

Hybrid non-animal modeling: A mechanistic approach to predict chemical hepatotoxicity

Elena Chung^{1,2,a}, Xia Wen^{3,a}, Xuelian Jia^{1,2}, Heather L. Ciallella⁴, Lauren M. Aleksunes³, and Hao Zhu^{1,2,}*

¹Department of Chemistry and Biochemistry, Rowan University, New Jersey, USA.

²Center for Biomedical Informatics and Genomics, Tulane University, New Orleans, Louisiana, USA.

³Department of Pharmacology and Toxicology, Rutgers University, Piscataway, New Jersey, USA.

⁴Department of Toxicology, Cuyahoga County Medical Examiner's Office, Cleveland, Ohio, USA.

^aContributed to the paper equally

*Corresponding author:

Hao Zhu

Telephone: (504) 988-3443

Email: h Zhu10@tulane.edu

1415 Tulane Ave

New Orleans, LA 70112

ABSTRACT

Developing mechanistic non-animal testing methods based on the adverse outcome pathway (AOP) framework must incorporate molecular and cellular key events associated with target toxicity. Using data from an *in vitro* assay and chemical structures, we aimed to create a hybrid model to predict hepatotoxicants. We first curated a reference dataset of 869 compounds for hepatotoxicity modeling and profiled them against PubChem for existing *in vitro* toxicity data. Of the 2,560 resulting assays, we selected the mitochondrial membrane potential (MMP) assay, a high-throughput screening (HTS) tool that can test chemical disruptors for mitochondrial function. Then, machine learning was applied to develop quantitative structure-activity relationship (QSAR) models with 2,536 compounds tested in the MMP assay for screening new compounds. The MMP assay results, including QSAR model outputs, yielded hepatotoxicity predictions for reference set compounds with a Correct Classification Ratio (CCR) of 0.59. The predictivity improved by including 37 structural alerts (CCR = 0.8). We validated our model by testing 37 reference set compounds in human HepG2 hepatoma cells and reliably predicted these compounds for hepatotoxicity (CCR = 0.79). This study introduces a novel AOP modeling strategy that combines public HTS data, computational modeling, and experimental testing to predict chemical hepatotoxicity.

Keywords

hepatotoxicity; computational modeling; adverse outcome pathway; oxidative stress; mitochondrial dysfunction

MAIN TEXT

1. Introduction

Hepatotoxicity, often referred to as drug-induced liver injury (DILI), is a leading cause of acute liver failure in the United States and Europe (Chalasani et al., 2008; Germani et al., 2012; Lee, 1993, 2013; Weiler et al., 2020) and the primary reason for drug attrition during development (Björnsson, 2016; David & Hamilton, 2010; McNaughton et al., 2014). The idiosyncratic metabolic nature of DILI continues to prompt the cessation of clinical trials and even the recalls of drugs post-marketing (M. Chen et al., 2011). Given the severity and potential for DILI to hinder drug development, risk assessments are essential to mitigate potential life-threatening adverse reactions to drugs, environmental agents, and other xenobiotics (Senior, 2014). As the pathogenesis of DILI is one of the most complex toxicity phenomena, there are currently no applicable alternative models available for its assessment. DILI involves numerous cellular and biochemical processes, and many compounds that induce hepatotoxicity have poorly understood mechanisms. Given the concern for hepatotoxicity across diverse areas of toxicology, comprehensive assessments are imperative.

High-throughput screening (HTS) facilitates the assessment of a vast array of compounds for their specific biological activity. The United States Environmental Protection Agency's (US EPA) Toxicity Forecaster (ToxCast) comprises chemical screening data from HTS approaches. It plays a vital role in the Toxicology in the 21st Century (Tox21) program developed by the EPA, Food and Drug Administration (FDA), and National Center of Advancing Translational Sciences (NCATS) (Dix et al., 2007; Judson et al., 2010). While ToxCast HTS assays have significantly advanced our ability to assess chemical toxicities (Fox et al., 2012; Knight et al., 2009), no individual assay can replace *in vivo* testing for hepatotoxicity (Ekins, 2014).

Regulatory agencies and pharmaceutical companies have prioritized developing and implementing non-animal testing methodologies, such as computational approaches and HTS, that can minimize

the use of animal testing (Höfer et al., 2004; Krebs et al., 2019; Stucki et al., 2022). Depending solely on synthesizing and testing new compounds to gather experimental HTS data can still be expensive and time-consuming, especially when multiple assays are necessary to evaluate many new compounds. Computational modeling, such as quantitative structure-activity relationship (QSAR) models, can effectively and efficiently predict the biological activities of many new compounds across simple toxicity mechanisms such as skin sensitization (Cronin & Basketter, 1994), skin and eye irritation or corrosion (Patlewicz et al., 2003), and protein binding (Dimitrov et al., 2016; Gutsell & Russell, 2013; Sedykh et al., 2013). However, QSAR models have encountered challenges predicting complex toxicity endpoints, such as hepatotoxicity (Ekins, 2014). Based on our previous studies, integrating computational approaches and HTS assays for modeling complicated toxicity endpoints (e.g., estrogen receptor agonists, *in vivo* acute toxicity) holds promise for modeling hepatotoxicity (Jia et al., 2022; Kim et al., 2016; Ciallella et al., 2021b; Russo et al., 2019; H. Zhu et al., 2014).

In recent years, the adverse outcome pathway (AOP) framework has gained much attention as a strategy for conducting risk assessment (Ankley et al., 2010; Hartung, 2009; Krewski et al., 2009; Perkins et al., 2019; Tollefsen et al., 2014; Vinken, 2013). The Organization for Economic Cooperation and Development (OECD) established AOP guidelines that structure the progression of chemical toxicity occurrence from initial chemical exposure (i.e., the molecular initiating event (MIE)) to the following key events (KEs) and the manifestation of an adverse outcome (AO) (Allen et al., 2014; OECD, 2017). Alongside this, the AOP-Wiki (<https://aopwiki.org/>) has emerged as the central platform for sharing AOP knowledge, encompassing both OECD-endorsed AOPs and those contributed by the scientific community. This platform features established AOPs and those under development for hepatotoxicity (Arnesdotter et al., 2021, 2022; Gijbels & Vinken, 2017).

However, none of the current AOPs include the decrease in MMP as the KE linked from the MIE associated with hepatotoxicophores, leading to liver injury (AO). Our previous research employed a single assay (i.e., the antioxidant response element (ARE) assay to probe oxidative stress) and structural alerts (i.e., hepatotoxicophores) for hepatotoxicity modeling (Jia et al., 2022). However, the dataset's predictivity and coverage were limited to compounds that showed active results in this assay. Consequently, a substantial portion of the predicted compounds yielded inconclusive results. While oxidative stress is a recognized biomarker for mitochondrial toxicity, a deeper mechanistic understanding of mitochondrial roles in chemical-induced hepatotoxicity requires additional *in vivo* investigations. Mitochondrial stress in hepatotoxicity disrupts liver cell mitochondrial function, reducing mitochondrial membrane potential (MMP), impairing energy production, initiating cell death pathways, and causing liver damage (Ramachandran et al., 2018). The HTS MMP assay, extensively used in previous studies to identify mitochondrial toxicants, has not been applied to link this toxicity mechanism with hepatotoxicity or to predict potential hepatotoxicants (Huang et al., 2016; Seal et al., 2022; R. Zhang et al., 2023). Therefore, incorporating another essential assay that reflects mitochondrial activity linked to the oxidative stress pathway and supplementing our modeling framework with additional data can significantly improve the accuracy and interpretability of hepatotoxicity predictions.

This study established an enhanced computational toxicity modeling workflow based on an AOP framework by creating a correlation between chemical structures, low-cost toxicity assays, and human hepatotoxicity by prioritizing and modeling the MMP assay from profiling a hepatotoxicity dataset (outlined in Figure 1). We developed machine learning models to predict the activity of new compounds when experimental MMP assay data were unavailable. By combining structural alerts with assay testing and prediction results, the hepatotoxicity of new compounds can be

predicted by revealing their toxicity mechanisms. Collectively, the hybrid, mechanism-driven model elucidates a pathway from initial off-site events to eventual organism-level toxicity.

2. Methods

2.1. Hybrid modeling for human hepatotoxicity

The framework presented in Figure 1 describes the proposed hybrid modeling approach outlined in this study. The hepatotoxicity reference dataset served as the foundation for identifying structural alerts and profiling relevant bioassays, with the MMP assay selected for QSAR modeling. The QSAR modeling process, visually represented in purple, and structural alerts, shown in red, were integrated to create the hybrid model indicated by a combination of red and purple. The hybrid model was subsequently validated for its ability to predict human hepatotoxicity through experimental testing, as shown in red and green.

2.2. Profiling PubChem assays for hepatotoxicity data collection

The reference dataset used for modeling primarily contained the drug-induced liver injury rank (DILIrank) dataset, which the US FDA has extensively curated (M. Chen et al., 2016). The DILIrank dataset comprised 1,036 FDA-approved drugs classified into four groups based on their level of DILI concern: “most,” “less,” “no,” and “ambiguous.” The “ambiguous-DILI concern” group with 254 compounds was excluded from modeling due to the lack of conclusive causal evidence. The remaining compounds were categorized into hepatotoxic and non-hepatotoxic classifications. Specifically, 470 compounds from the “most-” and “less-DILI concern” classes were classified as hepatotoxic. The 312 compounds from the “no-DILI concern” group were classified as non-hepatotoxic. The chemical structures of the 782 compounds underwent curation and standardization using the CASE Ultra v1.8.0.0 DataKurator tool. This process involved the removal of duplicate structures, inorganic compounds, and mixtures.

After curating and standardizing the DILIrank dataset, 678 compounds were identified, comprising 432 hepatotoxic and 246 non-hepatotoxic compounds, causing an unbalanced distribution for the final hepatotoxicity reference set. To address the unequal distribution of toxic and non-toxic compounds in the dataset, an additional 191 non-hepatotoxic compounds were collected from previous studies (M. Chen et al., 2011; Ekins et al., 2010; Fourches et al., 2010; Kim et al., 2016; Liew et al., 2011; Liu et al., 2015; Mulliner et al., 2016). The final balanced hepatotoxicity reference set comprised 869 unique compounds, with 432 known hepatotoxins that afflict humans with DILI (Jia et al., 2022), including anti-cancer, anti-retroviral, and cardiac drugs (M. Chen et al., 2016), along with 437 non-hepatotoxins (Table S1, Supplementary Excel file). We generated a bioprofile for the compounds in the hepatotoxicity reference set using *in vitro* bioassay hit calls. Our automated profiling tool extracted information from *in vitro* assays on PubChem for these 869 compounds. The Python code used for profiling is openly accessible on our GitHub repository at <https://github.com/zhu-research-group/HTSProfiling/> (Russo & Zhu, 2022). The bioprofile categorizes the activity of compound-bioassay pairs as ‘active’ (1), ‘inactive’ (-1), ‘inconclusive’ (0), ‘unspecified’ (0), or ‘untested’ (0) (Y. Wang et al., 2009). We eliminated assays with fewer than five active responses and a total of 100 responses (combining active and inactive) within the profile. The remaining assays were then ranked based on their correlation with hepatotoxic compounds.

2.3. QSAR model development

The issue of missing data hinders the ability to discern relationships between compound-bioassay pairs. Traditional imputation methods, like random sampling, can be inadequate and potentially problematic. Therefore, we used QSAR modeling for a more accurate imputation of biological data that was congruent with previous computational toxicology studies (Russo et al., 2019; Zhao

et al., 2020). First, the curated dataset, which excluded inconclusive results, was balanced using a one-sided undersampling method. This method involved randomly removing inactive compounds to equalize the number of active and inactive compounds in the dataset for training the models (Zakharov et al., 2014).

Five distinct machine learning algorithms were then employed to create QSAR models for the selected PubChem assay to fill data gaps for new compounds. These algorithms consisted of AdaBoost decision tree (ADA), Bernoulli Naïve Bayes (BNB), k -nearest neighbors (k NN), random forest (RF), and support vector machines (SVM). The QSAR modeling process used a publicly available workflow (Ciallella, Chung, et al., 2022) implemented with Python v3.9.4 and scikit-learn v0.24.1 (<http://scikit-learn.org/>) (Pedregosa et al., 2011). The ADA algorithm assigns sample weights in the training set and trains subsequent trees while increasing the weights of misclassified instances (Freund & Schapire, 1997; Hastie et al., 2009). It repeats this process until it achieves the desired accuracy and combines the trees to predict instances. The BNB algorithm, commonly employed in classification tasks, assumes feature independence and uses Bayes' theorem to predict the chemical activity of a target compound (Manning et al., 2009). The k NN algorithm predicts the chemical activity of the target compound by identifying the k nearest compounds in a training set, where the number k is a parameter that determines the size of the neighborhood used for prediction (Cover & Hart, 1967). It then assigns the compound to the most common activity among these neighbors. The RF algorithm constructs decision tree ensembles with varying training data subsets and input features. Then, it aggregates the predictions to improve accuracy and reduce overfitting (Breiman, 2001). The SVM algorithm analyzes the active and inactive training set compounds to find the best way to form a plane that separates two classes of compounds based on their features (Vapnik, 2000).

