Nuclear Receptor Resource
White Paper
Gene Expression: Nuclear Receptors
J.P. Vanden Heuvel, INDIGO Biosciences Inc., State College PA

Gene Expression: Nuclear Receptors

1. Overview

There are four main classes of receptors, Ion channels, G-protein coupled (GPCR), receptor tyrosine kinase and soluble (also called intracellular) receptors. The first three have ligands that are mainly present in the extracellular space (water soluble ligands). The binding of ligand causes a change in the shape/function of the protein that generates a signal inside the cell (increased ions, second messengers, phosphorylation of substrate, see Figure 1). A common feature of the xenobiotic ligands of these receptors is that they are lipophilic compounds that can diffuse across the cell membrane with moderate ease. The interaction with the xenobiotic receptor may occur predominantly in the cytosol (AhR) or in the nucleus (ER, PXR). The ability of xenobiotic receptors to regulate gene expression makes them potent regulators of the cell cycle and apoptosis. In addition, many cytochrome P450 enzymes are regulated by

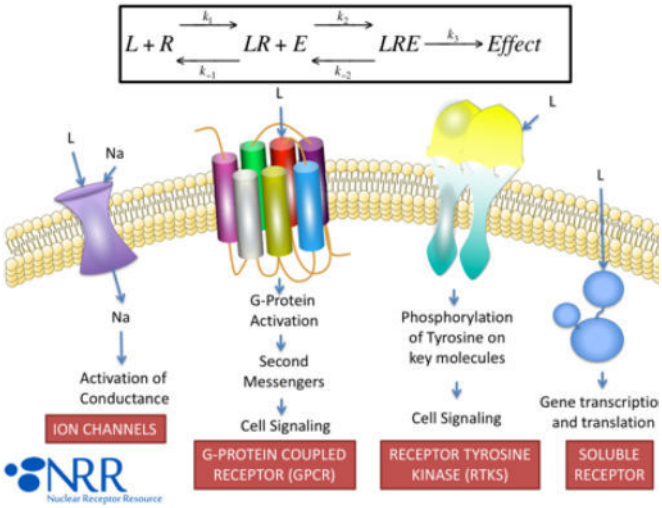

Figure 1. Type of receptors and signal transduction

xenobiotic receptors. In addition to affecting the metabolism of endogenous compounds, enzyme induction is an important aspect of drug-drug interactions and toxicity.

2. Nomenclature

Unlike receptors found on the cell surface, members of the nuclear hormone receptor (NHR or NR) superfamily are restricted to metazoan organisms such as nematodes, insects, and vertebrates. These proteins are intracellular transcription factors that directly regulate gene expression in response to lipophilic molecules. They affect a wide variety of functions, including fatty acid metabolism, reproductive development, and detoxification of foreign substances. To date, over 300 NHRs have been cloned, 48 in humans. Early classification of these receptors was based on ligands, DNA binding properties or other functional characterization. A more systematic classification has been proposed, based on sequence similarity, much like that employed for cytochrome P450s. Phylogenic analysis has shown six subfamilies (NR1-6) with various groups and individual genes (see Figure 2). The aryl-hydrocarbon receptor (AhR)

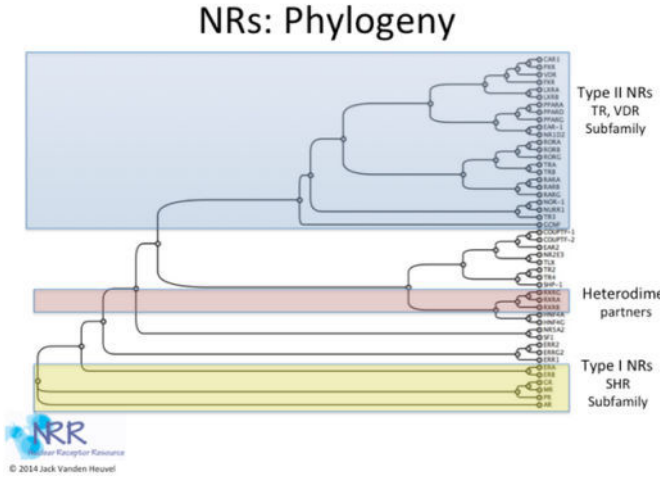

Figure 2. Nuclear receptor phylogenic tree.

