Understanding nuclear receptors using computational methods

Ni Ai[1], Matthew D. Krasowski[2,3], William J. Welsh[1] and Sean Ekins[1,4,5]

[1] Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, Piscataway, NJ 08854, USA
[2] Department of Pathology, University of Pittsburgh, Pittsburgh, PA, 15213, USA
[3] Department of Pathology, Division of Clinical Chemistry, Toxicology and Therapeutic Drug Monitoring Laboratory, University of Pittsburgh Medical Center Presbyterian/Shadyside, Pittsburgh, PA, 15261, USA
[4] Collaborations in Chemistry, Jenkintown, PA 19046, USA
[5] Department of Pharmaceutical Sciences, University of Maryland, Baltimore, MD, USA

Nuclear receptors (NRs) are important targets for therapeutic drugs. NRs regulate transcriptional activities through binding to ligands and interacting with several regulating proteins. Computational methods can provide insights into essential ligand–receptor and protein–protein interactions. These in turn have facilitated the discovery of novel agonists and antagonists with high affinity and specificity as well as have aided in the prediction of toxic side effects of drugs by identifying possible off-target interactions. Here, we review the application of computational methods toward several clinically important NRs (with special emphasis on PXR) and discuss their use for screening and predicting the toxic side effects of xenobiotics.

Introduction

Nuclear receptors (NRs) are ligand-dependent transcription factors that regulate the expression of a variety of important target genes involved in a wide spectrum of developmental and physiological processes [1]. In addition to ligand binding, the transcriptional activities of NRs are also modulated through a range of regulating proteins termed as coactivator and corepressor [2–4]. The ligand-binding domain (LBD) of NRs is responsible for both ligand recognition and regulation of protein–protein interactions (Fig. 1a) [5]. Upon agonist binding, conformational changes are induced in the LBD, particularly the AF-2 region, which leads to the dissociation of a corepressor and recruitment of a coactivator (Fig. 1b) (reviewed in [6]). This contributes to downstream gene activation.

NRs represent one of the most important targets for therapeutic interventions for multiple diseases, including cancer, inflammation and metabolic diseases (such as metabolic syndrome) [7]. Understanding xenobiotic interactions with NRs is also important in the context of endocrine disruptors and environmental toxicity assessment [8]. It is therefore important not only to identify synthetic compounds that mimic the cognate NR ligand activity, but also to develop synthetic compounds that selectively modulate the pharmacology of NRs in a cell-type and/or tissue-selective manner to exert the desired therapeutic effects while avoiding potentially undesirable off-target effects (reviewed in [9–15]).

Different computational methods have emerged aiming at understanding and modeling the functional activities of NR modulators at the molecular level. Generally these computational approaches fall into two categories, ligand-based and receptor-based approaches, although more recently there have been efforts to combine these usually distinct approaches [16]. Ligand-based methods essentially focus on molecular similarity, which implies molecules with similar features exhibit similar biological responses. It is a particularly valuable approach to identify compounds if structural information for a receptor is unavailable. By contrast, receptor-based (also synonymous with target-based) methods require the three-dimensional structure of the protein targets predominantly generated from X-ray crystallography, NMR structures or homology modeling, to address the fundamental question of how a potential ligand might bind to the receptor. Both ligand- and receptor-based strategies have been widely applied to advance the understanding of various aspects of

Corresponding author: Ekins, S. (ekinssean@yahoo.com)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2009.03.003

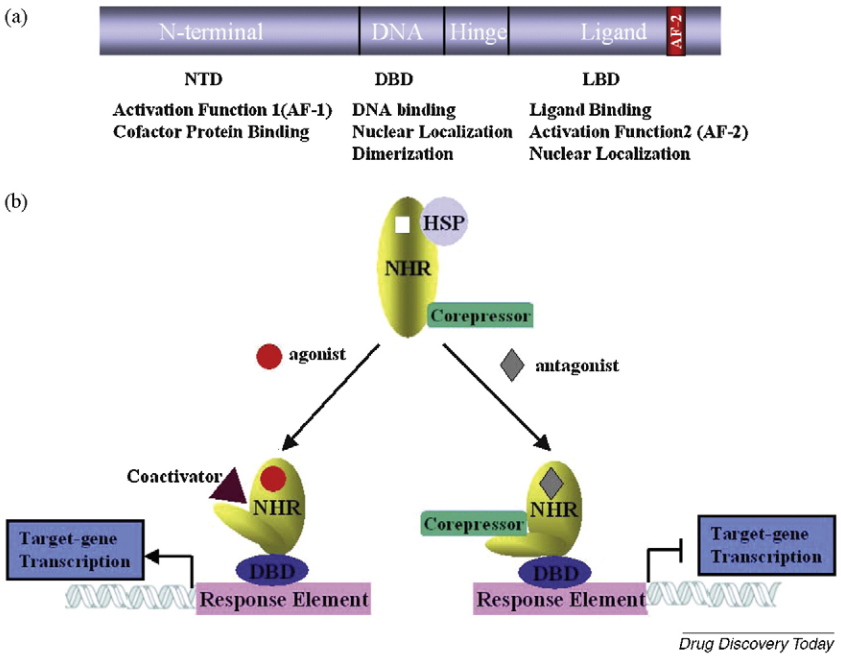

FIGURE 1

Domain structures of NRs and mechanism of action upon ligand binding. **(a)** The NR superfamily shares a common domain structure consisting of a NH_2-terminal domain (NTD), a central DNA-binding domain (DBD), a carboxy-terminal ligand-binding domain (LBD) and a hinge domain between DBD and LBD. Functions for each domain are also listed. **(b)** General model for transcriptional activation and repression in the presence of agonist and antagonist. Upon agonist binding, heat shock protein (HSP) and corepressor are dissociated from receptor. Conformational changes occur in the LBD during coactivator recruitment, then activation complexes are formed with other cofactor proteins to turn on target gene expression. Antagonism of NRs is complex and not completely understood. Here we present one known silencing mechanisms associated with antagonist binding. Antagonist binding may induce a difference conformation of LBD, therefore prohibiting coactivator binding or promoting the recruitment of corepressors. It is important to note many NRs work as homo- or heterodimers (and possibly higher order multimers). The monomer is displayed here for simplicity.

pharmacology in NRs [17,18]. In this review, we focus on several key NRs including the androgen receptor (AR; NR3C4), estrogen receptors α and β (ERα and ERβ; NR3A1 and NR3A2), pregnane X receptor (PXR; NR1I2), farnesoid X receptor (FXR; NR1H4), liver X receptors α and β (LXRα and LXRβ; NR1H3 and NR1H2) and vitamin D receptor (VDR; NR1I1). We will describe the process of method development and optimization to accommodate distinct receptor features of NRs, detail the success of computational methods and finally discuss the application of computational strategies to examine adverse effects of drugs.

