Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Chem Res Toxicol. Author manuscript; available in PMC 2018 March 27.

Published in final edited form as:

Chem Res Toxicol. 2015 April 20; 28(4): 797–809. doi:10.1021/acs.chemrestox.5b00017.

Site of Reactivity Models Predict Molecular Reactivity of Diverse Chemicals with Glutathione

Tyler B. Hughes[*,†], **Grover P. Miller**[*,‡], and **S. Joshua Swamidass**[*,†]

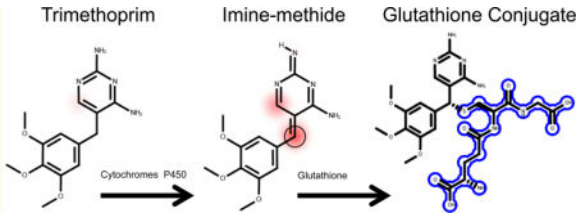
[†]Department of Pathology and Immunology, Washington University School of Medicine, Campus Box 8118, 660 S. Euclid Ave., St. Louis, Missouri 63110, United States

[‡]Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, United States

Abstract

Drug toxicity is often caused by electrophilic reactive metabolites that covalently bind to proteins. Consequently, the quantitative strength of a molecule's reactivity with glutathione (GSH) is a frequently used indicator of its toxicity. Through cysteine, GSH (and proteins) scavenges reactive molecules to form conjugates in the body. GSH conjugates to specific atoms in reactive molecules: their sites of reactivity. The value of knowing a molecule's sites of reactivity is unexplored in the literature. This study tests the value of site of reactivity data that identifies the atoms within 1213 reactive molecules that conjugate to GSH and builds models to predict molecular reactivity with glutathione. An algorithm originally written to model sites of cytochrome P450 metabolism (called XenoSite) finds clear patterns in molecular structure that identify sites of reactivity within reactive molecules with 90.8% accuracy and separate reactive and unreactive molecules with 80.6% accuracy. Furthermore, the model output strongly correlates with quantitative GSH reactivity data in chemically diverse, external data sets. Site of reactivity data is nearly unstudied in the literature prior to our efforts, yet it contains a strong signal for reactivity that can be utilized to more accurately predict molecule reactivity and, eventually, toxicity.

Graphical abstract



[*]**Corresponding Authors:** (T.B.H.) tyler@wustl.edu. (G.P.M.) MillerGroverP@uams.edu. (S.J.S.) swamidass@wustl.edu.

Supporting Information

Tables S1 and S2 define the molecule and atom level descriptors, respectively, used by the XenoSite reactivity model. Table S3 lists the groups of descriptors used in sensitivity analysis. The AMD Registry numbers file lists the reaction and molecule registry numbers of our data set, based on the June 2014 AMD release, as well as a column of binary values indicating whether a molecule is reactive with GSH. This material is available free of charge via the Internet at http://pubs.acs.org.

Notes

The authors declare no competing financial interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

INTRODUCTION

It costs about one billion dollars to bring a single drug to market.[1,2] Efforts to make the process more efficient may be necessary for survival of the pharmaceutical industry.[3,4] An estimated 50% of drug leads fail due to low efficacy, and 40% fail due to toxicity issues;[5] both of these issues are often associated with failure to properly predict the impact of drug metabolism.[6] In fact, drug-induced liver injury (DILI) is the most common culprit for withdrawal of already approved drugs from the market and termination of a drug's clinical investigation.[7] DILI is responsible for 50% of acute liver failure cases as well as 15% of liver transplants within the United States.[7] This drug toxicity often reflects formation of electrophilic reactive metabolites, such as quinones or epoxides, which covalently bind to proteins or DNA.[7,8] These metabolites typically bind nucleophilic sites in proteins, such as thiols,[9] and then cause adverse drug reactions by eliciting an immune response.[7,10,11]

Consequently, efforts to model and understand the formation of reactive metabolites could improve the efficiency of drug development and the safety of future medicines. Specifically, a system capable of modeling how metabolism gives rise to reactive metabolites would be valuable in drug development and discovery. The first part of this system, quantitative models of metabolism, already exists. Several studies, by our group and others, have already demonstrated that computational models can predict how molecules are metabolized.[12–17] In contrast to metabolism, comparatively little has been done to model molecule reactivity. Until now, there have been no mathematical models of molecule reactivity that work on diverse molecules. This study uses tools previously used to model metabolism to build effective models of reactivity. Ultimately, we hope to combine this reactivity model with metabolism models and other key factors, like dose,[18] to better understand and predict toxicity.

Here, we focus on modeling the reactivity of molecules with glutathione (GSH). In the liver, reactive metabolites are trapped by GSH, the most abundant peptide in the body (Figure 1).[19–22] Experimental systems exploit this fact to monitor reactive metabolite formation by detecting metabolites conjugated to GSH.[19,23,24] In comparison to *in vitro* approaches, computational methods can more quickly and cheaply predict molecular properties relevant to drug discovery and development.

Prior work in modeling reactivity has been limited in critical ways. The presence of specific structural alerts in molecules is commonly used to flag potentially toxic molecules, yet these alerts do not distinguish between reactive and nonreactive molecules that contain the same substructure.[25] For example, all Michael acceptors are flagged as problematic even though many are not actually reactive.[24,26] By contrast, quantitative structural–activity relationship (QSAR) models can correctly identify reactive molecules with the same alert, but their utility extends only to closely related molecules.[19,24] Others have suggested that indices derived from quantum simulations may predict reactivity, but this idea has been validated only in limited studies that do not include GSH.[27–31] A computational way of predicting the reactivity of a large range of diverse molecules would be a major advance in the field.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

In this study, we constructed a model to predict the GSH reactivity of diverse molecules. Our approach implemented several critical advances over prior methods. First, unlike QSAR methods trained on small sets of structurally similar molecules, our model was tuned to the structural data from over 1400 diverse molecules, orders of magnitude more molecules than published QSAR methods. Second, we modeled reactivity at an atom level, using a combination of topological and quantum descriptors. This fine-grained approach started with the reactivity of atoms instead of molecules to make structurally localized predictions about the source of a molecule's reactivity. Third, we used a deep neural network to find a mapping between these descriptors and molecule reactivity and atom-level sites of reactivity or SOR.[32] We found success with this approach when predicting P450 metabolism.[12] When applied in this study, the strategy was capable of encoding nonlinear relationships and simultaneously making SOR predictions for each atom in a molecule along with GSH reactivity predictions for the molecule as a whole. The validation of those models demonstrated the ability to model effectively GSH reactivity of diverse chemicals, both identifying reactive molecules and sites of reactivity.

MATERIALS AND METHODS

Glutathione Site of Reactivity Training Data

As a starting point, we mined a large chemically diverse training data set from the Accelrys Metabolite Database (AMD). Reactive molecules were identified from 1281 reactions of molecules with GSH. Reactions were validated by checking that each reaction's starting molecule did not contain GSH and the product molecule did contain GSH. An automated algorithm used the structure of the starting and product molecules to identify the reactive atom within each reactive molecule. The final data set included 1213 reactive molecules, each with atoms marked if they conjugated to GSH. Structurally similar but unreactive molecules were mined from the full reaction network for each reactive molecule. From this network, metabolic parent and sibling molecules were identified. After excluding molecules already identified as reactive, the remaining 271 molecules were marked as unreactive. Each one is metabolically studied and chemically similar to a reactive molecule in the data set.

This set of molecules contained a wide range of chemically diverse molecules, including epoxides, classically defined Michael acceptors, and more. Unfortunately, our license for the AMD data did not allow us to publish the structures of the entire data set. The registry numbers for all molecules are included in the Supporting Information, and this is enough information to reconstruct the database and replicate our results.

External Reactivity Data

We assessed the predictive ability of the reactivity model against two published external quantitative data sets. The first study focused on 10 substituted *p*-benzoquinone compounds, measuring both the rate at which molecules react with GSH (log k_{GSH}) and rat hepatocyte viability (the concentration at which 50% lethality is observed, log LC_{50}).[24] A second study measured the reactivity of 38 structurally diverse contact allergens molecules with several amino acids and GSH, as measured by their percent depletion of the trapping agent after 24 h or 15 min, respectively.[23]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Comparison QSAR Models

As a baseline for comparison, QSAR models reported in the literature might be expected to predict GSH reactivity.[19,24] Those studies involved small training sets with limited chemical diversity. Consequently, we did not expect these models to generalize well to diverse chemicals, but they are still useful examples against which to compare our approach. Specifically, we selected two QSAR models trained on the reactivity and toxicity of 10 *p*-benzoquinones.[24] The source study included a total of eight QSAR models, from which we selected the two reported by the authors to have the highest performance on their training data

$$\log \mathrm{LC}_{50}=24.54+17.7E_{\mathrm{LUMO}}+3.36(E_{\mathrm{LUMO}})^2 \quad (1)$$

$$\log k_{\mathrm{GSH}}=-18.38-16.78E_{\mathrm{LUMO}}-3.19(E_{\mathrm{LUMO}})^2 \quad (2)$$

The authors excluded one molecule as an outlier when training eq 1, whereas eq 2 was based on the reactivity of all 10 molecules. The other models in this study were inapplicable because they used the electron density on specific atoms within the benzoquinone core structure, a quantity that was not computable for diverse chemical structures.