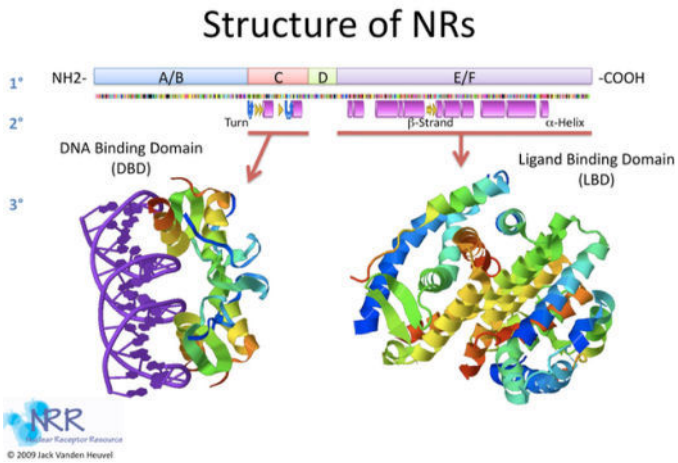

Figure 3. Structure of nuclear receptors.

fits many of the same characteristics as the NR family, although it is structurally unrelated. This receptor will be described in following a detailed examination of the NRs.

3. NR Functional Domains

Most NRs have the same basic structure, as shown in Figure 3. The most highly conserved region is the C4 zinc finger domain, which is a site-specific DNA binding motif. Each of the different domains of NRs contribute to its overall function as a receptor and a signal transduction molecule. There are some "exceptions to the rule" which will be described subsequently.

a. N-terminal regulatory domain (AB domain)

The N terminus of the NR, sometimes called the modulator, hypervariable or A/B domain, has transactivation activity, termed activation function 1 (AF-1). This acidic AD is ligand-independent, or constitutively functional. The A/B domain's sequence and length are highly variable between receptors (i.e. GR versus RXR) and among receptor subtypes (RXRα versus β). In addition, this region is the most frequent site of alternative splicing and secondary start sites and contains a variety of kinase recognition sequences. For these reasons, it is thought that the variable N-terminal sequences may be responsible for the receptor-, species-, and cell type-specific effects as well as promoter context-dependent properties of NR transactivation

b. DNA Binding Domain (C Domain)

NRs bind to hormone response elements (HREs) in their target promoters through the DNA binding domain (DBD) or C domain. Composed of two zinc fingers, the DBD is the most conserved region within the NR superfamily. The first zinc finger contains the proximal-or P-box region, an alpha helix that which is responsible for high-affinity recognition of the "core half-site" of the response element. Located within the second zinc finger is the distal or D-box, an α-helix which lies perpendicular to P-box helix, and is a site that mediates receptor dimerization. NRs bind to DNA as heterodimers, homodimers, or monomers, depending on the class of NR (see Figure 4). The steroid hormone receptors GR, PR, ER, AR and MR (receptors for

