Chemical Research in Toxicology

Article

pubs.acs.org/crt

Predicting Rat and Human Pregnane X Receptor Activators Using Bayesian Classification Models

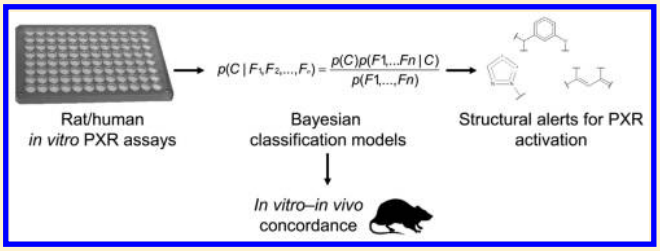
Mohamed Diwan M. AbdulHameed,*,† Danielle L. Ippolito,‡ and Anders Wallqvist*,†

†Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, 504 Scott Street, Fort Detrick, Maryland 21702, United States

‡U.S. Army Center for Environmental Health Research, 568 Doughten Drive, Fort Detrick, Maryland 21702, United States

Supporting Information

ABSTRACT: The pregnane X receptor (PXR) is a ligand-activated transcription factor that acts as a master regulator of metabolizing enzymes and transporters. To avoid adverse drug−drug interactions and diseases such as steatosis and cancers associated with PXR activation, identifying drugs and chemicals that activate PXR is of crucial importance. In this work, we developed ligand-based predictive computational models for both rat and human PXR activation, which allowed us to identify potentially harmful chemicals and evaluate species-specific effects of a given compound. We utilized a large publicly available data set of nearly 2000 compounds screened in cell-based reporter gene assays to develop Bayesian quantitative structure−activity relationship models using physicochemical properties and structural descriptors. Our analysis showed that PXR activators tend to be hydrophobic and significantly different from nonactivators in terms of their physicochemical properties such as molecular weight, logP, number of rings, and solubility. Our Bayesian models, evaluated by using 5-fold cross-validation, displayed a sensitivity of 75% (76%), specificity of 76% (75%), and accuracy of 89% (89%) for human (rat) PXR activation. We identified structural features shared by rat and human PXR activators as well as those unique to each species. We compared rat *in vitro* PXR activation data to *in vivo* data by using DrugMatrix, a large toxicogenomics database with gene expression data obtained from rats after exposure to diverse chemicals. Although *in vivo* gene expression data pointed to cross-talk between nuclear receptor activators that is captured only by *in vivo* assays, overall we found broad agreement between *in vitro* and *in vivo* PXR activation. Thus, the models developed here serve primarily as efficient initial high-throughput *in silico* screens of *in vitro* activity.



1. INTRODUCTION

The pregnane X receptor (PXR) [gene symbol: NR1I2] is a member of the nuclear receptor family of ligand-activated transcription factors that includes other receptors such as the vitamin D and thyroid hormone receptors.[1] Unlike other nuclear receptors that interact selectively with ligands containing specific structural features, PXR is a promiscuous protein that acts as a sensor for a wide array of xenobiotics and endogenous chemicals.[1,2] PXR activators include diverse classes of chemicals such as bile acids, steroid hormones, fat-soluble vitamins, prescription drugs, herbal formulations, pesticides, and environmental chemicals.[3] Activation of PXR by its ligands leads to the induction of Phase-I, -II, and -III metabolizing enzymes and transporters that aid in the metabolism and clearance of those ligands through feed-forward mechanisms.[1] PXR target genes include several cytochrome P450 (CYP) enzymes, glutathione S-transferases, sulfotransferases (SULT), and transporters such as ATP binding cassette proteins.[4] Several *in vitro* and *in vivo* studies have identified PXR as the major regulator of CYP3A4, which metabolizes more than half of all approved drugs.[5−7] Induction of major metabolizing enzymes and transporters not only affects the activating ligands but also can affect endogenous molecules, such as bile acids and steroid hormones, and thereby alter normal physiological processes.[8] PXR activation can also lead to adverse drug−drug interactions between co-administered drugs via increased drug metabolism, which leads to toxic metabolite accumulation or increased transport-mediated drug efflux, which in turn decreases the concentration of co-administered drugs.[9] For example, clinical studies show that long-term administration of the human PXR (hPXR) agonist rifampicin abolishes the antihypertensive effect of the co-administered drug verapamil.[10,11] Other studies show that rifampicin decreases the efficacy of co-administered oral contraceptives and HIV protease inhibitors through PXR-mediated CYP3A4 induction.[12,13] Importantly, PXR has been associated with other diseases such as liver steatosis, bone disorders, and cancers.[14] Thus, identifying PXR activation is essential in studies of absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox), as well as for risk assessment of environmental chemicals.

Received: June 28, 2016
Published: September 7, 2016

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

Experimental screening and identification of PXR activators is time-consuming. Computational approaches provide an alternative and efficient way to identify PXR activators.[15] Ekins et al. utilized pharmacophore modeling to identify a PXR pharmacophore with four hydrophobic features and one hydrogen-bond acceptor feature, highlighting the hydrophobic nature of the PXR active site.[16] Structure-based approaches such as docking are less effective in identifying PXR activators than are ligand-based approaches.[3,17] Ung et al. reported the development of quantitative structure−activity relationship (QSAR) models using three different machine learning approaches on a set of 98 hPXR activators collected from the literature, with the best classifier correctly predicting 85% of the activators.[18] Others have developed QSAR models by using the same data set but with other descriptors and machine learning approaches such as Random Forest.[19] Pan et al. reported the identification of new hPXR activators, such as fluticasone, using a Bayesian QSAR model developed on a data set of 177 ligands.[20] Their best Bayesian model had a specificity of 92% and an accuracy of 69% when evaluated with a test set of 145 molecules.[20] Dybdahl et al. described QSAR model development for hPXR binding by using Leadscope structural features and partial logistic regression, which resulted in an overall accuracy of 84% based on cross-validation analysis.[21] Matter et al. utilized a data set of 434 molecules and developed classification models based on decision trees.[22] Recently, Shi et al. reported a Bayesian classification model for hPXR activators by using a large set of 532 compounds, which included those used in earlier computational studies along with additional compounds collected from recent literature data.[23] Although their data set is not publicly available, it represents the largest data set used for developing computational models for hPXR activation thus far. Their best classifier showed an accuracy of 92.7% in leave-one-out cross-validation analysis.[23]

One of the limitations of earlier computational work is that the data used to build the models were collected from the literature and represent results from different experimental groups using different assay formats, reporter genes, etc. Using data generated from the same assay, preferably from the same group, will provide consistent quality data for model development. Moreover, many PXR ligands exhibit species-specific effects, which were not analyzed in earlier computational QSAR studies. Such effects result from the ligand-binding domain of PXR, which is less conserved than its DNA-binding domain. For example, rifampicin strongly activates human PXR but not rat PXR; pregnenalone 16α-carbonitrile activates rat PXR but not human PXR; and progesterone activates both rat and human PXR. To address this data gap, Shukla et al. recently carried out a systematic screening of a large chemical library (nearly 2000 compounds) for rPXR and hPXR activation using quantitative high-throughput screening.[24] This PubChem data set, which is now publicly available, represents the largest set of compounds screened for PXR activation.[25] Moreover, because PXR is a transcription factor, its activation will be reflected in the increased or decreased expression of its target genes. The computational studies reported thus far have not utilized this characteristic of PXR to study *in vitro*−*in vivo* concordance.

Here, we utilized the PubChem data set to develop Bayesian computational models that predict rat and human PXR activators. The main difference between this work and earlier studies is that we utilized a large data set of compounds screened in the same assay to develop the PXR models. To the best of our knowledge, there are also no earlier reports of rat PXR activation models. Our analysis shows that hydrophobic compounds are preferred as PXR ligands, with notable differences between PXR activators and nonactivators in terms of molecular properties such as molecular weight, number of rings, and solubility. We developed Bayesian models with the extended connectivity fingerprint 4 (ECFP4) fingerprint along with molecular properties, and we used 5-fold cross-validation to evaluate the models. Overall, our models displayed an accuracy of 89% in cross-validation analysis. We further evaluated the models by using Y-randomization and external testing with a ToxCast data set. We identified the key structural features associated with rat as well as human PXR activators. Finally, we analyzed DrugMatrix, a large *in vivo* toxicogenomic data set, and found reasonably good *in vitro*−*in vivo* concordance for PXR activation. The approach used in this work provides a framework for integrated chemoinformatic−toxicogenomic analyses. The models developed in this work could be used as screening tools to evaluate the species-specific PXR activation potential of chemicals.