Computational methods to understand NR pharmacology and evolution

NR ligands typically occupy a hydrophobic pocket that lies within the core of the NR LBD (reviewed in [19]). In contrast to the extensively studied ligand–protein recognition inside the ligand-binding pocket (LBP), the ligand entry or exit mechanisms to and from the binding site of NRs are poorly understood because there is not an obvious entry or exit route on the surface. Molecular dynamics (MD) simulations have emerged as a powerful approach to elucidate various potential dissociation routes of endogenous ligands and synthetic modulators for several NRs, such as the retinoic acid receptors (RARs; NR1B1, NR1B2 and NR1B3), thyroid receptors α and β (TRα and TRβ; NR1A1 and NR1A2), VDR and ERs [20–26]. Results from these computational simulations have revealed that significant conformational rearrangements of the receptor occur upon ligand binding and release. In addition to specific receptor and ligand types, computational simulations have indicated that the quaternary state of the receptor is one of the key factors that influences the prevalence of a particular pathway, implying that dimerization of NRs plays a role in ligand dissociation [20]. Several groups have suggested that it may be possible to develop pharmaceutically relevant ligands that could interact preferentially with NRs in specific oligomeric states, representing a new type of NR modulator [20].

Transcriptional activities of NRs are regulated by complex functional interactions of NRs with ligand, DNA, coactivators and corepressors. Recently a comparison of multiple crystal structures of LBDs revealed distinct receptor conformations under different ligand conditions (ligand-free, agonist-bound or antagonist bound) and coactivator/corepressor binding [27]. These structures make it possible to understand the ligand specificity of individual NRs at the atomic level. In the process they also identified helix 12 (H12) in the AF-2 site of the LBD as the molecular switch for NR activation. H12 is allosterically controlled by the binding of different ligands because it plays a structural role through its localization in the receptor, forming a corresponding site on the protein surface for cooperative coactivator/corepressor binding, in turn modulating the transcriptional effects of NRs. Several pieces of evidence suggest that receptor-specific recruitment or interaction

with different transcriptional cofactors is necessary for gene activation in different cell types and leads to distinct physiological consequences [28–30].

Computational methods have been employed to understand the structural basis for recognition and specificity among interacting partners that are required for the precise regulation of transcriptional activities of NRs. Wang *et al*. constructed a structural model for the PXR LBD with two different interacting domains (ID1 and ID2) from corepressors using homology modeling [31]. MD simulations on these two models predicted preferential binding toward ID2 over ID1 by PXR and revealed the key interactions, which was further supported and validated experimentally. These studies provide insights into the molecular interactions that direct the assembly of PXR and the corepressor. This may in turn have important implications for understanding the role of corepressors in regulating the biological activities of PXR, facilitating the design of therapeutic modulators promoting corepressor binding. Recently, similar MD simulations were performed on various oligomerization states of PXR to understand how the AF-2 region rearranges to form an 'active-capable' conformation on the receptor surface for obligate contacts with transcriptional coactivators [32]. These results revealed highly correlated motions by helices that comprise the AF-2 region, and indicated a path transmitting long-range motions from the PXR ligand scaffold to the surface (AF-2 region). On the basis of observations from computational simulations, it was also suggested that PXR formed a heterotetramer with RXR, instead of a heterodimer, to interact with coactivators [32]. Increased motion or flexibility/distortion of H12 on the AF-2 region was also demonstrated by MD performed with ERα after coactivator binding [32], as well as with AR and mutations associated with androgen insensitivity syndrome [33].

Computational methods have also shown their applicability to study adaptive evolutionary changes of the LBP of different NRs that correspond to varying ligand specificities across species. Reschly *et al*. recently combined computational approaches with *in vitro* studies to explore the structural basis of ligand preference across vertebrate species for NRs involved in the cholesterol metabolism pathway, identifying three patterns of coevolution with bile salts, the major elimination metabolites of cholesterol [34,35]. FXRs were suggested to have different specificities for primary bile salts across species which was achieved by altering the shape and size of the LBP. Human FXR has a curved binding pocket best suited for the bent steroid ring configuration typical of evolutionarily more recent bile acids. VDRs have recently acquired sensitivity to lithocholic acid, a toxic secondary bile salt, by changing the entrance to the LBP. PXRs have expanded their specificity for bile salts, from a narrow selectivity for planar bile salts in zebrafish to broad specificity for a wide range of bile salt structures in human, by substantial increases in the volume of the LBP [36]. The analysis of crystal structures of these three human NRs and LXR shows substantial differences in the volume of the ligand-binding cavity (PXR $\gg$ LXR > FXR > VDR) and distances between a crucial arginine residue (Arg-328 in human FXR) and a key glutamine or histamine residue (His-444 in human FXR) (FXR > LXR > VDR > PXR) [34]. Using evidence from homology modeling and docking, Reschly *et al*. proposed that FXR, VDR and PXR each acquired sensitivity to bile salts at separate points in vertebrate evolution and then adapted in different ways to the evolutionary changes in bile salt structure and metabolism [34]. In total, these combined *in vitro–in silico* observations present a new integrated picture of coevolution of a biochemical pathway and protein structure.

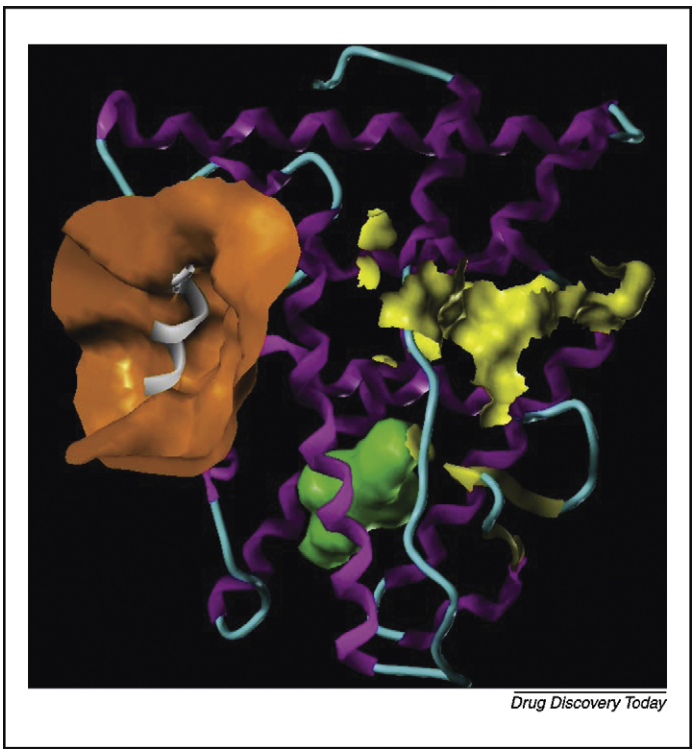

FIGURE 2

Functional sites on the LBD of NRs. The ER is used as an example. The ligand-binding pocket is shown as a green surface. The cofactor protein-binding site is colored in orange. The computationally identified possible steroid-receptor-specific functional site is shown with a yellow surface. The helices of ER are rendered by a ribbon representation and colored in magenta, while the coactivator motif is in white.

Computational analysis of three-dimensional coordinates of protein structures has helped to identify evolutionarily important residues and correlate them with important NR functions, such as dimerization and DNA binding [37]. Recently a new functional site located on the receptor surface was identified for steroid receptors in the NR3 family using the evolutionary trace computational method, which integrates evolution of protein sequence, structures and functions [38]. As shown in Fig. 2, this new site is different from amino acid residues in the AF-2 region, dimerization domains, or LBP and comprises residues specific for steroid receptors. Five of nine residues were found to correlate with certain human diseases, which suggests a biological role *in vivo* [39–43]. The biological relevance of the new functional site was further confirmed by experimental mutations of these residues in ER, showing that they disrupted signaling by allosterically influencing estradiol binding and coactivator recruitment [38]. Results from this computational study have also expanded our knowledge of protein–protein interactions involved in NR signaling.

