Food and Chemical Toxicology ■■ (2015) ■■–■■

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

Developing a QSAR model for hepatotoxicity screening of the active compounds in traditional Chinese medicines

Shan-Han Huang [a], Chun-Wei Tung [a,b], Ferenc Fülöp [c], Jih-Heng Li [a,b,*]

[a] Ph.D. Program in Toxicology and School of Pharmacy, College of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan
[b] National Environmental Health Research Center, National Health Research Institutes, Taiwan
[c] Institute of Pharmaceutical Chemistry, University of Szeged, Hungary

ARTICLE INFO

Article history:
Received 18 August 2014
Accepted 16 January 2015
Available online

Keywords:
Traditional Chinese medicine
QSAR
Drug-induced liver injury (DILI)
Hepatotoxicity

ABSTRACT

The perception that natural substances are deemed safe has made traditional Chinese medicine (TCM) popular in the treatment and prevention of disease globally. However, such an assumption is often misleading owing to a lack of scientific validation. To assess the safety of TCM, *in silico* screening provides major advantages over the classical laboratory approaches in terms of resource- and time-saving and full reproducibility. To screen the hepatotoxicity of the active compounds of TCMs, a quantitative structure–activity relationship (QSAR) model was firstly established by utilizing drugs from the Liver Toxicity Knowledge Base. These drugs were annotated with drug-induced liver injury information obtained from clinical trials and post-marketing surveillance. The performance of the model after nested 10-fold cross-validation was 79.1%, 91.2%, 53.8% for accuracy, sensitivity, and specificity, respectively. The external validation of 91 well-known ingredients of common herbal medicines yielded a high accuracy (87%). After screening the TCM Database@Taiwan, the world's largest TCM database, a total of 6853 (74.8%) ingredients were predicted to have hepatotoxic potential. The one-hundred chemical ingredients predicted to have the highest hepatotoxic potential by our model were further verified by published literatures. Our study indicated that this model can serve as a complementary tool to evaluate the safety of TCM.

1. Introduction

Traditional Chinese medicines (TCMs) are widely used in the ethnic Chinese population. In Taiwan, the trend in the utilization of TCMs has been observed with a mean increment of 1,671,476 of new users yearly under the National Health Insurance Research Database from 1996 to 2001 (Chen et al., 2007). Another study showed the crude utilization of TCMs increased from 36,372 in 1997 to 41,823 in 2003 (Chang et al., 2008). Recently, it has also been found that more Asian and non-Asian consumers and patients are using TCMs in the United States (Ko, 2004). The increasing popularity of TCMs or herbal medicines in Western society is based on the perceived effectiveness of TCMs in the treatment and prevention of disease (Chang et al., 2008) or a belief that these medications are natural and therefore safe for better control of a disease and its management (Stickel et al., 2005). While TCM is a common remedy to treat major diseases of the respiratory system (22.1%), musculoskeletal system and connective tissue (18.1%) in the Chinese population (Chang et al., 2008), it is regarded as a complementary and alternative medicine in Western society. TCM is regulated as a dietary supplement under the Dietary Supplement Health and Education Act of 1994 in the United States, natural health products under the Natural Health Products Regulations in 2004 in Canada, therapeutics goods under the Therapeutic Goods Acts 1989 in Australia, and "regular" medicinal products under the Traditional Herbal Registration in 2005 in the European Union (Jordan et al., 2010). In other words, TCM is only loosely regulated in comparison with other drugs. However, TCMs may possess adverse reactions, as all drugs do.

In fact, TCMs have been reported to cause adverse reactions due to manufacturing/quality problems (adulteration, heavy metal content, improper processing/preparation, and substitution/misidentification) or active/toxic ingredients (inherent toxicity, overdose toxicity, idiosyncratic reactions, and drug–herb interactions) (Ko, 2004). Furthermore, there may be a causal relationship between certain TCM ingredients and specific organ toxicity (Chen et al., 2011b; Jordan et al., 2010; Ko, 2004), but the safety assessment of TCM, which is costly and time-consuming, has not been systematically investigated (Kruger and Mann, 2003).

