

## Research Article

# LASSO-ing Potential Nuclear Receptor Agonists and Antagonists: A New Computational Method for Database Screening

Sean Ekins,<sup>1</sup> Michael-R. Goldsmith,<sup>2</sup> Aniko Simon,<sup>3</sup> Zsolt Zsoldos,<sup>3</sup>  
Orr Ravitz,<sup>3</sup> and Antony J. Williams<sup>4</sup>

<sup>1</sup> Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay-Varina, NC 27526, USA

<sup>2</sup> National Exposure Research Laboratory, US-Environmental Protection Agency, 109 TW Alexander Drive, Research Triangle Park, NC 27711, USA

<sup>3</sup> SimBioSys, Inc., 135 Queen's Plate Drive, Suite 520, Toronto, ON, Canada M9W 6V1

<sup>4</sup> Royal Society of Chemistry, 904 Tamaras Circle, Wake Forest, NC 27587, USA

Correspondence should be addressed to Sean Ekins; [ekinssean@yahoo.com](mailto:ekinssean@yahoo.com) and Michael-R. Goldsmith; [goldsmith.rocky@epa.gov](mailto:goldsmith.rocky@epa.gov)

Received 1 February 2013; Accepted 28 May 2013

Academic Editor: Gabriela Mustata Wilson

Copyright © 2013 Sean Ekins et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nuclear receptors (NRs) are important biological macromolecular transcription factors that are implicated in multiple biological pathways and may interact with other xenobiotics that are endocrine disruptors present in the environment. Examples of important NRs include the androgen receptor (AR), estrogen receptors (ER), and the pregnane X receptor (PXR). In this study we have utilized the Ligand Activity by Surface Similarity Order (LASSO) method, a ligand-based virtual screening strategy to derive structural (surface/shape) molecular features used to generate predictive models of biomolecular activity for AR, ER, and PXR. For PXR, twenty-five models were built using between 8 to 128 agonists and tested using 3000, 8000, and 24,000 drug-like decoys including PXR inactive compounds ( $N = 228$ ). Preliminary studies with AR and ER using LASSO suggested the utility of this approach with 2-fold enrichment factors at 20%. We found that models with 64–128 PXR actives provided enrichment factors of 10-fold (10% actives in the top 1% of compounds screened). The LASSO models for AR and ER have been deployed and are freely available online, and they represent a ligand-based prediction method for putative NR activity of compounds in this database.

## 1. Introduction

The nuclear receptor (NRs) family of transcription factors are important targets for therapeutic interventions for multiple diseases [1] and also may interact with other xenobiotics that are endocrine disruptors present in the environment [2]. It is therefore important to identify compounds that may specifically bind NRs and act as endocrine disruptors and develop synthetic compounds that can selectively (in a cell-type and/or tissue-selective manner) modulate NR pharmacology (reviewed in [3–9]). NRs including the androgen receptor

(AR; NR3C4), estrogen receptors  $\alpha$  and  $\beta$ , (ER $\alpha$  and ER $\beta$ ; NR3A1 and NR3A2) and pregnane X receptor (PXR; NR1I2) are particularly important as both therapeutic targets and for xenobiotics to mediate off-target effects.

The ERs are activated by 17 $\beta$ -estradiol while the AR is activated by testosterone and dihydrotestosterone and these receptors are transcriptional regulators of many genes [10] with important physiological functions [11–16]. The human PXR [17–19] similarly transcriptionally regulates genes involved in xenobiotic metabolism and excretion, as well as other cellular processes, including apoptosis [20–24]. Human

PXR is a broad specificity NR, binding a wide variety of molecules [25] and the activation of this NR can cause drug-drug interactions [23].

Multiple QSAR and machine learning models have been described for these NRs, to address endocrine disruptor risk assessment [26–28] and toxicological screening [29]. For example, a recent QSAR analysis of 74 natural or synthetic estrogens provided information on structural features for the activation of ER $\alpha$  and ER $\beta$  [30]. Nonlinear statistical machine learning methods have been applied to separate NR activators from nonactivators [31]. A virtual screening protocol identified ER $\beta$  specific ligands from a plant product-based database [32]; from 12 candidates evaluated by a fluorescence polarization binding assay, 3 had >100-fold selectivity to ER $\beta$  over ER $\alpha$ . The same approach has also been used to find compounds with good selectivity for ER $\alpha$  over ER $\beta$  [32, 33]. Bisson et al. have used computational methods that led to a nonsteroidal antiandrogen with improved AR antagonistic activity based on an initial screening of FDA approved new drugs [33]. Several groups have published datasets or performed modeling on ER and AR and these data are readily available for further evaluation with new modeling methods [34–38].

While the crystal structures of human PXR [39–43] have led to a greater understanding of the ligand binding domain (LBD) and ligand-receptor interactions [39–45], ligand-based computational models possess the key features for predicting binding [46–49]. PXR pharmacophores have been used to predict interactions for antibiotics [50] verified *in vitro*, and machine learning methods have also been evaluated [25, 38, 51–53]. Several protein-based docking studies have also been used to predict PXR agonists [25, 54–56], although machine learning methods appear to be advantageous to date.

We have recently described troubleshooting various computational methods [57] and specifically compared different methods for PXR [25]. There is a continuous search for new methods that might offer advantages for computational modeling to overcome some of these limitations and specifically for NRs [58]. A ligand-based software called LASSO (Ligand Activity by Surface Similarity Order) has been described that is focused on similarities in biomolecular activity rather than structural similarity [59]. The key components describing LASSO are the 23 kinds of Interacting Surface Point Type (ISPT) molecular descriptors (see Supplemental Table 1 available online at <http://dx.doi.org/10.1155/2013/513537>), which capture the essence of the surface point information in a feature vector containing the counts of each surface point type and create the feature vector for that ligand. This vector serves as the descriptor of that molecule with the assumption that ligands with similar feature vectors will have similar activity. A key property of the LASSO descriptor is its conformation independence which is due to the fact that it is defined by the number and type of Interacting Surface Points and not by their relative spacial distribution. LASSO has been shown to be able to readily screen over 1 million structures/minute, identify active molecules by enriching screened databases, and provide a means for scaffold hopping [59]. The current study applies LASSO to various NR datasets to generate

models, validate them, and make the models available on a public website to illustrate how the method can be used. This work can be considered an extension of our previous troubleshooting studies [25, 57].

