HHS Public Access

Author manuscript

Toxicol Appl Pharmacol. Author manuscript; available in PMC 2016 March 31.

Published in final edited form as:

Toxicol Appl Pharmacol. 2014 October 1; 280(1): 177–189. doi:10.1016/j.taap.2014.07.009.

Prediction of binding affinity and efficacy of thyroid hormone receptor ligands using QSAR and structure based modeling methods

Regina Politi[1,2], **Ivan Rusyn**[2,*], and **Alexander Tropsha**[1,*]

[1]Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, University of North Carolina, Chapel Hill, NC, 27599, United States

[2]Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC, 27599, United States

Abstract

The thyroid hormone receptor (THR) is an important member of the nuclear receptor family that can be activated by endocrine disrupting chemicals (EDC). Quantitative Structure-Activity Relationship (QSAR) models have been developed to facilitate the prioritization of THR-mediated EDC for the experimental validation. The largest database of binding affinities available at the time of the study for ligand binding domain (LBD) of THRβ was assembled to generate both continuous and classification QSAR models with an external accuracy of R^2=0.55 and CCR=0.76, respectively. In addition, for the first time a QSAR model was developed to predict binding affinities of antagonists inhibiting the interaction of coactivators with the AF-2 domain of THRβ (R^2=0.70). Furthermore, molecular docking studies were performed for a set of THRβ ligands (57 agonists and 15 antagonists of LBD, 210 antagonists of the AF-2 domain, supplemented by putative decoys/non-binders) using several THRβ structures retrieved from the Protein Data Bank. We found that two agonist-bound THRβ conformations could effectively discriminate their corresponding ligands from presumed non-binders. Moreover, one of the agonist conformations could discriminate agonists from antagonists. Finally, we have conducted virtual screening of a chemical library compiled by the EPA as part of the Tox21 program to identify potential THRβ-mediated EDCs using both QSAR models and docking. We concluded that the library is unlikely to have any EDC that would bind to the THRβ. Models developed in this study can be employed either to identify environmental chemicals interacting with the THR or, conversely, to eliminate the THR-mediated mechanism of action for chemicals of concern.

[*]**Contact information:** Alexander Tropsha, PhD, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599; alex_tropsha@unc.edu; Phone: 919-966-2955; Fax: 919-966-0204, or *Ivan Rusyn, MD, PhD, Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC 27599; iir@unc.edu; Phone: 919-843-2596; Fax: 919-843-2596.

The authors declare no further competing financial interests.

Keywords

Endocrine disrupting chemicals; Thyroid hormone receptor; Quantitative structure–activity relationships modeling; Docking; Virtual screening

Introduction

Endocrine disrupting chemicals (EDCs) are natural or synthetic compounds that have the potential to interfere with the endocrine system, often through imitating or blocking endogenous hormones(Rogers *et al.*, 2013). Fetal and early life exposures appear to have more severe effects than exposure in adulthood on developmental, reproductive, cardiovascular, metabolic, and immune systems(Birnbaum and Fenton, 2003; Rubin and Soto, 2009).

EDCs may act via multiple pathways; however one privileged route is through their direct interaction with nuclear receptors (NRs), which leads to perturbation or modulation of downstream gene expression. The thyroid hormone receptors (THR) are important members of the NR family and act as regulators of metabolism, fetal development, bone remodeling, cardiac function, and mental status. Thus, maintenance of normal thyroid function is essential for psychological and physiological human well-being. Long-term exposure to thyroid-disrupting chemicals may potentially result in hypothyroidism and have other significant consequences for human health(Boas *et al.*, 2012).

The majority of THR responses are induced by the thyroid hormone T3(Harvey and Williams, 2002). There are two main isoforms of THR (THRα and THRβ), and each form can be alternatively spliced and differentially localized across tissue types (Izumo and Mahdavi, 1988; Williams, 2000). THRα_1 is highly expressed in cardiac and skeletal muscles accounting for cardiac responses to the endogenous T3. On the other hand, most of the hormonal effects in the liver (including the influence on the cholesterol metabolism), brain and other tissues are mediated through THRβ1(Takeda *et al.*, 1992; Forrest and Vennstrom, 2000). Thus, the ability to recognize environmental chemicals causing THRβ-mediated endocrine disruption is highly important. In addition, agonists and antagonists of THRβ can be used therapeutically for treating several thyroid and non-thyroid disorders. For example, THRβ antagonists serve as therapies for thyrotoxicosis whereas highly selective agonists are used to treat metabolic disorders such as obesity, for lowering cholesterol to treat hyperlipidemia, for amelioration of depression, and for stimulation of bone formation in osteoporosis(Grover *et al.*, 2004; Shoemaker *et al.*, 2012). However, development of newer compounds with increased selectivity is required to achieve higher precision of action and avoid adverse effects such as cardiotoxicity mediated by THRs(Forrest and Vennstrom, 2000).

Several functional domains of THR have been identified, which include a DNA binding domain, a ligand binding domain (LBD), a ligand-independent transactivation domain (termed activation function 1 or AF-1), and a ligand-inducible coactivator binding domain (termed activation function 2 or AF-2)(Kumar and Thompson, 1999). In the absence of the ligand, co-repressor proteins are bound to THR preventing transcriptional activation(Chen

and Evans, 1995). Ligand binding to the LBD causes dissociation of co-repressors and allows recruitment of co-activator proteins to the AF-2 domain to regulate gene transcription(Ribeiro *et al.*, 1998). Many ligands that bind to the LBD have been reported in the literature; for instance, several selective ligands for THRβ have been identified based on the structural similarity to the endogenous thyroid receptor hormones, T4 and T3 (Carlsson *et al.*, 2002; Ye *et al.*, 2003; Hangeland *et al.*, 2004; Hedfors *et al.*, 2005; Garcia Collazo *et al.*, 2006; Koehler *et al.*, 2006; Li *et al.*, 2006; Garg *et al.*, 2007; Malm *et al.*, 2007).

Recently, a series of THR antagonists have been identified that inhibit THR-coactivator interaction by binding to the AF-2 domain (Arnold *et al.*, 2005; Arnold *et al.*, 2006; Arnold *et al.*, 2007a; Arnold *et al.*, 2007b; Estebanez-Perpina *et al.*, 2007a; Hwang *et al.*, 2009; Hwang *et al.*, 2011; Hwang *et al.*, 2012; Hwang *et al.*, 2013). These compounds may be useful to treat metabolic disorders and hyperthyroidism without affecting thyroid hormone levels; however additional studies revealed significant dose-related cardiotoxicity, suspected to arise from ion-channel inhibition(Arnold *et al.*, 2005; Arnold *et al.*, 2006; Arnold *et al.*, 2007a; Arnold *et al.*, 2007b; Estebanez-Perpina *et al.*, 2007a; Hwang *et al.*, 2009; Hwang *et al.*, 2011; Hwang *et al.*, 2012; Hwang *et al.*, 2013).

Computational methods such as Quantitative Structure Activity Relationship (QSAR) modeling have been widely used to prioritize chemicals for *in vivo* or *in vitro* testing that may pose endocrine disruption hazard (Lo Piparo and Worth, 2010; Tsakovska *et al.*, 2011). Several groups have reported QSAR models for the LBD of THR. These models are summarized in Table 1. Although these previous models were reported to have significant predictive power, all of them were created using relatively small datasets with limited chemical diversity and consequently, these models had a limited applicability domain (AD) (Tropsha, 2010). In addition, these previous models were developed to predict THR binding affinity but none was capable of distinguishing the type of functional activity, or efficacy (i.e., agonism *vs.* antagonism) of the ligands. Finally, to date no QSAR studies have been reported to predict biological activity of compounds that bind to the AF-2 domain.

In this study, we have assembled the largest dataset (as compared to all data reported in the open literature) of ligands tested for their interaction with the THR, including data on the $THR\beta_1$ binding affinity (129 compounds binding at the LBD and 181 compounds binding at the AF-2 domain) and functional activity (57 agonists/15 antagonists binding at the LBD, 210 antagonists binding at the AF-2 domain). Using OECD(Organization for Economic Cooperation and Development)-compliant predictive QSAR modeling workflow(Tropsha, 2010), we have developed both continuous and categorical QSAR models for ligands of both the LBD and the AF-2 domain. Furthermore, we have identified co-crystalized complexes between THR and several ligands in the Protein Data Bank that allowed the use of molecular docking to classify ligands binding at the LBD as either agonists or antagonists. The predictive models developed in this study are suitable to use in virtual screening to identify putative THR binders and non-binders in environmental chemical libraries and classify them into agonists or antagonists. Here we present an example of such a study using the EPA Tox21 database of suspected endocrine disrupting chemicals. In addition, these models can be employed to exclude THR-mediated mechanism of endocrine disruption for environmental chemicals of concern.

Materials and Methods

Datasets

THRβ binding affinity—129 unique organic compounds with known binding affinity to the LBD of the THRβ were collected from ChEMBL(Gaulton *et al.*, 2012). Their binding affinities and structures were verified against published literature(Hedfors *et al.*, 2005; Liu and Gramatica, 2007; Malm *et al.*, 2007; Valadares *et al.*, 2007; Vedani *et al.*, 2007; Du *et al.*, 2008). The affinity data were reported as IC_{50} values determined from the radioligand binding assay as described by Ye(Ye *et al.*, 2003). For the AF-2 domain, 210 ligands were found; however, only 181 had well-defined IC_{50} values(Arnold *et al.*, 2007b; Hwang *et al.*, 2009; Hwang *et al.*, 2012) (see Table 2). The IC_{50} values associated with inhibition of co-regulatory peptide SRC2-2 binding to THRβ were determined using fluorescence polarization(Arnold *et al.*, 2007b; Hwang *et al.*, 2009; Hwang *et al.*, 2012).

Collected IC_{50} values varied from 0.0191 nM to 32 μM and from 0.310 μM to 100 μM for LBD and AF-2 domains, respectively. The IC_{50} values were converted to $-\log IC_{50}$ (pIC_{50}). As can be seen in the Supplemental Figure 1, pIC_{50} values of ligands for both domains showed normal distribution; however, most of the AF-2 domain ligands were not very potent. The chemical structures for 129 and 181 compounds able to bind at the LBD and AF-2 domains, respectively, are included in the Supplemental Table 1.

THRβ functional activity—57 known agonists and 15 antagonists for the LBD were obtained from ChEMBL(Gaulton *et al.*, 2012) and their functional annotation was verified using published literature(Carlsson *et al.*, 2002; Ye *et al.*, 2003; Hedfors *et al.*, 2005; Garcia Collazo *et al.*, 2006; Garg *et al.*, 2007). In addition, 5101 presumed decoys were obtained from the DUD-E (Database of Useful Decoys: Enhanced) database (Mysinger *et al.*, 2012). Presumed decoys are defined as chemicals that have similar physical properties but are topologically dissimilar from the known ligand structures and are expected not to bind to the respective receptor. For the AF-2 domain, 210 known antagonists were obtained from the published literature (Arnold *et al.*, 2005; Arnold *et al.*, 2006; Arnold *et al.*, 2007a; Arnold *et al.*, 2007b; Estebanez-Perpina *et al.*, 2007a; Hwang *et al.*, 2009; Hwang *et al.*, 2011; Hwang *et al.*, 2012; Hwang *et al.*, 2013). Since there were no publicly available decoy datasets for AF-2 domain antagonists at the time of this study, the DUD-E package was used to create decoys. All data used for docking studies are summarized in Table 2. Note that datasets used for QSAR and docking validations studies were different, but there was some overlap. For QSAR modeling, we used chemicals with known binding affinities/inhibition activities whereas for docking studies we only needed to know if compounds were agonists, antagonists, or (presumed) decoys, i.e., the knowledge of binding affinity was not essential. The chemical structures of agonists, antagonists and presumed decoys for LBD and AF-2 domain of THRβ are included in Supplemental Tables 2 and 3, respectively.

The 3D conformations of the THRβ ligands and presumed decoys/non-binders were prepared using the LigPrep wizard in Schrodinger Suite (http://www.schrodinger.com/). The ionization state of each molecule was calculated assuming the pH value of 7.0±2.0.

