Fundamental Toxicological Sciences (Fundam. Toxicol. Sci.)
Vol.6, No.8, 327-332, 2019

The Japanese Society of Toxicology

Fundamental Toxicological Sciences
URL : http://www.fundtoxicolsci.org/index_e.html

Original Article

Development of a hepatotoxicity prediction model using *in vitro* assay data of key molecular events

Koji Jojima[1,2,3], Takashi Yamada[1] and Akihiko Hirose[1,2]

[1]Division of Risk Assessment, National Institute of Health Sciences, 3-25-26, Tonomachi, Kawasaki-ku, Kawasaki 210-9501, Japan
[2]Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, Japan
[3]Present address: Chemicals Assessment and Research Center, Chemicals Evaluation and Research Institute, Japan, 1600 Shimotakano, Sugito-machi, Kitakatsushika-gun, Saitama 345-0043, Japan

(Received November 18, 2019; Accepted November 21, 2019)

ABSTRACT — In this study, we developed screening-level hepatotoxicity prediction models using test data on *in vitro* assays, which measure key events at molecular levels that are possibly linked to hepatotoxicity. Hepatotoxic chemicals were retrieved from repeated-dose toxicity databases of the Hazard Evaluation Support System Integrated Platform and the Toxicogenomics Project. *In vitro* assay data with specified protein targets likely leading to hepatotoxicity were selected using the hepatotoxic chemicals. In total, 47 *in vitro* assays were selected for constructing the hepatotoxicity prediction models. Then, two predictive models were constructed. Model A returns "Hepatotoxic" if the query chemical is tested, and the test result is "Active" in any of the selected *in vitro* assays. Model B returns "Hepatotoxic" if an analog of the query chemical is tested, and the test result is "Active" in any of the selected *in vitro* assays. External validation of the two models was performed using repeated-dose toxicity test data from the Toxicity Reference Database. Model A and Model B had sensitivity values of 0.67 and 0.72 and specificity values of 0.74 and 0.72, respectively. Our models could predict the hepatotoxic chemicals underlying the toxic mechanisms that are not established by the existing knowledge base model. On the other hand, false negatives were found to involve mechanisms requiring metabolic activation. Because our hepatotoxicity prediction model is based on the biological activity of key molecular events leading to the toxicity endpoint, scientific justification would be more acceptable as adverse outcome pathway information becomes more available.

Key words: Risk assessment, Screening tool, Hepatotoxicity, Repeated-dose toxicity

INTRODUCTION

Recent international interest in chemical safety regulations is focused on establishing an assessment strategy for identifying and assessing adverse effects of chemicals without relying solely on animal testing. *In silico* approaches are regarded as important, and further development is desired in this field.

Among *in silico* toxicity prediction models, knowledge-based models are based on chemical structure rules associated with toxicity established by experts. Compared with statistical-based prediction models, knowledge-based models have a mechanistic linkage between chemical structure and toxicity endpoint, making it easier for users to understand the basis of the prediction. However, one drawback of knowledge-based toxicity prediction models is that they are less sensitive—that is, they are likely to miss potentially toxic chemicals. This is pre-

Correspondence: Koji Jojima (E-mail: jojima-koji@ceri.jp)

sumably because it is difficult to establish various toxicity mechanisms of structurally diverse chemical substances. From the viewpoint of regulatory safety assessment, the low sensitivity of predictive models is a serious problem in practical use. Hence, it is necessary to develop a highly sensitive toxicity prediction model while retaining the advantages of a knowledge base.

In the process of constructing a knowledge-based toxicity prediction model with high sensitivity, it should be noted that the conventional approach based only on structural alerts has several limitations. Therefore, in recent years, attention has focused on using information on key molecular events, e.g., interactions between chemical substances and cellular biomolecules, leading to the development of toxicity (Dix *et al*., 2007; Sakatis *et al*., 2012; Liu *et al*., 2015). Such an approach would have the potential to further improve knowledge-based toxicity prediction models.

