Chapter 11

In Silico Models for Hepatotoxicity

Mark Hewitt and Katarzyna Przybylak

Abstract

In this chapter we review the challenges of predicting human hepatotoxicity. Principally, this is our partial understanding of a very complex biochemical system and our ability to emulate that in a predictive capacity. We give an overview of the published modeling approaches in this area to date and discuss their design, strengths, and weaknesses. It is interesting to note the shift during the period of this review in the direction of evidenced-based approaches including structural alerts and pharmacophore models. Proposals on how best to utilize the data emerging from modeling studies are also discussed.

Key words Liver, Hepatotoxicity, In silico or computational prediction, QSAR, Expert system

1 Introduction

Toxicity of new medicinal compounds to the liver is perhaps the most significant hurdle to overcome during drug development. Often termed "drug-induced liver injury (DILI)," these adverse effects can range in nature from subtle elevations in serum enzymes, to acute and chronic hepatocellular injuries (steatosis, necrosis, cirrhosis), cholestatic injuries, and neoplasia [1]. Unfortunately, DILI accounts for a significant proportion of drugs (>25 %) being terminated during development or withdrawn from the market [2].

Given the protective/metabolic function of the liver, it is perhaps not surprising that hepatotoxicity is frequently encountered. Given the livers high blood flow and first-pass metabolism it is a certainty that a proportion of the diverse pharmaceutical products in use today are hepatotoxic (via metabolic conversion). Unwanted interaction between the liver and pharmaceuticals is a major hurdle which can often result in the loss of drug efficacy and/or hepatotoxicity. Despite preclinical and clinical safety assessments, liver toxicity remains a main cause of drug development failures and subsequent market withdrawal due to the poor predictivity of idiosyncratic toxicity in animal models [3, 4].

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_11, © Springer Science+Business Media New York 2016

The need to predict whether a new drug is likely to lead to hepatotoxicity is clear. Information relating to the likelihood of liver toxicity is critical in order to increase patient safety, reduce the frequency of drug withdrawals/failures and to further increase our understanding of liver toxicity.

Interestingly, despite a clear need to predict these effects, computational studies in this area have only started to emerge in the last decade [1, 5]. Such methods are well-suited to the rapid screening of large numbers of compounds, offering significant time and cost savings over traditional animal-based screening approaches. Furthermore, computational screening has been successfully established for other endpoints, including skin sensitization and mutagenicity [6, 7]. When coupled with supporting in vitro data they provide a powerful tool capable of predicting toxicity and, in certain cases, determining the mechanism of that toxicity. However, as stated, computational models for DILI have only recently started to surface and those that have been published are often limited in their scope and predictive capability.

The reason for this is simple; predicting toxicity to the liver is far from simple!

The task of predicting DILI is difficult because (a) the liver is an intricate and complex organ with numerous biological and metabolic pathways that can lead to downstream toxicities and (b) many of these toxicological pathways are poorly understood or remain unknown. Furthermore, as already introduced, DILI can take many forms and range in severity. With the absence of a single "catch all" biomarker that can be used as a metric of hepatotoxicity, actually measuring these affects in patients is very challenging.

Furthermore, toxicity to the liver can occur in a dose dependent manner (termed intrinsic toxicity) or in a non-dose dependent manner (termed idiosyncratic toxicity) [8]. Typically, intrinsic liver toxicity accounts for approximately 80 % of cases, where the observed toxicity can be related to a particular mechanism of action (pharmacological, toxicological, or chemical) triggered by the drug or its metabolite(s). Idiosyncratic toxicity is very difficult to predict and is thankfully a relatively rare occurrence. The susceptibility of particular patients to idiosyncratic DILI has been the focus of much research [9], but the prediction of idiosyncratic effects remains a herculean task.

2 Prediction of Hepatotoxicity

It is crucial to develop predictive screening systems and mechanistic models capable of detecting hepatotoxicity as early as possible in the drug development process. However, accurate prediction of organ toxicity is very challenging due to the complexity of the underlying mechanisms, which are very often not known.

Moreover, the lack of specific and selective biomarkers that can be used to detect hepatotoxicity leads to a shortage of reliable in vivo and in vitro data from which to derive predictive models. Most likely as a result of these limitations, the first in silico models were described in the literature only at the beginning of the last decade [10, 11]. The bulk of available computational models for liver toxicity have been published more recently [1].

Published models for the prediction of hepatotoxicity can be classified as belonging to one of two approaches [12]:

(A) The development of statistically based structure–activity relationship (SARs) of varying complexity. This modeling approach utilizes existing DILI data to derive a model able to predict a quantitative estimation of hepatotoxicity.

(B) The development of qualitative "models" based on expert knowledge, directly related to chemical structure and molecular features. Most often, these qualitative approaches result in the development of structural alerts or three-dimensional pharmacophore models.

These models can be further subdivided based upon (1) the endpoint being modeled (general hepatotoxicity or a specific aspect (e.g., steatosis)), (2) the type of variable(s) (descriptors) used to develop the model, or (3) the type of data being modeled (in vivo or in vitro). Figure 1 depicts how the 21 published models that are the subject of this chapter can be divided using these four differentiating criteria.

Statistical models are generally built from a training dataset of chemical structures and their associated toxicity data, expressed either in quantitative or qualitative terms, using an appropriate algorithm. Therefore, they are often referred to as "(quantitative) structure–activity relationships" ((Q)SARs). In contrast, expert systems apply expert knowledge to a predictive environment and are usually not statistically based. The knowledge is based on the observed toxicity of known compounds, together with an understanding of toxicological mechanisms, metabolism and chemical reactivity [13].

The development of statistical models is usually faster (if suitable data are available) than that of expert systems, since expert systems require extensive study and integration with existing literature sources and are usually evidence-based (examples and supporting documentation is supplied along with a prediction). Therefore, statistical models are the most common. Approximately 75 % of the existing predictive models for liver toxicity have been developed using an array of different statistical methodologies, including discriminant analysis [14], Bayesian models [15, 16], Artificial Neutral Networks (ANN) [14], k-Nearest Neighbor (kNN) [17, 18], Random Forest (RF) [18, 19], and specialist QSAR software [20].

In terms of endpoint, most in silico models are focused towards the prediction of general hepatotoxicity (positive/negative irrespective of the mechanism/toxicity outcome) [5, 10, 14, 15, 18, 19, 21–26]. However, it is important to stress that this trend seems to be changing in recent years as the number of approaches considering more specific endpoints is increasing. Examples of these specific endpoints include elevations of liver serum enzymes [17], cholestasis and jaundice [20], hepatosteatosis [27, 28], and hepatic histopathologic effects including hypertrophy, injury, and proliferative lesions [29].

It is interesting to see that the majority of in silico approaches have utilized variables representing only chemical structure [10, 11, 14–17, 20, 21, 23–25]. It is perhaps not surprising given that QSAR models traditionally relate chemical structure to observed activity, but it seems here that the complex nature of the liver may warrant the use of biological descriptors to describe the biological process/systems at work. Only three models, discussed later, employed both chemical and biological descriptors and are referred to as hybrid models [18, 19, 29].

Finally, considering the nature of endpoint data used for modeling, most models have been developed using in vivo data. This can be broken down further into human data [10, 14–17, 19, 20, 24–26] and animal data [18, 29] which may be further subdivided into data from different species [23]. Only two models have been built using in vitro data [11, 21] and a further two models utilizing both in vitro and in vivo data [23, 30]. The 21 in silico models considered in this chapter can be subdivided by their differentiating characteristics as described by Fig. 1. The models will be discussed in the context of these categories and the strengths and weaknesses of different modeling methods will be highlighted. Potential future developments in the area are also speculated.

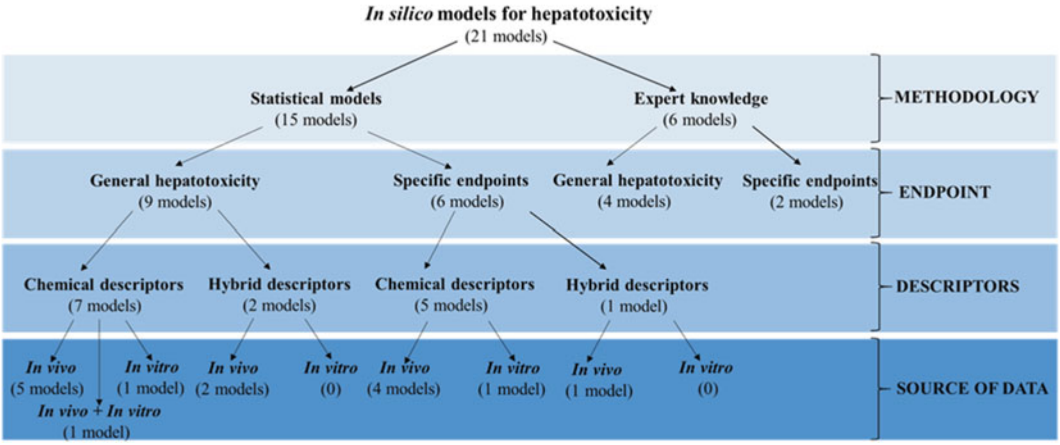


Fig. 1 Summary of published in silico models for predicting liver toxicity between 2000 and 2015

2.1 Statistically Derived Quantitative Models

A large proportion of the published hepatotoxicity models are statistical in their nature. The predictive element of these models is the statistical correlation of toxicity with one or more dependent variables. The approach used to identify and model this correlation varies considerably both in terms of methodology and complexity. Usually, statistically derived models are developed using sophisticated modeling software and tools.

The premise of any (Q)SAR model is the relationship between chemical structure (described using a number of descriptors) and biological activity (e.g., liver toxicity). This enables predictions of such activity to be made for new substances based on their chemical structure. The algorithms used to construct these models comprise of simple linear regression, complex multi-variant data modeling, data mining, and classification approaches [31]. Every statistical model has to be internally and externally validated to show its true predictive power and reliability [32, 33]. The predictive performance is usually evaluated by sensitivity (correctly predicted positive chemicals), specificity (correctly predicted negative chemicals) and accuracy (correctly predicted positive and negative chemicals). High sensitivity and specificity of a model guarantees correct classification of toxicologically active and inactive compounds. Therefore, it is the most important feature when aiming to detect potential hepatotoxic drugs in early drug development, since the consequences of misclassifying a toxic (positive) drug are severe (i.e., the possibility of a toxic drug reaching clinical trials) [14]. Of course, in drug development poor specificity can be a significant problem since many negative compounds may be dropped from further development unnecessarily.

Table 1, at the end of this section, describes the 15 diverse hepatotoxicity models discussed in this chapter and gives details of the methodologies employed, the endpoint modeled, the type of descriptors utilized, and the source of hepatotoxicity data.

2.1.1 Statistical Models for In Vivo General Hepatotoxicity Endpoint Using Chemical Descriptors

As already stated, most of the available in silico models have been developed based on in vivo data and are used to predict a general hepatotoxicity endpoint. These models consider intrinsic hepatotoxicity, idiosyncratic hepatotoxicity or a combination of these. The majority have been developed based only on the chemical features of the training set.

