3. DISCUSSION

3.1. *Pem* and *OTEX are* androgen-regulated homeobox genes belonging to the *PEPP* subfamily

3.1.1. Characterization of the androgen-regulated murine Pem promoter

Expression of the murine *Pem* gene is regulated by androgens in the epididymis and testis (Sutton et al., 1998; Rao et al., 2002). The *Pem* promoter was cloned and analyzed by cell-based transactivation tests. A strong androgen response was observed in CV-1 cells transiently transfected with the AR. These effects were androgen preferential as progestins and especially glucocorticoids elicited less response in CV-1 cells transfected with the corresponding receptor. This is in line with the *in vivo* situation since the *Pem* gene is mainly transcribed in the testis and epididymis under androgen control, and in the ovary with possibly an implication of progesterone. The preferential stimulation by R1881 suggested the presence of androgen-selective DNA response elements in the mouse *Pem* promoter. A sequence analysis revealed the presence of related motifs, of which two formed specific complexes with the AR in the EMSA (Barbulescu et al., 2001). They were entirely conserved in the rat *Pem* promoter, suggesting an important regulatory role. An alignment of the *Pem* elements with the consensus SRE showed that they did not fit the classical inverted repeat with a three nucleotide-spacing model (Table 2). Pem ARE-1 was composed of two half-sites including the G and C contact bases but separated by five nucleotides which is unusual. Pem ARE-2 was composed of two half-sites spaced by three nucleotides in which four out of six possible positions were maintained in the direct repeat mode (Table 2).

The functionality of these elements was confirmed in transactivation studies. Alteration of Pem ARE-1 or Pem ARE-2 led to a marked loss of the androgen response of the *Pem* promoter. As the Pem ARE-2 mutation had more impact, it is most likely determinant in the androgen control of the *Pem* promoter, in line with its stronger AR-binding properties in the EMSA as a single element (Barbulescu et al., 2001). The additional presence of Pem ARE-1 appeared however necessary to yield the full response, suggesting cooperative interactions between both elements. Indeed Barbulescu et al. (2001) reported a cooperative binding of the tandem elements in EMSA. These data support the notion that synergy between weak AREs

is the basis for a strong, selective androgen response (Scheller et al., 1998). It is also worth noting that Pem ARE-1 overlaps a potential initiator element, which might have implications on how *Pem* expression is controlled *in vivo*. In fact, the *Pem* gene is one of the very few which are both androgen-regulated and TATA-less. Other examples are the *AR* gene itself which is transcriptionally up- or down-regulated by androgens, depending on the cell context (Quarmby et al., 1990; Shan et al., 1990) and the newly discovered human homeobox gene *OTEX* (see below).

3.1.2. Identification of a human member of the PEPP gene family: *OTEX*

In search of a human orthologue of murine *Pem*, a novel human homeobox gene was identified and named *OTEX*, based on its expression in ovary, testis and epididymis. A sequence comparison indicated that *OTEX* belonged to the recently defined *PEPP* subfamily of *paired*-like homeobox genes, of which *Pem* is a founding member. The *PEPP* genes probably derived from a common ancestor related to the *Drosophila aristaless* gene. Several family members are known in mice but only one, *ESXR1*, had been characterized until now in humans. In addition to *OTEX*, *THG1* has also been identified as a new human member of the *PEPP* subfamily and *Ehox* and the RIKEN clone NM_029203 as two new murine members.

The existence of a common ancestor for the *PEPP* genes is supported by several facts. Firstly, all *PEPP* genes are located on chromosome X and, in the case of human members, are clustered in the Xq22-Xq24 region. Secondly, their HD-coding region is interrupted by two introns, as is the case for *aristaless*. This rare intron is not found in the coral *Aquapora millepora PaxD* or in human *SHOX*, despite the comparatively high amino acid sequence identity of the respective HDs. Thirdly, an Arg is found at position 58 of the encoded HD (except for the RIKEN clone NM_029203) and in addition, OTEX shares with Pem, Psx2 and Ehox the presence of a Lys at position 50 which is essential for DNA-binding specificity. Finally, the expression profiles of *PEPP* members partially overlap as they are often detected in ovary, testis and placenta.

The *PEPP* genes are not the closest relatives of *aristaless*. Human *ARIX* (Zellmer et al., 1995), *ALX3* (Wimmer et al., 2002), *ALX4* (Wuyts et al., 2000), as well as *ARX* (accession number AY038071) are 72% to 85% identical in their deduced HD to *aristaless*. *ARX* is located on chromosome X whereas the other paralogues are on chromosomes 1 and 11. *SHOX* is another human homeobox gene mapping to

chromosome X and could have been considered as a possible *PEPP* member. It is expressed in a variety of tissues and human genetic studies have shown that its mutation leads to multiple phenotypes including short stature and wrist deformity (Rao et al., 1997). However in the four human aristaless paralogues (ARIX, ALX3, ALX4, ARX) and in SHOX, position 58 of the HD is occupied by Lys rather than Arg. HD position 50 which is implicated in DNA-binding specificity is a Gln, which is never found in the HDs of PEPP proteins. Therefore the genes coding for these five different proteins are not considered to be members of the *PEPP* family.