Repeated-dose toxicity is one of the key regulatory endpoints in the hazard assessment of chemicals. It is, however, one of the rate-limiting factors for risk assessment because of time and cost considerations. In this study, we focused on the toxicity of the liver, which is the primary target of chemical substances, and developed a screening-level hepatotoxicity prediction model using highly reliable *in vivo* toxicity databases and *in vitro* test data measuring key molecular events.

MATERIALS AND METHODS

Gathering *in vivo* hepatotoxicity data

The *in vivo* hepatotoxicity training data set was obtained from the repeated-dose toxicity databases of the Hazard Evaluation Support System (HESS) Integrated Platform provided by National Institute of Technology and Evaluation (https://www.nite.go.jp/chem/qsar/hess.html) and the Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (Open TG-GATEs) provided by the National Institutes of Biomedical Innovation, Health, and Nutrition (National Institute of Biomedical Innovation, 2007) (HESS–TGP data set; HESS, approximately 700 tests; Open TG-GATEs, approximately 150 tests), both of which include high-quality test data to assess hepatotoxicity (dose-response data of blood biochemistry, organ weights, and pathological findings) and are accessible to original toxicity test reports (Abe *et al*., 2012; Igarashi *et al*., 2015). A substance with a lowest observed effect level (LOEL) in liver of $\leq$ 50 mg/kg/day with significant changes in liver weight and histopathology is defined as hepatotoxicity positive, whereas a substance with a no observed effect level of 1000 mg/kg/day or an LOEL of 1000 mg/kg/day is regarded as hepatotoxic negative. As a result, 170 positive substances and 173 negative substances were selected. As a data set for external validation, hepatotoxicity positive and negative substances were collected according to the above criteria from repeated-dose toxicity studies included in the Toxicity Reference Database (ToxRefDB; approximately 500 studies) provided by the US Environmental Protection Agency (EPA) (Martin *et al*., 2009). Overall, 128 positive and 72 negative substances were obtained (ToxRefDB data set).

Collection of *in vitro* test data relevant to *in vivo* hepatotoxicity

We searched PubChem BioAssay for test data from *in vitro* studies measuring key molecular events that may be associated with hepatotoxicity (Wang *et al*., 2010; Kim *et al*., 2016). Data were included if they satisfied the following criteria [conditions (1) to (4)]:

(1) The protein target for the assay is specified.
(2) The number of tested compounds is $\geq$ 100.
(3) The number of positive cases in the *in vitro* assay and *in vivo* hepatotoxicity is $\geq$ 4.
(4) The number of hepatotoxicity positive/number of positive cases in the *in vitro* assay is $\geq$ 0.7.

Furthermore, *in vitro* test data related to hepatotoxicity were selected using target protein information described in the hepatotoxicity ToxList of gene pathway analysis software Ingenuity Pathway Analysis (Qiagen Bioinformatics, Redwood City, CA, USA). As a result, 47 *in vitro* test data related to liver toxicity were obtained. They include 24 tests for signal perturbation of nuclear receptors and aryl hydrocarbon receptor, 12 tests for cytochrome P450 (CYP) inhibition, and 11 tests for measurement of various cellular signal activation or inhibition (Table 1).

Development of hepatotoxicity prediction model using test data of *in vitro* assay relevant to hepatotoxicity

A hepatotoxicity prediction model was constructed using the 47 *in vitro* assay test data collected. In principle, when a query (prediction target) substance is included in each *in vitro* test data, the test result (Active or Inactive) is determined as positive or negative (Model A). This prediction method is established based on the hypothesis that the occurrence of key molecular events measured in the collected *in vitro* tests will lead to the development of hepatotoxicity. However, because not all query substances have *in vitro* test data, a different model was developed (Model B). Model B returns "Hepatotoxic" if an analog

Table 1. Selected *in vitro* test data used for constructing the hepatotoxicity prediction model.