## 2. Materials and Methods

*2.1. Training and Test Set Molecular Dataset Selection.* One of the goals of this study was to determine what level of enrichment for binders (at weak or strong binding threshold) can be afforded using the ChemSpider LASSO descriptors (ligand-based approach) and compared with enrichment from a structure-based docking approach (eHiTS). For AR, the dataset consisting of 203 molecules with relative binding affinities and activity threshold classes of (a) strong, (b) moderate, (c) weak, and (d) inactive/nonbinding ligands, we evaluated the ability of LASSO to differentiate both (a) strong and (a + b) strong and moderate binders (all others were considered to be nonbinding). The training set for AR, derived with the LASSO descriptors, was obtained from the DUD set [60] and differ considerably from the test set. To evaluate the LASSO descriptors for the ER dataset consisting of 50 molecules with 15 “hits” (i.e., considerably weak binders) and 35 “nonhits” for the estrogen binding that differs considerably from the training set obtained again, we used the DUD ER (default or agonist and antagonist) as a training set [60].

In addition to the ChemSpider LASSO approach for the AR test set we used the eHiTS structure-based (molecular docking) screening strategy on two conformations of the AR (using PDB structures 2AMA and 1XNN) and reported the minimum score across the two conformations examined (this approach was used to add flexibility to the receptor). Similarly, for the ER dataset we docked against two functionally distinct conformations of the estrogen receptor (3ERT and 1GWR).

*2.2. Datasets for LASSO Modeling: Structure File Preparation.* The rat ER binding dataset ( $K_i$  values for 50 compounds of environmental relevance [35]) was obtained from EPA's DSSTox database (<http://epa.gov/ncct/dsstox/> [34]). This dataset contains 15 industrial chemical “binders” (i.e., non-therapeutic) with significantly weaker binding affinities than what would be desired for drug lead candidates (i.e., 3–5-fold weaker binding affinity than the natural ligand 17 $\beta$ -estradiol). Similarly, the NCTR's rat AR activity dataset (competitive inhibition assays), also used in this study, contains 146 AR binders and 56 nonbinders (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/edkb/index.htm> [37, 38]). All structures were imported into MOE and geometry optimized using the MMFFx forcefield in MOE (Chemical Computing Group, Montreal, Canada).

Three human PXR datasets were used, namely, dataset 1 represented 80 actives  $EC_{50} < 100 \mu\text{M}$  and 64 inactives  $EC_{50} > 100 \mu\text{M}$  that were drug-like molecules. The SMILES string for each molecule named or CAS number provided was obtained by downloading from either PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) or ChemSpider (<http://www.chemspider.com/>) or sketched using the BUILDER module

of SYBYL [56]. Dataset 2 represented 93 actives and 75 inactives that were drug-like molecules from a previous study [61]. The molecular structures encoded as SMILES strings [62] were downloaded from the supplementary information tables in the original publication [61]. Dataset 3 represented 30 actives and 89 inactives from a dataset of steroidal compounds (namely, androstanes, estratrienes, pregnanes and bile salts) as well as the ligands used in the crystal structures with hPXR activation determined by a luciferase-based reporter assay [25]. Human PXR activation was determined by a luciferase-based reporter assay as has been previously described in these and other publications.

**2.3. LASSO Models for ER and AR.** The methodology of LASSO has already been previously described in detail, [59] and the method performance in terms of diversity of test set and % enrichment of a database has also already been evaluated for the DUD set in paper just mentioned and also for other targets published elsewhere ([http://www.simbiosys.com/ehits\\_lasso/ehits\\_lasso\\_table.html](http://www.simbiosys.com/ehits_lasso/ehits_lasso_table.html)) to examine the performance of eHiTS LASSO, with this endocrine panel subset of target proteins of the total ~48 nuclear receptors. We used the newly assembled directory of useful decoys (DUD) [60] dataset to augment both the KIERBL and NCTR AR datasets.

**2.4. PXR Models: Method I.** The previously mentioned three PXR datasets were received from three different sources described earlier. Set 1, called: "hpxr\_test," contained 80 actives and 64 inactives or decoys; set 2, called: "hpxr\_train," contained 93 actives and 75 inactives; and finally set 3, called: "PXR119-class," contained 30 actives with 89 inactives. Out of these three data sets, 7 screening prediction models were built using only the actives (the inactives were automatically generated by the software).

The following models were developed. Model 1 was trained on the first dataset (hpxr\_test, 80 ligands) and tested with the other two sets (123 ligands). Model 2 was trained on the second dataset (hpxr\_train, 93 ligands) and tested with the other two sets (110 ligands). Model 3 was trained on the third dataset (PXR119-class, 30 ligands) and tested with the other two sets (173 ligands). Models 4–6 were trained on sets 1 and 2, that is, 173 ligands, sets 1 and 3, that is, 110 ligands, and sets 2 and 3, that is, 123 ligands and tested with the remaining one set of actives (i.e., 30, 93, and 80 ligands, resp.). Model 7 was trained on all actives (1, 2, and 3) and tested on the same. This was done as an extreme case to see the maximum potential training effect.

**2.5. PXR Models: Method II.** A second method for creating LASSO prediction models for the PXR test case was also investigated. Actives from the 3 datasets were all merged, resulting in an SDF file with 203 ligands with relative binding affinities and activity threshold classes (a) strong, (b) moderate, (c) weak, and (d) inactive/nonbinding ligands. To determine how many actives are needed to be selected for a good LASSO prediction model and also to see if the source of the actives is important, 25 LASSO models were developed

(Supplementary Table 2). Prediction models were built by selecting 8, 16, 32, 64, and 128 actives, starting from positions first, ninth, seventeenth, thirty-third, and sixty-fourth in the merged actives file.