2. MATERIALS AND METHODS

2.1. Data Set and Preprocessing. We retrieved rat and human PXR activation screening data for 2864 compounds from PubChem, a public repository of experimental screening data for millions of compounds across various biological targets.[25] The PubChem assay IDs for rat and human PXR activation screening data are AID651751 and AID720659, respectively.[26,27] These are cell-based and quantitative high-throughput screening assays that utilize the luciferase reporter gene system. Details of the assay have been reported in earlier publications from the National Institutes of Health Chemical Genomics Center group.[24,28] Each compound in these data sets is given a PubChem activity score depending on its type of concentration−response curve, maximal response (efficacy), and concentration at half-maximal activity. A compound with a PubChem activity score of >40, 1−39, or 0 is denoted as *active*, *inconclusive*, or *inactive*, respectively. A compound with single-point activity is denoted as inconclusive.

We processed the rat and human PXR data separately. First, the compounds were checked for duplicates. If duplicated compounds had the same activity, then the duplicate was removed; otherwise, both compounds were removed from the data set. Next, we used Pipeline Pilot protocols to remove salts as well as mixtures and standardize the molecules.[29] We retained the active and inactive compounds and removed the inconclusive compounds from the data set. After preprocessing, we obtained 2079 compounds (111 actives and 1968 inactives) for rPXR and 1830 compounds (180 actives and 1650 inactives) for hPXR on the PXR screening data sets.

2.2. Chemical Space Networks and Diversity Analysis. Chemical space networks provide better visualization of a given data set than do traditional coordinate-based chemical space representations. In addition, they offer insight into the diversity of the data set.[30] We calculated the similarity values of each compound with every other compound in the data set, using the ECFP4 fingerprint available in Pipeline Pilot, version 9.2, and created similarity matrices for the rat and human PXR data sets.[31] The Tanimoto coefficient was used as the similarity measure. We used an R script to convert the similarity matrix into a simple interaction format (SIF) network file format.[32] In this network, every node represents a compound and every edge represents a Tanimoto coefficient. We retained edges with a Tanimoto coefficient of >0.5. The resulting network was visualized in Cytoscape.[33] We calculated the network density as the number of observed edges divided by the number of possible edges. The latter was calculated as

$$\text{number of possible edges} = \frac{n(n-1)}{2} \quad (1)$$

where n is the number of nodes in the network.

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

2.3. Model Building. Model building involves calculating molecular descriptors and utilizing them with a suitable classification approach. We utilized the physicochemical properties and structural fingerprints of a compound as molecular descriptors for model development. The eight physicochemical properties used as descriptors in this study were molecular weight, AlogP, number of rings, number of rotatable bonds, number of hydrogen-bond acceptors, number of hydrogen-bond donors, solubility, and molecular polar surface area. The ECFP with a diameter of 4, 6, 8, or 10 (i.e., ECFP4,-6,-8,-10), functional class fingerprint (FCFP4), connectivity fingerprint with AlogP atom types (LCFP4), path fingerprints (EPFP4, FPFP4, and LPFP4), and MDL public keys were used as the structural fingerprint descriptors. We used Pipeline Pilot, version 9.2, for descriptor calculation and model building.

We used the Bayesian classification approach implemented in Pipeline Pilot for building classification models. This is one of the most popular classification approaches used in multiple drug design and ADME/Tox studies to distinguish sets of active and inactive compounds.[34−41] This approach has the ability to classify large data sets, handle unbalanced data sets with a small number of active compounds, and identify the top features that make major contributions to the model; in addition, it requires no tuning parameters.[34] Details of the Bayesian classifier approach have been described earlier.[42] Briefly, this approach uses Bayes' theorem and a "learn-by-example" model to predict the likelihood that a given compound is active. It calculates the frequency of occurrence of each molecular feature in the active compounds compared with all compounds in the data set and generates as output a Laplacian-adjusted probability estimate, which is a relative predictor of the likelihood of compounds being from the active set.[36,42] The Laplacian correction accounts for differences in the sampling frequencies of each feature.[36]

2.4. Model Validation. We carried out 5-fold cross-validation to validate the model. In this procedure, the data set was split into five groups and one group was left out; subsequently, the model built from the compounds in the remaining four groups was used to predict the compounds in the left out group. Once we completed this cycle of prediction by leaving out each of the five groups, we calculated the model evaluation parameters such as sensitivity, specificity, accuracy, balanced accuracy, and kappa. Sensitivity (also known as the recall or true positive rate) is the ability to correctly predict positive results; specificity (also known as the true negative rate), the ability to correctly predict negative results; accuracy, the total percentage correctly predicted; and kappa, a measure that compares the probability of correct prediction with the probability of correct prediction by chance. These parameters are defined as follows:

$$\text{sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\text{balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2} \tag{5}$$

$$\text{kappa} = \frac{\text{accuracy} - Pr(e)}{1 - Pr(e)} \tag{6}$$

In eqs 1−4, TP refers to true positive, TN, true negative, FP, false positive, and FN, false negative. Pr(e), which is an estimate of correct prediction by chance, is calculated as follows:

$$Pr(e) = \frac{(TP + FN)(TP + FP) + (FP + TN)(TN + FN)}{(TP + TN + FP + TN)^2} \tag{7}$$

We generated the receiver operating characteristic (ROC) curve and calculated the area under the curve (AUC). We further validated the model by repeatedly shuffling the activity values 100 times and generating models by using randomly generated data (Y-randomization). This approach is useful for confirming that the models are not obtained by chance correlations.

2.5. External Test Data Set. An external test set of 2540 compounds from ToxCast was used for evaluation.[43] ToxCast is a U.S. Environmental Protection Agency initiative to screen and prioritize chemicals based on their bioactivity profile across multiple *in vitro* assays.[44] The PXR activation profile was screened as a part of ToxCast using the cellular biosensor system Factorial (Attagene Inc., Research Triangle Park, NC).[45] We used 2540 compounds that were active in both CIS and TRANS Attagene assays for PXR activation.[45] We preprocessed this data set and removed 830 overlapping compounds that were also present in the hPXR data set as well as duplicate molecules. The final data set comprised 1677 compounds with 648 actives and 1029 inactives. As a further test of external validation, we used the data set of Benod et al., who recently reported 27 PXR activators.[46]

2.6. Comparison of *in Vitro* and *in Vivo* Data. We utilized DrugMatrix, a publicly available toxicogenomics database that contains gene expression data obtained from Sprague−Dawley rats after exposure to a range of chemicals at different doses and time intervals.[47] We downloaded the DrugMatrix data from the National Institute of Environmental Health Sciences server and focused on liver data generated by using the Affymetrix rat 230 2.0 GeneChip array.[48] We followed the preprocessing protocol as described in our earlier study.[49,50] Briefly, we used the R/BioConductor package *affy* and *ArrayQualityMetrics* to perform quantile normalization and assess the quality of the microarray data. We removed outlier arrays and renormalized the data. We used the *MAS5calls* function in the *affy* package to obtain "Present/Absent" calls for each probe set and removed probe sets that were "Absent" in all replicates across all chemical exposures. We used the BioConductor *genefilter* package and performed gene-level filtering to remove genes showing low variance across chemical exposures. After calculating the average intensity of replicates of a chemical exposure condition, for each gene we computed log ratios between treatments and their corresponding controls. To select unique chemical exposure conditions, we chose chemical exposures with >1 day of exposure at the highest dose tested. This gave us a data set of 8992 genes and 170 chemical exposure conditions.

Ihunnah et al. previously listed the known PXR target genes.[4] We identified 17 genes that matched our preprocessed DrugMatrix data. We converted rat Affymetrix probe IDs to rat and human gene symbols using the BioConductor/R packages *annotate* and *biomaRt*.[51] The human *CYP3A4* mapped to two rat genes, namely, *Cyp3a9* and *Cyp3a23/3a1*. We clustered the log ratio matrix of 170 chemical exposures across 18 genes using the R hierarchical clustering function *hclust* in the *stats* package.[32] We used the Euclidean distance and complete linkage method to perform the clustering.

3. RESULTS AND DISCUSSION

PXR is an important target in drug design and toxicology, with species-specific differences due to its variable ligand-binding domain. For example, the DNA-binding domains for hPXR and rPXR have 96% sequence similarity, whereas the ligand-binding domains of these two species have only 76% sequence similarity.[52] Although most earlier experimental and computational studies have predominantly focused on identifying hPXR activators, the importance of the rat as a model organism warrants the characterization of rPXR activators to identify species-specific effects. Here, we utilized the largest publicly available PXR screening data set to develop computational models that predict hPXR and rPXR activation.