Virtual screening methods to identify NR modulators

Ligand specificity of NRs is crucial in terms of both cellular transcription and therapeutic applications. To understand the

essential interactions that determine specificity, it is important to identify the key ligand structural features and crucial receptor residues that are involved. Computational methods have been widely applied to derive knowledge of such ligand–protein interactions and identify novel NR ligands, such as endocrine disrupting compounds (EDCs). EDCs are xenobiotics that activate NRs and disrupt crucial physiological functions by mimicking cognate ligands, posing a potential risk to wildlife and humans. Quantitative Structure–Activity Relationship (QSAR) and machine learning models are useful tools to rapidly extract the structural characteristics for binding from a set of active ligands using statistical techniques and specific molecular descriptors that may represent physicochemical properties, molecular shape or other properties. Multiple QSAR and machine learning models have been published for several NRs, including ER, AR and PXR, primarily to address the increasing need not only for high-throughput endocrine disruptor risk assessment [44–46], but also for toxicological screening [47] in combination with crystal structures and other *in silico* methods. A recent analysis of 74 natural or synthetic estrogens by a QSAR model yielded useful information on structural features for the preferential activation of the ERα and ERβ subtypes [48]. In addition to QSAR models, nonlinear statistical machine learning methods have been successfully applied to separate NR activators from nonactivators [49].

The activation of PXR regulates the expression of metabolizing enzymes such as cytochrome P450 enzymes (CYP3A4, CYP2B6 and CYP2C8/9) and glutathione-*S*-transferases, as well as important drug transporters (P-glycoprotein, multidrug resistance protein as well as others) [47]. Because the CYP enzymes metabolize the majority of clinically important drugs, inadvertent upregulation by PXR agonists may increase the metabolism and excretion of other coadministered therapeutic agents and cause undesirable drug–drug interactions or the generation of toxic levels of a drug metabolite. It is important to identify molecules that interact with PXR early in the drug development process, while some environmental compounds may also activate PXR [50]. Computational methods continue to play a significant role in identifying PXR activators [45,51,52] and assist in avoiding unexpected drug–drug interactions before human clinical testing. For example, a recent study used docking to propose candidate molecules that reduced PXR activity by forming destabilizing interactions [53]. Ung *et al.* explored the application of three machine learning methods for predicting PXR activators [52]. Their results indicated that the support vector machine (SVM) method had the best performance with overall accuracy around 80%. More recent machine learning analyses have used the same training set to create Tree and SVM models which were then tested with a large external test sets, providing the most exhaustive evaluations of PXR models to date [16,54]. The selected descriptors in these models were in good accord with those used in previous pharmacophore and QSAR modeling, as well as interactions suggested in X-ray crystallography studies.

Some NRs display broad ligand selectivity such that they can promiscuously bind and be activated by an array of structurally diverse ligands. This characteristic is correlated with their biological functions as xenobiotic sensors (e.g. PXR) or lipid sensors (e.g. peroxisome-proliferator activated receptor, PPARs, NR1C1, NR1C2 and NR1C3) [55]. Many proteins have been found to possess intrinsic disorder in some part of their sequence [56]. These disordered areas lack a rigid three-dimensional structure

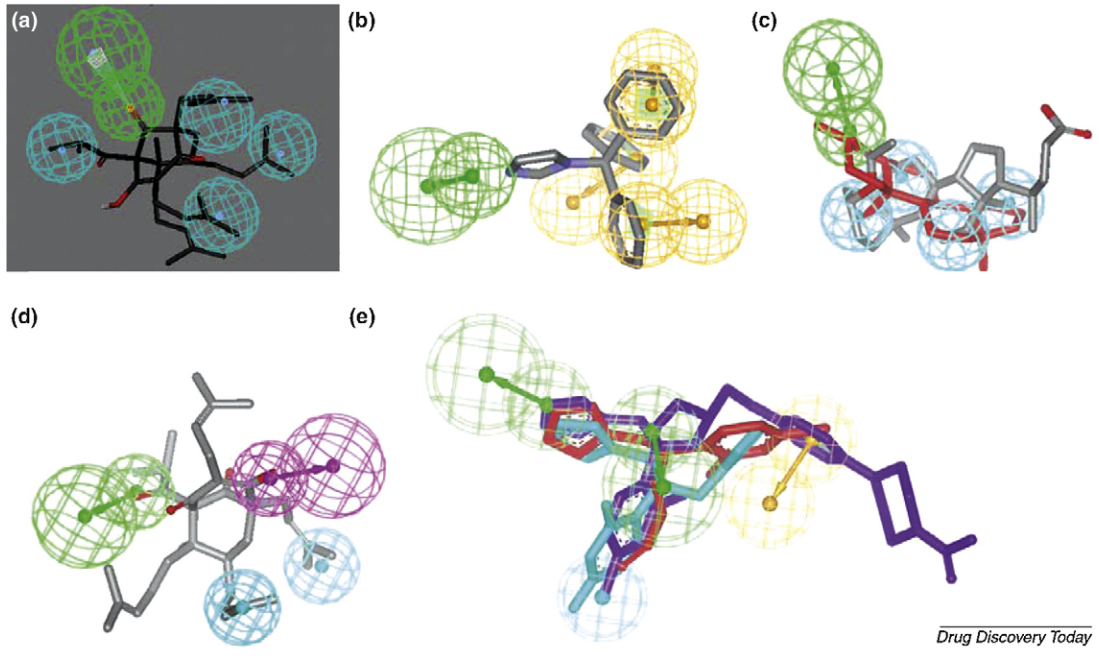

FIGURE 3

(a)–(d) 3D agonist pharmacophores derived from different training sets [51,70]. **(e)** Antagonist pharmacophore [70,86]. All pharmacophores were generated with Catalyst (Accelrys, San Diego). Pharmacophore features represent: green = hydrogen bond acceptor; cyan = hydrophobic, orange = ring hydrophobic and purple = hydrogen bond donor.

and frequently are not resolved in X-ray or NMR structures. These disordered regions can perform various important biological functions [57] such as protein–protein interactions [58], DNA binding, cell signaling [59] and protein–ligand interactions [60]. It was recently found that there is considerable variability in predicted intrinsic disorder within NRs, particularly in the D-domain (a region of NRs known to be involved in DNA recognition and heterodimerization, coactivator/corepressor interactions and protein–protein interactions) and LBD, which is probably an additional important evolutionary force in shaping protein–protein interactions, promiscuity and NR function [61].