Four types of chemical descriptors were generated for all compounds in Python v3.9.4 using the RDKit package v2021.03.1 (<http://www.rdkit.org/>): Molecular ACCess System (MACCS), extended-connectivity fingerprints (ECFPs), functional-class fingerprint (FCFPs), and RDKit. MACCS keys, ECFPs, and FCFPs are types of binary vector-based chemical fingerprints. Specifically, the MACCS keys were represented as 166-bit vectors identifying two-dimensional substructures within molecules (Leach & Gillet, 2007), while both the ECFPs and FCFPs were represented as 1,024-bit binary vectors (Rogers & Hahn, 2010). ECFPs capture atom properties within a substructure of a molecule, and FCFPs characterize general functional groups of the atoms (Rogers & Hahn, 2010). All ECFPs and FCFPs were calculated using a bond radius of 3 (Russo et al., 2018). A total of 208 RDKit molecular descriptors were calculated using the RDKit package, including information about the compounds' compositions and topological states (e.g., molecular weight and polar surface area). Individual QSAR models were developed using a combination of a chemical descriptor type and a machine learning algorithm. To mitigate the potential bias in individual model predictions, consensus models were developed by averaging individual models' prediction outcomes (Chung et al., 2023; Ciallella et al., 2021a; Golbraikh et al., 2017; Jia et al., 2021, 2022; W. Wang et al., 2015).

A five-fold cross-validation procedure was applied to evaluate the models' performance (Tropsha et al., 2003). During this procedure, the assay dataset was randomly split into five subsets, each comprising an equal number of compounds. One subset (20% of the total training set compounds) was used for predictions. The remaining four subsets (80% of the total training set compounds) were combined as a training set to develop the QSAR models. This procedure was completed after five iterations, once each compound was used for prediction once in a test set. The statistical metrics for each iteration were calculated and averaged to estimate the overall performance. Our

previous publication provided details about the QSAR modeling workflow and the modeling algorithms (Ciallella, Chung, et al., 2022), which we successfully applied in previous studies (Chung et al., 2023; Ciallella et al., 2021b; Jia et al., 2022; Russo et al., 2018). The Python code for the QSAR modeling workflow is publicly accessible from https://github.com/zhu-research-group/auto_qsar/.

2.3.1. External QSAR model evaluation

We obtained another dataset (AID 1347389) from PubChem, which includes 346 natural product compounds that underwent HTS MMP testing. These compounds were sourced from the NCATS-Canvass Library. After removing compounds that overlapped with the balanced QSAR model training set (AID 720635) and standardizing the chemical structures, the dataset comprised 51 active and 209 inactive compounds, totaling 260 unique compounds. This dataset was used as an external test set to evaluate the generated QSAR models.

We implemented a Principal Component Analysis (PCA) using 206 MOE two-dimensional descriptors within the Molecular Operating Environment (MOE) software v2020.09 to visualize the chemical space of the balanced training set and external test set. We selected the top three most important principal components to explain the total variance and to represent the large and diverse chemical space.

2.4. Experimental validation

2.4.1. Cell culture

Human hepatocellular carcinoma HepG2 cells were obtained from American Type Culture Collection (Manassas, VA) and cultured in Dulbecco's modified Eagle medium (ThermoFisher Scientific, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS, Atlanta Biologicals, Norcross, GA) and 1% penicillin-streptomycin.

2.4.2. Cell viability assays

HepG2 cells (8×10^3 /well) were seeded in 96-well plates overnight and then treated with the tested compounds (0-100 μ M) in culture media containing 1% FBS for 24 hours. Cell viability was determined using propidium iodide (PI) staining. Briefly, cells were stained with PI (1 μ g/ml, Sigma) and Hoechst 33342 (2 μ M, Nexcelom Bioscience, Lawrence, MA). The fluorescence from PI staining (nonviable cells) was acquired using a Cytation 5 Cell Imaging Reader (Agilent Technologies Inc., Wilmington, DE) fitted with PI (PI staining) and DAPI filter cubes (Hoechst 33342). The percentage of nonviable cells was calculated and normalized to the total number of cells stained with the Hoechst 33342 dye. Typically, five to seven concentrations spanning 1 to 3 logarithmic units were tested. Concentration-dependent data for each chemical and the thresholds at which greater than 10% of cells were stained PI-positive are included in Table S3, Supplementary Excel file. DMSO (20%) was used as a positive control and typically achieved more than 90% of cells being PI-positive stained at 24 hours.

2.4.3. Mitochondrial membrane potential (MMP) assay

MitoTracker Red probes (ThermoFisher Scientific, Carlsbad, CA), containing a thiol-reactive chloromethyl moiety, were used to label mitochondria and determine the MMP. Briefly, HepG2 cells (8×10^3 /well) were seeded in 96-well plates overnight and treated with tested compounds (0-100 μ M) in culture media containing 1% FBS for 24 hours. Following treatment, cells were incubated with MitoTracker Red probes (20 nM) and Hoechst 33342 (2 μ M) for 45 minutes. The fluorescence of MitoTracker Red staining was acquired using a Cytation 5 Cell Imaging Reader fitted with Texas Red (MitoTracker Red staining) and DAPI filter cubes (Hoechst 33342). The percentage of MitoTracker-stained cells was calculated and normalized to the total number of cells stained with Hoechst 33342 dye. Five non-toxic concentrations (where dead cells were <10% of

the control) of each chemical, determined from the PI analysis, were selected for further investigation in the MMP assay. Concentration-dependent data for each chemical and the thresholds at which there was a 25% loss of MitoTracker Red fluorescence intensity are included in Table S3, Supplementary Excel file. Sorafenib (5-25 μ M) was used as a positive control and typically achieved over a 50% loss of MitoTracker Red fluorescence at 24 hours.

2.5. Statistical parameters for assessing *in vitro-in vivo* correlations

Sensitivity, specificity, correct classification rate (CCR), positive predictive value (PPV), and accuracy (ACC) were used to evaluate the QSAR model performance and the correlations between the *in vitro* assay response and the known human hepatotoxicity endpoint. Sensitivity is the ability to predict active compounds correctly (Equation 1). Specificity is the ability to predict inactive compounds correctly (Equation 2). The CCR is the average of sensitivity and specificity, representing the overall balanced accuracy of the predictions (Equation 3). The PPV is the proportion of correctly classified active predictions (Equation 4). Accuracy (ACC) is the proportion of correctly predicted compounds (Equation 5). True positives (TP) were correctly identified as human hepatotoxicants, and true negatives (TN) were correctly identified as non-hepatotoxic compounds. False positives (FP) mistakenly indicated hepatotoxicity, and false negatives (FN) mistakenly indicated the absence of hepatotoxicity for the predicted compounds.

$$\text{sensitivity} = \text{true positives (TP)} / (\text{true positives (TP)} + \text{false negatives (FN)}) \quad (1)$$

$$\text{specificity} = \text{true negatives (TN)} / (\text{true negatives (TN)} + \text{false positives (FP)}) \quad (2)$$

$$\text{CCR} = (\text{sensitivity} + \text{specificity}) / 2 \quad (3)$$

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP}) \quad (4)$$

$$ACC = (TP + TN)/(TP + TN + FP + FN) \quad (5)$$

2.6. Characterizing chemical structural alerts

CASE Ultra software v.1.9.0 (MultiCASE Inc., Beachwood, OH) was employed to identify chemical substructures statistically associated with hepatotoxins. The training set compounds were randomly selected from the reference set and the test set compounds constituted the remainder. Hepatotoxicity classifications were determined based on the potential structural alerts of training set compounds. To establish standardized criteria, we considered structural alerts with a PPV greater than 0.68 that appeared in at least five compounds (Jia et al., 2022). Using these identified structural alerts, we predicted the toxicity of the compounds in the test set. We then used the results to refine the mechanistic hepatotoxicity model further.

3. Results

3.1. Overview of the hepatotoxicity reference dataset

The hepatotoxicity reference set defines the applicability domain within the chemical space and influences the model development process. To assess the structural diversity of our reference set, we employed the Tanimoto similarity metric based on MACCS fingerprints (Bajusz et al., 2015). This metric, which measures chemical similarity on a scale from 0 (i.e., no similarity) to 1 (i.e., identical compounds), analyzes molecular fingerprints for each compound to all other compounds in the reference set. Compounds were identified as structurally distinct if their Tanimoto coefficient, indicating similarity with their nearest neighbor compound in the reference set, was less than 0.7 (L. Zhang et al., 2013). The resulting average similarity score of 0.32 suggests moderate similarity among compounds in the reference set. Therefore, our hepatotoxicity reference set demonstrated significant structural diversity (Figure S1, Supplementary Word file).

3.2. Profiling and assay filtering

The *in vitro* toxicity data for compounds in the reference set were extracted from PubChem using an automatic profiling tool (Russo & Zhu, 2022). The resulting bioprofile encompassed 2,560 assays with at least five active compound-bioassay pairs (Figure 2). This initial bioprofile used for the following mechanistically related bioassay selections contained 36,593 active testing results and 146,011 inactive results for the 869 compounds (Figure 2). The imbalanced active/inactive ratio reflected the nature of high-throughput bioassay data in toxicity testing (i.e., predominately comprising inactive rather than active results) (Ciallella & Zhu, 2019; H. Zhu et al., 2014; H. Zhu, 2020).

We conducted a selective filtering process to identify which assays to consider in our hybrid modeling process. Irrelevant bioassays and bioassays that provided insufficient data on hepatotoxicity were filtered out. This filtering process also involved analyzing correlations between *in vitro* assay responses and *in vivo* hepatotoxicity. We focused on HTS assays that met the following criteria: the assay must have at least 100 responses (active or inactive), a PPV above 0.65, and a minimum of 25 correctly predicted hepatotoxic drugs (i.e., true positives). This filtration resulted in 73 assays. Further selection excluded nuclear receptors and cytochrome P450 (CYP) enzymes because their generalized mechanisms are not specific to toxicity, which narrowed the bioprofile to 10 assays (Table 1). Most assays represented events in pathways relevant to toxicity (i.e., mitochondrial and cell-death pathways) and were considered for mechanistic modeling. Fisher's exact test was conducted to assess the correlation of these assays with *in vivo* hepatotoxicity. This test procedure evaluated the likelihood of the observed responses against the expected proportion of known *in vivo* hepatotoxic drugs. Five *in vitro* assays showed a statistically significant relationship to human hepatotoxicity ($p < 0.05$) (Table 1).

310 Oxidative stress, a complex biological process, promotes inflammation and fibrosis and has been
311 implicated in several disease processes within organ systems such as the heart, kidney, and liver
312 (Daenen et al., 2019; Stocker & Keaney, 2004; X. Wang & Michaelis, 2010; R. Zhu et al., 2012).
313 Several studies connect mitochondrial oxidative stress to hepatotoxicity (Jaeschke et al., 2012;
314 Mansouri et al., 2018; Pessayre et al., 1999). Notably, associations were reported among
315 idiosyncratic DILI, mitochondrial dysfunction, and hepatocyte apoptosis (Boelsterli & Lee, 2014;
316 Boelsterli & Lim, 2007; K. Wang, 2014). For example, mitochondrial dysfunction can lead to
317 apoptosis through uncoupling of oxidative phosphorylation (Wittig et al., 2006), inhibition of
318 respiratory chain complexes (i.e., complex I/III) (R. Guo et al., 2018; Kühlbrandt, 2015), excess
319 cellular calcium ions (Matuz-Mares et al., 2022; Vogel et al., 2006), and activation of c-Jun N-
320 terminal kinases (Chambers & LoGrasso, 2011; Labbe et al., 2008; Win et al., 2018). Additionally,
321 several studies have explored the impact of the antioxidant response element (ARE) signaling
322 pathway on oxidative stress and showed a high association with hepatotoxicity through
323 computational and experimental research (Jia et al., 2022; Kim et al., 2016). Given these findings,
324 we selected the assay for chemical disruptors of the MMP (PubChem Assay Identifier (AID)
325 720635) as the key event for hepatotoxicity model development in this study.

326 The selection of the MMP assay for toxicity modeling is supported by its role as the AOP-Wiki's
327 KE 1170, which comprises several AOPs. Among them, AOP328, modeling excessive reactive
328 oxygen species (ROS) generation, is directly linked to mitochondrial dysfunction and has been
329 incorporated as a KE of this toxicity pathway, leading to mortality. Comparably, AOP387 models
330 the deposition of ionizing energy leading to population decline through mitochondrial dysfunction.
331 Although these two AOPs are not explicitly labeled as causes of liver toxicity, they collectively
332 highlight the significance of decreased MMP in various toxicity pathways, leading to population-

level toxicity. Therefore, our findings offer a promising new direction for constructing hepatotoxicity pathways through the loss of MMP, providing detailed insights into the toxicity phenomenon in organisms.

3.3. Data preprocessing

When modeling with HTS bioassay data, missing data points are common, as illustrated in Figure 2 (Ciallella & Zhu, 2019; H. Zhu, 2020; H. Zhu et al., 2014). QSAR approaches have been used and proven effective in predicting bioassay outcomes to resolve this issue (Chung et al., 2023; Ciallella et al., 2021b). The MMP assay contained results for 229 of the 869 compounds in the reference set (coverage = 26.4%). With 640 compounds lacking experimental MMP results, we developed QSAR models to fill these data gaps.

The MMP assay dataset from PubChem initially comprised 5,106 compounds. After curation and standardization, this was refined to 4,794 unique compounds, of which 1,268 were active and 3,526 were inactive. In this method, all 1,268 active compounds were retained, and 2,258 of the 3,526 inactive compounds were randomly removed. The balanced dataset containing 2,536 compounds was used to train the QSAR models.

The top three principal components, accounting for 53.28% of the total variance, were plotted to form the three-dimensional chemical space for the balanced training dataset ($n = 2,536$) and the external test set ($n = 260$). This visualization revealed a broad and diverse chemical space covered by these two datasets (Figure S2, Supplementary Word file). However, some outliers from both datasets were observable.