Descriptors

Our study utilized a total of 116 different descriptors, including both atom- and molecule-level descriptors. These descriptors were computed using in-house software that takes as input SDF files with 3D coordinates (generated using Open Babel) and explicit hydrogens.[33] Descriptor subgroups include topological, quantum chemical reactivity, and molecule-level descriptors.[34] Table 1 provides a condensed summary of quantum chemical reactivity descriptors discussed in this article; a comprehensive table is available in the Supporting Information. The bulk of our descriptors have been previously shown to be useful for the XenoSite metabolism model, although we supplement them with new reactivity descriptors in the current study.[12] Several of these descriptors have been proposed as reactivity indices, such as the energies of the lowest unoccupied and highest occupied molecular orbitals (E_{LUMO} and E_{HOMO}), the maximum nucleophilic and electrophilic delocalizabilities (max[$D^{\mathrm{N}}(r)$] and max[$D^{\mathrm{E}}(r)$]), and the maximum self-polarizability (max[$\pi^{S}(r)$]). Moreover, $D^{\mathrm{N}}(r)$, $D^{\mathrm{E}}(r)$, and $\pi^{\mathrm{S}}(r)$ have been proposed as atom-level reactivity indices that may predict sites of GSH reactivity.[35–37] These reactivity descriptors are calculated from self-consistent field computations using MOPAC, a semiempirical quantum chemistry modeler, using the PM7 force field and an implicit solvent model.[38,39]

We do not use fingerprints or fingerprint similarity as a descriptor. Fingerprints are a very powerful and easy way to implicitly encode molecular shape, which makes them particularly useful in predicting protein–ligand binding. As useful as fingerprints are in virtual screening,[40] off-target prediction,[41–43] and high-throughput screening analysis,[44,45] their use to predict reactivity would be problematic. Instead of overall shape, it is specific substructures in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

molecules that give rise to GSH reactivity. Furthermore, very similar molecules can have very different GSH reactivity, as observed for unreactive acetaminophen and reactive NAPQI despite a difference of just two hydrogens and, consequently, a very high similarity by fingerprints.[46] This example is not an anomaly: virtually all Michael acceptors have a "sibling" molecule with additional hydrogens that is not reactive. It is not surprising, therefore, that no reported methods use fingerprints to predict reactivity.

Combined Atom- and Molecule-Level Reactivity Model

We built a model for atom and molecule reactivity using a neural network with one input layer, one hidden layer, and two output layers (Figure 2). This network is similar to commonly used neural networks that have an input, hidden, and output layer. One of the output layers represents molecule-level predictions as molecule reactivity scores (MRS); the other output layer represents atom-level predictions as atom reactivity scores (ARS). We trained this model in a two-stage process.

First, we trained the atom network in which each atom within a molecule was a candidate reactive atom, which we defined as the site of binding to GSH. Every atom had a vector of numbers, or descriptors, such that each entry of the vector described a chemical property of that atom. The data set was a matrix, organized as one row per atom and one column per descriptor. An additional binary target vector labeled the experimentally observed reactive atoms with a 1. We used the Pybel python library to identify topologically equivalent atoms and label atoms equivalent to reactive atoms as reactive for the purposes of training.[47] The weights of the network were tuned using gradient descent on the cross-entropy error so that sites of reactivity scored a higher ARS than other atoms. These ARS ranged from zero to one, reflecting the probability that an atom is reactive.

Next, the molecule-output node was trained to compute the MRS. The data matrix was composed of one row per molecule and one column per descriptor. The binary target vector labeled the reactive molecules with a 1. A logistic regressor found a scoring function that gave reactive molecules high scores and unreactive molecules low scores, as represented by the MRS, which ranged from zero to one. The descriptors for each molecule were the top five ARS corresponding to the scores of the five atoms predicted to be the most reactive within a molecule as well as all molecule-level descriptors. Less accurate variations that excluded either the molecule or bonds were considered in the Results and Discussion section.

RESULTS AND DISCUSSION

In the following sections, four strategies assessed the quality of reactivity models built using this approach. First, we determined the accuracy of the ARS computed by the model in accurately identifying the site of reactivity (SOR) within reactive molecules. Second, we evaluated the accuracy of the MRS computed by the model in separating reactive and unreactive molecules. Third, in both cases, we assessed the plausibility of the model by determining the descriptors upon which the model relies. Fourth, we assessed the reactivity model using external quantitative data from the literature.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Accuracy in Identifying Sites of Reactivity

One goal was to accurately identify the atom within a reactive molecule that covalently conjugated with GSH. Knowledge of the specific site of conjugation in a molecule, its SOR, can be used to guide modifications of a reactive molecule to make it less reactive. In addition, SOR predictions lead to specific and testable hypotheses about the mechanism of a molecule's reactivity. To date, none of the reported GSH reactivity models identify SORs in reactive molecules and thus the model for SOR prediction in this study is the first of its kind.

Once trained, the model computed an ARS for each atom in a test molecule. These scores ranged between zero and one and can be regarded as the probability that the corresponding atom reacts with GSH in our data set. Within reactive molecules, the ARS reported by the model should clearly distinguish between reactive and unreactive atoms, thereby accurately identifying SORs.

We estimated the accuracy of our model using cross-validation. In this approach, molecules were separated into metabolically related clusters based on connections through metabolic reactions in the database. Each cluster of molecules was removed from the training set one at a time. The remainder of the molecules was used to train a model and make predictions on all of the molecules present in the cluster left out of the training process. In this way, predictions were made on all molecules in the training data. In each cross-validation fold, the model predictions for test molecules then did not depend on training data from the same or closely related molecules. Next, for each molecule, we quantified the separation of known SORs from unreactive atoms using the area under the ROC curve (AUC). The whole data set performance was quantified by averaging the molecule-level AUCs for each molecule in the data set.

This cross-validation strategy quantitatively measured the accuracy of the ARS scores from the model. Currently, the only other published methods of predicting reactive atoms are in the quantum modeling literature, which proposes several reactivity indices.[27–31] The accuracy of these indices in identifying the SOR was a useful baseline against which to compare ARS performance. This comparison revealed several critical observations. First, the ARS reported by the model more accurately separated reactive and unreactive atoms than any other method, with a cross-validated average AUC of 90.8% (Figure 3).

Second, ARS accuracy was greater than any of the individual descriptors including those specifically designed to predict reactivity. For example, self-polarizability, $\pi^S(i)$, separated reactive and unreactive atoms with an accuracy of 75.3%. Likewise, the nucleophilic and electrophilic delocalizabilities ($D^N(i)$ and $D^E(i)$), also known as fukui reactivity, separated reactive and unreactive atoms, but they did so with reduced accuracies of 72.4 and 70.5%, respectively. Third, a logistic regressor for ARS (ARS[LR] in the figure) was more accurate than the reactivity indices, 88.7%, yet it was less accurate than the model derived from a neural network. The 18.6% reduction in error achieved by using the neural network indicated a significant nonlinear component in the reactivity model that was consistent with previous work on the best models predicting sites of metabolism.[48]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Descriptors Driving Atom Reactivity Performance

The identification of essential descriptors gave insight into how the model made predictions, why the model was sensible, and what could be done to improve the model. The importance of specific descriptors for identifying SORs was quantifiable through a permutation sensitivity test.[49] First, a model was built using all of the training data, and the performance on training data was recorded. Next, the importance of individual descriptors (or groups of descriptors) was quantified by measuring the drop in performance of the model on the training data when the descriptor values were randomly shuffled. For the test in this study, the specific descriptors and groups of descriptors were included in the Supporting Information, and only the most salient results are presented here. For each set of descriptors, we repeated the randomization process 10 times and recorded the average performance drop. The larger the drop in performance, the more important the descriptor was to the model.

Surprisingly, topological descriptors were much more important than quantum chemical descriptors in predicting atom-level reactivity (Figure 4). The first and third most important descriptors were the element and hybridization state of the test atom, associated with performance drops of 14.3 and 8.1%, respectively. Ten of the 12 most important descriptors were topological. In contrast, the model was not strongly dependent on quantum chemical descriptors. The most important quantum chemical descriptor, demonstrating a performance drop of 8.1%, was the maximum bond distance between the test atom and a covalently bound hydrogen, a known proxy for the strength of the atom to hydrogen covalent bond.[17,50,51] The next most important quantum descriptor was the density of the E_{LUMO} molecular orbital on the test atom, with an accuracy of only 1.2%. Strikingly, the model did not rely strongly on classic quantum reactivity indices like $D^{E}(r)$, $D^{N}(r)$, or $\pi^{S}(r)$, which were associated with performance drops, respectively, of only 0.5, 0.1, and 0.0% (data not shown in figure).

This outcome may seem to contradict observations in the prior section wherein these reactivity indices could identify SORs with some accuracy. In this case, the predictive values of the descriptors depended on their combined contribution to model performance. None of the individual topological descriptors accurately identified reactive atoms. Instead, the combination of several topological descriptors collectively appeared to very accurately identify reactive atoms such that the information conveyed in the quantum descriptors was not necessary. Importantly, this result suggested that accurate models of GSH reactivity can be constructed using only topological descriptors without requiring a time- and resource-consuming quantum simulation. In a broader context, similar topology-only models may quantify other types of reactivity as well; however, those applications are beyond the scope of this study and left for future work.