A phylogenetic analysis of PEPP HDs using human ARX HD as outgroup reveals the existence of five subgroups in this family (Figure 22): (i) mouse and rat Pem; (ii) murine Psx1 and Psx2; (iii) human THG1, mouse Ehox and the mouse RIKEN clone NM_029203; (iv) human OTEX; (v) human ESXR1 and mouse Esx1. In the phylogenetic tree, OTEX is more distant from Pem than Psx1 and Psx2 are. On the other hand OTEX shares with Pem the presence of a Lys at HD position 50 and has a tissue expression profile closely resembling that of Pem. In addition, both the *OTEX* and *Pem* genes are under androgen control and are localized on syntenic human and mouse regions of chromosome X. Altogether this suggests that they may have comparable functions despite the low sequence conservation and therefore could be orthologues. Indeed, the *PEPP* genes have a high propensity to diverge (Maiti et al., 1996b), which is also documented by the limited sequence identity seen outside the HD. This suggests that PEPP proteins may engage interactions with different partners to modulate biological events. A rapid evolution of proteins of the reproductive system has been demonstrated in other species (Swanson et al., 2001) and might build up species barriers which are of importance in the process of speciation in the animal kingdom.

3.1.3. Androgen regulation of human OTEX

OTEX expression was found to be under androgen control in a prostate carcinoma cell line expressing the AR. The *OTEX* upstream region shows little sequence conservation with the rat or mouse *Pem* promoter. The rodent *Pem* promoters display very high DNA sequence identity and the functional AREs identified in the mouse are entirely conserved in the rat. This was not the case for the *OTEX* promoter in which the mouse AREs were not maintained. Potential degenerated SREs could be found at other locations but the determination of their functionality awaits further analysis.

Altogether the tissue distribution profile and the androgenic regulation in PC3/ARwt cells make it likely that similarly to murine and rat *Pem*, AR and PR directly control the expression of the human *OTEX* gene.

3.1.4. Possible function of OTEX

Based on its sequence similarity to homeobox proteins especially in the HD, OTEX is likely to act as a transcription factor. An exclusively nuclear localization was evidenced by immunofluorescence experiments and mutational analysis identified the region responsible for nuclear import. Several residues of this region were maintained among most PEPP proteins, suggesting that each possessed an NLS at the C-terminal end of their HD. This does not necessarily imply that all of them are nuclear proteins and Esx1, for instance, was primarily detected in the cytoplasm, due to the presence of an inhibitory motif C-terminal of its HD (Yan et al., 2000). On the other hand, experimental evidence shows a functional NLS to exist at the C-terminal end of the HD of the human PDX-1 and mouse Crx homeobox proteins (Moede et al., 1999; Fey et al., 2000).

No inherent transactivation activity could be determined for OTEX in a one-hybrid assay, suggesting that additional factors may be necessary. Like for Esx1 (Yan et al., 2000), no evidence for homodimer formation of OTEX was obtained. It can however not entirely be ruled out as presence of the DNA-binding element might be necessary for this to take place. Alternatively, a heterodimeric partner or an additional stimulus (e.g. phosphorylation event) might be essential. Another possibility is that OTEX is a transcriptional repressor, as described for other paired-like homeobox proteins (Norris et al., 2001; Dasen et al., 2001). The repression of proliferation of CV-1 cells transiently expressing OTEX is in line with this. It might also be that OTEX plays a role in cell differentiation, as is the case for many homeobox proteins, and thereby reduces proliferation. Indeed tumor suppressor activity has been reported for the homeobox genes *CDX2*, *NKX3.1* and *BARX2* (Abate-Shen, 2002). Overexpression of these three genes reduces proliferation of transformed cell lines as seen for *OTEX* (Abate-Shen, 2002). Additional studies are under way in order to identify the DNA response elements recognized by OTEX as a first step towards identifying its target genes and elucidating its function.

To date not much is known about the biological function of the other PEPP family members in mice and humans. The generation of mice deficient for Pem or Psx2 was

not conclusive as only a limited change was observed, suggesting subtle phenotypic effects that cannot be detected under usual rearing conditions. Alternatively, this might point to redundant functions, in line with the overlapping expression patterns of some *PEPP* family members. Redundancy has already been evidenced for the homeobox genes *Alx3* and *Alx4*, *Nkx2.5* and *Nkx2.6*, and *Hoxa-9* and *Hoxd-9* for which severe phenotypic alterations have only been detected in double knock-out mice (Beverdam et al., 2001; Fromental-Ramin et al., 1996; Tanaka et al., 2001).