The above 25 models were then tested for enrichment factor using the actives from the total active set and leaving out the ones used for training (this was 8, 16, 32, 64 or 128 ligands, resp.) mixed with drug-like decoys, that were obtained from another recent screening study [63]. To assess the effect of the size of the decoy set upon the prediction model, random 3000 (3 k), 8000 (8 k), and the whole 24,000 (24 k) decoy sets were used. In each case the decoys from all three sets received (228 structures in total) were added into the decoy test set.

### 3. Results

The enrichment plots shown for AR (Figure 1) and ER (Figure 2) with the percent actives recovered versus percent of dataset reveal an enrichment of ~2-fold at 20% of the dataset coverage regardless of whether a ligand or structure-based approach was used. For the AR dataset if the interaction threshold is specified as strong or strong + moderate, different levels of enrichment are incurred by either ChemSpider LASSO or eHiTS results. This translates into an improved performance of either ligand or structure-based screening approaches to bin molecules with stronger interactions (cyan and purple) than those substantially altered through the addition of weaker binding classes (magenta and yellow). Interestingly, in terms of the early-recognition problem, eHiTS is more sensitive (4-fold at 20%) than LASSO (1.5-fold at 20%); however, all 15 actives are captured by the LASSO descriptor within the first 37% of the dataset (with a minimum value of LASSO = 0.07) at considerably lower computational cost. A means of incorporating this into a real scenario would be to screen ChemSpider for AR with a descriptor above a threshold (in this case 0.07) from a specific dataset on ChemSpider and follow up these "hits" only with a more costly structure-based approach.

For the ER dataset where all 15 binders are in fact weak binders (i.e., 3 to 5 orders of magnitude weaker binders than the natural ligand  $17\beta$ -estradiol) the default LASSO descriptors outperform (3-fold enrichment) the structure-based approach (2-fold enrichment) and the agonist-trained LASSO method outperforms the antagonist LASSO method (most likely due to a large diversity among antagonists than agonists). Here we can see that even for weakly interacting partners (i.e., low affinity binders for ER) we can still obtain enrichment that is substantially better than random.

These tandem virtual screening approaches combine computationally efficient ligand-based ChemSpider LASSO descriptors (since ChemSpider is at its core a rich and diverse collection of chemical structures, these were used in order to produce LASSO predictions for over 14 million compounds against a series of 40 targets including AR and ER. A LASSO search feature was added to ChemSpider to allow users to search the database by LASSO value (see Figure 3(a)). Scientists can readily search for the top 1000 compounds (or

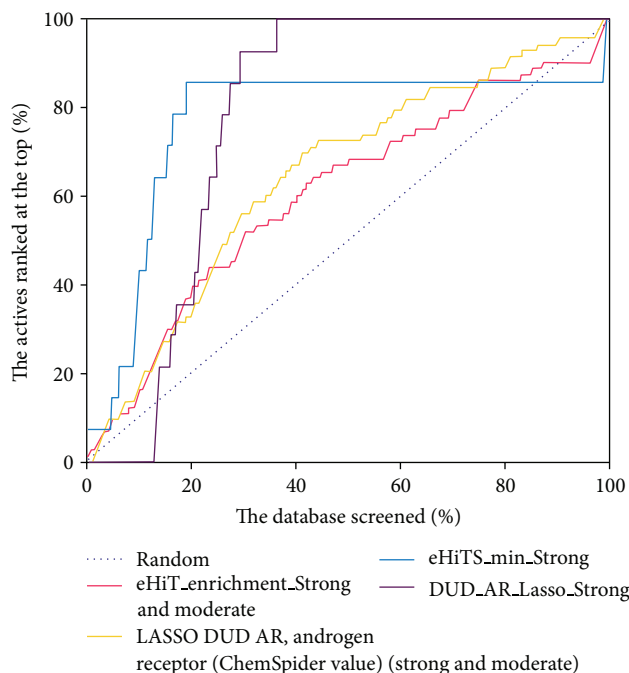


FIGURE 1: Enrichment plot for AR. Legend: blue dotted line: random hit rate, pink line: eHiTS docking score-based enrichment for strong and moderate binders (i.e., strong  $\text{Log}_{IC_{50}}$  mean =  $-7.7$ , stdev =  $0.7$ , moderate  $\text{Log}_{IC_{50}}$  mean =  $-5.1$ , stdev =  $0.6$ , probe DHT) from two crystal structures (1XNN and 2AMA), yellow line: LASSO DUD AR androgen receptor trained ligands (ChemSpider value with strong and moderate binding ligands as “hit” criteria), light blue: eHiTS minimum energy docking score from two crystal structures using strong hit criteria only (crystal structures 1XNN and 2AMA), purple: DUD LASSO using strong binders hit criteria only.

less) with the highest LASSO value for a particular target of interest. An advanced search in ChemSpider can combine LASSO value searches with other parameters such as molecular weight, rule of 5 values, and specific data sources (e.g., selecting molecules from commercial data sources only) prior to more costly structure-based virtual screening strategies, dramatically improving virtual screening and “early-recognition problem” workflow efficiency.

Piggy-backing more costly structure-based virtual screening strategies on top of an initial screen dramatically assists in virtual screening endeavors and the early-recognition problem.

We have also shown an example of a molecule, mibolerone, a strong AR and ER binder based on LASSO (Figure 3(b)) which is known as a potent AR binder [64]. The LASSO surface point type values are shown in Supplemental Table 3 and more visually in Supplemental Figure 2.

