Computational Toxicology 1 (2017) 39–48

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

QSAR development and profiling of 72,524 REACH substances for PXR activation and CYP3A4 induction

S.A. Rosenberg [a], M. Xia [b], R. Huang [b], N.G. Nikolov [a,1], E.B. Wedebye [a,1], M. Dybdahl [a,*,1]

[a] Division of Diet, Disease Prevention and Toxicology, National Food Institute, Technical University of Denmark, Mørkhøj Bygade 19, 2860 Søborg, Denmark
[b] National Center for Advancing Translational Sciences, National Institutes of Health, 9800 Medical Center Drive, Rockville, MD 20850, USA

ARTICLE INFO

Article history:
Received 21 November 2016
Received in revised form 16 January 2017
Accepted 19 January 2017
Available online 24 January 2017

Keywords:
PXR
CYP3A4
QSAR
REACH
Screening

ABSTRACT

The Pregnane X Receptor (PXR) is a key regulator of enzymes, for example the cytochrome P450 isoform 3A4 (CYP3A4), and transporters involved in the metabolism and excretion of xenobiotics and endogenous compounds. Activation of PXR by xenobiotics causes altered protein expression leading to enhanced or decreased turnover of both xenobiotics and endogenous compounds. This can potentially result in perturbations of normal physiology and adverse effects. Identification of PXR activating and CYP3A4 inducing compounds is included in drug-discovery programs but we still need similar information for the remaining tens-of-thousands of man-made compounds to which humans are potentially exposed. In the present study, we used high-throughput *in vitro* assay results for 2816 drugs to develop four quantitative structure-activity relationship (QSAR) models with binary outputs for binding to the human PXR ligand binding domain, full-length human and rat PXR activation and human CYP3A4 induction, respectively. Rigorous cross- and blinded external validations demonstrated four robust and highly predictive models with balanced accuracies ranging from 75.4% to 92.7%. The models were applied to screen 72,524 substances pre-registered under the EU chemicals regulation, REACH, and the models could predict 52.5% to 71.9% of the substances within their respective applicability domains. These predictions can, for example, be used for priority setting and in weight-of-evidence assessments of chemicals. Statistical analyses of the experimental drug dataset and the QSAR-predicted set of REACH substances were performed to identify similarities and differences in frequencies of overlapping positive results for PXR binding, PXR activation and CYP3A4 induction between the two datasets.

1. Introduction

The nuclear receptor (NR) superfamily is a large group of transcription factors that control expression of multiple genes involved in a broad range of biological processes, such as development, homeostasis and metabolism. The transcriptional activity of NRs is primarily regulated through ligand binding [1]. The Pregnane X Receptor (PXR), first described by Kliewer and colleagues in 1998, is a member of the NR superfamily [2,3]. PXR is mainly expressed in the liver, intestine and kidneys, and plays a key role in the regulation of genes involved in the metabolism and efflux of endogenous hormones and xenobiotic molecules [3–5]. The genes regulated by PXR include genes encoding enzymes, such as cytochrome P450s (CYPs), glucuronyltransferases and sulfotransferases, as well as transporters, such as P-glycoprotein and multidrug resistance proteins [2,3,6–8]. The ligand-binding domain (LBD) of PXR is large and flexible, and can change its shape to accommodate structurally diverse molecules including steroids, bile acids, antibiotics, statins, and pesticides [9,10]. A considerable amount of inter-species variation has been observed in the PXR LBD with human, rabbit and rat sharing roughly 75–80% amino acid identity [11,12]. There are numerous examples of differences in ligand binding to PXR and resulting downstream transcription

Abbreviations: AD, applicability domain; AOP, adverse outcome pathway; CYP, cytochrome P450; CYP3A4, human cytochrome P450 isoform 3A4; DTU, Technical University of Denmark; Food, National Food Institute; hPXR, full-length human Pregnane X Receptor; hPXR-LBD, human Pregnane X Receptor Ligand Binding Domain; IATA, Integrated Approaches to Testing and Assessment; LBD, Ligand Binding Domain; LPDM, Leadscope® Predictive Data Miner; NCATS, National Center for Advancing Translational Sciences; NIH, National Institute of Health; NR, nuclear receptor; PLR, partial logistic regression; PRS, Pre-Registered Substances; PXR, Pregnane X Receptor; QSAR, quantitative structure-activity relationship; *q*HTS, quantitative high-throughput screening; REACH, Registration, Evaluation, Authorisation & restriction of CHemicals; rPXR, full-length rat Pregnane X Receptor; RXRα, Retinoid X Receptor α; SD, standard deviation; TR-FRET, time-resolved fluorescence resonance energy transfer; XRE, Xenobiotic Response Element.

* Corresponding author.
E-mail address: mdyb@food.dtu.dk (M. Dybdahl).
[1] Contributed equally.

http://dx.doi.org/10.1016/j.comtox.2017.01.001
2468-1113/

of enzymes and transporters between species, which complicates the extrapolation of results from *in vivo* animal studies to humans [11,13–15].