3.4. QSAR modeling and validation results of the MMP assay

We explored various methods to balance the training dataset and conducted an initial pilot study to compare QSAR modeling outcomes using balanced and imbalanced datasets (Zakharov et al.,

2014). The results showed no significant difference in overall model performance, as measured by CCR and PPV, though specificity increased by 13% at the expense of a 17% decrease in sensitivity. Given our aim to have the resulting models reliably predict activity/toxicity, using a balanced dataset was expected to enhance our success rate while likely preserving the original chemical diversity (Zakharov et al., 2014).

We developed 20 classification models by combining five machine learning algorithms (ADA, BNB, *k*NN, RF, SVM) with four chemical descriptor types (MACCS, ECFP6, FCFP6, rdkit). However, we observed poor predictions of active compounds with some algorithm-descriptor combinations (Bender, 2011; Zheng & Tropsha, 2000). We omitted three models that yielded subpar cross-validation sensitivities below 0.7: BNB-rdkit (sensitivity = 0.46), *k*NN-ECFP6 (sensitivity = 0.53), and *k*NN-FCFP6 (sensitivity = 0.58). The 17 remaining QSAR models were combined into a consensus model to enhance the overall performance and prediction reliability. Figure 3 displays the performance metrics of the individual and consensus QSAR models through five-fold cross-validation and external predictions. After cross-validation, the individual models achieved consistently good performance, sensitivity, specificity, and CCR, ranging between 0.73-0.86, 0.71-0.84, and 0.73-0.85, respectively. The consensus model, which was created by averaging the predictive values from each of the 17 models, demonstrated superior performance compared to one model alone (Figure 3a). The consensus model showed the best performance for the five-fold cross-validation, achieving a CCR value of 0.84, indicating it is the most suitable for predicting new compounds.

When evaluating an external test set, each model predicted actives with a sensitivity ranging from 0.65 to 0.96 (Figure 3b). The CCR varied between 0.56 and 0.74, and the specificity ranged from 0.22 to 0.71. The model's lower performance on external validation, compared to cross-validation,

can be partially explained by the disparity in the active-to-inactive compound ratios. The training set had a balanced distribution, whereas the external validation set had an active ratio of 1:4.1. This imbalance indicates that an insufficient range of inactive compounds in the training data can compromise the model. Compared to HTS bioassay data in toxicity testing, which typically has a 1:7 active-to-inactive ratio, downsampling the inactive compounds underscores the predictions of active compounds since active compounds are more critical for mechanistic modeling (Ciallella, Russo, et al., 2022; H. Zhu, 2020).

In some studies, excluding test chemicals that show low similarity to the training set chemicals could improve the model predictivity (Netzeva et al., 2005; Sahigara et al., 2012). We defined an applicability domain for the QSAR models by calculating a consensus value, the average probability across all 17 models for each compound, both during cross-validation and on the external test set (Sahigara et al., 2012; Tong et al., 2005). Compounds with a consensus value above 0.6 were classified as active, those below 0.4 as inactive, and those between 0.4 and 0.6 were excluded. However, defining an applicability domain did not show any advantages in improving the performance of the current model in this study.

Additionally, the natural compounds and their derivatives have distinct chemical structures and properties that differentiate them from most drugs, which can attenuate the predictivity of the models. For example, natural product compounds often have larger molecular sizes, reduced hydrophobicity, and fewer aromatic rings than fully synthetic drugs (Stratton et al., 2015). Moreover, the CCR of 0.67 of the consensus predictions aligned closely with the best individual models. The predictions for the natural products, though not as sensitive as those for drug molecules, were still suitable for external predictions when bolstered by consensus predictions.

3.5. Structural alerts for hepatotoxicity

Identifying structural alerts can reveal potential chemical features associated with hepatotoxicity. These specific substructures of the toxicants can trigger toxicity pathways, such as the oxidative stress pathway, by exhibiting specific initial structural or physicochemical properties. Therefore, these structural alerts can represent the MIEs, such as off-target bindings, that induce a toxic cellular response (Allen et al., 2014).

Previously, we developed a predictive hepatotoxicity model with the antioxidant response element (ARE) assay and structural alerts associated with liver injury (Jia et al., 2022). Here, we used the same strategy to identify structure alerts responsible for potential hepatotoxicity but with a larger hepatotoxicity dataset. The reference set was expanded with 12 additional compounds from the prior study, forming a hepatotoxicity set with 881 compounds. To focus on the predictivity of toxic compounds, we included a higher proportion of active, potentially toxic compounds in the test set. Therefore, the hepatotoxicity set had a training set of 702 compounds ($\approx 80\%$) and a test set of 179 compounds ($\approx 20\%$). The CASE-Ultra software was used to identify substructures statistically correlated with hepatotoxicity from toxic compounds in the training set. The analysis identified 37 potential structural alerts (Table S2, Supplementary Word file). Each alert was present in at least five compounds and exhibited a PPV > 0.68 . When predicting the test set compounds using structural alerts, the results showed suitable identification of potential hepatotoxic compounds (sensitivity = 0.75 and PPV = 0.94).

Structural alerts such as aniline/anilide (alert 22/24), triazole (alert 22), nitronium ion (alert 6), and imidazole derivatives (alert 31) induce ROS production in liver cells, causing oxidative stress (Table S2, Supplementary Word file). Oxidative stress can lead to rapid depolarization of the inner MMP. For example, triazole impairs the function of complexes I/III in the electron transport chain, increasing mitochondrial ROS production and causing apoptosis (Haegler et al., 2017). Nitric

oxide derivatives trigger apoptosis through nitrosative stress-mediated mitochondrial membrane depolarization via ROS generation (Langer et al., 2008). Imidazole derivatives exhibit cytotoxicity indicative of mitochondrial toxicity (Haegler et al., 2017), and aniline induces oxidative stress in hepatocytes (Y. Wang et al., 2016).

3.6. Hybrid mechanistic hepatotoxicity model and experimental *in vitro* validation

Our AOP-based hybrid model can predict hepatotoxicity by combining structural alerts and the MMP assay results, providing a mechanistic understanding of hepatotoxic outcomes (Figure 1). Compounds with structural alerts and active MMP assay results were labeled “hepatotoxic.” Those lacking structural alerts and showing inactive MMP assay results were labeled “non-hepatotoxic.” Compounds with structural alerts but inactive MMP assay results suggested triggering of other hepatotoxicity pathways. Compounds without structural alerts but with active MMP assay results indicated missing structural alerts (i.e., hepatotoxicants with unique structures in the training set) and the possibility of limited *in vivo* bioavailability for these toxicants. For these reasons, compounds falling into these two categories (i.e., conflicting between structure alerts and MMP assay results) were labeled “inconclusive.”

Our hybrid mechanistic model exhibited an enhanced correlation with hepatotoxicity compared to using only the MMP assay results (Figure 4). The model's sensitivity, specificity, and CCR improved for the training set from 0.40 to 0.68, 0.74 to 0.92, and 0.57 to 0.83, respectively. The sensitivity of the test set significantly increased from 0.5 to 0.71. However, because it had a higher proportion of active compounds, the specificity slightly decreased from 0.62 to 0.5. Overall, CCR moderately increased from 0.56 to 0.6. These data indicate that the loss of MMP is a critical mechanism of hepatotoxicity.

447 To validate the predictions of our model and demonstrate the value of incorporating low-cost
448 experimental testing like HTS assays into computational studies, we experimentally tested a set of
449 37 drugs, comprising 32 known hepatotoxic and five non-toxic compounds, *in vitro* using a
450 functionally equivalent MMP assay. These drugs were selected from a test set with established
451 hepatotoxicity outcomes but had not been previously evaluated using the MMP assay. As
452 illustrated in Figure 5, relying only on QSAR model predictions can effectively predict
453 hepatotoxicity, achieving a sensitivity of 100% and indicating that our model can identify potential
454 hepatotoxic compounds without experimental testing. However, as indicated in Figure 3, the model
455 predictions showed an emphasis on active compounds and a moderate ability to predict non-toxic
456 compounds (specificity = 0.5).

457 When the experimental MMP assay results replaced QSAR predictions in our hybrid model, the
458 predictions were more balanced, with a sensitivity of 0.83 and a specificity of 0.75. A total of 15
459 hepatotoxic compounds and three non-hepatotoxic compounds were accurately identified. There
460 were only four misclassified compounds: tamoxifen (PubChem Compound Identifier (CID)
461 2733526), fenofibrate (CID 3339), and desvenlafaxine (CID 125017), which were false negatives,
462 and panobinostat (CID 6918837), which was the only false positive. The remaining 15 compounds
463 were inconclusive due to discrepancies between structural alerts and the MMP assay testing
464 results. For example, progesterone (CID 5994) and pazopanib (CID 10113978) were inconclusive
465 because they were inactive *in vitro* but contained structural alerts. Our experimental results showed
466 that these two compounds significantly increased the number of dead cells in the PI assay at higher
467 concentrations, indicating direct cytotoxic effects or other modes of action not detected by the
468 MMP assay (Table S3, Supplemental Excel File). Most inconclusive compounds may indicate
469 hepatotoxicity through different mechanisms.

The false positive result for panobinostat conceivably occurs because extensive metabolism by CYP3A4 compromises its bioavailability (Van Veggel et al., 2018). HepG2 cells express lower levels of CYP3A4 compared to primary human hepatocytes, potentially leading to an incomplete assessment of panobinostat's metabolic transformation (Wilkening et al., 2003). Additionally, the high protein binding rate (74 - 83%) suggests that panobinostat may exhibit different effects in HepG2 cells than *in vivo* due to its restricted bioavailability (Van Veggel et al., 2018). A future study could improve the current model by incorporating a metabolism factor. This addition may address the issue of the three false negatives, likely due to toxicity mechanisms not currently captured by the model. For example, tamoxifen is metabolized by CYP enzymes into N-oxide, α -hydroxy, and 4-hydroxy forms that interact with protein and DNA, potentially leading to hepatocarcinogenic effects (Fan & Bolton, 2001; Park et al., 2005). Similarly, fenofibrate may cause hepatotoxicity from an immune response to its altered metabolites or protein conjugates in the liver rather than from direct chemical reactivity or mitochondrial interference (Ahmad et al., 2017).

Desvenlafaxine is primarily eliminated through the urine, unchanged, as a glucuronide metabolite or as an oxidative metabolite, N,O-didesmethylvenlafaxine (DeMaio et al., 2011). The glucuronidation of desvenlafaxine, catalyzed by uridine 5'-diphospho-glucuronosyltransferase (UGT) enzyme isoforms such as UGT1A1, 1A3, 2B4, 2B15, and 2B17, can be influenced by genetic polymorphisms, which may play a toxicological role in idiosyncratic DILI (DeMaio et al., 2011; Kutsuno et al., 2014; Miners et al., 2002). Moreover, no structural alerts were identified for these three compounds, indicating the diverse chemical structures of hepatotoxicants and the need to enlarge the current training set further.

4. Discussion

In this study, we used an automatic data mining approach to reveal an HTS assay representing a potential toxicity mechanism underlying DILI. The loss of MMP was identified as a potential KE in hepatotoxicity, suggesting the possibility of developing new pathways for understanding hepatotoxicity. For example, our findings can associate MMP reduction with lower cell viability, inferring a possible connection to irreversible cell damage and activation of cell death processes (e.g., apoptosis and necrosis) (Sun et al., 2005). However, this requires further refinement of hepatotoxicity data during the data mining and modeling processes and additional toxicity assay data for target compounds in future studies. Nevertheless, mitochondrial dysfunction is widely recognized as a key mechanism in DILI (Mihajlovic & Vinken, 2022), and its value in predicting DILI was confirmed in this study. Our study demonstrates that computational methods combined with publicly available HTS datasets offer a cost-effective approach to exploring hepatotoxicity pathways, particularly for developing new AOPs targeting specific toxicity mechanisms. The model developed in this study will be scalable, allowing for the integration of additional hepatotoxicity pathways as more public data on other hepatotoxicity mechanisms becomes available.

The QSAR model predictions allowed for the initial assessments of potentially toxic compounds and demonstrated high sensitivity. HTS assays typically contain more inactive compounds than active compounds, and machine learning models tend to favor the majority class (Magana-Mora & Bajic, 2017). An imbalanced training set would likely predict more compounds as inactive, necessitating extensive experimental testing to correct the QSAR model predictions. Therefore, we used a balanced dataset to enhance the prediction of active/toxic compounds and help elucidate potential pathways of hepatotoxicity. Additionally, this approach demonstrates the importance of incorporating low-cost testing as a post-requisite to validate toxic predictions.

Although not improving the prediction accuracy, integrating experimental MMP assay testing results to replace QSAR predictions renders the hybrid model more suitable for toxicity evaluations, including regulatory assessments. This approach ensures a balanced prediction capability for both hepatotoxic and non-toxic compounds, evidenced by its sensitivity, specificity, and CCR, each denoted by values 0.83, 0.75, and 0.79, respectively. A key advantage of the hybrid model, which incorporated experimental MMP assay data, is its ability to predict toxicity and accurately determine the underlying mechanisms involved. For instance, compounds diclofenac (CID 3033) and sertraline (CID 68617) were previously implicated in inducing oxidative stress, corroborated by our model predictions (Boelsterli, 2003; Li et al., 2012).