The remarkably narrow tissue distribution of *OTEX* and the other *PEPP* members in steroid hormone target organs points to a role in reproduction. *Pem*, *Esx1* and *ESXR1* are all expressed in the testis as is *THG1* (personal observation), and *Pem* is additionally expressed in the epididymis. *OTEX* and *Pem* transcripts are also expressed in the ovary and, interestingly, a form of premature ovarian failure (POF1) maps to the chromosomal region where the three human *PEPP* genes are located (Krauss et al., 1987). The expression of *OTEX* in the ovary and mammary gland furthermore suggests that female sexual hormones, e.g. progestins may also be involved in the control of this gene.

In the probable absence of a direct orthologue of OTEX in mice and due to the possible functional overlap of the PEPP members, the elucidation of the biological role of OTEX in processes as diverse as human reproduction, embryonic development and possibly carcinogenesis will represent a challenging task.

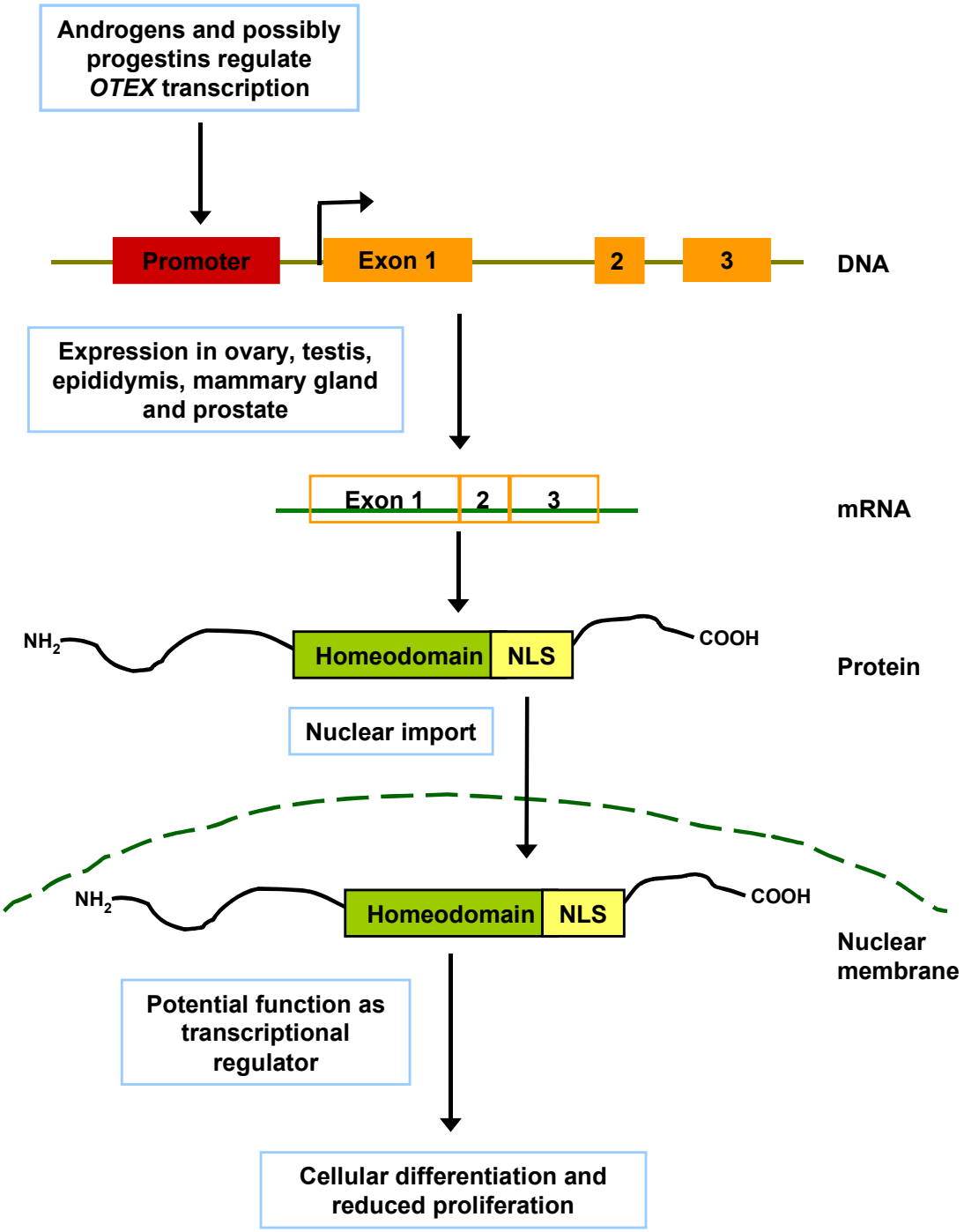

Figure 30: Schematical representation of OTEX regulation and action. *See text for discussion.*

3.2. DNA response elements are modulators of AR function

A long-standing question in gene regulation by steroids is how receptors that recognize the same SRE can specifically regulate different sets of target genes. One possibility is that unique variations in the sequence of the consensus SRE are instrumental for receptor specificity.