The analysis of available crystal structures of PXR cocrystallized with different ligands indicated the LBP of PXR is a very large and flexible hydrophobic site, which may account for its promiscuity in binding structurally diverse ligands [62–68]. Receptor flexibility observed in the PXR LBD represents a considerable challenge for binding pose prediction and evaluation by docking methods and is an important consideration for applying receptor-based approaches to PXR. The majority of computational studies on PXR have therefore focused on ligand-based approaches to distinguish PXR activators and nonactivators, such as pharmacophores and descriptor-based statistical models (as described above). Pharmacophores represent the geometric arrangement of essential structural features of ligands for binding to protein targets. The first pharmacophore study of PXR used 12 published ligands to develop a model that revealed the structural determinants necessary for binding and suggested its potential application as a structural filter before *in vitro* determination [51] (Fig. 3). As more three-dimensional receptor structures cocrystallized with various ligands became available, new structural information on the LBP was integrated to develop a receptor-based pharmacophore model that labeled the area prohibiting the ligand binding (excluded volumes) and important residues involved in the binding [69]. To date, a consensus pharmacophore suggests that multiple hydrophobic features and one hydrogen bond acceptor are common for PXR agonists [70]. Recently, we have evaluated a hybrid method combining ligand-based SVM models with molecular docking in an attempt to improve predictions [16]. A two-dimensional pharmacophore model has been used by Lemaire *et al.* to filter a commercial compound database for previously unknown PXR activators [71]. The resulting compounds were docked into a PXR crystal structure and scored for ligand–receptor interactions. Experimental data from *in vitro* and *in vivo* studies corroborated their predictions and identified more active compounds. It is important to note that across species, PXR shows a high degree of sequence diversity in the LBD, resulting in marked differences in ligand selectivity, which probably also correlates with the evolutionary pressure of adaptation to toxic compounds throughout development. A pharmacophore analysis performed on the same 16 ligands for mammalian (human, mouse and rat), chicken, frog

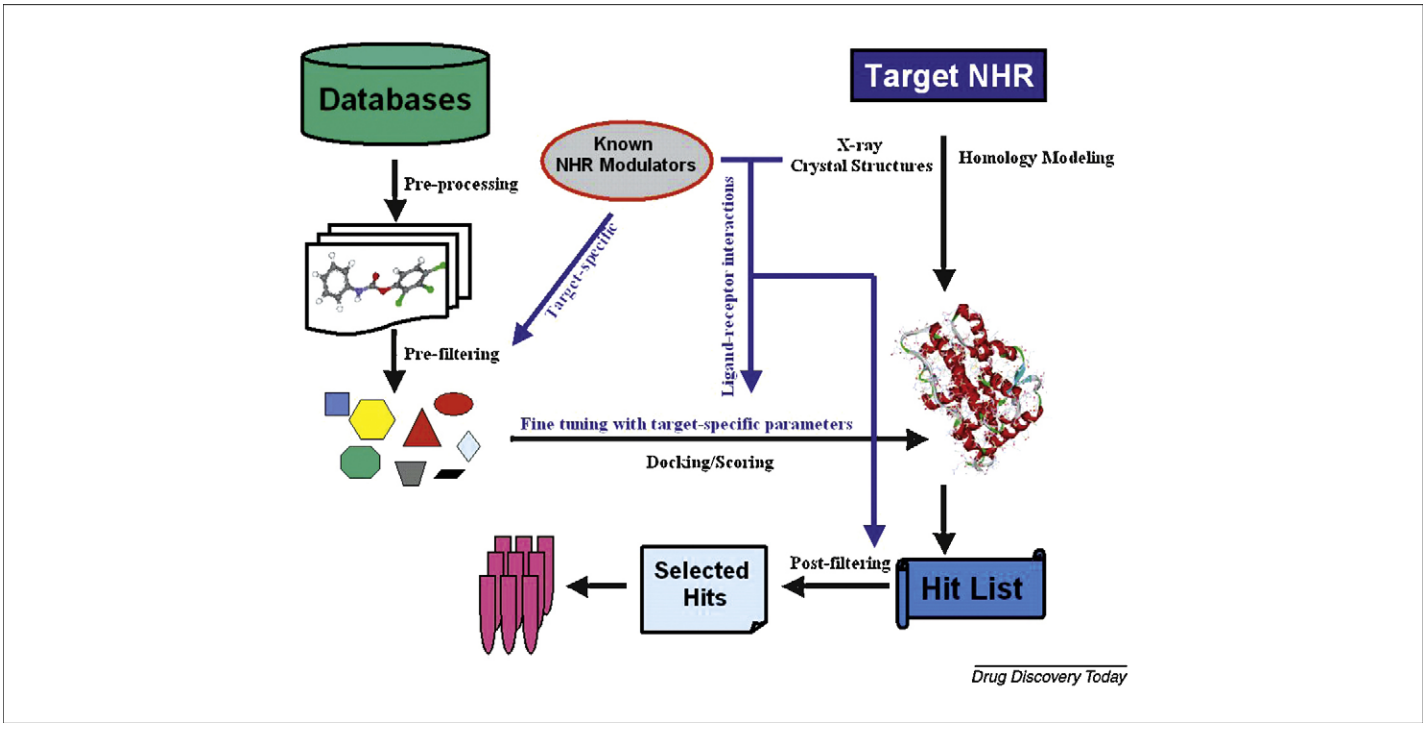

FIGURE 4

A virtual screening scheme for NRs. First, the corporate or commercial databases need to undergo prescreening processing that incorporates input file formatting, prefiltering for the drug-like compounds, exploring the conformational space of compounds, as well as their protonation, tautomeric and stereochemical stages. Studies have identified the 'hidden' impact of database preprocessing on VS results [91]. It is common that the number of screening compounds in the database can reach thousands to millions. Because docking/scoring is the most CPU-intensive and rate-limiting step, it is useful to apply certain prescreening filters to speed up the process. Generally these filters can be derived from structural and activities information from known modulators based on pharmacophore modeling and structural similarity. Structural information about the target receptor can be extracted from resolved X-ray crystallography structures or homology modeling. Abundant knowledge about ligand–protein interactions can be utilized to customize or fine-tune important parameters for the docking/scoring step of VS. The list of hits will be then post-processed by target-specific filters. This step tries to increase the retrieval rate of the active compounds from the whole database and minimize the occurrence of false positives. Finally selected hits that pass through all requirements are suggested to be evaluated biologically.

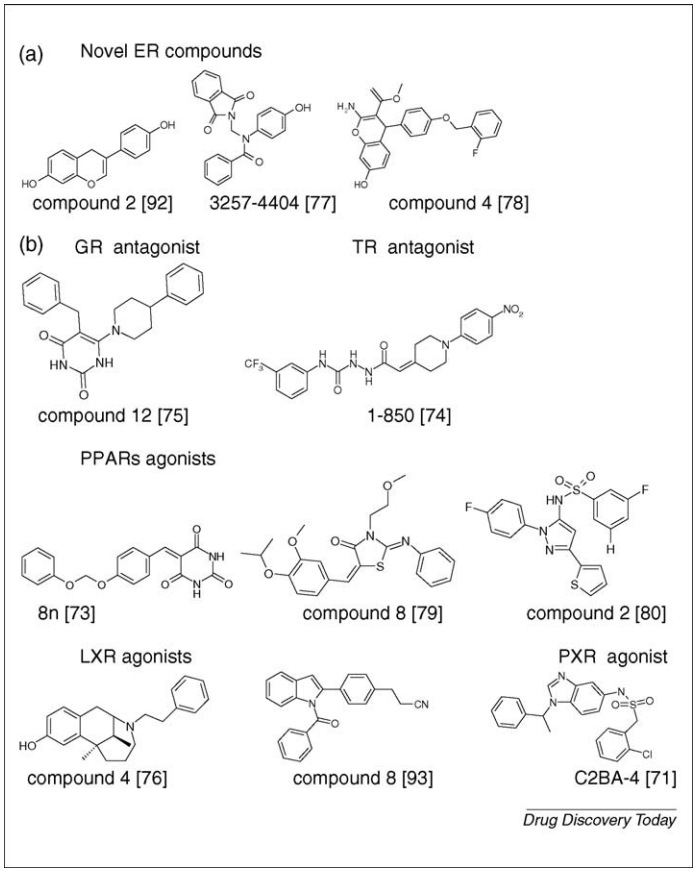

FIGURE 5

Structural drawings of some computationally discovered NR modulators using identifiers in the original references. These compounds bind to the LBP of the corresponding NRs.

and zebrafish PXRs also helped to understand the evolutionary history influencing ligand specificity [72] as mammals had similar pharmacophore arrangements while other species differed widely.