Our study aimed to predict chemical hepatotoxicity using a hybrid AOP framework, focusing on a prevalent mechanism potentially leading to hepatotoxicity. Integrating additional relevant assays, such as the ARE assay results, enables a deeper understanding of the toxicity mechanisms of oxidative stressors and can potentially lead to better predictions (Begriche et al., 2011; Tang et al., 2014). We previously obtained ARE testing results for 14 of our 37 experimental validation compounds (Jia et al., 2022). By integrating two key events (i.e., MMP and ARE) into the hybrid model, we predicted a compound as hepatotoxic if one of these two assays showed active results and the compound had an identified structural alert. Predicted non-toxicants had inactive results in both assays, and no structural alerts were identified. The remaining were inconclusive. This multi-assay hybrid model included an additional effective and cost-efficient experimental testing procedure, strengthening the predictions by increasing prediction accuracy to 79% with only one misclassified compound. As mentioned above, the misclassification of panobinostat in both the MMP and ARE assay results may be attributed to the inability of HepG2 cells to fully recapitulate hepatic metabolism.

With both assays strongly related to oxidative stress mechanisms, the multi-assay hybrid model better reflected the nature of toxicity pathways and significantly enhanced confidence in predicting hepatotoxicity potential. For example, crizotinib (CID 11626560) impacts oxidative responses by activating the ARE/Nrf2 pathway and decreasing mitochondrial function in the human hepatocyte cell line (L. Guo et al., 2021), potentially leading to hepatotoxicity. Similarly, diclofenac (CID 3033) was reported to activate the ARE/Nrf2 pathway, affecting mitochondrial function and promoting ROS generation (Herpers et al., 2016).

Sorafenib (CID 216239), a drug used to treat kidney, liver, and thyroid cancers, provides a compelling example of our model's predictive ability. This drug demonstrates hepatotoxicity through mitochondrial oxidative stress, which underpins its anti-cancer action. Sorafenib was shown to increase oxidative stress levels by reducing glutathione levels, a double-edged sword that both aids in cancer cell destruction and elevates the risk of liver injury (Duval et al., 2019). The case of fatal hepatotoxicity and renal failure reported by Fairfax et al. further underscored the intricate relationship between its therapeutic and side effects (Fairfax et al., 2012). The significance of the structural alerts in sorafenib is especially revealing insights into its anti-cancer and hepatotoxic effects. The drug possesses structural alerts 2, 21, and 24, as shown in Figure 5. Among them, alert 24 indicates a cyclohexyl urea structure that may be crucial for anti-tumor effectiveness and selectivity, as the diaryl urea structure and cyclohexyl moiety are vital to inhibiting multiple kinases, a sought-after mechanism in cancer therapy (F. Chen et al., 2019; Lowinger et al., 2002; Lu et al., 2015). However, the same structural feature may contribute to its hepatotoxic potential, exhibiting a dual functionality that is both beneficial and risky. Therefore, explaining these structural alerts is essential in evaluating a balance between the therapeutic efficacy and potential hepatotoxicity of drug molecules.

Beyond structural alerts, prospective chemical information can be used to improve the current model. For instance, compounds like ionophores can undermine mitochondrial membrane integrity and contribute to hepatotoxicity through severe mitochondrial uncoupling of oxidative phosphorylation (Song & Villeneuve, 2021). Classical uncouplers for ionophoric activity, characterized by their salicylate and benzimidazole structures, impair mitochondrial function by dissipating the proton gradient (Battaglia et al., 2005; Kessler et al., 1976; Terada, 1990) and can precipitate the collapse of MMP (Fiskum et al., 2000). Some chemicals in the hepatotoxicity set, such as mebendazole (CID 4030) (with a benzimidazole structure), salsalate (CID 5161) (containing two salicylate molecules), and clofibrate (CID 2796) (an aromatic monocarboxylic acid), may function as ionophore uncouplers. Future studies would benefit from considering uncoupling activity when expanding the training data (i.e., the hepatotoxicity reference set) to include more compounds. Additionally, the hepatotoxicity reference set, primarily sourced from the DILIrank dataset, does not include specific phenotypic information (e.g., acute liver failure, cirrhosis) but can be updated as new data become available.

Additionally, the induction of oxidative stress and impaired mitochondrial function do not always go hand in hand, as seen with boceprevir (CID 10324367) (Baines, 2010; Chu et al., 2021). Boceprevir may trigger hepatotoxicity through the oxidative stress pathway indicated by activation of NRF2/ARE signaling. Refining the hepatotoxicity profile with the ARE model for such compounds, our hybrid modeling approach integrates relevant assays into a novel multi-assay hybrid AOP model. This approach marks a meaningful advancement beyond traditional computational methods that can predict chemical hepatotoxicity across diverse mechanisms. To further explore the additional hepatotoxicity mechanisms, we examined the bioprofile of 20 hepatotoxic compounds inactive in the MMP assay. The analysis revealed active responses in the

assays belonging to other hepatotoxicity pathways. These postulated hepatotoxicity mechanisms include inhibition of CYP isoenzymes (AID 678712, AID 678713, AID 678715, AID 678716, and AID 678717) (Feng & He, 2013; Xuan et al., 2016), histone lysine methyltransferase G9a inhibition (AID 504332) (Y. Zhang et al., 2020), D2 dopamine receptor antagonism (AID 485344) (Abdel-Salam et al., 2013; Todorović Vukotić et al., 2021), metabolic stability in liver microsomes, as evidenced by GSH adduct formation (AID 678721) (Srivastava et al., 2010). These mechanisms may explain why some toxic compounds do not induce the oxidative stress pathways and warrant further investigation. Future research will explore expanding potential AOP frameworks to assess toxicities in other systems, such as developmental and reproductive systems, to predict potential toxicants through mechanistically understanding chemical/drug-induced toxicity pathways leading to hazardous effects.

5. Conclusions

Our study integrated machine learning and other modeling techniques to simulate the AOP framework, specifically for hepatotoxicity modeling and prediction. This novel modeling workflow commences with analyzing fundamental chemical structures, progressing to low-level and cost-effective *in vitro* testing, and ultimately extrapolating to human hepatotoxicity prediction. The developed hybrid model can predict potential toxicants that cause liver injury through oxidative stress. Incorporating low-cost experimental testing into the predictive process improves the model's performance. Moreover, including an additional assay targeting a similar but distinct toxicity mechanism further improves the predictive strength of the model. This study introduced a novel hybrid modeling approach designed to efficiently extract data from dynamic, unstructured public resources and use curated data, facilitating the modeling of complex toxicity endpoints. The model elucidates toxicity mechanisms, making it well-suited for regulatory risk assessments and

prioritizing hazardous chemicals for experimental testing. Moreover, this AOP modeling framework holds promise for adaptation into non-animal models, reducing animal usage in toxicity testing. Overall, our findings represent a significant advancement in traditional toxicity evaluation, advocating for a paradigm shift towards computational and alternative risk assessment methods in the era of big data.

Acknowledgments

This work was partially supported by the National Institute of General Medical Sciences (Grant R01GM148743), the National Institute of Child Health and Human Development (Grant UHD113039), the National Science Foundation (Grant 2402311), and the National Institute of Environmental Health Sciences (Grants R01ES031080, R35ES031709, UL1TR003017, and P30ES005022).

CRedit Authorship Contribution Statement

Elena Chung: Methodology, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Xia Wen:** Validation, Writing – original draft, Writing – review & editing. **Xuelian Jia:** Data curation, Methodology. **Heather L. Ciallella:** Software, Methodology. **Lauren M. Aleksunes:** Resources, Writing – review & editing, Supervision, Funding acquisition. **Hao Zhu:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Article Notes

The authors declare that they have no competing interests.

References

- Abdel-Salam, O. M. E., Youness, E. R., Khadrawy, Y. A., & Sleem, A. A. (2013). Brain and liver oxidative stress after sertraline and haloperidol treatment in mice. *Journal of Basic and Clinical Physiology and Pharmacology*, 24(2), 115–123. <https://doi.org/10.1515/jbcpp-2012-0022>
- Ahmad, J., Odin, J., Hayashi, P. H., Chalasani, N., Fontana, R. J., Barnhart, H., Cirulli, E. T., Kleiner, D. E., & Hoofnagle, J. H. (2017). Identification and Characterization of Fenofibrate-Induced Liver Injury. *Digestive Diseases and Sciences*, 62(12), 3596–3604. <https://doi.org/10.1007/s10620-017-4812-7>
- Allen, T. E. H., Goodman, J. M., Gutsell, S., & Russell, P. J. (2014). Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. *Chemical Research in Toxicology*, 27(12), 2100–2112. <https://doi.org/10.1021/tx500345j>
- Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrano, J. A., Tietge, J. E., & Villeneuve, D. L. (2010). Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry*, 29(3), 730–741. <https://doi.org/10.1002/etc.34>
- Arnesdotter, E., Gijbels, E., dos Santos Rodrigues, B., Vilas-Boas, V., & Vinken, M. (2022). Adverse Outcome Pathways as Versatile Tools in Liver Toxicity Testing. In E. Benfenati (Ed.), *In Silico Methods for Predicting Drug Toxicity* (pp. 521–535). Springer US. https://doi.org/10.1007/978-1-0716-1960-5_20
- Arnesdotter, E., Spinu, N., Firman, J., Ebbrell, D., Cronin, M. T. D., Vanhaecke, T., & Vinken, M. (2021). Derivation, characterisation and analysis of an adverse outcome pathway network for human hepatotoxicity. *Toxicology*, 459, 152856. <https://doi.org/10.1016/j.tox.2021.152856>
- Baines, C. P. (2010). Role of the Mitochondrion in Programmed Necrosis. *Frontiers in Physiology*, 1, 156. <https://doi.org/10.3389/fphys.2010.00156>

653 Bajusz, D., Rácz, A., & Héberger, K. (2015). Why is Tanimoto index an appropriate choice for
654 fingerprint-based similarity calculations? *Journal of Cheminformatics*, 7(1), 20.
655 <https://doi.org/10.1186/s13321-015-0069-3>

656 Battaglia, V., Salvi, M., & Toninello, A. (2005). Oxidative Stress Is Responsible for Mitochondrial
657 Permeability Transition Induction by Salicylate in Liver Mitochondria. *Journal of Biological*
658 *Chemistry*, 280(40), 33864–33872. <https://doi.org/10.1074/jbc.M502391200>

659 Begriche, K., Massart, J., Robin, M.-A., Borgne-Sanchez, A., & Fromenty, B. (2011). Drug-induced
660 toxicity on mitochondria and lipid metabolism: Mechanistic diversity and deleterious
661 consequences for the liver. *Journal of Hepatology*, 54(4), 773–794.
662 <https://doi.org/10.1016/j.jhep.2010.11.006>

663 Bender, A. (2011). Bayesian Methods in Virtual Screening and Chemical Biology. In J. Bajorath (Ed.),
664 *Chemoinformatics and Computational Chemical Biology* (pp. 175–196). Humana Press.
665 https://doi.org/10.1007/978-1-60761-839-3_7

666 Björnsson, E. S. (2016). Hepatotoxicity by Drugs: The Most Common Implicated Agents. *International*
667 *Journal of Molecular Sciences*, 17(2), 224. <https://doi.org/10.3390/ijms17020224>

668 Boelsterli, U. A. (2003). Diclofenac-induced liver injury: A paradigm of idiosyncratic drug toxicity.
669 *Toxicology and Applied Pharmacology*, 192(3), 307–322. [https://doi.org/10.1016/S0041-](https://doi.org/10.1016/S0041-008X(03)00368-5)
670 [008X\(03\)00368-5](https://doi.org/10.1016/S0041-008X(03)00368-5)

671 Boelsterli, U. A., & Lee, K. K. (2014). Mechanisms of isoniazid-induced idiosyncratic liver injury:
672 Emerging role of mitochondrial stress. *Journal of Gastroenterology and Hepatology*, 29(4), 678–
673 687. <https://doi.org/10.1111/jgh.12516>

674 Boelsterli, U. A., & Lim, P. L. K. (2007). Mitochondrial abnormalities—A link to idiosyncratic drug
675 hepatotoxicity? *Toxicology and Applied Pharmacology*, 220(1), 92–107.
676 <https://doi.org/10.1016/j.taap.2006.12.013>

677 Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
678 <https://doi.org/10.1023/A:1010933404324>

- Chalasani, N., Fontana, R. J., Bonkovsky, H. L., Watkins, P. B., Davern, T., Serrano, J., Yang, H., & Rochon, J. (2008). Causes, Clinical Features, and Outcomes From a Prospective Study of Drug-Induced Liver Injury in the United States. *Gastroenterology*, 135(6), 1924–1934.e4. <https://doi.org/10.1053/j.gastro.2008.09.011>
- Chambers, J. W., & LoGrasso, P. V. (2011). Mitochondrial c-Jun N-terminal Kinase (JNK) Signaling Initiates Physiological Changes Resulting in Amplification of Reactive Oxygen Species Generation. *The Journal of Biological Chemistry*, 286(18), 16052–16062. <https://doi.org/10.1074/jbc.M111.223602>
- Chen, F., Fang, Y., Zhao, R., Le, J., Zhang, B., Huang, R., Chen, Z., & Shao, J. (2019). Evolution in medicinal chemistry of sorafenib derivatives for hepatocellular carcinoma. *European Journal of Medicinal Chemistry*, 179, 916–935. <https://doi.org/10.1016/j.ejmech.2019.06.070>
- Chen, M., Suzuki, A., Thakkar, S., Yu, K., Hu, C., & Tong, W. (2016). DILrank: The largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discovery Today*, 21(4), 648–653. <https://doi.org/10.1016/j.drudis.2016.02.015>
- Chen, M., Vijay, V., Shi, Q., Liu, Z., Fang, H., & Tong, W. (2011). FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discovery Today*, 16(15), 697–703. <https://doi.org/10.1016/j.drudis.2011.05.007>
- Chu, Q., Gu, X., Zheng, Q., Wang, J., & Zhu, H. (2021). Mitochondrial Mechanisms of Apoptosis and Necroptosis in Liver Diseases. *Analytical Cellular Pathology (Amsterdam)*, 2021, 8900122. <https://doi.org/10.1155/2021/8900122>
- Chung, E., Russo, D. P., Ciallella, H. L., Wang, Y.-T., Wu, M., Aleksunes, L. M., & Zhu, H. (2023). Data-Driven Quantitative Structure–Activity Relationship Modeling for Human Carcinogenicity by Chronic Oral Exposure. *Environmental Science & Technology*, 57(16), 6573–6588. <https://doi.org/10.1021/acs.est.3c00648>

703 Ciallella, H. L., Chung, E., Russo, D. P., & Zhu, H. (2022). Automatic quantitative structure-activity
 704 relationship modeling to fill data gaps in high-throughput screening. In *High-Throughput*
 705 *Screening Assays in Toxicology* (2nd ed.). Springer US.