Here it was demonstrated that DNA elements do not merely position the AR receptor to the promoter but represent, beside the ligand and the receptor-associated proteins, additional interacting partners with important regulatory functions. The androgen-selective AREs and the non-selective SREs represent two classes of response elements recognized by the AR for stimulation of target genes. Using mutational analysis, Barbulescu et al. (2001) have previously shown the role of the DNA half-sites in AR binding to vary, depending on the element. Evidence is presented here that the two element classes differently modulate AR function.

The Pem AREs mediate a much stronger androgen response in transactivation assays than the SREs, despite being less well bound by the AR. The basis for selectivity is therefore not enhanced binding to DNA elements but rather altered intra- and inter-domain interactions as well as differential cofactor recruitment. A model in which DNA elements impart conformational modifications onto bound AR thus leading to such novel interactions may therefore be postulated (Figure 31).

3.2.1. DNA response elements alter AR conformation

Limited protease treatments of DNA/AR complexes showed that there were differences in accessibility to chymotrypsin and trypsin when the AR was attached to different DNA elements. In the chymotrypsin experiments the ARE-associated AR displayed a higher accessibility than the SRE-bound AR. Trypsin digestions also discriminated between the two element classes, even though Pem ARE-2 had an intermediary profile. The cleavage sites were difficult to localize precisely, but all were in the AR N-terminal domain. The C1 band was generated through cutting near the N-terminal end and the T1 and T2 bands through cutting close to the C-terminal end of the AR N-terminal domain. This suggests that the DNA elements induced changes outside of the DBD and that inter-domain communications existed in the AR. A further possibility is that different sets of coregulators were recruited by the DNA/AR complexes, thereby altering protease accessibility.