Accuracy in Identifying Reactive Molecules

Another key goal of an effective method is to accurately distinguish between GSH reactive and unreactive molecules. Thus far, there have been a few published QSAR models for GSH reactivity. All of these efforts used linear or quadratic regression to map quantum chemical parameters, like the energy of the highest occupied molecular orbital (E_{HOMO}), to the quantitative GSH reactivity of small sets of closely related molecules.[19,24] The authors of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

these studies were quick to point out that these QSAR models were trained on very small data sets of similar molecules, so they are not suited to predict the reactivity of a structurally diverse set of molecules. In contrast, our approach will effectively work across the full range of chemical diversity encountered in drug development programs.

Individual atom reactivity contributes to the overall reactivity of a molecule and thus the model was designed to consider this aggregate property and predict overall reactivity to score test molecules. The reactivity score, the MRS, ranged from zero to one. This value reflected the probability of a test molecule being reactive in our data set. The respective reactivity scores should distinguish between reactive and unreactive molecules across the entire data set. The accuracy of these predictions was quantified and assessed by cross-validation. As discussed for atom reactivity model assessment, individual models were built after excluding each cluster of related molecules in turn. The MRS was computed for each molecule within the cluster left out of the training. In this case, the accuracy was quantified by measuring the area under the ROC curve, the AUC, across the entire data set. As a baseline against which to compare model performance, we also computed the performance of both QSAR models of reactivity published in the literature[24] and reactivity indices proposed in the literature,[35–37] including individual molecule descriptors and the maximum of each atom descriptor associated with the molecule atoms.

From this assessment (Figure 5), several patterns were evident. First, the MRS can distinguish reactive and unreactive molecules with a reasonable accuracy of 80.6%. Our model outperformed several closely related variants, including versions using only ARS output (max[ARS] and MRS[Atom only]) or molecule descriptors (MRS[Molecule only]). This observation suggests that a combination of atom- and molecule-level information was necessary to most accurately predict molecule reactivity. Moreover, the addition of hidden nodes at the top level did not improve accuracy (data not shown), suggesting that this stage of the network does not require nonlinearity to improve accuracy. Second, the model MRS outperformed two published QSAR models for reactivity that both yielded an accuracy of only 65.1%. This result was not surprising, as one of the known limitations of QSAR models is their limited domain of applicability. As expected, QSAR models trained only on a small number of benzoquinones do not generalize to a large, diverse set of molecules. Third, the model MRS outperformed individual quantum chemical descriptors by a wide margin. The quantum chemical descriptors E_{LUMO}, E_{HOMO}, max[$D^{N}(r)$], max[$\pi^{S}(r)$], and max[$D^{E}(r)$] have accuracies of 64.7, 58.2, 56.4, 52.0, and 50.4%, respectively. Fourth, reactive and nonreactive molecules had average scores of 0.85 and 0.65, respectively, which suggests that the model more confidently identified reactive molecules than nonreactive molecules. This could be a consequence of error in the training data arising from difficulty in accurately identifying nonreactive molecules.

An important caveat in this assessment was the reliability of labeling unreactive molecules in the training data. There was a potential for error arising from the difficulty of extracting negative data from literature-derived sources. It was necessary to make assumptions to overcome this shortcoming. We assumed that molecules did not react with GSH if none of the metabolites identified in the literature are GSH conjugates. However, despite this assumption, the absence of GSH conjugates in the database was not strong evidence proving

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a molecule is not reactive. Not all studies look for GSH conjugates, and, consequently, some reactive molecules are incorrectly labeled as unreactive in the training data.

There are two potential consequences for this intrinsic limitation and the error that it introduced into the training data. First, the MRS scores are not directly interpretable as the probability of a reaction occurring between the molecule and GSH in an experimental benchtop study. Rather, the scores are the probability of the molecule being labeled reactive in our data set and, by extension, being recognized as reactive in the literature. Of course, we still hope that this score will correlate closely with experimental measures of reactivity, as will be tested in a subsequent section. Second, the potential error due to mislabeling unreactive molecules could lead to cross-validation performances either over- or under-estimating the true generalization accuracy. Nonetheless, cross-validation analyses provide critical insight into the relative performance of different methods and the importance of individual descriptors. In this case, cross-validation experiments provided the best assessment of models on a wide range of chemically diverse molecules because of the size of the data set.

Descriptors Driving Molecule Reactivity Performance

As pointed out for atom reactivity, knowledge of the useful descriptors provides both a deeper understanding of the model and insight into ways of improving the model. We could once again use a permutation sensitivity analysis to identify important descriptors, but, in this case, a simpler method is available. The final node of the model was a logistic regressor trained using normalized inputs so that the importance of individual descriptors was clear from the weight of the respective nodes. An advantage of weight inspection over permutation sensitivity analysis is that the sign of the weight indicates whether an increase in the descriptor value will favor or disfavor a higher reactivity score. This outcome can facilitate an assessment of the plausibility of the model.

The model used a combination of atom- and molecule-level descriptors that confirmed observations about the overall performance of model variants (Figure 6). Encouragingly, three of the top descriptors were immediately understandable. First, as expected, the reactivity of the most reactive atom in the molecule (the highest ARS) was an essential feature. Second, the energy of the LUMO orbital was negatively associated with reactivity, a finding consistent with frontier molecular orbital theory. Third, the size of the smallest ring was another important descriptor, which reflects the high enthalpy gain associated with opening three-member rings during bond formation with GSH.[52] Lastly, the weights of the remaining reactivity indices were similarly understandable. In particular, the utility of $\max[D^{N}(r)]$ agreed with prior studies that have implicated the importance of this descriptor in modeling molecule GSH reactivity. Likewise, $\max[D^{E}(r)]$ correlated with the reactivity of molecules containing electrophilic properties and thus, as expected, this descriptor was not a strong predictor.

Collectively, these results increase our confidence in the ability of the model to learn generalizable rules governing relationships between test molecules and reactivity from the data. Descriptors of high importance were consistent with our knowledge of theoretical and experimental knowledge of chemical reactivity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Performance on External Data Sets

We tested reactivity models against two external quantitative data sets to further assess their performance. We stress that these are quantitative data sets, where the degree of reactivity with GSH was measured yet the specific site of reactivity within reactive molecules was not identified. In contrast, the training data was not quantitative by including only binary values labeling atoms and molecules as reactive. Consequently, we did not expect to observe exceedingly strong correlations between the model output and these quantitative data sets per se. Nonetheless, statistically significant correlations are critically important because they ensure that errors in training data are not so significant that they prevent the construction of a useful model.

The first external data set on 10 benzoquinones included both rat toxicity data (measured as the LC_{50}) and experimentally measured rates of reactivity with GSH. In this set of molecules, the GSH reactivity correlated closely with toxicity, so we determined correlations between MRS and either set of experimental values (Figure 7). In both cases, MRS correlated very closely to experimental values with R^2 values of 0.78 and 0.70 and significant p-values of 0.0015 and 0.0025. These findings are striking because they demonstrate the ability of the model trained on qualitative structural data to yield relative quantitative values for the reactivity of molecules.

By comparison, the source paper for this data reported QSAR models with correlations of 0.9 and 0.8. These correlations were higher than those for our model; however, this comparison needs to be interpreted in light of several issues. First, these QSAR models were trained on the data on which they were assessed and not cross-validated. Consequently, their correlations were higher than the true generalization accuracy of the respective models. Second, these QSAR models were overfit to benzoquinones and do not generalize to diverse chemicals, as the authors point out in their study. Third, when evaluating our reactivity model, it is significant that the model did not currently use any quantitative data in the training process, which may limit model performance. In the future, we plan to include quantitative data for building models to improve predictions beyond the insights gained from qualitative data sets. Considering all of these points, the model performance was nearly the same as the QSAR models, showing a statistically significant fit to quantitative data despite being trained only on qualitative structural data. We consider this finding a strong validation of the model.

The second external data set included reactivity of 38 structurally diverse contact allergens with cysteine, lysine, histidine, and GSH.[23] We considered the performance of our models in predicting reactivity between test molecules with all four of these molecule traps. In the reported assays, reactivity was measured as the percent depletion of the trapping agent at 15 min (for GSH) or 24 h (for cysteine, lysine, and histidine). As a single-point measurement, there is more experimental noise in this study than the first external study.

In this study, analysis of all four assays data demonstrated that the model MRS correlated with the experimental data better than the traditional reactivity QSAR models (Table 2 and Figure 8). MRS correlated with GSH depletion with a significant Pearson correlation (R = 0.43), whereas the QSAR model had a nonsignificant correlation (R = 0.30). Of the three

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

amino acids, the highest correlation corresponded to cysteine, which was a soft nucleophile. This observation was reasonable given that the model was trained on reactivity data for the cysteine-containing GSH. Although lysine is a hard nucleophile, XenoSite scores still correlate significantly, suggesting that our model may be sophisticated enough to predict reactivity with both soft and hard nucleophiles. In contrast, histidine is much less nucleophilic than the other amino acids and thus the correlation was weak.