When we used LASSO models with hPXR in method I (Figure 4) we found the best results with Model 1 which suggested 40% of the ligands can be pushed into the top 10% of the screened database resulting with an enrichment factor 4-fold better than random (Figure 5). In Method II we found the same enrichment factor using 64 actives in a 24 thousand compound decoy set (Figure 6). Another way to evaluate

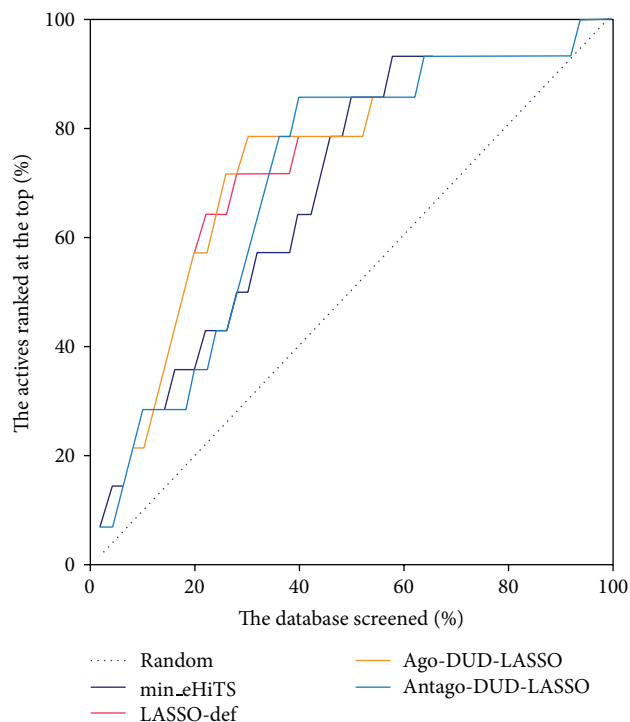
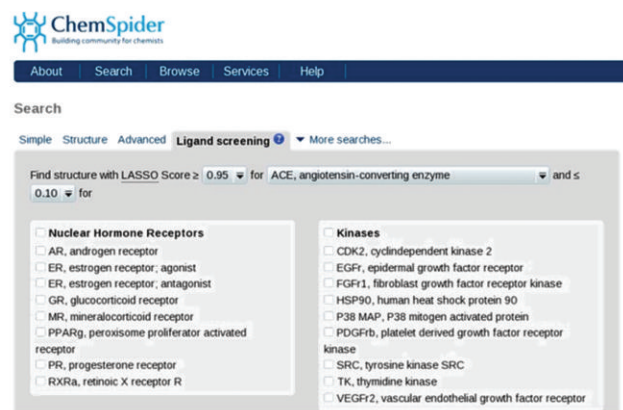


FIGURE 2: Enrichment plot for ER. Legend: purple dotted line: random hit rate, blue line: the minimum eHiTS docking score reported for two different crystal structures (Min\_eHiTS, crystal structure = 3ERT and 1GWR), pink: using the default LASSO descriptors (LASSO-def), yellow: using the DUD AR agonist-trained descriptors as found on ChemSpider (agonist DUD LASSO), light blue: using the DUD AR antagonist trained descriptors as found on ChemSpider (antagonist DUD LASSO).

the models is to present the statistics for using dataset 1 to predict dataset 2 ( $N = 168$ ) for which we obtained sensitivity 12%, specificity 99%, accuracy 51%, and Matthews correlation 0.2. Using dataset 2 to predict dataset 1 gives similar results. These results suggest that the models could identify potential human PXR agonists in databases similar to other target proteins [59].

## 4. Discussion

For both the AR and ER ligands the main objective was to see how ligand-based screening tools, such as LASSO’s ChemSpider implementation, perform such that they could be used for prioritizing chemicals for testing. The AR dataset contained a mixture of drug-like and environmental receptor modulators, whereas the ER dataset contained primarily environmental chemicals. Even in light of the relatively weak binding affinity of the “actives,” that is,  $K_i$  of  $10^{-4}$ – $10^{-6}$ , while these would be poor candidates for lead optimization into drugs, they still pose an interaction potential with biological systems such as NRs if they bioaccumulate. Using these leads from LASSO screening with other methods such as molecular docking or free-energy perturbation simulations may also be useful. The validation of the approaches outlined above was pursued by



(a)

SimBioSys LASSO

Descriptors: 0.0,0.0,0.0,0.0,0.2,0.0,1.2,28.0,0.1,0.0,1.1,2.0,0.0

Category	Target	PDB Code	LASSO Score
Nuclear Hormone Receptors	MR, mineralocorticoid receptor	2aa2	0.88
Nuclear Hormone Receptors	AR, androgen receptor	1xq2	0.42
Nuclear Hormone Receptors	ER, estrogen receptor, agonist	1i2i	0.41
Nuclear Hormone Receptors	PR, progesterone receptor	1sr7	0.28
Nuclear Hormone Receptors	GR, glucocorticoid receptor	1m2z	0.10
Metalloenzymes	ACE, angiotensin-converting enzyme	1o86	0.03
Nuclear Hormone Receptors	RXRa, retinoic X receptor R	1mvc	0.03
Nuclear Hormone Receptors	PPARg, peroxisome proliferator activated receptor	1fm9	0.02
Kinases	SRC, tyrosine kinase SRC	2src	0.01
Other Enzymes	HMGR, hydroxymethylglutaryl-CoA reductase	1hw8	0.01

(b)

FIGURE 3: (a) LASSO models implemented in ChemSpider. (b) LASSO scores for Mibolerone, a strong AR and ER binder found with a LASSO search on ChemSpider. The LASSO score is significant only for 4 targets: MR, AR, ER and PR while the rest of the 40 targets are all 0.10 or below for this molecule.

examining two real datasets. These were the FDA's NCTR AR [38] and 50 environmental molecules evaluated for ER binding affinity [35]. In addition we have used multiple sets of PXR agonists described previously. Our results show enrichments of between 2-fold and 4-fold depending on the NR. For PXR there have been numerous recent studies using different machine learning methods and descriptors [25, 54, 56], and while the Matthews correlation coefficient in this study is lower than those in previous studies, the level of enrichment from between 4 fold (40% of actives in the top 10% of compounds screened) and 10-fold (10% actives in the top 1% of compounds screened) was very encouraging.