THRβ conformations for docking—Three THRβ crystal structures were used for docking. Two had agonists bound at the LBD: [4-(3-benzyl-4-hydroxybenzyl)-3,5-dimethylphenoxy]acetic acid (PDB ID: 1Q4X) (Borngraeber *et al.*, 2003) and [4-(4-hydroxy-3-isopropyl-phenoxy)-3,5-dimethyl-phenyl]-6-azauracil (PDB ID: 1N46) (Dow *et al.*, 2003). The third structure was of THRβ with an antagonist bound at the AF-2 domain that directly inhibits the association of the receptor with its essential co-activators (PDB ID: 2PIN) (Estebanez-Perpina *et al.*, 2007b). This was the only structure of THRβ with an antagonist at the AF-2 domain, which was found in the PDB at the time of this study. The protein-ligand crystal structures are shown in the Supplemental Figure 2. The crystallographic structures found in the PDB were treated using the Protein Preparation wizard in Schrodinger Suite (http://www.schrodinger.com/) (Mouchlis *et al.*, 2010). A grid for each protein was calculated with the Grid Generation module in Schrodinger Suite, with the binding site defined as the location of the respective co-crystallized ligands.

Data curation—Stuctures for all compounds employed in this study were manually examined and curated according to the guidelines described elsewhere (Fourches *et al.*, 2010). All curated structures were stored in the SDF format for further analysis.

Molecular Descriptors

2D molecular descriptors for compounds represented with explicit hydrogen atoms were computed using Dragon software (version 5.4; Talete s.r.l., Milan, Italy). Descriptors with low variance (standard deviation lower than 0.01) or missing values were removed. Furthermore, if the squared correlation coefficient (R^2) between values of two descriptors over the entire data set exceeded 0.95, one of the descriptors was removed. The final descriptor set used in this study for ligands binding at the LBD or AF-2 domains contained 508 and 493 descriptors, respectively. The descriptors were range-scaled to the [0, 1] interval.

QSAR Modeling Approaches

Training, test, and external evaluation sets selection—Random Forest (RF) machine learning technique implemented in R was used for modeling. To avoid well-documented limitations of QSAR models developed with training sets only (Golbraikh and Tropsha, 2002), each dataset (consisting of 129 unique organic compounds for the LBD and of 181 unique compounds for the AF-2 domain) was subjected to 5-fold external cross-validation procedure as detailed elsewhere (Sedykh *et al.*, 2011). Specifically, each dataset was randomly partitioned into five subsets of similar size. Models were then independently developed such that compounds in four of the five subsets were used as modeling set and compounds in the remaining subset were used as an external evaluation set. The data within each modeling set were further divided into multiple pairs of training and test sets for internal validation. Individual models were developed based on each internal training set and internally validated by predicting the corresponding test set. All individual models showing acceptable performance on internal training/test sets were retained in an ensemble for the application to the external evaluation set; thus, the latter set was not used in any way in model development or internal validation. This procedure was repeated five times such that

each of the five subsets were employed as an external evaluation set once while the remaining subsets were used as a modeling set.

Validation of QSAR models—All reported model predictivity measures such as specificity, sensitivity and CCR for classification models and R^2 for continuous models were obtained from 5-fold external cross-validation. Specificity denotes the true negative rate, or the rate of correctly predicted compounds with low activity (pIC_{50}>10nM for ligands of the LBD). Conversely, sensitivity, or true positive rate denotes the rate of correctly predicted compounds with high activity (pIC_{50}≤10nM). CCR is the average of the rates correctly predicted within each class.

$$\mathrm{CCR}=[\mathrm{specificity}+\mathrm{sensitivity}]/2 \quad (1)$$

R^2 is the slope of a regression line (observed *vs*. predicted) forced through the origin.

Robustness of QSAR models—Y-randomization (randomization of response) is widely used to establish robustness of QSAR models(Rucker *et al*., 2007). The process involves rebuilding models using randomized activities of the training set and subsequent assessment of the model statistics. It is expected that models obtained for the training set with randomized activities should have significantly lower accuracy than those built using a training set with real activities. If this condition is not satisfied, models built for a training set with real activities are considered as not reliable and should be discarded. Y-randomization was applied to all training/test dataset divisions considered in this study following protocols as described previously (Golbraikh and Tropsha, 2002).

Docking Studies

All THRβ agonists and antagonists, as well as their respective presumed decoys/non-binders detailed in Table 2, were docked into the corresponding protein structures using Glide SP (Friesner *et al*., 2004) with default setting for flexible ligand docking as implemented in Schrodinger Suite. Protein conformations with each cognate ligand used for docking studies with the LBD or the AF-2 domain are shown in the Supplemental Figure 2. Glide results for each ligand are grouped and sorted by Emodel (scoring function used by Schrodinger Suite for selecting the "best" pose of a ligand), and then the ligand blocks are sorted according to the empirical scoring function (gscore) of the top member. This scoring function approximates the ligand binding free energy. The performance of the docking models in this study were assessed not only by ability to correctly distinguish known ligands from a larger pool of compounds but also discriminate between agonists and antagonists when possible.

To evaluate the efficiency of docking methods, the enrichment factors and receiver operating characteristic (ROC) curves were calculated. The enrichment factor (EF; Equation 2) reflects how many seed compounds (or known ligands) were found within a defined “early recognition” fraction of the ranked list relative to a random distribution:

$$\mathrm{EF}=H_{\mathrm{scr}}/H_{\mathrm{tot}} \times D_{\mathrm{tot}}/D_{\mathrm{scr}} \quad (2)$$

where H_{scr} is the number of target-specific ligands recovered at a specific % level of the ligand/decoy datasets; H_{tot} is the total number of ligands for the target; D_{scr} is the number of compounds screened at a specific % level of the database; D_{tot} is the total number of compounds in the database. The ROC curves were generated by plotting sensitivity (Equation 3) against [1 – specificity (Equation 4)] for a binary classifier system as its discrimination threshold is varied. In the case of virtual screening for recovering the i^{th} known active from the inactive decoys (or presumed decoys), the sensitivity and specificity were defined as follows:

$$\text{Sensitivity(SE)} = H_{\text{scr}}/H_{\text{tot}} \quad (3)$$

$$\text{Specificity(SP)} = \frac{(D_{\text{tot}} - H_{\text{tot}}) - (D_{\text{scr}} - H_{\text{scr}})}{D_{\text{tot}} - H_{\text{tot}}} \quad (4)$$

The area under the ROC curve is the metric widely accepted for assessing the likelihood that a screening method assigns a higher rank to known actives than to inactive compounds. The area under the curve values at a specific percentage of the ranked database were calculated from Equation (5) where n is the total number of known actives in the screening database.

$$\text{Area under the curve} = \sum_{i=1}^{n} [\text{SE}_i(\text{SP}_{i+1} - \text{SP}_i)] \quad (5)$$

Virtual screening

Out of 129 compounds used to develop QSAR model for the LBD, 101 compounds did not have a functional annotation as either agonists or antagonists. These 101 compounds were used as an external set for virtual screening to predict their annotation using docking.

Recently, approximately 8000 chemicals were screened by the EPA as part of their Tox21 program (Tice *et al.*, 2013) in order to detect compounds with potential to interfere with the endocrine system. One of the assays used was the luciferase reporter gene assay that was developed based on the TH-responsive rat pituitary tumor GH3 cell line, which constitutively expresses both THR isoforms (Freitas *et al.*, 2011). This assay was confirmed as an *in vitro* tool for identification and quantification of specific THR-disrupting chemicals (Freitas *et al.*, 2011). Moreover, this assay allowed discriminating between THR agonists and antagonists. Chemicals tested in this assay were virtually screened using both QSAR and molecular docking models developed in this study.

Results and Discussion

Testing of the previously published models

For this study we have assembled the largest dataset of IC_{50} values for a series of THRβ ligands reported thus far (Table 2). This dataset included compounds used in previously published models of THRβ ligands that were developed using both conventional QSAR (Liu and Gramatica, 2007; Valadares *et al.*, 2007) and multidimensional QSAR with docking

(Vedani *et al.*, 2007) approaches (cf. Table 1). This allowed us to test these previously published models for their ability to predict new data, i.e., those compounds in our collection that were not employed in previous studies. Specifically, we have evaluated a multidimensional QSAR model developed by the Vedani group (Vedani *et al.*, 2007), which is publicly available via their web portal (http://www.biograf.ch/index.php?id=projects&subid=virtualtoxlab) as well as the MLR QSAR model developed by Gramatica group (Liu and Gramatica, 2007). Specific DRAGON descriptors and respective regression coefficients employed by the latter model were published (Liu and Gramatica, 2007) allowing us both to recreate the model and apply it to new compounds. We have identified 47 new compounds in our extended dataset tested for their binding with the LBD using the same experimental assay (Carlsson *et al.*, 2002; Hangeland *et al.*, 2005; Garcia Collazo *et al.*, 2006; Koehler *et al.*, 2006; Garg *et al.*, 2007; Malm *et al.*, 2007) that were not part of the datasets used in previous modeling studies. For both models, no correlation between predicted and experimental data for compounds not used in developing the original model was observed (Supplemental Figure 3). Our further analysis of the Gramatica model (Liu and Gramatica, 2007) revealed that descriptors were selected to optimize the MLR model using the entire dataset and then coefficients were adjusted for the respective training set. Such an approach using variable selection does not prove the external predictivity of the models since the training set should be first set aside and then followed by descriptor selection. Only then should the training set model be validated using a completely independent test set, which was never used in any way for model development (Dearden *et al.*, 2009).

Parameters used in the model published by the Vedani group were not available making it impossible for us to reproduce the model. However, our negative external validation results indicate that most likely this model also suffers from overfitting. Thus, negative results of preliminary evaluating the best published models reinforced our desire to develop new models of THRβ binding using both an expanded dataset (Table 2) and rigorous external validation procedures.

QSAR Modeling of Ligand Binding Domain (LBD)

We have developed continuous QSAR models using the RF modeling approach and Dragon descriptors. The model showed relatively low but statistically significant external predictive accuracy for experimental *vs.* predicted pIC_{50} (R^2=0.55 for 5-fold external cross validation) (Figure 1a). Figure 1b shows absolute prediction errors, ΔpIC_{50} (calculated as absolute difference between experimental and predicted pIC_{50}) plotted *vs.* experimental pIC_{50}. Larger errors were observed for compounds with the highest and the lowest activities. Furthermore, the panel in Figure 1b shows frequency distribution of absolute prediction errors, ΔpIC_{50}. Overall, about 50% of compounds had an absolute prediction error of greater than 0.5 pIC_{50} units.

Next, we built a binary classification QSAR model (Figure 2) discriminating strong *vs.* weak binders. The threshold to classify THRβ ligands into strong and weak binders was set to pIC_{50} of 8 (i.e., 10 nM), below which the compounds were classified as weak binders and above which they were classified as strong binders; this threshold was chosen to achieve a

balance between the numbers of strong and weak binders. Because compounds close to the margin were expected to be predicted with less confidence we have evaluated the effect of changing the threshold on the prediction performance. While alternative thresholds yielded models with slightly higher CCR (data not shown), the underlying datasets were unbalanced with respect to the ratio of strong *vs*. weak binders. The models developed with pIC_{50} threshold of 8 showed good separation between weak and strong THRβ binders with a CCR of 0.76 and balanced specificity and sensitivity of 0.7 and 0.83, respectively (Figure 2a). Since most of the ligands in the dataset were within the AD, filtering by the AD did not have a significant effect on the predictive accuracy. To better understand where the largest mis-predictions were most likely to occur, the frequency of mis-predicted compounds at each activity bin was calculated (Figure 2b). As expected, the largest percentage of mis-predicted compounds using the classification model was found around the threshold value. Conversely, the largest percent of mis-predictions for the continuous model was found at the margins of the activity distribution (Supplemental Figure 1) and the predicted activities ranged from pIC_{50} values of 5.8 to 9.8, which was more narrow than the experimental range of pIC_{50} values between 4.5 and 10.7 (Figure 1a).