706 Ciallella, H. L., Russo, D. P., Aleksunes, L. M., Grimm, F. A., & Zhu, H. (2021a). Predictive modeling of
 707 estrogen receptor agonism, antagonism, and binding activities using machine- and deep-learning
 708 approaches. *Laboratory Investigation*, 101(4), 490–502. [https://doi.org/10.1038/s41374-020-](https://doi.org/10.1038/s41374-020-00477-2)
 709 00477-2

710 Ciallella, H. L., Russo, D. P., Aleksunes, L. M., Grimm, F. A., & Zhu, H. (2021b). Revealing Adverse
 711 Outcome Pathways from Public High-Throughput Screening Data to Evaluate New Toxicants by
 712 a Knowledge-Based Deep Neural Network Approach. *Environmental Science & Technology*,
 713 55(15), 10875–10887. <https://doi.org/10.1021/acs.est.1c02656>

714 Ciallella, H. L., Russo, D. P., Sharma, S., Li, Y., Slotter, E., Sweet, L., Huang, H., & Zhu, H. (2022).
 715 Predicting Prenatal Developmental Toxicity Based On the Combination of Chemical Structures
 716 and Biological Data. *Environmental Science & Technology*, 56(9), 5984–5998.
 717 <https://doi.org/10.1021/acs.est.2c01040>

718 Ciallella, H. L., & Zhu, H. (2019). Advancing Computational Toxicology in the Big Data Era by
 719 Artificial Intelligence: Data-Driven and Mechanism-Driven Modeling for Chemical Toxicity.
 720 *Chemical Research in Toxicology*, 32(4), 536–547.
 721 <https://doi.org/10.1021/acs.chemrestox.8b00393>

722 Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information*
 723 *Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>

724 Cronin, M. T. D., & Basketter, D. A. (1994). Multivariate Qsar Analysis of a Skin Sensitization Database.
 725 *SAR and QSAR in Environmental Research*, 2(3), 159–179.
 726 <https://doi.org/10.1080/10629369408029901>

727 Daenen, K., Andries, A., Mekahli, D., Van Schepdael, A., Jouret, F., & Bammens, B. (2019). Oxidative
728 stress in chronic kidney disease. *Pediatric Nephrology*, 34(6), 975–991.
729 <https://doi.org/10.1007/s00467-018-4005-4>

730 David, S., & Hamilton, J. P. (2010). Drug-induced Liver Injury. *US Gastroenterology & Hepatology*
731 *Review*, 6, 73–80.

732 DeMaio, W., P. Kane, C., Nichols, A. I., & Jordan, R. (2011). Metabolism Studies of Desvenlafaxine.
733 *Journal of Bioequivalence & Bioavailability*, 03(07). <https://doi.org/10.4172/jbb.1000076>

734 Dimitrov, S. D., Diderich, R., Sobanski, T., Pavlov, T. S., Chankov, G. V., Chapkanov, A. S., Karakolev,
735 Y. H., Temelkov, S. G., Vasilev, R. A., Geroval, K. D., Kuseva, C. D., Todorova, N. D., Mehmed,
736 A. M., Rasenberg, M., & Mekenyan, O. G. (2016). QSAR Toolbox – workflow and major
737 functionalities. *SAR and QSAR in Environmental Research*, 27(3), 203–219.
738 <https://doi.org/10.1080/1062936X.2015.1136680>

739 Dix, D. J., Houck, K. A., Martin, M. T., Richard, A. M., Setzer, R. W., & Kavlock, R. J. (2007). The
740 ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. *Toxicological*
741 *Sciences*, 95(1), 5–12. <https://doi.org/10.1093/toxsci/kfl103>

742 Duval, A. P., Troquier, L., de Souza Silva, O., Demartines, N., & Dormond, O. (2019). Diclofenac
743 Potentiates Sorafenib-Based Treatments of Hepatocellular Carcinoma by Enhancing Oxidative
744 Stress. *Cancers*, 11(10), Article 10. <https://doi.org/10.3390/cancers11101453>

745 Ekins, S. (2014). Progress in computational toxicology. *Journal of Pharmacological and Toxicological*
746 *Methods*, 69(2), 115–140. <https://doi.org/10.1016/j.vascn.2013.12.003>

747 Ekins, S., Williams, A. J., & Xu, J. J. (2010). A Predictive Ligand-Based Bayesian Model for Human
748 Drug-Induced Liver Injury. *Drug Metabolism and Disposition*, 38(12), 2302–2308.
749 <https://doi.org/10.1124/dmd.110.035113>

750 Fairfax, B., Pratap, S., Roberts, I., Collier, J., Kaplan, R., Meade, A., Ritchie, A., Eisen, T., Macaulay, V.,
751 & Protheroe, A. (2012). Fatal case of sorafenib-associated idiosyncratic hepatotoxicity in the

752 adjuvant treatment of a patient with renal cell carcinoma. *BMC Cancer*, 12(1), 590.
 753 <https://doi.org/10.1186/1471-2407-12-590>
 754 Fan, P. W., & Bolton, J. L. (2001). Bioactivation of Tamoxifen to Metabolite E Quinone Methide:
 755 Reaction with Glutathione and DNA. *Drug Metabolism and Disposition*, 29(6), 891–896.
 756 Feng, S., & He, X. (2013). Mechanism-based inhibition of CYP450: An indicator of drug-induced
 757 hepatotoxicity. *Current Drug Metabolism*, 14(9), 921–945.
 758 <https://doi.org/10.2174/138920021131400114>
 759 Fiskum, G., Kowaltowski, A. J., Andreyev, A. y., Kushnareva, Y. E., & Starkov, A. A. (2000). [21]—
 760 Apoptosis-Related Activities Measured with Isolated Mitochondria and Digitonin-Permeabilized
 761 Cells. In J. C. Reed (Ed.), *Methods in Enzymology* (Vol. 322, pp. 222–234). Academic Press.
 762 [https://doi.org/10.1016/S0076-6879\(00\)22023-5](https://doi.org/10.1016/S0076-6879(00)22023-5)
 763 Fourches, D., Barnes, J. C., Day, N. C., Bradley, P., Reed, J. Z., & Tropsha, A. (2010). Cheminformatics
 764 Analysis of Assertions Mined from Literature That Describe Drug-Induced Liver Injury in
 765 Different Species. *Chemical Research in Toxicology*, 23(1), 171–183.
 766 <https://doi.org/10.1021/tx900326k>
 767 Fox, J. T., Sakamuru, S., Huang, R., Teneva, N., Simmons, S. O., Xia, M., Tice, R. R., Austin, C. P., &
 768 Myung, K. (2012). High-throughput genotoxicity assay identifies antioxidants as inducers of
 769 DNA damage response and cell death. *Proceedings of the National Academy of Sciences*,
 770 109(14), 5423–5428. <https://doi.org/10.1073/pnas.1114278109>
 771 Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an
 772 Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
 773 <https://doi.org/10.1006/jcss.1997.1504>
 774 Germani, G., Theocharidou, E., Adam, R., Karam, V., Wendon, J., O’Grady, J., Burra, P., Senzolo, M.,
 775 Mirza, D., Castaing, D., Klempnauer, J., Pollard, S., Paul, A., Belghiti, J., Tsochatzis, E., &
 776 Burroughs, A. K. (2012). Liver transplantation for acute liver failure in Europe: Outcomes over

777 20years from the ELTR database. *Journal of Hepatology*, 57(2), 288–296.
778 <https://doi.org/10.1016/j.jhep.2012.03.017>

779 Gijbels, E., & Vinken, M. (2017). An Update on Adverse Outcome Pathways Leading to Liver Injury.
780 *Applied in Vitro Toxicology*, 3(4), 283–285. <https://doi.org/10.1089/aivt.2017.0027>

781 Golbraikh, A., Wang, X. S., Zhu, H., & Tropsha, A. (2017). Predictive QSAR Modeling: Methods and
782 Applications in Drug Discovery and Chemical Risk Assessment. In J. Leszczynski, A.
783 Kaczmarek-Kedziera, T. Puzyn, M. G. Papadopoulos, H. Reis, & M. K. Shukla (Eds.), *Handbook*
784 *of Computational Chemistry* (pp. 2303–2340). Springer International Publishing.
785 https://doi.org/10.1007/978-3-319-27282-5_37

786 Guo, L., Gong, H., Tang, T.-L., Zhang, B.-K., Zhang, L.-Y., & Yan, M. (2021). Crizotinib and Sunitinib
787 Induce Hepatotoxicity and Mitochondrial Apoptosis in L02 Cells via ROS and Nrf2 Signaling
788 Pathway. *Frontiers in Pharmacology*, 12.
789 <https://www.frontiersin.org/articles/10.3389/fphar.2021.620934>

790 Guo, R., Gu, J., Zong, S., Wu, M., & Yang, M. (2018). Structure and mechanism of mitochondrial
791 electron transport chain. *Biomedical Journal*, 41(1), 9–20.
792 <https://doi.org/10.1016/j.bj.2017.12.001>

793 Gutsell, S., & Russell, P. (2013). The role of chemistry in developing understanding of adverse outcome
794 pathways and their application in risk assessment. *Toxicology Research*, 2(5), 299–307.
795 <https://doi.org/10.1039/c3tx50024a>

796 Haegler, P., Joerin, L., Krähenbühl, S., & Bouitbir, J. (2017). Hepatocellular Toxicity of Imidazole and
797 Triazole Antimycotic Agents. *Toxicological Sciences*, 157(1), 183–195.
798 <https://doi.org/10.1093/toxsci/kfx029>

799 Hartung, T. (2009). Toxicology for the twenty-first century. *Nature*, 460(7252), Article 7252.
800 <https://doi.org/10.1038/460208a>

801 Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class AdaBoost. *Statistics and Its Interface*, 2(3),
802 349–360. <https://doi.org/10.4310/SII.2009.v2.n3.a8>

803 Herpers, B., Wink, S., Fredriksson, L., Di, Z., Hendriks, G., Vrieling, H., de Bont, H., & van de Water, B.
804 (2016). Activation of the Nrf2 response by intrinsic hepatotoxic drugs correlates with suppression
805 of NF- κ B activation and sensitizes toward TNF α -induced cytotoxicity. *Archives of Toxicology*,
806 90(5), 1163–1179. <https://doi.org/10.1007/s00204-015-1536-3>

807 Höfer, T., Gerner, I., Gundert-Remy, U., Liebsch, M., Schulte, A., Spielmann, H., Vogel, R., & Wettig,
808 K. (2004). Animal testing and alternative approaches for the human health risk assessment under
809 the proposed new European chemicals regulation. *Archives of Toxicology*, 78(10), 549–564.
810 <https://doi.org/10.1007/s00204-004-0577-9>

811 Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S. A., Attene-Ramos, M., Zhao, T., Austin, C. P., &
812 Simeonov, A. (2016). Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction
813 and mechanism characterization. *Nature Communications*, 7(1), Article 1.
814 <https://doi.org/10.1038/ncomms10425>

815 Jaeschke, H., McGill, M. R., & Ramachandran, A. (2012). Oxidant stress, mitochondria, and cell death
816 mechanisms in drug-induced liver injury: Lessons learned from acetaminophen hepatotoxicity.
817 *Drug Metabolism Reviews*, 44(1), 88–106. <https://doi.org/10.3109/03602532.2011.602688>

818 Jia, X., Ciallella, H. L., Russo, D. P., Zhao, L., James, M. H., & Zhu, H. (2021). Construction of a Virtual
819 Opioid Bioprofile: A Data-Driven QSAR Modeling Study to Identify New Analgesic Opioids.
820 *ACS Sustainable Chemistry & Engineering*, 9(10), 3909–3919.
821 <https://doi.org/10.1021/acssuschemeng.0c09139>

822 Jia, X., Wen, X., Russo, D. P., Aleksunes, L. M., & Zhu, H. (2022). Mechanism-driven modeling of
823 chemical hepatotoxicity using structural alerts and an in vitro screening assay. *Journal of*
824 *Hazardous Materials*, 436, 129193. <https://doi.org/10.1016/j.jhazmat.2022.129193>

825 Judson, R. S., Houck, K. A., Kavlock, R. J., Knudsen, T. B., Martin, M. T., Mortensen, H. M., Reif, D.
826 M., Rotroff, D. M., Shah, I., Richard, A. M., & Dix, D. J. (2010). In Vitro Screening of
827 Environmental Chemicals for Targeted Testing Prioritization: The ToxCast Project.
828 *Environmental Health Perspectives*, 118(4), 485–492. <https://doi.org/10.1289/ehp.0901392>

829 Kessler, R. J., Tyson, C. A., & Green, D. E. (1976). Mechanism of uncoupling in mitochondria:
830 Uncouplers as ionophores for cycling cations and protons. *Proceedings of the National Academy*
831 *of Sciences of the United States of America*, 73(9), 3141–3145.