The molecular descriptors used in eHiTS LASSO are independent of ligand conformation and have been shown to successfully enrich screened databases across a wide range of target families [59]. Lying somewhere between a 2D and a 3D descriptor the ISPT descriptor does not contain any shape or 2D connectivity information. There may however be some molecular size information implicit in the descriptor due to capturing the counts of surface points and larger molecules will have more surface points than smaller molecules (and eHiTS LASSO may be somewhat sensitive to this).

The relatively high speed of eHiTS LASSO on a single CPU [59] makes it an ideal tool to be used as a predocking screen. From a troubleshooting perspective, eHiTS LASSO will return a high percentage of false positives, due to not

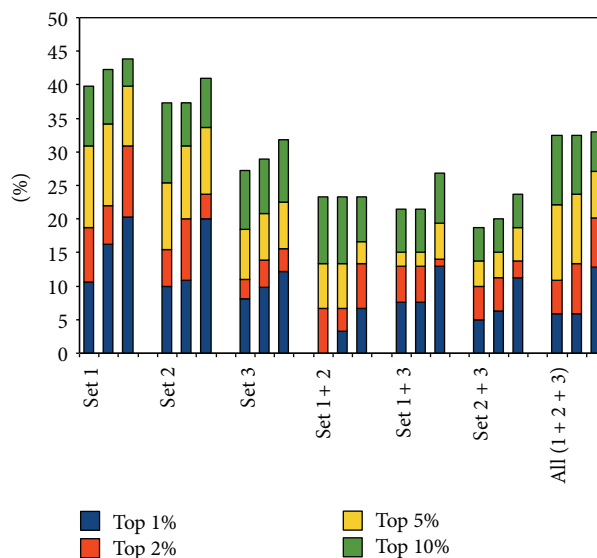


FIGURE 4: PXR LASSO 7 models derived in training method I.

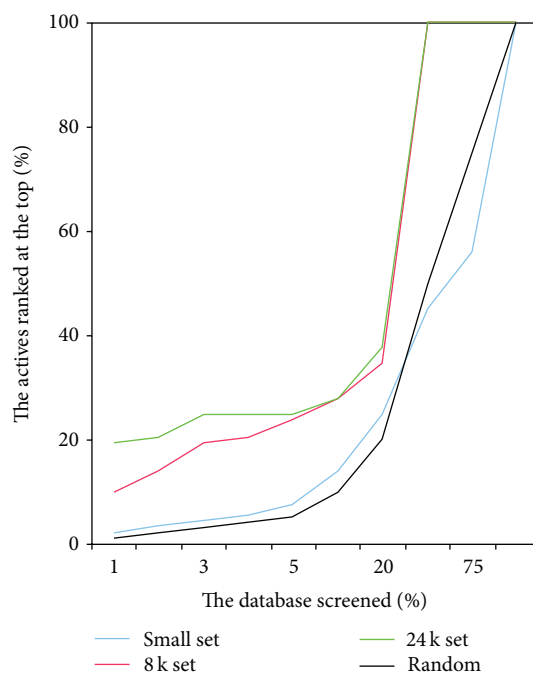


FIGURE 5: PXR LASSO enrichment for model from set 1.

considering 3D relationships of surface properties. Because of this, it will also return a higher percentage of different scaffolds, enabling scaffold hopping. It is also important to note that LASSO would not be able to differentiate stereoisomerism apart from, perhaps, diastereomeric pairs which have structurally (configurationally) different features rather than conformationally different features, for which this method is conformation invariant.

Taking the results of eHiTS LASSO and feeding the top  $N\%$  into a docking program would allow the docking program to weed out many of the false positives binders. For this

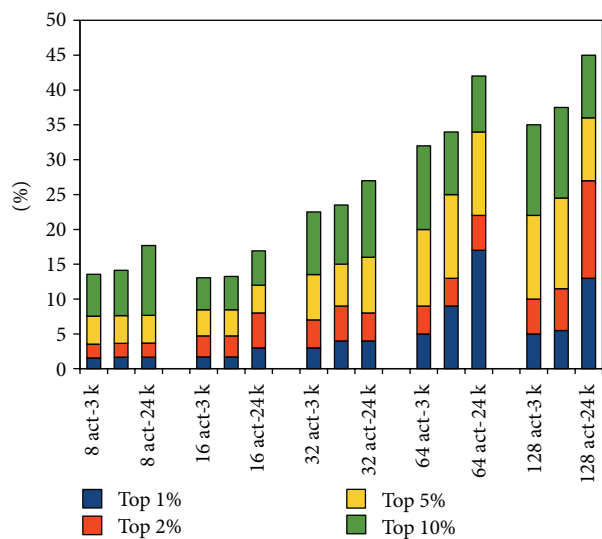


FIGURE 6: PXR LASSO average of 5 models over 5 runs derived in training method II.

reason, eHiTS LASSO is currently integrated with the commercially available eHiTS docking tool and can be readily used as a predocking screening tool for large virtual screens.

From the current study we have shown significant enrichments when testing computational models for AR, ER, and PXR. While AR and ER predictions are currently already implemented in ChemSpider, it is clear that adding predicted values for PXR and other NRs as they become available would be beneficial to the community in terms of accessing an open source of chemical structures with pregenerated descriptors. It should be noted however that the generation of model data for a database as large as that hosted by ChemSpider (now well over 25 million compounds) is not a small undertaking and consumes a significant amount of compute time, data preparation, and handling in order to deliver the models to the community for consumption.

The use of such ligand-based computational methods as exemplified by LASSO in this study could also be useful for the design and selection of chemical products that are less hazardous to human health and the environment. This may make them useful in green chemistry [65] ([http://www.epa.gov/gcc/pubs/about\\_gc.html](http://www.epa.gov/gcc/pubs/about_gc.html)) as well as in biomedical research. The ready accessibility of such NR binding predictions from computational models like LASSO will be key in future for both pharmaceutical and environmental applications, and databases like ChemSpider can have an important role in providing them to the public as a predocking criteria, as we have demonstrated in this study. This study used published rat ER, AR and human PXR data. LASSO could also be applied to build models for the same NRs across multiple species, such that they could be used to estimate interspecies variation in ligand binding.