In addition to building the binary QSAR model we have also evaluated the accuracy of predicting the compound class when predicted continuous values were converted to respective classes using the same threshold of pIC_{50}=8. Interestingly, we found that this results in the slight increase of classification accuracy (CCR of 0.80 without the AD) (Figure 3a). Although the prediction accuracy did not change much, the number of misclassified chemicals at each activity bin was reduced and the maximum percentage of such chemicals per bin was 40% (Figure 3b), as opposed to 52% in predictions resulting from the classification model (Figure 2b). This analysis suggests that the most accurate predictions for an external dataset can be obtained by using continuous model and then converting the predicted values into the respective categories, i.e., high or low affinity binders (using the pIC_{50} threshold of 8).

QSAR Modeling of the AF-2 Domain dataset

Recent studies have identified a novel set of 210 compounds that disrupt the interaction between $THR\beta_1$ and its co-activator, SRC2, in the presence of T3 (Arnold *et al*., 2005; Arnold *et al*., 2006; Arnold *et al*., 2007a; Arnold *et al*., 2007b; Estebanez-Perpina *et al*., 2007a; Hwang *et al*., 2009; Hwang *et al*., 2011; Hwang *et a.l*, 2012; Hwang *et al*., 2013); all of these studies were conducted by the same group. However, the IC_{50} values have been reported for only 181 compounds (Table 2 and Supplemental Table 1). Predictive models using this data were developed using the RF machine learning technique and the Dragon 2D descriptors as described in Methods. Figure 4a shows good external predictive accuracy for experimental *vs*. predicted pIC_{50} (R^2=0.70 for 5-fold external cross validation). Frequency distribution of the absolute prediction errors, shown in Figure 4b, indicates a relatively small range of errors with mean absolute error of 0.24 when more than 70% of compounds have ΔpIC_{50} of less than 0.3. Y-randomization testing, as briefly described in Methods, was conducted for the AF-2 domain model. We have demonstrated that the accuracy of a QSAR model built with the original data was significantly higher than that of models built using Y-

randomized datasets (p-value <0.05). This result implies that the high accuracies of QSAR models of the AF-2 dataset were not due to a spurious correlation.

Model Interpretation for the Ligand Binding and the AF-2 Domains

Previously, we have developed QSAR models for predicting the estrogen receptor (ER) binding affinity (Zhang *et al.*, 2013). By analyzing these QSAR models we have extracted significant chemical features that influence ER binding affinity; these features can be used to suggest structural modifications to diminish the ER binding potential of chemicals. To analyze chemical features that influence THR binding affinity, descriptors were ranked based on their importance: a Z-score was calculated for each descriptor and the values were normalized. Figure 5 shows the 35 most statistically significant descriptors ($p<0.05$ by Z-test) for the LBD dataset based on a continuous model. For these descriptors we have compared mean values for strong and weak THR binders defined by the activity threshold of 1nM and 1μM, respectively (those with a binding affinity below 1 nM were considered strong binders and those with a binding affinity above 1 μM were considered weak binders). From these values, we have calculated the impact (i.e., the difference between the mean descriptor values for all strong *vs.* all weak binders) of individual descriptors on the relative binding affinity to THRβ. Figure 5a shows that many descriptors exhibited substantially different mean values between the two groups, with most descriptors having higher values for strong binders. Such variation implies that these descriptors could potentially serve as determinants of the THR binding affinity. Indeed, Figure 5b shows patterns of chemical descriptor profiles for a typical strong binder in comparison with a typical weak THR binder. This plot illustrates that not only average but also individual chemical descriptor profiles show appreciable divergence between values of most descriptors for strong *vs.* weak binders.

Figure 6 shows the 27 most statistically significant descriptors ($p<0.05$ by Z-test) for the AF-2 dataset based on the respective continuous model. For these descriptors we have compared mean values for the strong and weak AF-2 domain binders defined by activity threshold of 1μM and 10μM, respectively (i.e., those with the a binding affinity below 1μM are considered strong and those with a binding affinity above 10μM are considered weak) and calculated the impact of individual descriptors on relative binding affinity to the AF-2 domain as described above for the LBD model. Although the ligand descriptors for the AF-2 domain dataset is very difficult since most of the compounds in the dataset are not very potent, Figure 6a shows that many descriptors still exhibited substantially different mean values between the two groups of binders. Figure 6b shows patterns of chemical descriptor profiles for the strongest and weakest THR binders (pIC_{50} above 6.5 or below 4, respectively).

In many cases, the nature of Dragon descriptors employed in this study does not allow for straightforward model interpretation in terms of chemical functional groups. However, some Dragon descriptors found to be statistically significant in QSAR models could still be useful in assessing what chemical modification may affect THR binding potential. Several examples of such modifications affecting the binding affinity of known LBD and AF-2 domain binders are shown in Table 3. The underlying descriptors may have either positive or negative impacts on binding affinity as shown in Figures 5a and 6a. For instance, one of the

modifications shown in Table 3 is the substitution of CF_2HO for hydrogen that resulted in the increase of both experimental (from 8.54 to 9.1) and predicted pIC_{50} (from 7.76 to 8.65). The edge adjacency index EEig03d (Eigenvalue 3 from edge adjacency matrix weighted dipole moments) was found to be the second most important descriptor that had a positive impact on compound binding affinity to the LBD of the THRβ (cf. Figure 5). This descriptor correlates with molecular polarity; its value increased from 0.44 to 0.74 as a result of the structural modification shown in the first example in Table 3. Interestingly, in the previous QSAR study of selective ligands for THRβ using Dragon descriptors(Vedani *et al.*, 2007), descriptor of the same type was found to significantly affect binding affinity of chemicals to the LBD of THRβ.

To summarize this part, the RF modeling procedure allows estimating the relative importance of the descriptors employed by the model (Kuz'min *et al.*, 2011). By interpreting the significant molecular descriptors, it is possible to gain insight into structural features responsible for the binding affinity of ligands to THR. Furthermore, the interpretation of molecular descriptors can help with guiding structural modifications that will modulate binding affinity of the ligands to THR. However, although the examples discussed above illustrate that individual highly significant descriptors resulting from QSAR modeling may be helpful in interpreting or predicting the observed changes in binding affinity when pairs of ligands are compared, such an analysis should be conducted with an extreme care. The reason for this cautionary note is that the underlying models employing these significant descriptors are created as a result of multivariate statistical analysis of the experimental data and the significance of individual descriptors is evaluated only in the context of other descriptor variables contributing to the model. If this were not true it would be very easy to obtain a high accuracy of discrimination between two ligand classes with single variables that would be naturally revealed by RF, which is obviously not the case. These considerations explain why it is not always true that statistically significant descriptors resulting from multivariate model are found universally significant at the level of an individual pair or a group of compounds.

Indeed, there are statistically significant descriptors found among those shown in Figures 5a and 6a that nevertheless have small differences in their average values for the strongest and the weakest binders. One should also remember that when a new chemical feature is added to a compound, not only does it cause a change in a particular descriptor value but other descriptors change their values as well. For instance, consider the first example in Table 3, where the value of the edge adjacency index EEig03d is increased. This change results in 30 of the 35 significant descriptors (i.e., 86%) for the same pair of compounds to change their values as well. Some of these changes are expected, i.e., the values for those descriptors that generally have positive impact on the binding affinity (cf. Figure 5a) do increase; however, there are also descriptors that change their values against such expectations. For instance, Supplemental Figure 5 shows a comparison of descriptor profiles for the three LBD compounds shown in Table 3 that have different activity. Indeed, the trend in changing descriptor values from compound to compound is for the most part as expected from the analysis of their relative impact on binding affinity, but there are cases when the trend is reversed. For example, the value of descriptor nCb- (cf. Table 3) increases with activity as expected from Figure 5a. However, descriptor BLTA96, that is supposed to have the same

tendency as descriptor nCb- to increase its value with the increase in binding affinity, has the largest value for a compound with the lowest binding affinity.

Similarly, two significant descriptors found as a result of model generation for the AF-2 domain and chosen for illustration purposes show positive correlation between binding affinity and their relative impact. However, while the activity value predicted by the model is quite accurate (predicted value of pIC_{50} is 5.62 *vs.* an experimental value of 6.04), even significant individual descriptors (i.e., chemical alerts) can not be used to predict the direction of the change in activity. This analysis suggests a concordant utility of chemical alerts resulting from model interpretation and quantitative predictions of a binding affinity (or the binder category) made by the models. For instance, structural alerts resulting from model interpretation can drive the design of compounds with a lower toxicity that are modified to exclude a specific structural group (recognized as a toxicity alert). However, each single descriptor (mapped on a structural alert) in a compound can not be viewed in isolation from other descriptors that also change their values in response to such modification. The global effect of any single sructural modification can only be predicted based on multivariate QSAR model; this task can not be accomplished by looking at a structural alert alone. This analysis suggests the synergistic use of the chemical alerts and QSAR models for designing novel compounds and predicting their toxicity, respectively.

Docking Studies

It is important to predict the functional behavior, or efficacy of chemicals, i.e., whether binders will act as receptor agonists or antagonists. We have recently employed structure-based docking studies using agonist- or antagonist-bound conformations of the x-ray characterized estrogen receptor (ER)(Zhang *et al.*, 2013). We found that a molecular docking approach could discriminate known receptor ligands from presumed non-binders. Most importantly, agonist-bound ER conformations were also able to discriminate between agonists and antagonists.

Herein, we have employed a similar molecular docking approach in an attempt to discriminate binders *vs.* non-binders for both the LBD and AF-2 domain of THR as well as agonists *vs.* antagonists for the LBD (only antagonists for the AF-2 domain have been reported). To this end, we have used x-ray characterized THR structures deposited to the PDB. To the best of our knowledge no receptor conformation with an antagonist bound at the LBD was deposited into the PDB by the time of this study. Docking studies were conducted with 57 THR agonists, 15 antagonists and 5101 presumed decoys for the LBD as well as using 210 antagonists and 12249 presumed decoys for the AF-2 domain. Compounds binding to the LBD were docked to two agonist THRβ protein conformations (see Methods), and enrichment factors and AUCs were calculated to establish the discriminatory power of this structure-based functional annotation of the THRβ ligands (Figure 7). We found that these protein conformations were able to discriminate their corresponding ligands (agonists) from the presumed decoys effectively (Figure 7a and 7b). Moreover, all protein conformations could successfully enrich their corresponding ligands with high selectivity (EF_{max} of 91) (Figure 7a and 7b). These results indicate that each receptor conformation is capable of accurately recognizing the type of molecules it is expected to bind.

Importantly, one of the agonist conformations of THRβ, i.e., PDB 1Q4X, was capable of separating agonists from antagonists (Figure 7a, AUCs for agonists and antagonists were 0.9 and 0.58, respectively). Another agonist conformation of THRβ, i.e., PDB 1N46, was also capable of separating binders, either agonists or antagonists, from their presumed decoys; however the AUC for agonists was only slightly better than for antagonists (Figure 7b, AUCs for agonist and antagonists are 0.94 and 0.90, respectively).

Structure-based docking studies performed at the AF-2 domain, i.e., with PDB 2PIN resulted in failure to discriminate between binders (antagonists) and presumed decoys (Figure 7c). The AF-2 domain contains a hydrophobic groove located at the receptor surface. It has been stated in the literature that docking becomes very challenging when the binding pocket is shallow and solvent exposed (Schulz-Gasch and Stahl, 2003); such exercises often lead to an inaccurate ranking of ligands. Proper parameterization of solvation effects is a major limitation of current scoring functions used to rank ligand binding. In addition, it was previously shown that nonspecific receptor ligands, such as those used in the case of the AF-2 domain are mostly false positives (Schapira *et al.*, 2003).

Virtual Screening

After putative THRβ binders were selected by the QSAR-based predictions of binding affinity, the docking models were used to establish the functional activity of each binder. The agonist conformation, PDB 1Q4 demonstrated the best performance in discriminating agonist and antagonist ligands bound at the LDB of THRβ; it was therefore selected to characterize 101 ligands with unknown functional activity from the compounds used for QSAR modeling. The results of virtual screening are shown in Figure 8. In the process of docking known agonists and antagonists into 1Q4X the first known antagonist had a docking *gscore* of −9.95. The *gscore* of the first known agonist docked into 1Q4X was −13.96 and the *gscore* of the first active compound out of all 101 screened compounds was −14.18. Analysis *of gscores* for the active compounds screened and the ROC curves for known agonists docked (Figures 7a and 7b) lead us to assume that the actives are most likely to be agonists.