One of the first published in silico models was developed by Cheng and Dixon (ID 1 in Table 1) and is predicting intrinsic liver toxicity in humans [10]. Data for 382 drug and drug-like compounds (of various therapeutic classes) were collected from the literature. Amongst them, there were 149 chemicals which caused dose-dependent hepatocellular, cholestatic, neoplastic and other liver injuries. The authors employed a modeling method known as recursive partitioning (RP) [34, 35] with an ensemble approach [36], wherein the overall model is actually an average of numerous

Table 1
Statistically based in silico models to predict hepatotoxicity available in the literature

ID	Statistical method	Endpoint	Descriptors	Type and size of data	Validation	Predictive performance	Ref
1	Ensemble recursive partitioning	Dose-dependent hepatotoxicity	Chemical: 25 1D and 2D descriptors filtered by Monte Carlo	In vivo data for 382 chemicals	IV: LOO, leave 10 % out EV: 54 chemicals	IV: 78 % (LOO), 76 % (leave 10 % out) SEN; 90 % (LOO), 75 % (leave 10 % out) SPE; 85 % (LOO), 76 % (leave 10 % out) ACC EV: 81 % ACC	[10]
2	Molecular interaction fields and SIMCA	Cell proliferation; LDH activity; ATP levels; Caspases 3 and 7 levels	Chemical: steric and electrostatic IFO descriptors from SYBYL	In vitro (human hepatocyte cells) for 654 chemicals	CV for 27 NSAIDs	IV: 52 % ACC CV: 93 % SEN; 85 % SPE; 83 % ACC (for 6 NSAIDs)	[11]
3	QSAR (MLR using Sigma Stat)	LC_{50} (mM)	Chemical: log P EHOMO and μ calculated by MOPAC	In vitro for rat and human cells for 12 halobenzenes	Not reported	$r^2 = 0.966$ for rat cell, 0.993 for human cell, 0.846 for induced rat cell	[21]
4	LDA, ANN, OneR	Idiosyncratic hepatotoxicity	Chemical: 3D RDF descriptors	In vivo human for 74 drugs	CV, EV for 13 drugs and three pairs of similar chemicals with opposite toxic potential	LDA model Train: 88 % SEN, 93 % SPE, 90 % ACC Test: 84 % SEN, 91 % SPE, 88 % ACC ANN model Train: 92 % SEN, 90 % SPE, 91 % ACC Test: 75 % SEN, 80 % SPE, 78 % ACC OneR model Train: 77 % SEN, 100 % SPE, 84 % ACC Test: 75 % SEN, 98 % SPE, 81 % ACC EV: for 3 pairs: 83 % ACC; for 13 drugs 69 % ACC	[14]

5	QSAR software: MC4PC, BioEpisteme, MDL-QSAR, Leadscope	(1) Hepatobiliary: liver disorders; jaundice and cholestasis; liver enzymes; gall bladder disorders; bile duct disorders, (2) urinary tract.	Chemical 2D descriptors	In vivo human for 1660 chemicals	IV: LMO, LOO EV for 18 toxic chemicals	IV: for LMO 39 % SEN, 86 % SPE For consensus model: 56 % SEN, 78 % SPE EV: 89 % ACC	[20, 45, 46]
6	Ligand-based Bayesian model	Idiosyncratic hepatotoxicity	Chemical 2D descriptors and FCFP of maximum diameter of 6	In vivo human for 295 chemicals	EV for 237 chemicals	EV: 56 % SEN, 67 % SPE, 60 % ACC	[15]
7	kNN	Five liver serum enzymes: ALP, ALT, AST, LDH, GG	Chemical: MolConnZ-topological indices and Dragon descriptors	In vivo human for 490 chemicals	IV, Y-randomization, EV	EV: for composite liver enzymes: 74 % SEN, 94 % SPE, 84 % ACC	[17]
8	SVM and clustering by chemical similarity	Hepatotoxicity	Chemical: 2D molecular fragments and Dragon descriptors	In vivo and in vitro human, rodent, and non-rodent for 951 chemical	Fivefold CV, EV for 246 chemicals and 18 chemicals toxic in non-rodents	Fivefold CV internal: 62–67 % ACC Fivefold CV external: 62–67 % ACC EV: for 246 chemicals 65–67 % ACC	[23]
9	kNN, SVM, RF, DWD	Hepatotoxicity	Chemical: Dragon and MOE Biological: toxicogenomics	In vivo rat for 127 chemicals	Fivefold CV	For chemical descriptors: CV: 45–56 % SEN, 60–77 % SPE, 55–61 % ACC For biological descriptors: CV: 57–67 % SEN, 77–84 % SPE, 69–76 % ACC For hybrid descriptors: CV: 76–77 % ACC	[18]

(continued)

Table 1 (continued)

ID	Statistical method	Endpoint	Descriptors	Type and size of data	Validation	Predictive performance	Ref
10	Ensemble of mixed learning algorithms	Hepatotoxicity	Chemical: PaDEL descriptors	In vivo human for 1087 chemicals	Fivefold CV, Y-randomization, EV for 187 chemicals divided into three sets	CV: 64 % SEN, 63 % SPE, 64 % ACC EV1: 68 % SEN, 71 % SPE, 70 % ACC EV2: 64 % SEN, 37 % SPE, 51 % ACC EV3: 62 % SEN, 62 % SPE, 62 % ACC	[24]
11	13 QSAR models developed using Bayesian methodology	13 hepatotoxic side effects	Chemical: functional class fingerprints (FCFP_6)	In vivo human for 888 chemicals	IV: LOO EV for three sets: LTKD, Pfizer, and O'Brien	LOO for 13 models: >71 % SEN, >94 % SPE, >93 % ACC EV LTKD: 66 % SEN, 67 % SPE, 66 % ACC EV Pfizer: 52 % SEN, 73 % SPE, 60 % ACC EV O'Brien: 56 % SEN, 93 % SPE, 70 % ACC	[16]
12	Machine learning methodology DT	Hepatotoxicity	Chemical: 82 Mold descriptors	In vivo human for 197 chemicals	Tenfold CV EV for three sets	CV: 58 % SEN, 78 % SPE, 70 % ACC EV1: 66 % SEN, 72 % SPE, 69 % ACC EV2: 58 % SEN, 67 % SPE, 62 % ACC EV3: 61 % SEN, 66 % SPE, 63 % ACC	[25]
13	RF	Hepatotoxicity	Chemical: CDK, Dragon, MOE Biological: HIATs	In vivo human for 292 chemicals	Fivefold CV	CV for chemical descriptors: 67 % SEN, 54 % SPE, 60 % ACC CV for biological descriptors: 67 % SEN, 87 % SPE, 77 % ACC CV for hybrid descriptors: 71 % SEN, 74 % SPE, 73 % ACC	[19]

14	Six machine learning analysis	Hepatic histopathologic effects: hypertrophy, injury and proliferative lesions	Chemical: 726 descriptor Biological: 124 bioactivity from ToxCast21	In vivo rat	Tenfold CV	CV: 84 % ACC for hypertrophy; 80 % ACC for injury; and 80 % ACC for proliferative lesions	[29]
15	PLS-DA	LXR binding potential involved in liver steatosis	Chemical: 6 PaDEL and 5 RDKit	LXRβ binding affinity for 356 LXR binders	Not reported		[52]

IV internal validation, *EV* external validation, *CV* cross-validation, *LOO* leave-one-out, *SEN* sensitivity, *SPE* specificity, *ACC* accuracy, *SIMCA* Soft Independent Modeling of Class Analogy, *LDH* lactate dehydrogenase, *ATP* adenosine triphosphate, *IFO* Idiotropic Field Orientation, *NSAIDs* nonsteroidal anti-inflammatory drugs, *QSAR* quantitative structure–activity relationship, *MLR* multiple linear regression, *EHOMO* energy of highest occupied molecular orbital, *LDA* linear discriminant analysis, *ANN* artificial neural networks, *RDF* radial distribution function, *LMO* leave many out, *FCFP* functional class fingerprint, *kNN* k-Nearest Neighbor, *ALP* alkaline phosphatase, *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *GG* glutamyl transpeptidase, *SVM* support vector machine, *RF* random forest, *DWD* distance weighted discrimination, *MOE* molecular operating environment, *DT* decision tree, *CDK* chemistry development kit, *HIAT* hepatocyte imaging assay technology models developed from random subsets of the training set. The RP technique involves the use of a decision tree to split the training dataset into predominantly toxic or predominantly nontoxic molecules based on the independent variables. Twenty-five descriptors were selected from 1D molecular similarity scores and 2D structural information using a Monte Carlo linear regression algorithm. As a result, 151 different trees were generated with the RP approach. A compound was predicted using each of the 151 trees as being toxic or nontoxic and then the ensemble average was used to obtain the final prediction. Leave-one-out (LOO) and leave-10 %-out validation techniques yielded an overall concordance of 85 % and 76 %, respectively. The external validation of 54 compounds (23 toxic) gave a similar order of accuracy (81 %). This study showed the usefulness of the ensemble approach, using a diverse training dataset to build a model that can be applied to a broad range of chemical classes. Furthermore, a measure of predictive confidence is also supplied. However, a potential drawback of an ensemble approach is observed when the combination of models makes the method less transparent and more difficult (or impossible) to investigate the underlying mechanisms.

The next model (ID 4), developed by Cruz-Monteagudo, employed a number of different modeling methods to predict hepatotoxicity; linear discriminant analysis (LDA), artificial neural networks (ANN), and machine learning algorithms [14]. In this study, 33 compounds associated with idiosyncratic hepatotoxicity and 41 chemicals not associated with liver toxicity were collected from the literature. The models used 3D Radial Distribution Function (RDF) descriptors, which give information about interatomic distances in the entire molecule, ring types, planar and nonplanar systems, atom types, and bond distances. The best predictive performance was obtained with the LDA model, which correctly classified 86.4 % of compounds. Furthermore, based on the LDA model, a "desirability" analysis was performed in order to ascertain the characteristics, or descriptor values, that a drug candidate should have to ensure a lower idiosyncratic hepatotoxicity potential. For the external validation, two small datasets were used. The first set consisted of three pairs of chemically and pharmacologically related drugs having opposite observed toxicological profiles, including toxic troglitazone vs. nontoxic pioglitazone (insulin resistance drugs), toxic tolcapone vs. nontoxic entacapone (catechol-*O*-methyltransferase (COMT) inhibitors), and toxic clozapine vs. nontoxic olanzapine (psychotropic drugs). In this case, LDA and OneR predicted hepatotoxicity with the same accuracy of 83.3 %. The second external set was created from 13 published drugs, all hepatotoxic, and was used to validate the LDA model. Nine out of the 13 drugs were classified correctly and provide evidence that the computational approaches could be applied in early drug discovery to minimize the selection of chemicals with idiosyncratic hepatotoxicity.

Another model (ID 6) for idiosyncratic hepatotoxicity was developed by Ekins et al. [15]. They used a training set of 295 compounds (containing 158 DILI-inducers) and an external validation set of 237 molecules (114 DILI-inducers) to develop a liver toxicity prediction model using a Bayesian classification approach [37]. 2D molecular descriptors and extended connectivity functional class fingerprints of maximum diameter 6 (ECFC_6) were used to differentiate the active from inactive molecules and also to highlight chemical substructures known to be important for DILI, such as ketones, diols, and α-methyl styrene. In addition, the authors applied SMILES Arbitrary Target Specification (SMARTS) filters published by several pharmaceutical companies to all 532 molecules to evaluate whether such reactive substructures could be readily detected by any of these filters. The best predictivity was obtained for the Bayesian model which correctly classified 56.0 % of active chemicals and 66.7 % of inactive compounds. The external validation resulted in 59.9 % accuracy. Regarding the SMARTS filters, the Abbott filters resulted in more stringent classification, giving a reasonable sensitivity of 66.9 %, but a relatively low specificity of 40.3 %. A significant outcome of this study was the provision of the structural and DILI classification data that can be used as a foundation for developing future computational models, as well as filters, in the early stages of the drug development process. It is evident that approaches such as the one above are not yet capable of delivering acceptable levels of predictivity. However, their potential application of drug screening makes them of great interest.

