Article

Modeling the Bioactivation and Subsequent Reactivity of Drugs

Tyler B. Hughes, Noah Flynn, Na Le Dang, and S. Joshua Swamidass*
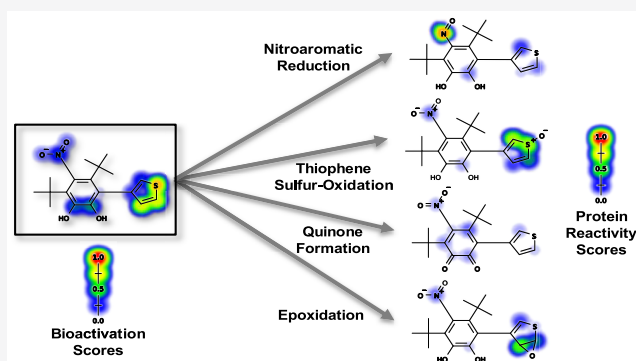
ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆂🅸 Supporting Information

ABSTRACT: Electrophilically reactive drug metabolites are implicated in many adverse drug reactions. In this mechanism— termed bioactivation—metabolic enzymes convert drugs into reactive metabolites that often conjugate to nucleophilic sites within biological macromolecules like proteins. Toxic metabolite-product adducts induce severe immune responses that can cause sometimes fatal disorders, most commonly in the form of liver injury, blood dyscrasia, or the dermatologic conditions toxic epidermal necrolysis and Stevens−Johnson syndrome. This study models four of the most common metabolic transformations that result in bioactivation: quinone formation, epoxidation, thiophene sulfur-oxidation, and nitroaromatic reduction, by synthesizing models of metabolism and reactivity. First, the metabolism models predict the formation probabilities of all possible metabolites among the pathways studied. Second, the exact structures of these metabolites are enumerated. Third, using these structures, the reactivity model predicts the reactivity of each metabolite. Finally, a feedfoward neural network converts the metabolism and reactivity predictions to a bioactivation prediction for each possible metabolite. These bioactivation predictions represent the joint probability that a metabolite forms and that this metabolite subsequently conjugates to protein or glutathione. Among molecules bioactivated by these pathways, we predicted the correct pathway with an AUC accuracy of 89.98%. Furthermore, the model predicts whether molecules will be bioactivated, distinguishing bioactivated and nonbioactivated molecules with 81.06% AUC. We applied this algorithm to withdrawn drugs. The known bioactivation pathways of alclofenac and benzbromarone were identified by the algorithm, and high probability bioactivation pathways not yet confirmed were identified for safrazine, zimelidine, and astemizole. This bioactivation model—the first of its kind that jointly considers both metabolism and reactivity—enables drug candidates to be quickly evaluated for a toxicity risk that often evades detection during preclinical trials. The XenoSite bioactivation model is available at http://swami.wustl.edu/xenosite/p/bioactivation.

INTRODUCTION

Adverse Drug Reactions (ADRs) are a major challenge for global public health. Independent investigations of diverse populations implicated ADRs in 6.5%,[1] 8.1%,[2] 8.7%,[3] and 12.8%[4] of hospital admissions. Similarly, a meta-analysis of U.S. hospital prospective studies found that 6.7% of patients had severe ADRs, with a fatality rate of 0.32%.[5] Extrapolating from those results, the FDA estimated that annual ADRs in the U.S. cause over 2,216,000 hospitalizations and more than 106,000 deaths.[6] Furthermore, predicated on the meta-analysis's accuracy, the FDA proposed that ADRs are the fourth leading cause of death in the U.S., exceeding automobile deaths, diabetes, AIDS, pulmonary disease, and pneumonia.[6] Some ADRs are traceable to the pharmacological effects of certain drugs, and this mechanistic understanding can inform efforts to reduce risk.[7] However, a subset of ADRs—idiosyncratic adverse drug reactions (IADRs)—has elusive etiologies.

These IADRs strike seemingly at random, with unpredictable and often severe symptoms. Most commonly, IADRs cause liver disorders but can also induce dangerous skin diseases, including Stevens-Johnson syndrome and toxic epidermal necrolysis, as well as dangerous blood disorders such as agranulocytosis or aplastic anemia.[8−18] In the U.S., IADRs are responsible for about half of all acute liver failure cases and 15% of liver transplants.[19] Nevertheless, IADRs are rare overall, only occurring in about 1 in 10,000 to 1 in 100,000 patients.[20] As a result, many IADR-causing drugs can slip through all stages of preclinical trials, which even in their largest phase generally only have about 3,000 patients.[8] After approval and market release, however, exposure to much larger patient populations can reveal a drug's hidden risk. Indeed, already-approved drugs are most commonly withdrawn from the market due to intolerable numbers of IADR cases.[21−25] Even if a drug is not withdrawn, it may be
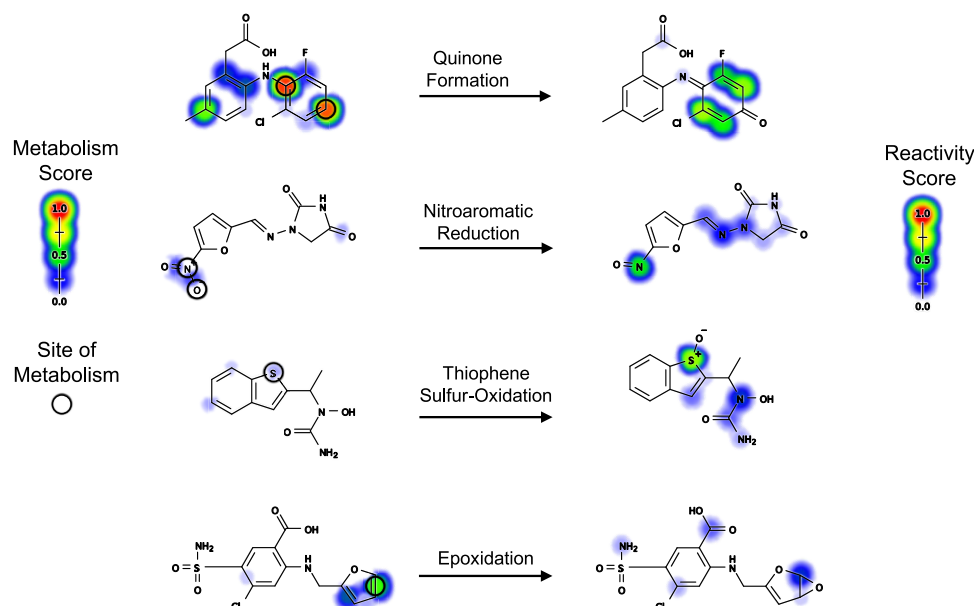
Figure 1. This study models four common bioactivation pathways: quinone formation, nitroaromatic reduction, thiophene sulfur-oxidation, and epoxidation. Top row, lumiracoxib, a cyclooxygenase-2 selective inhibitor, was withdrawn from several countries after several cases of severe liver damage.[27,28] This toxicity was traced to the formation of a reactive quinone-imine metabolite that conjugates to off-target proteins, inducing deleterious immune responses.[27,28] Second row, nitrofurantoin, an antibiotic, carries a risk of acute liver failure,[29] which is thought to be caused by reduction of nitrofurantoin's nitroaromatic group to a reactive nitroso.[30] Third row, zileuton, a 5-lipoxygenase inhibitor used to treat asthma, has been restricted in its use due to rare cases of severe hepatotoxicity, which has been traced to oxidation of the sulfur in its thiophene motif, producing a highly reactive sulfur-oxide.[31] Bottom row, furosemide, a diuretic, confers a risk of idiosyncratic hepatitis due to production of a reactive epoxide metabolite.[32−36]

labeled with a "black box" warning by the FDA, significantly curtailing its profitability.[26]

Devising early detection methods for IADRs would reduce patient morbidity and mortality. Furthermore, pharmaceutical developers would avoid heavily investing in drugs that ultimately are too risky to use. Unfortunately, IADRs are generally intractable to study and difficult to replicate in humans or animal models. It is not clear why IADRs only affect some individuals, or why the same drug can cause different IADRs in different patients. However, a growing body of evidence suggests that many IADRs are induced by a specific mechanism: bioactivation, the focus of this study (Figure 1).

In bioactivation, enzymes convert drugs into electrophilically reactive metabolites that covalently bind to nucleophilic sites within biological macromolecules, including DNA and (off-target) proteins. Metabolite-DNA adducts may be mutagenic or even carcinogenic,[19,37,38] and metabolite-protein adducts can disrupt protein function or trigger toxic immune responses.[39−41] In fact, many IADRs have been linked to overzealous autoimmune attacks set off by the production and covalent binding of reactive metabolites. Due to the threat conferred by bioactivation, drug developers strive to avoid advancing candidates that produce reactive metabolites, in order to reduce the risk of investing in IADR-causing drugs.

Screening assays for reactive metabolites are often used but have some limitations. Reactive metabolite trapping studies[42] and covalent binding studies[43,44] are commonly used techniques to detect and understand bioactivation. Trapping studies detect if conjugates have formed and can be used to characterize reactive metabolite structure. Typically, a trapping agent, e.g., the use of glutathione (GSH) or cyanide as proxies for protein and DNA, that has a high likelihood of conjugating to reactive

metabolites is selected.[45] Formation of a GSH conjugate indicates the presence of a reactive metabolite, and the GSH conjugate can be detected via mass spectrometry.[42] On the other hand, covalent binding studies are able to quantify the extent of conjugation through the use of a radiolabeled drug.[46] In general, screening assays may not accurately reflect endogenous metabolism, consume time and resources, and require physical synthesis of each compound under consideration.

Instead, computational models have the potential to rapidly screen possible structures for bioactivation risk, thereby flagging problematic molecules or providing a short list of molecules for experimental validation. A widely used and simple method for identifying problematic structures is to cross-reference a drug candidate to a list of structural alerts.[32] Usually, structural alerts contain substructures that have known bioactivation mechanisms that contribute to documented toxicity risk. However, structural alerts have limited utility—they are not sufficient to declare a molecule as toxic, and they can misclassify toxic molecules as safe. Since structural alerts are determined retrospectively, they have no predictive power for new and understudied motifs.[47] However, there is a lack of more complex models that can generalize to understudied cases and account for contextual nuances of drug metabolism.

In this study, we build a model that jointly models metabolism and reactivity, thereby producing bioactivation predictions. We model four types of metabolism that often produce reactive metabolites: quinone formation, nitroaromatic reduction, thiophene sulfur-oxidation, and epoxidation (Figure 1). These pathways are chosen because we have well-developed models for these metabolic routes, including an accurate model of quinone formation,[48] epoxidation,[49] and a phase I metabolism model that includes predictions for nitroaromatic reduction and

thiophene sulfur-oxidation.[50] Since the reaction type is accounted for during prediction, we can infer the structure of the resultant product using a structure inference model.[51] Next, to each inferred structure we apply a previously developed model for predicting reactivity to protein or GSH. Once we have predictions regarding metabolism of the input molecule and reactivity of its inferred metabolites, we can use these predictions to train a deep neural network that will predict bioactivation at both the molecule- and pathway-level. Molecule-level bioactivation aims to predict whether the input molecule will undergo bioactivation, while pathway-level bioactivation aims to predict which of the metabolic transformations and inferred structures will lead to bioactivation.

■ METHODS

Bioactivation Training Data. We assemble a heterogeneous data set of bioactivation reactions from the literature-derived Accelrys Metabolite Database (AMD). Each reaction takes place in humans, human cells, or human liver microsomes. Four types of metabolic transformations are extracted: quinone formation, epoxidation, nitro-aromatic reduction, and thiophene sulfur-oxidation. Each of these metabolic transformations are well-known bioactivation mechanisms, and in previous work, we built models that predict whether molecules will be subject to each type of metabolism. Quinone formation[48] and epoxidation[49] were modeled independently, and nitroaromatic reduction and thiophene sulfur-oxidation were included in a model of diverse phase I reactions.[50]

In total, we extract from the experimental data 210 quinone formations, 174 epoxidations, 4 nitroaromatic reductions, and 15 thiophene sulfur-oxidations. For each parent molecule, we enumerate all possible metabolites for each of these four pathways, producing 7580 quinone formations, 6100 epoxidations, 48 nitroaromatic reductions, and 15 thiophene sulfur-oxidations. We then select for bioactivated parent molecules by taking the set of reactions for each of the four types of metabolic transformations and retaining those parent molecules whose possible metabolites are known to directly conjugate to protein or glutathione (GSH). The bioactivated parent molecule's metabolic transformation and resultant, conjugating metabolite are labeled as an experimentally observed bioactivation pathway. However, the same parent molecule can be present multiple times if it is indicated with multiple possible, experimentally observed bioactivation pathways. To handle this case, we merge all duplicate parent molecules into a single representation per unique parent molecule, with all of its experimentally observed bioactivation pathways labeled.