As mentioned above, the luciferase reporter gene *in vitro* assay was used as part of the EPA Tox21 screening program (Tice *et al.*, 2013) to identify and quantify the potency of specific THR disrupting chemicals. Moreover, this assay was reported to be capable of detecting both agonists and antagonists (Freitas *et al.*, 2011). Out of 8000 chemicals screened, 629 were reported as either agonists or antagonists (Tice *et al.*, 2013). Both QSAR models and docking were used to assess the 629 chemicals from this screening dataset. However, no correlation was found between the reported activities and binding affinities predicted by QSAR models (Supplemental Figure 6a). In addition, we could not classify the Tox21 chemicals into actives and inactives similar to the classification done for the modeling set because data for the Tox21 chemicals were reported using different units than the binding affinity expressed as pIC_{50} for the modeling set. Regardless of this limitation, none of the chemicals was predicted as active when the classification model was applied to the same dataset. Furthermore, docking studies also failed to identify any of the Tox21 compounds as binders (Supplemental Figure 6b).

These negative results could be, of course, explained by the limited prediction power of our computational models when applied to the external Tox21 dataset; in fact, we have reported above the inability of models previously published by other groups to accurately predict new compounds reported in the literature. However, there are several considerations that should be discussed here to support the notion that our case is different. One important difference is that all of the data used to develop and validate previous models were obtained in THR binding experiments so we still suggest that previous QSAR models failed to predict new compounds due to issues related to the over fitting of such data. In contrast, the reporter gene assay used to evaluate compounds in the Tox21 library measures luciferase activity on the lysed cells and not direct binding affinity as was done for the training set chemicals. We suggest that the lack of a plausible outcome for our computational predictions could be explained by the incompatibility of biological assays used for the training and Tox21 datasets as well as by mechanisms influencing the outcome of the gene reporter assays for most Tox21 compounds (with possible exception of T3 and T4 used to validate the assay(Freitas *et al.*, 2011)) that do not involve binding to the THR.

To support this supposition, Figure 9 shows the results of a network analysis of compound chemical similarity conducted with an open source software, gephi (https://gephi.org/). Each node represents a chemical from either the modeling or Tox21 datasets, respectively color-coded blue or red. An edge connects chemicals with the Tanimoto similarity score higher than 0.8. This figure shows that chemical space occupied by the modeling dataset is relatively small: all the chemicals in the modeling set are highly similar and form a relatively tight cluster except for very few compounds with low activity. This result is expected since the active site of the THR is uniquely suited to accommodate compounds with similar scaffolds (most of the known active chemicals are derivatives of an endogenous hormone, T3). Chemical space, occupied by both the modeling and virtual screening Tox21 sets, is visualized in Figure 9, which shows that chemicals tested by Tox21 are very different from those in the modeling dataset. This observation explains why these chemicals were not accurately predicted by the QSAR models built with the current training set, and indeed the absolute majority of active Tox21 chemicals are found outside of the AD for the QSAR models. Furthermore, results from the docking studies, which do not have the AD limitation by default, further suggest that Tox21 compounds do not bind to the THR. Thus, our computational analysis raises an important question as to whether the results of the luciferase gene reporter assays can be unambiguously interpreted in terms of compound binding to the THR receptors. The luciferase gene reporter assays may be still valuable to detect EDCs acting via THR-dependent mechanisms. However, we believe that additional experimental studies using chemicals from our modeling sets with known THR binding affinities to either the LBD or AF-2 domain are needed to validate these assays in terms of detecting EDC chemicals acting via direct interaction with the THR.

Conclusions

We have described the generation of both continuous and classification QSAR models for the LBD of THRβ based on the largest dataset available at the time of this study, which included 129 chemicals. In addition, a QSAR study was performed for the first time to predict biological activity for compounds binding at the AF-2 domain of THRβ. All models

were developed using an OECD-compliant validation workflow and analyzed for their most statistically significant chemical descriptors. As a result of this analysis we have concluded that the task of rational design of new compounds with the desired THR binding by targeted modification of an existing compound cannot be accomplished on the basis of THR-dependent structural toxicity alerts revealed by QSAR model interpretation. We showed that although individual alerts could be revealed by mapping some of the significant molecular descriptors onto functional chemical groups, the multivariate nature of the underlying statistical QSAR models cannot reliably single out one or several descriptors. In fact, most of the focused structural modifications affect multiple descriptor values (cf. Supplemental Figure 5 and Table 3). However, such individual alerts could be very helpful in suggesting specific design strategies, and the impact of such design on the modified compound binding affinity can be evaluated by the QSAR model.

The majority of the previous studies reported in the literature pursued either statistical QSAR modeling or chemical alert strategies (often combined with read across (Low *et al.*, 2013)) in evaluating the toxicity of new chemicals. As discussed above, the analysis of QSAR modeling results in our study suggests the synergistic use of chemical alerts (revealed by model interpretation) and QSAR models for designing novel compounds and predicting their toxicity.

In addition to ligand based strategies, structure-based docking was used to complement QSAR models by predicting the functional activity (agonism or antagonism) of compounds able to bind at the LBD of THRβ. The two agonist protein conformations tested were able to discriminate corresponding ligands from presumed decoys. Moreover, one of the agonist conformations was also able to discriminate between agonists and antagonists. This is the second time where we show the successful use of the structure-based method as a complement to ligand-based approaches for evaluating potential endocrine disruptors (cf. our previous study of ER ligands (Zhang *et al.*, 2013)). However, we also present the limitations of this method, which are related to the type of the binding pocket being evaluated (i.e., a surface or buried binding site). Thus, structure-based docking studies performed with the surface-exposed AF-2 domain resulted in failure to discriminate between binders (antagonists) and presumed decoys probably due to problem of active-site definitions and the scoring of ligand binding complicated by solvation effects. Conversely, despite the fact that most antagonists of the AF-2 domain were not very potent, the respective QSAR model had a good external predictive accuracy for experimental *vs.* predicted pIC_{50} (R^2=0.70 for 5-fold external cross validation) with a mean absolute error of 0.24 when more than 70% of compounds had a ΔpIC_{50} of less than 0.3. This result encourages us to use the QSAR model for virtual screening of chemical libraries to identify hits with an ability to treat metabolic disorders and hyperthyroidism without affecting thyroid hormone levels. Finally, the computational strategies employed in both this and the previous (Zhang *et al.*, 2013) studies should be expanded to other nuclear receptors involved in the endocrine disruption caused by environmental chemicals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to Dr. Yen Low for technical advice, and Drs. Alexander Sedykh and Denis Fourches for fruitful discussions and helpful suggestions. The authors greatly appreciate fruitful discussions with Drs. Kevin Crofton and Keith Houck from EPA and the editorial assistance provided by Mr. Jason Harris. The authors also acknowledge the financial support by grants from NIH (GM076059) and EPA (RD83382501).

List of abbreviations

AD	Applicability Domain
AUC	Area Under the Curve
T3	3,5,3’-triiodo-L-thyronine
THR	Thyroid Hormone Receptor
EDCs	Endocrine Disrupting Chemicals
EF	Enrichment Factor
EPA	US Environmental Protection Agency
PDB	Protein Data Bank
QSAR	Quantitative Structure-Activity Relationships
RF	Random Forest
ROC	Receiver Operating Characteristic
SDF	Structure-Data File

References

Arnold LA, Estebanez-Perpina E, Togashi M, Jouravel N, Shelat A, McReynolds AC, Mar E, Nguyen P, Baxter JD, Fletterick RJ, Webb P, Guy RK. Discovery of small molecule inhibitors of the interaction of the thyroid hormone receptor with transcriptional coregulators. J. Biol. Chem. 2005; 280:43048–43055. [PubMed: 16263725]

Arnold LA, Estebanez-Perpina E, Togashi M, Shelat A, Ocasio CA, McReynolds AC, Nguyen P, Baxter JD, Fletterick RJ, Webb P, Guy RK. A high-throughput screening method to identify small molecule inhibitors of thyroid hormone receptor coactivator binding. Sci STKE. 2006:3.

Arnold LA, Kosinski A, Estebanez-Perpina E, Fletterick RJ, Guy RK. Inhibitors of the interaction of a thyroid hormone receptor and coactivators: preliminary structure-activity relationships. J. Med. Chem. 2007a; 50:5269–5280. [PubMed: 17918822]

Arnold, La; Kosinski, A.; Estébanez-Perpiñá, E.; Fletterick, RJ.; Guy, RK. Inhibitors of the interaction of a thyroid hormone receptor and coactivators: preliminary structure-activity relationships. J. Med. Chem. 2007b; 50:5269–5280. [PubMed: 17918822]

Birnbaum LS, Fenton SE. Cancer and developmental exposure to endocrine disruptors. Environ. Health Perspect. 2003; 111:389–394. [PubMed: 12676588]

Boas M, Feldt-Rasmussen U, Main KM. Thyroid effects of endocrine disrupting chemicals. Mol. Cell. Endocrinol. 2012; 355:240–248. [PubMed: 21939731]

Borngraeber S, Budny MJ, Chiellini G, Cunha-Lima ST, Togashi M, Webb P, Baxter JD, Scanlan TS, Fletterick RJ. Ligand selectivity by seeking hydrophobicity in thyroid hormone receptor. Proc. Natl. Acad. Sci. U. S. A. 2003; 100:15358–15363. [PubMed: 14673100]

Carlsson B, Singh BN, Temciuc M, Nilsson S, Li Y-L, Mellin C, Malm J. Synthesis and Preliminary Characterization of a Novel Antiarrhythmic Compound (KB130015) with an Improved Toxicity Profile Compared with Amiodarone. J. Med. Chem. 2002; 45:623–630. [PubMed: 11806713]

Chen JD, Evans RM. A transcriptional co-repressor that interacts with nuclear hormone receptors. Nature. 1995; 377:454–457. [PubMed: 7566127]

Dearden JC, Cronin MT, Kaiser KL. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). SAR QSAR Environ. Res. 2009; 20:241–266. [PubMed: 19544191]

Dow RL, Schneider SR, Paight ES, Hank RF, Chiang P, Cornelius P, Lee E, Newsome WP, Swick AG, Spitzer J, Hargrove DM, Patterson TA, Pandit J, Chrunyk BA, LeMotte PK, Danley DE, Rosner MH, Ammirati MJ, Simons SP, Schulte GK, Tate BF, DaSilva-Jardine P. Discovery of a novel series of 6-azauracil-based thyroid hormone receptor ligands: potent, TR beta subtype-selective thyromimetics. Bioorg. Med. Chem. Lett. 2003; 13:379–382. [PubMed: 12565933]

Du J, Qin J, Liu H, Yao X. 3D–QSAR and molecular docking studies of selective agonists for the thyroid hormone receptor beta. J. Mol. Graph. Model. 2008; 27:95–104. [PubMed: 18436460]

Estebanez-Perpina E, Arnold LA, Jouravel N, Togashi M, Blethrow J, Mar E, Nguyen P, Phillips KJ, Baxter JD, Webb P, Guy RK, Fletterick RJ. Structural insight into the mode of action of a direct inhibitor of coregulator binding to the thyroid hormone receptor. Mol. Endocrinol. 2007a; 21:2919–2928. [PubMed: 17823305]

Estebanez-Perpina E, Arnold LA, Jouravel N, Togashi M, Blethrow J, Mar E, Nguyen P, Phillips KJ, Baxter JD, Webb P, Guy RK, Fletterick RJ. Structural insight into the mode of action of a direct inhibitor of coregulator binding to the thyroid hormone receptor. Mol. Endocrinol. 2007b; 21:2919–2928. [PubMed: 17823305]

Forrest D, Vennstrom B. Functions of thyroid hormone receptors in mice. Thyroid. 2000; 10:41–52. [PubMed: 10691312]

Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J. Chem. Inf. Model. 2010; 50:1189–1204. [PubMed: 20572635]

Freitas J, Cano P, Craig-Veit C, Goodson ML, Furlow JD. Detection of thyroid hormone receptor disruptors by a novel in vitro reporter gene assay. Toxicol. In Vitro. 2011; 25:257–266. [PubMed: 20732405]

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 2004; 47:1739–1749. [PubMed: 15027865]

Garcia Collazo A-M, Koehler KF, Garg N, Färnegårdh M, Husman B, Ye L, Ljunggren J, Mellström K, Sandberg J, Grynfarb M, Ahola H, Malm J. Thyroid receptor ligands. Part 5: novel bicyclic agonist ligands selective for the thyroid hormone receptor beta. Bioorg. Med. Chem. Lett. 2006; 16:1240–1244. [PubMed: 16338239]

Garg N, Li YL, Garcia Collazo AM, Litten C, Ryono DE, Zhang M, Caringal Y, Brigance RP, Meng W, Washburn WN, Agback P, Mellstrom K, Rehnmark S, Rahimi-Ghadim M, Norin T, Grynfarb M, Sandberg J, Grover G, Malm J. Thyroid receptor ligands. Part 8: Thyromimetics derived from N-acylated-alpha-amino acid derivatives displaying modulated pharmacological selectivity compared with KB-141. Bioorg. Med. Chem. Lett. 2007; 17:4131–4134. [PubMed: 17543524]

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012; 40:D1100–D1107. [PubMed: 21948594]

Golbraikh A, Tropsha A. Beware of q2! J. Mol. Graph. Model. 2002; 20:269–276. [PubMed: 11858635]

Grover GJ, Egan DM, Sleph PG, Beehler BC, Chiellini G, Nguyen NH, Baxter JD, Scanlan TS. Effects of the thyroid hormone receptor agonist GC-1 on metabolic rate and cholesterol in rats and primates: selective actions relative to 3,5,3'-triiodo-L-thyronine. Endocrinology. 2004; 145:1656–1661. [PubMed: 14701670]

Hangeland JJ, Doweyko AM, Dejneka T, Friends TJ, Devasthale P, Mellstrom K, Sandberg J, Grynfarb M, Sack JS, Einspahr H, Farnegardh M, Husman B, Ljunggren J, Koehler K, Sheppard C, Malm J, Ryono DE. Thyroid receptor ligands. Part 2: Thyromimetics with improved selectivity for the thyroid hormone receptor beta. Bioorg. Med. Chem. Lett. 2004; 14:3549–3553. [PubMed: 15177471]

Hangeland JJ, Friends TJ, Doweyko AM, Mellström K, Sandberg J, Grynfarb M, Ryono DE. A new class of high affinity thyromimetics containing a phenyl-naphthylene core. Bioorg. Med. Chem. Lett. 2005; 15:4579–4584. [PubMed: 16099653]

Harvey CB, Williams GR. Mechanism of thyroid hormone action. Thyroid. 2002; 12:441–446. [PubMed: 12165104]

Hedfors Å, Appelqvist T, Carlsson B, Bladh L-G, Litten C, Agback P, Grynfarb M, Koehler KF, Malm J. Thyroid Receptor Ligands. 3. Design and Synthesis of 3,5-Dihalo-4-alkoxyphenylalkanoic Acids as Indirect Antagonists of the Thyroid Hormone Receptor. J. Med. Chem. 2005; 48:3114–3117. [PubMed: 15857115]

Hwang JY, Arnold La, Zhu F, Kosinski A, Mangano TJ, Setola V, Roth BL, Guy RK. Improvement of pharmacological properties of irreversible thyroid receptor coactivator binding inhibitors. J. Med. Chem. 2009; 52:3892–3901. [PubMed: 19469546]

Hwang JY, Attia RR, Carrillo AK, Connelly MC, Guy RK. Synthesis and evaluation of methylsulfonylnitrobenzamides (MSNBAs) as inhibitors of the thyroid hormone receptor-coactivator interaction. Bioorg. Med. Chem. Lett. 2013; 23:1891–1895. [PubMed: 23414840]

Hwang JY, Attia RR, Zhu F, Yang L, Lemoff A, Jeffries C, Connelly MC, Guy RK. Synthesis and evaluation of sulfonylnitrophenylthiazoles (SNPTs) as thyroid hormone receptor-coactivator interaction inhibitors. J. Med. Chem. 2012; 55:2301–2310. [PubMed: 22324546]

Hwang JY, Huang W, Arnold LA, Huang R, Attia RR, Connelly M, Wichterman J, Zhu F, Augustinaite I, Austin CP, Inglese J, Johnson RL, Guy RK. Methylsulfonylnitrobenzoates, a new class of irreversible inhibitors of the interaction of the thyroid hormone receptor and its obligate coactivators that functionally antagonizes thyroid hormone. J. Biol. Chem. 2011; 286:11895–11908. [PubMed: 21321127]

Izumo S, Mahdavi V. Thyroid hormone receptor alpha isoforms generated by alternative splicing differentially activate myosin HC gene transcription. Nature. 1988; 334:539–542. [PubMed: 2841611]

Koehler K, Gordon S, Brandt P, Carlsson B, Backsbro-Saeidi A, Apelqvist T, Agback P, Grover GJ, Nelson W, Grynfarb M, Farnegardh M, Rehnmark S, Malm J. Thyroid receptor ligands. 6. A high affinity "direct antagonist" selective for the thyroid hormone receptor. J. Med. Chem. 2006; 49:6635–6637. [PubMed: 17154490]

Kumar R, Thompson EB. The structure of the nuclear hormone receptors. Steroids. 1999; 64:310–319. [PubMed: 10406480]

Kuz'min VE, Polishchuk PG, Artemenko AG, Andronati SA. Interpretation of QSAR Models Based on Random Forest Methods. Mol. Inf. 2011; 30:593–603.

Li Y-L, Litten C, Koehler KF, Mellström K, Garg N, Garcia Collazo AM, Färnegård M, Grynfarb M, Husman B, Sandberg J, Malm J. Thyroid receptor ligands. Part 4: 4'-amido bioisosteric ligands selective for the thyroid hormone receptor beta. Bioorg. Med. Chem. Lett. 2006; 16:884–886. [PubMed: 16303304]

Liu H, Gramatica P. QSAR study of selective ligands for the thyroid hormone receptor beta. Bioorg. Med. Chem. 2007; 15:5251–5261. [PubMed: 17524652]

Lo Piparo, E.; Worth, A. Tools for predicting developmental and reproductive toxicity. Luxembourg: Institute for health and consumer protection, JointResearch Centre, European Commission; 2010. (http://ihcp.jrc.ec.europa.eu/our_labs/predictive_toxicology/doc/EUR_24522_EN.pdf)

Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, Tropsha A. Integrative chemical-biological read-across approach for chemical hazard classification. Chem. Res. Toxicol. 2013; 26:1199–1208. [PubMed: 23848138]

Malm J, Gordon S, Brandt P, Carlsson B, Agback P, Backbro Saeidi A, Sandberg J. Thyroid receptor ligands. Part 7: Indirect antagonists of the thyroid hormone receptor with improved affinity. Bioorg. Med. Chem. Lett. 2007; 17:2018–2021. [PubMed: 17254783]

Mouchlis VD, Mavromoustakos TM, Kokotos G. Molecular Docking and 3D–QSAR CoMFA Studies on Indole Inhibitors of GIIA Secreted Phospholipase A(2). J. Chem. Inf. Model. 2010; 50:1589–1601. [PubMed: 20795712]

Mysinger MM, Carchia M, Irwin JJ, Shoichet BK. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. J. Med. Chem. 2012; 55:6582–6594. [PubMed: 22716043]

Ribeiro RC, Apriletti JW, Wagner RL, West BL, Feng W, Huber R, Kushner PJ, Nilsson S, Scanlan T, Fletterick RJ, Schaufele F, Baxter JD. Mechanisms of thyroid hormone action: insights from X-ray crystallographic and functional studies. Recent Prog. Horm. Res. 1998; 53:351–392. discussion 392–354. [PubMed: 9769715]

Rogers JA, Metz L, Yong VW. Review: Endocrine disrupting chemicals and immune responses: a focus on bisphenol-A and its potential mechanisms. Mol. Immunol. 2013; 53:421–430. [PubMed: 23123408]

Rubin BS, Soto AM. Bisphenol A: Perinatal exposure and body weight. Mol. Cell. Endocrinol. 2009; 304:55–62. [PubMed: 19433248]

Rucker C, Rucker G, Meringer M. y-Randomization and its variants in QSPR/QSAR. J. Chem. Inf. Model. 2007; 47:2345–2357. [PubMed: 17880194]

Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. J. Med. Chem. 2003; 46:3045–3059. [PubMed: 12825943]

Schulz-Gasch T, Stahl M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. J. Mol. Model. 2003; 9:47–57. [PubMed: 12638011]

Sedykh A, Zhu H, Tang H, Zhang LY, Richard A, Rusyn I, Tropsha A. Use of in Vitro HTS-Derived Concentration-Response Data as Biological Descriptors Improves the Accuracy of QSAR Models of in Vivo Toxicity. Environ. Health Perspect. 2011; 119:364–370. [PubMed: 20980217]

Shoemaker TJ, Kono T, Mariash CN, Evans-Molina C. Thyroid Hormone Analogues for the Treatment of Metabolic Disorders: New Potential for Unmet Clinical Needs? Endocr. Pract. 2012; 18:954–964. [PubMed: 22784847]

Takeda K, Sakurai A, DeGroot LJ, Refetoff S. Recessive inheritance of thyroid hormone resistance caused by complete deletion of the protein-coding region of the thyroid hormone receptor-beta gene. J. Clin. Endocrinol. Metab. 1992; 74:49–55. [PubMed: 1727829]

Tice RR, Austin CP, Kavlock RJ, Bucher JR. Improving the human hazard characterization of chemicals: a Tox21 update. Environ. Health Perspect. 2013; 121:756–765. [PubMed: 23603828]

Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol. Inf. 2010; 29:476–488.

Tsakovska, I.; Pajeva, I.; Alov, P.; Worth, A. Chapter 6 - Recent Advances in the Molecular Modeling of Estrogen Receptor-Mediated Toxicity. In: Christo, C., editor. Advances in Protein Chemistry and Structural Biology. Academic Press; 2011. p. 217-251.