Key event	No.
Estrogen receptor signaling antagonization	4
Androgen receptor signaling antagonization	4
Retinoic acid receptor signaling agonization	1
Retinoic acid receptor signaling antagonization	2
Constitutive androstane receptor signaling agonization	2
Constitutive androstane receptor signaling antagonization	1
Retinoid X receptor signaling agonization	1
Pregnane X receptor signaling agonization	1
Farnesoid X receptor signaling antagonization	1
Peroxisome proliferator-activated receptor gamma signaling agonization	1
Vitamin D receptor signaling antagonization	1
Thyroid hormone receptor beta signaling antagonization	1
Glucocorticoid receptor signaling antagonization	1
Aryl hydrocarbon receptor signaling agonization	3
Cytochrome P450 (CYP) 19A1 inhibition	1
CYP1A2 inhibition	3
CYP2D6 inhibition	2
CYP2C9 inhibition	2
CYP2C19 inhibition	1
CYP3A4 inhibition	3
Antioxidant Response Element signaling activation	3
Multiple drug resistance 1 interaction	1
Activator protein 1 signaling activation	1
Polo-like kinase inhibition	1
Heat shock protein beta 1 signaling activation	1
P53 signaling activation	1
Interleukin 8 secretion stimulation	1
Interleukin 1 beta signaling inhibition	1
Aldehyde dehydrogenase 1, member A1 inhibition	1

of a query chemical (similarity degree of ≥ 0.9) is tested and the test result is "Active" in any of the selected *in vitro* assays. Molecular similarity was calculated using Morgan fingerprint (Feature Definition, radius 3, 2048 bits) and Dice algorithm. These models were built using Python programming language (version 2.7.12). RDKit (version 2017.09.3), an open source chemoinformatics software, was used to generate Morgan fingerprints and calculate Dice coefficients.

RESULTS

Validation of Models A and B

To verify the prediction performance of Models A and B, the HESS–TGP data set was applied for internal validation, and the ToxRefDB data set was used for external validation. Derek Nexus, a knowledge-based toxicity prediction model (Lhasa Limited, Leeds, UK), was used for comparison of predictive performance. In Derek Nexus prediction, it was judged as positive when the level of likelihood of liver toxicity alert was equal to or higher than Equivocal (the middle of seven stages). In addition, the prediction performance when either Model A or Derek Nexus was predicted to be positive was also verified (Model A + Derek). The cross-tabulation table of each model and the predictive performance for the HESS–TGP data set are shown in Tables 2 and 3.

Table 2. Cross-tabulation of actual values versus predictions for hepatotoxicity for the HESS–TGP data set.

		Prediction (Model A)		Prediction (Model B)	
		Positive	Negative	Positive	Negative
In vivo hepatotoxicity	Positive	84	86	88	82
	Negative	27	146	37	136

Table 3. Parameters of predictive performance of Models A and B for the HESS–TGP data set.

	Model A	Model B	Derek	Model A + Derek
Sensitivity	0.49	0.52	0.44	0.72
Specificity	0.84	0.79	0.82	0.69
PPV	0.76	0.70	0.72	0.70
F measure	0.60	0.60	0.55	0.71
Accuracy	0.67	0.65	0.63	0.71

For the HESS–TGP data set, the sensitivity was 0.49 for Model A. As Model B had four true positives more than Model A, the sensitivity increased to 0.52. The specificity was 0.84 for Model A. In Model B, the specificity decreased to 0.79 because false positives increased by 10 substances compared with Model A. In addition, Model B was lower than Model A in terms of positive predictive value (PPV), F measure, and accuracy. In comparison with Derek Nexus, both Model A and Model B have higher sensitivity. In Model A, all evaluation indices were higher than those of Derek Nexus. In Model A + Derek, the specificity decreased compared with that for each model, but the sensitivity exceeded 0.7, and the F measure and accuracy were highest.

To identify the *in vitro* test data that contributed to the prediction of liver toxicity, F measures were calculated for each *in vitro* test data for Model A. Based on the results, the *in vitro* test data that resulted in the highest F measure was on suppression of thyroid hormone receptor (Thrb) signal, followed by androgen receptor (AR) antagonist and CYP2C9 inhibition. The sensitivity of each of the top 10 *in vitro* assay data with high F measures was low, but the PPV was as high as 0.75 or more (data not shown).

Table 4. Cross-tabulation of actual values versus predictions for hepatotoxicity for the Toxicity Reference Database (ToxRefDB) data set.