3.1. Overview of PubChem PXR Data Set. After preprocessing and removing duplicates, mixtures, and incon-

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

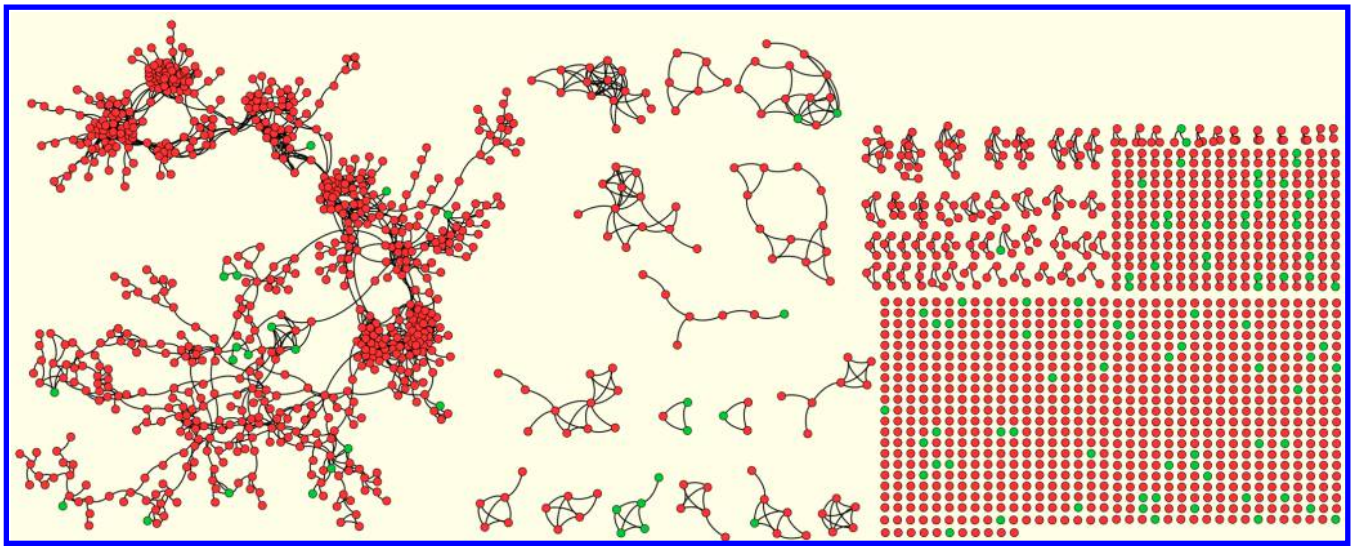


Figure 1. Chemical space networks of 2079 rat pregnane X receptor (rPXR) compounds. The actives are colored in green, and the inactives, in red. Each node in the network represents a compound, and the edge between two nodes represents the similarity between them (i.e., similar nodes are connected). Isolated nodes indicate lack of similarity with other compounds in the network at the given Tanimoto coefficient threshold of 0.5.

sistent or inconclusive data points, we obtained a final data set of 2079 and 1830 compounds for rPXR and hPXR, respectively (Figure S1, Supporting Information). The rPXR data set included 203 approved drugs, 353 agrochemical compounds, and 35 carcinogens or mutagens (Table S1, Supporting Information). The hPXR data set included 151 approved drugs, 280 agrochemical compounds, and 36 carcinogens or mutagens (Table S2, Supporting Information). We utilized a network approach to analyze the chemical space and provide a visual summary of the activity landscape and diversity of the data set.[30] Figure 1 and Figure S2 (Supporting Information) show the chemical space networks of the rPXR and hPXR data sets, respectively. In these networks, nodes represent compounds and edges between nodes represent the similarity between them. Nodes representing similar molecules are more connected to each other than to other nodes. Isolated nodes indicate a lack of similarity with other compounds in the network at the given Tanimoto coefficient threshold. The topological properties of the network provide information about the underlying data set. For example, in a less diverse data set, most nodes will be connected to each other and form a densely connected network with few isolated nodes. The chemical space network of compounds in the rPXR (hPXR) data set had 964 (888) connected components and 767 (722) isolated nodes. The network density was 0.1% for both the rPXR and hPXR networks, indicating that these are diverse data sets. The set of approved drugs represents a diverse chemical space; therefore, to gauge the diversity of the PXR networks in relation to a reference set, we created the chemical space network for all approved drugs (n = 1789), using the same parameters as those used in generating the PXR networks (i.e., ECFP4 fingerprint and a Tanimoto threshold of 0.5). The chemical space network of approved drugs had 1061 connected components and 860 isolated nodes with a network density of 0.1%, indicating that the PXR data sets are comparably diverse. These networks can also be used to delineate structure−activity relationships and identify activity cliffs among the tested compounds (Figure S3, Supporting Information).

Figure 2 shows the overlap between the rPXR and hPXR data sets with 1707 compounds (>80%) present in both. The

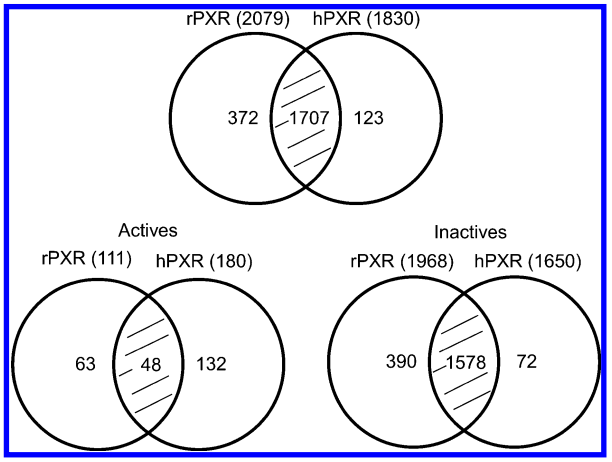


Figure 2. Venn diagram showing the overlap of compounds between rat pregnane X receptor (rPXR) and human PXR (hPXR) data sets.

majority of these compounds were inactives (1578), and only 48 were actives. Among the 132 active compounds unique to hPXR, 72 were inactive and 60 were inconclusive in the rPXR assay. Among the 63 active compounds unique to rPXR, 9 were inactive and 54 were inconclusive in the hPXR assay. Figure S4 (Supporting Information) provides examples of compounds that activated either rPXR or hPXR alone or both.

3.2. Analysis of Molecular Properties of Rat and Human PXR Activators and Nonactivators. We analyzed the variation in the following eight physiochemical properties between the PXR activators and nonactivators in both species: molecular weight (MW), log of the octanol/water partition coefficient (AlogP), number of rings (nRing), solubility, polar surface area, hydrogen-bond acceptors (HBA) and donors (HBD), and number of rotatable bonds. Figure 3 and Figure S5 (Supporting Information) show box plots with the median, quartiles, and extreme values of these eight properties. We used the nonparametric Mood's median test to determine whether the difference between the medians of PXR activators and nonactivators for each physiochemical property was significant. Our analysis showed that, with the exception of hydrogen-bond

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

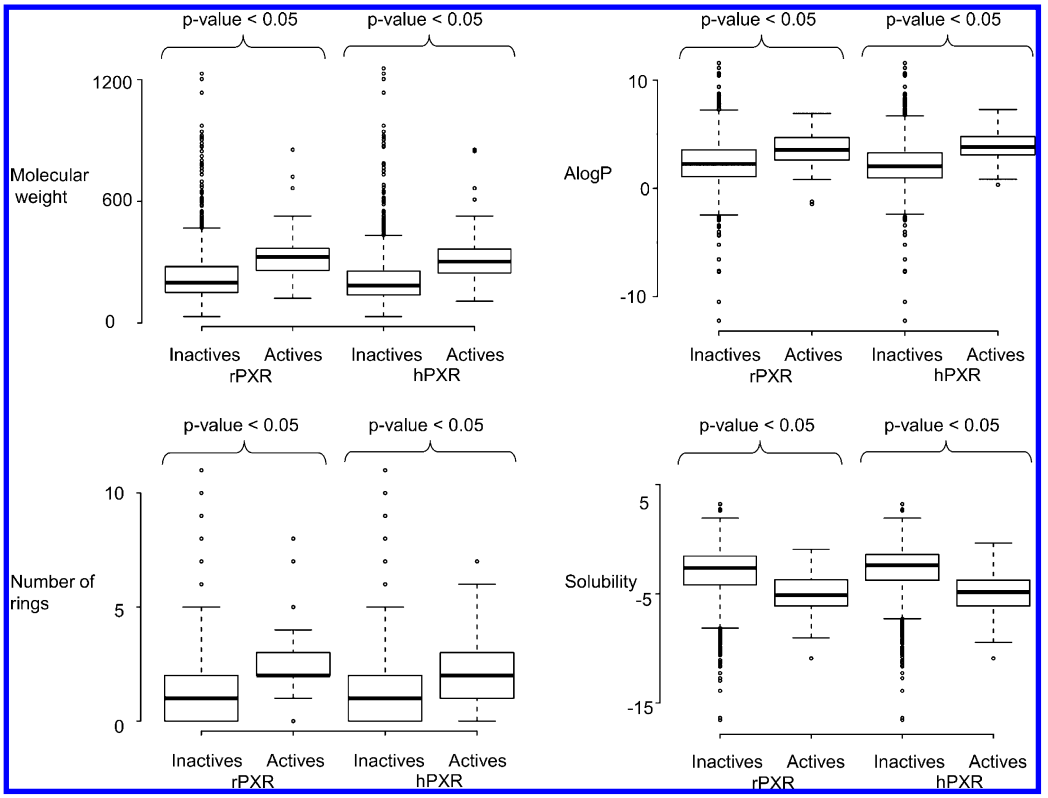


Figure 3. Box plots showing the distribution of four molecular properties (molecular weight, AlogP, number of rings, and solubility) among actives and inactives in rat pregnane X receptor (rPXR) and human PXR (hPXR) data sets.