The moderate correlation of 0.43 between the GSH reactivity and the XenoSite reactivity model scores might seem like a weak result. However, this outcome should be interpreted in light of several key points. First, model training relied solely on qualitative data and thus may not yield the most optimal model performances against quantitative data. This shortcoming will be addressed in future studies. Second, a lower correlation was expected in this assessment because single-point measurements of reactivity are very noisy. Third, despite issues with that data set, our trained model yielded statistically significant predictions for 38 structurally diverse molecules. Fourth, the final model performed better that traditional QSAR models, which are not-significantly correlated with the GSH reactivity. In the context of these points, our results provide strong evidence that this model represents a significant advance in reactivity modeling. In these experiments, we demonstrate that the model correlated well with quantitative reactivity data, including GSH reactivity data, reactivity with specific amino acids, and rat toxicity for molecules known to be toxic by a reactive mechanism. These results support the idea that the model was not overfitting the training data and that the data limitations were not preventing the construction of an accurate reactivity model.

Performance on Drug–Metabolite Pairs

Preliminary results from our studies indicate that the reactivity model MRS accurately distinguished drugs from their reactive metabolites. For example, our model assigned acetaminophen a MRS score of 0.75, whereas its reactive metabolite NAQPI scored 0.99. Keeping in mind the average MRS scores of reactive and nonreactive molecules (0.85 and 0.65, respectively), this result is encouraging but not definitive. Though a systematic study of reactive drug metabolite prediction is beyond the scope of this article, we consider two drugs, trimethoprim and felbamate, and their known reactive metabolites (Figure 9).

The metabolism of the antibacterial trimethoprim forms a reactive iminoquinone methide metabolite that likely contributes to trimethoprim hypersensitivity reactions.[25,53] The reactive metabolite conjugates to GSH or *N*-acetyl cysteine in human and rat liver microsomes.[54] Our model successfully distinguished trimethoprim from its toxic metabolite, with an increase in the MRS score from 0.30 to 0.99. Additionally, the most reactive atom on the metabolite predicted by the model corresponded to the same atom known to be conjugated to both GSH and *N*-acetyl cysteine, and, similarly, the second most reactive atom was the same one reported experimentally as a conjugate of *N*-acetyl cysteine.[54] The anticonvulsant felbamate sometimes causes hepatotoxicity and aplastic anemia, and such adverse reactions have been traced to a reactive α,β-unsaturated aldehyde metabolite.[25] Our model assigned this reactive metabolite a very high MRS score of 0.96, whereas felbamate

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

only scores 0.33. Moreover, the atom receiving the highest score was the actual atom known to conjugate to GSH.

These findings are encouraging because they show that the model can correctly identify which of two very similar molecules is reactive. This is a key quality control check that suggests the model is sensibly encoding reactivity.

NON-OBVIOUS PREDICTIONS

One hope is that this approach would identify reactive molecules that do not contain obviously reactive groups. We qualitatively and quantitatively assessed our approach in this task by identifying molecules that are correctly predicted as reactive but do not contain an epoxide or a Michael acceptor structure, two chemical groups widely recognized as reactive. One prior study listed 17 substructures that should be considered Michael acceptors.[26] To this list, we added an epoxide alert, which is also commonly reactive. This list contains α,β-unsaturated carbonyls, the classically defined Michael acceptor. We identified 622 reactive molecules within our training data that do not contain an epoxide or any of these specific substructures. For the purpose of this analysis, we considered these molecules to be non-obvious reactive molecules.

First, we observed that our model correctly identified the SORs in these molecules. The average site AUC was 87.6%, comparable to the performance across the entire database of 90.8%. Specific examples, drawn from these non-obvious molecules, demonstrate that the model correctly predicted the mechanism of reactivity (Figure 10). Second, MRS could identify non-obvious reactive molecules, but it did so with a molecule AUC of only 67%. The reduced power to identify non-obvious reactive molecules is expected because these molecules are more difficult. Nonetheless, the model still identifies the correct SOR with nearly the same accuracy. These two observations demonstrate our approach generalizes beyond simple substructure matching to identify reactive molecules that do not match commonly used structural alerts.

These results are encouraging, suggesting a possible role for the model as a component of a weight-of-evidence strategy for regulatory risk assessment. While the model's results are not definitive, they could build evidence for specific mechanisms of reactivity and toxicity. Either more expensive computational studies or benchwork could then test these mechanisms to build additional evidence. Because the model's predictions are mechanistic, drilling down to specific reactive atoms, this approach could support the mechanism and pathway centered risk assessments that regulatory agencies are developing.

CONCLUSIONS

This study demonstrates a new approach to modeling reactivity using structural reactivity data. The reactivity model is trained on site of reactivity data and identifies with 90.8% accuracy the sites of reactivity within reactive molecules and separates reactive and unreactive molecules with 80.6% accuracy. Furthermore, the model predictions strongly correlate with quantitative GSH reactivity data in chemically diverse, external data sets. This predictive ability is especially encouraging because (1) it is trained only on qualitative data,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

regardless of whether an atom is reactive, and (2) the model is generalizable across broad areas of chemical space. In contrast, traditional QSAR models correlate significantly only with their own training data and fail to generalize for structurally diverse molecules. The current model does have some key limitations. First, the reported model is not trained on all available reactivity data. Specifically, it does not currently make use of quantitative reactivity data (like that in the external data sets). Expanding this effort to include quantitative data when available will likely improve the quality of the model. Second, the current model can detect only molecules reactive with GSH, a soft nucleophile. Some important reactive species do not efficiently react with GSH but, nonetheless, covalently bind DNA or proteins. In the future, we plan to overcome this limitation by including molecules that react with other types of trapping agents, including cyanide, proteins, and DNA. Finally, in the current form, this model does not consider how metabolism gives rise to reactive species. We plan to combine this reactivity model with metabolism models into a system without this limitation. While such a comprehensive model still lies in the future, this study represents a significant step forward by demonstrating that site of reactivity data yields accurate molecule reactivity predictions. Site of reactivity data is nearly unstudied in the literature, but it contains a strong signal for reactivity that can be utilized to more accurately predict molecule reactivity and could also be a useful component of mechanism-based predictions of toxicity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Jed Zaretzki and Na Le Dang for helpful discussions and Matthew Matlock for prior development of the XenoSite algorithm, which is extended in this work. Figures 1, 2, and 7–10 utilized OEDepict.[63] Computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially funded by NIH grant nos. 1S10RR022984-01A1 and 1S10OD018091-01.

Funding

We thank the Department of Immunology and Pathology at the Washington University School of Medicine for its generous support of this work.

ABBREVIATIONS

$\pi^S(r)$	self-polarizability
AMD	Accelrys Metabolite Database
ARS	atom reactivity scores
AUC	area under the receiver operating characteristic curve
$D^E(r)$	electrophilic delocalizability
DILI	drug-induced liver injury
$D^N(r)$	nucleophilic delocalizability

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

E_{HOMO}	energy of the highest occupied molecular orbital
E_{LUMO}	energy of the lowest unoccupied molecular orbital
GSH	glutathione
MRS	molecule reactivity scores
NAPQI	*N*-acetyl-*p*-benzoquinone imine
QSAR	quantitative structural–activity relationship
ROC	receiver operating characteristic
SOR	sites of reactivity

References

1. Hughes J, Rees S, Kalindjian S, Philpott K. Principles of early drug discovery. Br J Pharmacol. 2011; 162:1239–1249. [PubMed: 21091654]
2. Adams CP, Brantner VV. Spending on new drug development. Health Econ. 2010; 19:130–141. [PubMed: 19247981]
3. Borhani DW, Shaw DE. The future of molecular dynamics simulations in drug discovery. J Comput-Aided Mol Des. 2012; 26:15–26. [PubMed: 22183577]
4. Kessel M. The problems with today's pharmaceutical business—an outsider's view. Nat Biotechnol. 2011; 29:27–33. [PubMed: 21221096]
5. DiMasi JA. Success rates for new drugs entering clinical testing in the United States. Clin Pharmacol Ther. 1995; 58:1–14. [PubMed: 7628176]
6. Kuppens IE, Breedveld P, Beijnen J, Schellens J. Modulation of oral drug bioavailability: from preclinical mechanism to therapeutic application. Cancer Invest. 2005; 23:443–464. [PubMed: 16193644]
7. Srivastava, A., Maggs, J., Antoine, D., Williams, D., Smith, D., Park, B. Adverse Drug Reactions. Uetrecht, J., editor. Springer; London: 2010. p. 165-194.
8. Kalgutkar AS, Didiuk MT. Structural alerts, reactive metabolites, and protein covalent binding: how reliable are these attributes as predictors of drug toxicity? Chem Biodiversity. 2009; 6:2115–2137.
9. Wu Z, Han M, Chen T, Yan W, Ning Q. Acute liver failure: mechanisms of immune-mediated liver injury. Liver Int. 2010; 30:782–794. [PubMed: 20492514]
10. Knowles SR, Uetrecht J, Shear NH. Idiosyncratic drug reactions: the reactive metabolite syndromes. Lancet. 2000; 356:1587–1591. [PubMed: 11075787]
11. Numata K, Kubo M, Watanabe H, Takagi K, Mizuta H, Okada S, Kunkel SL, Ito T, Matsukawa A. Overexpression of suppressor of cytokine signaling-3 in T cells exacerbates acetaminophen-induced hepatotoxicity. J Immunol. 2007; 178:3777–3785. [PubMed: 17339476]
12. Zaretzki J, Matlock M, Swamidass SJ. XenoSite: accurately predicting CYP-mediated sites of metabolism with neural networks. J Chem Inf Model. 2013; 53:3373–3383. [PubMed: 24224933]
13. Campagna-Slater V, Pottel J, Therrien E, Cantin LD, Moitessier N. Development of a computational tool to rival experts in the prediction of sites of metabolism of xenobiotics by p450s. J Chem Inf Model. 2012; 52:2471–2483. [PubMed: 22916680]
14. Rydberg P, Gloriam DE, Zaretzki J, Breneman C, Olsen L. SMARTCyp: a 2D method for prediction of cytochrome P450-mediated drug metabolism. ACS Med Chem Lett. 2010; 1:96–100. [PubMed: 24936230]
15. Zaretzki J, Rydberg P, Bergeron C, Bennett KP, Olsen L, Breneman CM. RS-Predictor models augmented with SMARTCyp reactivities: robust metabolic regioselectivity predictions for Nine CYP isozymes. J Chem Inf Model. 2012; 52:1637–1659. [PubMed: 22524152]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

