Journal of Pharmaceutical Sciences xxx (2016) 1-9

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Pharmacokinetics, Pharmacodynamics and Drug Transport and Metabolism

Development of a Support Vector Machine-Based System to Predict Whether a Compound Is a Substrate of a Given Drug Transporter Using Its Chemical Structure

Atsushi Ose [1], Kota Toshimoto [2, 3], Kazushi Ikeda [2], Kazuya Maeda [4], Shuya Yoshida [5], Fumiyoshi Yamashita [5], Mitsuru Hashida [5], Takashi Ishida [2], Yutaka Akiyama [2], Yuichi Sugiyama [3, *]

[1] *Development Planning, Clinical Development Center, Asahi Kasei Pharma Corporation, 1-105 Kanda Jinbocho, Chiyoda-ku, Tokyo 101-8101, Japan*
[2] *Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan*
[3] *Sugiyama Laboratory, RIKEN Innovation Center, RIKEN Cluster for Industry Partnerships, RIKEN, 1-6, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan*
[4] *Laboratory of Molecular Pharmacokinetics, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[5] *Department of Drug Delivery Research, Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshidashimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan*

ARTICLE INFO

Article history:
Received 28 October 2015
Revised 31 March 2016
Accepted 22 April 2016

Keywords:
in silico modeling
QSAR
transporters
computational ADME
high-throughput technologies

ABSTRACT

The aim of this study was to develop an *in silico* prediction system to assess which of 7 categories of drug transporters (organic anion transporting polypeptide [OATP] 1B1/1B3, multidrug resistance-associated protein [MRP] 2/3/4, organic anion transporter [OAT] 1, OAT3, organic cation transporter [OCT] 1/2/ multidrug and toxin extrusion [MATE] 1/2-K, multidrug resistance protein 1 [MDR1], and breast cancer resistance protein [BCRP]) can recognize compounds as substrates using its chemical structure alone. We compiled an internal data set consisting of 260 compounds that are substrates for at least 1 of the 7 categories of drug transporters. Four physicochemical parameters (charge, molecular weight, lipophilicity, and plasma unbound fraction) of each compound were used as the basic descriptors. Furthermore, a greedy algorithm was used to select 3 additional physicochemical descriptors from 731 available descriptors. In addition, transporter nonsubstrates tend not to be in the public domain; we, thus, tried to compile an expert-curated data set of putative nonsubstrates for each transporter using personal opinions of 11 researchers in the field of drug transporters. The best prediction was finally achieved by a support vector machine based on 4 basic and 3 additional descriptors. The model correctly judged that 364 of 412 compounds (internal data set) and 111 of 136 compounds (external data set) were substrates, indicating that this model performs well enough to predict the specificity of transporter substrates.

Introduction

When clarifying the pharmacokinetic properties of drugs, it is widely known that the functions of drug transporters must be taken into consideration and also the functions of metabolic enzymes. Drug transporters are classified into ATP-binding cassette and solute carrier families of transporters on the basis of their structural and functional characteristics. Drug transporters are expressed in various important tissues dominating drug disposition and elimination, such as epithelial cells of the intestine and kidney, hepatocytes, and brain capillary endothelial cells.[1-9] The accumulated evidence has demonstrated that drug transporters play

Abbreviations used: AUC, area under the curve; BCRP, breast cancer resistance protein; DDI, drug–drug interaction; f_{up}, protein unbound fraction in plasma; ITC, International Transporter Consortium; log D, octanol-water distribution coefficient; MATE, multidrug and toxin extrusion; MDR1/P-gp, multidrug resistance protein 1/ P-glycoprotein; MRP, multidrug resistance-associated protein; OATP, organic anion transporting polypeptide; OAT, organic anion transporters; OCT, organic cation transporter; ROC, receiver operating characteristic; SVM, support vector machine.

The authors Atsushi Ose and Kota Toshimoto contributed equally.

This article contains supplementary material available from the authors by request or via the Internet at http://dx.doi.org/10.1016/j.xphs.2016.04.023.

* *Correspondence to:* Yuichi Sugiyama (Telephone: +81-45-503-9211; Fax: +81-45-503-9190).

E-mail address: ychi.sugiyama@riken.jp (Y. Sugiyama).

http://dx.doi.org/10.1016/j.xphs.2016.04.023
0022-3549/

pivotal roles in the pharmacokinetics of substrate drugs including as active barriers to drug absorption, in efficient elimination of drugs from the liver and kidney, and in regulation of exposure of pharmacological and toxicological targets to various drugs.

The so-called "transporter white paper"[10] authored by the International Transporter Consortium (ITC) and the regulatory (draft) guidance and guidelines[11-13] emphasize the need to evaluate the risk of transporter-mediated drug–drug interactions (DDIs) for the following drug transporters: multidrug resistance protein 1 (MDR1), breast cancer resistance protein (BCRP), organic anion transporting polypeptide (OATP) 1B1 and 1B3, organic anion transporter (OAT) 1 and 3, organic cation transporter (OCT) 2, and multidrug and toxin extrusion (MATE) 1 and 2-K, in the process of drug development. Furthermore, the ITC[14] and regulatory authorities[15,16] have proposed that the effects of genetic polymorphisms in OATP1B1 and BCRP on the pharmacokinetics of their substrate drugs should be investigated in a phase I study. In such situations, if we can predict whether a compound acts as a substrate for a given drug transporter based on its chemical structure alone, we can anticipate not only the possible involvement of transporters in the drug pharmacokinetics but also the necessity of clinical DDIs and pharmacogenetic studies at the discovery stage of drug development without any *in vitro* data for transporters. However, because the substrate specificities of drug transporters are in general very broad, it is difficult to intuitively predict whether a compound is a substrate for each transporter. Thus, methods for predicting recognition of substrates by clinically important drug transporters have been desired.