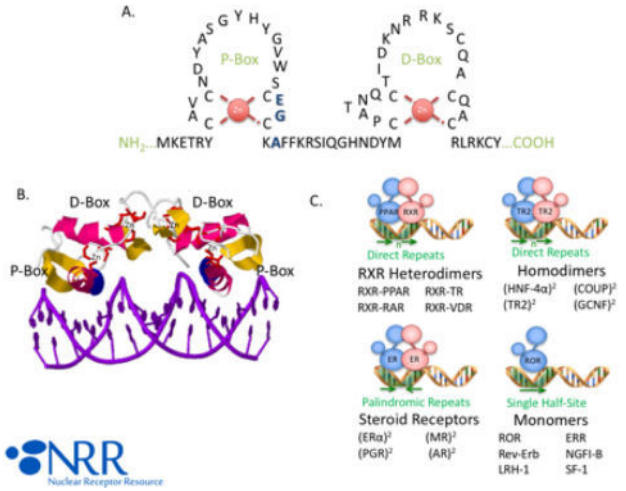

Figure 4. DNA binding domain

glucocorticoid, progesterone, estrogen, androgen and mineralocorticoids, respectively) bind to DNA as homodimers and recognize a palindromic response element. However, thyroid, retinoid, vitamin D and peroxisome proliferator receptors (TR, RAR, VDR and PPAR), as well as most orphan receptors, bind to DNA as a heterodimer with retinoid-x-receptor (RXR). The three dimensional structure of the RXR heterodimer complex produces different DNA binding affinities. Response elements may be direct repeats (DR_x, AGGTCA-N_x-AGGTCA, where N is any nucleotide and x is any number of residues from 0-10), everted repeats (ER_x, ACTGGA-N_x-AGGTCA) or inverted repeat (IR_x, AGGTCA-Nx-ACTGGA).

c. Hinge region (D domain).

Immediately adjacent to the DNA binding domain is the D or hinge domain. This particular region has an ill-defined function. The hinge domain contains the carboxy-terminal extension (CTE) of the DBD, which may be involved in recognizing the extended 5' end of the HRE. The D-domain appears to allow for conformational changes in the protein structure following ligand binding. Also, this region may contain nuclear localization signals and protein-protein interaction sites.

d. Ligand binding domain (LBD, E/F Domain)

The sequence of the ligand binding domain (LBD) or E/F domain varies substantially between NRs, but they all share a common structure of 11-13 α-helices organized around a hydrophobic binding pocket. Residues within the binding pocket confer specificity, determining whether the LBD will accept steroid hormones, retinoid compounds or the host of xenobiotic ligands that affect receptor function. Ligand-dependent activation requires the presence of activation function 2 (AF-2), located at the extreme C terminus of the NR. LBDs also contain nuclear localization signals, protein interaction with dimerization motifs for heat shock proteins, coregulators and other transcription factors.

e. Exceptions to the Rule

An important exception to the classic model of NR activation comes from receptors that do not require DNA binding. DAX-1 (NR0B1) and SHP (NR0B2) lack DBDs completely, although they are able to regulate gene expression. These receptors bind to other NRs and affect their ability to bind to DNA and affect gene transcription.

4. Basic Mechanism of Gene Regulation by NRs

The mechanism of action of nuclear hormone receptors can take one of two basic forms, that of steroid hormone receptors (SHRs) or that of retinoid/thyroid/Vitamin D receptors (See Figure 5). In the absence of ligand, the transcriptionally inactive SHRs MR, PR, GR, AR and ER are sequestered in a large complex comprising the receptor, heat shock protein-90 (HSP90), Hsp70, FKBP52/51 and possibly other proteins. The cellular localization of this

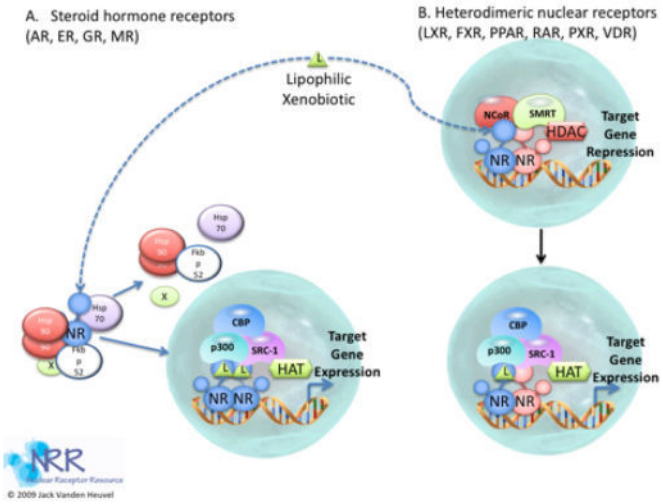

Figure 5. Mechanism of gene regulation by NRs

inactive complex is somewhat controversial and cytoplasmic or nuclear localization may be observed depending on the cell type and the conditions examined; however, the central dogma is that SHRs are cytosolic in the un-liganded form. One consequence of hormone binding to receptor is a distinct conformational change in receptor structure (discussed below). This conformational change marks the beginning of the signal transduction process. In the case of the GR subfamily (GR, AR, MR, PR), hormone binding elicits a dissociation of hsps and the release of a monomeric receptor from the complex. Genetic analysis and *in vitro* protease digestion experiments indicate that the conformational changes in receptor structure induced by agonists are similar but distinct from those produced by antagonists. However, both conformations appear to be incompatible with hsp binding.