Exploring the premise that no single learning algorithm is optimal for toxicity modeling problems, Liew et al. applied an ensemble of mixed learning algorithms and mixed features to develop a model to predict hepatic adverse effects (ID 10) [24]. The authors obtained the list of available drugs on the market from the US Food and Drug Administration (US FDA) Orange Book [38], which were then screened for adverse hepatic effects by checking the reports on adverse reaction in each drug's monograph. A final set of 1274 drugs was obtained which were split into a modeling set of 1087 and a validation set of 187 compounds. Using PaDEL descriptors [39] calculated for the training set, a total of 617 base classifiers were selected using three algorithms: support vector machine (SVM), k-nearest neighbor (kNN), and Naiv̈e Bayes (NB). The remaining 187 compounds were divided into three different external validation sets. Two of them were aimed at verifying the model's ability to predict "severely" toxic compounds and structurally similar chemicals but of opposing toxicity status. The outcome of this was that 22 of 23 withdrawn drugs or those with black warnings were predicted correctly. However, for the structurally similar chemicals with opposite hepatotoxicity potential, only 30 % of nontoxic drugs were predicted correctly. The inability of the model to separate the non-hepatoxic chemicals was probably due to the similarity of the true negative compounds to positive training compounds, coupled with the inherent difficulty to separate highly similar compounds by QSAR, which by definition expects that structurally related chemicals have similar activities. The third external set of 120 drugs gave the most reliable evaluation of model performance resulting in a sensitivity of 81.9 %, specificity of 64.6 % and overall accuracy of 75 %. The ensemble model was able to identify the positive compounds quite well, but it was less successful in classifying negative chemicals, especially when they were structurally similar. In general, this study again demonstrated the usefulness of an ensemble methodology when applied to large and diverse datasets similarly to the Cheng and Dixon study [10].

It is very important, especially in the case of such a complex endpoint as hepatotoxicity, to correctly annotate a drugs' potential to induce toxicity. The accuracy and utility of a predictive model depends largely on how to annotate the potential of a drug to cause hepatotoxicity in a reliable and consistent way. To address this issue, Chen et al. used the high quality US FDA-approved drug labeling DILI dataset to construct a QSAR model for hepatotoxicity (ID 12) [25]. Within this dataset most DILI-concern drugs are (1) withdrawn from the market; (2) labeled with a boxed warning; or (3) indicated in the warning and precautions section. The authors divided the 387 drugs into a training set of 197 drugs (containing 81 positives) and test dataset of 190 drugs (95 positives). They then used a Decision Tree (DT) algorithm and Mold molecular descriptors to develop a QSAR model to predict hepatotoxicity in humans. The model consisted of six decision trees using 82 descriptors. Its predictive performance was first assessed by tenfold cross validation giving an overall accuracy of 69.7 %. Then external validation was undertaken applying the test set and two additional (independent) validation datasets: Green dataset consisting of 214 hepatotoxins and 114 drugs with no evidence of hepatotoxicity [22] and the Xu dataset consisting of 132 hepatotoxins and 109 negative compounds [40]. The accuracy obtained in each external validation was between 61.6 and 68.9 %. The external validation also showed that the drugs with consistent annotations among these three validation sets were better predicted (69.1 % accuracy) than drugs with inconsistent annotations (58.8 % correctly predicted). Finally, the applicability of the model was examined. To this aim, 2000 repetitions of cross-validation based on the training set were performed to identify therapeutic subgroups in which the QSAR model had higher or lower accuracy than the overall accuracy. As a result, 22 therapeutic subgroups with high-prediction confidence and 18 therapeutic categories with low prediction confidence were identified. Some drugs in the higher confidence subgroups, such as: analgesic, antibacterial agents and antihistamines, are well documented either to cause or not to cause DILI. Focusing only on the therapeutic categories with high prediction confidence, the accuracy of model increased to 73.6 %. So, the therapeutic categories can be used to define the chemical structure space, where the model has better predictive power. This study demonstrates that using relatively large datasets with high quality annotations and focusing on the therapeutic subgroups where the model performs best is crucial in developing reliable predictive models, especially for very complex endpoint, such as liver toxicity.

2.1.2 Statistical Models for In Vitro General Hepatotoxicity Using Chemical Descriptors

Considering the scarcity of in vitro data, only one study employed such data to predict general hepatotoxicity (ID 3). It is not a typical in silico predictive model, as it focuses mostly on the validation of the in vitro method itself using isolated hepatocytes, which includes QSARs examining physicochemical properties of chemical congeners responsible for observed cytotoxic activity [21]. The authors investigated the molecular mechanism of hepatotoxicity for 12 halobenzenes in rat and human hepatocytes. A relatively good correlation ($r^2 = 0.90$) between LC_{50} measured in phenobarbital (PB)-induced rat hepatocytes and in vivo toxicity in PB-induced male Sprague-Dawley (SD) rats was found. Moreover, the QSAR was used to identify the metabolic activating pathway in halobenzene toxicity. It was found that toxicity in normal rat and human hepatocytes was strongly correlated with hydrophobicity ($\log P$), ease of oxidation (energy of Highest Occupied Molecular Orbital (EHOMO)) and the asymmetric charge distribution according to the arrangement of halogen substituents (dipole moment, μ). This suggests that the mechanism of toxicity is similar in both species and involves the interaction between halogens and cytochrome CYP450 for oxidation. In the case of PB-induced rat hepatocytes, halobenzene toxicity was correlated only with $\log P$ and dipole moment, but not EHOMO. This can indicate that ease of oxidation is no longer of significance in the underlying toxicity. This study is significant as it allows for better understanding of hepatotoxic mechanism(s) for that class of chemical. This knowledge is critical for the future prediction of hepatotoxicity.

2.1.3 Statistical Models for In Vivo and In Vitro General Hepatotoxicity Using Chemical Descriptors

Only a single example could be found where a combination of in vivo and in vitro data was used to develop a computational model for hepatotoxicity (ID 8) [23]. Given the success of ensemble modeling approaches previously applied, pooling together all supporting or descriptive data seems a logical step in order to try to explain and increase user confidence when predicting complex endpoints. Fourches et al. constructed a large and diverse dataset for liver toxicity using a novel approach of text mining from the published literature. The authors extracted 14,000 assertions linking compounds to different degrees, or types, of hepatotoxicity (from the cellular level to the whole organ) across different species:

including humans and rodents (mostly rat and mouse). A final dataset of 951 compounds was obtained following a data curation process. The data were classified into "class 1" consisting of 248 chemicals inducing liver effects in humans only and "class 2" consisting of 283 compounds inducing no liver toxicity in humans, but causing liver effects in rodents. The authors used hierarchical cluster analysis to identify groups of chemicals sharing similar molecular motifs corresponding to similar liver effect profiles in humans and rodents. As reported by Liew et al. [24] in their previous study, Fourches et al. again identified clusters of structurally similar molecules that possessed different liver effect profiles. This presents a significant challenge for modeling approaches fundamentally based on the premise that structurally similar compounds should act in a similar manner. It is possible that, descriptor-based approaches such as these are not sensitive enough to distinguish these compounds and opens the door to structural alert-based approaches which are discussed later in this chapter.

In addition, the authors also developed Support Vector Machine (SVM)-based models to predict whether a compound would be expected to produce adverse liver effects in humans. Predictive performance was assessed by internal and external five-fold cross-validation, giving accuracies ranging from 61.9 to 67.5 % and 55.7–72.6 % for internal and external validation, respectively. After removal of structural outliers using an implementation of the applicability domain, an accuracy of 67.8 % was obtained for an external validation dataset of 222 compounds.

Further examination of the external validation set highlighted 18 chemicals reported as liver toxicants in non-rodents only. This study confirmed low cross-species concordance of liver effects (40–45 %), which is in agreement with previous investigations [41, 42]. On the other hand, it showed the reasonably good predictivity of cheminformatics techniques using data generated by automated text mining with limited manual curation. The data mining technique seems to be feasible to search for the evidence of toxicity for compounds of interest that can be used to create in silico models.

2.1.4 Statistical Models for In Vivo Specific Hepatotoxicity Endpoints Using Chemical Descriptors

Hepatotoxicity is a complex beast, a result of multiple mechanisms, many of which are still poorly understood or are not yet known. Moreover, there are various types of liver injury which can occur, such as acute and chronic hepatocellular injuries (steatosis, necrosis, cirrhosis); cholestatic injuries; neoplasia; and elevated levels of liver serum enzymes (aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP)) [8, 43]. That given, "global" modeling of general hepatotoxicity seems almost like trying to paint the Mona Lisa using only one brush with a single color. Much information would be lost. If you truly aim to be able to understand and predict hepatotoxicity with confidence, it seems logical that models should be developed for specific endpoints of liver injury initiated by a single mechanism of action. Indeed, the focus in many areas of toxicity is shifting in the direction of trying to predict single molecular initiating events (MIEs) which then, once triggered, cause a cascade of effects leading to one or more toxicity outcomes. Such information is being termed an Adverse Outcome Pathway (AOP) (*see* also Chapter 14). Indeed, an AOP specifically for liver steatosis is one such development by the Organisation for Economic Cooperation and Development (OECD) [44]. A battery of such models used in combination would provide an incredibly powerful tool.

The US FDA conducted a three-part investigation to create a human health effects database and subsequently developed QSAR models to predict the hepatobiliary (liver enzyme disorders, cytotoxic injury, cholestasis and jaundice, bile duct disorders, gall bladder disorders) and urinary tract (acute renal disorders, nephropathies, bladder disorders, kidney function tests, blood in urine, urolithiases) adverse effects of drugs. Furthermore, they described specific properties of drugs that caused these adverse effects (ID 5) [20, 45, 46]. A dataset of about 1660 chemical structures was constructed from two pharmaceutical post-market surveillance databases maintained by the US FDA: a Spontaneous Reporting System (SRS) and an Adverse Event Reporting System SRS (AERS), and from the published literature. Five specific endpoints were considered: liver enzyme disorders, cytotoxic injury, cholestasis and jaundice, bile duct and gall bladder disorders. The authors employed four QSAR modeling programs to construct predictive models and model performance was optimized by adjusting the ratio of active to inactive drug molecules in the training sets. An average sensitivity of 39.3 % and specificity of 86.5 % was obtained in the internal leave many out (LMO) validation procedure of the four programs. To improve the low sensitivity, consensus models were constructed by a combination of two programs. This resulted in an average sensitivity and specificity of 56.2 % and 78.4 %, respectively. In the external validation of 18 new drugs, which were removed from market because of serious hepatotoxicity effects, 16 compounds were predicted correctly by at least one program, but only two drugs were assigned as hepatotoxic by all four programs. These studies demonstrated that QSAR technology is a useful (albeit data-hungry) tool providing decision support information in drug discovery. However, given its multifaceted nature, prediction of hepatotoxicity remains a significant challenge and the use of multiple models in combination could be a method of increasing performance and user confidence. Moreover, the US FDA study also provided molecular insights into the mechanisms responsible for some adverse effects, and this was investigated further in the third part of this study [46].