Table 1. Evaluation of Bayesian Classification Models Using Various Descriptors and 5-Fold Cross-Validation

species	model[a]	accuracy	sensitivity	specificity	kappa	AUC
rPXR	ECFP4	0.83	0.59	0.84	0.20	0.79
	ECFP6	0.85	0.59	0.86	0.23	0.79
	ECFP8	0.87	0.56	0.88	0.25	0.78
	ECFP10	0.86	0.55	0.88	0.24	0.78
	MDL keys	0.76	0.61	0.77	0.14	0.72
	DES8	0.75	0.86	0.74	0.19	0.85
	DES8 + ECFP4	0.83	0.77	0.84	0.27	0.89
	DES8 + ECFP6	0.84	0.74	0.85	0.28	0.88
	DES8 + ECFP8	0.85	0.74	0.85	0.28	0.88
	DES8 + MDL keys	0.78	0.80	0.77	0.21	0.85
hPXR	ECFP4	0.77	0.58	0.79	0.22	0.77
	ECFP6	0.82	0.52	0.85	0.26	0.77
	ECFP8	0.80	0.54	0.83	0.25	0.76
	ECFP10	0.76	0.56	0.78	0.19	0.76
	MDL keys	0.78	0.59	0.80	0.24	0.73
	DES8	0.73	0.88	0.71	0.28	0.86
	DES8 + ECFP4	0.82	0.75	0.83	0.37	0.89
	DES8 + ECFP6	0.82	0.75	0.83	0.37	0.88
	DES8 + ECFP8	0.82	0.75	0.82	0.36	0.88
	DES8 + MDL keys	0.78	0.76	0.78	0.31	0.87

[a]ECFP: extended connectivity fingerprint; DES8: eight molecular properties, namely, molecular weight, log of the octanol/water partition coefficient, number of rings, solubility, polar surface area, hydrogen-bond acceptors and donors, and number of rotatable bonds.

donors ($p > 0.3$), all other properties were significantly different between PXR activators and nonactivators in both species. For example, the mean and median molecular weights of rPXR activators were 325 and 326, respectively, whereas those for rPXR nonactivators were 228 and 199 (Figure 3; Table S3, Supporting Information). Our results are in contrast to those of an earlier report by Shi et al., who used the same physicochemical properties such as molecular weight and number of rotatable bonds but did not observe any significant difference between hPXR activators and nonactivators.[23] Sample size could have influenced the results, given that the earlier study used a data set of 532 compounds whereas we used a larger data set of 1830 compounds. PXR activators had more hydrogen-bond acceptors than hydrogen-bond donors,

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

consistent with earlier reports.[16,18] Overall, the significant differences in molecular properties show that PXR activators of both species tend to be heavier (larger molecular weight), more hydrophobic (higher AlogP, more rings, and lower solubility), and more flexible (higher number of rotatable bonds) than nonactivators (Figure 3). X-ray crystal structures of PXR show that the ligand-binding site is flexible and predominantly composed of hydrophobic residues with few polar residues, in agreement with the observed differences in the molecular properties of the ligands. We also analyzed whether any of these eight properties differed significantly between rPXR and hPXR activators. Among these properties, the number of rings was significantly different ($p < 0.05$) between rPXR and hPXR activators (Figure S6). We observed that 75% of rPXR activators had two or more rings, whereas this was true for only 50% of hPXR activators.

3.3. Bayesian Classifier for Rat and Human PXR Activators. To overcome the overall imbalance between inactives and actives in the assay data sets, we developed our classification models by using the Bayesian approach. We utilized 5-fold cross-validation and explored model building by using various descriptors, including structural fingerprints (ECFP4, -6, -8, -10, FCFP4, LCFP4, EPFP4, FPFP4, LPFP4, and MDL keys) and the eight molecular properties mentioned above (Table 1; Table S4, Supporting Information). We included structural fingerprints to help us identify key molecular scaffolds that contribute to the activity. Using molecular properties alone as descriptors led to classifiers with higher sensitivity but lower specificity and accuracy, whereas using structural fingerprints alone generated classifiers with lower sensitivity but higher specificity and accuracy. We observed optimal classification performance, as noted by the high values of the ROC-AUC, balanced accuracy (i.e., the average of sensitivity and specificity), and kappa for a model based on a combination of ECFP4 fingerprint and the eight molecular properties (Figure 4, Table 1). The rPXR model had slightly higher sensitivity and lower kappa values compared with the hPXR model. The hPXR data included more actives (180, ~10% of the data) than did rPXR data (111, ~5% of the data), which partly explains the difference between the models (e.g., lower kappa values). We further analyzed whether the model performed better than random chance by shuffling the activity values (Y-randomization) 100 times. The random models were associated with sensitivity and specificity values of ~50% and a kappa value of zero (Figure 4). Hence, our model parameters were not obtained by random chance.

We explored whether the model was truly species-specific by interchanging the species data, i.e., by using the rPXR model to predict hPXR data and vice versa (Table S5, Supporting Information). The balanced accuracy and kappa values decreased when the model developed with the data from one species was used to predict data for the other species. This indicated that the models captured species-specific effects.

Finally, we evaluated the models by using two external data sets that were not used to construct the models. The ToxCast data set consists of 2540 compounds screened for hPXR activation, using an assay format different from that of the PubChem assay. We found 830 overlapping compounds among those screened in both data sets. First, we analyzed the agreement between the two assays by using these overlapping compounds. The PubChem assay showed an accuracy of 82% in predicting the ToxCast assay results. Table S6 (Supporting Information) shows a list of the 830 overlapping compounds

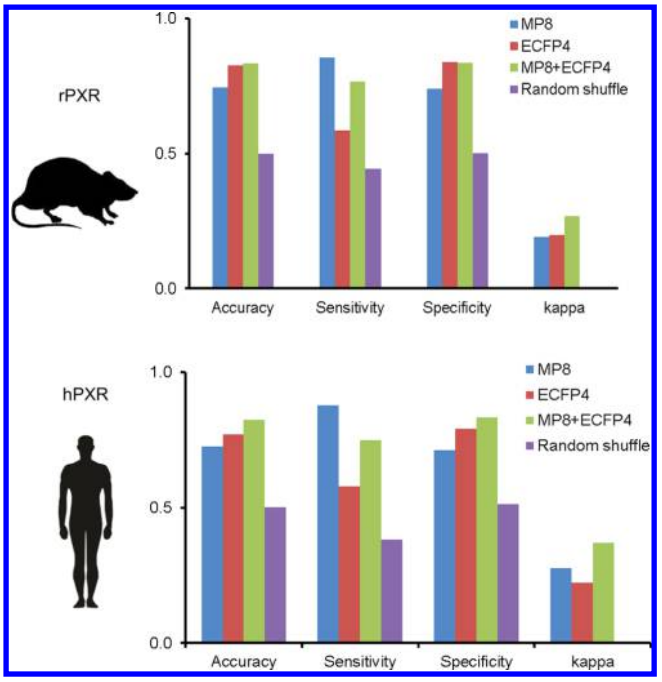


Figure 4. Performance of models in 5-fold cross-validation of rat pregnane X receptor (rPXR) and human PXR (hPXR) data sets. The models were generated by using the eight molecular properties (MP8) alone, extended connectivity fingerprint-4 (ECFP4) alone, a combination of MP8 and ECFP4, or random shuffling of activity values.

along with their activity values. After removing these overlapping compounds, we used the remaining 1677 compounds (648 active and 1029 inactive) as the external test data set. The accuracy, sensitivity, specificity, and kappa values of the hPXR model for the ToxCast data set were 77, 61, 87, and 49%, respectively. Ng et al. reported similar results when they used ToxCast as the external test data set for their estrogen binding model.[53] Similarly, we utilized a second set of 27 PXR activators recently reported by Benod et al. and obtained a sensitivity of 63%.[46]

3.4. Analysis of Key Structural Features. In constructing our Bayesian models, a score was assigned to each structural feature, using the Laplacian-adjusted probability estimate. This score is representative of the relative contribution of the structural feature to the final classification model. The higher a feature's score, the more likely that it contributes to "PXR-agonist" likeness, whereas the lower its score, the less likely it is to do so. One of the advantages of developing Bayesian classifiers by using structural fingerprints is that we can use the score associated with each feature and identify the top structural fragments that contribute to a model. Figures S7 and S8 (Supporting Information) show the top 15 structural fragments associated with actives and inactives of rPXR and hPXR, respectively. We compared the top features from rPXR and hPXR models. Figure 5 lists the top structural features common to both models as well as top features that occur only in one of the models.