The TR, RAR and VDR receptors do not avidly interact with hsps and are localized predominantly in the nucleus in the absence of ligand. Some unliganded NRs of this class may interact with DNA and act as transcription repressors. This may be the result of interaction with co-repressor proteins. An interesting exception to this observation is the constitutively active receptor (CAR) that is transcriptionally active in the absence of its ligand. Hormone induced conformational changes also occur upon activation of this class of NR, suggesting that alteration of receptor shape by ligands is a key step in the activation pathway.

Evidence suggests that receptors of the GR subfamily (SHRs) cooperatively bind to DNA as homodimers. The TR, RAR, VDR, PPAR and most of the orphan receptors form heterodimers with other members of the intracellular receptor superfamily. TR, RAR, PPAR and VDR can utilize RXRs as partners for heterodimer formation. The DNA site of contact depends on certain sequences within the C-domain, namely the proximal (P-box) and distal (D-box) zinc finger motifs (see description of the C-domain above). The P-box determines the half-site recognized, while the D-box determines the spacing between half-sites. Following activation, the SHRs receptors are capable of interacting with DNA, and both classes of NRs (SHRs and TR/RAR) can now recruit co-activators. The DNA bound NR complex is now a substrate for general transcription apparatus and the initiation of transcription commences.

5. Transcriptional co-regulatory proteins

Nuclear receptors bound to hormone response elements recruit a significant number of other proteins (referred to as transcription co-regulators) which facilitate or inhibit the transcription of the associated target gene into mRNA. The function of these co-regulators are varied and include chromatin remodeling (making the target gene either more or less accessible to transcription) or a bridging function to stabilize the binding of other co-regulatory proteins (see Figure 6).

That chromatin structure is important in gene regulation is underscored by its importance in development. The activation of developmentally-regulated genes may occur as a result of sequential changes in chromatin structure. Recent studies have shown that coactivators and corepressors mediate communication between upstream regulatory proteins and RNA pol II. These transcription *coregulators* carry DNA modifying activities, in particular histone acetylation and deacetylation capabilities. Ultimately, the end result of this complex, tiered approach to transcription regulation is that a particular gene is expressed in a proper temporal and spatially-localized manner and is sensitive to small differences in levels of an extracellular signaling molecule. Three means by which chromatin structure may be modified are briefly described below and include acetylation of histone tails, other post-translational modification (phosphor-ylation, methylation, and ubiquitinization) of the core histones and DNA methylation.

Histones are modified on specific lysine residues in the tail region and are the targets of histone acetyltransferase (HAT)- and histone deacetylase (HDAC)-containing co-regulators.

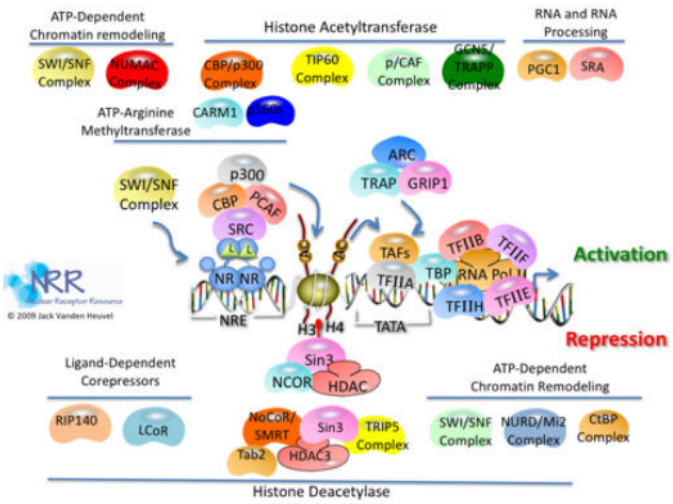

Figure 6. Nuclear receptors and co-regulatory proteins

The acetylation state of histones affects chromatin structure on several levels. First, acetylation of histone tails may reduce the stability of interaction with nucleosomal DNA. Acetylated histones wrap DNA less tightly and are more mobile than are hypoacetylated histone tails. Second, acetylation may disrupt the protein-protein interactions between histones, in particular the H3-H4 junction. Third, acetylated histones are less able to interact with adjacent nucleosomal arrays, thereby decreasing the stability of the compacted fiber. Last, acetylation of the core histones decreases the association of the linker histone H1, further destabilizing the higher order structure. Ultimately the end result of acetylating histone is an increase in transcription factor access to nucleosomal DNA. This fact may be observed by DNase sensitivity assays where regions of sensitivity correlate with histone acetylation.