In contrast with an experimentally observed bioactivation pathway, a parent molecule might undergo one of the foir studied metabolic transformations, but the resultant metabolite is not known to conjugate to protein or GSH. Cases where a bioactivated parent molecule only contains bioactivation pathways would not be a good test of the model's ability to predict the correct pathway within the bioactivated parent molecules. Instead, we want the model to clearly distinguish between pathways that are experimentally observed to bioactivate and pathways that are experimentally unobserved to bioactivate. As a final step, we filter out any molecules that did not have at least one experimentally unobserved bioactivation pathway. This procedure produces a total of 332 bioactivated training molecules.

We also select molecules not experimentally observed to be bioactivated. In this case, we are referring to all molecules that could potentially undergo at least one of the four metabolic transformations being studied that precede reactive metabolite formation, but whose resultant metabolites are not experimentally observed to conjugate to protein or GSH. To select these molecules, we start with the same set of reactions from the AMD. Next, we filter out all molecules that are intrinsically reactive: those that are known to directly conjugate to protein or GSH. Additionally, we remove all molecules that belong to the set of bioactivated molecules. Finally, we select all molecules with at least one possible bioactivation pathway among the four metabolic

transformations considered in this study. This procedure produces a large pool of 30690 negative molecules.

From the pool of negative molecules, we randomly select 332 molecules—equivalent to the total number of bioactivated molecules—to form a final data set of 664 molecules. The remaining negative molecules are randomly batched into external test sets, each consisting of 332 molecules, with the same proportion of molecules representing each pathway as the negative molecules from the training set. None of these molecules are considered during training. Instead, they represent additional tests for the final trained model to evaluate how well the model generalizes to unseen cases.

Under the AMD licensing agreement, we are not able to share the exact chemical structures of the product metabolites and their corresponding reactions used in the data set. However, we provide all reaction and molecule AMD registry numbers, as well as the parent molecule structures, in the Supporting Information, which is enough information to reconstruct the data set and replicate our results.

DrugBank. DrugBank is used as another source of external data and was downloaded from www.drugbank.ca (accessed 2020-07-03). We later use the DrugBank external data set during evaluation of the final bioactivation model's ability to generate hypotheses for the toxicity drivers of withdrawn drugs. To identify withdrawn, small-molecule drugs, we filter entries designated by both the "small-molecule" annotation and the "withdrawn" group. This release of DrugBank includes the structures of 221 drugs. Prior to submission to the model, we remove any drugs also present in our training data, resulting in a total of 201 molecules. We provide all withdrawn drugs assessed by the model and their structural information in the Supporting Information.

Bioactivation Descriptors. For this study, we synthesize several previous models of metabolism and reactivity to design specific bioactivation descriptors that are inputs to a neural network (Figure 2).
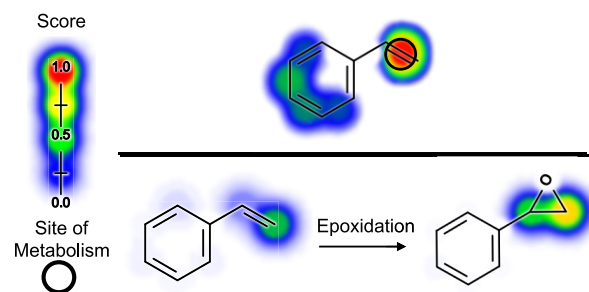
Figure 2. Bioactivation descriptors were computed using site-level metabolism predictions, metabolite structure predictions, and reactivity predictions. Top, the predictions generated by a previously developed epoxidation model[49] are visualized on styrene. The colored shading indicates site-level epoxidation scores, which reflect the probability that an epoxide will form at each possible location within styrene. Bottom, using a previously developed metabolite structure generator,[51] the exact epoxide structure corresponding to the highest site-level epoxidation score was generated. Next, we applied our previously published reactivity model,[52] which has proved useful in other studies as well,[53,54] to predict the atom-level reactivity of both the substrate and the metabolite. Finally, by tracking each atom's reactivity score in the metabolite from the corresponding atom in the substrate, we calculated atom-level reactivity deltas. Metabolism renders the two carbon atoms more reactive, with predictions increasing from 0.19 to 0.41 and from 0.56 to 0.73.

Using previously designed models for quinone formation,[48] epoxidation,[49] nitroaromatic reduction, and thiophene sulfur-oxidation,[50] formation scores are computed for each possible transformation for a given input molecule. The formation score for each reaction type represents the likelihood of the parent molecule undergoing the respective metabolic transformation. Next, the actual structures of each of these possible metabolites are generated using an in-house metabolite structure predictor.[51] After enumeration of the metabolite
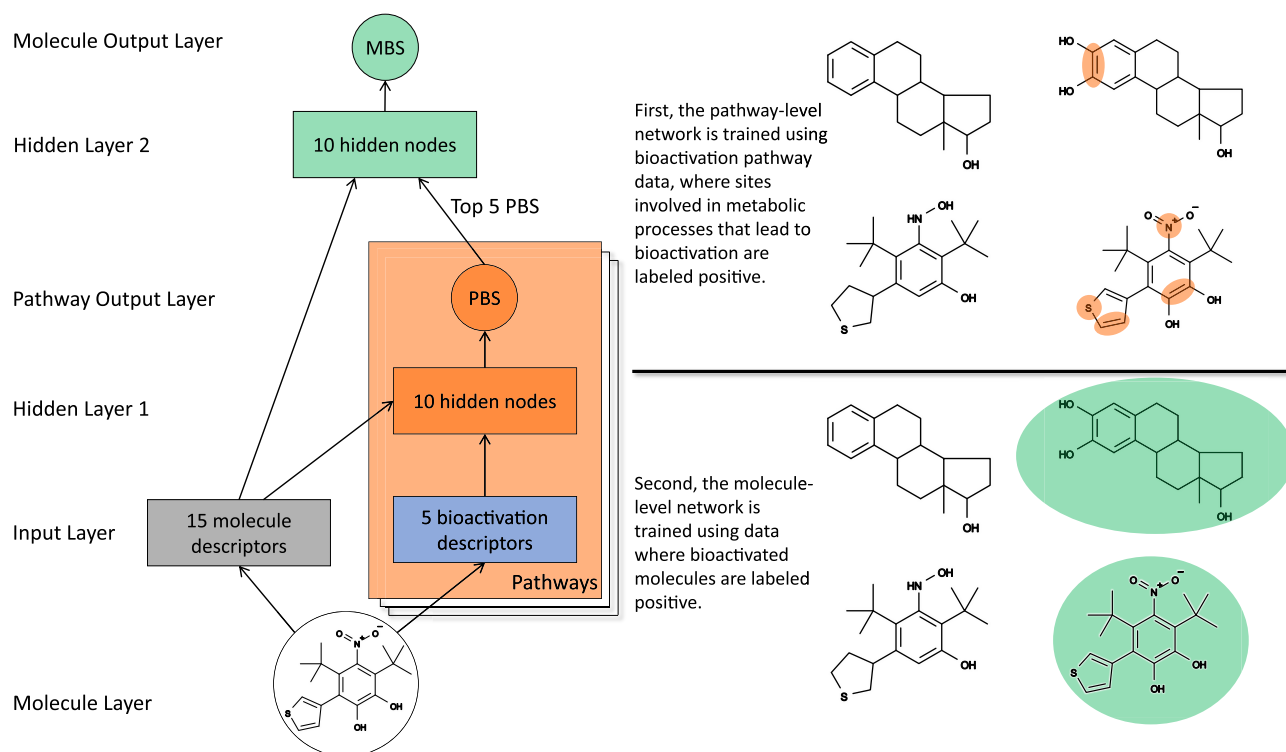
Figure 3. Structure of the bioactivation model. The diagram on the left demonstrates the flow of data through the model. The model consists of 1 molecule layer, 1 input layer, 2 hidden layers, and 2 output layers. First, several descriptors are calculated from an input molecule's structure. Each pathway has its own set of bioactivation descriptors, and these descriptors are submitted to the hidden layer, which computes pathway bioactivation scores (PBSs or pathway score) for the input molecule and each of its pathways. Each pathway score ranges from zero to one, reflecting the probability of a specific bioactivation mechanism at a specific site within the input molecule. For this study, we enumerated four types of metabolic transformations that precede bioactivation, including all possible quinone formations, epoxidations, nitroaromatic reductions, and thiophene sulfur-oxidations. Next, molecule descriptors and the top-five pathway scores are submitted to a second hidden layer, which computes a molecule bioactivation score (MBS or molecule score) for the input molecule. Molecule scores also represent the probability of a molecule undergoing bioactivation by any of the pathways considered and, like pathway scores, ranges from zero to one. A chemical structure is represented by the molecule node. The other circles are probabilistic scores between 0 and 1. Blocks are vectors of real numbers. The stack of plates marked "Pathways" represents that the pathway-level module is replicated across all possible metabolites and their corresponding pathways. On the right, site-level data are illustrated on the top (with sites of predicted bioactivation circled), and molecule-level data are illustrated on the bottom (with the molecules of predicted bioactivation circled). Model output for two pairs of highly similar molecules are illustrated.

structures, the atom-level reactivity scores with respect to both glutathione (GSH) and protein are computed for both the substrate and the product molecule.[52] Next, the atom-level reactivity deltas between the metabolite and substrate are calculated for both GSH and protein by subtracting the atom's reactivity prediction in the product from the atom's reactivity prediction in the substrate. Finally, a bioactivation score is computed by multiplying the maximum atom-level reactivity delta by the corresponding formation score.

In some cases, molecules being evaluated were present in the training sets for the models used to generate the formation scores or atom-level reactivity scores. To ensure unbiased evaluation of each molecule, we use versions of the phase I metabolism models and reactivity models that were retrained with all of the molecules in their original training sets except for the molecule being evaluated. The predictions from the newly trained models are then used to derive the bioactivation descriptors. We also compute simple molecule descriptors, such as molecular weight and the total number of atoms. Overall, we generate 20 descriptors for each possible metabolite, including 5 bioactivation descriptors (Table S1) and 15 molecule descriptors (Table S2).

Combined Path- and Molecule-Level Bioactivation Model. Descriptors are computed to create inputs for machine learning algorithms, which find mappings between vectors of numbers, known as features, and labeled examples, known as targets. In this context, the descriptors that compose our features and our targets are a binary column indicating whether a pathway was experimentally observed as a bioactivation pathway in the AMD.

The bioactivation model's architecture is representative of a feedforward neural network, in which an input is passed into a layer of fully connected neurons. Each neuron has a weight parameter associated with each of its connections to the input or neurons in the previous hidden layer. An affine combination is computed whereby the weight parameter is multiplied by the input feature to a neuron and adjusted by a learnable bias parameter. The affine combination is passed into a nonlinear function, such as a sigmoid function or a rectified linear unit (ReLU), and the neuron output is the input to each neuron in the next hidden layer. The final layer aggregates the outputs of the previous layer and results in a final model prediction. In comparison, a logistic regressor computes an affine combination between the input features and the model parameters, which is then fed into a sigmoid function that squashes it in the range of 0 to 1. For both methods, a loss function is used to compute the error between the model output and the true label. The computed error is used to update the model's parameters to minimize the loss function.

The bioactivation model is a feedforward neural network, with 1 molecule layer, 1 input layer, 2 hidden layers, and 2 output layers (Figure 3). The first output layer calculates pathway bioactivation scores (PBSs or pathway score), and the second output layer computes a single molecule bioactivation score (MBS or molecule score) for each input molecule.