16. Jones JP, Mysinger M, Korzekwa KR. Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction. Drug Metab Dispos. 2002; 30:7–12. [PubMed: 11744605]
17. Kim DN, Cho KH, Oh WS, Lee CJ, Lee SK, Jung J, No KT. EaMEAD: activation energy prediction of cytochrome P450 mediated metabolism with effective atomic descriptors. J Chem Inf Model. 2009; 49:1643–1654. [PubMed: 19545128]
18. Yu K, Geng X, Chen M, Zhang J, Wang B, Ilic K, Tong W. High daily dose and being a substrate of cytochrome P450 enzymes are two important predictors of drug-induced liver injury. Drug Metab Dispos. 2014; 42:744–750. [PubMed: 24464804]
19. Asturiol, D., Worth, A. The Use of Chemical Reactivity Assays in Toxicity Prediction. JRC Scientific and Technical Reports. 2011. http://publications.jrc.ec.europa.eu/repository/handle/JRC65567
20. Fujimoto K, Kishino H, Yamoto T, Manabe S, Sanbuissho A. In vitro cytotoxicity assay to evaluate the toxicity of an electrophilic reactive metabolite using glutathione-depleted rat primary cultured hepatocytes. Chem–Biol Interact. 2010; 188:404–411. [PubMed: 20846519]
21. James LP, Mayeux PR, Hinson JA. Acetaminophen-induced hepatotoxicity. Drug Metab Dispos. 2003; 31:1499–1506. [PubMed: 14625346]
22. Mitchell J, Jollow D, Potter W, Gillette J, Brodie B. Acetaminophen-induced hepatic necrosis. IV. Protective role of glutathione. J Pharmacol Exp Ther. 1973; 187:211–217. [PubMed: 4746329]
23. Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. Development of a peptide reactivity assay for screening contact allergens. Toxicol Sci. 2004; 81:332–343. [PubMed: 15254333]
24. Chan K, Jensen N, O'Brien PJ. Structure–activity relationships for thiol reactivity and rat or human hepatocyte toxicity induced by substituted *p*-benzoquinone compounds. J Appl Toxicol. 2008; 28:608–620. [PubMed: 17975849]
25. Stepan AF, Walker DP, Bauman J, Price DA, Baillie TA, Kalgutkar AS, Aleo MD. 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. 2011; 24:1345–1410. [PubMed: 21702456]
26. Schultz TW, Yarbrough JW, Hunter RS, Aptula AO. Verification of the structural alerts for Michael acceptors. Chem Res Toxicol. 2007; 20:1359–1363. [PubMed: 17672510]
27. Bultinck P, van Neck D, Acke G, Ayers PW. Influence of electron correlation and degeneracy on the Fukui matrix and extension of frontier molecular orbital theory to correlated quantum chemical methods. Phys Chem Chem Phys. 2012; 14:2408–2416. [PubMed: 22249745]
28. Fukui K, Yonezawa T, Shingu H. A molecular orbital theory of reactivity in aromatic hydrocarbons. J Chem Phys. 1952; 20:722–725.
29. Roy R, Krishnamurti S, Geerlings P, Pal S. Local softness and hardness based reactivity descriptors for predicting intra-and intermolecular reactivity sequences: carbonyl compounds. J Phys Chem A. 1998; 102:3746–3755.
30. Morell C, Grand A, Toro-Labbe A. New dual descriptor for chemical reactivity. J Phys Chem A. 2005; 109:205–212. [PubMed: 16839107]
31. Chattaraj PK, Maiti B, Sarkar U. Philicity: a unified treatment of chemical reactivity and selectivity. J Phys Chem A. 2003; 107:4973–4975.
32. Baldi, P., Brunak, S. Bioinformatics: The Machine Learning Approach. The MIT Press; Cambridge, MA: 2001.
33. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. J Cheminf. 2011; 3:33.
34. Rydberg P, Rostkowski M, Gloriam DE, Olsen L. The contribution of atom accessibility to site of metabolism models for cytochromes P450. Mol Pharmaceutics. 2013; 10:1216–1223.
35. Karelson M, Lobanov VS, Katritzky AR. Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev. 1996; 96:1027–1044. [PubMed: 11848779]
36. Cronin, MTD., Madden, JC., editors. Silico Toxicology: Principles and Applications. Royal Society of Chemistry; Cambridge: 2010.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

37. Schwöbel JA, Koleva YK, Enoch SJ, Bajot F, Hewitt M, Madden JC, Roberts DW, Schultz TW, Cronin MT. Measurement and estimation of electrophilic reactivity for predictive toxicology. Chem Rev. 2011; 111:2562–2596. [PubMed: 21401043]
38. Stewart JJ. MOPAC: a semiempirical molecular orbital program. J Comput-Aided Mol Des. 1990; 4:1–103. [PubMed: 2197373]
39. Hostaš J, Řezáć J, Hobza P. On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions. Chem Phys Lett. 2013; 568:161–166.
40. Nasr R, Swamidass S, Baldi P. Large scale study of multiple-molecule queries. J Cheminf. 2009; 1:7.
41. Schneider G, Tanrikulu Y, Schneider P. Self-organizing molecular fingerprints: a ligand-based view on drug-like chemical space and off-target prediction. Future Med Chem. 2009; 1:213–218. [PubMed: 21426077]
42. Zhu Q, Lajiness M, Ding Y, Wild D. WENDI: a tool for finding non-obvious relationships between compounds and biological properties, genes, diseases and scholarly publications. J Cheminf. 2010; 2:6.
43. Keiser M, Roth B, Armbruster B, Ernsberger P, Irwin J, Shoichet B. Relating protein pharmacology by ligand chemistry. Nat Biotechnol. 2007; 25:197–206. [PubMed: 17287757]
44. Glick M, Klon A, Acklin P, Davies J. Enrichment of extremely noisy high-throughput screening data using a naive Bayes classifier. J Biomol Screening. 2004; 9:32.
45. Crisman T, Jenkins J, Parker C, Hill W, Bender A, Deng Z, Nettles J, Davies J, Glick M. "Plate cherry picking": a novel semi-sequential screening paradigm for cheaper, faster, information-rich compound selection. J Biomol Screening. 2007; 12:320.
46. Albano E, Rundgren M, Harvison P, Nelson S, Moldeus P. Mechanisms of *N*-acetyl-*p*-benzoquinone imine cytotoxicity. Mol Pharmacol. 1985; 28:306–311. [PubMed: 4033631]
47. O'Boyle NM, Morley C, Hutchison GR. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. Chem Cent J. 2008; 2:1–5. [PubMed: 18234100]
48. Sorich MJ, Miners JO, McKinnon RA, Winkler DA, Burden FR, Smith PA. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. J Chem Inf Comput Sci. 2003; 43:2019–2024. [PubMed: 14632453]
49. Hunter A, Kennedy L, Henry J, Ferguson I. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. Comput Methods Programs Biomed. 2000; 62:11–19. [PubMed: 10699681]
50. Ziólkowski J. New relation between ionic radii, bond length, and bond strength. J Solid State Chem. 1985; 57:269–290.
51. Brown I, Shannon R. Empirical bond-strength-bond-length curves for oxides. Acta Crystallogr, Sect A: Cryst Phys, Diffr, Theor Gen Crystallogr. 1973; 29:266–282.
52. Thaens D, Heinzelmann D, Böhme A, Paschke A, Schüürmann G. Chemoassay screening of DNA-reactive mutagenicity with 4-(4-nitrobenzyl) pyridine-application to epoxides, oxetanes, and sulfur heterocycles. Chem Res Toxicol. 2012; 25:2092–2102. [PubMed: 22889134]
53. Kalgutkar AS, Gardner I, Obach RS, Shaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J, Harriman SP. A comprehensive listing of bioactivation pathways of organic functional groups. Curr Drug Metab. 2005; 6:161–225. [PubMed: 15975040]
54. Damsten MC, de Vlieger JS, Niessen WM, Irth H, Vermeulen NP, Commandeur JN. Trimethoprim: novel reactive intermediates and bioactivation pathways by cytochrome P450s. Chem Res Toxicol. 2008; 21:2181–2187. [PubMed: 18816075]
55. Scicinski J, Oronsky B, Taylor M, Luo G, Musick T, Marini J, Adams CM, Fitch WL. Preclinical evaluation of the metabolism and disposition of RRx-001, a novel investigative anticancer agent. Drug Metab Dispos. 2012; 40:1810–1816. [PubMed: 22699395]
56. Sun H, Yost GS. Metabolic activation of a novel 3-substituted indole-containing TNF-α inhibitor: dehydrogenation and inactivation of CYP3A4. Chem Res Toxicol. 2007; 21:374–385. [PubMed: 18095656]
57. Shetty HU, Zoghbi SS, Siméon FG, Liow JS, Brown AK, Kannan P, Innis RB, Pike VW. Radiodefluorination of 3-fluoro-5-(2-(2-[^{18}F](fluoromethyl)-thiazol-4-yl) ethynyl) benzonitrile