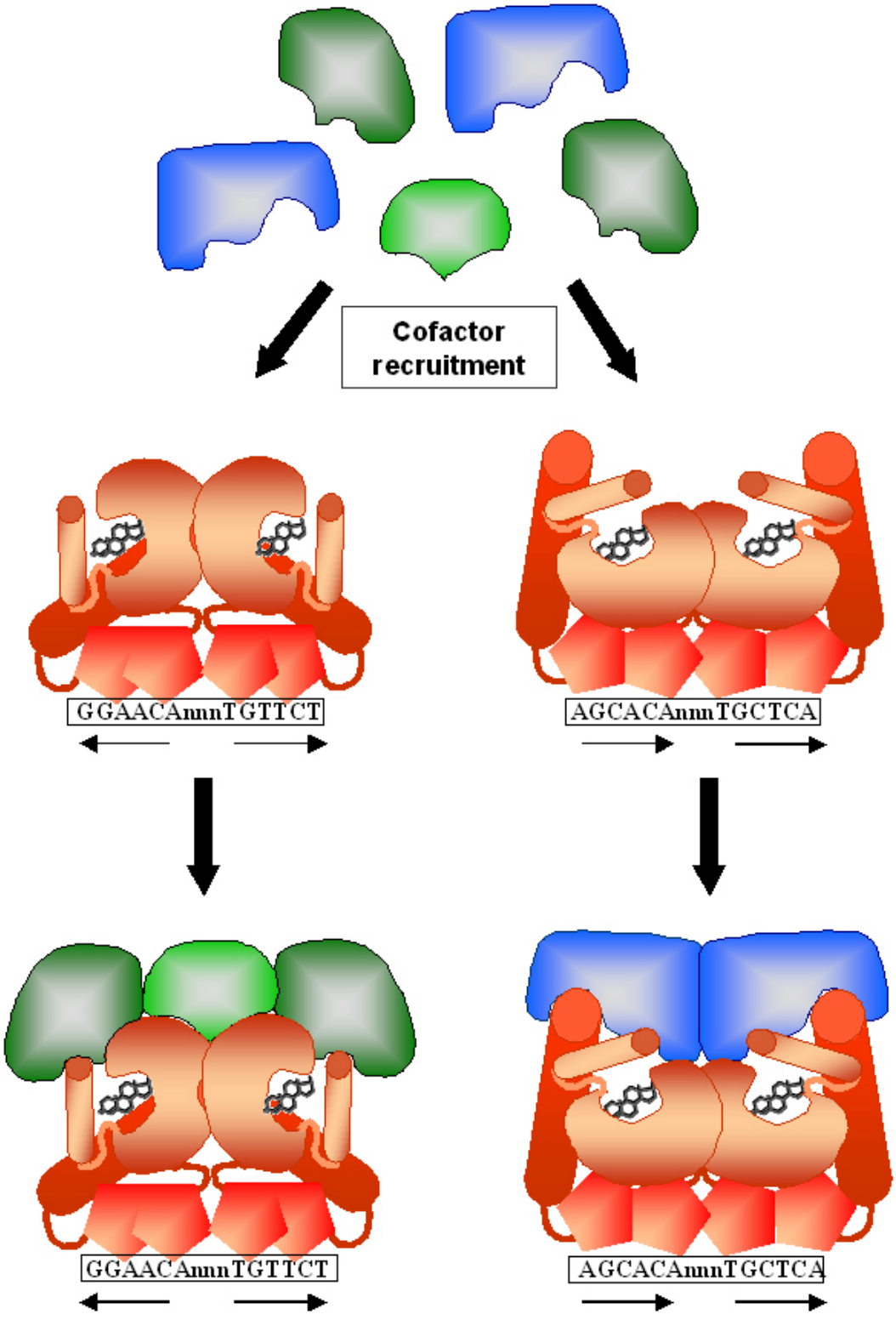

Figure 31: Model of DNA element-induced conformational changes onto the AR. *The two classes of response elements, SREs (left) and AREs (right) induce distinct structures of the AR homodimer (in red). The element-modulated receptors recruit different sets of cofactors (green and blue), which leads to distinct transcriptional responses.*

3.2.2. The role of the DBD dimerization interface differs depending on the DNA element

The AR D box is part of the dimerization interface of the DBD. Disruption of salt bridges by mutating specific residues in this region of the GR or MR receptor leads to enhanced activity on multiple SREs, indicating a role in restraining transcriptional synergy rather than in stabilizing the receptor homodimer (Liu et al., 1996; Chen et al., 1997). The equivalent mutations of the AR, R598D and D600R, increased its activity mainly on the non-selective SREs. Far less (R598D) or no effects (D600R) were observed when testing the selective Pem AREs. This and the studies comparing single and tandem repeats of response elements document that self-synergy effects varied depending on response elements and opens up the possibility that control of AR activity may be achieved at this level.

Another residue of the AR D box likely to play a role in receptor dimerization is Ala at position 596. The A596T mutation has been identified in the AR of patients suffering from Reifenstein syndrome, a form of partial androgen insensitivity (Kaspar et al., 1993). Due to its inability to bind, the mutant receptor cannot transactivate promoters containing a single response element. It is however fully active on promoters with multiple response elements (Kaspar et al., 1993). AR A596T possessed a reduced transactivation potential towards the selective AREs. The situation was not the same for the SREs. Little change was noted for the CRISP-1 SRE and a much enhanced activity for the 2.43 SRE. This indicated that here also, disruption of the dimerization interface had very different effects on AR function, depending on the response elements analyzed. Even though the results were biased by the lower expression of the mutant AR as compared to the wild-type form, it can be concluded that all the selective AREs behaved similarly. Interestingly, all D box mutations tested had the strongest stimulating effects on transactivation potential mediated by the 2.43 SRE, which was originally identified as the optimal AR binding sequence using an *in vitro* selection procedure (Roche et al., 1992). Conversely, the weak AR binders generally responded less or even negatively to disruption of the dimerization interface. Altogether, the results show the effects of the mutations to be context-dependent and are in line with an influence of DNA elements on the contacts formed within the dimer interface. Previous studies on the ER and GR have shown this interface to form only after DNA binding, implying that its conformation is not fixed beforehand (Schwabe et al., 1995; Hard et al., 1990). By affecting the dimerization interface, DNA elements

might furthermore allosterically influence the position and activity of the synergy control motifs described in the AR N-terminal domain (Iniguez-Lluhi and Pearce, 2000). These motifs are operative only when the dimerization interface is intact and have recently been shown to harbor sites for sumoylation (Poukka et al., 1999). This opens the possibility that dimerization and sumoylation are interconnected mechanisms controlling self-synergy.

3.2.3. DNA elements differentially transmit cofactor effects

The changes in AR conformation brought about by DNA elements might in turn affect the recruitment and activity of coregulatory proteins. TIF2 and ARA55 have been identified as potent AR cofactors (Haendler, 2002; Heinlein et al., 2002). Their ectopic expression was found to enhance AR activity stimulated by 0.1 nM and higher androgen concentrations in transactivation assays. The peak noted at 0.1 nM for TIF2 overexpression showed the role of this coactivator to be more important at low than at high ligand concentrations. In absolute terms, the effects of both cofactors were much more pronounced on the AREs than on the SREs, with Pem ARE-1 exhibiting the strongest response to ectopic expression of both TIF2 and ARA55. Interestingly, the cofactor-enhanced response of the SREs attained about the levels reached by AREs without added cofactors. This demonstrates that the respective strengths of response elements may be dramatically influenced by cofactor activity. The fact that the effects of cofactors known to associate with the AR N-terminal region or LBD are affected by the response element recognized by the DBD indicates that indeed information originating from DNA elements can be transmitted to distal domains. Since the contact interface between nuclear receptors and cofactors is generally small (Feng et al., 1998), it is conceivable that even limited conformational changes affected recruitment of regulatory proteins. An influence of DNA elements on TIF2 association with the ER (Loven et al., 2001; Wood et al., 2001) as well as on SRC-1 interaction with the thyroid hormone receptor (Takeshita et al., 1998) have also been reported. A function of DNA elements in inducing changes in receptor conformation leading to selective cofactor recruitment may thus represent a general mechanism for control of gene expression.