Valadares NF, Castilho MS, Polikarpov I, Garratt RC. 2D QSAR studies on thyroid hormone receptor ligands. Bioorg. Med. Chem. 2007; 15:4609–4617. [PubMed: 17467994]

Vedani A, Zumstein M, Lill Ma, Ernst B. Simulating alpha/beta selectivity at the human thyroid hormone receptor: consensus scoring using multidimensional QSAR. ChemMedChem. 2007; 2:78–87. [PubMed: 17096449]

Williams GR. Cloning and characterization of two novel thyroid hormone receptor beta isoforms. Mol. Cell. Biol. 2000; 20:8329–8342. [PubMed: 11046130]

Ye L, Li Y-L, Mellström K, Mellin C, Bladh L-G, Koehler K, Garg N, Garcia Collazo AM, Litten C, Husman B, Persson K, Ljunggren J, Grover G, Sleph PG, George R, Malm J. Thyroid receptor ligands. 1. Agonist ligands selective for the thyroid receptor beta1. J. Med. Chem. 2003; 46:1580–1588. [PubMed: 12699376]

Zhang L, Sedykh A, Tripathi A, Zhu H, Afantitis A, Mouchlis VD, Melagraki G, Rusyn I, Tropsha A. Identification of putative estrogen receptor-mediated endocrine disrupting chemicals using QSAR- and structure-based virtual screening approaches. Toxicol. Appl. Pharmacol. 2013; 272:67–76. [PubMed: 23707773]

Highlights

- This is the largest curated dataset for ligand binding domain (LBD) of the THRβ
- We report the first QSAR model for antagonists of AF-2 domain of THRβ.
- A combination of QSAR and docking enables prediction of both affinity and efficacy.
- Models can be used to identify environmental chemicals interacting with THRβ.
- Models can be used to eliminate the THRβ-mediated mechanism of action

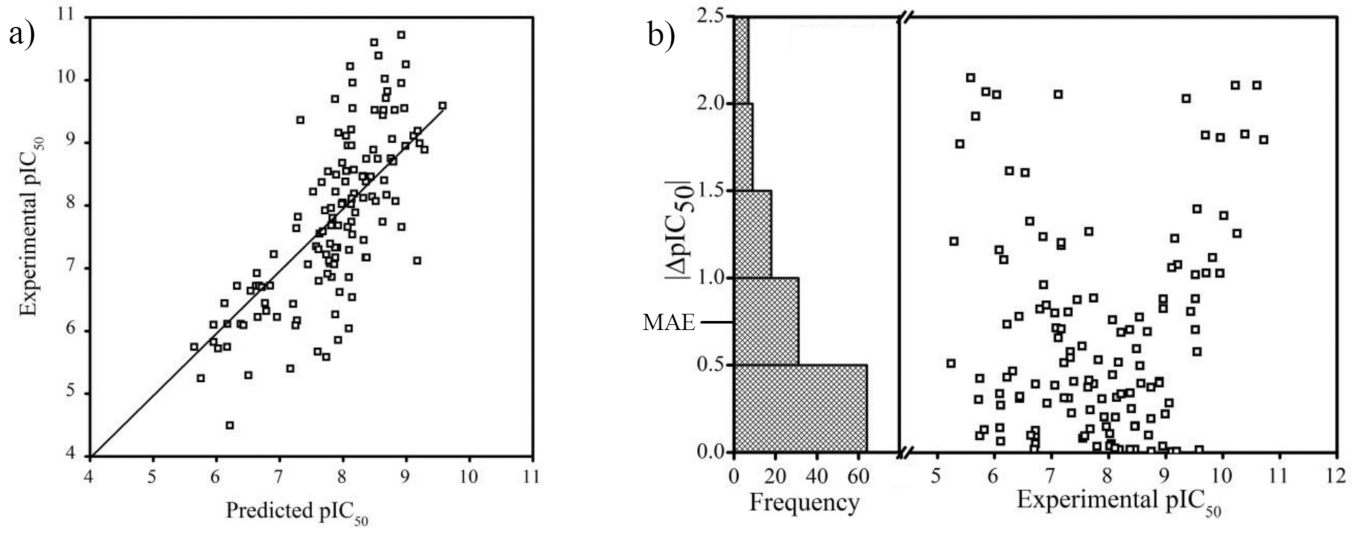


Figure 1.

Results of continuous QSAR modeling of the LBD dataset (see Table 2): a) experimental *vs.* predicted pIC_{50} (R^2=0.55); b) absolute prediction errors, ΔpIC_{50}, calculated as absolute difference between experimental and predicted pIC_{50} *vs.* experimental pIC_{50}. The left panel in Figure 1b shows frequency distribution of ΔpIC_{50}; the ΔpIC_{50} value corresponding to the mean absolute error (MAE=0.68) is identified by a horizontal line.

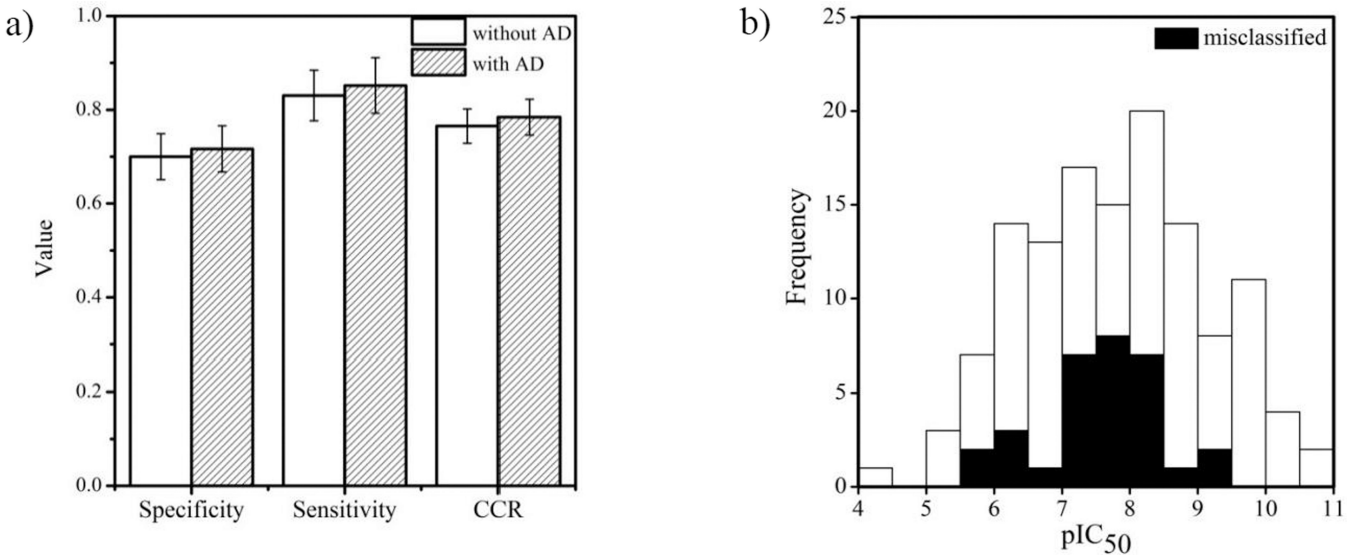


Figure 2.

Results of binary QSAR modeling (two categories of compounds are defined using threshold of pIC_{50} = 8) for the LBD dataset (see Table 2): a) External prediction accuracy estimated from 5-fold external cross validation with and without filtering by the applicability domain); b) Frequency distribution of all compounds across activity bins (white bars); frequency of misclassified compounds in each activity bin is also shown (black bars).

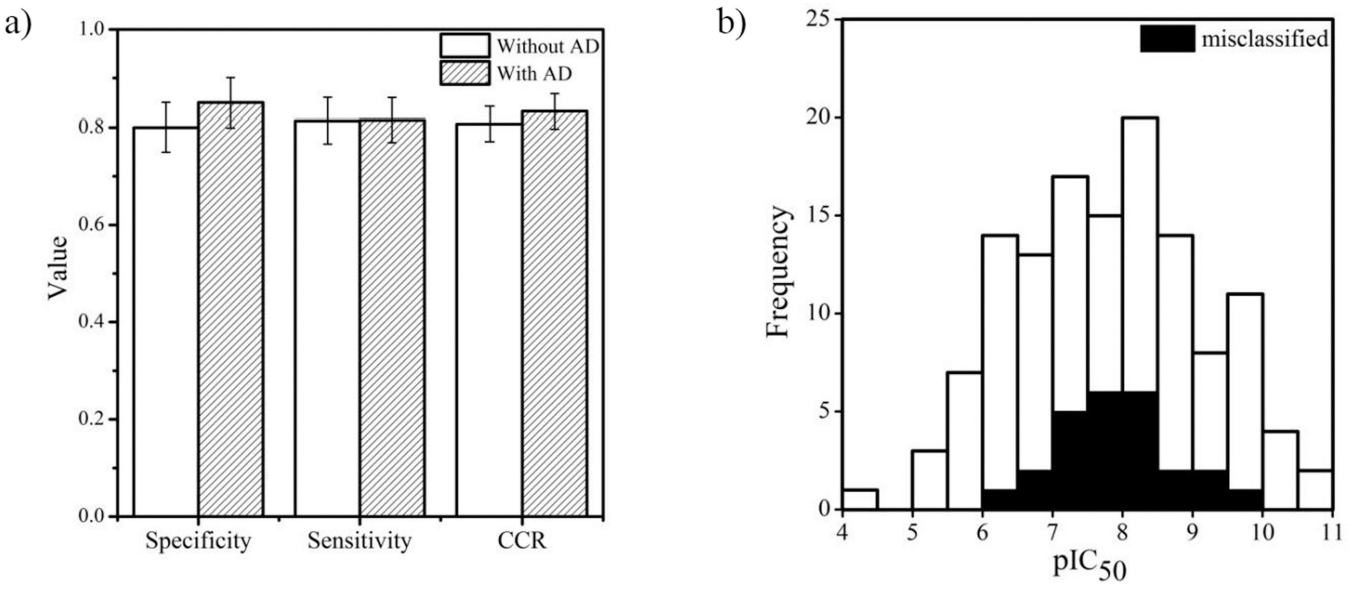


Figure 3.

Accuracy of binary classification of compounds when compounds are assigned into two classes based on pIC_{50} values (using cutoff of pIC_{50}=8) predicted by continuous QSAR model for the LBD dataset. a) External prediction accuracy estimated from 5-fold external cross validation with and without filtering by the applicability domain. b) Frequency distribution of all compounds across activity bins (white bars); frequency of misclassified compounds in each activity bin is also shown (black bars).

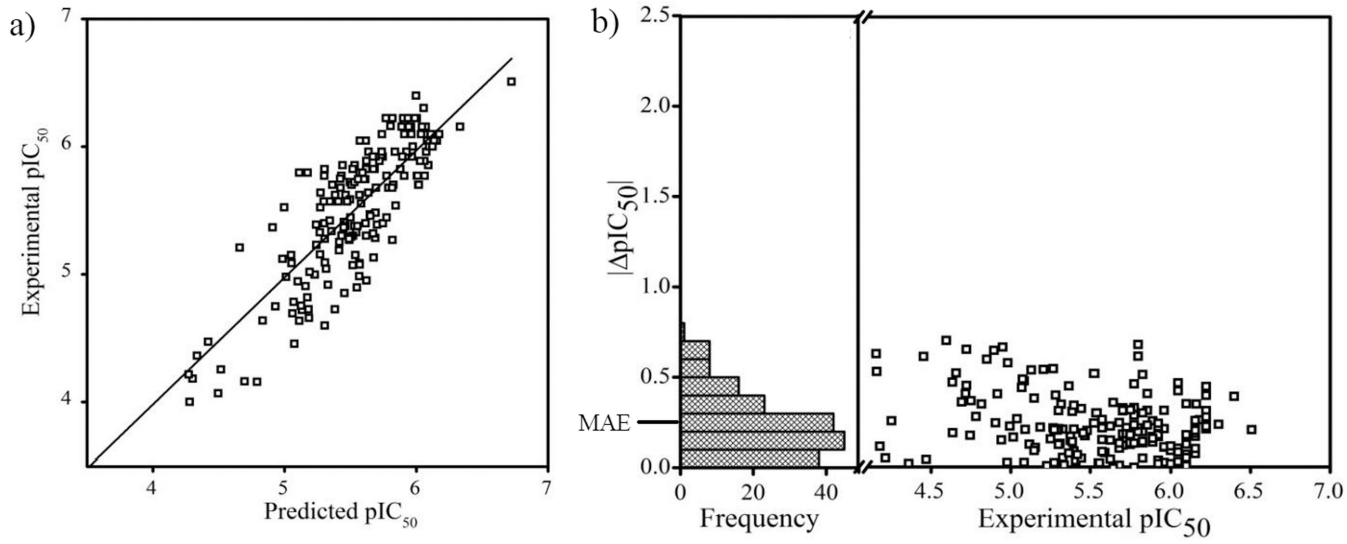


Figure 4.

Results of continuous QSAR modeling of the AF-2 dataset (see Table 2): a) experimental *vs.* predicted pIC_{50} (R^2=0.70); b) Absolute prediction errors, ΔpIC_{50}, calculated as absolute difference between experimental and predicted pIC_{50} *vs.* experimental pIC_{50}. The left panel in Figure 4b shows frequency distribution of ΔpIC_{50}; the ΔpIC_{50} value corresponding to the mean absolute error (MAE=0.24) is identified by a horizontal line.