## Abbreviations

AR: Androgen receptor  
DUD: Directory of useful decoys

ER: Estrogen receptor  
ISPT: Interacting Surface Point Type  
LASSO: Ligand Activity by Surface Similarity Order  
LBD: Ligand binding domain  
PXR: Pregnane X receptor  
QSAR: Quantitative Structure Activity Relationship  
SNNs: Stuttgart Neural Network Simulator.

## Supporting Information

The supplemental files contain (I) the 23 Surface Point Types used in LASSO with related descriptions, (II) the model building details for PXR (III) the LASSO 6.1 surface point types for Mibolerone, (IV) a visualization of the generalized surface-point types from LASSO for a histidine-like fragment as visualized in CheVi and (V) Mibolerone displayed in SimBioSys' CheVi 3D desktop visualization tool, showing the 3D structure, color-coded interaction surface of the molecule, and the surface point representation.

## Disclaimer

This document has been subjected to review by the US Environmental Protection Agency and approved for publication.

## Conflict of Interests

A. J. Williams is employed by the Royal Society of Chemistry which owns ChemSpider and associated technologies. S. Ekins and M. R. Goldsmith were on the advisory board for ChemSpider from June 2007 until May 2011. A. Simon, Z. Zsoldos, and O. Ravitz are employed by SimBioSys Inc. which owns LASSO and eHiTS.

## References

- [1] A. G. Smith and G. E. O. Muscat, "Skeletal muscle and nuclear hormone receptors: implications for cardiovascular and metabolic disease," *International Journal of Biochemistry and Cell Biology*, vol. 37, no. 10, pp. 2047–2063, 2005.
- [2] A. K. Hotchkiss, C. V. Rider, C. R. Blystone et al., "Fifteen years after "wingspread"—environmental endocrine disrupters and human and wildlife health: where we are today and where we need to go," *Toxicological Sciences*, vol. 105, no. 2, pp. 235–259, 2008.
- [3] S. Tenbaum and A. Baniahmad, "Nuclear receptors: structure, function and involvement in disease," *International Journal of Biochemistry and Cell Biology*, vol. 29, no. 12, pp. 1325–1341, 1997.
- [4] A. Zimmer and C. Gespach, "Bile acids and derivatives, their nuclear receptors FXR, PXR and ligands: role in health and disease and their therapeutic potential," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 8, no. 5, pp. 540–563, 2008.
- [5] P. L. Feldman, M. H. Lambert, and B. R. Henke, "PPAR modulators and PPAR pan agonists for metabolic diseases: the next generation of drugs targeting peroxisome proliferator-activated receptors?" *Current Topics in Medicinal Chemistry*, vol. 8, no. 9, pp. 728–749, 2008.