Rogers et al. employed the US FDA Human Liver Adverse Effects Database (HLAED) containing 490 chemicals with five serum enzyme markers of liver toxicity: ALP, ALT, AST, lactate dehydrogenase (LDH), and γ-glutamyl transpeptidase (GGT) to build QSAR models using a kNN method (ID 7) [17]. Approximately 200 compounds covering a wide range of clinical data, structural similarity, and balanced (40/60) active/inactive ratios were selected for modeling and divided into multiple training/test and external validation sets. Since the kNN technique is based on interpolating activities of the nearest neighbors, it was necessary to introduce an applicability domain to avoid making predictions for compounds that differed substantially from the training set molecules [47]. Four hundred topological descriptors generated by MolConnZ (eduSoft LC, Ashland, VA) and 1664 Dragon descriptors (v.5.4, Talete SRL, Milano, Italy) were used to construct the models for the five endpoints as well as for the composite liver endpoint created from all five liver enzymes endpoints. Sensitivities >73 % and specificities >94 % were obtained in external validations. It was interesting to note that only three endpoints (ALT, AST, and the composite score) had a relatively broad coverage among the 490 drugs in the database. This is in agreement with the fact that ALT and AST are routine, widely used clinical chemistry biomarkers for liver toxicity. The examination of the applicability of these developed models, using three chemical databases: World Drug Index (WDI), Prestwick Chemical Library (PCL), and Biowisdom Liver Intelligence Module, showed low coverage. For example, 80 % of chemicals in the WDI database were outside the applicability domain of the models. The authors also verified the predictions for compounds from these three external datasets, by comparing model-based classification with reports in the publically available literature. For many compounds, the predictions could not be verified, because of the lack of reports of toxicity in the literature. This is a common problem encountered in many hepatotoxicity modeling studies. The lack of data is a limiting factor as is the questionable quality and relevance of what is available.

The model for the composite endpoint was also further validated using five pairs of structurally similar chemicals with opposing liver toxicity effects. The outcome of this external validation was equivocal. Two pairs were outside of the models applicability domain and only one pair was predicted correctly. Building on the similar experiences noted above, this may suggest that in some cases chemical mechanism(s) alone may not account for the toxic potential. It is possible in these cases that the differential toxicity may arise from metabolic transformations, complex disease pathways, or other risk factors dependent on genetic polymorphism and/or environmental conditions. This study clearly illustrates that the limitations of in silico methodologies result from their restricted applicability domains as well as a lack of understanding of the complexities of human risk factors and DILI pathways.

Liu et al. utilized the clinical and post-marketing data from the computer-readable side effect resource (SIDER) database [48] and identified 13 types of hepatotoxic side effects (HepSEs) based on MedDRA ontology, including bilirubinemia, cholecystitis, cholelithiasis, cirrhosis, elevated liver function tests, hepatic failure, hepatic necrosis, hepatitis, hepatomegaly, jaundice, liver disease, fatty liver, and liver function test abnormalities [16]. Firstly, these 13 side effects were used to discriminate drugs that do and do not cause DILI using the Liver Toxicity Knowledge Base Benchmark Dataset (LTKB-BD) [49] and the PfizerData [22]. For the LTKB-DB, classification accuracy was 91 %; for the PfizerData the accuracy was significantly lower (74 %). In the next step, using the SIDER database, QSAR models for every HepSEs were generated using a Bayesian methodology and these were then combined to form a DILI prediction system (DILIps) (ID 11). Finally, the authors implemented a "rule of three" (RO3) criterion (a chemical being positive in at least three HepSEs) into DILIps which increased classification accuracy. The predictive performance of DILIps was examined using three external databases: LTKB-DB, PfizerData and a dataset published by O'Brien et al. [50] and yielded prediction accuracies of 60–70 %.

Liu et al. also applied the RO3 criterion to drugs in DrugBank to investigate their DILI potential in terms of protein targets and therapeutic categories. Two therapeutic categories showing a higher risk for causing DILI were identified (anti-infective for systemic use and musculoskeletal system drugs). These findings are consistent with current knowledge that most of the anti-infective drugs are very often associated with liver injuries. One hundred thirty-four protein targets related to drugs inducing liver toxicity have been identified using pathway analysis and co-occurrence text mining with most of these targets being associated with multiple HepSEs. This study provides an interesting example of the translation of clinical observations into an in silico tool which can be used to screen and prioritize new drug candidates or chemicals and to avoid those that might cause hepatotoxicity.

In recent years, a number of new initiatives and international projects have been undertaken to develop in silico models to predict the harmful effects of chemicals to humans considering different endpoints such as liver injury. One such example is the COSMOS project [51] (belonging to the larger research initiative—SEURAT-1). The main aim of COSMOS is to develop publically available tools and workflows to predict the safety to humans following the use of cosmetic ingredients. Among them is the development of computational methods to evaluate the potential of chemicals to bind to liver X receptor (LXR), activation of which leads to liver steatosis (ID 15) [52]. Using different techniques such as molecular modeling to assess the LXR binding potential and applying PaDEL or RDKit descriptors, QSAR models based on Partial Least Squares Discriminant Analysis (PLS-DA) were developed and implemented into the freely available KNIME Platform [52]. These models, used together with the molecular modeling methods and structural alerts as discussed within this chapter, are forming integrated in silico strategies for screening of potential steatosis inducers.

2.1.5 Statistical Models for In Vitro Specific Hepatotoxicity Endpoints Using Chemical Descriptors

Only one in silico model (ID 2) has been found that predicts in vitro specific hepatotoxicity endpoints measured by cell proliferation, lactate dehydrogenase (LDH) for membrane integrity, intracellular ATP levels for cell vitality, and levels of caspases 3 and 7 for cell apoptosis [11]. The authors applied molecular interaction fields (Idiotropic Field Orientation for Comparative Molecular Field Analysis (IFO-CoMFA)) as structural descriptors and Soft Independent Modeling of Class Analogy (SIMCA) to classify the hepatotoxicity of 654 drugs from the Sigma-RBI Library of Pharmaceutically Active Compounds (LOPAC) [11]. Each of the four assays showed good discrimination between the toxic and nontoxic chemicals. The greatest accuracy of 52 % was obtained for a hierarchical ranking model, which combined all four assays (again demonstrating that ensemble/consensus models show promise). A significant improvement in predictive performance (accuracy of 88 %) was achieved with a model constructed for a set of 27 nonsteroidal anti-inflammatory drugs (NSAIDs) using data from the LDH assay. The cross-validation confirmed the good performance of this model giving an accuracy of 71 % and 83 % for a training set of 21 NSAIDs and a test set of six NSAIDs, respectively. The poor predictivity of the global IFO-SIMCA approach for the large, diverse dataset of biologically active compounds and significant improvement for single pharmacological class chemicals' model showed that for endpoints based on specific cytotoxicity indicators only models for closely related class of chemicals may be useful. This possibly indicates that they are applicable only to a single mechanism of action within structurally related compounds. This is the main limitation of this approach, as it constricts the applicability of the model. However, local models such as this often demonstrate superior levels of predictivity, hence are useful in limited chemical space.

2.1.6 Statistical Models for In Vivo General Hepatotoxicity Using Hybrid Descriptors

Significant progress has been made in analytical and biomedical techniques in recent years which has resulted in the development of hundreds of new high-throughput screening (HTS) assays. The US Environment Protection Agencies (EPA's) Toxicity Forecaster (ToxCast) program uses these HTS assays to screen environmental chemicals for bioactivity [53–55]. Within two phases of this program, 1057 chemicals were measured using more than 800 HTS assay endpoints including biochemical assays, cell-based assays, cell-free assays, and multiplexed transcription reporter assays. These data provide valuable information about the molecular mechanism(s) of toxicity and help to identify the pathways related to adverse effects. Three studies using both chemical and biological descriptors have been identified. The main objective of these studies was to investigate if chemical descriptors and biological descriptors could be complementary in the prediction of hepatotoxicity.

One of the first studies applying chemical and biological descriptors to develop models for hepatotoxicity was conducted by Low et al. (ID 9) [18]. In contrast to many other in silico studies, the authors utilized only the animal data obtained from subchronic (28 days of treatment) assay in rats for 127 drugs studied in the Japanese Toxicogenomics Project [56]. The chemical was assigned as a liver toxicant if it exhibited histopathological characteristics of hepatotoxicity. Conversely, a compound was deemed non-hepatotoxic if it did not result in adverse histopathological features. When the observations were inconclusive, serum chemical indicators including ALT, AST, ALP, TBL, and gamma-glutamyl transpeptidase (GGT) were considered. The authors built conventional QSAR models using only chemical descriptors. They then applied toxicogenomic data to differentiate the hepatotoxins from non-hepatotoxins and finally hybrid hepatotoxicity classifiers were developed. For modeling purposes, statistical methodologies including: kNN, SVM, RF and Distance Weighted Discrimination (DWD) were applied using internal and a fivefold external cross-validation. The evaluation of predictivity showed that the accuracy of QSAR models based on chemical descriptors was generally poor (55–61 %). Conversely, models employing 85 selected toxicogenomics descriptors showed significantly improved predictive performance with accuracies as high as 76 %. The authors examined the spatial distribution of compounds in their chemical and toxicogenomics descriptor space which showed that 50 % of structurally similar pairs of compounds had opposing toxicities. On the other hand, amongst pairs of compounds with the most similar gene expression profilers, only 23 % exhibited opposing toxicity. It shows that pairs of compounds with similar gene expression profiles are more likely to have the same hepatotoxicity potential than pairs of chemically similar compounds. Of note here is that when hybrid models, combining both chemical and biological descriptors, were constructed they demonstrated similar accuracy (68–77 %) to those models based only on toxicogenomics data but the use of both chemical and biological descriptors provides additional insights into understanding DILI. The study confirmed that hepatotoxicity is a very complex endpoint and cannot be predicted effectively based only on the chemical characteristics of drugs. Such hybrid models look very promising as predictive and prioritization tools and allow for a better understanding of the mechanisms of hepatotoxicity.