Ring scaffolds such as triazole and tricyclic rings dominated the top predictive features for both rPXR and hPXR actives. This agrees with previous reports; for example, N-substituted azoles are a well-known class of PXR activators.[54] The top features for the PXR inactives in both species were predominantly linear molecules such as propyl, butyl groups

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

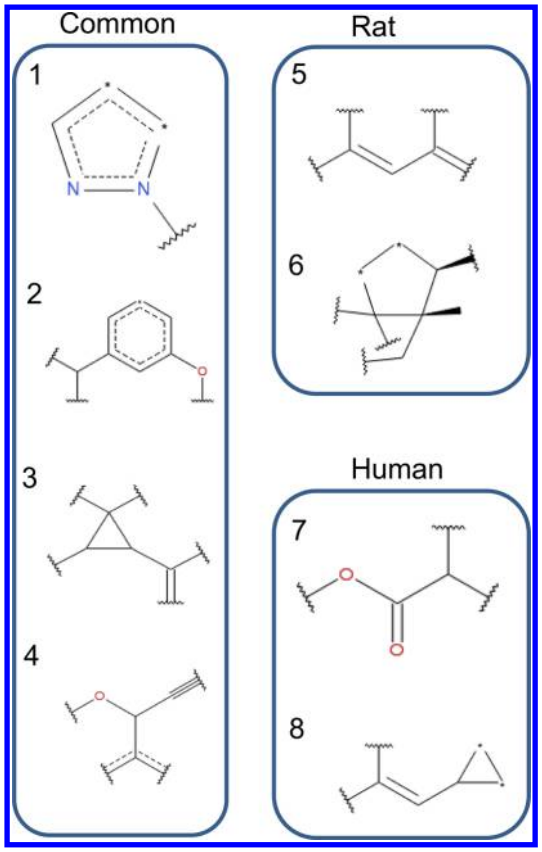
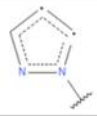
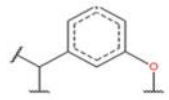
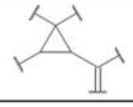
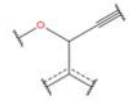
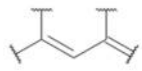
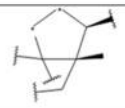
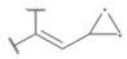


Figure 5. Top structural features associated with both rat pregnane X receptor (rPXR) and human PXR (hPXR) actives or with actives of only one species.

with acid, or hydroxyl and phenol groups. Among the actives, the four most commonly occurring predictive features for rPXR and hPXR activation were the five-membered triazole ring, substituted cyclopropane, the phenoxy group, and methoxyprop-1-yne. The partial cyclohexene-like feature was identified as important for rPXR activation but not for hPXR activation. This scaffold maps to steroidal molecules such as hydroxylprogesterone and mifepristone, which were active only in the rPXR assay. Features such as esters and substituted vinylcyclopropane were identified as the top features unique to the hPXR model. Nitro-phenol and sulfonyl groups were identified as the top features associated with only hPXR inactives and not with rPXR inactives. We utilized the top features as structural alerts and calculated the percentage of compounds that matched with these alerts among active and inactive compounds in the rPXR and hPXR data (Table 2). We found that common alerts such as the triazole ring mapped to >5% of rat and human PXR activators. Some top features such as partial cyclohexene mapped to 21.6% of rPXR activators but only to 10.6% of hPXR activators. These findings highlight the potential utility of using the features listed in Table 2 as structural alerts to screen external databases.

3.5. Analysis of *in Vivo* Toxicogenomic Data. The bulk of the experimental work used to identify PXR activators employs *in vitro* assays because *in vivo* animal studies are inherently low-throughput assays. Currently, there is a major impetus from regulatory agencies and industry to utilize *in vitro* and *in silico* approaches as alternative testing strategies to animal studies. Understanding *in vitro*−*in vivo* concordance should help clarify the utility and limitations of each approach

Table 2. Bayesian Scores of Top Alerts Mapped to rPXR and hPXR Data

No	Sub-structure	Bayesian score	
		Rat	Human
1		1.4	1.2
2		1.4	1.3
3		1.3	1.4
4		1.5	1.4
5		1.4	0.3
6		1.4	0.0
7		0.2	1.4
8		0.0	1.4

and suggest when alternative approaches are appropriate. PXR is an exemplary candidate for these studies because activation of PXR leads to increased expression of its target genes, which can be detected by analyzing gene expression data sets. To this end, we utilized DrugMatrix, a large *in vivo* toxicogenomics data set of gene expression data obtained from rats after exposure to diverse chemicals, to examine the *in vitro*−*in vivo* concordance for PXR activators.

We focused on 170 DrugMatrix liver chemical exposures with dosing regimens longer than 1 day of exposure at the highest available dose, to map the expression profiles of 18 known PXR target genes in the DrugMatrix data. Chemicals with the same mechanism of action are expected to have similar gene expression patterns and thus exhibit similarities based on these patterns. We clustered the chemical exposures by the transcriptional fold-change values for the 18 PXR target genes. Figure 6 shows the clustering results with two main compound/exposure groups. We used our rPXR model to predict the activity of the DrugMatrix compounds in these clusters, with the aim of associating them with *in vivo* activity. Three compounds (lipopolysaccharide and two lead compounds) were excluded from prediction. We found that 43% of compounds associated with *Cluster 2* were predicted to be active, whereas 30% of compounds associated with *Cluster 1* were predicted to be actives. Because PXR is a major regulator

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

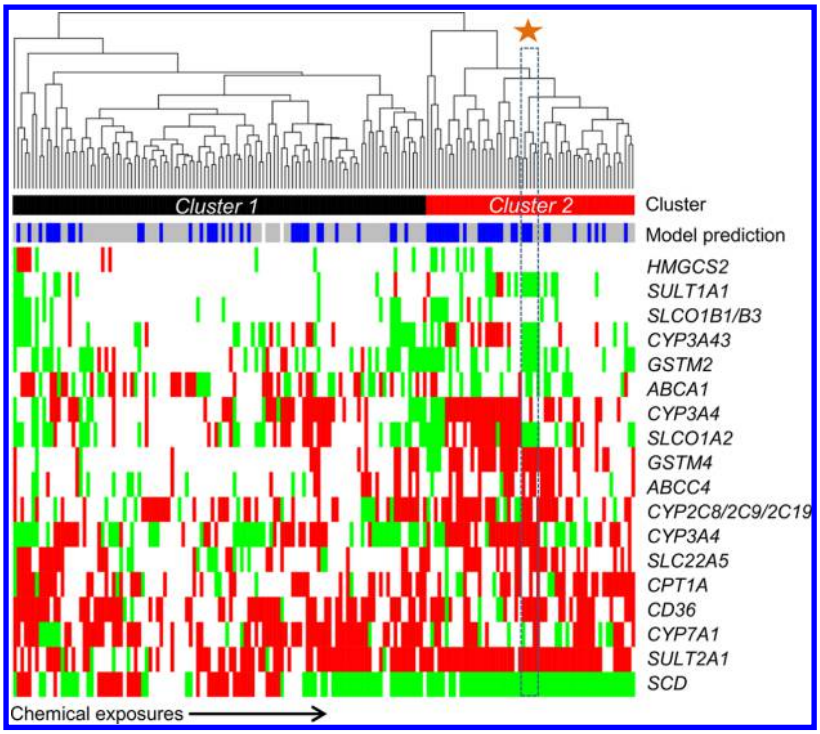


Figure 6. Hierarchical clustering of 170 DrugMatrix chemicals in rat liver data across 18 pregnane X receptor (PXR) target genes. The cluster membership of the 170 chemicals is indicated in top bar labeled "cluster". The chemicals predicted as actives and inactives by the rat pregnane X receptor (rPXR) classification model are shown in dark blue and gray, respectively, in the bar labeled "Model prediction". The three excluded chemicals are shown in white. Genes with $\log_2$ fold-change ratios greater than 0.6 are colored in red, those with $\log_2$ fold-change ratios less than 0.6 are in green, and those with $\log_2$ fold-change ratios between 0.6 and −0.6 are in white.