		Prediction (Model A)		Prediction (Model B)	
		Positive	Negative	Positive	Negative
In vivo hepatotoxicity	Positive	88	40	95	33
	Negative	19	53	20	52

Table 5. Parameters of predictive performance of Models A and B for Toxicity Reference Database (ToxRefDB) data set.

	Model A	Model B	Derek	Model A + Derek
Sensitivity	0.69	0.74	0.43	0.81
Specificity	0.74	0.72	0.78	0.61
PPV	0.82	0.83	0.78	0.79
F measure	0.75	0.78	0.55	0.80
Accuracy	0.71	0.74	0.55	0.74

The cross-tabulation table of each model and the predictive performance for the ToxRefDB data set are shown in Tables 4 and 5. The external validation results demonstrate that Model A had a sensitivity of 0.69 and a specificity of 0.74, indicating that it had higher sensitivity but lower specificity than when the HESS–TGP data set is used. Meanwhile, Model B had a higher sensitivity but lower specificity compared with Model A. Moreover, Model B had a higher PPV, F measure, and accuracy compared with Model A. The specificity level of Model A and Model B were lower than that of Derek Nexus. Conversely, the sensitivity of Derek Nexus was less than 0.5, whereas the sensitivity of Model A and Model B was approximately 0.7.

In terms of the *in vitro* test data that contributed to the prediction of hepatotoxicity for the ToxRefDB data set, the AR and Thrb antagonist tests showed higher F measures, as determined in the HESS–TGP data set. Moreover, the test data on the signal disturbance of nuclear receptors such as estrogen receptor and constitutive androstane receptor were ranked high. The sensitivity of each of the top 10 data with high F values was low, but the PPV was ≥ 0.83, which was higher than that of the HESS–TGP data set.

DISCUSSION

In this study, we developed a hepatotoxicity prediction model using *in vitro* test data on key molecular events that are part of adverse outcome pathways. We used the databases of Japanese HESS and Open TG-GATEs repeated-dose toxicity studies as *in vivo* toxicity data sources to search for relevant *in vitro* test data of key molecular events and used the US ToxRefDB for external validation. The predictive performance was examined and then compared with that of the Derek Nexus knowledge base.

The prediction performance of Model A exceeded that of Derek Nexus in terms of all evaluation indices (sensitivity, specificity, PPV, F measure, and accuracy) for internal validation as well as all metrics except specificity for external validation. Furthermore, the combination of Model A and Derek Nexus was more sensitive than Model A or Derek Nexus alone.

Model A had a high PPV in both internal and external validation. This suggests that the selected key molecular events are likely to contribute to liver toxicity. For example, there are several reports suggesting that signal disturbance of hepatic nuclear receptors affects various endogenous metabolism pathways in the liver and is associated with hypertrophy and proliferation (Thole *et al*., 2004; Lin *et al*., 2008; Wagner *et al*., 2011; Huang *et al*., 2013; Pierre *et al*., 2014). Some chemical substances can act as inhibitors of CYP species, thereby suppressing the metabolism of endogenous ligands in the liver, and other substances may act as substrates for CYP species to produce reactive metabolites, leading to cellular dysfunction and injury in the liver (Feng and He, 2013).

The developed model is more sensitive than the knowledge-based Derek Nexus, but there are still several false negatives. This indicates that there is still some missing key molecular event information relevant to hepatotoxicity. However, this is not the only reason for the low sensitivity. In the HESS–TGP data set, when a false negative in Model A and a true positive in Derek Nexus were examined, six had alerts for "Halobenzene," "Halogenated hydrocarbon," or "Aromatic nitro compound" by Derek Nexus. Many of the halides are thought to require metabolic activation by CYP2E1 for the development of hepatotoxicity; Model A, however, does not include a test for CYP2E1. "Aromatic nitro compound" is also thought to involve reactive metabolites in the development of hepatotoxicity. Thus, considering these factors, one of the characteristics of the substances that are false negative in Model A was considered to be involved in metabolic activation.