A second study employing hybrid descriptors was conducted by Zhu et al. (ID 13) [19]. The authors constructed models based on chemical descriptors and in vitro cell-imaging information taken from human hepatocyte imaging assay technology (HIAT) that measures the intensity of biochemical indicators, such as lipids, glutathione (GSH), reactive oxygen species (ROS) [40]. The models were built based on a dataset of 292 diverse chemicals (156 positive) using RF and fivefold cross validation methodologies. For each model the applicability domain was defined to control the distance between the predicted compound and its closest neighbor in the dataset. The main purpose of this research was comparing the prediction performance of models with a single type of descriptor (chemical or HIAT) with hybrid models. The hybrid models were constructed by combination of HIAT descriptors with chemical descriptors calculated using three programs (CDK-HIAT, Dragon-HIAT, and MOE-HIAT). These three hybrid models were combined into a consensus model. The models with chemical descriptors alone showed the poorest predictivity with accuracies between 57 % (for CDK descriptors) and 63 % (for MOE descriptors). Similar to the study conducted by Low et al. [18], this research confirmed that structural properties alone are incapable of capturing the complex mechanisms of liver toxicity. The highest accuracy (77 %) and specificity (87 %) were obtained from the HIAT model. However, the consensus hybrid model showed the greatest sensitivity (74 %). Since the HIAT model had the highest specificity and consensus model-best sensitivity, both models were applied together to distinguish liver toxicants from nontoxic chemicals. Ninety-eight of 158 DILI-inducers and 96 of 136 non-inducers were predicted correctly by both models. Careful investigation of the 39 false negative compounds revealed that at least three types of mechanisms are not captured by the models: (1) drugs that may cause liver toxicity only in high dosage, e.g., naltrexone; (2) metabolic activation, e.g., tianeptine; and (3) blockage of bile secretion, e.g., norethindrone. Ideally, QSAR models should be mechanistically interpretable to help understand the underlying mechanisms of toxicity. In this study, the distribution of molecular fragments among the toxic and nontoxic chemicals was investigated together with the analysis of biological descriptors. Forty-seven molecular fragments showed a significantly higher probability of being present in DILI-inducers than in non-inducers. Most of these fragments were associated with amine-derivatives, aromatic rings and alkyl chloride fragments. Furthermore, three of HIAT descriptors: the tetramethylrhodamine methyl ester (TMRM) intensity, ROS and a reduced intracellular GSH level were ranked as the most important indicators of DILI. These findings proved, for example, that the redox cycling of nitroaromatic drugs can generate reactive oxygen species (represented as ROS intensity HIAT descriptor) which are indicators of oxidative stress in hepatocytes. A further HIAT descriptor, TMRM, is an indicator of mitochondrial abnormality which can generate superoxide and damage endogenous macromolecules. This study showed that chemical and biological descriptors can be complementary and enhances the prediction accuracy of hepatotoxicity and can aid in rational mechanistic interpretation.

2.1.7 Statistical Models for In Vivo Specific Hepatotoxicity Endpoints Using Hybrid Descriptors

A recent study conducted by Liu et al. utilized the in vitro bioactivity data from ToxCast together with chemical structure descriptors for 677 chemicals to predict in vivo hepatotoxicity (ID 14) [29]. Of the 677 compounds, 214 were classified as hepatotoxic based on rat liver histopathological observations in chronic studies and were categorized into three hepatotoxicity groups: (1) hypertrophy (161), (2) injury (101), and (3) proliferative lesions (99). The remaining 463 chemicals were classified as non-hepatotoxic. The authors built the models using six machine learning algorithms: LDA, NB, SVM, classification and regression trees (CART), kNN, and an ensemble of these classifiers (ENSMB). Three types of descriptors were used to build the models: 726 chemical descriptors from QikProp, OpenBabel, PaDEL, and PubChem; 125 ToxCast HTS bioactivity descriptors and hybrid descriptors (the combination of chemical and bioactivity descriptors). Because of the skewed ratio of positive to negative chemicals in every hepatotoxicity category, undersampled, balanced datasets have been prepared: 160 positive and negative chemicals for hypertrophy, 100 positive and negative chemicals for injury, and 90 positive and negative chemicals for proliferative lesions. For each of the three categories, classifiers of hepatotoxicity were built using imbalanced and balanced datasets for three types of descriptors: chemical, bioactivity, and hybrid. Predictive performance was evaluated using tenfold cross-validation and repeated 100 times. For each step in the cross-validation loop, the subset of best descriptors was filtered. The best predictive accuracy for hypertrophy (84 %), injury (80 %) and proliferative lesions (80 %) was obtained for hybrid descriptors. Using undersampled balanced datasets improved the sensitivity, but reduced the specificity of classifiers compared to the imbalanced datasets.

In general, classifiers with bioactivity descriptors have better specificity than models with chemical descriptors only, but have lower sensitivity. However, the best predictive statistics in terms of balanced accuracy, sensitivity and specificity were obtained for hybrid classifiers for both balanced and imbalanced datasets. This study showed that using both types of descriptors is more relevant for building predictive models, since they reflect the synergies between structural features, molecular mechanisms and cellular functions. The interpretation of these selected descriptors is important for the understanding of underlying mechanisms of hepatotoxicity and can help to establish the adverse outcome pathways (AOPs) as highlighted previously in this chapter. The analysis of the descriptors suggested that the classifiers may be related to AOPs initiated by the pregnane X receptor (PXR), farnesoid X receptor (FXR), and vitamin D receptor (VDR). Overall, this study demonstrates the usefulness of HTS assays for characterizing the in vivo hepatotoxicity and the benefit of using both types of descriptors reflecting bioactivity and chemical structure.

2.1.8 Statistical Models Summary

The performance of statistical models generally suffers when predicting complex toxicity endpoints such as hepatotoxicity, a phenotype with multiple complex mechanisms and many that remain unknown. This literature review of the existing statistical models for predicting hepatotoxicity has confirmed that there is no easy solution to the problem of correctly identifying hepatotoxins. The shortage of reliable data, the lack of sensitive biomarkers and the multifaceted nature of hepatotoxicity itself, all serve to complicate an already complex problem. Since hepatotoxicity is so complex a phenomenon, it could not be predicted with high confidence based solely on the structural properties of the chemicals. It was found that the application of both chemical and biological information together and modeling specific endpoints of liver injury, initiated by a single mechanism of action rather than the effect as a whole, can significantly improve the identification of potential hepatotoxins. Moreover, multiple studies showed that the ensemble methodology that combines different models had improved the final performances when compared with the best performing individual model.

2.2 Qualitative (Expert Knowledge-Based) Models

In contrast to the quantitative models discussed up to this point, a number of qualitative approaches have also been explored. These are summarized later in this section by Table 2 following the discussion of these models.

2.2.1 Development of Structural Alerts

The development of structural alerts has been an area of considerable interest in recent years. Their transparency and ability to incorporate (or elucidate) mechanistic information offers an advantage over other, statistically derived, approaches.

2.2.1.1 Egan et al. (2004): Structural Alerts for Hepatotoxicity

Over a decade ago, Egan et al. provided an excellent review of in silico methods to predict various aspects of drug safety (ID 1 in Table 2) [5]. The authors own contribution to this review was the development of a structural alert-based approach for the prediction of liver toxicity. From a dataset of 244 drugs (54 of which were withdrawn from the market or abandoned during development owing to hepatotoxicity) a series of 74 computational alerts were developed. These alerts were based on an extensive review of the literature and were often accompanied with mechanistic reasoning for their observed hepatotoxicity. It is interesting to note that 56 of the 74 alerts were based on functional groups and were related to the formation of reactive (or otherwise toxic)

Table 2
Table summarizing expert knowledge-based models for liver toxicity

ID	Endpoint	Type and size of data	No. of structural alerts	Validation	Predictive performance	Ref
1	Hepatotoxicity	In vivo human data for 244 compounds	74 developed	No data	No data	[5]
2	Hepatotoxicity	In vivo data for 1266 compounds	38 developed	External validation using 626 chemicals	SEN (46 %), SPE (73 %), and ACC (56 %)	[22]
3	Hepatotoxicity	In vivo human data for 951 compounds	16 developed	N/A	N/A	[30]
4	Hepatosteatosis	PDB and ChEMBL	N/A	Validation using the 251 ChEMBL compounds and 951 Fourches et al. dataset	N/A	[28]
5	Hepatotoxicity	In vivo human data for 577 compounds	12 molecular fragments	Not reported	Not reported	[29]
6	Steatosis	Pharmacophore built on the three most active agonists	None—pharmacophore model	External validation using a test set of 21 agonists	N/A	[27]

metabolites. The remainder were based on whole molecule similarity and were more complex, often with limited or no mechanistic rationale. No attempt was made here to assess their predictive performance since the authors aim was to extract and investigate structural alerts for hepatotoxicity.

Unlike the statistical models in the previous section of this chapter, qualitative methods such as structural alerts are not statistically derived models. In fact, they should not be considered as “models” at all. They serve as a direct link showing that a particular molecular fragment/feature is associated with observed hepatotoxicity. No quantitative measure is provided. Interest in structural alerts is increasing. Since they are developed in an evidence-based manner and may contain mechanistic information, they are completely transparent and user confidence in their application is generally higher than that of statistical models.

This is not to say that structural alerts are simple to generate. Each structural alert must be carefully defined. Too general in nature and it will be flagged up in almost all compounds and will not differentiate toxicity classes. Too specific (rigid) may restrict its application to a single compound and not extend to derivatives containing the actual fragment initiating the toxicity. All of this, coupled with the need to research and define mechanistic rationale makes structural alert definition a complex and time-consuming task.

Irrespective of their origins, the beauty of structural alerts is that they can be coded into computational systems which allow for rapid screening of compound libraries. Egan et al. packaged the knowledge extracted from the literature, linked this to defined structural alerts and developed a system capable of making mechanistically supported predictions of likely hepatotoxicity in humans.

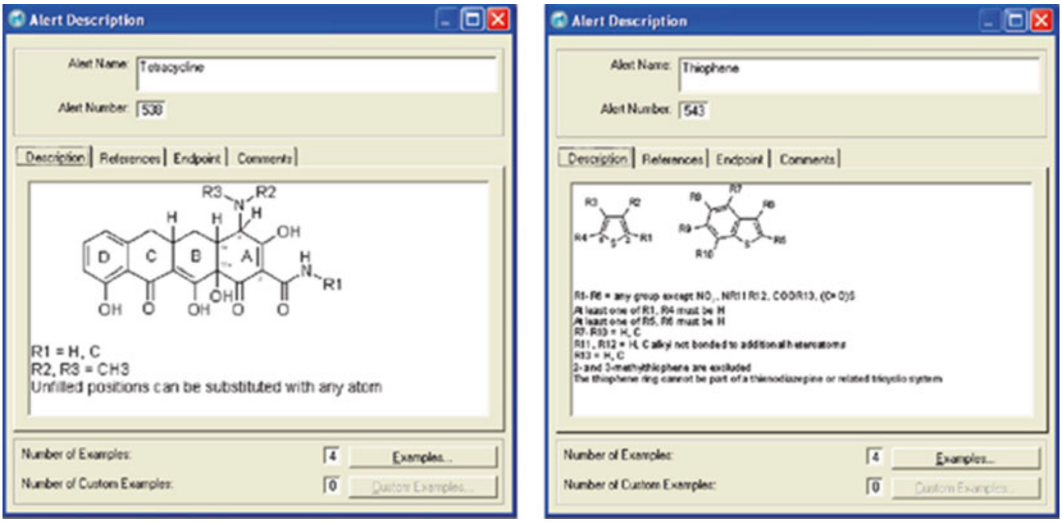
2.2.1.2 Greene et al. (2010): The Interest in Structural Alerts Grows

Green et al. further develop the concept of generating structural alerts for hepatotoxicity (ID 2) [22]. The authors highlight the presence of Derek for Windows (DfW), a commercial prediction system developed by Lhasa Ltd. [56]. In recent years this has been rebranded as Derek Nexus as already introduced in Chapter 10. This knowledge-based expert system emulates human reasoning and utilizes the approach described by Egan et al. [5] to make predictions based on structural alerts and associated mechanistic knowledge. Version 8 of this software contained structural alerts for several endpoints, many of which were well established (e.g., carcinogenicity). However, at the time this study was performed, only two structural alerts for hepatotoxicity were present in DfW's knowledgebase.