Among the TCM-induced toxic effects, hepatotoxicity is one of the major concerns (Kane et al., 1995; Teschke, 2014; Wang et al., 2014). TCM-induced liver injuries can result from direct, dose-dependent hepatotoxicity or idiosyncratic reactions, and a number of them are associated with serious hepatotoxic events such as acute

* Corresponding author. Ph.D. Program in Toxicology and School of Pharmacy, College of Pharmacy, Kaohsiung Medical University, Kaohsiung, Taiwan. Tel.: +886 7 3121101 ex 2651; fax: +886 7 3210683.
E-mail address: jhlitox@kmu.edu.tw (J.-H. Li).

http://dx.doi.org/10.1016/j.fct.2015.01.020
0278-6915/

Please cite this article in press as: Shan-Han Huang, Chun-Wei Tung, Ferenc Fülöp, Jih-Heng Li, Developing a QSAR model for hepatotoxicity screening of the active compounds in traditional Chinese medicines, Food and Chemical Toxicology (2015), doi: 10.1016/j.fct.2015.01.020

liver failure. The liver plays a pivotal role in intermediary metabolism, energy exchanges, and the biotransformation of xenobiotics. Hepatotoxicity is an important cause of failure in both the clinical and post-approval stages of drug development and poses a substantial challenge for the pharmaceutical industry (Cheng, 2009). The severity of drug-induced liver injury can range from steatosis to fatal liver failure (Lee, 2003). However, the mechanism of TCM-induced liver injuries remains to be elucidated and the risk factors of hepatotoxicity are not well-defined (Licata et al., 2013; Murray et al., 2008; Pittler and Ernst, 2003; Stickel et al., 2005). Even with the gold standard testing methods such as regulatory animal toxicity studies, the sensitivity of hepatotoxicity prediction is only 52% (O'Brien et al., 2006). It is an unmet need to identify these toxic reactions early and efficiently in the drug development process.

In 2011 the U.S. Food and Drug Administration (FDA) published guidelines for assessing the potential of a drug to cause severe drug-induced liver injury (DILI) in premarketing clinical evaluation (Chen et al., 2011a). The representative methods include, but are not limited to, quantitative structure–activity relationship (QSAR) assessments, *in vitro* assays, high-content screening (HCS) assays and 'omics' studies (Chen et al., 2011a). Traditionally, *in vitro* testing strategies have been developed to predict clinical outcomes of DILI with a multiparametric analysis of xenobiotic toxicity, such as cell viability, nuclear morphology, mitochondrial function, intracellular calcium concentration, and oxidative stress at the single-cell level using a HCS system during the preclinical stages (Tolosa et al., 2012; Xu et al., 2008). These studies aim to correlate the drug-induced toxicity with the underlying mechanism(s), and have a moderate sensitivity and a high specificity (Persson et al., 2013; Xu et al., 2008). However, a battery of assays is inevitably needed, which is time- and resource-consuming, to determine the toxicity of each individual new compound.

Computational modeling, with the advantages of full reproducibility, saving of time, money and energy, and reduction of animal sacrifices, is useful to evaluate the efficacy, metabolism, and "general toxicity" of active pharmaceutical ingredients (API) and is widely used in many pharmaceutical companies (Gibb, 2008). *In silico* methods have been successfully used as guides for assessment of mutagenicity and skin sensitization (Aptula et al., 2005; Custer and Sweder, 2008; Dimitrov et al., 2005; Nandy et al., 2013; Roberts et al., 2006; Valencia et al., 2013; Valerio and Cross, 2012). However, there are still no standard QSAR models to assess DILI. Greene et al. identified structure–activity relationships for chemicals that have the potential to cause hepatotoxicity. Four classes of hepatotoxicity were assigned, namely, no evidence, weak evidence, animal hepatotoxicity and human hepatotoxicity (the concordance, sensitivity and specificity were 56%, 46% and 73%, respectively) (Greene et al., 2010). It would be a challenge for this system to modulate the balance between sensitivity (picking out all the positive compounds) and specificity (identifying all of the negative compounds). The assessment of chemicals using *in silico* methods such as QSAR based on *in vitro* cytotoxicity or rodent data cannot be readily extrapolated to human hepatotoxicity (Jordan et al., 2010; O'Brien et al., 2006). Rodgers et al. developed a QSAR model from the dataset of liver adverse effects of drugs (AEDs) that were identified from an FDA spontaneous reporting database of human liver AEDs (the sensitivity and specificity were 73.7% and 94.4%, respectively) (Rodgers et al., 2010). The data of this composite liver endpoint model were derived from several biomarkers. However, alanine aminotransferase (ALT) levels could be elevated due to other indirect factors; aspartate aminotransferase (AST) activity is also known to fluctuate throughout the day and increases with exercise.