The absorption, distribution, metabolism, and excretion (ADME) viewpoint may be lacking as a false positive factor that reduces the performance of the hepatotoxicity prediction model developed in this study. One of them concerns the effect of oral bioavailability. As chemical descriptors affecting oral bioavailability, topological polar surface area (TPSA) and rotatable bond count

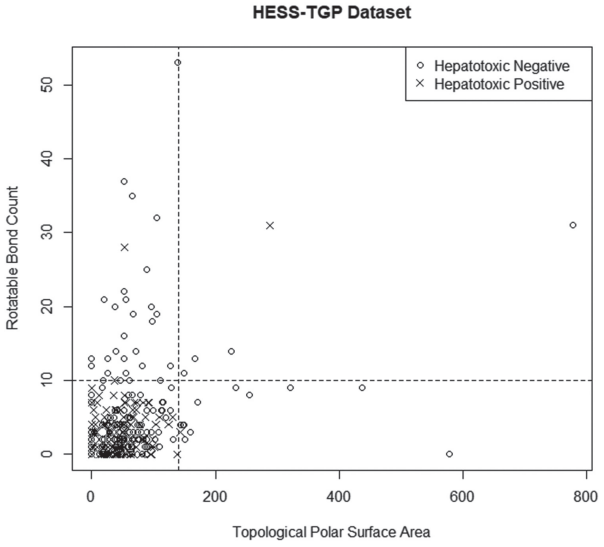

Fig. 1. Scatter plot of topological polar surface area (TPSA) and rotatable bond count (RBC) for chemicals in the HESS–TGP data set. Hepatotoxic positive chemicals are shown as crosses (×), and hepatotoxic negative chemicals are indicated by open circles (○). It was reported that compounds meeting the criteria of ≤ 10 RBC and ≤ 140 TPSA will have a high oral bioavailability in rats (Veber *et al*., 2002).

(RBC) have been proposed (Veber *et al*., 2002). It has been shown that oral availability decreases when TPSA > 140 or when RBC > 10. In the analysis of the HESS–TGP data set, most of the substances with RBC > 10 or TPSA > 140 are hepatotoxicity negative (Fig. 1). Prediction of substances with low oral bioavailability based on these descriptors may improve the prediction performance of our hepatotoxicity model. Furthermore, the lack of consideration of the effects of excretion was another false positive factor. In the HESS–TGP data set and ToxRefDB data set, there were a total of 22 substances with sulfo groups. Of these, 21 were negative for hepatotoxicity, but five substances were false positive in Model A. Substances with sulfo groups, such as sulfate conjugate metabolites, are considered to be rapidly excreted from the body, and as a result, the hepatotoxic effect is reduced. The possibility of reducing false positives was facilitated by considering ADME.

In summary, Model A has a wider chemical space and better prediction performance compared with the existing knowledge-based toxicity prediction model. In addition, Model B can be improved by investigating the optimal similarity settings. It is also projected that the prediction performance can be further improved by incorporating the still missing toxicity mechanisms and ADME not established by this model.

ACKNOWLEDGEMENTS

This work was supported by a Health and Labour Sciences Research Grant (H30-Kagaku-Sitei-005) (2018-2020) for "Development of evaluation strategies for human health risks of chemical substances based on advancement and practical application of in silico prediction technology" from the Ministry of Health, Labour and Welfare, Japan.

Conflict of interest---- The authors declare that there is no conflict of interest.

REFERENCES

Abe, T., Kobayashi, K., Nishikawa, S., Yamada, T., Hasegawa, R., Sakuratani, Y., Yamada, J., Hirose, A., Kamata, E., Yoshida, M., Sunouchi, M., Honma, M., Yamashita, T., Sakai, K., Yamazoe, Y. and Hayashi, M. (2012): Development of hazard evaluation support system (HESS). Pharm Stage, **12**, 39-47.

Dix, D.J., Houck, K.A., Martin, M.T., Richard, A.M., Setzer, R.W. and Kavlock, R.J. (2007): The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol. Sci., **95**, 5-12.

Feng, S. and He, X. (2013): Mechanism-based inhibition of CYP450: an indicator of drug-induced hepatotoxicity. Curr. Drug Metab., **14**, 921-945.

Heller, S.R., McNaught, A., Pletnev, I., Stein, S. and Tchekhovskoi, D. (2015): InChI, the IUPAC International Chemical Identifier. J. Cheminform., **7**.