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

([^{18}F]SP203), a radioligand for imaging brain metabotropic glutamate subtype-5 receptors with positron emission tomography, occurs by glutathionylation in rat brain. J Pharmacol Exp Ther. 2008; 327:727–735. [PubMed: 18806125]

58. LeBlanc A, Sleno L. Atrazine metabolite screening in human microsomes: detection of novel reactive metabolites and glutathione adducts by LC-MS. Chem Res Toxicol. 2011; 24:329–339. [PubMed: 21361395]
59. Xu L, Woodward C, Khan S, Prakash C. In vitro metabolism of BIIB021, an inhibitor of heat shock protein 90, in liver microsomes and hepatocytes of rats, dogs, and humans and recombinant human cytochrome P450 isoforms. Drug Metab Dispos. 2012; 40:680–693. [PubMed: 22217465]
60. deBethizy JD, Rickert DE. Metabolism of nitrotoluenes by freshly isolated Fischer 344 rat hepatocytes. Drug Metab Dispos. 1984; 12:45. [PubMed: 6141911]
61. Frankmoelle WP, Medina JC, Shan B, Narbut MR, Beckmann H. Glutathione S-transferase metabolism of the antineoplastic pentafluorophenylsulfonamide in tissue culture and mice. Drug Metab Dispos. 2000; 28:951–958. [PubMed: 10901706]
62. Kalgutkar AS, Driscoll J, Zhao SX, Walker GS, Shepard RM, Soglia JR, Atherton J, Yu L, Multib AE, Munchhof MJ, Reiter LA, Jones CS, Doty JL, Trevena KA, Shaffer CL, Ripp SL. A rational chemical intervention strategy to circumvent bioactivation liabilities associated with a nonpeptidyl thrombopoietin receptor agonist containing a 2-amino-4-arylthiazole motif. Chem Res Toxicol. 2007; 20:1954–1965. [PubMed: 17935300]
63. OEDepict. OpenEye Scientific Software, Inc; Santa Fe, NM: 2014. version 1.7.4.5www.eyesopen.com

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

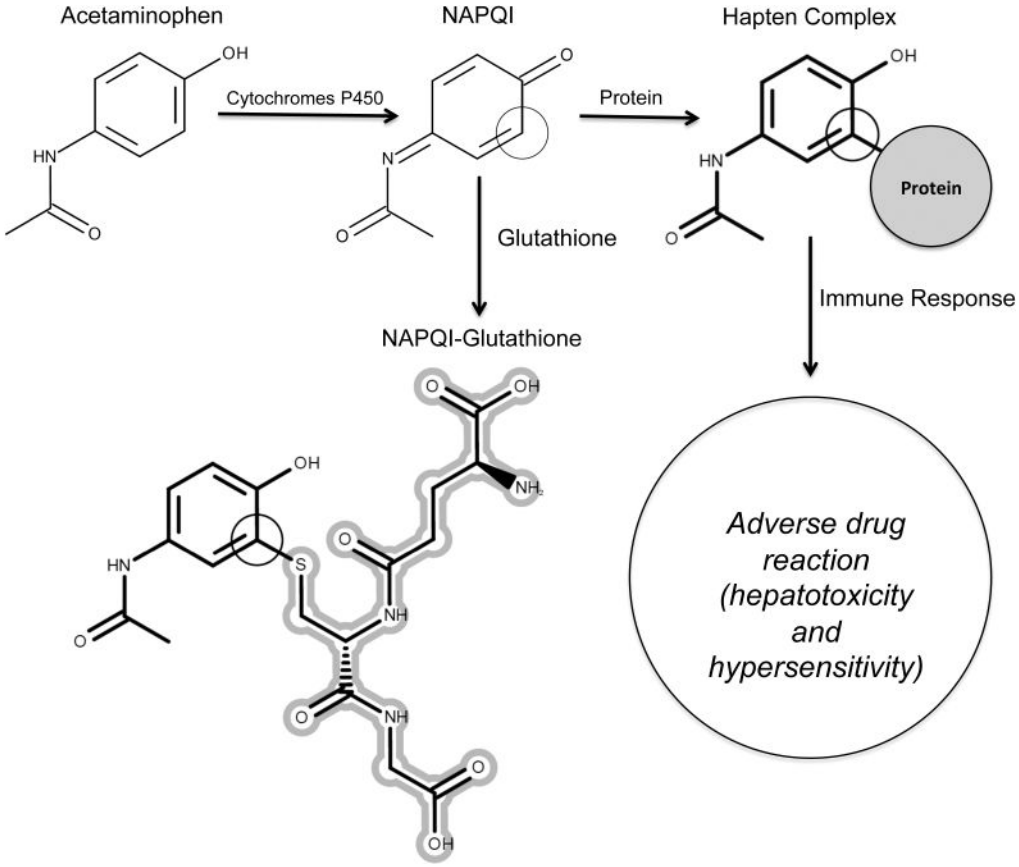


Figure 1.

Adverse drug reactions are often cased by reactive metabolites. Acetaminophen is metabolized by cytochromes P450 to *N*-acetyl-*p*-benzoquinone imine (NAPQI). NAPQI is electrophilically reactive and covalently binds to nucleophilic sites within proteins, eliciting an immune response. Glutathione (GSH and outlined in gray) protects the body from this adverse drug reaction by scavenging electrophiles like NAPQI, to which GSH binds at its site of reactivity (circled atom). Thus, a site of GSH conjugation is a likely site of protein conjugation, and identifying these sites of reactivity offers information about the mechanism of metabolite toxicity. Several methods have been published that can predict how P450s metabolize molecules. This study, however, focuses on modeling the reactivity of molecules with GSH but not the metabolism of molecules into reactive species.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

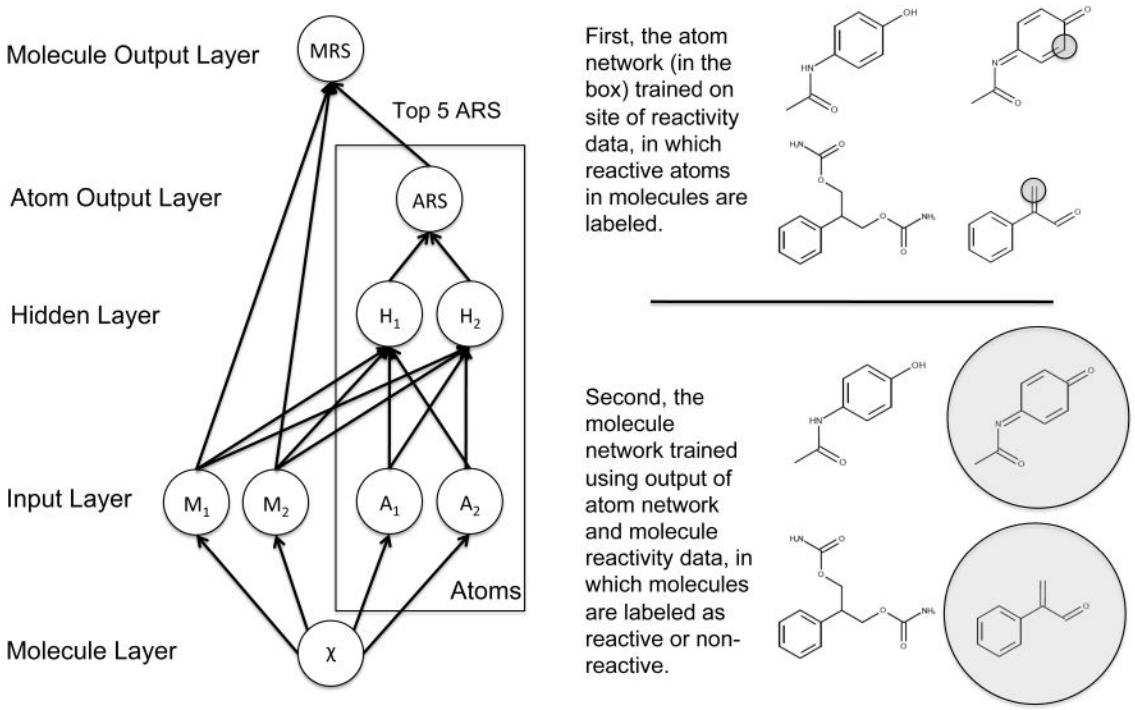


Figure 2.