- [6] X. Ma, J. R. Idle, and F. J. Gonzalez, "The pregnane X receptor: from bench to bedside," *Expert Opinion on Drug Metabolism and Toxicology*, vol. 4, no. 7, pp. 895–908, 2008.
- [7] M. Bertolotti, C. Gabbi, C. Anzivino, L. Carulli, P. Loria, and N. Carulli, "Nuclear receptors as potential molecular targets in cholesterol accumulation conditions: insights from evidence on hepatic cholesterol degradation and gallstone disease in humans," *Current Medicinal Chemistry*, vol. 15, no. 22, pp. 2271–2284, 2008.
- [8] E. E. Baulieu, M. Atger, and M. Best Belpomme, "Steroid hormone receptors," *Vitamins and Hormones*, vol. 33, pp. 649–736, 1975.
- [9] D. V. Henley and K. S. Korach, "Endocrine-disrupting chemicals use distinct mechanisms of action to modulate endocrine system function," *Endocrinology*, vol. 147, no. 6, pp. S25–S32, 2006.
- [10] M. D. Krasowski, E. J. Reschly, and S. Ekins, "Intrinsic disorder in nuclear hormone receptors," *Journal of Proteome Research*, vol. 7, no. 10, pp. 4359–4372, 2008.
- [11] A. C. W. Pike, A. M. Brzozowski, R. E. Hubbard et al., "Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist," *EMBO Journal*, vol. 18, no. 17, pp. 4608–4618, 1999.
- [12] A. M. Brzozowski, A. C. W. Pike, Z. Dauter et al., "Molecular basis of agonism and antagonism in the oestrogen receptor," *Nature*, vol. 389, no. 6652, pp. 753–758, 1997.
- [13] J. W. R. Schwabe, L. Chapman, J. T. Finch, D. Rhodes, and D. Neuhaus, "DNA recognition by the oestrogen receptor: from solution to the crystal," *Structure*, vol. 1, no. 3, pp. 187–204, 1993.
- [14] J. S. Sack, K. F. Kish, C. Wang et al., "Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 9, pp. 4904–4909, 2001.
- [15] C. E. Bohl, C. Chang, M. L. Mohler et al., "A ligand-based approach to identify quantitative structure-activity relationships for the androgen receptor," *Journal of Medicinal Chemistry*, vol. 47, no. 15, pp. 3765–3776, 2004.
- [16] P. L. Shaffer, A. Jivan, D. E. Dollins, F. Claessens, and D. T. Gewirth, "Structural basis of androgen receptor binding to selective androgen response elements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 14, pp. 4758–4763, 2004.
- [17] G. Bertilsson, J. Heidrich, K. Svensson et al., "Identification of a human nuclear receptor defines a new signaling pathway for CYP3A induction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 21, pp. 12208–12213, 1998.
- [18] B. Blumberg, W. Sabbagh Jr., H. Juguilon et al., "SXR, a novel steroid and xenobiotic-sensing nuclear receptor," *Genes and Development*, vol. 12, no. 20, pp. 3195–3205, 1998.
- [19] S. A. Kliewer, J. T. Moore, L. Wade et al., "An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway," *Cell*, vol. 92, no. 1, pp. 73–82, 1998.
- [20] S. Verma, M. M. Tabb, and B. Blumberg, "Activation of the steroid and xenobiotic receptor, SXR, induces apoptosis in breast cancer cells," *BMC Cancer*, vol. 9, article 3, 2009.
- [21] D. Gupta, M. Venkatesh, H. Wang et al., "Expanding the roles for pregnane X receptor in cancer: proliferation and drug resistance in ovarian cancer," *Clinical Cancer Research*, vol. 14, no. 17, pp. 5332–5340, 2008.
- [22] J. Zhou, M. Liu, Y. Zhai, and W. Xie, "The antiapoptotic role of pregnane X receptor in human colon cancer cells," *Molecular Endocrinology*, vol. 22, no. 4, pp. 868–880, 2008.
- [23] A. Biswas, S. Mani, M. R. Redinbo et al., "Elucidating the 'Jekyll and Hyde' nature of PXR: the case for discovering antagonists," *Pharmaceutical Research*, vol. 26, no. 8, pp. 1807–1815, 2009.
- [24] B. L. Urquhart, R. G. Tirona, and R. B. Kim, "Nuclear receptors and the regulation of drug-metabolizing enzymes and drug transporters: implications for interindividual variability in response to drugs," *Journal of Clinical Pharmacology*, vol. 47, no. 5, pp. 566–578, 2007.
- [25] S. Ekins, S. Kortagere, M. Iyer et al., "Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR," *PLoS Computational Biology*, vol. 5, no. 12, Article ID e1000594, 2009.
- [26] N. Ai, R. K. DeLisle, S. J. Yu, and W. J. Welsh, "Computational models for predicting the binding affinities of ligands for the wild-type androgen receptor and a mutated variant associated with human prostate cancer," *Chemical Research in Toxicology*, vol. 16, no. 12, pp. 1652–1660, 2003.
- [27] M. N. Jacobs, "In silico tools to aid risk assessment of endocrine disrupting chemicals," *Toxicology*, vol. 205, no. 1-2, pp. 43–53, 2004.
- [28] W. Tong, R. Perkins, L. Xing, W. J. Welsh, and D. M. Sheehan, "QSAR models for binding of estrogenic compounds to estrogen receptor  $\alpha$  and  $\beta$  subtypes," *Endocrinology*, vol. 138, no. 9, pp. 4022–4025, 1997.
- [29] S. Ekins, L. Mirny, and E. G. Schuetz, "A ligand-based approach to understanding selectivity of nuclear hormone receptors PXR, CAR, FXR, LXR $\alpha$ , and LXR $\beta$ ," *Pharmaceutical Research*, vol. 19, no. 12, pp. 1788–1800, 2002.
- [30] T. Z. Bao, G.-Z. Han, J.-Y. Shim, Y. Wen, and X.-R. Jiang, "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*, vol. 147, no. 9, pp. 4132–4150, 2006.
- [31] D. Plewczynski, M. Von Grotthuss, S. A. H. Spieser et al., "Target specific compound identification using a support vector machine," *Combinatorial Chemistry and High Throughput Screening*, vol. 10, no. 3, pp. 189–196, 2007.
- [32] L. Zhao and R. D. Brinton, "Structure-based virtual screening for plant-based EB $\beta$ -selective ligands as potential preventative therapy against age-related neurodegenerative diseases," *Journal of Medicinal Chemistry*, vol. 48, no. 10, pp. 3463–3466, 2005.
- [33] W. H. Bisson, A. V. Cheltsov, N. Bruey-Sedano et al., "Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 29, pp. 11927–11932, 2007.
- [34] S. Laws et al., DSSTox EPA Estrogen Receptor Ki Binding Study (Laws et al.) Database—(KIERBL): SDF file and documentation. 2009 [http://www.epa.gov/ncct/dsstox/sdf\\_kierbl.html](http://www.epa.gov/ncct/dsstox/sdf_kierbl.html).
- [35] S. C. Laws, S. Yavanxay, R. L. Cooper, and J. C. Eldridge, "Nature of the binding interaction for 50 structurally diverse chemicals with rat estrogen receptors," *Toxicological Sciences*, vol. 94, no. 1, pp. 46–56, 2006.
- [36] J. R. Rabinowitz, S. B. Little, S. C. Laws, and M.-R. Goldsmith, "Molecular modeling for screening environmental chemicals for estrogenicity: use of the toxicant-target approach," *Chemical Research in Toxicology*, vol. 22, no. 9, pp. 1594–1602, 2009.
- [37] H. Fang, W. Tong, W. S. Branham et al., "Study of 202 natural, synthetic, and environmental chemicals for binding to