We trained this network in two stages. In the first stage, we trained the pathway-level network to compute accurate PBS values. For a given input molecule, each of its experimental or model-inferred pathways is
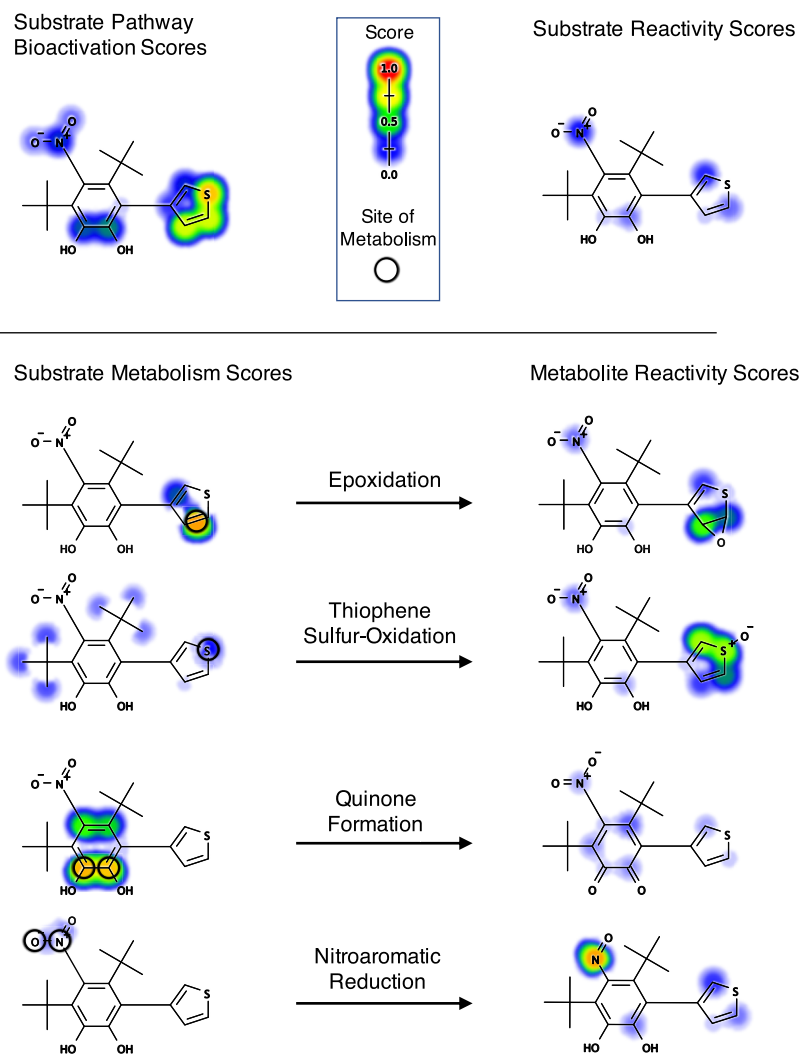
Figure 4. Visualization of pathway bioactivation is elusive because it requires consolidating multiple PBSs that can overlap at the same site. Epoxidations and nitroaromatic reductions entail bond cleavage. However, quinones take place on atom pairs, and thiophene sulfur-oxidation occurs on a single sulfur atom. Consequently, while each of these predictions types can be visualized on separate structures (Figure 1), this paradigm does not make for easily understood predictions. Instead, we first map all PBSs to the atom level. This is straightforward: for a quinone pair prediction, each atom in the pair would be assigned the initial score. Similarly, both atoms of a bond prediction (nitroaromatic reduction and epoxidation) would be assigned the same initial score. While this mapping moves closer to an interpretable result, the same atom can still be subject to multiple possible bioactivation pathways. For instance, the atoms making up sites of aromatic epoxidation are often themselves possible sites of quinone formation. Therefore, in the last step prior to visualization, we compute the scores on each atom by using the probabilistic OR function across all predictions that included the current atom under consideration. Top left, an example of the resulting, final visualization of pathway bioactivation. Top right and bottom, the underlying metabolism and reactivity processes that constitute some of the possible bioactivation pathways extending from the visualized substrate.

considered as a possible bioactivation pathway. Each pathway is represented by a vector containing the bioactivation descriptors that define it. Thus, the data set passed into the pathway-level network is a matrix with one column per bioactivation descriptor and one row per pathway. Using the first hidden layer, the pathway-level network learns an association between each pathway and its bioactivation descriptors against a binary target vector where the experimentally observed bioactivation pathways are labeled with a 1.

To train the model, we use gradient descent on the cross-entropy error.[55] In this procedure, the model's weights are gradually adjusted to assign high pathway scores for experimentally observed bioactivation pathways and low pathway scores for all other pathways. Each pathway score ranges from 0 to 1 and reflects the probability of a specific bioactivation event at specific sites within a molecule, where each pathway is represented by its five bioactivation descriptors. Effectively, the pathway score predicts which pathway(s) will lead to bioactivation, where a pathway consists of the input molecule, one of the four

metabolic transformations, and the structure and conjugation potential of the resultant metabolite.

The described procedure can generate several different PBSs, which may share some atoms in common if the different pathway's involve multiple metabolic transformations occurring at the same sites. For example, the atoms that make up sites of aromatic epoxidation are often themselves possible sites of quinone formation. This phenomenon can make visualization of multiple PBSs, each of which designates a prediction regarding a different bioactivation pathway, a difficult task. To aid interpretability, we devised a method for visualizing PBSs for a given input molecule (Figure 4). We first map all PBSs to the atom-level. This is straightforward: for a quinone pair prediction, each atom in the pair would be assigned the PBS. Similarly, both atoms of a bond prediction (nitroaromatic reduction and epoxidation) would be assigned the same PBS. This results in a vector of PBSs for each atom in the molecule, where each vector entry represents the PBS for each of the atom's possible bioactivation pathways. Next, we compute
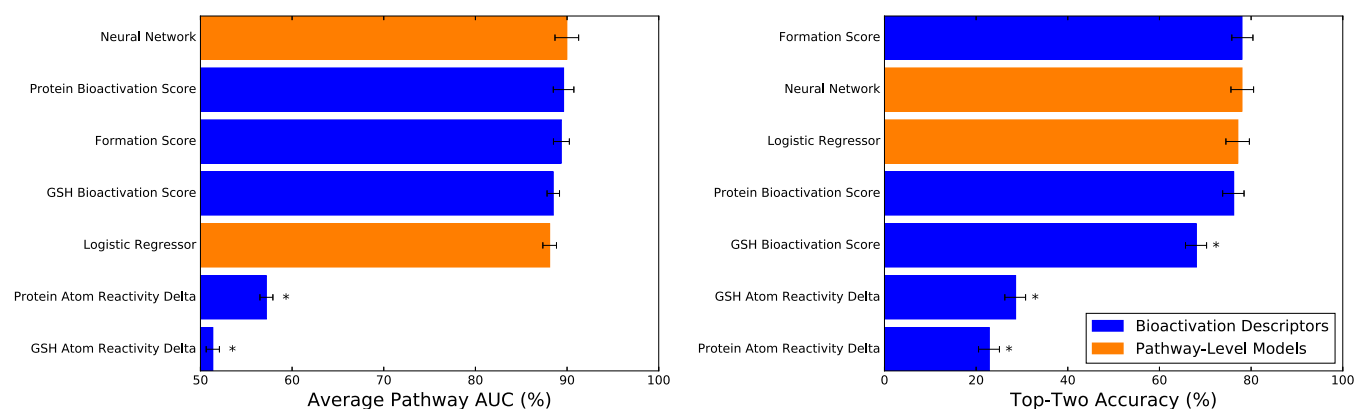
Figure 5. Model produced accurate cross-validated bioactivation pathway predictions. Right, the top-two metric was computed across 332 bioactivated molecules. The top-two metric reflects the percentage of correctly predicted molecules, where a correct prediction is defined as a molecule for which any of its bioactivated pathways received the highest or second-highest score for all possible pathways within that molecule. Left, across the same set of bioactivated molecules, the average path AUC was measured by calculating how often bioactivated pathways received higher scores than all other possible pathways. For both metrics, the performances of the five bioactivation descriptors were reported, as well as the cross-validated scores produced by training with either a neural network or a logistic regressor. Asterisks denote performances that were significantly worse than the highest-scoring method, as determined by a paired *t*-test using Bonferonni's correction.[61] In this context, the paired instances are the bioactivated parent molecules and the computed value of the average path AUC or top-two metric when the bioactivated molecule is processed by either the highest-scoring model or one of the other individual models, whose performance is being compared against the highest-scoring model. The error bars represent 95% two-sided confidence intervals.[62]

the final scores on each atom by using the probabilistic OR function across the vector of predictions assigned to that atom.

In the second training stage, we trained the molecule-level output layer to compute MBS values. Using the second hidden layer, the molecule-level network learns an association between the 15 molecule descriptors, the top-five pathway scores, and the input molecule's classification as being experimentally observed to bioactivate or not. Each row of the input data matrix represents a molecule, and each column represents either a molecule descriptor or one of the top-five pathway scores. Similar to the first stage, the weights of the network are trained using gradient descent on the cross-entropy error to assign bioactivating molecules with higher molecule scores than those for nonbioactivating molecules. Each molecule score reflects the probability of the input molecule undergoing bioactivation by any of the pathways considered and, like pathway scores, ranges from zero to one.

To produce pathway and molecule scores for the entire data set, we use a standard practice in machine learning for simulating performance on external data: cross-validation. In this procedure, any metabolically related molecules are withheld together, and the model is trained on the remaining data. Molecules were separated into metabolically related clusters based on connections through metabolic reactions in the AMD—each cluster was comprised of a molecule and all of its detected parent and sibling molecules. Next, the trained model predicts the pathway scores of the withheld molecules. In total, there are 569 groups of related molecules, so the cross-validation procedure entails training 569 individual models. This process guarantees that each molecule's predictions are computed by a model that does not contain information about that molecule or closely related molecules.

■ RESULTS AND DISCUSSION

The bioactivation model's performance and capabilities are investigated in the following sections. First, we focus on the model's pathway bioactivation scores (PBSs or pathway score). Using several metrics, we quantify how well pathway scores predict the correct pathway(s) within bioactivated molecules. Next, we turn our attention to the model's second output layer, which produces molecule bioactivation scores (MBSs or molecule score). Using similar methodologies as the pathway-level analysis, we measure the performance of molecule scores by several standards. Finally, we use the final bioactivation model to

enumerate hypotheses for the toxicity drivers of drugs with currently unknown or poorly understood toxicity mechanisms.

Bioactivation Pathway Prediction Accuracy. For bioactivated molecules, knowledge of their specific bioactivation pathways yields potentially fundamental insights about their possible toxicity mechanisms. Bioactivation takes place at specific sites within molecules and forms specific reactive metabolite structures. Knowledge of these sites and subsequent reactive structures can potentially guide rational modifications to prevent bioactivation while hopefully retaining a drug's pharmacological effect.

In the following experiments, we compare the neural network's performance on various metrics to that of a simpler model—a logistic regressor—trained with identical inputs and cross-validation folds. We also calculate the performance of each descriptor by treating it as a model and simply using its raw values as predictions. These comparisons indicate which descriptors are informative, and whether anything is gained by agglomerating several properties using the neural network or logistic regressor. We use several metrics to assess each model's performance at predicting bioactivation pathways.

First, we compute the "top-two" performance, a commonly used metric in site of metabolism prediction studies (Figure 5).[49,56−59] This metric counts a molecule as accurately predicted if any of its experimentally observed bioactivated pathways receive the highest or second-highest PBS for the entire molecule. The total number of correct predictions is divided by the total number of bioactivated molecules and multiplied by 100 to produce the percentage of correct predictions.

Second, we compute the "average path AUC" for calculating the area under the receiver operating characteristic curve (ROC AUC) for each bioactivated molecule, followed by averaging these AUCs. Recall that each bioactivated molecule has multiple metabolites and pathways that were enumerated in the construction of the bioactivation training data, and at least one of the pathways is experimentally observed to bioactivate. Note that a given bioactivated molecule can have multiple pathways specific to one of the four metabolic transformations
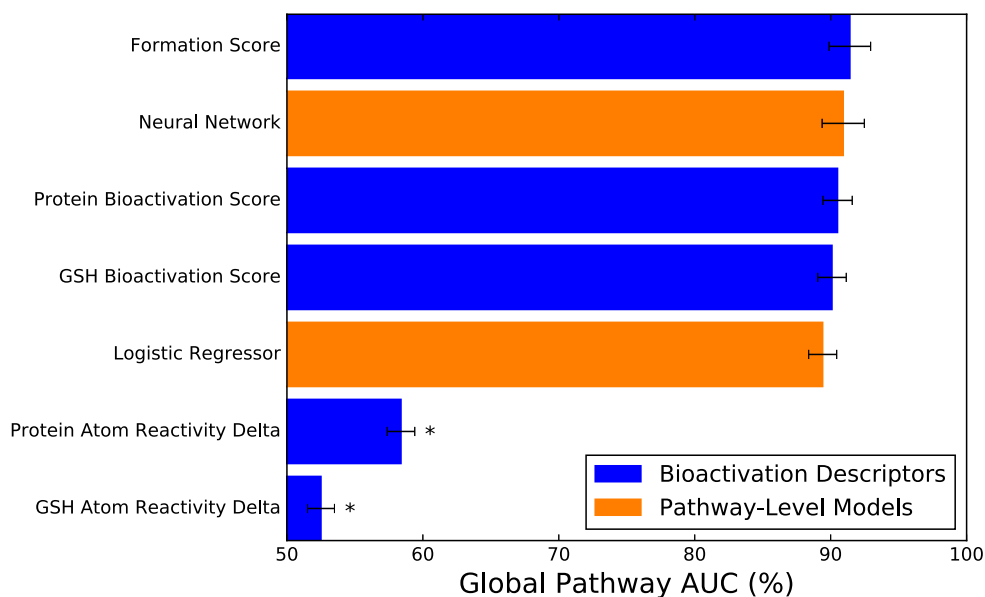
Figure 6. Binary-class global pathway AUC was computed for several methods across all bioactivated molecules. Asterisks denote performances that were statistically significantly worse than the highest-scoring method, using a false positive rate paired *t*-test.[61] In this context, the paired instances are the experimental or model-inferred pathways and their computed PBS when the pathway is processed by either the highest-scoring model or one of the other individual models, whose global pathway AUC performance is being compared against the highest-scoring model. The error bars represent 95% two-sided confidence intervals computed using the method specified in Cortes and Mohri, which requires the error rate and the numbers of positive and negative samples.[62] With respect to each model's ROC curve, the threshold for optimal binarization is calculated and applied to classify each sample, from which the error rate can be computed.