Virtual screening (VS) of large compound databases, which incorporates molecular docking methods and mathematical scoring functions, has been widely adopted to predict ligand–protein interactions and identify new drug leads for many NRs, such as ER, progesterone receptor (PR, NR3C3), glucocorticoid receptor (GR, NR3C1), TR, PPARs and LXR [73–80]. Because specificity for a given subtype of NR is of considerable importance for therapeutic applications, it is important to tailor the VS scheme with target-specific structural information, for example key hydrogen bonding residues, distance cutoff for important interactions and ligand–receptor complementarities to discover subtype-selective compounds. Using a VS protocol customized with ERβ-specific parameters, Zhao *et al.* identified ERβ-specific ligands from a plant product-based database [81]. Binding affinity and selectivity of the 12 candidates from VS were evaluated by a fluorescence polarization-binding assay. Three of 12 compounds displayed over 100-fold selectivity to ERβ over ERα. Similarly, Knox *et al.* proposed a target-specific VS scheme optimized for ERα. Nineteen known ERα inhibitors were used to calibrate the scoring functions for specific ERα activity. Three of seven tested compounds bound to ERα and exhibited good selectivity of ERα over ERβ [78,82]. A similar strategy (Fig. 4) can be utilized to discover novel compounds that interact with other NRs with high specificities and affinity such as the agonists and antagonists shown in Figs 5,6.

The LBP has been the predominant focus for the identification of NR modulators. Computational methods, combined with experimental approaches, have shown that there are other sites

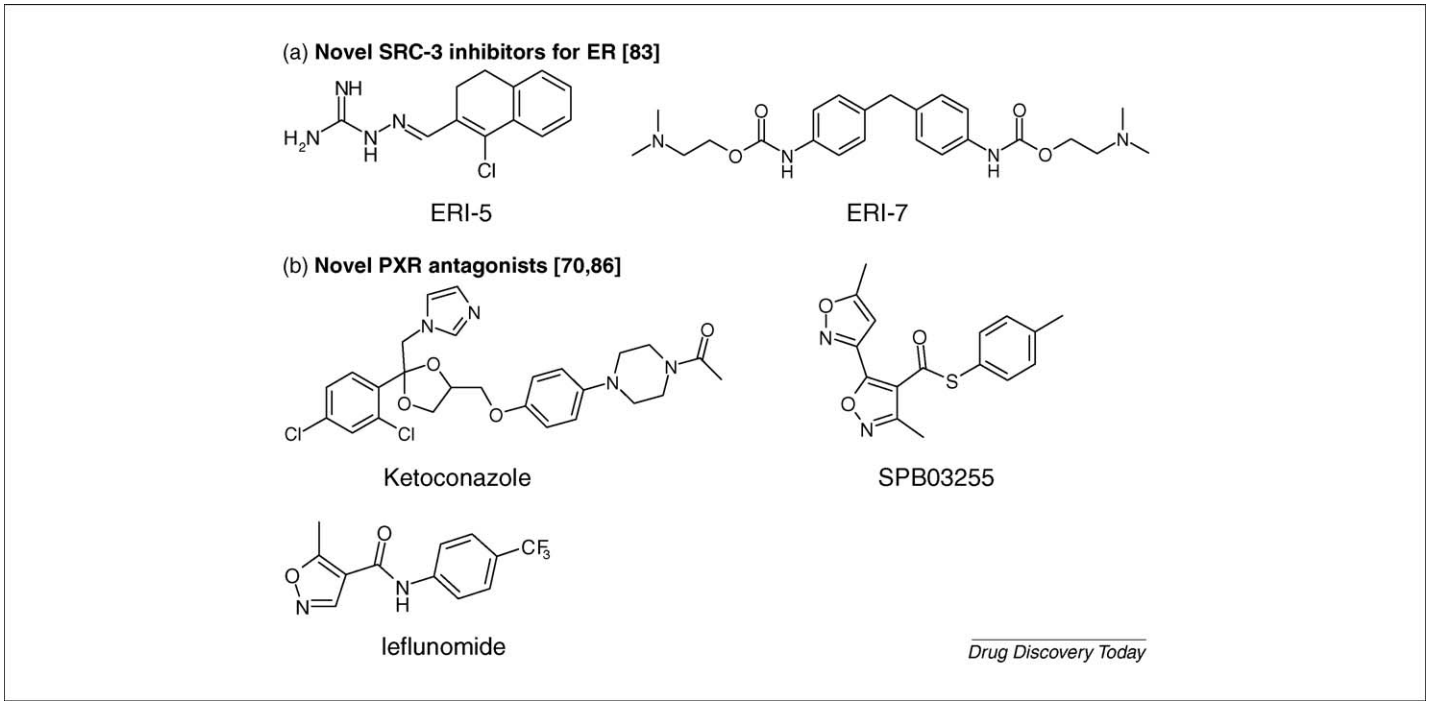

FIGURE 6

Structural drawings of some computationally discovered NR antagonists using identifiers in the original references. These compounds bind to the alternative site on the surface of NRs.

on the receptor surface possible for small-molecule binding to alter NR biological activities. The AF-2 site is a conserved region among NRs for cofactor protein binding, and is therefore a potential site for designing novel NR modulators which directly interfere with essential protein–protein interactions. Recently, a VS scheme was integrated with a cell-based assay to identify novel ERα antagonists that may disrupt coactivator binding [83]. Through a molecular docking study, compounds from the VS were found to selectively fit into the AF-2 site on the ER surface because of their unique shape and charge properties [83]. A mammalian two-hydrid assay also confirmed that selected compounds disrupted the receptor coactivator interactions without displacing estradiol binding to ER. The best compound was not structurally similar to known antiestrogens, which suggested a novel class of ERα modulators as an alternative to the current antiestrogen therapy.

Ketoconazole and miconazole have been shown as GR antagonists of dexamethasone binding (while fluconazole had no effect), repressing PXR, CAR and downstream gene expression [84]. Mutagenesis data have indicated that the AF-2 region is a potential binding site for azole antifungals (ketoconazole, enilconazole and fluconazole) acting as PXR antagonists [85]. Computational docking results revealed these PXR antagonists partially occupied the same hydrophobic groove where the coactivator motif binds to receptor, antagonizing the essential protein–protein interaction ([70] and references therein). On the basis of these PXR antagonists, a pharmacophore model was developed to elucidate the important structural features for binding. When this model was combined with docking studies and biological testing, it enabled discovery of several more potent nonazole PXR antagonists (included in Fig. 6) which included commercially available synthetic compounds and the FDA approved prodrug leflunomide (Fig. 6), confirmed experimentally *in vitro* [86]. These small molecules also had good ligand efficiencies (as they were more potent on a per heavy atom basis) compared with ketoconazole when determined using the published approaches (Fig. 7) [87]. This suggests that smaller molecules could also be effective antagonists with optimal protein–ligand interactions. It is also important to consider that the antagonism of PXR or other NRs could occur via interactions with other proteins that interact with PXR or at other surface sites beyond those currently known. Novel PXR antagonists provide a possible small-molecule intervention to control drug metabolism and transport by reducing the activation of these genes during therapeutic treatment.