832 Kim, M. T., Huang, R., Sedykh, A., Wang, W., Xia, M., & Zhu, H. (2016). Mechanism Profiling of
833 Hepatotoxicity Caused by Oxidative Stress Using Antioxidant Response Element Reporter Gene
834 Assay Models and Big Data. *Environmental Health Perspectives*, 124(5), 634–641.
835 <https://doi.org/10.1289/ehp.1509763>

836 Knight, A. W., Little, S., Houck, K., Dix, D., Judson, R., Richard, A., McCarroll, N., Akerman, G., Yang,
837 C., Birrell, L., & Walmsley, R. M. (2009). Evaluation of high-throughput genotoxicity assays
838 used in profiling the US EPA ToxCast™ chemicals. *Regulatory Toxicology and Pharmacology*,
839 55(2), 188–199. <https://doi.org/10.1016/j.yrtph.2009.07.004>

840 Krebs, A., Waldmann, T., Wilks, M. F., Vugt-Lussenburg, B. M. A. van, Burg, B. van der, Terron, A.,
841 Steger-Hartmann, T., Ruegg, J., Rovida, C., Pedersen, E., Pallocca, G., Luijten, M., Leite, S. B.,
842 Kustermann, S., Kamp, H., Hoeng, J., Hewitt, P., Herzler, M., Hengstler, J. G., ... Leist, M.
843 (2019). Template for the description of cell-based toxicological test methods to allow evaluation
844 and regulatory use of the data. *ALTEX - Alternatives to Animal Experimentation*, 36(4), Article 4.
845 <https://doi.org/10.14573/altex.1909271>

846 Krewski, D., Andersen, M. E., Mantus, E., & Zeise, L. (2009). Toxicity Testing in the 21st Century:
847 Implications for Human Health Risk Assessment. *Risk Analysis*, 29(4), 474–479.
848 <https://doi.org/10.1111/j.1539-6924.2008.01150.x>

849 Kühlbrandt, W. (2015). Structure and function of mitochondrial membrane protein complexes. *BMC*
850 *Biology*, 13(1), 89. <https://doi.org/10.1186/s12915-015-0201-x>

851 Kutsuno, Y., Itoh, T., Tukey, R. H., & Fujiwara, R. (2014). Glucuronidation of Drugs and Drug-Induced
852 Toxicity in Humanized *UDP-Glucuronosyltransferase 1* Mice. *Drug Metabolism and*
853 *Disposition*, 42(7), 1146–1152. <https://doi.org/10.1124/dmd.114.057083>

854 Labbe, G., Pessayre, D., & Fromenty, B. (2008). Drug-induced liver injury through mitochondrial
 855 dysfunction: Mechanisms and detection during preclinical safety studies. *Fundamental & Clinical*
 856 *Pharmacology*, 22(4), 335–353. <https://doi.org/10.1111/j.1472-8206.2008.00608.x>

857 Langer, D. A., Das, A., Semela, D., Kang-Decker, N., Hendrickson, H., Bronk, S. F., Katusic, Z. S.,
 858 Gores, G. J., & Shah, V. H. (2008). Nitric Oxide Promotes Caspase-Independent Hepatic Stellate
 859 Cell Apoptosis Through the Generation of Reactive Oxygen Species. *Hepatology (Baltimore,*
 860 *Md.)*, 47(6), 1983–1993. <https://doi.org/10.1002/hep.22285>

861 Leach, A. R., & Gillet, V. J. (2007). *An Introduction to Chemoinformatics*. Springer.

862 Lee, W. M. (1993). Acute Liver Failure. *New England Journal of Medicine*, 329(25), 1862–1872.
 863 <https://doi.org/10.1056/NEJM199312163292508>

864 Lee, W. M. (2013). Drug-induced Acute Liver Failure. *Clinics in Liver Disease*, 17(4),
 865 10.1016/j.cld.2013.07.001. <https://doi.org/10.1016/j.cld.2013.07.001>

866 Li, Y., Couch, L., Higuchi, M., Fang, J.-L., & Guo, L. (2012). Mitochondrial Dysfunction Induced by
 867 Sertraline, an Antidepressant Agent. *Toxicological Sciences*, 127(2), 582–591.
 868 <https://doi.org/10.1093/toxsci/kfs100>

869 Liew, C. Y., Lim, Y. C., & Yap, C. W. (2011). Mixed learning algorithms and features ensemble in
 870 hepatotoxicity prediction. *Journal of Computer-Aided Molecular Design*, 25(9), 855.
 871 <https://doi.org/10.1007/s10822-011-9468-3>

872 Liu, R., Yu, X., & Wallqvist, A. (2015). Data-driven identification of structural alerts for mitigating the
 873 risk of drug-induced human liver injuries. *Journal of Cheminformatics*, 7(1), 4.
 874 <https://doi.org/10.1186/s13321-015-0053-y>

875 Lowinger, T. B., Riedl, B., Dumas, J., & Smith, R. A. (2002). Design and Discovery of Small Molecules
 876 Targeting Raf-1 Kinase. *Current Pharmaceutical Design*, 8(25), 2269–2278.
 877 <https://doi.org/10.2174/1381612023393125>

878 Lu, Y.-Y., Zhao, C.-R., Wang, R.-Q., Li, W.-B., & Qu, X.-J. (2015). A novel anticancer diarylurea
879 derivative HL-40 as a multi-kinases inhibitor with good pharmacokinetics in Wistar rats.
880 *Biomedicine & Pharmacotherapy*, 69, 255–259. <https://doi.org/10.1016/j.biopha.2014.11.003>

881 Magana-Mora, A., & Bajic, V. B. (2017). OmniGA: Optimized Omnivariate Decision Trees for
882 Generalizable Classification Models. *Scientific Reports*, 7(1), Article 1.
883 <https://doi.org/10.1038/s41598-017-04281-9>

884 Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrieval*.

885 Mansouri, A., Gattolliat, C.-H., & Asselah, T. (2018). Mitochondrial Dysfunction and Signaling in
886 Chronic Liver Diseases. *Gastroenterology*, 155(3), 629–647.
887 <https://doi.org/10.1053/j.gastro.2018.06.083>

888 Matuz-Mares, D., González-Andrade, M., Araiza-Villanueva, M. G., Vilchis-Landeros, M. M., &
889 Vázquez-Meza, H. (2022). Mitochondrial Calcium: Effects of Its Imbalance in Disease.
890 *Antioxidants*, 11(5), 801. <https://doi.org/10.3390/antiox11050801>

891 McNaughton, R., Huet, G., & Shakir, S. (2014). An investigation into drug products withdrawn from the
892 EU market between 2002 and 2011 for safety reasons and the evidence used to support the
893 decision-making. *BMJ Open*, 4(1), e004221. <https://doi.org/10.1136/bmjopen-2013-004221>

894 Mihajlovic, M., & Vinken, M. (2022). Mitochondria as the Target of Hepatotoxicity and Drug-Induced
895 Liver Injury: Molecular Mechanisms and Detection Methods. *International Journal of Molecular*
896 *Sciences*, 23(6), 3315. <https://doi.org/10.3390/ijms23063315>

897 Miners, J. O., McKinnon, R. A., & Mackenzie, P. I. (2002). Genetic polymorphisms of UDP-
898 glucuronosyltransferases and their functional significance. *Toxicology*, 181–182, 453–456.
899 [https://doi.org/10.1016/S0300-483X\(02\)00449-3](https://doi.org/10.1016/S0300-483X(02)00449-3)

900 Mulliner, D., Schmidt, F., Stolte, M., Spirk, H.-P., Czich, A., & Amberg, A. (2016). Computational
901 Models for Human and Animal Hepatotoxicity with a Global Application Scope. *Chemical*
902 *Research in Toxicology*, 29(5), 757–767. <https://doi.org/10.1021/acs.chemrestox.5b00465>

903 Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T. D., Gramatica, P., Jaworska, J. S.,
 904 Kahn, S., Klopman, G., Marchant, C. A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G. Y.,
 905 Perkins, R., Roberts, D., Schultz, T., Stanton, D. W., van de Sandt, J. J. M., Tong, W., ... Yang,
 906 C. (2005). Current status of methods for defining the applicability domain of (quantitative)
 907 structure-activity relationships. The report and recommendations of ECVAM Workshop 52.
 908 *Alternatives to Laboratory Animals: ATLA*, 33(2), 155–173.
 909 <https://doi.org/10.1177/026119290503300209>
 910 OECD. (2017). *Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated*
 911 *Approaches to Testing and Assessment (IATA)*. OECD. <https://doi.org/10.1787/44bb06c1-en>
 912 Park, B. K., Kitteringham, N. R., Maggs, J. L., Pirmohamed, M., & Williams, D. P. (2005). The role of
 913 metabolic activation in drug-induced hepatotoxicity. *Annual Review of Pharmacology and*
 914 *Toxicology*, 45(1), 177–202. <https://doi.org/10.1146/annurev.pharmtox.45.120403.100058>
 915 Patlewicz, G., Rodford, R., & Walker, J. D. (2003). Quantitative structure-activity relationships for
 916 predicting skin and eye irritation. *Environmental Toxicology and Chemistry*, 22(8), 1862–1869.
 917 <https://doi.org/10.1897/01-439>
 918 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M.,
 919 Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011).
 920 Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*, 6.
 921 Perkins, E. J., Ashauer, R., Burgoon, L., Conolly, R., Landesmann, B., Mackay, C., Murphy, C. A.,
 922 Pollesch, N., Wheeler, J. R., Zupanec, A., & Scholz, S. (2019). Building and Applying
 923 Quantitative Adverse Outcome Pathway Models for Chemical Hazard and Risk Assessment.
 924 *Environmental Toxicology and Chemistry*, 38(9), 1850–1865. <https://doi.org/10.1002/etc.4505>
 925 Pessayre, D., Mansouri, A., Haouzi, D., & Fromenty, B. (1999). Hepatotoxicity due to mitochondrial
 926 dysfunction. *Cell Biology and Toxicology*, 15(6), 367–373.
 927 <https://doi.org/10.1023/a:1007649815992>

928 Ramachandran, A., Visschers, R. G. J., Duan, L., Akakpo, J. Y., & Jaeschke, H. (2018). Mitochondrial
 929 dysfunction as a mechanism of drug-induced hepatotoxicity: Current understanding and future
 930 perspectives. *Journal of Clinical and Translational Research*, 4(1), 75–100.
 931 Rogers, D., & Hahn, M. (2010). Extended-Connectivity Fingerprints. *Journal of Chemical Information*
 932 *and Modeling*, 50(5), 742–754. <https://doi.org/10.1021/ci100050t>
 933 Russo, D. P., Strickland, J., Karmaus, A. L., Wang, W., Shende, S., Hartung, T., Aleksunes, L. M., &
 934 Zhu, H. (2019). Nonanimal Models for Acute Toxicity Evaluations: Applying Data-Driven
 935 Profiling and Read-Across. *Environmental Health Perspectives*, 127(4), 047001.
 936 <https://doi.org/10.1289/EHP3614>
 937 Russo, D. P., & Zhu, H. (2022). High-Throughput Screening Assay Profiling for Large Chemical
 938 Databases. In H. Zhu & M. Xia (Eds.), *High-Throughput Screening Assays in Toxicology* (pp.
 939 125–132). Springer US. https://doi.org/10.1007/978-1-0716-2213-1_12
 940 Russo, D. P., Zorn, K. M., Clark, A. M., Zhu, H., & Ekins, S. (2018). Comparing Multiple Machine
 941 Learning Algorithms and Metrics for Estrogen Receptor Binding Prediction. *Molecular*
 942 *Pharmaceutics*, 15(10), 4361–4370. <https://doi.org/10.1021/acs.molpharmaceut.8b00546>
 943 Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., & Todeschini, R. (2012). Comparison
 944 of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, 17(5),
 945 4791–4810. <https://doi.org/10.3390/molecules17054791>
 946 Seal, S., Carreras-Puigvert, J., Trapotsi, M.-A., Yang, H., Spjuth, O., & Bender, A. (2022). Integrating
 947 cell morphology with gene expression and chemical structure to aid mitochondrial toxicity
 948 detection. *Communications Biology*, 5(1), Article 1. <https://doi.org/10.1038/s42003-022-03763-5>
 949 Sedykh, A., Fourches, D., Duan, J., Hucke, O., Garneau, M., Zhu, H., Bonneau, P., & Tropsha, A. (2013).
 950 Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake,
 951 efflux and interactions. *Pharmaceutical Research*, 30(4), 996–1007.
 952 <https://doi.org/10.1007/s11095-012-0935-x>

953 Senior, J. R. (2014). Evolution of the Food and Drug Administration Approach to Liver Safety
 954 Assessment for New Drugs: Current Status and Challenges. *Drug Safety*, 37(1), 9–17.
 955 <https://doi.org/10.1007/s40264-014-0182-7>

956 Song, Y., & Villeneuve, D. L. (2021). AOP Report: Uncoupling of Oxidative Phosphorylation Leading to
 957 Growth Inhibition via Decreased Cell Proliferation. *Environmental Toxicology and Chemistry*,
 958 40(11), 2959–2967. <https://doi.org/10.1002/etc.5197>

959 Srivastava, A., Maggs, J. L., Antoine, D. J., Williams, D. P., Smith, D. A., & Park, B. K. (2010). Role of
 960 Reactive Metabolites in Drug-Induced Hepatotoxicity. In J. Uetrecht (Ed.), *Adverse Drug*
 961 *Reactions* (pp. 165–194). Springer. https://doi.org/10.1007/978-3-642-00663-0_7

962 Stocker, R., & Keaney, J. F. (2004). Role of Oxidative Modifications in Atherosclerosis. *Physiological*
 963 *Reviews*, 84(4), 1381–1478. <https://doi.org/10.1152/physrev.00047.2003>