or zero pathways specific to another type of metabolic transformation. Due to the discrepancy in pathway type representation within each bioactivated molecule, an appropriate multiclass metric is elusive because computing four ROC curves for each bioactivated molecule (one for each pathway type) results in undefined behavior where there is not a sufficient number of either bioactivated or nonbioactivated classes for the computation to make sense. Furthermore, we are more concerned with predicting whether a molecule will be bioactivated and, if so, which pathway leads to bioactivation—no matter the exact mechanism. For this reason, the pathway-level target vector is binary and not multiclass. Thus, we compute a single, binary-class ROC curve for each bioactivated molecule using binary labels for each of the molecule's pathways (whether the pathway has been experimentally observed to bioactivate or not) and the pathway-level predictions generated by the model. Finally, we compute the AUC for each of these ROC curves and aggregate them via an ordinary arithmetic mean to yield the average path AUC performance. The average path AUC is more sensitive than the top-two metric, because it considers the relative ranking of all pathways within each molecule, and we have often used variations of it in past work.[48,49,52,60]

However, five methods have average path AUC performances that are statistically equivalent by paired *t*-tests corrected for multiple comparison via Bonferroni's correction.[61] The neural network, the protein bioactivation score, the formation score, the GSH bioactivation score, and a logistic regressor had equivalent average path AUCs of 89.98%, 89.62%, 89.38%, 88.50%, and 88.10%, respectively. The top-two performances are slightly more informative, but nevertheless four of the five previously tied methods are also equivalent, with the formation score, the neural network, the logistic regressor, and the protein bioactivation score, having statistically comparable top-two performances of 78.01%, 78.01%, 77.11%, and 76.20%,

respectively. Only the GSH bioactivation score can be judged inferior, with its 68.07% top-two performance having a *p*-value of 0.002 when compared to the predictions of the top-scoring method using paired *t*-tests corrected for multiple comparison via Bonferroni's correction.[61]

To further evaluate the four methods with equivalent average path AUCs and top-two performances, we compute a third metric, the global pathway AUC (Figure 6). Unlike the previous metrics, this measure does not consider molecule identity and instead computes the AUC of a single, binary-class ROC curve across all possible bioactivation pathways within bioactivated molecules. However, the same five methods that had equivalent average path AUCs also had equivalent global pathway AUCs, with the formation score, the neural network, the protein bioactivation score, the GSH bioactivation score, and the logistic regressor performing at 91.41%, 90.92%, 90.51%, 90.10%, and 89.41%, respectively.

Unlike the average pathway AUC, we can compute a multiclass variant of the global pathway AUC and evaluate the pathway-level models for each metabolic transformation being studied. For this measure, we compute the AUC of four individual ROC curves—one for each metabolic transformation. The target vector is still binary—1 if bioactivated and 0 if not—and we use the term multiclass in this setting to refer to a global pathway AUC analysis on separate subsets of the pathway data that have been stratified across the four pathway types. Each ROC curve is created based on its metabolic transformation's possible bioactivation pathways within the set of bioactivated molecules, which are annotated based on whether the pathway has been experimentally observed to bioactivate or not, and the model-predicted pathway scores. Each molecule does not contribute just its maximum scored pathway to the global AUC calculation for a given metabolic transformation but instead contributes all the scored pathways and their possible metabolites. Each AUC, which is measured for one of the four

Table 1. Each Pathway-Level Model's Separate AUC Performance for Each of the Four Metabolic Transformations and Their Aggregate Multiclass AUC Performance

| model	epoxidation	nitroaromatic reduction	quinone formation	thiophene S-oxidation	multiclass global pathway AUC
formation score	92.8%	65.2%	90.1%	71.4%	91.2%
neural network	92.1%	65.9%	89.9%	64.3%	90.8%
protein bioactivation score	91.6%	73.2%	89.5%	42.3%	90.3%
GSH bioactivation score	91.5%	88.9%	89.3%	85.7%	90.3%
logistic regressor	90.5%	64.4%	88.4%	59.0%	89.2%
protein atom reactivity delta	70.4%	70.4%	56.2%	0.07%	62.5%
GSH atom reactivity delta	63.4%	89.2%	46.6%	42.6%	54.2%
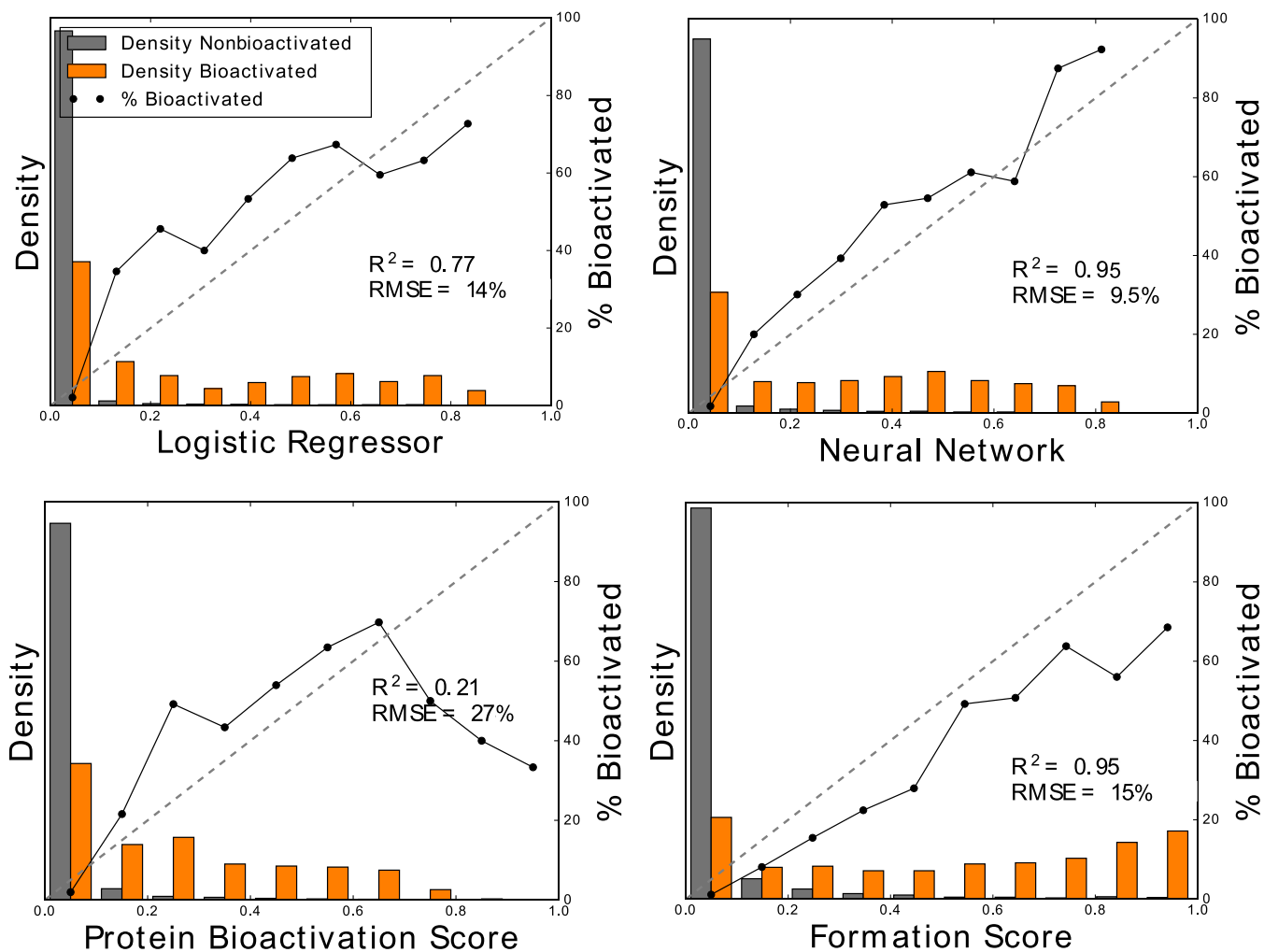
Figure 7. Of all methods, the neural network computed the most well-scaled predictions of bioactivation pathways, with predictions that were highly probabilistic. The bar graphs plot the normalized distributions for each method across 388 pathways within 332 bioactivated molecules. The solid lines plot in each bin the percentage of bioactivated pathways among all pathways in the corresponding score, denoted on the X-axis. The diagonal dashed lines indicate the ideal perfectly scale prediction. The neural network produced the best scaled prediction of all methods, indicated by the highest correlation to the best fit-line, and the lowest RMSE compared to a perfectly scaled prediction.

pathway types, is weighted by the prevalence of its corresponding pathway type in the pathway-level training set. For example, the weight of the AUC for the ROC curve specific to quinone forming pathways is 7,580 divided by 13,743. The multiclass AUC is then given by the sum of the four weighted AUCs.[63]

The pathway-level multiclass global pathway AUC for each model is reported in Table 1, along with each pathway type's individual global pathway AUC. The same five methods that had equivalent binary-class global pathway AUCs also had

equivalent multiclass global pathway AUCs, with the formation score, the neural network, the protein bioactivation score, the GSH bioactivation score, and the logistic regressor performing at 91.2%, 90.8%, 90.3%, 90.3%, and 89.2%, respectively. For each model, pathway performance was highest for both the epoxidation and quinone formation pathway types. In contrast, the pathway types for nitroaromatic reduction and thiophene S-oxidation had a lower, more varied performance. Both the nitroaromatic reduction and thiophene S-oxidation pathway types are much less frequently observed than the epoxidation
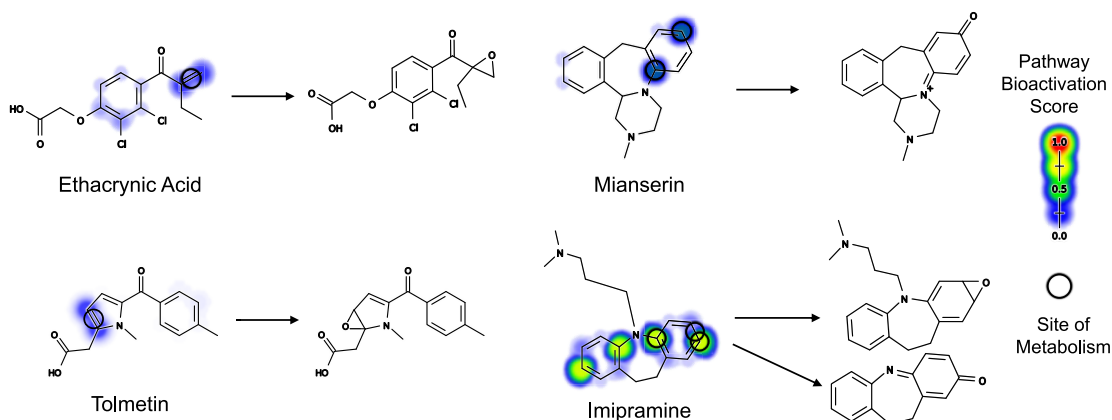
Figure 8. From top to bottom, left to right: ethacrynic acid,[64] a diuretic, mianserin,[65] an antidepressant, tolmetin,[66] a nonsteroidal anti-inflammatory, and imipramine,[33,67] an antidepressant. For each drug, the experimentally observed site of metabolism is circled in black, and the observed reactive metabolite is displayed. The magnitude of the PBS is indicated by the color shading gradient. Imipramine is known to produce two different reactive metabolites: an epoxide and a quinone, both of which are predicted and visualized. Notice that imipramine's plane of symmetry means that the highly predicted sites on the other side of the molecule merely indicate the same quinone pathway.

and quinone formation pathway types in the bioactivation training set, and the disparity in pathway type-specific results may be driven by the disparity in available data.

Ultimately, we place more emphasis on average pathway AUC because it more closely approximates the expected use of the pathway-level model. When inputting a test molecule to see which potential pathways and metabolites may result in bioactivation, a user is likely interested in whether the model correctly identifies the correct pathway(s) of bioactivation, relative to all other pathways inferred by the model. Average pathway AUC reflects this consideration.

It is intriguing that three of the bioactivation descriptors perform so well compared to training the logistic regressor or neural network. These performances are calculated simply by treating a descriptor as a model prediction, without any training. In previous studies, we have frequently compared our results to individual descriptors, and we have always found that machine learning outperforms the naive descriptor approach.[48,49,52,60] This is as we would expect, due to the greater flexibility of the machine learning approaches, which by design simultaneously consider many different chemical attributes.

One explanation for the above phenomenon is that, in the previous studies, the individual descriptors that were out-performed were either topological or quantum descriptors describing a basic feature of the molecule or its atoms. In this study, the three individual descriptors that perform nearly as well as the model are derived from the outputs of the previously trained site of metabolism or reactivity models. These bioactivation descriptors represent information learned by the previously trained models and may encode more information regarding bioactivation, and metabolic processes in general, compared to the more naive descriptors referenced in prior work. A broader comparison of global pathway AUC is made between several individual topological descriptors against the neural network, logistic regressor, and the top-three bioactivation descriptors (Figure S1). In this comparison, the individual topological descriptors all perform significantly worse, as expected and in keeping consistent with prior work.