Green et al. highlighted this shortfall and published a study aimed at developing a number of additional structural alerts. Importantly, this study investigated whether it is possible to use publically available data to develop structural alerts for hepatotoxic potential. This study goes into some detail of how a dataset of known hepatotoxins was divided into various chemical/therapeutic classes. This article also starts to introduce the concept of using structural similarity to generate structural alerts from clusters of structurally related compounds.

Thirty-eight new structural alerts were identified in this study based on human and/or animal data. Each was incorporated into a customized version of DfW (*see* Fig. 2) together with supporting examples and mechanistic information gathered from the literature. Importantly, these alerts were externally validated using a large Pfizer-developed dataset of 626 compounds (*see* Fig. 3 for examples of compounds containing identified alerts). The predictive performance of these alerts in the customized DfW knowledge base are summarized in Table 2.

The importance of developing structural alerts and embedding these into a tool such as DfW is clear. SARs in the form of structural alerts for complex endpoints can be elucidated from the open literature. The additional support of case studies and mechanistic rationale extracted from the literature is where a structural alert approach



Alert Description

Alert Name: Tetracycline

Alert Number: 538

Description | References | Endpoint | Comments
R3
R2
N
H
H
H
OH
D
C
B
A
H
N
R1
OH
OH
O
OH
O
O
O

R1 = H, C
R2, R3 = CH3
Unfilled positions can be substituted with any atom

Number of Examples: 4

Number of Custom Examples: 0

Examples...
Custom Examples...

Alert Description

Alert Name: Thiophene

Alert Number: 543

Description | References | Endpoint | Comments
R3
R2
R4
R1
R7
R8
R6
R9
R5
R10

R1- R6 = any group except NO_2, NR11R12, COOR13, ((O=O)S
At least one of R1, R4 must be H
At least one of R5, R6 must be H
R7- R10 = H, C
R11, R12 = H, C alkyl not bonded to additional heteroatoms
R13 = H, C
2- and 3-methylthiophene are excluded
The thiophene ring cannot be part of a thienodiazepine or related tricyclic system

Number of Examples: 4

Number of Custom Examples: 0

Fig. 2 Example alert describing SARs developed for tetracyclines and thiophenes. Reprinted with permission from Green et al. Chem. Res. Toxicol. 23, 1215–1222. Copyright 2015 American Chemical Society

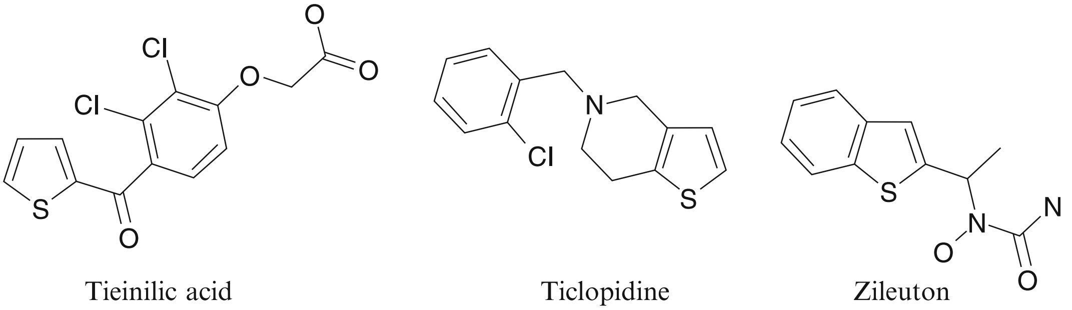


Fig. 3 Drugs containing a thiophene ring and associated with hepatotoxicity. Reprinted with permission from Green et al. Chem. Res. Toxicol. 23, 1215–1222. Copyright 2015 American Chemical Society

differs from traditional quantitative modeling. As a screening tool, a prediction along with transparent supporting evidence is very powerful. Of course, at the same time, the approach of developing structural alerts in this manner drives research into mechanisms of liver toxicity and injury which is of equal importance.

2.2.1.3 Hewitt et al. (2013): A Scheme for Generating Structural Alerts for Human Hepatotoxicity

Driven by the continued need to predict hepatotoxicity and the growing utilization of structural alerts, our contribution to this area has been in the development of a general scheme for structural alert development (ID 3) [30]. Focusing purely on publically accessible data, our aim was to develop an approach (using freely available tools) capable of yielding mechanistically supported structural alerts as previously described [5, 22]. Given the scarcity of high quality hepatotoxicity data, the broad spectrum of possible endpoints to consider and the complex nature of the mechanisms involved, defining such alerts is a considerable challenge. Furthermore, our focus was set solely on predicting human hepatotoxicity utilizing compiled clinical data for 951 structural diverse compounds. Given that hepatotoxicity is often not evident until identified during post-marketing surveillance, it seems logical to conclude that current histopathological liver findings in rats do not model the idiosyncratic effects seen in humans [41, 42]. Conversely, Lhasa Ltd. (the developers of Derek Nexus) recently presented a poster showing that the alerts available in Derek Nexus which are developed using human data cannot predict the liver findings in rats [57].

In our study, structural similarity scores were used to highlight chemical categories of structurally related (and hepatotoxic) compounds (using the freely available Toxmatch software [58]). Eighty-two such categories were identified and each was manually inspected for validity. Following this validation step, 16 unique structural categories were identified and researched in detail to propose a mechanistic rationale. The common structural fragment of each category was extracted and taken to be the structural alert for that class. Each alert was further validated by using that structural alert to repopulate the original category. Examination of the resulting hits proved useful in highlighting alerts that were too general or restricted in terms of their definition.

An example of an alert generated from a small chemical category (Table 3) is shown in Fig. 4. This category contains a number of phenothiazine derivatives commonly used as antipsychotics. The common structural fragment was extracted and formed the structural alert as shown in Fig. 4. Searching the literature for a mechanistic rationale to explain the observed hepatotoxicity for this chemical class quickly revealed multiple implications in mitochondrial toxicity (*see* Hewitt et al. for more details). As was often the case, categories contained one or more members which were recorded as non-hepatotoxins. Here, perphenazine was classified as such in the Fourches et al. dataset. However, further literature

Table 3
Showing the category members formed using structural alert 6 (depicted)

Compound	Hepatotoxicity
Chlorpromazine	Positive
Perazine	Positive
Perphenazine	Negative
Prochlorperazine	Positive
Thioridazine	Positive
Triflupromazine	Positive

(*See* also Fig. 4)

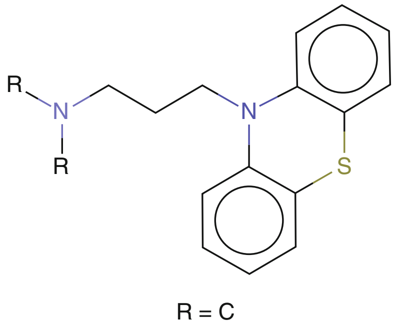


Fig. 4 Showing the category members formed using structural alert 6 (depicted) (*see* also Table 3)

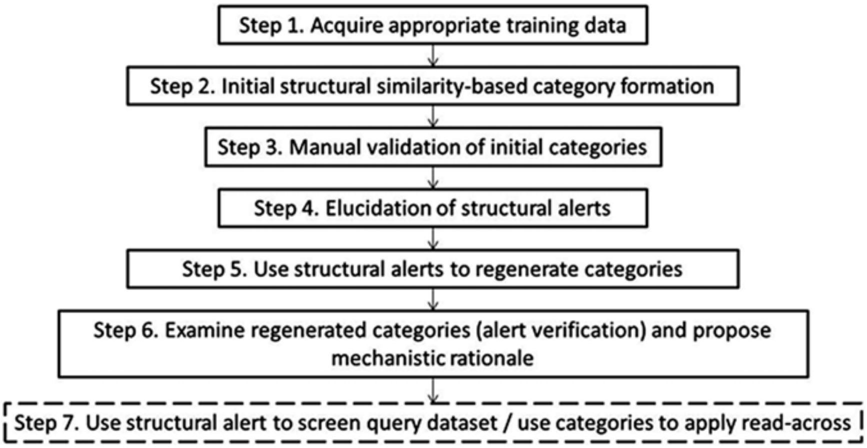

Step 1. Acquire appropriate training data
Step 2. Initial structural similarity-based category formation
Step 3. Manual validation of initial categories
Step 4. Elucidation of structural alerts
Step 5. Use structural alerts to regenerate categories
Step 6. Examine regenerated categories (alert verification) and propose mechanistic rationale
Step 7. Use structural alert to screen query dataset / use categories to apply read-across

Fig. 5 Strategy for the development of structural alerts for the prediction of hepatotoxicity (taken with permission from Hewitt et al. [30])

searching suggested this to be an incorrect classification since perphenazine has been associated with liver effects in humans.

As such, this is not solely a process of extracting knowledge from a given dataset, but acts to highlight instances where the literature can be used synergistically to support and extend our current knowledge.

The aim of the article by Hewitt et al. was not to create a comprehensive suite of hepatotoxicity alerts, but to develop and publish a generic scheme for their development using freely available tools. Given the limitations of publically assessable data and our incomplete understanding of hepatotoxicity, developing a system sufficiently capable of predicting hepatotoxicity in humans is a herculean task. A dynamic scheme such as that proposed by Hewitt et al., updated regularly with new data leading to new alerts and renewed mechanistic understanding, is likely to be the most productive approach.

The general 7-step strategy proposed in this work is summarized in Fig. 5. As with all modeling approaches, the first step is to acquire an appropriate dataset suitable for modeling (defined chemical structures, clear toxicity annotations, etc.). The second step is to form groupings of structurally related compounds (often termed chemical categories). A manual validation step is then required in order to remove any duplicate categories or those exhibiting too wide a range of chemical diversity. Step 4 is when each category is inspected and a common structural feature is identified. This feature becomes the structural alert. In order to assess the selectivity of the alerts generated, step 5 involves using these alerts to screen the original dataset. Step 6 then examines the resulting category members (which may contain compounds with the alert but not previously assigned to the category) This stage quickly highlights alerts that are too general in nature since the repopulated category tends to contain multiple new compounds (many of which often demonstrate no toxicity). If developed well, this category adds a supportive element to the alert demonstrating a category of example toxic compounds. The second stage of step 6 adds mechanistic support to the structural alert. Each alert (and its category members) is investigated in detail to define or propose a mechanistic basis for the toxicity observed. This stage is time consuming with no guarantee of success, but in most cases mechanistic rationale could be identified and this gives a much greater weighting (and user confidence) in their use. The final step proposed in the Hewitt et al. article (step 7) highlights that, at this stage, the structural alerts are read to be used to screen query datasets. Furthermore, it is stressed that the chemical categories themselves should not be forgotten and have a potential role in read-across; a process whereby measures of structural similarity can be used to match a query chemical to those in a library. These reference compounds (or category members) can then be used to estimate the properties/toxicity of the query compound based on their similarity.

As with the study by Greene et al., the power of structural alerts is their ability to be built into a platform capable of screening large numbers of compounds for the presence of each alert. The 16 alerts developed in his study were combined into a predictive tool and were made available on the predictive modeling platform developed within the eTOX Project [59]. Here, the structural alerts were coded as SMARTS and were incorporated into the KNIME platform [52]. This automated the screening procedure and allowed for an input file to be uploaded and rapidly screened.