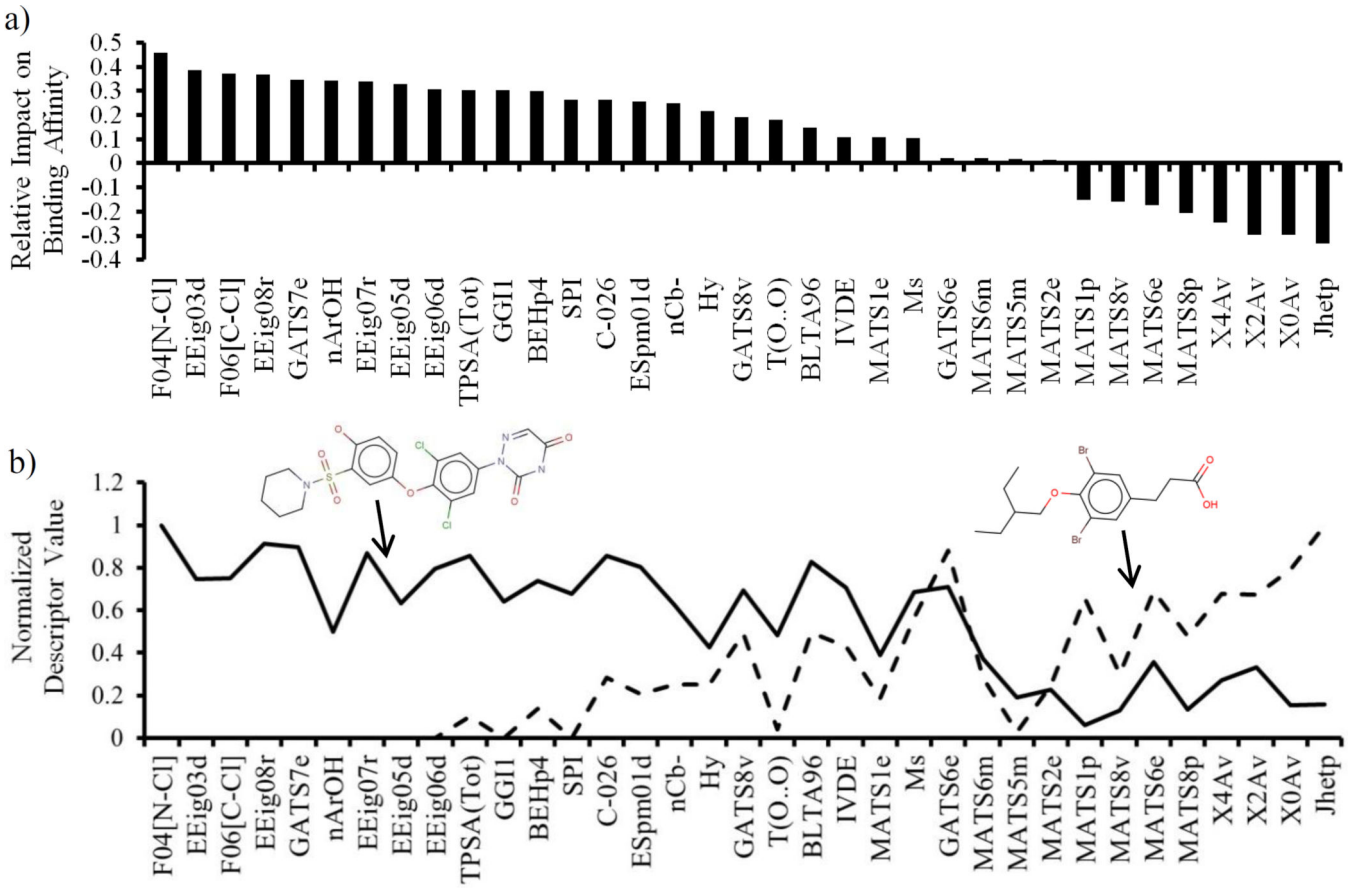


Figure 5.

The impact of 35 most significant chemical descriptors for the LBD dataset on ligand binding affinity based on the continuous model. a) Each bar shows the difference between mean descriptor values for strong (binding affinity equal or below 1 nM) *vs.* weak (binding affinity equal or above 1μM) binders. b) Comparison of the descriptor profiles for a strong (solid line) *vs.* a weak (dotted line) THRβ_1 binders.

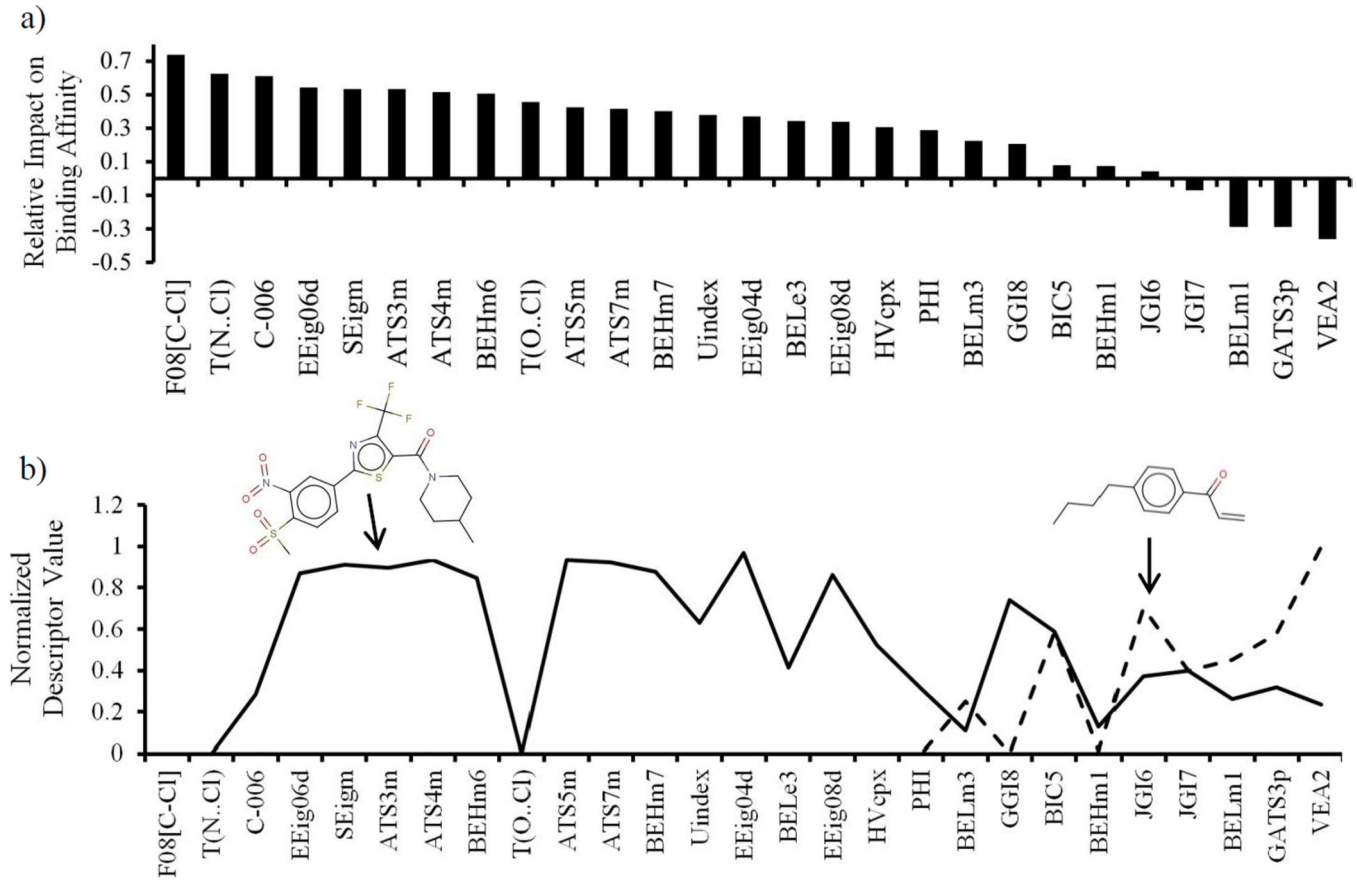


Figure 6.

The impact of 27 most significant chemical descriptors for the AF-2 dataset on ligand binding affinity based on continuous model. a) Each bar shows the difference between mean descriptor values for strong (binding affinity below 1μM) *vs.* weak (binding affinity above 10μM) binders. b) Comparison of the descriptor profiles for strong *vs.* weak AF-2 binders.

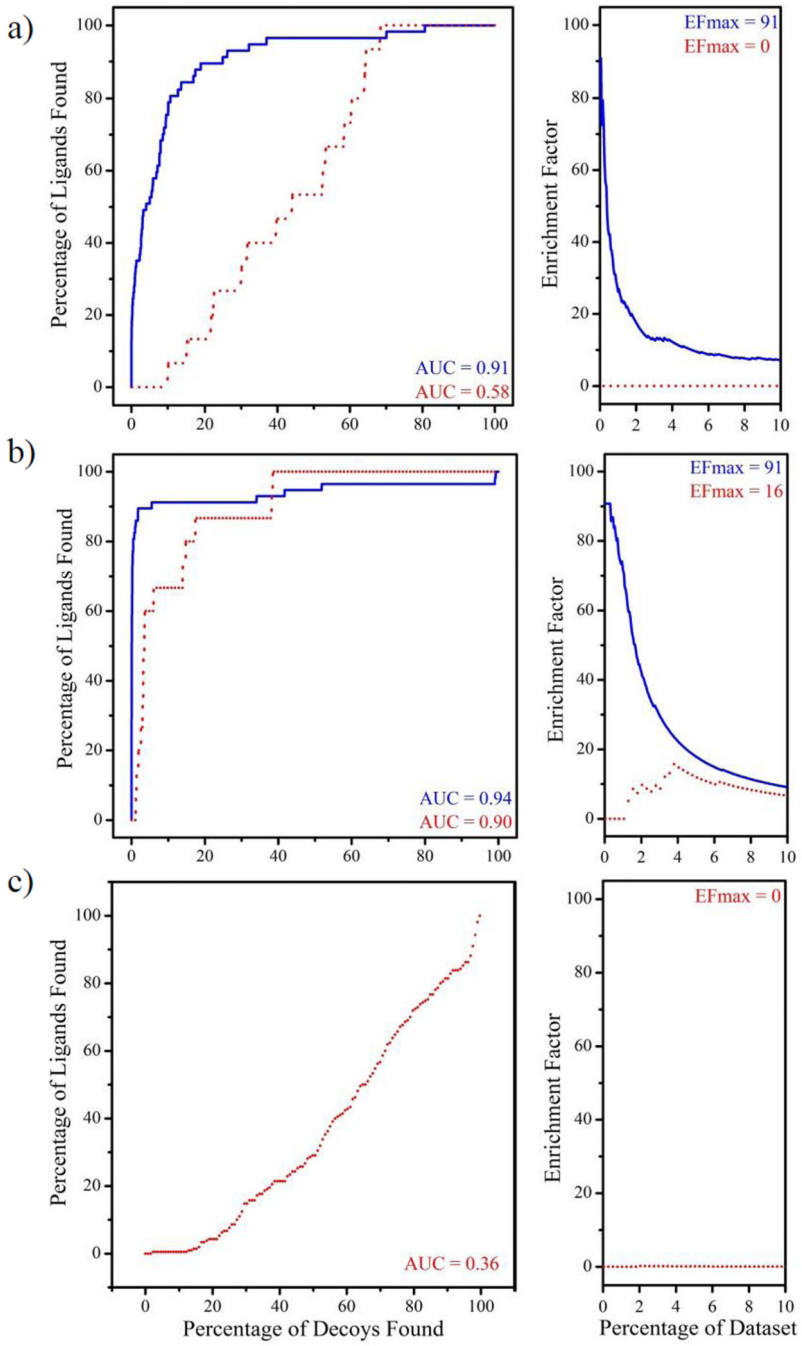


Figure 7.

ROC and enrichment factor curves obtained as a result of docking studies. a) Docking was done using a THRβ structure with an agonist bound to the LBD domain (PDB code 1Q4X). b) Docking was done using another THRβ structure with an agonist bound to the LBD domain (PDB code 1N46). c) Docking was done using the THRβ structure with an antagonist bound to the AF-2 domain (PDB code 2PIN). Blue solid lines in all plots indicate THRβ agonists and red dotted lines indicate THRβ antagonists. The numbers of known

THRβ ligands and presumed decoys/non-binders are shown in Table 2. The chemical structures are included in Supplemental Tables 2 and 3.