Additionally, it is noteworthy that the two bioactivation descriptors with much less predictive value are the two reactivity deltas. This suggests that, for this study's data set, merely calculating the max reactivity delta of a potential product is not

very informative for predicting bioactivation, because that metabolite may be very unlikely to occur. All three descriptors that match the model performances include information about the likelihood of the metabolic transformation: the formation score (which is just a prediction of metabolism and does not consider reactivity) and both "bioactivation scores", computed by multiplying the maximum atom reactivity increase by the formation score. Based on these observations, we infer that, for our data set, metabolism seems to be more important to evaluate than reactivity for predicting bioactivation. In future work, by expanding the model to include additional bioactivation pathways that produce metabolites with more diverse reactivity, we might observe an increased relevance of the reactivity computations.

As a final evaluation of the various best-performing methods, we compute reliability plots across all of the enumerated pathways within the bioactivated molecules (Figure 7). In this measure, the predictions of each method are distributed into ten bins between 0 and 1, each of width 0.1. Next, we compute the percentage of bioactivations in each bin and calculate the root-mean-square error (RMSE) between each of these percentages and the midpoint value of that bin. Lower RMSE values indicate prediction methods that correlate better to probabilities, with a perfectly scaled prediction having an RMSE of 0. To assess performance, we also compute the R^2 of the best fit line, which has a maximum possible result of 1 for a model that produces score bins that fit on a perfectly straight line.

Note that the RMSE and R^2 metrics are dependent on the binning strategy and can be sensitive to the concentration of data points in each bin. To resolve this issue, we also compare each method's calibration performance using the Brier score, which calculates the mean squared error between the model's predictions and the observed outcome. A lower Brier score designates improved model calibration and is preferred.

With an RMSE of only 9.5% and a Brier score of 0.0197, the neural network outperforms all other methods by both measures. In comparison, the Brier scores for the logistic regressor, the protein bioactivation score, and the formation score are 0.022, 0.0230, and 0.0258, respectively. The logistic regressor and the protein bioactivation score have much lower R^2 performances of 0.77 and 0.21, respectively. In contrast, the formation score's R^2 of 0.95 almost equals the neural network's
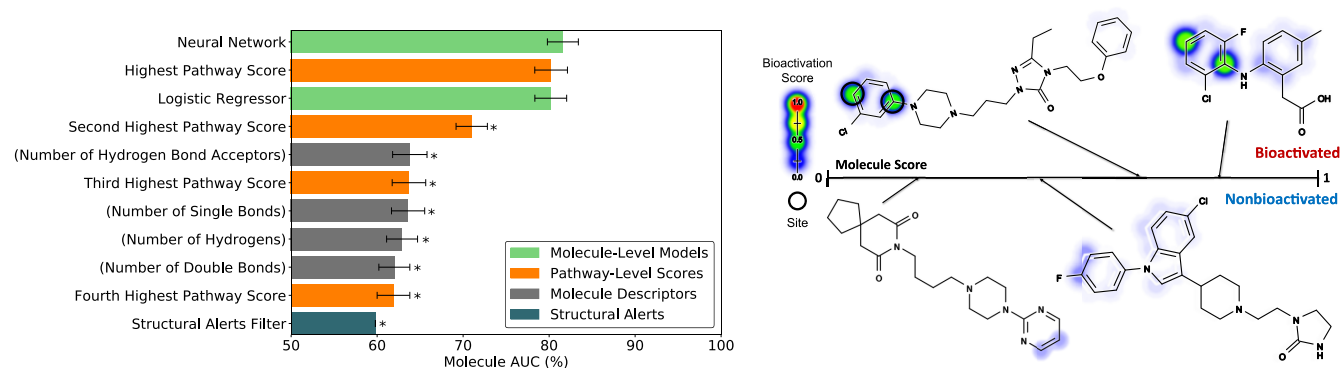
Figure 9. Molecule bioactivation scores (MBSs or molecule score) accurately identified bioactivated molecules. While all the previous AUC values are based on the predictions of the pathway-level model, the molecule-level AUC is based on the predictions of the molecule-level model and only considers the parent molecule and not the whole set of possible metabolites. Several methods are compared by their capacity to separate bioactivated and nonbioactivated molecules. These methods include training a neural network or logistic regressor on the top-five pathway-level scores and all molecule descriptors, using each molecule's top pathway-level scores as molecule predictions, and using molecule descriptors as molecule predictions. Based on false positive rate paired *t*-tests,[61] results are denoted with an asterisk if their performance is significantly lower than the best-performing method. In this context, the paired instances are the bioactivated or nonbioactivated molecules and their computed MBSs when processed by either the highest-scoring model or one of the other individual models, whose molecule AUC performance is being compared against the highest-scoring model. As a baseline comparison, an approach that uses structural alerts to filter molecules with at least one alert achieved a molecule AUC of 59.8%. The 35 structural alerts used for screening were derived from literature on minimizing idiosyncratic toxicity mediated by reactive metabolites.[68] The error bars represent 95% two-sided confidence intervals.[62] Right, examples of bioactivated and nonbioactivated molecules are visualized. From left to right, top to bottom: nefazodone (MBS: 0.74), lumiracoxib (MBS: 0.78), buspirone (MBS: 0.18), sertindole (MBS: 0.43). Each experimentally observed site of bioactivation is circled. For each molecule, the colored shading represents PBSs, which range from 0 to 1.
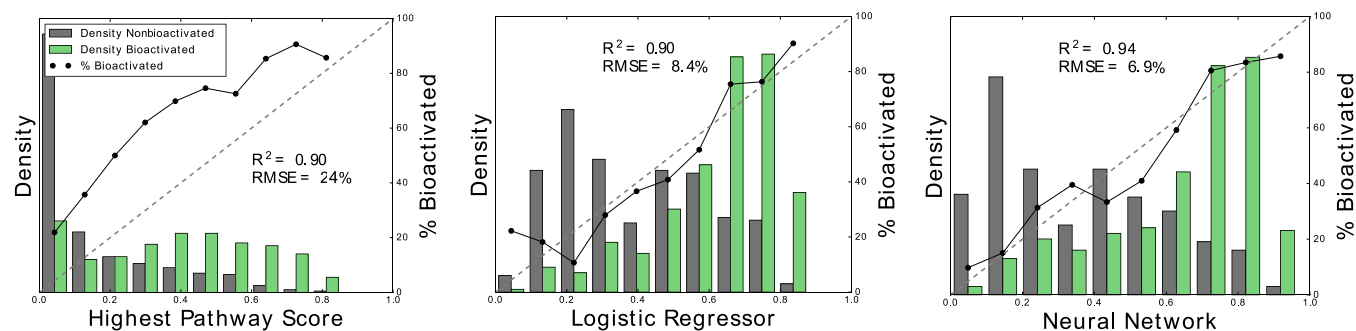
Figure 10. Of all methods, the neural network produces the best-scaled, highly probabilistic molecule predictions. The bar graphs plot the normalized distributions for each method across 388 pathways within 332 bioactivated molecules. The solid lines plot the percentage of bioactivated pathways among all pathways that were assigned the corresponding score bin, denoted on the *X*-axis. The diagonal dashed lines indicate the ideal perfectly scaled prediction. Of all three methods, the neural network has both the highest correlation to a best fit-line through respective binned percentages, indicated by the R^2 values, and the lowest RMSE compared to a perfectly scaled prediction.

R^2 of 0.95. Nevertheless, the formation score's RMSE of 15% indicates that it does not produce a probability as well-scaled as the neural network, with its much lower 9.5% RMSE.

The neural network is the only method that produces both a high R^2 and a low RMSE and Brier score, so we chose to use it for the rest of the study. However, we certainly acknowledge that valid arguments could also be made for choosing one of the simpler methods instead. In addition, models with good correlation but bad RMSE can still be useful in scenarios where binarizing the prediction is applicable and where there is flexibility in shifting the decision threshold for optimal binarization. Ultimately, we chose the neural network's scores because we felt that their probabilistic nature might be helpful for the next objective: predicting molecule bioactivation. To give an explicit example of the output given by the pathway-level neural network model, four drugs are visualized with their cross-validated PBS produced by the neural network (Figure 8).

Accuracy at Predicting Molecule Bioactivation. Given a list of drugs or drug candidates, a useful bioactivation model should accurately predict which structures will be bioactivated. Flagging these potentially toxic molecules and separating them from lower-risk compounds could enable problematic compounds to be triaged for more rigorous testing, set aside if there are acceptable alternatives, or rationally modified to prevent bioactivation while retaining efficacy. To enable the bioactivation model to make these molecule-level bioactivation predictions, we include in the training data molecules that are not bioactivated, despite having structures capable of forming quinones, epoxides, thiophene sulfur-oxides, or aromatic nitrosos.

After the first stage of training, which produces accurate pathway-level predictions, we perform a second training stage to tune the model to distinguish between bioactivated and nonbioactivated molecules. To assess performance at this objective, we compute the "molecule AUC": the AUC over all molecules in the training data set. While all the previous AUC values are based on the predictions of the pathway-level model, the molecule-level AUC is based on the predictions of the
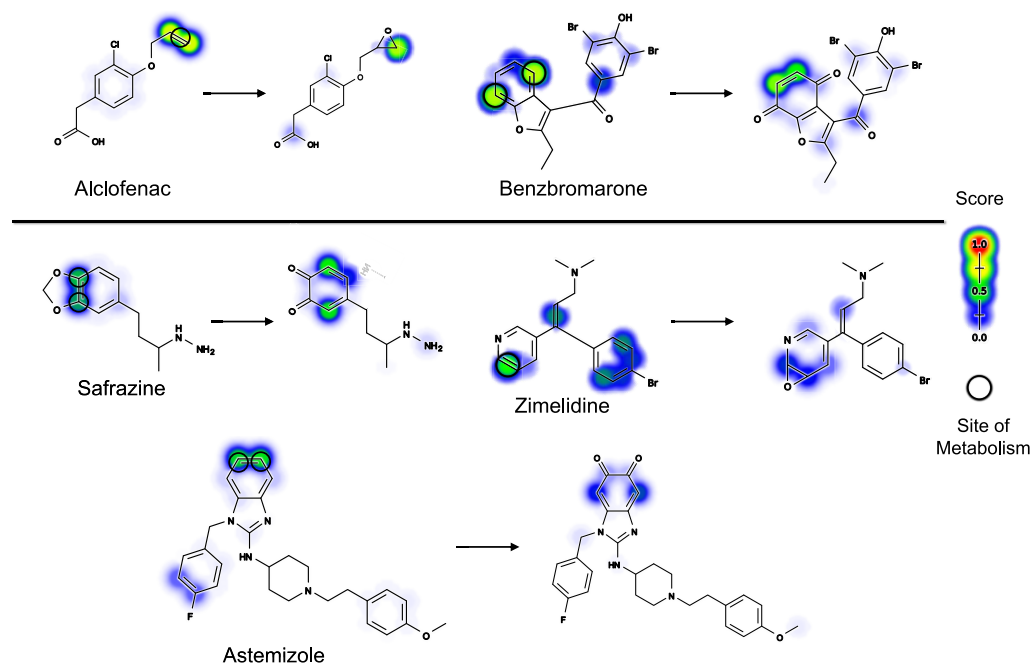
Figure 11. Top, examples of withdrawn drugs with high bioactivation predictions. We submitted the structures of 201 withdrawn drugs to the bioactivation model, the majority of which were downloaded directly from DrugBank and supplemented by literature review. In order to generate toxicity mechanism hypotheses for these drugs, we looked at molecules that received high molecule bioactivation scores yet have not been reported to form reactive metabolites (to the best of our knowledge). From left to right: alclofenac[71,72] (MBS: 0.75) and benzbromarone (MBS: 0.66). Bottom, hypothesized toxicity drivers of withdrawn drugs. The bioactivation model predicts novel bioactivation mechanisms for several drugs. Three examples are visualized with their metabolism predictions indicated by the colored shading on each parent drug structure, and the reactivity predictions are similarly visualized on the structure of the hypothesized metabolite. From left to right, top to bottom: safrazine (MBS: 0.65), zimelidine (MBS: 0.66), and astemizole (MBS: 0.69). For each pairing, the color shading gradient indicates PBS on the substrate and protein reactivity on the metabolite.

molecule-level model and only considers the parent molecule and not the whole set of possible metabolites. Specifically, we compute the AUC of a single ROC curve that is generated using the labels for each molecule in the training set (whether that molecule has been experimentally observed to bioactivate or not) and the molecule-level predictions generated by the model. We evaluate several models using this metric.

