Drug Metabolism and Pharmacokinetics xxx (2016) 1–7

Contents lists available at ScienceDirect

Drug Metabolism and Pharmacokinetics

journal homepage: http://www.journals.elsevier.com/drug-metabolism-and-pharmacokinetics

Regular Article

Development of pharmacophore-based classification model for activators of constitutive androstane receptor

Kyungro Lee [a], Hwan You [a], Jiwon Choi [b], Kyoung Tai No [a, b, *]

[a] Department of Biotechnology, College of Life Science and Biotechnology, Yonsei University, Seoul 03722, South Korea
[b] Bioinformatics & Molecular Design Research Center, Yonsei University, Seoul 03722, South Korea

ARTICLE INFO

Article history:
Received 5 August 2016
Received in revised form
21 September 2016
Accepted 10 November 2016
Available online xxx

Keywords:
Constitutive androstane receptor
Pharmacophore-based descriptor
Classification
Machine learning
Support vector machine

ABSTRACT

Constitutive androstane receptor (CAR) is predominantly expressed in the liver and is important for regulating drug metabolism and transport. Despite its biological importance, there have been few attempts to develop *in silico* models to predict the activity of CAR modulated by chemical compounds. The number of *in silico* studies of CAR may be limited because of CAR's constitutive activity under normal conditions, which makes it difficult to elucidate the key structural features of the interaction between CAR and its ligands. In this study, to address these limitations, we introduced 3D pharmacophore-based descriptors with an integrated ligand and structure-based pharmacophore features, which represent the receptor-ligand interaction. Machine learning methods (support vector machine and artificial neural network) were applied to develop an *in silico* model with the descriptors containing significant information regarding the ligand binding positions. The best classification model built with a solvent accessibility volume-based filter and the support vector machine showed good predictabilities of 87%, and 85.4% for the training set and validation set, respectively. This demonstrates that our model can be used to accurately predict CAR activators and offers structural information regarding ligand/protein interactions.

1. Introduction

The nuclear constitutive active/androstane receptor (CAR, NR1I3) is a key member of the nuclear receptor (NR) transcription factors, which regulate metabolic homeostasis when activated by their ligands. CAR serves as a xenobiotic sensing receptor that organizes the cellular defense system against endogenous and exogenous challenges by controlling the expression of genes encoding phase I oxidation enzymes (e.g., cytochrome P450s), phase II conjugation enzymes (e.g., UDP-glucuronosyltransferases), and phase III efflux transporters (e.g., multidrug resistance proteins) [1]. Drugs that bind to the CAR in hepatocytes can disturb the liver metabolic system through pharmacokinetic drug–drug interactions. CAR is also associated with several hepatic functions, including bilirubin metabolism, fatty acid oxidation, bile acid homeostasis, hormone homeostasis, gluconeogenesis regulation, and cell apoptosis and proliferation [2].

NRs are structurally conserved and are composed of three domains: a highly variable N-terminal DNA binding domain, hinge domain, and ligand binding domain (LBD). In general, the LBD of NRs is enclosed by 12 α–helices, where α12 containing the activation function (AF) domain is a crucial region for activation [3–5]. Despite the structural similarity of CAR with other NRs, CAR is constitutively activated in the absence of ligand binding [6]. Helices αX and αAF of CAR mediate the ligand-independent interaction with coactivators and confer constitutive activity [7,8]. Furthermore, the positive K195 residue in helix α5 preferentially interacts with the negatively charged carboxy-terminus and constructs the active conformation of αAF, which has relevance to the shortened loop between αX and αAF in addition to a shortened αAF compared to other NRs [9,10]. Additionally, a mutagenesis study showed that several amino acids within α3 (Asn165), α5 (Val199), α10 (Tyr326, Ile330, and Gln331), and αAF (Leu343 and Ile346) contributed to the constitutive activity of CAR and some residues within α3 (Ile164 and Asn165), α5 (Cys202 and His203), and α7 (Phe234 and Phe238) affected the selectivity of chemicals for CAR activation [11].

The activity of CAR is commonly determined using a cell-based reporter gene assay, which measures the expression of a reporter in

* Corresponding author.
E-mail addresses: ktno@bmdrc.org, ktno@yonsei.ac.kr (K.T. No).