Post-translational modifications are emerging as important regulatory steps in AR function. Among these, the process of sumoylation has been shown to repress AR transactivation (Poukka et al., 2000). Recent data indicate that Ubc9 and the PIAS

proteins are implicated in AR sumoylation as E2 conjugase and E3 ligases, respectively (Kotaja et al., 2002; Poukka et al., 1999). The effects of Ubc9 cotransfection differed, depending on both the hormone concentration and the DNA element used. A parallel enhancement of AR response was seen for the non-selective SREs. For the selective AREs little effect or even repression was observed at higher concentrations of the ligand. Concerning PIASxα cotransfection, no effect with the non-selective SREs was noted but a dramatic repression of AR activity was found when the selective AREs were tested. Altogether the data show that the response to cofactors implicated in protein sumoylation is different for the non selective SREs and the selective AREs. The parallel shift towards repression observed in all curves when comparing the effects of Ubc9 and PIASxα might be explained by the fact that Ubc9 acts upstream of different PIAS proteins, some with repressing and some with stimulating effects on AR activity, indicating a regulatory network to exist (Kotaja et al., 2002; Nishida et al., 2002; Kotaja et al., 2000; Gross et al., 2001). In addition, sumoylation-independent effects of Ubc9 on AR activity have also been described (Poukka et al., 1999), therefore the effects of both enzymes need not necessarily be identical. Finally, sumoylation affects the function of many proteins, including TIF2, leading to complex effects within the cell (Kotaja et al., 2002). The AR sumoylation sites have recently been identified within the synergy control motifs present in the N-terminal region. Mutation of these motifs has effects similar to mutations of the DBD dimer interface on AR activation, suggesting a functional link between both domains (Poukka et al., 2000). The recent discovery that ablation of the AR sumoylation sites leads to loss of recruitment of the corepressor SMRT might explain, at least partly, how control of self-synergy is achieved (Dotzlaw et al., 2002). Finally, the observation that sumoylation has also repressive effects on PR and GR function (Abdel-Hafiz et al., 2002; Tian et al., 2002), suggests this post-translational modification to represent a general mechanism for regulation of steroid receptor function. The finding that DNA elements differentially interpret AR sumoylation adds another level of complexity in transcriptional control.

In summary the data shows that DNA elements convey information to bound AR and act as regulators influencing the conformation of the receptor dimer, the effect of cofactors and of protein-modifying coregulators. As proposed for other transcription factors (Lefstin and Yamamoto, 1998), allosteric modification by DNA may represent an essential mechanism used to regulate the activity of genes in a cell-specific

manner and future studies will show whether modulation of transcription can be achieved at this level.