First, simply assigning each molecule its maximum cross-validated pathway score is the most intuitive approach. This highest pathway score method produces a molecule AUC of 80.23% (Figure 9). Alternatively, we perform a second training step to tune the weights of an additional model layer to distinguish bioactivated and nonbioactivated molecules. As descriptors, we use each molecule's top-five PBSs as well as all 15 molecule topological descriptors, for a total of 20 descriptors. With these descriptors as input, we compare a logistic regressor and a neural network with ten hidden nodes. The same cross-validation folds are used as for the pathway-level training: each group of similar molecules is withheld in turn, and a model is trained on the remaining data, ensuring that training and testing data is never mixed. The cross-validated scores produced by the logistic regressor and the neural network have molecule AUCs of 80.19% and 81.06%, respectively.

While molecule AUC fails to distinguish the highest pathway score, the logistic regressor, and the neural network, reliability plots indicate the best method: the neural network has both the best RMSE (6.9%) and the best R^2 (0.94) (Figure 10). In contrast, the reliability performances of the highest pathway score and the logistic regressor are, respectively, RMSEs of 8.4% and 24% and R^2 values of 0.90 and 0.90. Furthermore, the neural network achieves the lowest Brier score of 0.168, compared to

Brier scores of 0.184 and 0.179 for the highest pathway score and logistic regressor, respectively. The reliability plots also show that the neural network does the best job of assigning high scores to bioactivated molecules and low scores to nonbioactivated molecules. Consequently, similar to our decision on the path-level, we select the neural network for the final model structure, while acknowledging that this is a somewhat subjective choice. The final model is trained on the full bioactivation training data set.

As external tests, we use ten of the previously discussed batches of nonbioactivated molecules. Each batch includes 332 unique molecules not present in the training data. Using the final trained model, we predict the MBS of each molecule within each external test set. Next, for each test set, we measure the separation between the MBS of the test molecules and the cross-validated MBS of the bioactivated molecules.

Separation between the bioactivated molecules from the training set and the nonbioactivated molecules from the external sets could simply be due to a lack of calibration between the cross-validated model predictions and the final trained model predictions. Without additional positives being introduced via the external test sets, further evidence is required to demonstrate that the final trained model predictions do not need to be adjusted to compensate for potential differences relative to the cross-validated model predictions. On the bioactivated molecule training subset, we compare the cross-validated molecule predictions against the fully trained model predictions (Figure S5). As expected, there is an optimistic bias to the fully trained model predictions, which can be seen when comparing the line of best fit to the identity line. The optimistic bias results from the fully trained model seeing each instance during training. Several

notable outliers are present, most of which undergo observed quinone formation. These outliers may have manifested due to having moderate pathway-level scores that lead to greater sensitivity in bias when alternating between the cross-validated model and the fully trained model. The squared Pearson correlation coefficient of 0.87 (p-value ≤0.001) denotes a high correlation between the cross-validated model predictions and the fully trained model's predictions, which is evidence that the fully trained model's predictions are comparable to the cross-validated model predictions.

Over the ten external test sets, the molecule AUC is 83.29% ± 0.96%, equivalent to the accuracy of the MBS on the full cross-validated training set of 81.06%. The model successfully generalized to new data, assigning nonbioactivated molecules it had never seen before with much lower scores than bioactivated molecules.

Inferring the Toxicity Mechanism of Withdrawn Drugs. To investigate the model's utility, we explored its applicability to withdrawn, small-molecule drugs. Ideally, the model will be able to generate hypotheses regarding the underlying toxicity of the withdrawn drugs. After downloading a list of withdrawn drugs from DrugBank and removing any molecules found in our training data, we used the bioactivation model to enumerate hypotheses for the toxicity drivers of the withdrawn drugs. Each of these 201 molecules were submitted to our final trained model, thereby producing a molecule score for each molecule. Molecule-level bioactivation scores for each withdrawn drug are made available in Table S3.

The model outputs probabilistic molecule scores, but for the purposes of estimating the total number of bioactivated molecules within the list of withdrawn drugs, it was useful to define an exact score to binarize predictions. To define this cutoff, we analyzed the full ROC curve used to calculate the molecule AUC across all training molecules (Figure S2). While several methods for defining this cutoff are possible, in this study, we computed the cutoff that optimizes both sensitivity and specificity.[69] This analysis resulted in an MBS threshold of 0.57. Using this threshold, we predicted that among the data set of withdrawn drugs, 33 are bioactivated.

Upon investigation, we find that some of these predictions are consistent with previously reported experimental results that were not present in our training data (Figure 11, top). For example, alclofenac, a nonsteroidal anti-inflammatory drug, was introduced to the U.K. in 1972 but withdrawn in 1979[70] due to several incidences of hepatotoxicity and skin rashes.[71] Later, it was determined that an epoxidation reaction takes place at alclofenac's terminal olefin.[72] Furthermore, this epoxide metabolite directly conjugates to sulfhydryl nucleophiles[73,74] and consequently is a likely driver of alclofenac's adverse effects. Indeed, the bioactivation model predicts this epoxidation and assigns alclofenac a very high probability of bioactivation with an MBS of 0.75.

The bioactivation model also agrees with the experimental findings for the gout drug benzbromarone, withdrawn from Europe in 2004[75] due to cases of severe hepatotoxicity.[76−78] At the time, the mechanism of this toxicity was unclear. Later, experiments with human liver microsomes produced glutathione-conjugated metabolites, believed to be adducts of a reactive *ortho*-quinone.[79] This reactive metabolite could be the driver of the benzbromarone's idiosyncratic reactions. Surprisingly, the bioactivation model similarly predicts quinone formation at this ring, although it scores formation of a *para*-quinone as more likely than formation of the *ortho*-quinone. It is

not clear whether the human liver microsomes had the resolution to distinguish *ortho*- and *para*-quinones, so it seems that the bioactivation model's hypothesis essentially matches the experimentally observed reactive metabolite.

In the cases of alclofenac and benzbromarone, the bioactivation model reiterates toxicity mechanisms that have already been reported. These results are examples of the model successfully generalizing beyond its training data since these structures were not included at any training phase. Additionally, the agreement between experiment results and the computational modeling bolsters confidence that predictions of novel toxicity mechanisms for less well-understood drugs be experimentally validated in the future.

In other cases, the model's predictions represent new hypotheses for the mechanism's of a drug's toxicity (Figure 11, bottom). For example, the antidepressant safrazine, a monoamine oxidase inhibitor, was withdrawn due to hepatotoxicity[80] and has also been associated with neuropathy.[81] The model assigned safrazine an MBS of 0.65, above the threshold defined by the optimal point on the ROC curve of cross-validated scores. To the best of our knowledge, a toxicity mechanism has not been proposed for this abandoned drug. From the structural alert's perspective, one motif of concern might be safrazine's terminal hydrazine, a well-known structural alert.[32] However, the model does not predict a bioactivation pathway extending from the terminal hydrazine, which may be because activation of this structural alert was not one of the bioactivation pathways modeled in this study. Instead, the model predicts a bioactivation pathway on a different structural alert of safrazine: a 1,3-benzdioxole motif leading to a reactive quinone. Without experiments, it is difficult to say which bioactivation pathway is more likely, but understanding the toxicity of this drug could inspire a rational modification to improve its safety profile.

The bioactivation model also provides a new hypothesis for zimelidine's toxicity mechanism. Zimelidine was an antidepressant introduced to the U.K. market in 1982, only to be withdrawn in 1983[70] because of several reports of idiosyncratic neuropathy, including Guillain Barré syndrome.[82,83] Later, it was proposed[84] that these adverse affects were due to a bioactivation mechanism beginning with oxidation of zimelidine's aliphatic nitrogen's, ultimately producing both a reactive nitrone[85] metabolite and a reactive acrylaldehyde metabolite, which have both been identified in human urine.[86] The bioactivation model presents an alternative hypotheses: direct epoxidation of zimelidine's heterocyclic ring. This hypothesized reactive metabolite—formed in a single step—could form in higher quantities than the other reactive metabolites, which take at least three steps to form.

Finally, astemizole, an antihistamine, was approved in 1988 by the FDA. However, due to cardiotoxicity,[87] Johnson and Johnson chose to withdraw the drug globally in 1999.[88] As far as we can tell, the source of this cardiotoxicity was never discovered. While astemizole is not known to produce reactive metabolites, our model hypothesizes formation of a reactive quinone metabolite following a double hydroxylation. Further reinforcing the model's predictions, the main biotransformation pathways of emedastine, a structurally similar molecule to astemizole, are mediated by aromatic hydroxylations at the same proposed positions on its benzimidazole ring.[89]

To contextualize the bioactivation predictions on the discussed withdrawn drugs, we evaluated their Tanimoto similarity against the training data set on ECFP6-derived

fingerprints. Neither alclofenac, benzbromarone, safrazine, zimelidine, nor astemizole had a Tanimoto similarity greater than 0.5 with any of the training set molecules. Figure S3 displays the Tanimoto similarity distribution between each of the discussed DrugBank withdrawn drugs and the training set instances.

As seen for safrazine, zimelidine, and astemizole, the model can unearth explicit, testable hypotheses about the contexts and mechanisms that drive a molecule's incidence of toxicity. Further experimental validation of these bioactivation predictions is necessary and a possible next step in future work but is outside of the scope of this study.

Lastly, it is worthwhile to point out cases in which the model yields potential false negatives with respect to the literature. Twelve of the withdrawn drugs had an MBS below the bioactivation threshold but were annotated as being observed to undergo bioactivation according to the AMD. Of these 12, we found corresponding literature for six that referenced potential bioactivation via the four pathways studied—zomepirac[90] (MBS: 0.42), nimesulide[91] (MBS: 0.40), thioridazine[92] (MBS: 0.36), troglitazone[93] (MBS: 0.22), phenacetin[94] (MBS: 0.15), and thalidomide[95] (MBS: 0.07). Figure S4 displays the Tanimoto similarity distribution between each of the six false negatives and the training set instances. For the remaining six drugs, their withdrawal is associated with other causes, such as drug−drug interactions or formation of alternate reactive metabolites than those within the scope of this study. Identifying and addressing the model's limitations may aid in resolving these false negatives or better understanding why they occur.

MODEL LIMITATIONS

One limitation of the current model is that there are other bioactivation pathways beyond those included in this study. However, the approach demonstrated here is easily extendable to additional pathways by providing training data for those pathways and models for specific reactions of interest. For instance, the phase I metabolism model used for prediction of nitroaromatic reduction and sulfur oxidation also supports epoxidation and provides coverage of 92.3% of AMD phase I reactions,[50] and the structure inference model supports the phase I metabolism model's rule set coverage as well.[51] Another shortcoming is that we only considered one-step bioactivations, where a metabolic reaction produced a metabolite that conjugated to macromolecules. Going forward, by expanding the metabolism scores to included multiple steps, we plan to extend the ability of the model to pick up multistep bioactivations.

Additionally, the current modeling approach has some inherent limitations due to its reliance on a database of literature-derived reactions. For example, we label molecules as not bioactivated based on a lack of evidence in the database. However, the aphorism "absence of evidence is not evidence of absence" is certainly worth remembering, and we view our labeling paradigm more as a necessary assumption than an ideal data set, due to the paucity of alternatives. This caveat is reflected in the lower performance of the molecule-level scores compared to the pathway-level scores. Another data limitation is with respect to understudied or less common pathways, as seen in the model's worse performance when evaluating potential bioactivation pathways involving nitroaromatic reduction or thiophene sulfur-oxidation compared to those involving quinone formation or epoxidation. Hopefully, future experimental work will yield higher-quality data sets that include both

bioactivated and nonbioactivated molecules, with both categories tested by the same assays. Furthermore, the current approach is reliant on several underlying models of metabolism and reactivity. Updates to any of the underlying models will influence the values of the calculated bioactivation descriptors and, as a result, require retraining the bioactivation model.

In the current work, the max atom-level reactivity delta features had less relevance than the formation scores in predicting PBSs and MBSs. This could be a result of the types of reactivity present in the data set, but it could also be a result of how the reactivity is used. Future work could expand upon more complex feature engineering of reactivity. For example, a product metabolite can result in a large max atom-level reactivity delta without having a high reactivity. Potential avenues of experimentation include altering the type of aggregation function used, e.g., an average function instead of a max function, and the incorporation of reactivity delta features with a feature for the absolute value of the product's reactivity.

Finally, not all reactive metabolites are toxic. Sometimes, detoxification pathways such as glutathionation are able to effectively clear reactive metabolites before they cause deleterious effects. Other complex phenomenon that influence the likelihood of an adverse event include route of administration, coadministered medications, comorbidities, and genetic variants. We ultimately envision a broader toxicity model, where bioactivation predictions are combined with other important factors, like daily dose and rate of reaction, in order to build a model that explicitly predicts the toxicity risk of a certain molecule.