A number of co-activators that are recruited to transcription factors have intrinsic HAT activity. These include the p160 family, important in nuclear receptor (NR) function, and the more general coactivators CBP/p300 and PCAF. The p160 family includes steroid receptor coactivator-1 (SRC-1), transcription intermediary factor-2 (TIF2), glucocorticoid receptor interacting protein (GRIP), pCIP and many others. These proteins associate with the NR's activator function-2 (AF2) domain, a latent domain that is revealed upon ligand binding. Thus, in the presence of hormone, the activated receptor binding to DNA and the HAT activity is recruited to the histone core.

HDAC activity is equally important in the regulation of gene expression. Deacteylation results in an increase in histone/DNA interaction, internucleosomal interactions and higher order DNA structure (reversal of HAT activity). Similar to HAT-containing proteins, HDAC proteins are often parts of large complexes of proteins. Several co-repressors are associated with histone deacetylase 1 (HDAC1) including Sin3, N-CoR and SMRT. Transcriptional regulation by co-repressors complexes can be blocked by chemical HDAC inhibitors (Tricostatin A, trapoxin), showing the role of this enzymatic activity in gene control. The recruitment HDAC to DNA by NRs is often the reverse to that of HAT proteins. That is, co-repressors bind to NRs in the absence of ligand and are released upon addition of hormone.

6. Aryl hydrocarbon receptor

a. Introduction

As mentioned above, the aryl-hydrocarbon receptor (AhR) is a soluble receptor that has the same basic features of the NRs including ligand binding and specific DNA binding. However, the structural domains used to impart these activities is different. AhR belongs to a relatively small group of proteins, which includes AhR-nuclear translocator (Arnt), considered the bHLH-PAS family. The helix-loop-helix (HLH) family contains over 240 transcriptional regulatory proteins and is found in a variety of organisms ranging from yeast to humans In eukaryotes, these proteins play key roles in developmental processes, including neurogenesis, myogenesis, hematopoiesis, and pancreatic development. From a toxicological standpoint, of interest are the basic HLH (bHLH) and the bHLH Per-Arnt-Sim (bHLH-PAS) proteins. Many important transcription factors including myc, Max, sterol-responsive element-binding protein (SREBP) are bHLH family members and lack the PAS domain The bHLH-PAS family includes the aryl hdrocarbon receptor (AhR), AhR nuclear translocator (Arnt) and hypoxia inducible factor-1α (HIF1α).