of *CYP3A4* and *SULT2A1*, we analyzed the percentage of compounds that upregulate [$\log_2$(fold-change > 0.6)] these two genes in both clusters. We found that 65% (35%) and 90% (53%) of the compounds in *Cluster 2* (*Cluster 1*) upregulated *CYP3A4* and *SULT2A1*, respectively. We found that other targets such as the fatty acid transporter *CD36*, which is also upregulated by proteins such as the liver X receptor (LXR) and peroxisome proliferator activated receptor-γ (PPAR-γ),[55] was upregulated by roughly the same percentage of compounds in *Cluster 2* (47%) and *Cluster 1* (53%). On the basis of these observations, we effectively designated *Cluster 2* as the active set and *Cluster 1* as the inactive set.

Next, we mapped the rPXR (PubChem) compounds to the DrugMatrix data set and identified 60 compounds present in both data sets. Five of the six actives in this set clustered together and were present in *Cluster 2*. Lovastatin was the active compound that did not cluster with other actives. An analysis of all lovastatin exposures at different doses and time points showed that PXR was not activated under any condition (Figure S9, Supporting Information). Apart from lovastatin, other drugs of the statin class in DrugMatrix, such as simvastatin, cerivastatin, fluvastatin, and atorvastatin, were found in *Cluster 1*, suggesting that this class of compounds may not activate rPXR *in vivo*. An analysis of the literature showed that lovastatin has a rat-specific metabolic transformation reaction that could contribute to its inactivity toward PXR.[56] Conversely, 54 rPXR inactives were mapped to the DrugMatrix compounds, 39 of which were present in *Cluster 1*. Among the rPXR inactives present in *Cluster 2*, three were estrogen modulators, namely, diethylstilbestrol, β-estradiol, and tamoxifen. An analysis of the DrugMatrix clustering data showed that a set of estrogenic compounds including estradiol, mestranol, ethinyl estradiol, β-estradiol, and diethylstilbestrol were grouped together in *Cluster 2* (Figure 6, orange star). These compounds mainly target estrogen receptors and all activate key PXR target genes such as *CYP3A4*, *CD36*, and *SULT2A1* (Figure 6). Nuclear receptors such as the estrogen receptor, constitutive androstane receptor, and PXR are involved in crosstalk, which influences the expression of metabolism-related genes.[57] PXR-mediated gene transcription involves ligand binding, translocation to the nucleus, heterodimerization with the retinoid X receptor (RXR), recruitment of coactivators such as steroid receptor coactivators-1 (SRC-1) and PPAR-γ coactivator 1α (PGC-1α), and binding of this complex to xenobiotic response elements (XRE) located in the promoter region of target genes. The nuclear receptors have common coactivators, which could affect their function and contribute to the crosstalk among them. For example, PGC-1α acts a coactivator for number of other nuclear receptors including the estrogen receptor, PPAR-γ, constitutive androstane receptor, LXR, and farnesoid X receptor.[57] *In vitro* rPXR assays may not be able to capture the effects from such crosstalk, and this inability could be a potential limitation of the *in vitro* assay. Table 3 shows the overlap of PXR actives and inactives from the *in vitro* rPXR and *in vivo* DrugMatrix data. On the basis of our separation of active and inactive compounds *in vivo*, this calculation translates to a sensitivity of 83% (5/6), a specificity of 72% (39/54), and a balanced accuracy of 78%. This mapping is far from capturing all *in vivo* effects associated with PXR activation. Nevertheless, overall we find a high concordance between *in vitro* and *in vivo* PXR activation.

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

Table 3. Compounds Present in Both rPXR and DrugMatrix Data and Their PXR Activity[a]

CID	name	class	rPXR	DrugMatrix	hPXR
5833	spironolactone	drug	1	1	1
55245	mifepristone	drug	1	1	2
5743	dexamethasone	drug	1	1	–
28417	danazol	drug	1	1	–
68589	econazole	drug	1	1	–
53232	lovastatin	drug	1	2	1
7108	phenothiazine	–	2	1	–
1057	pyrogallol	–	2	1	2
5144	safrole	–	2	1	2
7858	allyl alcohol	–	2	1	2
2554	carbamazepine	drug	2	1	–
448537	diethylstilbestrol	drug	2	1	–
7048670	β-estradiol	drug	2	1	–
2733525	tamoxifen	drug	2	1	–
5280795	cholecalciferol	drug	2	1	–
6323490	rifabutin	drug	2	1	1
54900	raloxifene	drug	2	1	1
3033832	clomiphene	drug	2	1	1
3715	indomethacin	drug	2	1	2
443939	doxorubicin	drug	2	1	2
443939	epirubicin	drug	2	1	2
11080	1-naphthyl isothiocyanate	–	2	2	1
5280961	genistein	–	2	2	1
289	catechol	–	2	2	2
2406	bithionol	–	2	2	2
2723	chloroxylenol	–	2	2	2
5694	pirinixic acid	–	2	2	2
5921	*N*-nitrosodiethylamine	–	2	2	2
6212	chloroform	–	2	2	2
6228	*N*,*N*-dimethylformamide	–	2	2	2
6366	1,1-dichloroethene	–	2	2	2
8447	benzothiazyl disulfide	–	2	2	2
853433	isoeugenol	–	2	2	2
16220118	zomepirac	–	2	2	2
7577	4,4′-methylenedianiline	agrochemical	2	2	2
1983	acetaminophen	drug	2	2	2
2244	aspirin	drug	2	2	2
2265	azathioprine	drug	2	2	2
2578	carmustine	drug	2	2	2
2708	chlorambucil	drug	2	2	2
2796	clofibrate	drug	2	2	2
2797	clofibric acid	drug	2	2	2
3672	ibuprofen	drug	2	2	2
5147	salicylamide	drug	2	2	2
6013	testosterone	drug	2	2	2
11057	gentian violet	drug	2	2	2
18283	stavudine	drug	2	2	2
657298	propylthiouracil	drug	2	2	2
5281034	oxymetholone	drug	2	2	2
5284373	cyclosporin A	drug	2	2	2
456201	ketoconazole	drug	2	2	2
3365	fluconazole	drug	2	2	–
4133	methyl salicylate	drug	2	2	–
5282379	isotretinoin	drug	2	2	–
3339	fenofibrate	drug	2	2	–
3463	gemfibrozil	drug	2	2	–
39042	bezafibrate	drug	2	2	–
6014	promethazine	drug	2	2	–
8478	benzethonium chloride	–	2	2	–
54680691	oxytetracycline	–	2	2	–

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

Table 3. continued

[a]Active compounds are labeled with 1s and inactive compounds are labeled with 2s. Dashes indicate either that data are absent for those compounds or that their chemical class is unknown.

4. CONCLUSIONS

Identifying PXR activators is critical in drug design and toxicology studies. In this analysis, we employed the largest publicly available rPXR and hPXR screening data set to develop species-specific PXR classification models for predicting PXR activation. We used a network-based approach to analyze chemical space and qualify the diversity of the data set and structure−activity relationships. We analyzed the molecular properties of PXR activators and nonactivators and showed that PXR activators significantly differ from nonactivators in terms of molecular weight, logP, number of rings, number of rotatable bonds, and solubility (i.e., PXR activators tend be heavier, more hydrophobic, and more flexible). We also showed that activators of rPXR differ from hPXR in terms of the number of rings. We developed species-specific Bayesian classification models by utilizing molecular properties and structural fingerprints, and we systematically evaluated the performance of the models by using 5-fold cross-validation analysis, Y-randomization, and external test set validation. Our best models for rPXR and hPXR had balanced accuracy values of 81 and 79%, respectively. The developed rPXR model is the first of its kind and allows us to investigate species-specific effects, e.g., by analyzing the top structural features associated with PXR activators and identifying those that are either common to or unique for rPXR and hPXR activation. Finally, we utilized a large *in vivo* toxicogenomics data set and performed *in vitro*−*in vivo* comparisons. Analysis of overlapping compounds revealed a high level of *in vitro*−*in vivo* concordance with a balanced accuracy of 78%. We also highlighted the complexity of PXR signaling stemming from nuclear receptor crosstalk, which may affect *in vivo* results in ways that are not captured by the *in vitro* assay. Collectively, these findings suggest that our computational models could serve as efficient initial high-throughput *in silico* screens to predict rat and human PXR activators.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemrestox.6b00227.