CONCLUSION

Bioactivation entails two distinct concepts: metabolism and reactivity. Metabolism gives rise to electrophilic structures—reactive metabolites—that then conjugate to nucleophilic sites within biological macromolecules, such as proteins. Protein-metabolite adducts can incite toxic immune responses, the driving force for many idiosyncratic adverse drug reactions. This study constructs a novel, sophisticated bioactivation model by synthesizing several previous studies of both metabolism[48−51] and reactivity.[52,60] Four types of metabolism are included: quinone formation, epoxidation, thiophene sulfur-oxidation, and nitroaromatic reduction, some of the most common reactions that form reactive products. To predict bioactivation, the model first predicts the chances of any of these metabolic events occurring at all possible sites within a given input molecule. Next, the model enumerates the exact structures of those metabolites. These structures are assigned scores reflecting their chances of conjugating to protein or GSH. Finally, using these metabolism and reactivity scores as input, a feedforward neural network predicts bioactivation on both the pathway- and molecule-level.

Across 332 bioactivated molecules extracted from the literature, the model predicts the exact bioactivation pathway with 89.98% AUC. These pathway-level bioactivation predictions make a specific hypothesis about the mechanism of a molecule's toxicity. Furthermore, with 81.06% AUC the model separates bioactivated and nonbioactivated molecules. Molecule-level bioactivation predictions can be used to rapidly screen a large number of molecules for the key toxicity risk of bioactivation. To demonstrate, we use the bioactivation model to generate hypotheses for the toxicity drivers of several withdrawn drugs associated with idiosyncratic reactions with currently unknown etiology. By adding metabolism models for

additional bioactivation pathways, future work will easily expand the utility of the bioactivation model. Nevertheless, this study makes a major step forward toward a fully comprehensive model of metabolism and reactivity.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00417.

> "AMD_Registry_Numbers.csv" file, reaction registry numbers, molecule registry numbers, and a binary column that indicates bioactivation; "withdrawn_structures.sdf" file, withdrawn drugs assessed by model and information regarding their chemical structures, including but not limited to SMILES, INCHI identifier and key, Formula, DrugBank ID, Generic Name, and Synonyms; and "training_set_parents.sdf" file, parent molecule structures used in training data set (ZIP)
>
> Additional content on molecule-level bioactivation scores for all withdrawn drugs in data set, global pathway AUC comparison with topological descriptors, and molecule-level neural network AUC with extraction of optimal molecule-level score threshold and full listing of bioactivation and topological descriptors in Tables S1 and S2, respectively (PDF)

AUTHOR INFORMATION

Corresponding Author

S. Joshua Swamidass − *Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, United States;* ⓞ orcid.org/0000-0003-2191-0778; Email: swamidass@gmail.com

Authors

Tyler B. Hughes − *Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, United States;* ⓞ orcid.org/0000-0001-6221-9014

Noah Flynn − *Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, United States;* ⓞ orcid.org/0000-0002-8542-8887

Na Le Dang − *Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, United States;* ⓞ orcid.org/0000-0001-7458-1264

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemrestox.0c00417

Author Contributions

T.B.H. and N.F.: equal contribution as first author.

Notes

The content is completely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

ADR, adverse drug reaction; AMD, Accelrys Metabolite Database; AUC, area under the receiver operating characteristic curve; IADR, idiosyncratic adverse drug reaction; MBSs, molecule bioactivation scores; PBSs, pathway bioactivation scores; RMSE, root-mean-square error; ROC, receiver operating characteristic

REFERENCES

(1) Pirmohamed, M., James, S., Meakin, S., Green, C., Scott, A. K., Walley, T. J., Farrar, K., Park, B. K., and Breckenridge, A. M. (2004) Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ. [Br. Med. J.]* 329, 15−19.

(2) Chan, S. L., Ang, X., Sani, L. L., Ng, H. Y., Winther, M. D., Liu, J. J., Brunham, L. R., and Chan, A. (2016) Prevalence and characteristics of adverse drug reactions at admission to hospital: a prospective observational study. *Br. J. Clin. Pharmacol.* 82, 1636−1646.

(3) Oscanoa, T., Lizaraso, F., and Carvajal, A. (2017) Hospital admissions due to adverse drug reactions in the elderly. A meta-analysis. *Eur. J. Clin. Pharmacol.* 73, 759−770.

(4) Alexopoulou, A., Dourakis, S. P., Mantzoukis, D., Pitsariotis, T., Kandyli, A., Deutsch, M., and Archimandritis, A. J. (2008) Adverse drug reactions as a cause of hospital admissions: a 6-month experience in a single center in Greece. *Eur. J. Intern. Med.* 19, 505−510.

(5) Lazarou, J., Pomeranz, B. H., and Corey, P. N. (1998) Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA, J. Am. Med. Assoc.* 279, 1200−1205.

(6) U.S. Food and Drug Administration, *Preventable Adverse Drug Reactions: A Focus on Drug Interactions.* 2018. https://www.fda.gov/drugs/drug-interactions-labeling/preventable-adverse-drug-reactions-focus-drug-interactions (accessed 2020-07-05).

(7) Edwards, I. R., and Aronson, J. K. (2000) Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 356, 1255−1259.

(8) Lee, W. M. (2003) Drug-induced hepatotoxicity. *N. Engl. J. Med.* 349, 474−485.

(9) Kaplowitz, N. (2001) Drug-induced liver disorders. *Drug Saf.* 24, 483−490.

(10) Temple, R. J., and Himmel, M. H. (2002) Safety of newly approved drugs: implications for prescribing. *JAMA, J. Am. Med. Assoc.* 287, 2273−2275.

(11) Russo, M. W., Galanko, J. A., Shrestha, R., Fried, M. W., and Watkins, P. (2004) Liver transplantation for acute liver failure from drug induced liver injury in the United States. *Liver Transplantation* 10, 1018−1023.

(12) Ma, S., and Subramanian, R. (2006) Detecting and characterizing reactive metabolites by liquid chromatography/tandem mass spectrometry. *J. Mass Spectrom.* 41, 1121−1139.

(13) Tohkin, M., Ishiguro, A., Kaniwa, N., Saito, Y., Kurose, K., and Hasegawa, R. (2010) Prediction of severe adverse drug reactions using pharmacogenetic biomarkers. *Drug Metab. Pharmacokinet.* 25, 122−133.

(14) Uetrecht, J. P., Ma, H. M., MacKnight, E., and McClelland, R. (1995) Oxidation of aminopyrine by hypochlorite to a reactive dication: possible implications for aminopyrine-induced agranulocytosis. *Chem. Res. Toxicol.* 8, 226−233.

(15) Andrès, E., Federici, L., Weitten, T., Vogel, T., and Alt, M. (2008) Recognition and management of drug-induced blood cytopenias: the example of drug-induced acute neutropenia and agranulocytosis. *Expert Opin. Drug Saf.* 7, 481−489.

(16) Adams, D. A., and Perry, S. (1958) Agranulocytosis associated with thenalidine (sandostene) tartrate therapy: report of three cases. *JAMA, J. Am. Med. Assoc. 167*, 1207−1210.

(17) Stevenson, J., and Kennedy, A. (1961) A fatal case of agranulocytosis due to thenalidine tartrate complicated by acute renal failure and mycelial abscesses of brain. *Scott. Med. J. 6*, 522−525.

(18) Zolov, B. (1958) Agranulocytosis resulting from sandostene. *J. Maine Med. Assoc. 49*, 335−337.

(19) Srivastava, A., Maggs, J., Antoine, D., Williams, D., Smith, D., and Park, B. *Adverse Drug Reactions*; Springer: Berlin, 2010; pp 165−194.

(20) Bauman, J. N., Kelly, J. M., Tripathy, S., Zhao, S. X., Lam, W. W., Kalgutkar, A. S., and Obach, R. S. (2009) Can in vitro metabolism-dependent covalent binding data distinguish hepatotoxic from nonhepatotoxic drugs? An analysis using human hepatocytes and liver S-9 fraction. *Chem. Res. Toxicol. 22*, 332−340.

(21) Almario, E. E., Borlak, J., Suzuki, A., and Chen, M. (2017) Drug-Induced Liver Injury. *BioMed Res. Int. 2017*, 2461694.

(22) Guengerich, F. P. (2011) Mechanisms of drug toxicity and relevance to pharmaceutical development. *Drug Metab. Pharmacokinet. 26*, 3−14.

(23) Kaplowitz, N. (2013) Avoiding idiosyncratic DILI: two is better than one. *Hepatology 58*, 15−17.

(24) Deng, X., Luyendyk, J. P., Ganey, P. E., and Roth, R. A. (2009) Inflammatory stress and idiosyncratic hepatotoxicity: hints from animal models. *Pharmacol. Rev. 61*, 262−282.

(25) Smith, G. F. (2011) Designing drugs to avoid toxicity. *Prog. Med. Chem. 50*, 1−47.

(26) Frank, C., Himmelstein, D. U., Woolhandler, S., Bor, D. H., Wolfe, S. M., Heymann, O., Zallman, L., and Lasser, K. E. (2014) Era of faster FDA drug approval has also seen increased black-box warnings and market withdrawals. *Health Affairs 33*, 1453−1459.

(27) Kang, P., Dalvie, D., Smith, E., and Renner, M. (2009) Bioactivation of lumiracoxib by peroxidases and human liver microsomes: identification of multiple quinone imine intermediates and GSH adducts. *Chem. Res. Toxicol. 22*, 106−117.

(28) Li, Y., Slatter, J. G., Zhang, Z., Li, Y., Doss, G. A., Braun, M. P., Stearns, R. A., Dean, D. C., Baillie, T. A., and Tang, W. (2008) In vitro metabolic activation of lumiracoxib in rat and human liver preparations. *Drug Metab. Dispos. 36*, 469−473.

(29) Tujios, S., and Fontana, R. J. (2011) Mechanisms of drug-induced liver injury: from bedside to bench. *Nat. Rev. Gastroenterol. Hepatol. 8*, 202.

(30) Núñez-Vergara, L. J., Sturm, J., Olea-Azar, C., Navarrete-Encina, P., Bollo, S., and Squella, J. (2000) Electrochemical, UV-Visible and EPR studies on nitrofurantoin: Nitro radical anion generation and its interaction with glutathione. *Free Radical Res. 32*, 399−409.

(31) Joshi, E. M., Heasley, B. H., Chordia, M. D., and Macdonald, T. L. (2004) In vitro metabolism of 2-acetylbenzothiophene: relevance to zileuton hepatotoxicity. *Chem. Res. Toxicol. 17*, 137−143.

(32) Stepan, A. F., Walker, D. P., Bauman, J., Price, D. A., Baillie, T. A., Kalgutkar, A. S., and Aleo, M. D. (2011) Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. *Chem. Res. Toxicol. 24*, 1345−1410.

(33) Masubuchi, N., Makino, C., and Murayama, N. (2007) Prediction of in vivo potential for metabolic activation of drugs into chemically reactive intermediate: correlation of in vitro and in vivo generation of reactive intermediates and in vitro glutathione conjugate formation in rats and humans. *Chem. Res. Toxicol. 20*, 455−464.

(34) Wirth, P. J., Bettis, C. J., and Nelson, W. L. (1976) Microsomal metabolism of furosemide evidence for the nature of the reactive intermediate involved in covalent binding. *Mol. Pharmacol. 12*, 759−768.

(35) Qu, Q., Liu, J., Zhou, H.-H., and Klaassen, C. D. (2014) Nrf2 protects against furosemide-induced hepatotoxicity. *Toxicology 324*, 35−42.

(36) McMurtry, R. J., and Mitchell, J. R. (1977) Renal and hepatic necrosis after metabolic activation of 2-substituted furans and thiophenes, including furosemide and cephaloridine. *Toxicol. Appl. Pharmacol. 42*, 285−300.

(37) Martin, E. A., Rich, K. J., White, I. N., Woods, K. L., Powles, T. J., and Smith, L. L. (1995) 32P-postlabelled DNA adducts in liver obtained from women treated with tamoxifen. *Carcinogenesis 16*, 1651−1654.

(38) Boocock, D. J., Maggs, J. L., White, I. N., and Park, B. K. (1999) α-Hydroxytamoxifen, a genotoxic metabolite of tamoxifen in the rat: identification and quantification in vivo and in vitro. *Carcinogenesis 20*, 153−160.

(39) Aithal, G. P., Ramsay, L., Daly, A. K., Sonchit, N., Leathart, J., Alexander, G., Kenna, J. G., Caldwell, J., and Day, C. P. (2004) Hepatic adducts, circulating antibodies, and cytokine polymorphisms in patients with diclofenac hepatotoxicity. *Hepatology 39*, 1430−1440.

(40) Robin, M. A., Maratrat, M., Roy, M. L., Breton, F. P. L., Bonierbale, E., Dansette, P., Ballet, F., Mansuy, D., and Pessayre, D. (1996) Antigenic targets in tienilic acid hepatitis. Both cytochrome P450 2C11 and 2C11-tienilic acid adducts are transported to the plasma membrane of rat hepatocytes and recognized by human sera. *J. Clin. Invest. 98*, 1471−1480.