Computational methods to study toxicity profiles related to NR modulators

Selective estrogen receptor modulators (SERMs) such as tamoxifen have been developed to treat breast cancers and other diseases. However, several adverse effects are associated with treatment using these drugs, including cardiac abnormalities, thromboembolic disorders and ocular toxicity. To understand the molecular mechanism of adverse effects, considerable efforts have been taken to identify off-target interactions using computational approaches such as docking studies with structural similarity comparison methods for binding sites. The protein target with the highest similarity to the ERα LBP was sarcoplasmic reticulum (SR) Ca^{2+} ion channel ATPase protein (SERCA), which is a key mediator of cytosolic calcium levels by accumulating calcium in the lumen.

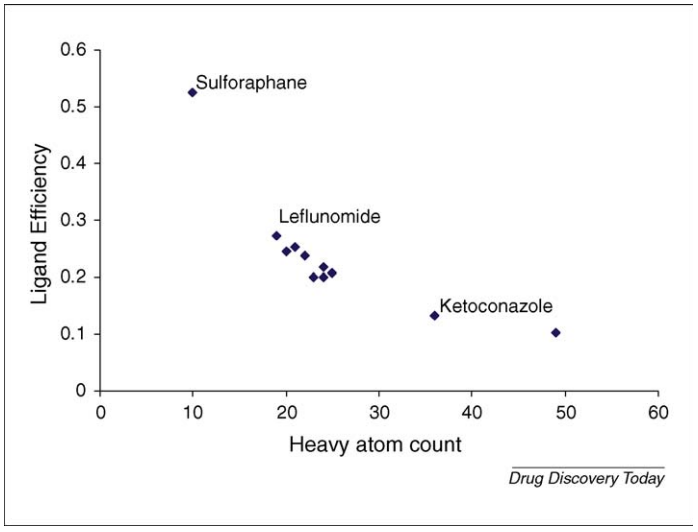

FIGURE 7

Ligand efficiency versus heavy atom count for PXR antagonists which highlights the relative positions of the three compounds of interest. When we consider the ligand efficiency log K_i/heavy atom count (no hydrogens) versus heavy atom count, there is an exponential decrease in efficiency between 10 and 20 heavy atoms, which is comparable with observations for much larger datasets across different targets [87].

Because the gradient concentration of calcium in the SR is important for muscle contraction, it is possible that the inhibition of SERCA by SERMs may cause the loss of calcium homeostasis in platelets and lead to the reported adverse effects, such as cardiac abnormalities [88].

Similar *in silico* approaches could therefore be used with other NRs and incorporated into the drug discovery process for the early identification of off-target adverse effects. Retrospectively, computational approaches are also able to discover endocrine effects for marketed drugs. These results can help to explain the observed side effects and provide an opportunity to optimize the pharmacological profiles of drugs to eliminate activity at the original target and ultimately enhance endocrine activities for related therapeutic indications (an example of drug repurposing [89]). This latter approach takes advantage of the presumably favorable bioavailability and toxicity profiles for marketed drugs to save time and cost during the drug development process. Bisson *et al.* demonstrated the application of this computational approach on a library of marketed oral drugs. Their study led to a nonsteroidal antiandrogen with improved AR antagonistic activity and markedly reduced antipsychotic effects [90].

Conclusions

NRs are important transcriptional factors that regulate several essential physiological processes involved in metabolism, development and systemic homeostasis. Transcriptional activities of NRs are guided by interactions with ligands and multiple cofactor proteins. In recent years, computational modeling of NRs has proved increasingly valuable to advance the understanding of NR pharmacology. Detailed insights about how ligand–protein, protein–protein, protein–corepressor and protein–coactivator interactions occur make it possible to predict potential off-target effects. We have shown that several agonists and antagonists were computationally discovered for NRs (Figs 5,6), which indicates

their potential utility for discovering potential therapeutics to treat NR-related diseases in the future.

Conflicts of interest

Sean Ekins is a consultant for Collaborations In Chemistry. Ni Ai, Matthew D. Krasowski and William J. Welsh have no conflicts of interest.

Acknowledgements

The authors kindly acknowledge the collaborations and the many valuable discussions with Drs Erica J. Reschly, Lee R. Hagey, Alan F. Hofmann, Sridhar Mani, Erin Schuetz, Ken Bachmann, Peter W. Swaan and our colleagues at UMDNJ.

SE also thanks Accelrys, San Diego, CA for making Discovery Studio Catalyst available.

Support for this work has been provided by the USEPA-funded Environmental Bioinformatics and Computational Toxicology Center (ebCTC), under STAR Grant number GAD R 832721-010, and by the Defense Threat Reduction Agency, under contract number HDTRA-BB07TAS020. This work was also funded by NIH-GM081394 from the National Institute of General Medical Sciences (to WJW). MDK is supported by NIH K08-GM074238. This work has not been reviewed by and does not represent the opinions of the funding agencies.

We thank the reviewers for their very useful comments.