- the androgen receptor," *Chemical Research in Toxicology*, vol. 16, no. 10, pp. 1338–1358, 2003.
- [38] D. Ding, L. Xu, H. Fang et al., "The EDKB: an established knowledge base for endocrine disrupting chemicals," *BMC Bioinformatics*, vol. 11, no. 6, article 5, 2010.
- [39] R. E. Watkins, P. R. Davis-Searles, M. H. Lambert, and M. R. Redinbo, "Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor," *Journal of Molecular Biology*, vol. 331, no. 4, pp. 815–828, 2003.
- [40] R. E. Watkins, J. M. Maglich, L. B. Moore et al., "2.1 Å crystal structure of human PXR in complex with the St. John's wort compound hyperforin," *Biochemistry*, vol. 42, no. 6, pp. 1430–1438, 2003.
- [41] Y. Xue, E. Chao, W. J. Zuercher, T. M. Willson, J. L. Collins, and M. R. Redinbo, "Crystal structure of the PXR-T1317 complex provides a scaffold to examine the potential for receptor antagonism," *Bioorganic and Medicinal Chemistry*, vol. 15, no. 5, pp. 2156–2166, 2007.
- [42] J. E. Chrencik, J. Orans, L. B. Moore et al., "Structural disorder in the complex of human pregnane X receptor and the macrolide antibiotic rifampicin," *Molecular Endocrinology*, vol. 19, no. 5, pp. 1125–1134, 2005.
- [43] D. G. Teotico, J. J. Bischof, L. Peng, S. A. Kliewer, and M. R. Redinbo, "Structural basis of human pregnane X receptor activation by the hops constituent colupulone," *Molecular Pharmacology*, vol. 74, no. 6, pp. 1512–1520, 2008.
- [44] R. E. Watkins, G. B. Wisely, L. B. Moore et al., "The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity," *Science*, vol. 292, no. 5525, pp. 2329–2333, 2001.
- [45] Y. Xue, L. B. Moore, J. Orans et al., "Crystal structure of the pregnane X receptor-estradiol complex provides insights into endobiotic recognition," *Molecular Endocrinology*, vol. 21, no. 5, pp. 1028–1038, 2007.
- [46] K. Bachmann, H. Patel, Z. Batayneh et al., "PXR and the regulation of apoA1 and HDL-cholesterol in rodents," *Pharmacological Research*, vol. 50, no. 3, pp. 237–246, 2004.
- [47] S. Ekins, C. Chang, S. Mani et al., "Human pregnane X receptor antagonists and agonists define molecular requirements for different binding sites," *Molecular Pharmacology*, vol. 72, no. 3, pp. 592–603, 2007.
- [48] S. Ekins and J. A. Erickson, "A pharmacophore for human pregnane X receptor ligands," *Drug Metabolism and Disposition*, vol. 30, no. 1, pp. 96–99, 2002.
- [49] D. Schuster and T. Langer, "The identification of ligand features essential for PXR activation by pharmacophore modeling," *Journal of Chemical Information and Modeling*, vol. 45, no. 2, pp. 431–439, 2005.
- [50] K. Yasuda, A. Ranade, R. Venkataraman et al., "A comprehensive in vitro and in silico analysis of antibiotics that activate pregnane X receptor and induce CYP3A4 in liver and intestine," *Drug Metabolism and Disposition*, vol. 36, no. 8, pp. 1689–1697, 2008.
- [51] A. Khandelwal, M. D. Krasowski, E. J. Reschly, M. W. Sinz, P. W. Swaan, and S. Ekins, "Machine learning methods and docking for predicting human pregnane X receptor activation," *Chemical Research in Toxicology*, vol. 21, no. 7, pp. 1457–1467, 2008.
- [52] S. Kortagere, D. Chekmarev, W. J. Welsh, and S. Ekins, "Hybrid scoring and classification approaches to predict human pregnane X receptor activators," *Pharmaceutical Research*, vol. 26, no. 4, pp. 1001–1011, 2009.
- [53] S. Ekins, E. J. Reschly, L. R. Hagey, and M. D. Krasowski, "Evolution of pharmacologic specificity in the pregnane X receptor," *BMC Evolutionary Biology*, vol. 8, no. 1, article 103, 2008.
- [54] S. Kortagere, D. Chekmarev, W. J. Welsh, and S. Ekins, "Hybrid scoring and classification approaches to predict human pregnane X receptor activators," *Pharmaceutical Research*, vol. 26, no. 4, pp. 1001–1011, 2009.
- [55] S. Kortagere, M. D. Krasowski, E. J. Reschly, M. Venkatesh, S. Mani, and S. Ekins, "Evaluation of computational docking to identify pregnane X receptor agonists in the toxcast database," *Environmental Health Perspectives*, vol. 118, no. 10, pp. 1412–1417, 2010.
- [56] A. Khandelwal, M. D. Krasowski, E. J. Reschly, M. W. Sinz, P. W. Swaan, and S. Ekins, "Machine learning methods and docking for predicting human pregnane X receptor activation," *Chemical Research in Toxicology*, vol. 21, no. 7, pp. 1457–1467, 2008.
- [57] S. Kortagere and S. Ekins, "Troubleshooting computational methods in drug discovery," *Journal of Pharmacological and Toxicological Methods*, vol. 61, no. 2, pp. 67–75, 2010.
- [58] N. Ai, M. D. Krasowski, W. J. Welsh, and S. Ekins, "Understanding nuclear receptors using computational methods," *Drug Discovery Today*, vol. 14, no. 9–10, pp. 486–494, 2009.
- [59] D. Reid, B. S. Sadjad, Z. Zsoldos, and A. Simon, "LASSO—ligand activity by surface similarity order: a new tool for ligand based virtual screening," *Journal of Computer-Aided Molecular Design*, vol. 22, no. 6–7, pp. 479–487, 2008.
- [60] N. Huang, B. K. Shoichet, and J. J. Irwin, "Benchmarking sets for molecular docking," *Journal of Medicinal Chemistry*, vol. 49, no. 23, pp. 6789–6801, 2006.
- [61] C. Y. Ung, H. Li, C. W. Yap, and Y. Z. Chen, "In silico prediction of pregnane X receptor activators by machine learning approaches," *Molecular Pharmacology*, vol. 71, no. 1, pp. 158–168, 2007.
- [62] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *Journal of Chemical Information and Computer Sciences*, vol. 28, pp. 31–36, 1988.
- [63] G. B. McGaughey, R. P. Sheridan, C. I. Bayly et al., "Comparison of topological, shape, and docking methods in virtual screening," *Journal of Chemical Information and Modeling*, vol. 47, no. 4, pp. 1504–1519, 2007.
- [64] L. R. Murthy, M. P. Johnson, and D. R. Rowley, "Characterization of steroid receptors in human prostate using mibolerone," *Prostate*, vol. 8, no. 3, pp. 241–253, 1986.
- [65] A. M. Voutchkova, T. G. Osimitz, and P. T. Anastas, "Toward a comprehensive molecular design framework for reduced hazard," *Chemical Reviews*, vol. 110, no. 10, pp. 5845–5882, 2010.





**Hindawi**  
Submit your manuscripts at  
<http://www.hindawi.com>