(41) Cribb, A. E., Nuss, C. E., Alberts, D. W., Lamphere, D. B., Grant, D. M., Grossman, S. J., and Spielberg, S. P. (1996) Covalent binding of sulfamethoxazole reactive metabolites to human and rat liver subcellular fractions assessed by immunochemical detection. *Chem. Res. Toxicol. 9*, 500−507.

(42) Bylund, J., Macsari, I., Besidski, Y., Olofsson, S., Petersson, C., Arvidsson, P. I., and Bueters, T. (2012) Novel bioactivation mechanism of reactive metabolite formation from phenyl methyl-isoxazoles. *Drug Metab. Dispos. 40*, 2185−2191.

(43) Isin, E. M., Elmore, C. S., Nilsson, G. N., Thompson, R. A., and Weidolf, L. (2012) Use of Radiolabeled Compounds in Drug Metabolism and Pharmacokinetic Studies. *Chem. Res. Toxicol. 25*, 532−542. PMID: 22372867.

(44) Zhang, D., Krishna, R., Wang, L., Zeng, J., Mitroka, J., Dai, R., Narasimhan, N., Reeves, R. A., Srinivas, N. R., and Klunk, L. J. (2005) Metabolism, pharmacokinetics, and protein covalent binding of radiolabeled MaxiPost (BMS-204352) in humans. *Drug Metab. Dispos. 33*, 83−93.

(45) Kalgutkar, A. S., Gardner, I., Obach, R. S., Shaffer, C. L., Callegari, E., Henne, K. R., Mutlib, A. E., Dalvie, D. K., Lee, J. S., Nakai, Y., O'Donnell, J. P., Boer, J., and Harriman, S. P. (2005) A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab. 6*, 161−225.

(46) Evans, D. C., Watt, A. P., Nicoll-Griffith, D. A., and Baillie, T. A. (2004) Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol. 17*, 3−16.

(47) Alves, V. M., et al. (2016) Alarms about structural alerts. *Green Chem. 18*, 4348−4360.

(48) Hughes, T. B., and Swamidass, S. J. (2017) Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism. *Chem. Res. Toxicol. 30*, 642−656.

(49) Hughes, T. B., Miller, G. P., and Swamidass, S. J. (2015) Modeling Epoxidation of Drug-like molecules with a Deep Machine Learning Network. *ACS Cent. Sci. 1*, 168−180.

(50) Dang, N. L., Matlock, M. K., Hughes, T. B., and Swamidass, S. J. (2020) The Metabolic Rainbow: Deep Learning Phase I Metabolism in Five Colors. *J. Chem. Inf. Model. 60*, 1146−1164.

(51) Hughes, T. B., Dang, N. L., Kumar, A., Flynn, N. R., and Swamidass, S. J. (2020) The Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites. *J. Chem. Inf. Model. 60*, 4702−4716.

(52) Hughes, T. B., Dang, N. L., Miller, G. P., and Swamidass, S. J. (2016) Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network. *ACS Cent. Sci. 2*, 529−537.

(53) Matlock, M. K., Hughes, T. B., Dahlin, J. L., and Swamidass, S. J. (2018) Modeling Small-Molecule Reactivity Identifies Promiscuous Bioactive Compounds. *J. Chem. Inf. Model. 58*, 1483−1500.

(54) Dang, N. L., Hughes, T. B., Miller, G. P., and Swamidass, S. J. (2018) Computationally Assessing the Bioactivation of Drugs by N-Dealkylation. *Chem. Res. Toxicol. 31*, 68−80.

(55) Ruder, S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747* 2016, https://arxiv.org/abs/1609.04747 (accessed 2021-01-25).

(56) Zaretzki, J., Matlock, M., and Swamidass, S. J. (2013) XenoSite: Accurately predicting CYP-mediated sites of metabolism with neural networks. *J. Chem. Inf. Model. 53*, 3373−3383.

(57) Cruciani, G., Baroni, M., Benedetti, P., Goracci, L., and Fortuna, C. G. (2013) Exposition and reactivity optimization to predict sites of metabolism in chemicals. *Drug Discovery Today: Technol. 10*, e155−e165.

(58) Zamora, I., Afzelius, L., and Cruciani, G. (2003) Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9. *J. Med. Chem. 46*, 2313−2324.

(59) Rydberg, P., Gloriam, D. E., Zaretzki, J., Breneman, C., and Olsen, L. (2010) SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism. *ACS Med. Chem. Lett. 1*, 96−100.

(60) Hughes, T. B., Miller, G. P., and Swamidass, S. J. (2015) Site of Reactivity Models Predict Molecular Reactivity of Diverse Chemicals with Glutathione. *Chem. Res. Toxicol. 28*, 797−809.

(61) Swamidass, S. J., Azencott, C.-A., Daily, K., and Baldi, P. (2010) A CROC stronger than ROC: measuring, visualizing and optimizing early retrieval. *Bioinformatics 26*, 1348−1356.

(62) Cortes, C., and Mohri, M. (2005) Confidence Intervals for the Area Under the ROC Curve. *Advances in Neural Information Processing Systems 17 - Proceedings of the 2004 Conference, NIPS 2004*, Vol. 17, pp 305−312.

(63) Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recognition Letters 27*, 861−874. ROC Analysis in Pattern Recognition

(64) Lim, H. K., and Silva, J. (2009) Structure elucidation of thioether conjugates of ethacrynic acid in NADPH-and glutathione-fortified human liver S9 by microbore LC-LTQ/Orbitrap. *Drug Metab. Rev.*, 136−137.

(65) Wen, B., and Fitch, W. L. (2009) Screening and characterization of reactive metabolites using glutathione ethyl ester in combination with Q-trap mass spectrometry. *J. Mass Spectrom. 44*, 90−100.

(66) Chen, Q., Doss, G. A., Tung, E. C., Liu, W., Tang, Y. S., Braun, M. P., Didolkar, V., Strauss, J. R., Wang, R. W., Stearns, R. A., Evans, D. C., Baillie, T. A., and Tang, W. (2006) Evidence for the bioactivation of zomepirac and tolmetin by an oxidative pathway. Identification of glutathione adducts in vitro in human liver microsomes and in vivo in rats. *Drug Metab. Dispos. 34*, 145−151.

(67) Masubuchi, Y., Igarashi, S., Suzuki, T., Horie, T., and Narimatsu, S. (1996) Imipramine-induced inactivation of a cytochrome P450 2D enzyme in rat liver microsomes: in relation to covalent binding of its reactive intermediate. *J. Pharmacol. Exp. Ther. 279*, 724−731.

(68) Kalgutkar, A. S., and Soglia, J. R. (2005) Minimising the potential for metabolic activation in drug discovery. *Expert Opin. Drug Metab. Toxicol. 1*, 91−142.

(69) Kumar, R., and Indrayan, A. (2011) Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr. 48*, 277−287.

(70) Bakke, O. M., Wardell, W. M., and Lasagna, L. (1984) Drug discontinuations in the United Kingdom and the United States, 1964 to 1983: issues of safety. *Clin. Pharmacol. Ther. 35*, 559−567.

(71) Leung, L., Kalgutkar, A. S., and Obach, R. S. (2012) Metabolic activation in drug-induced liver injury. *Drug Metab. Rev. 44*, 18−33.

(72) Slack, J., and Ford-Hutchinson, A. (1980) Determination of a urinary epoxide metabolite of alclofenac in man. *Drug Metab. Dispos. 8*, 84−86.

(73) Slack, J., Ford-Hutchinson, A., Richold, M., and Choi, B. (1981) Some biochemical and pharmacological properties of an epoxide metabolite of alclofenac. *Chem.-Biol. Interact. 34*, 95−107.

(74) Mercier, M., Poncelet, F., De Meester, C., McGregor, D., Willins, M., Léonard, A., and Fabry, L. (1983) In vitro and in vivo studies on the potential mutagenicity of alclofenac, dihydroxyalclofenac and alclofenac epoxide. *J. Appl. Toxicol. 3*, 230−236.

(75) World Health Organization, (2005) *Pharmaceuticals: restrictions in use and availability*.

(76) Arai, M., Yokosuka, O., Fujiwara, K., Kojima, H., Kanda, T., Hirasawa, H., and Saisho, H. (2002) Fulminant hepatic failure associated with benzbromarone treatment: a case report. *J. Gastroenterol. Hepatol. 17*, 625−626.

(77) Wagayama, H., Shiraki, K., Sugimoto, K., Fujikawa, K., Shimizu, A., Takase, K., Nakano, T., and Tameda, Y. (2000) Letters to the Editor: Fatal fulminant hepatic failure associated with benzbromarone. *J. Hepatol. 32*, 874.

(78) Suzuki, T., Kimura, M., Shinoda, M., Fujita, T., Miyake, N., Yamamoto, S., and Tashiro, K. (2001) A case of fulminant hepatitis, possibly caused by benzbromarone. *Nihon Shokakibyo Gakkai zasshi= The Japanese journal of gastro-enterology 98*, 421−425.

(79) McDonald, M. G., and Rettie, A. E. (2007) Sequential metabolism and bioactivation of the hepatotoxin benzbromarone: formation of glutathione adducts from a catechol intermediate. *Chem. Res. Toxicol. 20*, 1833−1842.

(80) Entzeroth, M., and Ratty, A. K. (2017) Monoamine Oxidase Inhibitors—Revisiting a Therapeutic Principle. *Open Journal of Depression 6*, 31.

(81) Fišar, Z. (2016) Drugs related to monoamine oxidase activity. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry 69*, 112−124.

(82) Dexter, S. (1984) Zimeldine induced neuropathies. *Hum. Toxicol. 3*, 141−143.

(83) Fagius, J., Osterman, P., Siden, A., and Wiholm, B. (1985) Guillain-Barré syndrome following zimeldine treatment. *J. Neurol., Neurosurg. Psychiatry 48*, 65−69.

(84) Zhou, S., Chan, E., Duan, W., Huang, M., and Chen, Y.-Z. (2005) Drug bioactivation covalent binding to target proteins and toxicity relevance. *Drug Metab. Rev. 37*, 41−213.

(85) Cashman, J. R., Yang, Z. C., and Hoegberg, T. (1990) Oxidation of N-hydroxynorzimeldine to a stable nitrone by hepatic monooxygenases. *Chem. Res. Toxicol. 3*, 428−432.

(86) Lundström, J., Högberg, T., Gosztonyi, T., and de Paulis, T. (1981) Metabolism of zimelidine in rat, dog and man. Identification and synthesis of the principal metabolites. *Arzneim.-Forsch. 31*, 486−494.

(87) Wiley, J. F., Gelber, M. L., Henretig, F. M., Wiley, C. C., Sandhu, S., and Loiselle, J. (1992) Cardiotoxic effects of astemizole overdose in children. *J. Pediatr. 120*, 799−802.

(88) Zhang, W., Roederer, M. W., Chen, W.-Q., Fan, L., and Zhou, H.-H. (2012) Pharmacogenetics of drugs withdrawn from the market. *Pharmacogenomics 13*, 223−231.

(89) Hamada, T., and Awata, N. (1990) Metabolism of a New Antiallergic Agent, Emedastine Difumarate, in Human. *Yakubutsu Dotai 5*, 871−881.

(90) Chen, Q., Doss, G. A., Tung, E. C., Liu, W., Tang, Y. S., Braun, M. P., Didolkar, V., Strauss, J. R., Wang, R. W., Stearns, R. A., Evans, D. C., Baillie, T. A., and Tang, W. (2006) Evidence for the bioactivation of zomepirac and tolmetin by an oxidative pathway: identification of glutathione adducts in vitro in human liver microsomes and in vivo in rats. *Drug Metab. Dispos. 34*, 145−151.

(91) Leone, A., Nie, A., Brandon Parker, J., Sawant, S., Piechta, L.-A., Kelley, M. F., Mark Kao, L., Jim Proctor, S., Verheyen, G., Johnson, M. D., et al. (2014) Oxidative stress/reactive metabolite gene expression signature in rat liver detects idiosyncratic hepatotoxicants. *Toxicol. Appl. Pharmacol. 275*, 189−197.

(92) Wen, B., and Zhou, M. (2009) Metabolic activation of the phenothiazine antipsychotics chlorpromazine and thioridazine to electrophilic iminoquinone species in human liver microsomes and recombinant P450s. *Chem.-Biol. Interact. 181*, 220−226.

(93) Tettey, J. N., Maggs, J. L., Rapeport, W. G., Pirmohamed, M., and Park, B. K. (2001) Enzyme-Induction Dependent Bioactivation of Troglitazone and Troglitazone Quinone In Vivo. *Chem. Res. Toxicol. 14*, 965−974. PMID: 11511170.

(94) Hinson, J. A. (1983) Reactive metabolites of phenacetin and acetaminophen: a review. *Environ. Health Perspect. 49*, 71−79.

(95) Wani, T. H., Chakrabarty, A., Shibata, N., Yamazaki, H., Guengerich, F. P., and Chowdhury, G. (2017) The Dihydroxy Metabolite of the Teratogen Thalidomide Causes Oxidative DNA Damage. *Chem. Res. Toxicol. 30*, 1622−1628. PMID: 28745489.