for TFIIF. Sequences within the C-terminal of AR-AF1 are also likely to be important for SMAD3 (amino acids 333-563) and STAT3 (amino acids 234-538) binding (Hayes et al. 2001, Ueda et al. 2002). ART-27 is a novel 157 amino acid protein identified in a yeast two-hybrid screen, and it has been shown to interact with amino acids 153-366 in the AR-NTD (Markus et al. 2002). The binding site for the corepressor SMRT has recently been mapped to the same region (amino acids 171-328), although more C-terminal sequences also play a role in binding (Dotzlaw et al. 2002). Thus, the binding sites for ART-27 and SMRT within the AR-transactivation domain potentially overlap with each other and with TFIIF, and it will be interesting to determine whether these proteins compete for receptor binding.

Overall, AR-dependent gene regulation will be the consequence of cooperation or synergism with other transcription factors (including tissue-specific proteins) and multiple interactions with the transcription machinery. Therefore, it is significant that the recruitment of protein complexes by the DNA-bound AR, containing components of the general transcriptional machinery (RNA polymerase II) and coactivators (p160, CBP and TRAP220), has been elegantly demonstrated recently by chromatin immunoprecipitation (ChIP) assays (Kang et al. 2002, Shang et al. 2002, Wang et al. 2002, Louie et al. 2003).

Interplay between protein conformation and protein–protein interactions

Binding of an agonist causes a conformational change in the LBD that results in a more compact structure and a realignment of helix 12 (Matias et al. 2000). Helix 12 consists of a conserved motif that contributes the core of the ligand-dependent activation function AF-2, and, together with residues in helices 3, 4 and 5, forms a binding pocket on the surface of the LBD that allows for the docking of a number of coactivator proteins via LxxLL motifs (where L represents leucine and X is any amino acid) (Darimont et al. 1998, Mcinerney et al. 1998). Mutating conserved hydrophobic residues within helix 12 disrupts the function of the full-length AR, while the isolated LBD interacts in a hormone-dependent manner with coactivator proteins in two-hybrid studies (see Bevan et al. 1999). However, demonstrating an independent AF2 activity for the AR has been difficult; at best, this represents a weak transactivation function. Consistent with this are the observations that the primary binding site for p160 coactivators maps to the AR-NTD/AF-1, and not the LBD (see above). In contrast, ligand-induced conformational changes within the LBD are important for the NTD/CTD interaction, which involves the first 21 amino acids of the receptor and residues within AF2 (He & Wilson 2002, Steketee et al. 2002).

Analysis of point mutations introduced into the AR-AF1 domain revealed two mutant polypeptides, M3 (M244A, L246A and V248A) and M6 (S159A and S161A), with reduced binding to TFIIF (RAP74) (Reid et al. 2002b, Betney & McEwan 2003). Interestingly, AF1-M3 and M6 do not share the same conformational properties. The mutations in M6 were in a six amino acid repeat sequence (PSTLSL), and the changes were predicted to alter the secondary structure of this sequence. In contrast, to both the wild-type and M3 polypeptides, AF1-M6 showed a more structured conformation, with the tryptophan residues being less surface exposed and the polypeptide showing an overall increase in protease resistance (R Betney & IJ McEwan 2003). We conclude from these studies that the hydrophobic residues mutated in AF1-M3 (M244A, L246A and V248A) may form part of the binding surface for TFIIF, while the residues mutated in M6 (S159A and S161A) have an indirect effect on binding by altering the structural flexibility of this domain.