References

1 Aranda, A. and Pascual, A. (2001) Nuclear hormone receptors and gene expression. *Physiol. Rev.* 81, 1269–1304
2 Horwitz, K.B. *et al.* (1996) Nuclear receptor coactivators and corepressors. *Mol. Endocrinol.* 10, 1167–1177
3 Chen, J.D. and Li, H. (1998) Coactivation and corepression in transcriptional regulation by steroid/nuclear hormone receptors. *Crit. Rev. Eukaryot. Gene Expr.* 8, 169–190
4 Glass, C.K. and Rosenfeld, M.G. (2000) The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev.* 14, 121–141
5 Shiau, A.K. *et al.* (1998) The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* 95, 927–937
6 Steinmetz, A.C.U. *et al.* (2001) Binding of ligands and activation of transcription by nuclear receptors. *Annu. Rev. Biophys. Biomol. Struct.* 30, 329–359
7 Smith, A.G. and Muscat, G.E. (2005) Skeletal muscle and nuclear hormone receptors: implications for cardiovascular and metabolic disease. *Int. J. Biochem. Cell Biol.* 37, 2047–2063
8 Hotchkiss, A.K. *et al.* (2008) Fifteen years after ''wingspread'' – environmental endocrine disrupters and human and wildlife health: where we are today and where we need to go. *Toxicol. Sci.* 105, 235–259
9 Tenbaum, S. and Baniahmad, A. (1997) Nuclear receptors: structure, function and involvement in disease. *Int. J. Biochem. Cell Biol.* 29, 1325–1341
10 Zimber, A. and Gespach, C. (2008) Bile acids and derivatives, their nuclear receptors FXR, PXR and ligands: role in health and disease and their therapeutic potential. *Anticancer Agents Med. Chem.* 8, 540–563
11 Feldman, P.L. *et al.* (2008) PPAR modulators and PPAR pan agonists for metabolic diseases: the next generation of drugs targeting peroxisome proliferator-activated receptors? *Curr. Top. Med. Chem.* 8, 728–749
12 Ma, X. *et al.* (2008) The pregnane X receptor: from bench to bedside. *Expert Opin. Drug Metab. Toxicol.* 4, 895–908
13 Bertolotti, M. *et al.* (2008) Nuclear receptors as potential molecular targets in cholesterol accumulation conditions: insights from evidence on hepatic cholesterol degradation and gallstone disease in humans. *Curr. Med. Chem.* 15, 2271–2284
14 Baulieu, E.E. *et al.* (1975) Steroid hormone receptors. *Vitam. Horm.* 33, 649–736
15 Henley, D.V. and Korach, K.S. (2006) Endocrine-disrupting chemicals use distinct mechanisms of action to modulate endocrine system function. *Endocrinology* 147, s25–32
16 Hybrid scoring and classification approaches to predict human pregnane x receptor activators. Kortagere S, Chekmarev D, Welsh WJ, Ekins S. *Pharm. Res.* 2009 Apr;26 (4), 1001–11
17 Ekins, S. *et al.* (2007) In silico pharmacology for drug discovery: applications to targets and beyond. *Br. J. Pharmacol.* 152, 21–37
18 Ekins, S. *et al.* (2007) In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Br. J. Pharmacol.* 152, 9–20
19 Moras, D. and Gronemeyer, H. (1998) The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* 10, 384–391
20 Sonoda, M.T. *et al.* (2008) Ligand dissociation from estrogen receptor is mediated by receptor dimerization: evidence from molecular dynamics simulations. *Mol. Endocrinol.* 22, 1565–1578
21 Blondel, A. *et al.* (1999) Retinoic acid receptor: a simulation analysis of retinoic acid binding and the resulting conformational changes. *J. Mol. Biol.* 291, 101–115
22 Kosztin, D. *et al.* (1999) Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. *Biophys. J.* 76, 188–197
23 Carlsson, P. *et al.* (2006) Unbinding of retinoic acid from the retinoic acid receptor by random expulsion molecular dynamics. *Biophys. J.* 91, 3151–3161
24 Martinez, L. *et al.* (2005) Molecular dynamics simulations reveal multiple pathways of ligand dissociation from thyroid hormone receptors. *Biophys. J.* 89, 2011–2023
25 Martinez, L. *et al.* (2006) Molecular dynamics simulations of ligand dissociation from thyroid hormone receptors: evidence of the likeliest escape pathway and its implications for the design of novel ligands. *J. Med. Chem.* 49, 23–26
26 Perakyla, M. (2009) Ligand unbinding pathways from the vitamin D receptor studied by molecular dynamics simulations. *Eur. Biophys. J.* 38, 185–198
27 Egea, P.F. *et al.* (2000) Ligand–protein interactions in nuclear receptors of hormones. *FEBS Lett.* 476, 62–67
28 Kawagoe, J. *et al.* (2007) Mechanism of the divergent effects of estrogen on the cell proliferation of human umbilical endothelial versus aortic smooth muscle cells. *Endocrinology* 148, 6092–6099
29 Shang, Y. and Brown, M. (2002) Molecular determinants for the tissue specificity of SERMs. *Science* 295, 2465–2468
30 Morganstein, D.L. and Parker, M.G. (2007) Role of nuclear receptor coregulators in metabolism. *Expert Rev. Endocrinol. Metab.* 2, 797–807
31 Wang, C.Y. *et al.* (2006) Structural model reveals key interactions in the assembly of the pregnane X receptor/corepressor complex. *Mol. Pharmacol.* 69, 1513–1517
32 Teotico, D.G. *et al.* (2008) Active nuclear receptors exhibit highly correlated AF-2 domain motions. *PLoS Comput. Biol.* 4, e1000111
33 Elhaji, Y.A. *et al.* (2006) Impaired helix 12 dynamics due to proline 892 substitutions in the androgen receptor are associated with complete androgen insensitivity. *Hum. Mol. Genet.* 15, 921–931
34 Reschly, E.J. *et al.* (2008) Evolution of the bile salt nuclear receptor FXR in vertebrates. *J. Lipid Res.* 49, 1577–1587
35 Reschly, E.J. *et al.* (2008) Ligand specificity and evolution of liver X receptors. *J. Steroid Biochem. Mol. Biol.* 110, 83–94
36 Krasowski, M.D. *et al.* (2005) Evolution of the pregnane X receptor: adaptation to cross-species differences in biliary bile salts. *Mol. Endocrinol.* 19, 1720–1739
37 Folkertsma, S. *et al.* (2005) The nuclear receptor ligand-binding domain: a family-based structure analysis. *Curr. Med. Chem.* 12, 1001–1016
38 Raviscioni, M. *et al.* (2006) Evolutionary identification of a subtype specific functional site in the ligand binding domain of steroid receptors. *Proteins: Struct. Funct. Bioinform.* 64, 1046–1057
39 Gad, Y.Z. *et al.* (2003) A novel point mutation of the androgen receptor (F804L) in an Egyptian newborn with complete androgen insensitivity associated with congenital glaucoma and hypertrophic pyloric stenosis. *Clin. Genet.* 63, 59–63
40 Alen, P. *et al.* (1999) The androgen receptor amino-terminal domain plays a key role in p160 coactivator-stimulated gene transcription. *Mol. Cell Biol.* 19, 6085–6097
41 Castagnaro, M. *et al.* (1993) Androgen receptor gene mutations and p53 gene analysis in advanced prostate cancer. *Verh. Dtsch. Ges. Pathol.* 77, 119–123
42 Chavez, B. *et al.* (2001) Eight novel mutations of the androgen receptor gene in patients with androgen insensitivity syndrome. *J. Hum. Genet.* 46, 560–565
43 Hiort, O. *et al.* (1993) Single strand conformation polymorphism analysis of androgen receptor gene mutations in patients with androgen insensitivity syndromes: application for diagnosis, genetic counseling, and therapy. *J. Clin. Endocrinol. Metab.* 77, 262–266
44 Ai, N. *et al.* (2003) Computational models for predicting the binding affinities of ligands for the wild-type androgen receptor and a mutated variant associated with human prostate cancer. *Chem. Res. Toxicol.* 16, 1652–1660