The structure of the reactivity model. This diagram shows how information flows through the model, which is composed of one input layer, one hidden layer, and two output layers. This model computed a prediction for each test molecule and atom in the test molecule. Atom reactivity scores (ARS) were computed with a neural network, with one output node, one hidden layer (with ten units), and one input layer. From the 3D structure of input molecule χ, 30 molecule-level and 86 atom-level descriptors were calculated (two input layer nodes for each category are displayed). The diagram shows only two hidden nodes, two molecule input nodes, and two atom input nodes for conciseness. The actual model had several additional nodes in each input and hidden layer. For each atom within χ, all 116 descriptors were fed into the 10 hidden layer nodes (two are displayed), which generated an ARS. The molecule reactivity score (MRS) of χ was computed from the top five ARS corresponding to the scores of the five atoms predicted to be the most reactive within a molecule and all molecule-level descriptors. The molecules on the top right illustrate atom-level data, with sites of reactivity circled, and the molecules on the bottom right illustrate molecule-level data, with reactive molecules circled.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

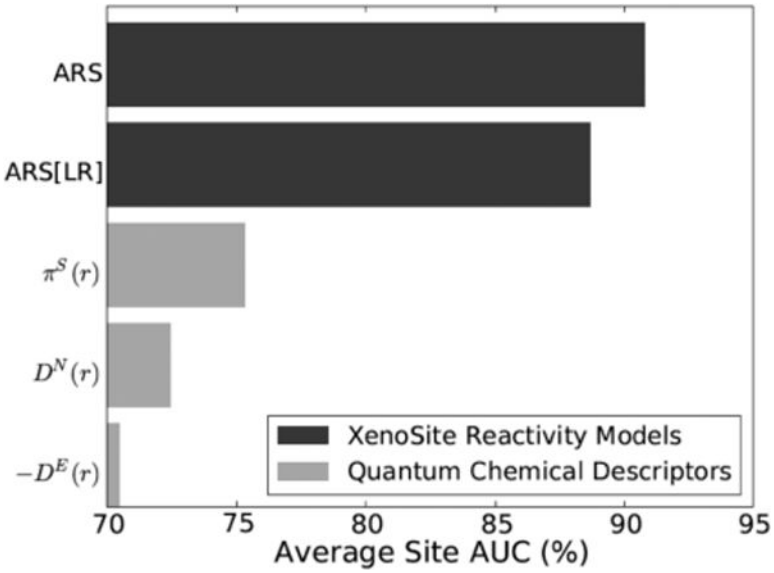


Figure 3.

Atom reactivity scores accurately identified sites of reactivity. For each prediction method, average site AUC was computed for 1213 reactive molecules, with their sites of conjugation to glutathione labeled. This metric reflected how often reactive atoms were ranked above unreactive atoms within reactive molecules. The cross-validated atom reactivity scores (ARS), generated by a neural network with 10 hidden nodes trained by gradient descent on the cross-entropy error, outperformed the cross-validated predictions of a logistic regressor (ARS[LR]). The performances of selected atom-level descriptors were also evaluated. The accuracy of the reactivity model exceeded that of $\pi^S(r)$, $D^N(r)$, and $D^E(r)$, three commonly used reactivity indices.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

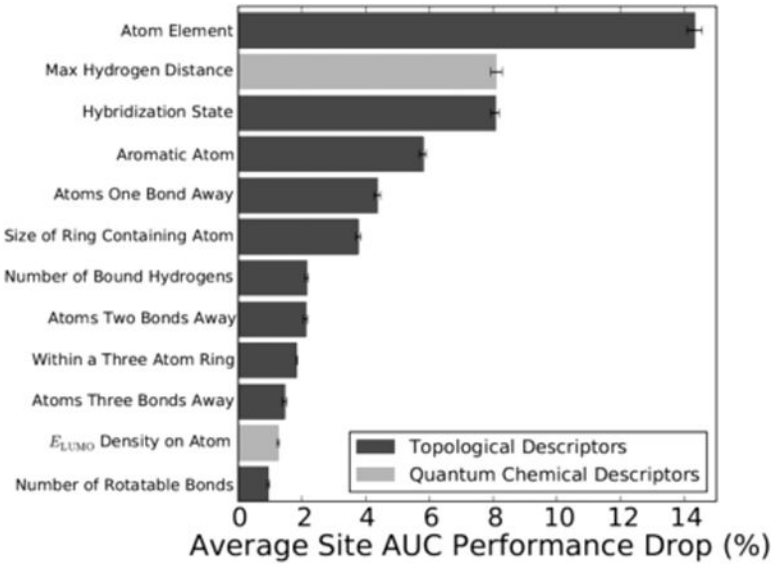


Figure 4.

Importance of specific descriptors to the atom reactivity model. A permutation sensitivity analysis quantified the importance of descriptors for the final trained atom reactivity model. This listing indicates the 12 most important descriptors in decreasing order of importance from top to bottom. The graph shows the model performance drop associated after permuting the associated descriptor values, averaging over 10 iterations. All top descriptors with the exception of two were topological; the remainder were derived from a quantum simulation of the molecule structure.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

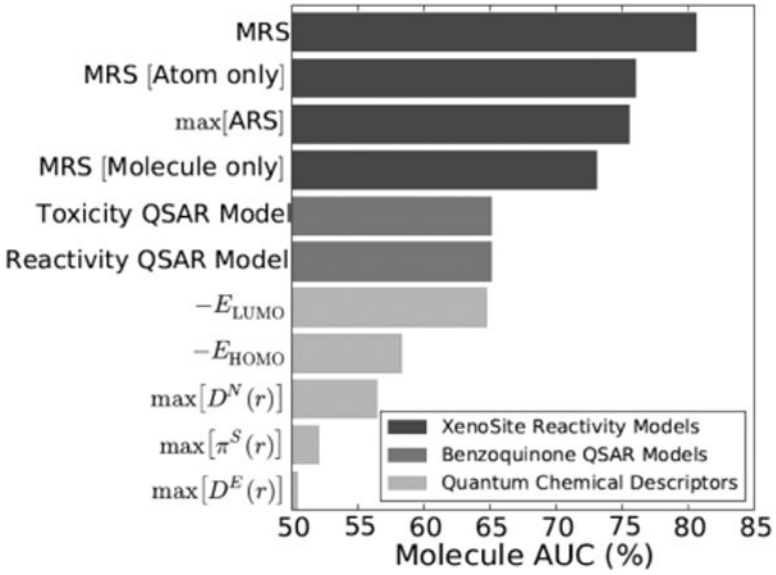


Figure 5.

The reactivity model accurately identified reactive molecules. Several prediction methods were compared based on their ability to identify reactive molecules. The data set included 1484 molecules, 1213 of which are reactive with glutathione and 271 of which are not reactive but are structurally similar to reactive molecules. Performance was measured by computing the area under the ROC curve (molecule AUC). The best performing approach computed a MRS using a logistic regressor based on molecule-level descriptors and the top five ARS scores associated with each atom of the molecule. Control models demonstrated lower accuracy using only atom-level information (MRS[Atom only] and max[ARS]) or molecule-level information (MRS[Molecule only]). Similarly, neither the two published QSAR models (eqs 1 and 2), several molecule descriptors, nor atom descriptors yielded models as accurate as our reactivity model.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

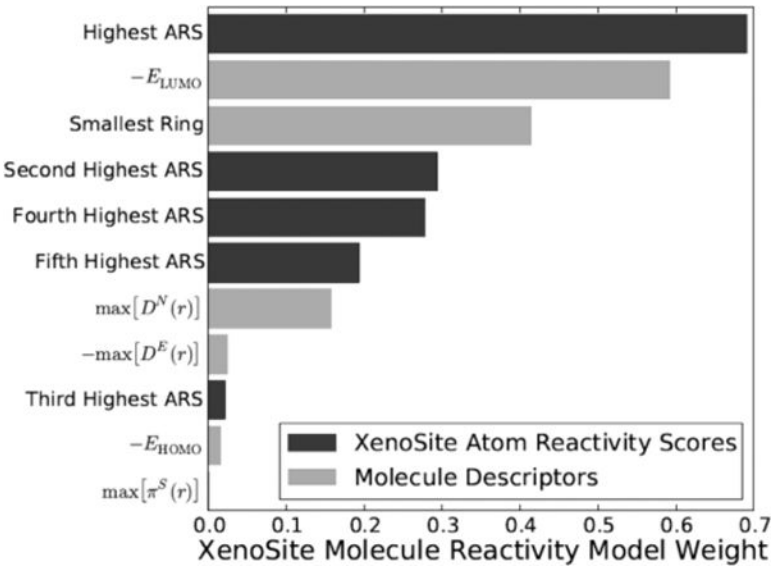


Figure 6.