Structural analysis of AR-AF1 has suggested that binding of TFIIF (RAP74) leads to a more structured conformation of this region, and that this may involve an increase in a-helix content, while comparison of the mutant polypeptides M3 and M6 suggests that TFIIF binding is sensitive to the flexible conformation of the AF1. Interestingly, in a ‘GST-pull down assay’, folding of the AR-AF1 domain by preincubating with TMAO or TFIIF (RAP74) resulted in a dramatic increase in the binding of the coactivator SRC-1a (Kumar et al. In press). The result was specific for TFIIF, as incubation with a non-interacting protein had little effect on SRC-1 binding. Similar findings have recently been reported for the GR-
AF1 (τ1) domain, where folding with TMAO significantly enhanced protein–protein interactions with the general transcription factor TBP and the coactivators CBP and SRC-1 (Kumar et al. 2001). These findings have led us to propose the following working model for the AR–AF1: protein–protein interactions result in induced folding of the TAD, which in turn leads to the generation of surfaces for further interactions and the assembly of a transcriptionally competent receptor complex (Fig. 4). The folding of the AR–AF1 in response to TFIIF-binding suggests that this is likely to be a high-affinity interaction that is enthalpy driven (see Wright & Dyson 1999). Thus, it will be important to determine what other interactions lead to a folding of AF1 and what the relative affinities of the different interactions are.

Conclusions and future work

The last five years has seen a dramatic increase in information that has helped increase our understanding of the molecular details of AR-dependent gene regulation. This has included the identification of binding partners that are a direct link to the transcriptional machinery or serve as bridging factors for the DNA-bound receptor (coactivators) as well as an increasing number of co-regulatory proteins. It will be important to investigate the binding affinity and the thermodynamic profile of these different AR–protein interactions and to correlate this binding with what occurs at androgen-regulated genes. In addition, it will be important to correlate the role of post-translational modifications (phosphorylation, acetylation and sumoylation) in receptor function, including protein–protein interactions. Recent studies have also started to address the question of the structure of the AR–NTD and AF1 domain and to examine the folding of the transactivation domain in response to specific protein–protein interactions and the consequences of this for the assembly of transcriptionally competent complexes. The determination of the three-dimensional structure of this domain will be a major achievement and will undoubtedly advance our understanding of the mechanism of AR-dependent gene regulation and help correlate mutations with structure and function. Ultimately, it is to be hoped that an increased understanding of the different steps in AR signalling will lead to improved or novel diagnostic and therapeutic strategies for combating AR-dependent disorders such as prostate cancer.

Acknowledgements

I would like to thank Dr Alasdair MacKenzie (University of Aberdeen) for help with navigating the Ensembl Genome Browser. Work in the author’s laboratory is supported by the Association for International Cancer Research, the Biotechnology and Biological Sciences Research Council, the Medical Research Council and the James Alexander Mairns Trust.

References

Almlof T, Wallberg AE, Gustafsson J-Å & Wright AP 1998 Role of important hydrophobic amino acids in the interaction between the glucocorticoid receptor tau 1-core activation domain and target factors. Biochemistry 37 9586–9594.
Baskakov IV, Kumar R, Srinivasan GJYS, Bolen DW & Thompson EB 1999 Trimethylamine N-oxide-induced


Dai JL & Burnstein KL 1996 Two androgen response elements in the androgen receptor coding region are required for cell-specific up-regulation of receptor messenger RNA. *Molecular Endocrinology* **10** 1582–1594.


Gaughan L, Logan IR, Cook S, Neal DE & Robson CN 2002 Tip60 and histone deacetylase 1 regulate androgen receptor activity through changes to the acetylation status of the receptor. *Journal of Biological Chemistry* **277** 25904–25913.


Gonzalez MI & Robins DM 2001 Oct-1 preferentially interacts with androgen receptor in a DNA-dependent manner that facilitates recruitment of SRC-1. Journal of Biological Chemistry 276 6420–6428.

Grad JM, Dai JL, Wu S & Burnstein KL 1999 Multiple androgen response elements and a Myc consensus site in the androgen receptor (AR) coding region are involved in androgen-mediated up-regulation of AR messenger RNA. Molecular Endocrinology 13 1896–1911.


Lee DK, Duan HO & Chang C 2001 Androgen receptor interacts with the positive elongation factor P-TEFb and enhances the efficiency of transcriptional elongation. Journal of Biological Chemistry 276 9978–9984.


McEwan IJ & Gustafsson J-A 1997 Interaction of the human androgen receptor transactivation function with the general transcription factor TFIIH. PNAS 94 8485–8490.


www.endocrinology.org