45 Jacobs, M.N. (2004) In silico tools to aid risk assessment of endocrine disrupting chemicals. *Toxicology* 205, 43–53
46 Tong, W. *et al.* (1997) QSAR models for binding of estrogenic compounds to estrogen receptor {alpha} and {beta} subtypes. *Endocrinology* 138, 4022–4025
47 Ekins, S. *et al.* (2002) A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXRa, and LXRb. *Pharm. Res.* 19, 1788–1800
48 Zhu, B.T. *et al.* (2006) Quantitative structure–activity relationship of various endogenous estrogen metabolites for human estrogen receptor {alpha} and {beta} subtypes: insights into the structural determinants favoring a differential subtype binding. *Endocrinology* 147, 4132–4150
49 Plewczynski, D. *et al.* (2007) Target specific compound identification using a support vector machine. *Comb. Chem. High Throughput Screen.* 10, 189–196
50 Coumoul, X. *et al.* (2002) PXR-dependent induction of human CYP3A4 gene expression by organochlorine pesticides. *Biochem. Pharmacol.* 64, 1513–1519
51 Ekins, S. and Erickson, J.A. (2002) A pharmacophore for human pregnane X receptor ligands. *Drug Metab. Dispos.* 30, 96–99
52 Ung, C.Y. *et al.* (2007) In silico prediction of pregnane X receptor activators by machine learning approaches. *Mol. Pharmacol.* 71, 158–168
53 Gao, Y.D. *et al.* (2007) Attenuating pregnane X receptor (PXR) activation: a molecular modelling approach. *Xenobiotica* 37, 124–138
54 Khandelwal, A. *et al.* (2008) Machine learning methods and docking for predicting human pregnane X receptor activation. *Chem. Res. Toxicol.* 21, 1457–1467
55 Noy, N. (2007) Ligand specificity of nuclear hormone receptors: sifting through promiscuity. *Biochemistry* 46, 13461–13467
56 Dunker, A.K. *et al.* (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59
57 Dunker, A.K. *et al.* (2002) Intrinsic disorder and protein function. *Biochemistry* 41, 6573–6582
58 Chen, J.W. *et al.* (2006) Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. *J. Proteome Res.* 5, 888–898
59 Iakoucheva, L.M. *et al.* (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323, 573–584
60 Dunker, A.K. *et al.* (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272, 5129–5148
61 Krasowski, M.D. *et al.* (2008) Intrinsic disorder in nuclear hormone receptors. *J. Proteome Res.* 7, 4359–4372
62 Watkins, R.E. *et al.* (2003) Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J. Mol. Biol.* 331, 815–828
63 Watkins, R.E. *et al.* (2003) 2.1A crystal structure of human PXR in complex with the St John's Wort compound hyperforin. *Biochemistry* 42, 1430–1438
64 Watkins, R.E. *et al.* (2002) Structural insights into the promiscuity and function of the human pregnane X receptor. *Curr. Opin. Drug Discov. Dev.* 5, 150–158
65 Watkins, R.E. *et al.* (2001) The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* 292, 2329–2333
66 Xue, Y. *et al.* (2007) Crystal structure of the PXR–T1317 complex provides a scaffold to examine the potential for receptor antagonism. *Bioorg. Med. Chem.* 15, 2156–2166
67 Chrencik, J.E. *et al.* (2005) Structural disorder in the complex of human pregnane X receptor and the macrolide antibiotic rifampicin. *Mol. Endocrinol.* 19, 1125–1134
68 Xue, Y. *et al.* (2007) Crystal structure of the pregnane X receptor–estradiol complex provides insights into endobiotic recognition. *Mol. Endocrinol.* 21, 1028–1038
69 Schuster, D. and Langer, T. (2005) The Identification of ligand features essential for PXR activation by pharmacophore modeling. *J. Chem. Inf. Model.* 45, 431–439
70 Ekins, S. *et al.* (2007) Human pregnane X receptor antagonists and agonists define molecular requirements for different binding sites. *Mol. Pharmacol.* 72, 592–603
71 Lemaire, G. *et al.* (2007) Discovery of a highly active ligand of human pregnane X receptor: a case study from pharmacophore modeling and virtual screening to ''in vivo'' biological activity. *Mol. Pharmacol.* 72, 572–581
72 Ekins, S. *et al.* (2008) Evolution of pharmacologic specificity in the pregnane X receptor. *BMC Evolut. Biol.* 8, 103
73 Sundriyal, S. *et al.* (2008) New PPARgamma ligands based on barbituric acid: virtual screening, synthesis and receptor binding studies. *Bioorg. Med. Chem. Lett.* 18, 4959–4962
74 Schapira, M. *et al.* (2003) Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7354–7359
75 Ray, N.C. *et al.* (2007) Discovery and optimization of novel, non-steroidal glucocorticoid receptor modulators. *Bioorg. Med. Chem. Lett.* 17, 4901–4905
76 Cheng, J.-F. *et al.* (2008) Combination of virtual screening and high throughput gene profiling for identification of novel liver X receptor modulators. *J. Med. Chem.* 51, 2057–2061
77 Wang, C.Y. *et al.* (2006) Identification of previously unrecognized antiestrogenic chemicals using a novel virtual screening approach. *Chem. Res. Toxicol.* 19, 1595–1601
78 Knox, A.J.S. *et al.* (2007) Target specific virtual screening: optimization of an estrogen receptor screening platform. *J. Med. Chem.* 50, 5301–5310
79 Derksen, S. *et al.* (2006) Virtual screening for PPAR modulators using a probabilistic neural network. *ChemMedChem* 1, 1346–1350
80 Lu, I.L. *et al.* (2006) Structure-based drug design of a novel family of PPARgamma; partial agonists: virtual screening, X-ray crystallography, and in vitro/in vivo biological activities. *J. Med. Chem.* 49, 2703–2712
81 Zhao, L. and Brinton, R.D. (2005) Structure-based virtual screening for plant-based ERbeta-selective ligands as potential preventative therapy against age-related neurodegenerative diseases. *J. Med. Chem.* 48, 3463–3466
82 Knox, A.J. *et al.* (2006) Estrogen receptors: molecular interactions, virtual screening and future prospects. *Curr. Top. Med. Chem.* 6, 217–243
83 Shao, D. *et al.* (2004) Identification of novel estrogen receptor alpha antagonists. *J. Steroid Biochem. Mol. Biol.* 88, 351–360
84 Duret, C. *et al.* (2006) Ketoconazole and miconazole are antagonists of the human glucocorticoid receptor: consequences on the expression and function of the constitutive androstane receptor and the pregnane X receptor. *Mol. Pharmacol.* 70, 329–339
85 Wang, H. *et al.* (2007) Activated PXR is a target for ketoconazole and its analogs. *Clin. Cancer Res.* 13, 2488–2495
86 Ekins, S. *et al.* (2008) Computational discovery of novel low micromolar human pregnane X receptor antagonists. *Mol. Pharmacol.* 74, 662–672
87 Reynolds, C.H. *et al.* (2008) Ligand binding efficiency: trends, physical basis, and implications. *J. Med. Chem.* 51, 2432–2438
88 Xie, L. *et al.* (2007) In silico elucidation of the molecular mechanism defining the adverse effect of selective estrogen receptor modulators. *PLoS Comput. Biol.* 3, e217
89 Chong, C.R. and Sullivan, D.J., Jr (2007) New uses for old drugs. *Nature* 448, 645–646
90 Bisson, W.H. *et al.* (2007) Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11927–11932
91 Knox, A.J. *et al.* (2005) Considerations in compound database preparation – ''hidden'' impact on virtual screening results. *J. Chem. Inf. Model.* 45, 1908–1919
92 Zhao, L. and Brinton, R.D. (2005) Structure-based virtual screening for plant-based ERβ-selective ligands as potential preventative therapy against age-related neurodegenerative diseases. *J. Med. Chem.* 48, 3463–3466
93 Kher, S. *et al.* (2007) 2-Aryl-*N*-acyl indole derivatives as liver X receptor (LXR) agonists. *Bioorg. Med. Chem. Lett.* 17, 4442–4446