http://dx.doi.org/10.1016/j.dmpk.2016.11.005
1347-4367/

the cell co-transfected with CAR and a reporter gene plasmid. The reporter assay has identified many compounds as agonists/inverse agonists of CAR, but the assay results do not demonstrate whether these compounds directly bind to the ligand-binding site of CAR. For example, phenobarbital indirectly augments the expression of the reporter genes by regulating epidermal growth factor receptor signaling, which is crucial for regulating the dephosphorylation of CAR at Thr38 and nuclear translocation of CAR [12,13]. As CAR is accumulated in the nucleus, up-regulation of CAR activity induces the transcription of drug metabolism and transporter-related genes. In addition to phenobarbital, phenytoin, triclocarban, galangin, chrysin, and baicalein are known to indirectly activate CAR [14–16]. Using the reporter assay, it is impossible to select the binders interacting with CAR-LBD among the activators of CAR. Two-hybrid assays and fluorescence resonance energy transfer assays were recently developed to verify the direct interaction between CAR and its ligands [16–18]. Since the activators in these assays are determined based on their function of dissociating the coactivator protein, TIF2 or PGC1, from CAR, further studies are needed to confirm the ligand binding site or a different activation site, particularly at the binding interface between CAR and coactivator proteins. Furthermore, a limited number of scaffolds, such as polychlorinated biphenyls and phthalates, have been tested in these assays, and the number of validated compounds is too small to be introduced in training data for developing machine-learning models. We collected data for chemicals regulating the activation of CAR, which include the results from yeast two-hybrid assays, fluorescence resonance energy transfer assays, and reporter gene assays, excluding known indirect activators and their analogs.

Machine-learning is a data-driven decision or prediction model-building method that is widely used in computer-aided molecular modeling. Diverse machine-learning methods have been used to infer the information of structure-based receptor-ligand binding. Although there are several machine-learning-based prediction models for the sister xenobiotic receptor pregnane X receptor (NR1I2) [19–21], only one machine-learning model for predicting the activity of CAR with its ligands has been developed and this model indicated the critical residues involved in CAR ligand binding [22]. Because CAR is constitutively active under normal conditions, it is difficult to determine whether the activation of CAR is caused by ligand binding.

In this study, we developed a general classification model derived from a support vector machine (SVM) and a solvent accessibility volume (SAVol)-based filter. Through validation with a predefined external validation set, the model was evaluated for its ability to classify compounds as activators/non-activators of CAR. The purpose of this study was to develop a prediction model and identify critical structural interactions between CAR and its ligands that highly contribute to binding affinity.

2. Materials and methods

To identify the activators of CAR and infer structural information regarding the interaction between CAR and its ligands, pharmacophore-based descriptors were introduced to represent the binding features. First, reliable binding poses of CAR ligands were selected through docking calculations. The poses of the ligands were used to generate a ligand-based pharmacophore. A receptor-based pharmacophore representing possible interactions between ligands and CAR was built in the ligand-binding site. The receptor and ligand-based pharmacophores were used to calculate descriptors representing the crucial interactions between a ligand and CAR. Machine-learning methods combined with a genetic algorithm (GA) were used to develop the binary classification model. The overall modeling procedure is depicted in Fig. 1.

2.1. Data set

A set of 548 compounds used in this study were collected from previous studies [17,18,22–55]. The compounds in the data set were classified in several ways as shown in Fig. 2. The compounds were classified as i) activators or non-activators and ii) binders or non-binders of CAR. By applying the biological and biophysical criteria (transcriptional activation and binding), the compounds were clustered into four groups, i) binders and activators, ii) binders and non-activators, iii) non-binders and activators, and iv) non-

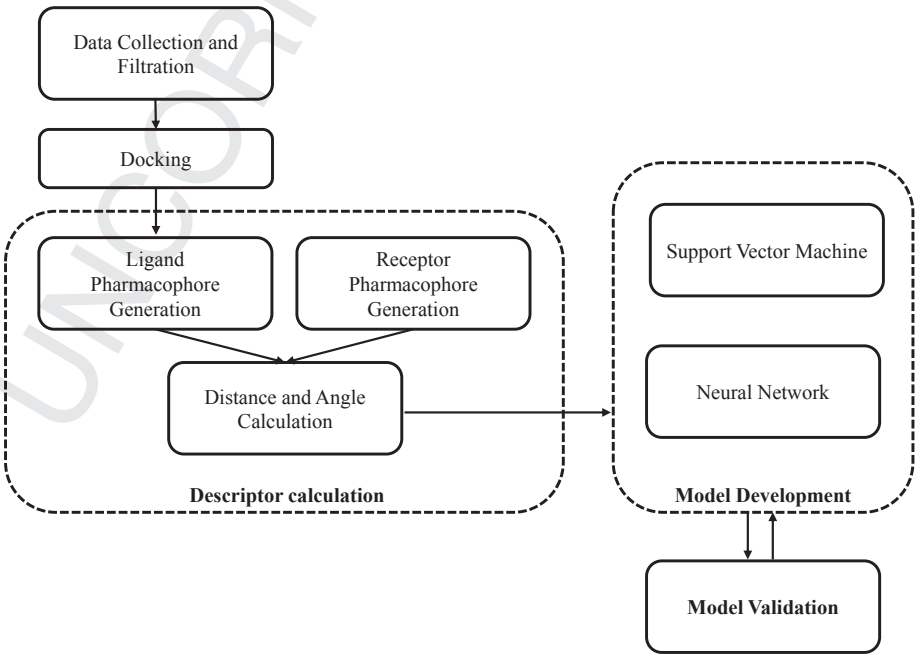

Fig. 1. Flowchart of pharmacophore-based classification model development in the present study.

Please cite this article in press as: Lee K, et al., Development of pharmacophore-based classification model for activators of constitutive androstane receptor, Drug Metabolism and Pharmacokinetics (2016), http://dx.doi.org/10.1016/j.dmpk.2016.11.005