Table S1: rPXR data of 2079 compounds with chemical class annotation (XLSX)

Table S2: hPXR data of 1830 compounds with chemical class annotation (XLSX)

Table S6: List of overlapping compounds between PubChem and ToxCast PXR assays (XLSX)

Table S3: Mean values of eight molecular properties for PXR activators and nonactivators. Table S4: Evaluation of Bayesian classification models using various descriptors and 5-fold cross-validation. Table S5: Evaluation of model performance by interchanging species data. Table S7: Model performance on external data sets. Figure S1: Preprocessing flowchart for rPXR and hPXR data. Figure S2: Chemical space network of hPXR data. Figure S3: Examples of structure−activity relationship and activity cliff identified in rPXR chemical space network. Figure S4: Examples of compounds that activated either rPXR or hPXR alone or both PXRs. Figure S5: Box plots of distribution of molecular properties among actives and inactives in rPXR and hPXR data sets. Figure S6: Box plots of distribution of molecular properties among actives in rPXR and hPXR data sets. Figure S7: Top structural features associated with rPXR actives and inactives. Figure S8: Top structural features associated with hPXR actives and inactives. Figure S9: Gene expression profile for lovastatin and dexamethasone at different doses and time points (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: mabdulhameed@bhsai.org. Phone: 301-619-1304 (M.D.M.A).

*E-mail: sven.a.wallqvist.civ@mail.mil. Phone: 301-619-1989 (A.W.).

Funding

The authors were supported by the Military Operational Medicine Research Program and the U.S. Army's Network Science Initiative, U.S. Army Medical Research and Materiel Command (USAMRMC, http://mrmc.amedd.army.mil), Ft. Detrick, MD.

Notes

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. Citations of commercial organizations or trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Drs. Ruifeng Liu and Patric Schyman for valuable discussions and comments on the manuscript.

■ ABBREVIATIONS

PXR, pregnane X receptor; CYP, cytochrome P450; QSAR, quantitative structure−activity relationship; ECFP, extended connectivity fingerprint; HC, hierarchical clustering

■ REFERENCES

(1) Kliewer, S. A., Goodwin, B., and Willson, T. M. (2002) The nuclear pregnane X receptor: a key regulator of xenobiotic metabolism. *Endocr. Rev. 23*, 687−702.

(2) Orans, J., Teotico, D. G., and Redinbo, M. R. (2005) The nuclear xenobiotic receptor pregnane X receptor: recent insights and new challenges. *Mol. Endocrinol. 19*, 2891−2900.

(3) Ekins, S., Kortagere, S., Iyer, M., Reschly, E. J., Lill, M. A., Redinbo, M. R., and Krasowski, M. D. (2009) Challenges predicting ligand-receptor interactions of promiscuous proteins: the nuclear receptor PXR. *PLoS Comput. Biol. 5*, e1000594.

(4) Ihunnah, C. A., Jiang, M., and Xie, W. (2011) Nuclear receptor PXR, transcriptional circuits and metabolic relevance. *Biochim. Biophys. Acta, Mol. Basis Dis. 1812*, 956−963.

(5) Lehmann, J. M., McKee, D. D., Watson, M. A., Willson, T. M., Moore, J. T., and Kliewer, S. A. (1998) The human orphan nuclear

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

receptor PXR is activated by compounds that regulate CYP3A4 gene expression and cause drug interactions. *J. Clin. Invest. 102*, 1016−1023.
(6) Xie, W., Barwick, J. L., Downes, M., Blumberg, B., Simon, C. M., Nelson, M. C., Neuschwander-Tetri, B. A., Brunt, E. M., Guzelian, P. S., and Evans, R. M. (2000) Humanized xenobiotic response in mice expressing nuclear receptor SXR. *Nature 406*, 435−439.
(7) Staudinger, J. L., Goodwin, B., Jones, S. A., Hawkins-Brown, D., MacKenzie, K. I., LaTour, A., Liu, Y., Klaassen, C. D., Brown, K. K., Reinhard, J., Willson, T. M., Koller, B. H., and Kliewer, S. A. (2001) The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. *Proc. Natl. Acad. Sci. U. S. A. 98*, 3369−3374.
(8) Hernandez, J. P., Mota, L. C., and Baldwin, W. S. (2009) Activation of CAR and PXR by Dietary, Environmental and Occupational Chemicals Alters Drug Metabolism, Intermediary Metabolism, and Cell Proliferation. *Curr. Pharmacogenomics Person. Med. 7*, 81−105.
(9) di Masi, A., De Marinis, E., Ascenzi, P., and Marino, M. (2009) Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. *Mol. Aspects Med. 30*, 297−343.
(10) Fuhr, U. (2000) Induction of drug metabolising enzymes: pharmacokinetic and toxicological consequences in humans. *Clin. Pharmacokinet. 38*, 493−504.
(11) Fromm, M. F., Busse, D., Kroemer, H. K., and Eichelbaum, M. (1996) Differential induction of prehepatic and hepatic metabolism of verapamil by rifampin. *Hepatology 24*, 796−801.
(12) Ma, X., Idle, J. R., and Gonzalez, F. J. (2008) The pregnane X receptor: from bench to bedside. *Expert Opin. Drug Metab. Toxicol. 4*, 895−908.
(13) Mallolas, J., Sarasa, M., Nomdedeu, M., Soriano, A., Lopez-Pua, Y., Blanco, J. L., Martinez, E., and Gatell, J. M. (2007) Pharmacokinetic interaction between rifampicin and ritonavir-boosted atazanavir in HIV-infected patients. *HIV Med. 8*, 131−134.
(14) Banerjee, M., Robbins, D., and Chen, T. (2015) Targeting xenobiotic receptors PXR and CAR in human diseases. *Drug Discovery Today 20*, 618−628.
(15) Ekins, S. (2014) Progress in computational toxicology. *J. Pharmacol. Toxicol. Methods 69*, 115−140.
(16) Ekins, S., and Erickson, J. A. (2002) A pharmacophore for human pregnane X receptor ligands. *Drug Metab. Dispos. 30*, 96−99.
(17) Kortagere, S., Chekmarev, D., Welsh, W. J., and Ekins, S. (2009) Hybrid scoring and classification approaches to predict human pregnane X receptor activators. *Pharm. Res. 26*, 1001−1011.
(18) Ung, C. Y., Li, H., Yap, C. W., and Chen, Y. Z. (2006) In silico prediction of pregnane X receptor activators by machine learning approaches. *Mol. Pharmacol. 71*, 158−168.
(19) Khandelwal, A., Krasowski, M. D., Reschly, E. J., Sinz, M. W., Swaan, P. W., and Ekins, S. (2008) Machine learning methods and docking for predicting human pregnane X receptor activation. *Chem. Res. Toxicol. 21*, 1457−1467.
(20) Pan, Y., Li, L., Kim, G., Ekins, S., Wang, H., and Swaan, P. W. (2011) Identification and validation of novel human pregnane X receptor activators among prescribed drugs via ligand-based virtual screening. *Drug Metab. Dispos. 39*, 337−344.
(21) Dybdahl, M., Nikolov, N. G., Wedebye, E. B., Jonsdottir, S. O., and Niemela, J. R. (2012) QSAR model for human pregnane X receptor (PXR) binding: screening of environmental chemicals and correlations with genotoxicity, endocrine disruption and teratogenicity. *Toxicol. Appl. Pharmacol. 262*, 301−309.
(22) Matter, H., Anger, L. T., Giegerich, C., Gussregen, S., Hessler, G., and Baringhaus, K. H. (2012) Development of in silico filters to predict activation of the pregnane X receptor (PXR) by structurally diverse drug-like molecules. *Bioorg. Med. Chem. 20*, 5352−5365.
(23) Shi, H., Tian, S., Li, Y., Li, D., Yu, H., Zhen, X., and Hou, T. (2015) Absorption, Distribution, Metabolism, Excretion, and Toxicity Evaluation in Drug Discovery. 14. Prediction of Human Pregnane X Receptor Activators by Using Naive Bayesian Classification Technique. *Chem. Res. Toxicol. 28*, 116−125.
(24) Shukla, S. J., Sakamuru, S., Huang, R., Moeller, T. A., Shinn, P., Vanleer, D., Auld, D. S., Austin, C. P., and Xia, M. (2011) Identification of clinically used drugs that activate pregnane X receptors. *Drug Metab. Dispos. 39*, 151−159.
(25) Wang, Y., Suzek, T., Zhang, J., Wang, J., He, S., Cheng, T., Shoemaker, B. A., Gindulyte, A., and Bryant, S. H. (2014) PubChem BioAssay: 2014 update. *Nucleic Acids Res. 42*, D1075−1082.
(26) *PubChem BioAssay Database, AID = 651751*, National Center for Biotechnology Information. https://pubchem.ncbi.nlm.nih.gov/bioassay/651751 (accessed May 16, 2016).
(27) *PubChem BioAssay Database, AID = 720659*, National Center for Biotechnology Information. https://pubchem.ncbi.nlm.nih.gov/bioassay/720659 (accessed May 16, 2016).
(28) Shukla, S. J., Huang, R., Austin, C. P., and Xia, M. (2010) The future of toxicity testing: a focus on in vitro methods using a quantitative high-throughput screening platform. *Drug Discovery Today 15*, 997−1007.
(29) Hassan, M., Brown, R. D., Varma-O'brien, S., and Rogers, D. (2006) Cheminformatics analysis and learning in a data pipelining environment. *Mol. Diversity 10*, 283−299.
(30) Zwierzyna, M., Vogt, M., Maggiora, G. M., and Bajorath, J. (2015) Design and characterization of chemical space networks for different compound data sets. *J. Comput.-Aided Mol. Des. 29*, 113−125.
(31) (2014) *Pipeline Pilot*, version 9.2, Biovia, San Diego, CA.
(32) (2015) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.
(33) Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res. 13*, 2498−2504.
(34) Xia, X., Maliski, E. G., Gallant, P., and Rogers, D. (2004) Classification of kinase inhibitors using a Bayesian model. *J. Med. Chem. 47*, 4463−4470.
(35) Clark, A. M., Dole, K., Coulon-Spektor, A., McNutt, A., Grass, G., Freundlich, J. S., Reynolds, R. C., and Ekins, S. (2015) Open Source Bayesian Models. 1. Application to ADME/Tox and Drug Discovery Datasets. *J. Chem. Inf. Model. 55*, 1231−1245.
(36) Ekins, S., Reynolds, R. C., Kim, H., Koo, M. S., Ekonomidis, M., Talaue, M., Paget, S. D., Woolhiser, L. K., Lenaerts, A. J., Bunin, B. A., Connell, N., and Freundlich, J. S. (2013) Bayesian models leveraging bioactivity and cytotoxicity information for drug discovery. *Chem. Biol. 20*, 370−378.
(37) Liu, L. L., Lu, J., Lu, Y., Zheng, M. Y., Luo, X. M., Zhu, W. L., Jiang, H. L., and Chen, K. X. (2014) Novel Bayesian classification models for predicting compounds blocking hERG potassium channels. *Acta Pharmacol. Sin. 35*, 1093−1102.
(38) Li, D., Chen, L., Li, Y., Tian, S., Sun, H., and Hou, T. (2014) ADMET evaluation in drug discovery. 13. Development of in silico prediction models for P-glycoprotein substrates. *Mol. Pharmaceutics 11*, 716−726.
(39) Tian, S., Wang, J., Li, Y., Xu, X., and Hou, T. (2012) Drug-likeness analysis of traditional Chinese medicines: prediction of drug-likeness using machine learning approaches. *Mol. Pharmaceutics 9*, 2875−2886.
(40) Wang, S., Li, Y., Wang, J., Chen, L., Zhang, L., Yu, H., and Hou, T. (2012) ADMET evaluation in drug discovery. 12. Development of binary classification models for prediction of hERG potassium channel blockage. *Mol. Pharmaceutics 9*, 996−1010.
(41) Chen, L., Li, Y., Zhao, Q., Peng, H., and Hou, T. (2011) ADME evaluation in drug discovery. 10. Predictions of P-glycoprotein inhibitors using recursive partitioning and naive Bayesian classification techniques. *Mol. Pharmaceutics 8*, 889−900.
(42) Singh, N., Chaudhury, S., Liu, R., AbdulHameed, M. D., Tawa, G., and Wallqvist, A. (2012) QSAR classification model for antibacterial compounds and its use in virtual screening. *J. Chem. Inf. Model. 52*, 2559−2569.
(43) (2015) *ToxCast & Tox21 Summary Files from invitrodb_v2*, U.S. EPA. http://www2.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data (accessed October 28, 2015).