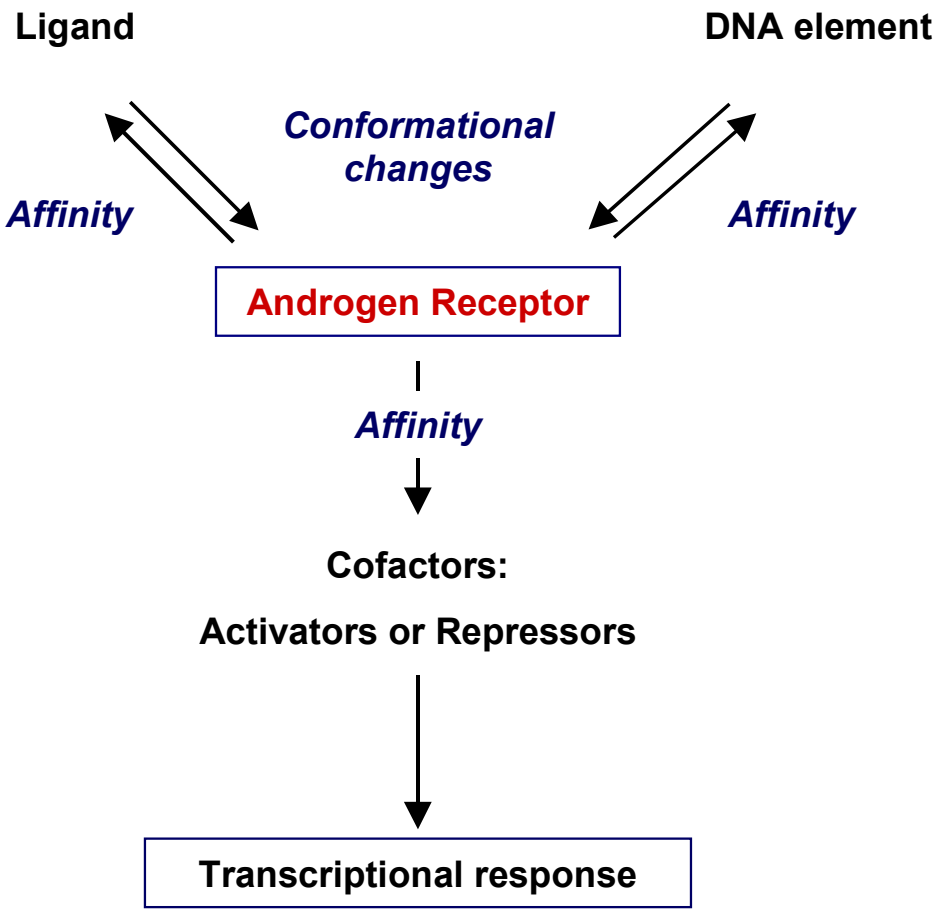

Figure 32: Model of AR function. *The affinity of the AR to a ligand and DNA element determines the probability of its binding to it. Each different ligand or DNA element may potentially induce distinct conformational changes onto the AR. These changes influence the affinity to cofactors. The cofactors act as integrators that translate the structural information into a transcriptional response by their interaction with the transcriptional machinery. The individual steps are hierarchical in that ligand binding is essential for DNA binding, which itself is essential to start transcription.*

3.3. The modulatory effects of DNA response elements have a different impact on AR, PR and GR function

A critical problem within transcription factor families is how diverse regulatory programs are directed by very related members. AR, PR and GR which are highly conserved in their 66 amino acid-long DBD recognize the same consensus DNA response element. Yet they regulate distinct target genes with precise specificity. GR is the most widely and abundantly expressed receptor of this group, therefore androgen or progesterone-specific response mechanisms must emphasize AR or PR action while minimizing the effects of the GR.

For the AR it could be shown that the DNA response element transmits information to the receptor by inducing alterations in receptor conformation, which in turn influences ligand-induced transcriptional response (see above). The signals originating from the DNA element and conveyed to the bound receptor are differently interpreted by the AR, PR and GR, despite their high sequence similarity. This is most apparent in the case of receptor self-synergy. All three receptors are capable to self-synergize, but this varies, depending on the response element. Pem ARE-1 allowed the highest synergistic effects for the AR and PR whereas it blocked GR self-synergy. GR only synergized on the consensus SRE. Thus the lack of self-synergy of the GR on AREs is one mechanism how these DNA response elements specify receptor selectivity.

The recently defined synergy control motif acts as a negative regulatory region which has no effect when the transcription factor is bound to a single DNA response element, but blocks synergistic activity resulting from cooperative DNA binding of a transcription factor to multiple sites (Iniguez-Lluhi and Pearce, 2000). AR and GR both harbor two such synergy control motifs whereas PR has only one (Iniguez-Lluhi and Pearce, 2000). Because synergy is a phenomenon only occurring at multiple response elements, the combination of different elements in one promoter can add another level of complexity to the control of gene expression. In fact, transcriptional activation relying on multiple and often non-consensus response elements appears to be a hallmark of androgen-dependent enhancers, as reported for the rat Pb (Schoenmakers et al., 1999), mouse sex-limited protein (Verrijdt et al., 2002), human prostate-specific antigen (Cleutjens et al., 1996) and mouse *Pem* promoters.

The DNA element dependence of receptor synergy indicates furthermore that the DBD serves other roles beside DNA binding. Using chimeras of AR and GR with swapped domains it has been shown that the AR DBD is essential but not sufficient

for androgen-selective activation (Scheller et al., 1998). The authors speculate that N-C terminal interactions leading to synergy between several AR homodimers bound to multiple weak AREs play an important role in receptor specificity. For this to take place the DBD must transmit information originating from the bound response element to various other receptor domains. Subtle distinctions in AR, PR and GR DBD must therefore exist to enforce transcriptional specificity. Indeed the three different DBD mutations studied demonstrated an altered element dependence of the three receptors. The three amino acids are all implicated in the dimerization function of the DBD (Kaspar et al. 1993; Luisi et al., 1991). For the AR, GR and PR the salt bridge mutations led to an enhanced activity on SREs. In contrast on AREs these mutations severely impaired PR but not AR function. The two salt bridges found in the AR homodimer are known to be important for restriction of self-synergy (Liu et al., 1996). This points to a mechanism in which inter- and intra- molecular interactions of the receptor molecules, influenced by the allosteric modulation of the DBD, control self-synergy. Mutation of the Ala residue in the D box had a different effect on the AR than on the PR and GR. The AR A596T mutation, which is responsible for the Reifenstein syndrome, did in contrast to what was seen for the equivalent GR and PR mutant not limit its transcriptional capacity on SREs. Interestingly, the GR A458T brought about a more severe activity decrease on the SREs than on the AREs, whereas the opposite was true for the equivalent A604T mutation of PR.