Despite the challenges in the extrapolation of a computational model to practical use of TCM, there are many advantages of such predictive approaches, especially for filling data gaps in relation to TCM safety. According to a recent report from the National Academy of Sciences, Toxicity Testing in the 21st Century (Gibb, 2008), an integrated approach has been developed for testing and assessment of potential toxicants to human health. In developing its vision for toxicity testing, the committee considered some options for toxicity testing, including minimizing the use of animals (Krewski et al., 2009). *In silico* screening could also play a role in predicting the likelihood that a particular compound will cause adverse health outcomes in humans (Krewski et al., 2009). To date, QSAR has become a common computational method to predict apical endpoints for regulatory applications. However, there is still no standard computational method to evaluate the safety of active ingredients of traditional Chinese medicines.

In this study, we aimed to establish a QSAR model based on the Liver Toxicity Knowledge Base (LTKB) (Chen et al., 2011a). Furthermore, we attempted to confirm whether identified ingredients of TCMs/common herbal drugs are associated with liver damage according to literature findings. Finally, we applied the QSAR model to predict the hepatotoxic potential of active compounds in the Traditional Chinese Medicine Database@Taiwan (Chen, 2011).

2. Materials and methods

2.1. Dataset

LTKB is a benchmark dataset that was developed by scientists at the National Center for Toxicological Research (NCTR), U.S. FDA. This dataset contains drugs that were reported to cause DILI in humans according to the FDA-approved prescription drug labels (Chen et al., 2011a). The LTKB has been used in previous studies that can be summarized as compounds of high lipophilicity, high daily dose and being a substrate of cytochrome P450 enzymes are associated with drug-induced liver injury (Chen et al., 2013a; Yu et al., 2014). Based on labeling description and severity levels, these drugs were divided into three categories, namely, most-DILI-concern, less-DILI-concern, and no-DILI-concern. The drugs were classified into the most-DILI-concern section if they were either withdrawn or assigned a black box warning for hepatotoxicity. They were labeled with a greater than moderate DILI severity if they caused fatal hepatotoxicity, acute liver failure, liver necrosis, jaundice, and hyperbilirubinemia in the warnings and precautions section. The drugs of no-DILI-concern had no DILI description mentioned on the labels (Chen et al., 2013a). The 2D structures of the drugs that were classified into the most-DILI-concern section (positive dataset, $n = 136$) and those of no-DILI-concern (negative dataset, $n = 65$) were downloaded from PubChem. Gemtuzumab was excluded because it has no 2D structure data in PubChem. PubChem is an open repository for experimental data and provides a significant, publicly accessible platform for mining the biological information of small molecules (Bolton et al., 2008). All chemical structures downloaded from the related websites were manually double-checked for the correctness and consistency of their molecular and structural representations, salts and charged groups.

2.2. Chemical descriptor

PaDEL-Descriptor is a freely available software package for calculating molecular descriptors and fingerprints, which are the final results transformed from chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment. PaDEL-Descriptor includes 863 descriptors (729 1D, 2D descriptors and 134 3D descriptors) and 10 types of fingerprint. The descriptors and fingerprints are calculated using the Chemistry Development Kit with some additional descriptors and fingerprints (Steinbeck et al., 2006). Finally, PaDEL afforded 1603 descriptors covering a wide variety of types containing 0D, 1D, 2D descriptors, and fingerprints, and the data applied to develop the QSAR model using WEKA software (Waikato Environment for Knowledge Analysis) (Hall et al., 2009).

2.3. Feature selection

For most QSAR modeling tools, irrelevant descriptors could significantly affect the performance if they are not removed prior to training. We used the WEKA filter named Remove Useless for feature selection. This filter can remove descriptors that do not vary at all or that vary too much, and the maximum variance percentage we allowed was 95%. After removing 541 irrelevant descriptors, the feature selection results with 1062 descriptors were used for the following analysis.

2.4. Random forests

The Random Forest (RF) algorithm, based on a large ensemble of decision trees, is an extensively used ensemble learning method (Breiman et al., 1984). The advantages of the RF classifier include avoiding overfitting problems, which is especially important for analyzing a small dataset (Amaratunga et al., 2008; Lin et al., 2004).

Please cite this article in press as: Shan-Han Huang, Chun-Wei Tung, Ferenc Fülöp, Jih-Heng Li, Developing a QSAR model for hepatotoxicity screening of the active compounds in traditional Chinese medicines, Food and Chemical Toxicology (2015), doi: 10.1016/j.fct.2015.01.020