2.2.1.4 Steinmetz et al. (2015): Focusing the Search

Working as part of the COSMOS Project, Steinmetz et al. (ID 4) [28] employed a slightly different approach to the problem. Instead of elucidating structural alerts and then investigating their mechanism(s) of action, they began with a known mechanism of interest (interaction with the retinoic acid receptor (RAR) which has been linked with liver steatosis) (It is interesting to note that the retinoid class was previously highlighted as a structural alert in

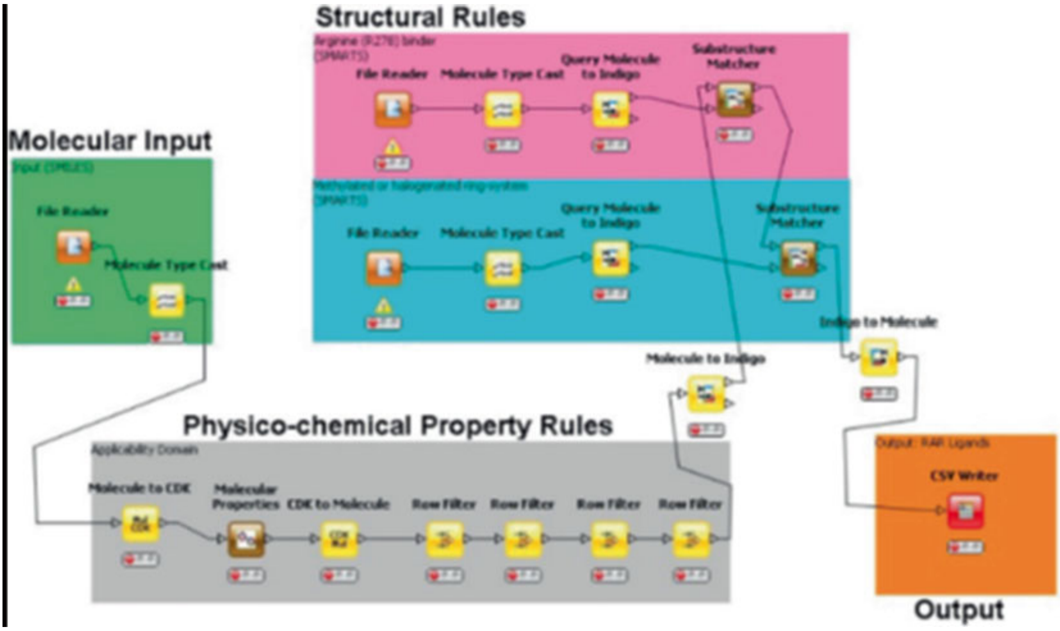


Fig. 6 KNIME workflow developed by Steinmetz et al. to predict RAR ligands (taken with permission from Steinmetz et al. [28])

Hewitt et al. [30].) Subsequent analysis then solely focuses on known RAR ligands to identify structural alerts for this mechanism of action. This is synonymous with the local versus global modeling approaches previously discussed with regards to the statistically derived models (multiple versus single mechanisms of action).

In contrast to previous works, Steinmetz at al. combined a small number of structural alerts together with a set of physicochemical property filters to highlight potential RAR ligands. These filters were based on the physicochemical characteristics of the known RAR ligands considered in the study.

Again, predictions were made via the development of a KNIME workflow containing the alerts as well as automated physicochemical property calculations and filters (*see* Fig. 6). The KNIME workflow then acts as a very powerful screening tool able to identify potential RAR ligands.

2.2.1.5 Liu et al. (2015): Boosting the Validity of Structural Alerts

The most recent example of structural alerts for human liver toxicity at the time of writing this chapter was an article by Liu et al. (ID 5) [29]. Their focus was on the validity of structural alerts. As stated in the article, a limitation of employing libraries of structural alerts is that they will effectively reduce the chemical space available for new drug discovery. Liu et al. highlight that more than half of the oral drugs currently on the market match to one or more structural alerts published for hepatotoxicity, suggesting that these alerts are either too general in their design or they are failing to take into account other factors, such as metabolism. They go on to discuss the development of robust, statistically validated, structural alerts.

In the publication of Hewitt et al., structural alerts were often developed using categories containing both hepatotoxic and non-hepatotoxic compounds. The conflicting "non-hepatotoxic" compounds could often be rebuttled following detailed literature searches suggesting these classifications to be false. Furthermore, with the dataset considered in the Hewitt et al. study, the absence of clinical reports for hepatotoxicity lead to a non-hepatotoxic classification.

Liu et al. proposed to ensure the relationship of alert and toxicity using a statistical approach (utilizing *p* values) to highlight the robustness of this relationship in a quantitative manner. Alerts based on categories containing nontoxic compounds will therefore show reduced statistics and less robustness than those based solely on toxic compounds. However, as mentioned previously, it is important to ensure the validity of the nontoxic classification before proceeding in this manner.

2.2.2 Development of Pharmacophore Models

As introduced earlier in this chapter, the development of pharmacophore models is another qualitative approach to the prediction of hepatotoxicity. It is important to stress from the outset that pharmacophore models, depending upon how they are utilized, can provide quantitative information. Pharmacophore models can be seen to extend the theory of structural alerts and transform the two-dimensional representation of a structural alert into a three-dimensional scaffold, overlaid with information of important physicochemical features. (This is not to be confused with chemotypes which are effectively two-dimensional structural alerts with encoded physicochemical data).

2.2.2.1 Tsakovska et al. (2014)

Tsakovska et al., partners in the COSMOS Project, recently published a pharmacophore study focussing on a particular mechanism of action thought to be a key factor in the elucidation of liver steatosis (ID 6) [27]. As in the Steinmetz et al. study, efforts are focused onto a single mechanism of action, in this case concentrating on the activation of the peroxisome proliferator-activated receptor gamma (PPARγ).

A pharmacophore model was developed following analysis of the interactions between PPARγ and the three most active full agonists (rosiglitazone and two compounds termed compound 544 and 570). The pharmacophore was evaluated using a dataset of full agonists and the pharmacophore features were evaluated.

The structure of one of the full PPARγ agonist (rosiglitazone) is shown in Fig. 7.

The three most active agonists are aligned on top of one another to define the characteristics of the PPARγ pharmacophore (Fig. 8). In this study, four polar atoms and functional groups

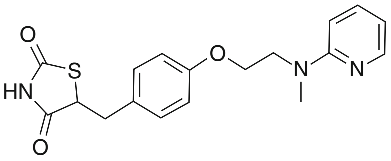


Fig. 7 Structure of rosiglitazone

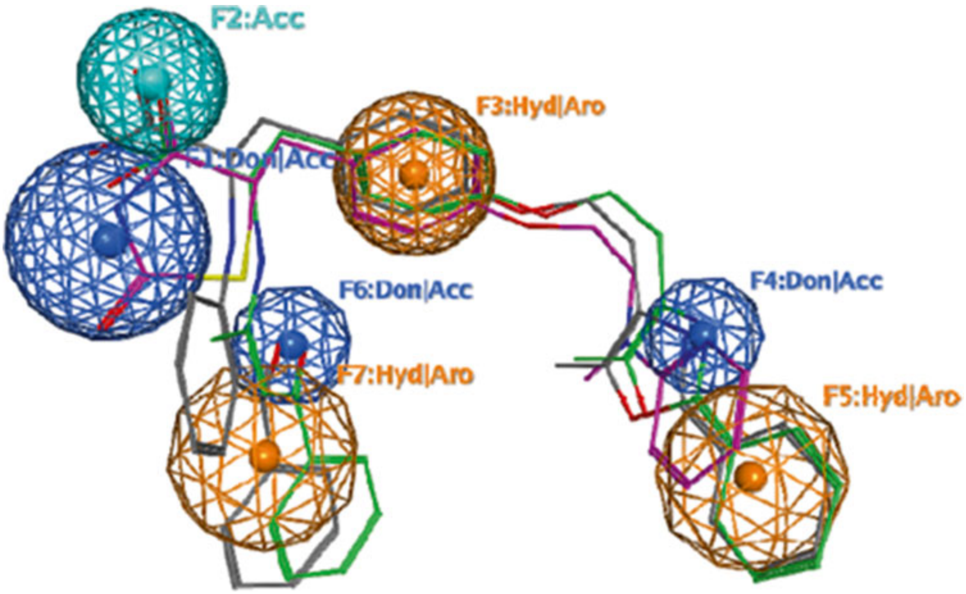


Fig. 8 Pharmacophore model of PPARγ full agonists (rosiglitazone, carbon atoms in *magenta*; compound 544, carbon atoms in *green*; compound 570, carbon atoms in *grey*) (taken from Tsakovska et al. [27])

capable of performing hydrogen bonding and ionic interactions (F1, F2, F4 and F6) and three hydrophobic and aromatic features (F3, F5 and F7) were determined to be important pharmacophore features of the most active agonists. The role of each feature and its interactions within the binding region of PPARγ are then considered.

This scaffold can be used to screen libraries of compounds for likely PPARγ binders. In its most simplistic form, the presence/absence of each pharmacophore feature can be used to predict activity. More complex application included assessment of the three-dimensional positioning of these features and the interactions these have with the PPARγ complexes.

Pharmacophore models extend beyond structural alerts in their ability to tease out information relating to the binding interactions between receptor and ligand. As such, if a particular interaction is known to be a prerequisite for activity, it can be explored and extended to find other groups/molecules which possess this ability. They therefore have a significant role in the drug development process given their possible applications in rational drug design.

3 Fitting Together the Different Pieces of the Puzzle and Future Directions

The mechanisms by which a compound can elicit toxicity to the liver are complex and diverse in nature. Attempting to then predict the hepatotoxicity of a new compound using a single approach is a very difficult task. It has already been seen that, on multiple occasions, authors have combined not only model predictions, but also model types in search of better and more reliable hepatotoxicity prediction [10, 24].

An emerging theme from all of these studies is that individual models have differing abilities to predict hepatotoxicity within a defined region of chemical space. As such, it is unlikely that a single model will ever be able predict such a complex endpoint as hepatotoxicity. Further integration of available datasets, mechanistic insights and available models for DILI is likely the only way to increase both prediction accuracy and application across chemical space. A system combining quantitative statistically derived models, structural alerts and pharmacophore models each bringing strengths (and weaknesses) is an exciting prospect and something that should be further explored. It is foreseeable that mechanistically based structural alerts could be used to screen large databases and populate a define category relating to a single mechanism of action. Local QSAR models could then be developed on this subset of data based on relevant descriptors. Moreover, it has been shown that most predictive methods discussed are based solely on descriptors of chemical structure and properties. Consideration and inclusion of biological information, such as toxicogenomics, can further help detect potential liver toxicants. Such biological descriptors may also provide further insights in the mechanisms at play in liver toxicity.

One of the major limitations currently is the lack of high quality hepatotoxicity data. To improve the prediction of potential hepatotoxins more effort should be focused towards developing specific and sensitive biomarkers for DILI. If this were possible, it would lead to more reliable hepatotoxicity data which then can be used for developing models to predict DILI. Similarly, a more detailed understanding of the mechanisms of liver injury would be of tremendous benefit and may invert the current approach of modeling with the subsequent addition of mechanistic reasoning. If we could better understand a causal mechanism of DILI (again relating to AOPs), perhaps we could design a model/alert based purely on the mechanism (e.g., what are the characteristics a chemical must possess in order to trigger mitochondrial toxicity?). These characteristics can then be used for screening and possibly further structural alert generation.