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740

(44) Shah, F., and Greene, N. (2014) Analysis of Pfizer compounds in EPA's ToxCast chemicals-assay space. *Chem. Res. Toxicol. 27*, 86−98.
(45) Martin, M. T., Dix, D. J., Judson, R. S., Kavlock, R. J., Reif, D. M., Richard, A. M., Rotroff, D. M., Romanov, S., Medvedev, A., Poltoratskaya, N., Gambarian, M., Moeser, M., Makarov, S. S., and Houck, K. A. (2010) Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. *Chem. Res. Toxicol. 23*, 578−590.
(46) Benod, C., Subra, G., Nahoum, V., Mallavialle, A., Guichou, J. F., Milhau, J., Robles, S., Bourguet, W., Pascussi, J. M., Balaguer, P., and Chavanieu, A. (2008) N-1H-benzimidazol-5-ylbenzenesulfonamide derivatives as potent hPXR agonists. *Bioorg. Med. Chem. 16*, 3537−3549.
(47) Ganter, B., Snyder, R. D., Halbert, D. N., and Lee, M. D. (2006) Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix database. *Pharmacogenomics 7*, 1025−1044.
(48) *DrugMatrix*, National Institutes of Environmental Health Research. https://ntp.niehs.nih.gov/drugmatrix/index.html.
(49) AbdulHameed, M. D., Tawa, G. J., Kumar, K., Ippolito, D. L., Lewis, J. A., Stallings, J. D., and Wallqvist, A. (2014) Systems level analysis and identification of pathways and networks associated with liver fibrosis. *PLoS One 9*, e112193.
(50) Tawa, G. J., AbdulHameed, M. D., Yu, X., Kumar, K., Ippolito, D. L., Lewis, J. A., Stallings, J. D., and Wallqvist, A. (2014) Characterization of chemically induced liver injuries using gene co-expression modules. *PLoS One 9*, e107230.
(51) Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc. 4*, 1184−1191.
(52) Jones, S. A., Moore, L. B., Shenk, J. L., Wisely, G. B., Hamilton, G. A., McKee, D. D., Tomkinson, N. C., LeCluyse, E. L., Lambert, M. H., Willson, T. M., Kliewer, S. A., and Moore, J. T. (2000) The pregnane X receptor: a promiscuous xenobiotic receptor that has diverged during evolution. *Mol. Endocrinol. 14*, 27−39.
(53) Ng, H. W., Doughty, S. W., Luo, H., Ye, H., Ge, W., Tong, W., and Hong, H. (2015) Development and Validation of Decision Forest Model for Estrogen Receptor Binding Prediction of Chemicals Using Large Data Sets. *Chem. Res. Toxicol. 28*, 2343−2351.
(54) Lemaire, G., Mnif, W., Pascussi, J. M., Pillon, A., Rabenoelina, F., Fenet, H., Gomez, E., Casellas, C., Nicolas, J. C., Cavailles, V., Duchesne, M. J., and Balaguer, P. (2006) Identification of new human pregnane X receptor ligands among pesticides using a stable reporter cell system. *Toxicol. Sci. 91*, 501−509.
(55) Zhou, J., Febbraio, M., Wada, T., Zhai, Y., Kuruba, R., He, J., Lee, J. H., Khadem, S., Ren, S., Li, S., Silverstein, R. L., and Xie, W. (2008) Hepatic fatty acid transporter Cd36 is a common target of LXR, PXR, and PPARgamma in promoting steatosis. *Gastroenterology 134*, 556−567.
(56) Halpin, R. A., Ulm, E. H., Till, A. E., Kari, P. H., Vyas, K. P., Hunninghake, D. B., and Duggan, D. E. (1993) Biotransformation of lovastatin. V. Species differences in in vivo metabolite profiles of mouse, rat, dog, and human. *Drug Metab. Dispos. 21*, 1003−1011.
(57) Pascussi, J. M., Gerbal-Chaloin, S., Duret, C., Daujat-Chavanieu, M., Vilarem, M. J., and Maurel, P. (2008) The tangle of nuclear receptors that controls xenobiotic metabolism and transport: crosstalk and consequences. *Annu. Rev. Pharmacol. Toxicol. 48*, 1−32.

DOI: 10.1021/acs.chemrestox.6b00227
Chem. Res. Toxicol. 2016, 29, 1729−1740