Altogether, for the AR, reduction of function was only seen for the A596T mutant, when tested with the AREs, whereas enhanced activity was obtained following most mutations. PR and GR behaved differently. This shows that the DNA elements differentially interpret equivalent modifications introduced into AR, PR and GR. It indicates the role of the dimerization domain to vary in each receptor and further suggests that the DNA element plays an essential role in determining receptor specificity. Additionally, other factors like coactivators, corepressors and receptor-modifying enzymes influence the transcriptional activity of the receptor on a given promoter, making gene regulation by steroid receptors even more complex. The DNA element and its promoter context determine the activation potential of the ligand-bound receptor, which can only attain full activity if the necessary cofactors are present. The integration of the individual differences at some or all of these levels of modulation might therefore be the key to receptor specificity.

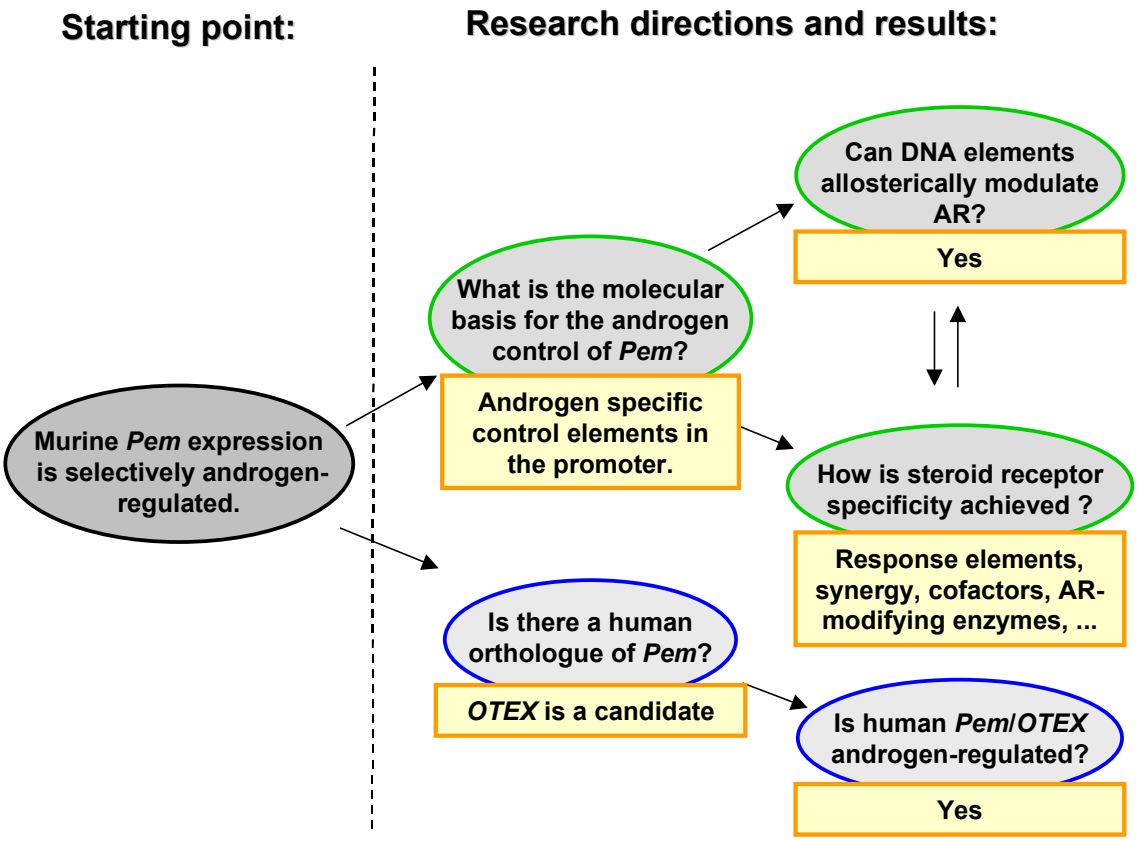

Figure 33: Schematic overview of the aim of this work and the main conclusions.

3.4. Outlook

It has been known for a long time that ligands regulate steroid receptor activity by inducing conformational changes which activate or repress their transcriptional potential. The finding that beside hormones, DNA elements may also allosterically modulate receptor conformation raises the possibility that the ligand- and DNA-induced changes are interdependent. It might thus be possible to identify ligands that specifically direct the AR to a given class of response elements and thereby activate only a subset of target genes. This might open new opportunities for manipulating the AR in therapeutic areas as diverse as hormone replacement therapy, male contraception and prostate carcinoma. By extension, other steroid receptors may also have to be reconsidered in this new light as targets for the control of selected sets of downstream genes.