964 Stratton, C. F., Newman, D. J., & Tan, D. S. (2015). Cheminformatic comparison of approved drugs from
 965 natural product versus synthetic origins. *Bioorganic & Medicinal Chemistry Letters*, 25(21),
 966 4802–4807. <https://doi.org/10.1016/j.bmcl.2015.07.014>

967 Stucki, A. O., Barton-Maclaren, T. S., Bhuller, Y., Henriquez, J. E., Henry, T. R., Hirn, C., Miller-Holt,
 968 J., Nagy, E. G., Perron, M. M., Ratzlaff, D. E., Stedeford, T. J., & Clippinger, A. J. (2022). Use of
 969 new approach methodologies (NAMs) to meet regulatory requirements for the assessment of
 970 industrial chemicals and pesticides for effects on human health. *Frontiers in Toxicology*, 4,
 971 964553. <https://doi.org/10.3389/ftox.2022.964553>

972 Sun, C.-K., Zhang, X.-Y., Sheard, P. W., Mabuchi, A., & Wheatley, A. M. (2005). Change in
 973 mitochondrial membrane potential is the key mechanism in early warm hepatic ischemia-
 974 reperfusion injury. *Microvascular Research*, 70(1), 102–110.
 975 <https://doi.org/10.1016/j.mvr.2005.04.003>

976 Tang, W., Jiang, Y.-F., Ponnusamy, M., & Diallo, M. (2014). Role of Nrf2 in chronic liver disease. *World*
 977 *Journal of Gastroenterology : WJG*, 20(36), 13079–13087.
 978 <https://doi.org/10.3748/wjg.v20.i36.13079>

979 Terada, H. (1990). Uncouplers of oxidative phosphorylation. *Environmental Health Perspectives*, 87,
980 213–218. <https://doi.org/10.1289/ehp.9087213>

981 Todorović Vukotić, N., Đorđević, J., Pejić, S., Đorđević, N., & Pajović, S. B. (2021). Antidepressants-
982 and antipsychotics-induced hepatotoxicity. *Archives of Toxicology*, 95(3), 767–789.
983 <https://doi.org/10.1007/s00204-020-02963-4>

984 Tollefsen, K. E., Scholz, S., Cronin, M. T., Edwards, S. W., de Knecht, J., Crofton, K., Garcia-Reyero,
985 N., Hartung, T., Worth, A., & Patlewicz, G. (2014). Applying Adverse Outcome Pathways
986 (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regulatory*
987 *Toxicology and Pharmacology*, 70(3), 629–640. <https://doi.org/10.1016/j.yrtph.2014.09.009>

988 Tong, W., Hong, H., Xie, Q., Shi, L., Fang, H., & Perkins, R. (2005). Assessing QSAR Limitations—A
989 Regulatory Perspective. *Current Computer - Aided Drug Design*, 1(2), 195–205.
990 <https://doi.org/10.2174/1573409053585663>

991 Tropsha, A., Gramatica, P., & Gombar, V. K. (2003). The Importance of Being Earnest: Validation is the
992 Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR &*
993 *Combinatorial Science*, 22(1), 69–77. <https://doi.org/10.1002/qsar.200390007>

994 Van Veggel, M., Westerman, E., & Hamberg, P. (2018). Clinical Pharmacokinetics and
995 Pharmacodynamics of Panobinostat. *Clinical Pharmacokinetics*, 57(1), 21–29.
996 <https://doi.org/10.1007/s40262-017-0565-x>

997 Vapnik, V. N. (2000). Methods of Pattern Recognition. In V. N. Vapnik (Ed.), *The Nature of Statistical*
998 *Learning Theory* (pp. 123–180). Springer. https://doi.org/10.1007/978-1-4757-3264-1_6

999 Vinken, M. (2013). The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology*,
1000 312, 158–165. <https://doi.org/10.1016/j.tox.2013.08.011>

1001 Vogel, F., Bornhövd, C., Neupert, W., & Reichert, A. S. (2006). Dynamic subcompartmentalization of the
1002 mitochondrial inner membrane. *Journal of Cell Biology*, 175(2), 237–247.
1003 <https://doi.org/10.1083/jcb.200605138>

1004 Wang, K. (2014). Molecular mechanisms of liver injury: Apoptosis or necrosis. *Experimental and*
 1005 *Toxicologic Pathology*, 66(8), 351–356. <https://doi.org/10.1016/j.etp.2014.04.004>

1006 Wang, W., Kim, M. T., Sedykh, A., & Zhu, H. (2015). Developing Enhanced Blood–Brain Barrier
 1007 Permeability Models: Integrating External Bio-Assay Data in QSAR Modeling. *Pharmaceutical*
 1008 *Research*, 32(9), 3055–3065. <https://doi.org/10.1007/s11095-015-1687-1>

1009 Wang, X., & Michaelis, E. (2010). Selective neuronal vulnerability to oxidative stress in the brain.
 1010 *Frontiers in Aging Neuroscience*, 2.
 1011 <https://www.frontiersin.org/articles/10.3389/fnagi.2010.00012>

1012 Wang, Y., Gao, H., Na, X., Dong, S., Dong, H., Yu, J., Jia, L., & Wu, Y. (2016). Aniline Induces
 1013 Oxidative Stress and Apoptosis of Primary Cultured Hepatocytes. *International Journal of*
 1014 *Environmental Research and Public Health*, 13(12), 1188.
 1015 <https://doi.org/10.3390/ijerph13121188>

1016 Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: A public
 1017 information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*,
 1018 37(Web Server issue), W623. <https://doi.org/10.1093/nar/gkp456>

1019 Weiler, N., Schlotmann, A., Anton Schnitzbauer, A., Zeuzem, S., & Welker, M.-W. (2020). The
 1020 Epidemiology of Acute Liver Failure. *Deutsches Ärzteblatt International*, 117(4), 43–50.
 1021 <https://doi.org/10.3238/arztebl.2020.0043>

1022 Wilkening, S., Stahl, F., & Bader, A. (2003). Comparison of primary human hepatocytes and hepatoma
 1023 cell line Hepg2 with regard to their biotransformation properties. *Drug Metabolism and*
 1024 *Disposition*, 31(8), 1035–1042. <https://doi.org/10.1124/dmd.31.8.1035>

1025 Win, S., Than, T. A., Zhang, J., Oo, C., Min, R. W. M., & Kaplowitz, N. (2018). New insights into the
 1026 role and mechanism of c-Jun-N-terminal kinase signaling in the pathobiology of liver diseases.
 1027 *Hepatology (Baltimore, Md.)*, 67(5), 2013–2024. <https://doi.org/10.1002/hep.29689>

1028 Wittig, I., Carrozzo, R., Santorelli, F. M., & Schägger, H. (2006). Supercomplexes and subcomplexes of
 1029 mitochondrial oxidative phosphorylation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*,
 1030 1757(9), 1066–1072. <https://doi.org/10.1016/j.bbabi.2006.05.006>

1031 Xuan, J., Chen, S., Ning, B., Tolleson, W. H., & Guo, L. (2016). Development of HepG2-derived cells
 1032 expressing cytochrome P450s for assessing metabolism-associated drug-induced liver toxicity.
 1033 *Chemico-Biological Interactions*, 255, 63–73. <https://doi.org/10.1016/j.cbi.2015.10.009>

1034 Zakharov, A. V., Peach, M. L., Sitzmann, M., & Nicklaus, M. C. (2014). QSAR Modeling of Imbalanced
 1035 High-Throughput Screening Data in PubChem. *Journal of Chemical Information and Modeling*,
 1036 54(3), 705–712. <https://doi.org/10.1021/ci400737s>

1037 Zhang, L., Fourches, D., Sedykh, A., Zhu, H., Golbraikh, A., Ekins, S., Clark, J., Connelly, M. C., Sigal,
 1038 M., Hodges, D., Guiguemde, A., Guy, R. K., & Tropsha, A. (2013). Discovery of Novel
 1039 Antimalarial Compounds Enabled by QSAR-Based Virtual Screening. *Journal of Chemical*
 1040 *Information and Modeling*, 53(2), 475–492. <https://doi.org/10.1021/ci300421n>

1041 Zhang, R., Chen, Z., Wang, B., Li, Y., Mu, Y., & Li, X. (2023). Modeling and Insights into the Structural
 1042 Characteristics of Chemical Mitochondrial Toxicity. *ACS Omega*, 8(35), 31675–31682.
 1043 <https://doi.org/10.1021/acsomega.3c01725>

1044 Zhang, Y., Xue, W., Zhang, W., Yuan, Y., Zhu, X., Wang, Q., Wei, Y., Yang, D., Yang, C., Chen, Y.,
 1045 Sun, Y., Wang, S., Huang, K., & Zheng, L. (2020). Histone methyltransferase G9a protects
 1046 against acute liver injury through GSTP1. *Cell Death & Differentiation*, 27(4), Article 4.
 1047 <https://doi.org/10.1038/s41418-019-0412-8>

1048 Zhao, L., Russo, D. P., Wang, W., Aleksunes, L. M., & Zhu, H. (2020). Mechanism-Driven Read-Across
 1049 of Chemical Hepatotoxicants Based on Chemical Structures and Biological Data. *Toxicological*
 1050 *Sciences: An Official Journal of the Society of Toxicology*, 174(2), 178–188.
 1051 <https://doi.org/10.1093/toxsci/kfaa005>

1052 Zheng, W., & Tropsha, A. (2000). Novel Variable Selection Quantitative Structure–Property Relationship
 1053 Approach Based on the k -Nearest-Neighbor Principle. *Journal of Chemical Information and*
 1054 *Computer Sciences*, 40(1), 185–194. <https://doi.org/10.1021/ci980033m>

1055 Zhu, H. (2020). Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annual Review of*
 1056 *Pharmacology and Toxicology*, 60(1), 573–589. [https://doi.org/10.1146/annurev-pharmtox-](https://doi.org/10.1146/annurev-pharmtox-010919-023324)
 1057 010919-023324

1058 Zhu, H., Zhang, J., Kim, M. T., Boison, A., Sedykh, A., & Moran, K. (2014). Big Data in Chemical
 1059 Toxicity Research: The Use of High-Throughput Screening Assays To Identify Potential
 1060 Toxicants. *Chemical Research in Toxicology*, 27(10), 1643–1651.
 1061 <https://doi.org/10.1021/tx500145h>

1062 Zhu, R., Wang, Y., Zhang, L., & Guo, Q. (2012). Oxidative stress and liver disease. *Hepatology*
 1063 *Research*, 42(8), 741–749. <https://doi.org/10.1111/j.1872-034X.2012.00996.x>

Table 1. Selected bioassays and their correlation with *in vivo* hepatotoxicity

PubChem Assay ID	Description	CCR	PPV	Specificity	Sensitivity	<i>p</i> value
*1346979	Caspase-3/7 induction in CHO-K1 cells by small molecules, qHTS cell viability counter screen	0.54	0.73	0.95	0.14	0.0005
*1347034	Caspase-3/7 induction in HepG2 cells by small molecules, qHTS assay: Summary	0.52	0.78	0.98	0.067	0.013
*720635	qHTS assay for small molecule disruptors of the mitochondrial membrane potential	0.53	0.67	0.95	0.11	0.014
*1224896	qHTS assay to identify small molecule agonists of H2AX: Summary	0.52	0.78	0.98	0.058	0.014
*1346981	Caspase-3/7 induction in HepG2 cells by small molecules, qHTS cell viability counter screen	0.52	0.72	0.97	0.072	0.022
1224874	qHTS RealTime-Glo MT Cell Viability Assay in HEK293 cells - 40 hour	0.53	0.67	0.94	0.12	0.065
720552	qHTS assay for small molecule agonists of the p53 signaling pathway: Summary	0.53	0.70	0.96	0.09	0.074
1224868	qHTS RealTime-Glo MT Cell Viability Assay in HEK293 cells - 32 hour	0.53	0.67	0.95	0.11	0.081
651631	qHTS assay for small molecule agonists of the p53 signaling pathway	0.52	0.69	0.97	0.08	0.15
651633	qHTS assay for small molecule agonists of the p53 signaling pathway - cell viability	0.52	0.67	0.96	0.08	0.23

*statistically relevant relationships to hepatotoxicity using Fisher's exact test ($p < 0.05$)

qHTS: quantitative high-throughput screening

Figures

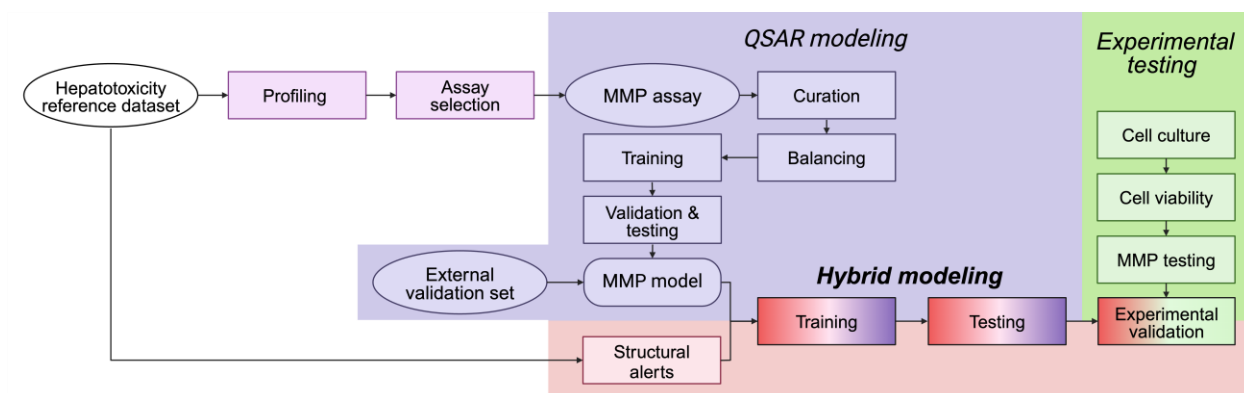


Figure 1. Overview of a hybrid modeling approach for human hepatotoxicity. Different colors represent the components that constitute the main framework of the hybrid model. Created with BioRender.com.