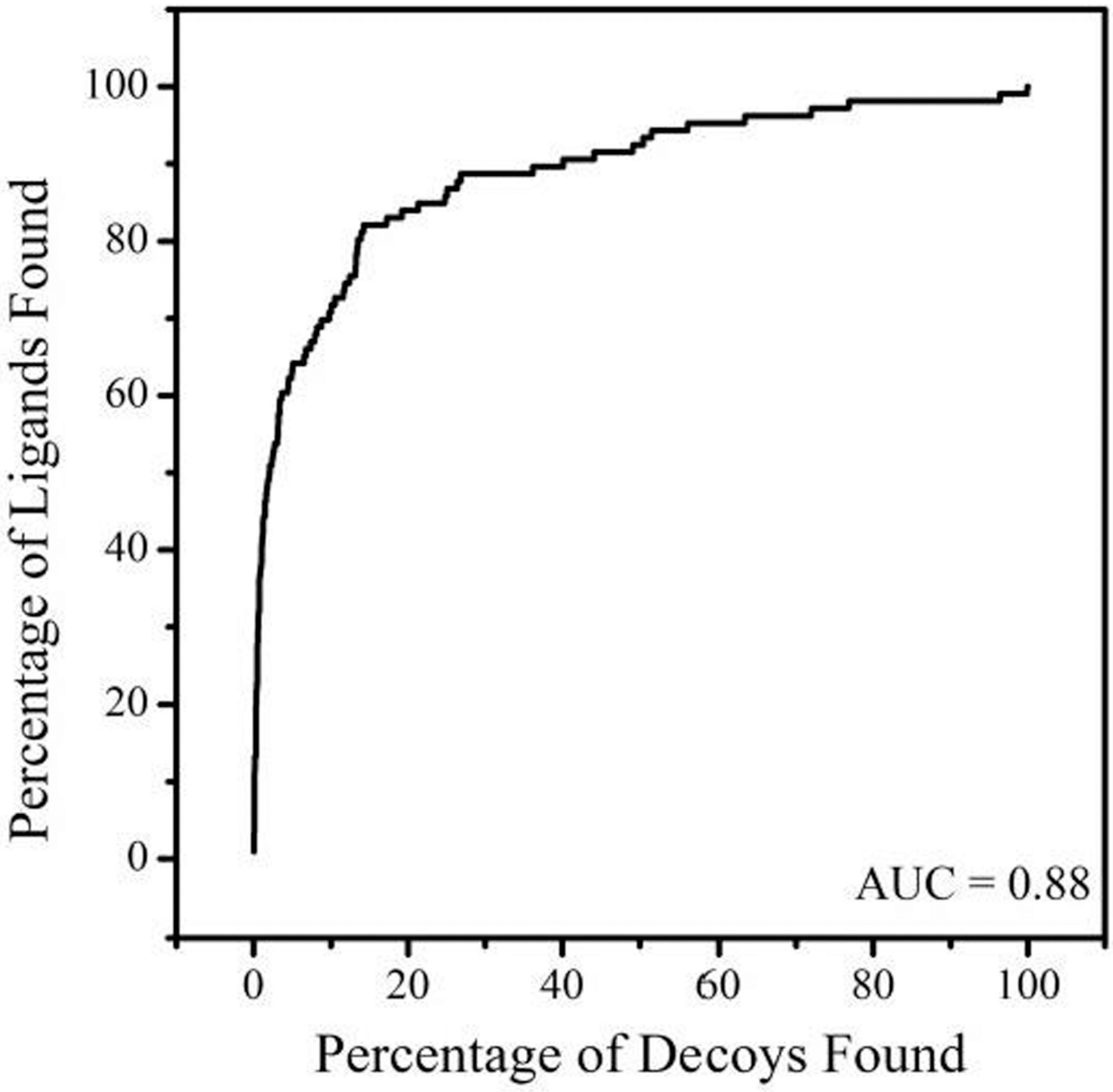


Figure 8.

ROC curve for docking of the 101 LBD ligands with unknown functional annotation (see Methods) using the THRβ structure with an agonist bound to the LBD domain (PDB code 1Q4X) that could discriminate agonists *vs.* antagonists (cf. Figure 7a).

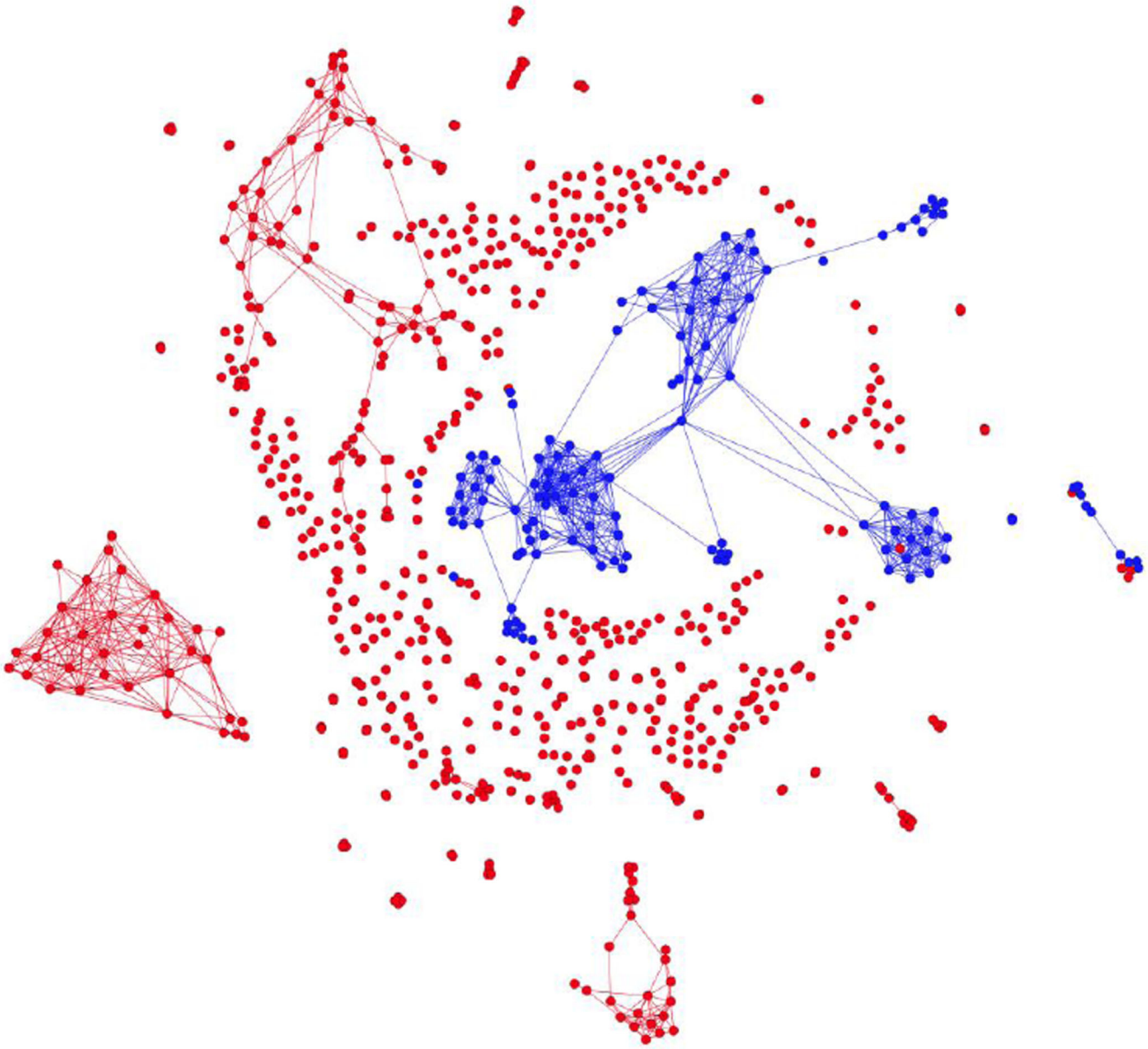


Figure 9.

Network analysis of a modeling dataset and compounds found to be active in an assay used by Tox21. Each node represents a chemical of either modeling or Tox 21 datasets color coded blue or red, respectively. An edge connects two chemicals with high similarity (Tanimoto coefficient of 0.8 or higher).

Table 1

Previous QSAR studies of THRβ_1

Reference/Website	Modeling method description	Number of compounds in dataset[a]	Reported prediction accuracy for test sets (R^2)
Liu and Gramatica., 2007	MLR with variable selection. (0–3)D Dragon descriptors	85(21)	0.73
VirtualToxLab Vedani et al., 2007	Multi-dimensional QSAR (Quasar and Raptor software)	82(18)	0.796
Valadares et al., 2007	Classical QSAR with 2D Dragon descriptors and Hologram QSAR (specialized fragment fingerprints)	68(13)	0.84
Du et al., 2008	3D QSAR (CoMFA and CoMSIA)	61(12)	0.68
Ren et al., 2007	Projection Pursuit Regression (PPR) with variable selection. CODESSA descriptors	80(13)	0.893

[a] number of compounds in test set is given in parenthesis

Table 2

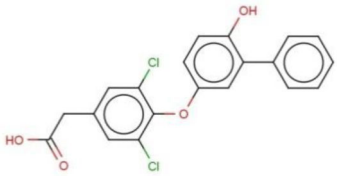
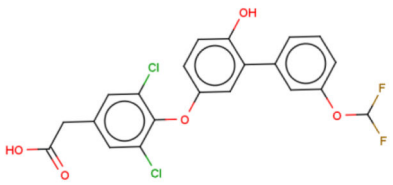
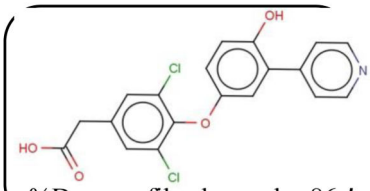
Modeling approaches and datasets of THRβ ligands employed in this study[a]

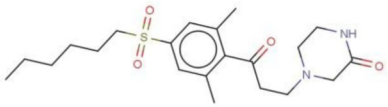
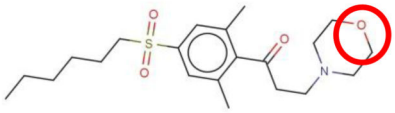
Modeling method	Data type	Number of compounds
QSAR (LBD)	pIC_{50}	129
QSAR (AF-2)	pIC_{50}	181
Docking (LBD)	Agonists	57
	Antagonists	15
	Presumed decoys	5101
Docking (AF-2)	Antagonists	210
	Presumed decoys	12249

[a]See Methods for data sources and for the additional information about these datasets

Table 3

Effect of structural modifications on binding affinity (based on the analysis of significant descriptors in QSAR models)

Domain	Parent Chemical	Modified Chemical*	Structural Features	
			High impact descriptor	Descriptor interpretation
LBD	 Experimental pIC_{50} = 8.54 Predicted pIC_{50} = 7.76 EEig03d = 0.44 nCb- = 0.75 BLTA96 = 0.28		EEig03d (positively correlated with the binding affinity)	Increase in molecular polarity
		 %Dscr profile changed = 86	BLTA96 (positively correlated with the binding affinity)	Substitution of N for aromatic carbon to increase MlogP
		Experimental pIC_{50} = 6.9 Predicted pIC_{50} = 7.75 EEig03d = 0.44 nCb- = 0.62 BLTA96 = 0.54 %Dscr profile changed = 63		

AF-2	 Experimental pIC_{50} = 5.26 Predicted pIC_{50} = 5.82 C-006=0.71 EEig06d = 0.83	 Experimental pIC_{50} = 6.04 Predicted pIC_{50} = 5.62 C-006=0.85 EEig06d = 0.64	C-006 (positively correlated with the binding affinity) EEig06d (positively correlated with binding affinity)	Add CH2RX functional group linked through carbon (R-any group,X-heteroatom) Increase in molecular polarity

*Red circle identifies a fragment in the modified compound that differs from the respective fragment in the parent compound.