The generation of predictive systems for liver toxicity is rapidly gaining pace. With emerging modeling methods, technologies and advances in all areas of science, it is likely that we are standing on the precipice of a modeling explosion. Careful consideration must now be made in how best to manage this emerging knowledge to best effect. In recent years, many regulatory agencies, institutions, EU Projects and working groups have established programs to help understand and detect DILI. These include the Virtual Liver Project (v-Liver™) established by US EPA [60], the Drug-Induced Liver Injury Network (DILIN) set up by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) in the USA [61], the Virtual Liver Network project initiated by the German Federal Ministry for Education and Research [62], and multiple EU Projects such as Mechanism based Integrated systems for the prediction of Drug Induced Liver Injury (MIP-DILI) [63]. Whilst duplication of effort is inevitable to some degree, what must be ensured is that both data and knowledge generated by these initiatives is shared. Just as combining models to form an ensemble seems to be beneficial for predictive performance, it is likely that a combined international ensemble effort is the only way we can successfully begin to tackle the prediction of liver toxicity in humans.

4 Conclusions

Hepatotoxicity has been a problem for many years. Unfortunately, the same can also be said for predictive models aimed at predicting these effects. It is only in the past decade that models/systems for predicting hepatotoxicity have started to emerge. It is fair to say that the modeling community are currently limited by the amount and quality/reliability of the data available to them. Coupled with an endpoint as complex as hepatotoxicity, the scale of the challenge is obvious. That said, it can be seen from the models discussed in this chapter that progress is being made, our knowledge of the processes behind liver toxicity is growing and our ability to tackle this problem is increasing. Given the diversity of the modeling approaches seen in these studies and the general transition towards ensemble/consensus approaches in this area, it is likely that the next decade will be equally as productive.

Acknowledgement

The funding provided by the eTOX project, grant agreement number 115002, under the Innovative Medicines Initiative Joint Undertaking (IMI-JU), the European Community's Seventh Framework Program (FP7/2007-2013) COSMOS project under grant agreement number 266835, and from Cosmetics Europe is gratefully acknowledged.

References

1. Przybylak KR, Cronin MTD (2012) In silico models for drug-induced liver injury—current status. Expert Opin Drug Metab Toxicol 8:201–217
2. Schuster D, Laggner C, Langer T (2005) Why drugs fail—a study on side effects in new chemical entities. Curr Pharm Des 11:3545–3559
3. Holt MP, Ju C (2006) Mechanisms of drug-induced liver injury. AAPS J 8:E48–E54
4. Kaplowitz N (2005) Idiosyncratic hepatotoxicity. Nat Rev Drug Discov 4:489–499
5. Egan WJ, Zlokarnik G, Grootenhuis PDJ (2004) In silico prediction of drug safety: despite progress there is abundant room for improvement. Drug Discov Today 1:381–387
6. Patlewicz G, Dimitrov SD, Low LK et al (2007) TIMES-SS-a promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. Regul Toxicol Pharmacol 48:225–239
7. Benigni R, Bossa C (2008) Structure alerts for carcinogenicity, and the salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res 659:248–261
8. Zimmerman HJ (1999) Hepatotoxicity: the adverse effects of drugs and other chemicals on the liver. Lippincott Williams & Wilkins, Philadelphia, PA
9. Li AP (2002) A review of the common properties of drugs with idiosyncratic hepatotoxicity and the "multiple determinant hypothesis" for the manifestation of idiosyncratic drug toxicity. Chem Biol Interact 142:7–23
10. Cheng A, Dixon SL (2003) In silico models for the prediction of dose-dependent human hepatotoxicity. J Comput Aided Mol Des 17:811–823
11. Clark RD, Wolohan PR, Hodgkin EE et al (2004) Modelling in vitro hepatotoxicity using molecular interaction fields and SIMCA. J Mol Graph Model 22:487–497
12. Marchant CA (2006) Virtual ADMET assessment. In: Testa B, Turski L (eds) Target selection and maturation. IOS Press, Amsterdam, p 237
13. Marchant CA, Fisk L, Note RR et al (2009) An expert system approach to the assessment of hepatotoxic potential. Chem Biodivers 6:2107–2114
14. Cruz-Monteagudo M, Cordeiro MN, Borges F (2008) Computational chemistry approach for the early detection of drug-induced idiosyncratic liver toxicity. J Comput Chem 29:533–549
15. Ekins S, Williams AJ, Xu JJ (2010) A predictive ligand-based Bayesian model for human drug-induced liver injury. Drug Metab Dispos 38:2302–2308
16. Liu Z, Shi Q, Ding D, Kelly R et al (2011) Translating clinical findings into knowledge in drug safety evaluation—drug induced liver injury prediction system (DILIps). PLoS Comput Biol 7(12):e1002310. doi:10.1371/journal.pcbi.1002310
17. Rodgers AD, Zhu H, Fourches D et al (2010) Modeling liver-related adverse effects of drugs using k-nearest neighbor quantitative structure-activity relationship method. Chem Res Toxicol 23:724–732
18. Low Y, Uehara T, Minowa Y et al (2011) Predicting drug-induced hepatotoxicity using qsar and toxicogenomics approaches. Chem Res Toxicol 24:1251–1262
19. Zhu XW, Sedykh A, Liu SS (2014) Hybrid in silico models for drug-induced liver injury using chemical descriptors and in vitro cell-imaging information. J Appl Toxicol 34:281–288
20. Matthews EJ, Ursem CJ, Kruhlak NL et al (2009) Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: part B. Use of (Q)SAR systems for early detection of drug induced hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54:23–42
21. Chan K, Jensen NS, Silber PM, O'Brien PJ (2007) Structure–activity relationships for halobenzene induced cytotoxicity in rat and human hepatoctyes. Chem Biol Interact 165:165–174
22. Greene N, Fisk L, Naven RT et al (2010) Developing structure-activity relationships for the prediction of hepatotoxicity. Chem Res Toxicol 23:1215–1222
23. Fourches D, Barnes JC, Day NC et al (2010) Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species. Chem Res Toxicol 23:171–183
24. Liew CY, Lim YC, Yap CW (2011) Mixed learning algorithms and features ensemble in hepatotoxicity prediction. J Comput Aided Mol Des 25:855–871
25. Chen M, Hong H, Fang H et al (2013) Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs. Toxicol Sci 136:242–249

26. Liu J, Mansouri K, Judson RS et al (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol 28:738–751
27. Tsakovska I, Al Sharif M, Alov P et al (2014) Molecular modelling study of the PPARγ receptor in relation to the mode of action/adverse outcome pathway framework for liver steatosis. Int J Mol Sci 15(5):7651–7666
28. Steinmetz FP, Mellor CL, Meinl T et al (2015) Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: using public data to build screening tools within a KNIME workflow. Mol Inform 34:171–178
29. Liu R, Yu X, Wallqvist A (2015) Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. J Cheminform 7:4
30. Hewitt M, Enoch SJ, Madden JC et al (2013) Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. Crit Rev Toxicol 43(7):537–555
31. Tralau T, Oelgeschläger M, Gürtler R et al (2015) Regulatory toxicology in the twenty-first century: challenges, perspectives and possible solutions. Arch Toxicol 89:823–850
32. Golbraikh A, Tropsha A (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des 16:357–369
33. Williams A, Tkachenko V, Lipinski C et al (2010) Free online resources enabling crowd-sourced drug discovery. Drug Discov World 10:33–39
34. Breiman L, Friedman JH, Olshen RA et al (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA
35. Hawkins DM, Kass GV (1982) Automatic interaction detection. In: Hawkins DH (ed) Topics in applied multivariate analysis. Cambridge University Press, Cambridge, UK, pp 269–302
36. Dixon SL, Villar HO (1999) Investigation of classification methods for the prediction of activity in diverse chemical libraries. J Comput Aided Mol Design 13:533–545
37. Xia X, Maliski EG, Gallant P, Rogers D (2004) Classification of kinase inhibitors using a Bayesian model. J Med Chem 47:4463–4470
38. Orange book: approved drug products with therapeutic equivalence evaluations. http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm
39. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32(7):1466–1474
40. Xu JJ, Henstock PV, Dunn MC et al (2008) Cellular imaging predictions of clinical drug-induced liver injury. Toxicol Sci 105:97–105
41. Olson H, Betton G, Stritar J et al (1998) The predictivity of the toxicity of pharmaceuticals in humans from animal data. An interim assessment. Toxicol Lett 10:535–538
42. Olson H, Betton G, Robinson D et al (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol 32:56–67
43. Farrell GC (1994) Drug-induced liver disease. Churchill Livingstone, New York
44. https://aopkb.org/index.html
45. Ursem CJ, Kruhlak NL, Contrera JF et al (2009) Identification of structure activity relationships for adverse effects of pharmaceuticals in humans. Part A: use of FDA post-market reports to create a database of hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54:1–22
46. Matthews EJ, Kruhlak NL, Benz RD et al (2009) Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: Part C. Use of QSAR and an expert system for the estimation of the mechanism of action of drug-induced hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54:43–65
47. Tropsha A, Golbraikh A (2007) Predictive QSAR modelling workflow, model applicability domains, and virtual screening. Curr Pharm Des 13:3494–3504
48. Kuhn M, Campillos M, Letunic I et al (2010) A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 6:6
49. Chen M, Vijay V, Shi Q, Liu Z, Fang H et al (2011) FDA-approved drug labelling for the study of drug-induced liver injury. Drug Discov Today 16:697–703
50. O'Brien PJ, Irwin W, Diaz D et al (2006) High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. Arch Toxicol 80:580–604
51. http://cosmostox.eu
52. http://knimewebportal.cosmostox.eu/webportal/#/Public/Nuclear%20Receptor%20Binding/LXR%20Binding%20Potential
53. Berthold MR, Cebron N, Dill F et al (2008) KNIME: the Konstanz Information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds) Studies in classification, data analysis, and knowledge organization. Springer, Berlin

54. Kavlock RJ, Chandler K, Houck KA et al (2012) Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25:1287–1302
55. Judson RS, Houck KA, Kavlock RJ et al (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118: 485–492
56. Uehara T, Ono A, Maruyama T et al (2010) The Japanese toxicogenomics project: application of toxicogenomics. Mol Nutr Food Res 54:218–227
57. Lhasa Ltd (2015) Analysis of human and in vivo data for hepatotoxicity modelling. http://www.lhasalimited.org/Public/Library/2015/Analysis%20of%20human%20and%20in%20vivo%20data%20for%20hepatotoxicity%20modelling.pdf
58. Ideaconsult Ltd (2012) Toxmatch structural similarity tool (version 1.07). http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/toxmatch
59. http://www.etoxproject.eu/
60. http://www.epa.gov/ncct/virtual_liver
61. Fontana RJ, Watkin PB, Bonkovsky HL et al (2009) DILIN Study Group, Drug-Induced Liver Injury Network (DILIN) prospective study: rationale, design and conduct. Drug Saf 32:55–68
62. http://www.virtual-liver.de
63. http://www.mip-dili.eu/