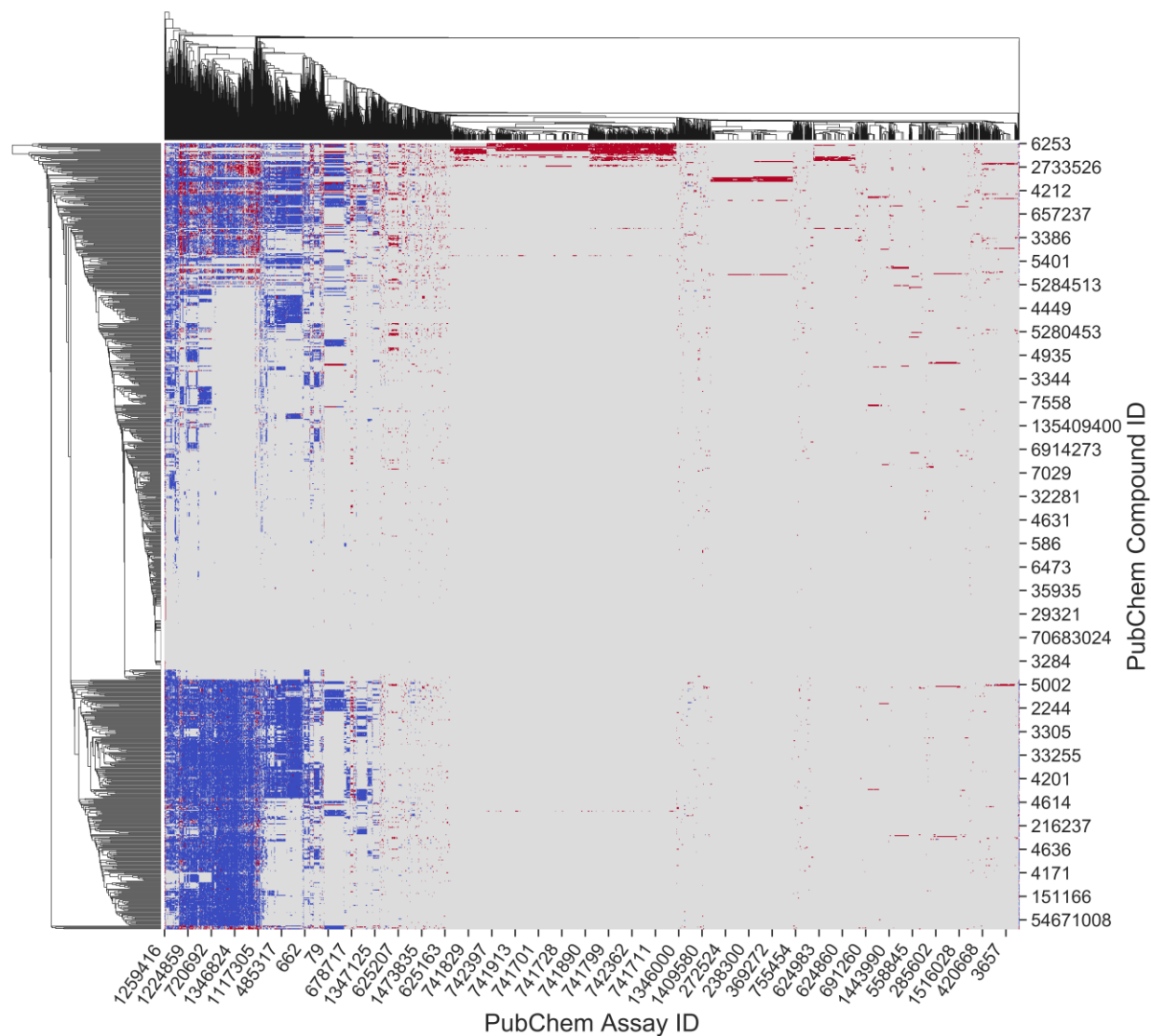


Figure 2. Bioprofile of 869 hepatotoxicity reference set compounds, including US Food and Drug Administration (FDA)-approved drugs. Active results are red squares, inactive results as blue squares, and inconclusive or untested results as gray squares.

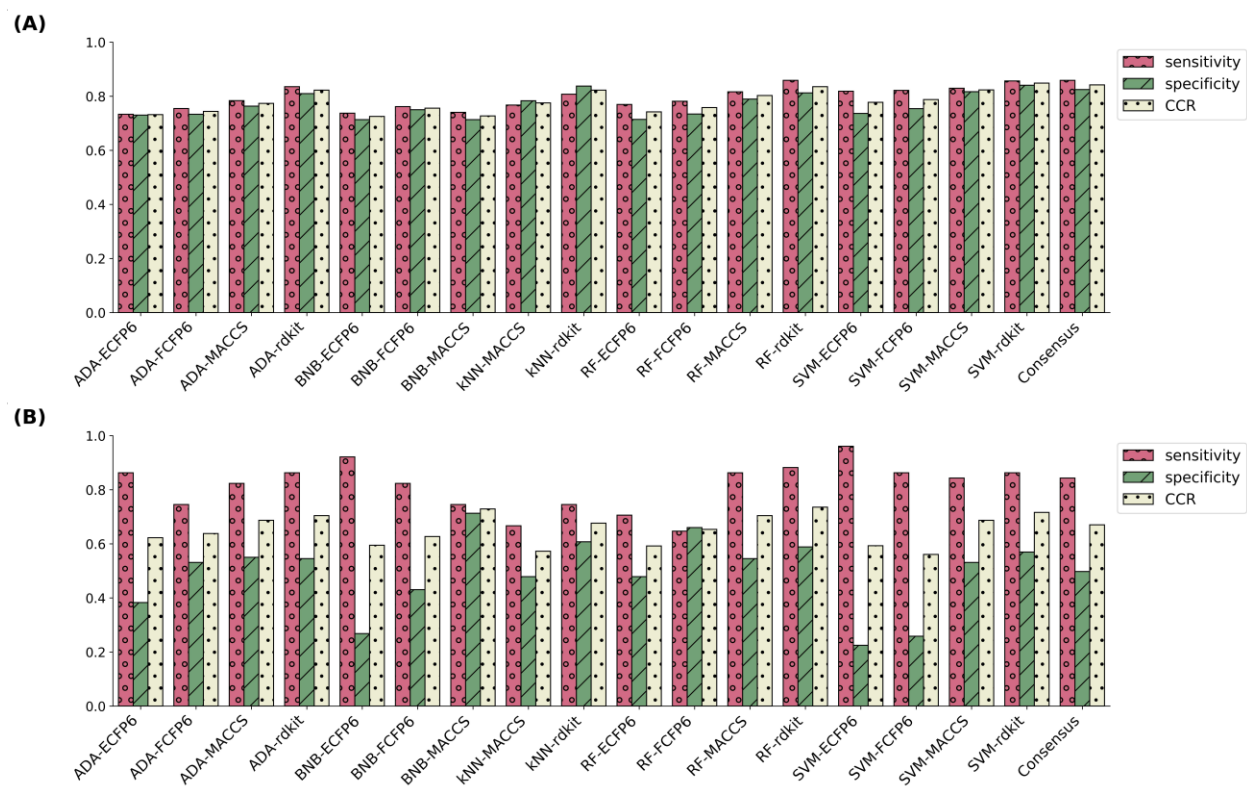


Figure 3. Performance of all resulting models. (A) cross-validation of 2536 training set compounds in AID 720635; (B) external validation of 260 unknown natural products.

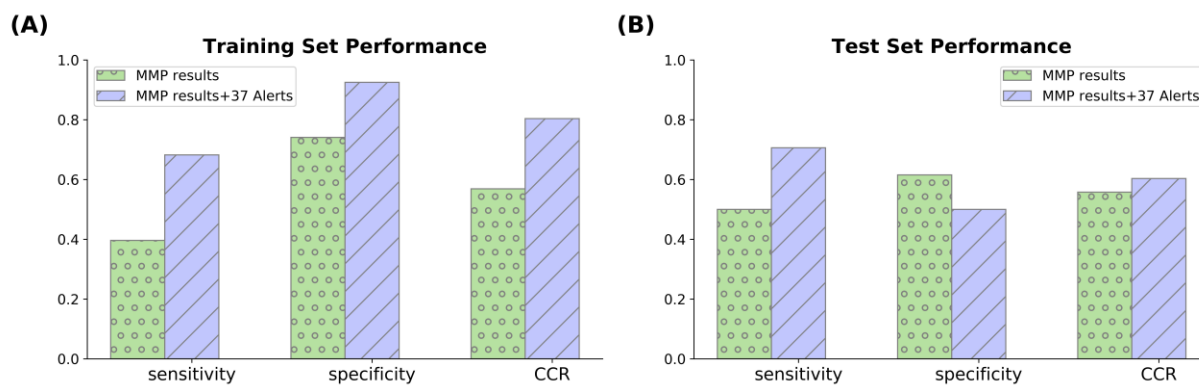


Figure 4. Comparative performance of hepatotoxicity predictions leveraging structural alerts and MMP assay results: predictions for (A) the 702 training set compounds and (B) 179 test set compounds shown as sensitivity, specificity, and CCR. The MMP results comprise MMP experimental outcomes and QSAR predictions.

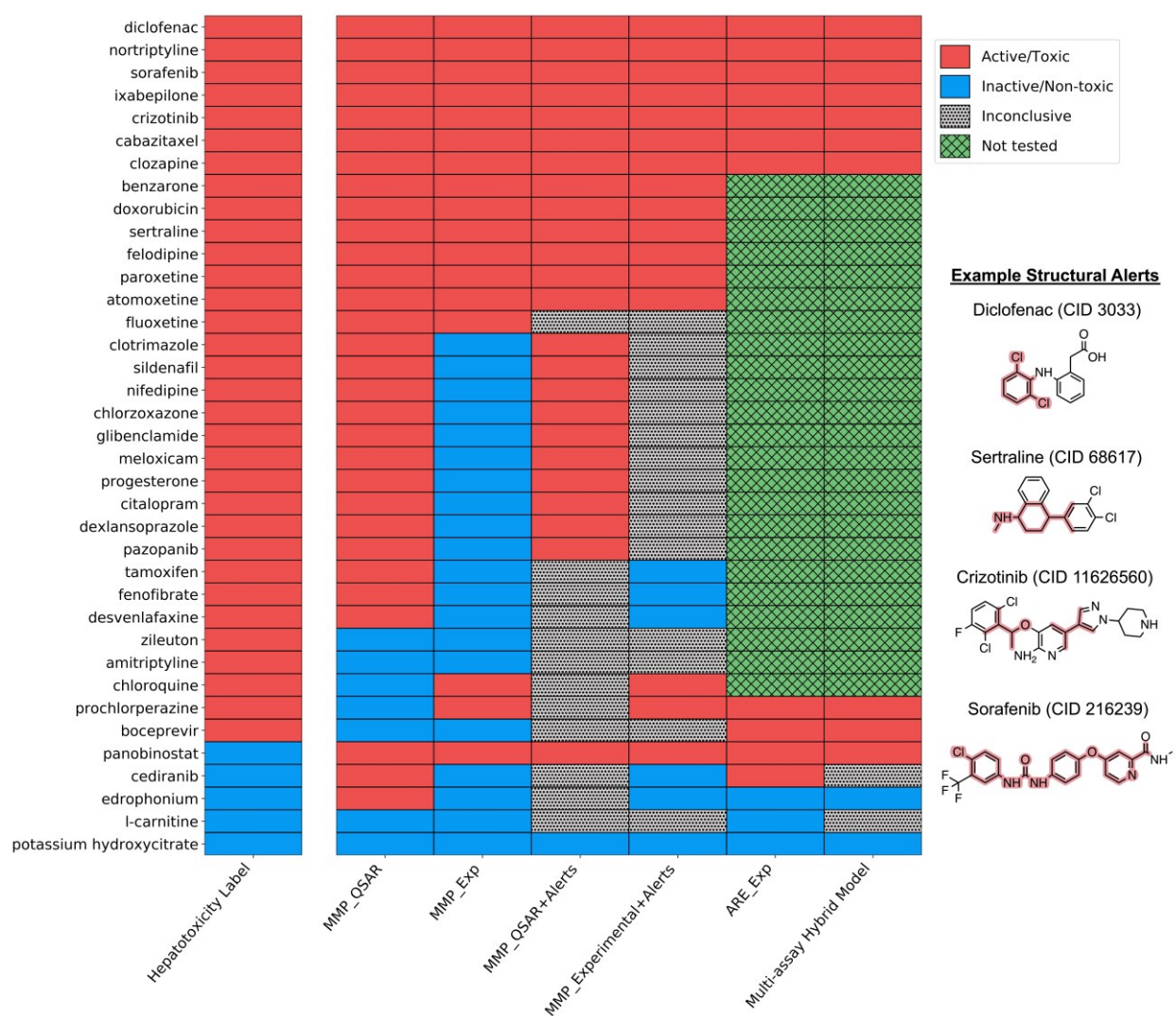


Figure 5. Hepatotoxicity predictions for 37 drug molecules using structure alerts and predictive/experimental outcomes of the mitochondrial MMP and oxidative stress ARE assays.