Huang, Y.Y., Gusdon, A.M. and Qu, S. (2013): Nonalcoholic fatty liver disease: molecular pathways and therapeutic strategies. Lipids Health Dis., **12**.

Igarashi, Y., Nakatsu, N., Yamashita, T., Ono, A., Ohno, Y., Urushidani, T. and Yamada, H. (2015): Open TG-GATEs: a large-scale toxicogenomics database. Nucleic Acids Res., **43**, D921-D927.

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J. and Bryant, S.H. (2016): PubChem substance and compound databases. Nucleic Acids Res., **44**, D1202-D1213.

Lin, H.Y., Yu, I.C., Wang, R.S., Chen, Y.T., Liu, N.C., Altuwaijri, S., Hsu, C.L., Ma, W.L., Jokinen, J., Sparks, J.D., Yeh, S. and Chang, C. (2008): Increased hepatic steatosis and insulin resistance in mice lacking hepatic androgen receptor. Hepatology, **47**, 1924-1935.

Liu, J., Mansouri, K., Judson, R.S., Martin, M.T., Hong, H., Chen, M., Xu, X., Thomas, R.S. and Shah, I. (2015): Predicting hepatotoxicity using ToxCast *in vitro* bioactivity and chemical structure. Chem. Res. Toxicol., **28**, 738-751.

Martin, M.T., Judson, R.S., Reif, D.M., Kavlock, R.J. and Dix, D.J. (2009): Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ. Health Perspect., **117**, 392-399.

National Center for Biotechnology Information. "The PubChem Project," Available: https://pubchem.ncbi.nlm.nih.gov/. (accessed 2019-1-30).

National Institute of Technology and Evaluation (NITE). Hazard evaluation support system integrated platform (HESS), Available: https://www.nite.go.jp/chem/qsar/hess.html. (accessed 2019-1-30).

Pierre, S., Chevallier, A., Teixeira-Clerc, F., Ambolet-Camoit, A., Bui, L.C., Bats, A.S., Fournet, J.C., Fernandez-Salguero, P., Aggerbeck, M., Lotersztajn, S., Barouki, R. and Coumoul, X. (2014): Aryl hydrocarbon receptor–dependent induction of liver fibrosis by dioxin. Toxicol. Sci., **137**, 114-124.

Pletnev, I., Erin, A., McNaught, A., Blinov, K., Tchekhovskoi, D. and Heller, S. (2012): InChIKey collision resistance: an experimental testing. J. Cheminform., **4**.

Sakatis, M.Z., Reese, M.J., Harrell, A.W., Taylor, M.A., Baines, I.A., Chen, L., Bloomer, J.C., Yang, E.Y., Ellens, H.M., Ambroso, J.L., Lovatt, C.A., Ayrton, A.D. and Clarke, S.E. (2012): Preclinical strategy to reduce clinical hepatotoxicity using *in vitro* bioactivation data for >200 compounds. Chem. Res. Toxicol., **25**, 2067-2082.

Thole, Z., Manso, G., Salgueiro, E., Revuelta, P. and Hidalgo, A. (2004): Hepatotoxicity induced by antiandrogens: A review of the literature. Urol. Int., **73**, 289-295.

Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W. and Kopple, K.D. (2002): Molecular properties that influence the oral bioavailability of drug candidates. J. Med. Chem., **45**, 2615-2623.

Wagner, M., Zollner, G. and Trauner, M. (2011): Nuclear receptors in liver disease. Hepatology, **53**, 1023-1034.

Wang, Y., Bolton, E., Dracheva, S., Karapetyan, K., Shoemaker, B.A., Suzek, T.O., Wang, J., Xiao, J., Zhang, J. and Bryant, S.H. (2010): An overview of the PubChem BioAssay resource. Nucleic Acids Res., **38**, D255-D266.

National Institute of Biomedical Innovation (2007), Toxicogenomics Project Toxicity Data Collection (in Japanese). https://toxico.nibiohn.go.jp/seika/01-exp_summary/TGP_ToxicologyData_summary.pdf

United States Environmental Protection Agency. ToxRefDB - Release user-friendly web-based tool for mining ToxRefDB. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NCCT&dirEntryId=227139