PXR is located in the cytoplasm and translocated to the nucleus upon ligand binding, and here the PXR-ligand complex heterodimerizes with the Retinoid X Receptor alpha (RXRα), another member of the NR superfamily. The PXR-RXRα heterodimer complexes with co-activators, and this multi-protein complex binds to the Xenobiotic Response Element (XRE) in the promoter region of target genes and induces their transcription leading to altered expression of their encoded proteins [2,3,16]. Because many of the proteins regulated by PXR are not only involved in the metabolism and transport of xenobiotics, but also of various endogenous compounds such as steroid and thyroid hormones, an altered protein expression upon xenobiotic exposure may interfere with the homeostatic balance of such endogenous compounds [17,18]. This interference can potentially affect normal physiological functions [2,19] and may result in adverse health effects. Findings from previous studies indicate that there is an association between PXR activation by environmental chemicals and adverse health effects [15,18,20,21]. The importance of PXR activation is also reflected in a number of suggested adverse outcome pathways (AOPs) available from the online AOP-Wiki [22], for example an AOP describing how activation of PXR and other related NRs upregulate thyroid hormone catabolism resulting in hypothyroidism and subsequent adverse neurodevelopmental outcomes [23]. The AOPs are envisioned to promote the industry's and regulators' use of results from alternative methods such as *in vitro* tests and computational models in chemical risk assessments to reduce, refine or replace traditional animal tests [24–26], for example by applying the AOP in an Integrated Approaches to Testing Assessment (IATA) context to support regulatory decisions [27].

PXR is also known to be involved in drug-drug interactions in which an administered drug affects the metabolism and excretion of a co-administered drug, leading to decreased efficacy or increased toxicity [2,28,29]. For this reason, attenuation of PXR activity has become an important focus area in early drug-discovery programs [30]. Similar to drug-drug interactions, an altered expression of enzymes and transporters through PXR activation upon xenobiotic exposure may cause changes in the response to other xenobiotic compounds.

Among the many PXR target genes is the gene encoding CYP3A4, an oxidizing enzyme involved in phase I metabolism of various compounds [4,31]. CYP3A4 is considered the main drug-metabolizing CYP isoform in the human liver and is involved in the metabolism of more than 50% of drugs on the market [2,5]. In most cases, CYP3A4 causes chemicals to become less biologically active and promotes their excretion; but in other cases it has the opposite effect causing bioactivation by converting them to metabolites that are more toxic than the parent molecule [32].

Because xenobiotic activation of PXR has the potential to alter normal physiology and lead to adverse effects, it is of great importance to identify chemicals that may act through this mechanism. In a study from 2011, Shukla and colleagues used four high-throughput *in vitro* assays to profile more than 2800 clinically-used and investigational drugs for their ability to bind to the human PXR-LBD, activate full-length human and rat PXR, and induce human CYP3A4 [14]. Chemicals in the ToxCast program [33], which include both drugs and environmental chemicals, have also been tested for these mechanisms in related assays [34]. However, we still need similar information for the remaining tens-of-thousands of xenobiotics to which humans are potentially exposed [35,36].

In the present study, we used the high-throughput *in vitro* data from Shukla et al. [14] to train and validate four Quantitative Structure-Activity Relationship (QSAR) models for human PXR-LBD binding, human and rat PXR activation, and human CYP3A4 induction, respectively. QSAR models are computational models that relate chemical structures to, e.g., a biological activity, and they can be used to predict the activity of an untested chemical based on its chemical structure (an introduction to QSAR can e.g. be found in [37,38]). In general, QSARs are rapid and cost-effective tools for predicting biological activities of chemical structures and can be used for virtual screening of single substances as well as large chemical inventories. The four developed models were applied to screen a structurally diverse library of 72,524 chemicals from the EU chemicals regulation REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) list of Pre-Registered Substances (PRS) [39,40], containing substances potentially present in our food, the environment and consumer products. These QSAR predictions can, e.g., be used, possibly together with other relevant data, 1) to identify and prioritize chemical substances for further testing and 2) in an IATA context, together with relevant AOP(s), to guide further testing and regulatory decisions in chemical risk assessments [25,27,41]. Furthermore, statistical analyses of the experimental drug dataset and the QSAR-predicted REACH PRS set were performed in order to elucidate similarities and differences in co-occurrences of overlapping positive results for PXR binding, PXR activation and CYP3A4 induction between the two chemical universes.

2. Materials and methods

2.1. Experimental datasets

We used four datasets containing chemical structure information and *in vitro* experimental data for a collection of 2816 clinically-used and investigational drugs to train and validate the QSAR models. The experimental data of the 2816 compounds included results from quantitative high-throughput screening (*q*HTS) for binding to the LBD of human PXR at the protein level (hPXR-LBD); activation of full-length human PXR (hPXR) and full-length rat PXR (rPXR) at the cellular level; and induction of human CYP3A4 at the cellular level (CYP3A4). All experimental data were generated by the National Center for Advancing Translational Sciences (NCATS) at the National Institute of Health (NIH). The compound collection, *q*HTS assays, and the classification of the *q*HTS results into actives, inconclusives and inactives have been described previously [14,42,43]. Briefly, actives showed binding to the hPXR-LBD, activation of hPXR and/or rPXR and/or induced transcription of CYP3A4 according to the applied assays. Inactives did not show activity in the given assay, and inconclusives showed equivocal activity results in the assays. Only the substances in each dataset classified as either active or inactive were used, i.e. substances with inconclusive experimental results were excluded. The experimental results for about one third of the substances in each of the four main datasets were masked by NIH NCATS and these compounds were used as external test sets for blinded external validations after the model development was finished. The selection of the test sets was designed and made by NIH NCATS scientists, who clustered all compounds in the dataset on structural similarity using the Euclidian distance and then, within each structure cluster and for each of the four endpoints, approximately one-third actives and one-third inactives were selected randomly. Thus the training and test sets are structurally comparable and have similar distributions of actives and inactives. NIH NCATS sent the training sets containing structure information and experimental results and the test sets containing only structure information to the National Food Institute (Food) at the Technical University of Denmark (DTU), who performed the structure preparations, the