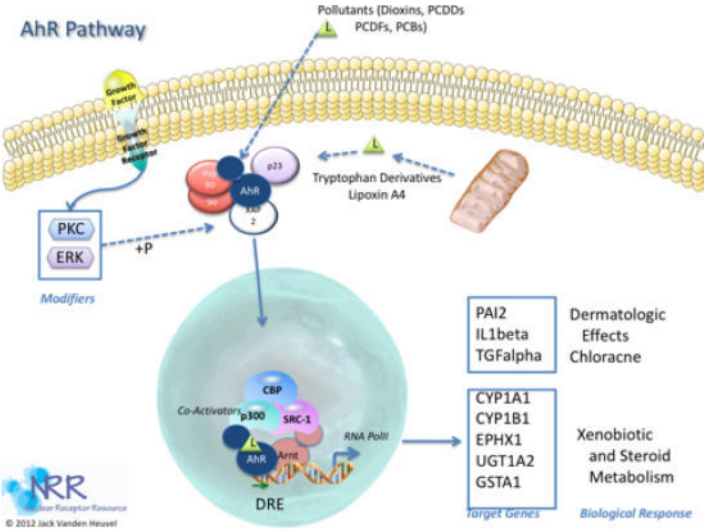

Figure 7. Mechanism of action of AhR

b. AhR Functional Domains

i. Basic Helix-Loop-Helix and DNA Binding

The basic HLH proteins are characterized by common possession of highly conserved bipartite domains for DNA binding and protein-protein interaction. A motif of mainly basic residues permits helix-loop-helix proteins to bind to a consensus hexanucleotide E-box (CANNTG). A second motif of primarily hydrophobic residues, the HLH domain, allows these proteins to interact and to form homo- and/or heterodimers. The dimerization motif contains about 50 amino acids and produces two amphipathic α-helices separated by a loop of variable length.

ii. Per-Arnt-Sim (PAS) Domain

The unique feature of bHLH-PAS proteins is the PAS domain, named for the first three proteins identified with this motif, Drosophila

Period (Per), human Arnt, and Drosophila Single-minded (Sim). The PAS domain is 260-310 amino acids long and can be subdivided into two well-conserved regions, PAS-A and PAS-B, separated by a poorly conserved spacer. The most definitive function of PAS is protein-protein interaction. The PAS domain is used to interact with other members of the bHLH-PAS family of proteins.

c. Mechanism of Action

The basic mechanism of action of AhR (Figure 7) has several features in common with both the steroid hormone receptors (cytoplasmic association with Hsps) as well as type II NRs (heterodimer formation). In the cytoplasm, AhR is inactive and associated with hsp90, p23 and XAP2. Upon ligand binding, the affinity for this hsp complex decreases and AhR crosses the nuclear membrane and interacts with Arnt. In mammalian cells, Arnt functions as a common dimerization partner with other bHLH-PAS proteins such as HIF and AhR. These heterodimers bind a sequence element related to the XRE (core GCGTG) or CME (core ACGTG). Co-regulatory proteins are brought to the AhR/Arnt heterodimer in a fshion very similar to the NRs.

7. Types of gene expression activity

When discussing a xenobiotic that binds with specificity to a receptor there are several classifications that are dependent on the biological response and should be addressed in a drug profiling study (Figure 8). An **agonist** is a ligand that increases the activity of a receptor above its basal, or constitutive level. A **full agonist** will bind to the LBD of the receptor and initiates a conformational change that results in high affinity interaction with transcriptional coactivator molecules. A **partial agonist** will also bind to the receptor, perhaps with the same affinity as a full agonist, but the resultant conformational change provides a weaker association with coactivator molecules. An **inverse agonist** is an agent that binds to the receptor but induces a pharmacological response opposite to that of an agonist. A prerequisite for an inverse agonist response is that the receptor must have high constitutive or basal activity in the absence of an identifiable ligand. The conformational change of the NR that results from an inverse agonist either decreases its association with coactivator molecules or increases affinity for corepressor proteins. There are several examples of inverse agonists of nuclear receptors, in particular of the estrogen receptor-related receptors (ERRα,β,ɣ) retinoic acid-related orphan receptors (RORα,β,ɣ) peroxisome proliferator-activated receptor δ (PPAR δ) and constitutive androstane receptor-1 (CAR-1). A receptor **antagonist** is a type of ligand that blocks or ameliorates agonist-mediated responses. A differentiating factor between an inverse agonist and an antagonist is that the latter functions in the presence of an agonist. **Full and partial antagonists** depend on the conformational shape that results from association with the receptor (see Figure 9). Full antagonists do not exhibit agonist properties whereas partial antagonist are also by definition partial agonists. It is important to note that while