In recent years, there have been various studies of *in silico* models for the prediction of protein–ligand interactions that include transporter–ligand interactions. *In silico* methods for the prediction of protein–ligand interactions can be divided into ligand-based and protein structure–based approaches. The transporter most investigated for predicting the substrates and/or inhibitors is P-gp, and investigators have reported the prediction of P-gp–ligand interaction using a ligand-based approach.[17,18] Some have predicted the substrates and/or inhibitors of P-gp and BCRP with a support vector machine (SVM).[19-21] Xue et al.[21] trained an SVM using a dataset of 116 substrates and 85 nonsubstrates of P-gp; using 5-fold cross-validation, the SVM had a prediction accuracy of 81% for substrates and 79% for nonsubstrates. In 2011, Wang et al.[20] compiled a large dataset of 206 substrates and 126 nonsubstrates of P-gp. Using the internal data set (212 compounds), they developed an SVM that possessed a prediction accuracy of 0.88 on the test data set (120 compounds). Recently, Hazai et al.[19] developed an SVM prediction model for BCRP substrates based on 164 substrates and 99 nonsubstrates. Their model had a prediction accuracy of 73% on an independent external data set. In contrast, there are few reports on the prediction of substrates for multiple transporters. Sedykh et al.[22] constructed three kinds of quantitative structure–activity relationship models, including an SVM, to predict the substrates and inhibitors of the intestinal transporters.

Over the past decade, many efforts have been made to develop an *in silico* system to predict the substrates of drug transporters especially listed in the ITC white paper and the regulatory guidance and guidelines on DDIs. However, most of these have focused on just 1 transporter (P-gp), and no *in silico* system is currently able to predict the substrates for multiple transporters. We previously established an *in silico* prediction system for major drug clearance pathways using machine learning approaches.[23,24] Our system has high predictive performance to classify 5 drug clearance pathways (metabolism by CYP3A4, CYP2C9, or CYP2D6; OATP-mediated hepatic uptake; or renal excretion). Therefore, we considered that similar strategies might be applied to the prediction of substrate recognition by multiple transporters. However, one of the potential problems for the prediction of transporter substrates is that while isoforms of CYP enzymes, which can or cannot recognize a drug as a substrate, have often been extensively characterized, there is little information about transporter nonsubstrates. Thus, we tried to create an expert concensus-based data set of nonsubstrates for each category of drug transporters by "expert opinion" to improve the prediction performance of the *in silico* system.

The aim of this study was to develop an *in silico* system to predict which of the 7 categories of drug transporters (OATP1B1/1B3, multidrug resistance-associated protein [MRP] 2/3/4, OAT1, OAT3, OCT1/2/MATE1/2-K, MDR1, and BCRP) can recognize a substrate from the chemical structure of the compound alone. Following our previous studies,[23,24] we initially tried to apply 2 prediction methods: (1) a rectangular method, which is visually intuitive and easy to understand; and (2) an SVM, which has a broad capacity for classification.

Methods

Data Set

The target transporters (those for which we wished to predict substrate specificity) were divided into 7 categories: OATP1B1/1B3, MRP2/3/4, OAT1, OAT3, OCT1/2/MATE1/2-K, MDR1, and BCRP. In some cases, multiple transporter isoforms were included in the same category because the literature suggested their substrate specificities to be very similar.

Internal Data Set

Information regarding the substrates for the 7 categories of transporters was collected through a TP search (http://togodb.dbcls.jp/tpsearch), and additional information was gathered using PubMed. Additional information about OATP1B1/1B3 substrates was collected from the PubMed database by text-mining technique developed by Yoshida et al.[25] Substrate information was collected through literature published until 2010.

External Data Set

Using PubMed, information regarding the substrates for the 7 categories of transporters was collected through literature published in 2011. Additional information about OAT1, OAT3, OCTs/MATEs, and BCRP was collected from the PubMed database (published until 2010) by text-mining technique[25] and that about MRP2/3/4 and MDR1 was taken from FUJITSU ADME database (Fujitsu Kyushu Systems Engineering Ltd., Fukuoka, Japan). Furthermore, among new chemical entities approved by the Pharmaceutical and Medical Devices Agency from 2013 to 2015, substrate information was taken from DIDB (http://www.druginteractioninfo.org/), PharmGKB (www.pharmgkb.org), and TransPortal (http://dbts.ucsf.edu/fdatransportal).

Internal and External Data Set

The following information regarding the physicochemical properties was then collected. (Note that compounds for which we had no physicochemical information or zwitterionic information were excluded from the data set.)

Physicochemical Descriptors

The 2-dimensional structures of the collected compounds were obtained from the PubChem Compound Database (http://pubchem.ncbi.nlm.nih.gov/). Their charge, molecular weight, and logD at pH 7.0 (lipophilicity) were obtained using SciFinder Scholar 2007 (Chemical Abstracts Service, Columbus, OH). Charge was defined based on the value of p*K*a. In addition, the protein-unbound