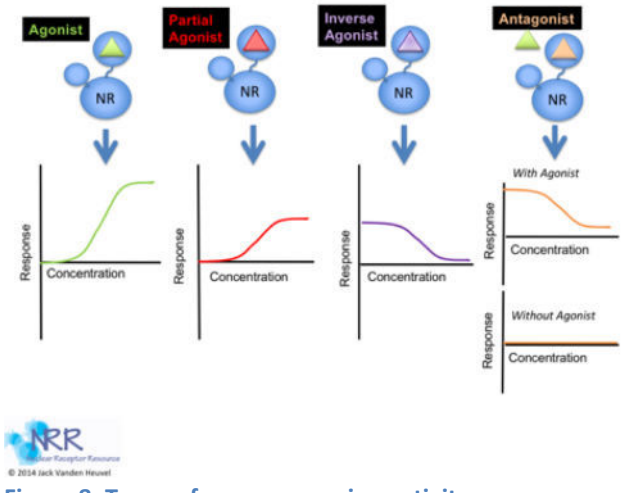

Figure 8. Types of gene expression activity

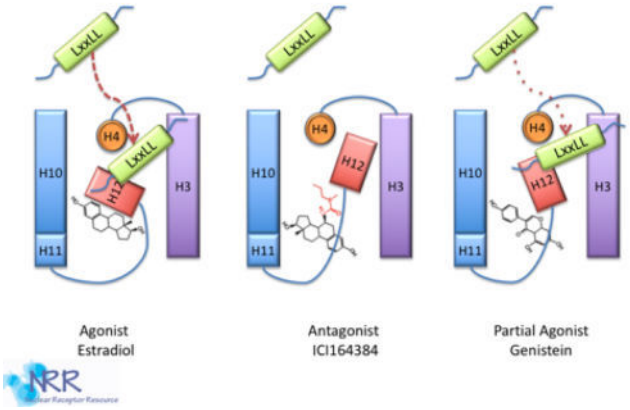

Figure 9. Conformational changes associated with NR agonist, antagonists and partial agonists

examining inverse agonists and antagonists toxicity is taken into account as a decrease in viability may present itself as either of these ligand-types.

Selective Receptor Modulators (SRMs) are compounds that result in a conformational change between full agonists and antagonists and hence have varying degrees of coactivator/corepressor affinity (i.e. are partial agonists). This results in tissue-specific responses that are dependent on the cellular context, in particular the concentration of transcriptional cofactors. A number of xenobiotics that work through nuclear receptors or AhR display an agonist response in some tissue while an antagonistic response in other tissues. The mechanism of action of SRMs may vary depending on the chemical structure of the ligand and the receptor involved, however it is thought that many SRMs work by promoting a conformation of the receptor that is closely balanced between agonism and antagonism. In tissues where the concentration of coactivator proteins is higher than corepressors, the equilibrium is shifted in the agonist direction. Conversely in tissues where corepressors dominate, the ligand behaves as an antagonist. A classic example of SRM activity is with the estrogen receptor (ER) antagonist tamoxifen which is a pure antagonist in breast tissue but a partial agonist in the uterus.

8. References

There are several general references on NR gene regulation that the reader is directed, including:

Chen, J. D. (2000). Steroid/nuclear receptor coactivators. Vitamins and Hormones, 58, 391-448.

Chiang, J. Y. (2005). Nuclear receptor regulation of lipid metabolism: Potential therapeutics for dyslipidemia, diabetes, and chronic heart and liver diseases.. Current Opinion in Investigational Drugs (London, England : 2000), 6(10), 994-1001.

Di Croce, L., Okret, S., Kersten, S., Gustafsson, J. A., Parker, M., Wahli, W., et al. (1999). Steroid and nuclear receptors. Villefranche-Sur-Mer, france, may 25-27, 1999. The EMBO Journal, 18(22), 6201-10.

Francis, G. A., Fayard, E., Picard, F., & Auwerx, J. (2003). Nuclear receptors and the control of metabolism. Annual Review of Physiology, 65, 261-311.

Kumar, R., & Thompson, E. B. (1999). The structure of the nuclear hormone receptors.. Steroids, 64(5), 310-9.

Mangelsdorf, D. J., & Evans, R. M. (1995). The RXR heterodimers and orphan receptors. Cell, 83(6), 841-50.

Nakata, K., Tanaka, Y., Nakano, T., Adachi, T., Tanaka, H., Kaminuma, T., et al. (2006). Nuclear receptor-mediated transcriptional regulation in phase I, II, and III xenobiotic metabolizing systems.. Drug Metabolism and Pharmacokinetics, 21(6), 437-57.

Tobin, J. F., & Freedman, L. P. (2006). Nuclear receptors as drug targets in metabolic diseases: New approaches to therapy.. Trends in Endocrinology and Metabolism: TEM, 17(7), 284-90.