Importance of descriptors to molecular reactivity score. The weights of the final model nodes revealed the relative contribution of individual descriptors to model performance. The descriptors were normalized before training so that the magnitude of the weights directly measured the importance of each descriptor. The values for the five ARS scores were included, as well as those of selected reactivity indices. As we would hope, a significant weight was placed on all of the ARS descriptors. Moreover, the qualitative contributions of specific quantum descriptors were within expectations based on frontier orbital theory. This analysis increased confidence in the ability of the model to sensibly generalize toward external data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

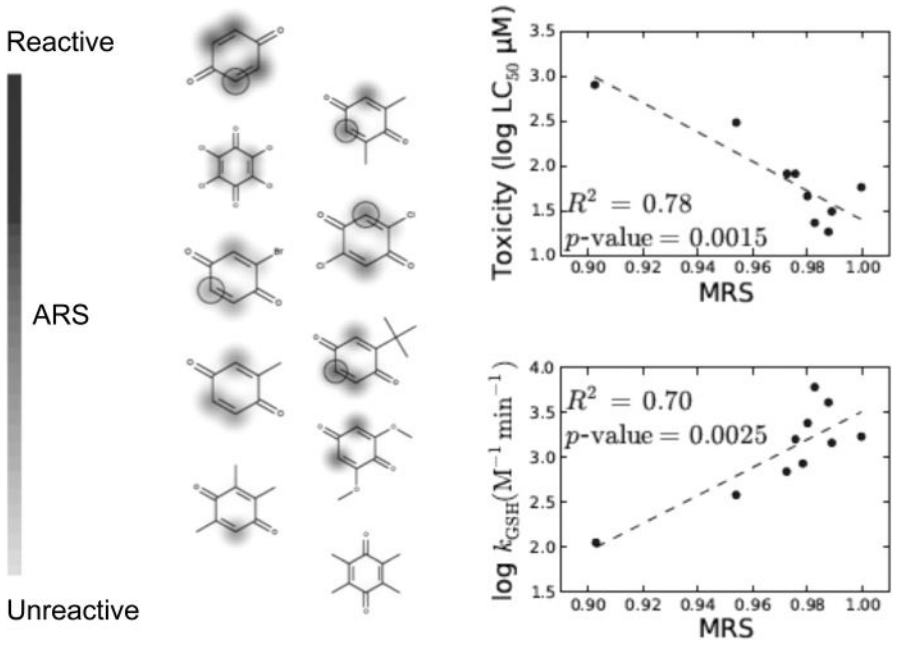


Figure 7.

Molecule reactivity scores correlated with glutathione reactivity and toxicity of substituted quinones. The model molecule reactivity scores (MRS) correlated closely with hepatocyte toxicity (LC_{50}, top graph) and the rate of reactivity with GSH (k_{GSH}, bottom graph) of 10 substituted *p*-benzoquinones.[24] The left panel illustrates all 10 test molecules and sorts them by MRS computed by a model trained without using these molecules. For each molecule, the shading intensity represents atom reactivity scores (ARS), which range from 0 to 0.41. Circled atoms are labeled as reactive in our training data set.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

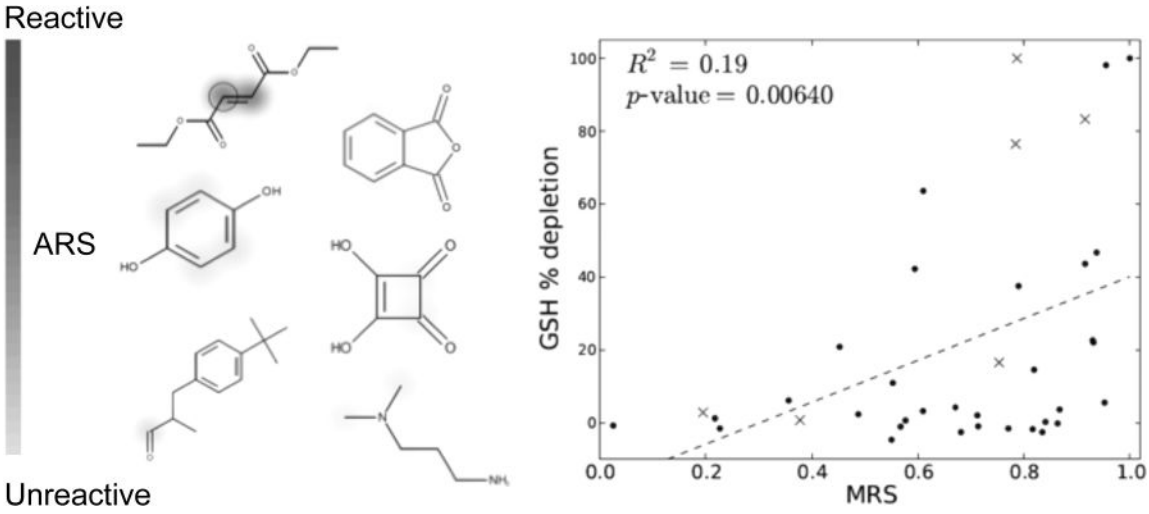


Figure 8.

Reactivity scores correlated with the nucleophile reactivity of structurally diverse contact allergens. Model performance was assessed using an external data set with 38 molecules.[23] The *y* axis is the percent depletion of GSH after 15 min incubation with each molecule. The *x* axis indicates molecule reactivity scores (MRS). For test molecules present in the training data set, the appropriate cross-validated predictions were extracted. The significance of Pearson correlation is reflected by the *p*-values. To the left, six example molecules are visualized with scaled ARS (which range from 0 to 0.43) and are sorted by MRS, which correspond to the data points marked with an × in the right panel plot. The six corresponding ×'s are in the same horizontal order as that of the visualized molecules. MRS significantly correlated with GSH reactivity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

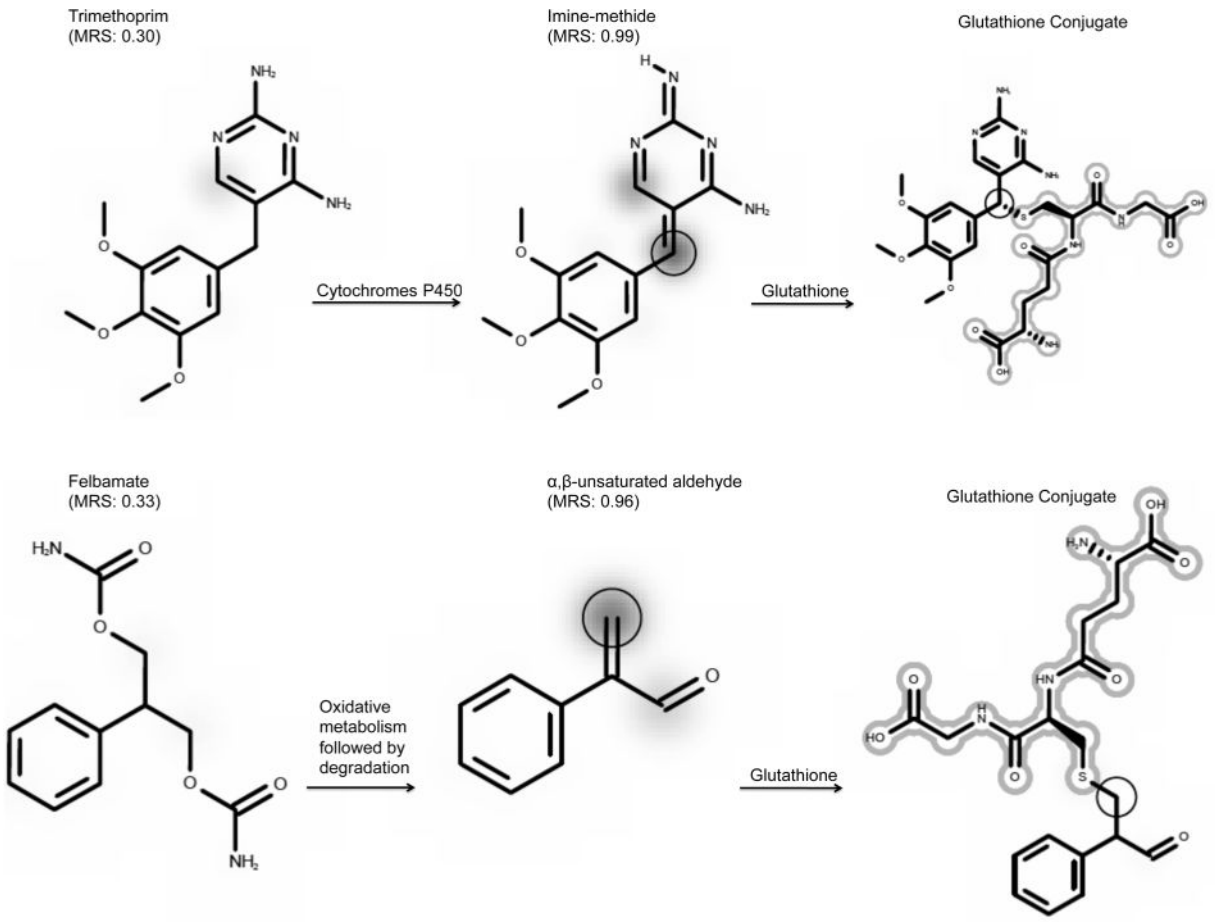


Figure 9.

Molecule reactivity scores distinguished drugs and their reactive metabolites. For each molecule, the shading represents atom reactivity scores (ARS), which ranged from 0 to 0.768. The structures of trimethoprim and felbamate are shown alongside their reactive metabolites and subsequent GSH conjugates. Circled atoms are labeled as reactive in our training data set. For molecules present in the training set, cross-validated predictions are displayed. In these cases, the predictions are obvious to an organic chemist, but they illustrate key points of the method's behavior. First, it can distinguish accurately between very similar molecules that are reactive and nonreactive. Second, in the current form, it cannot predict how metabolism gives rise to reactive species.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

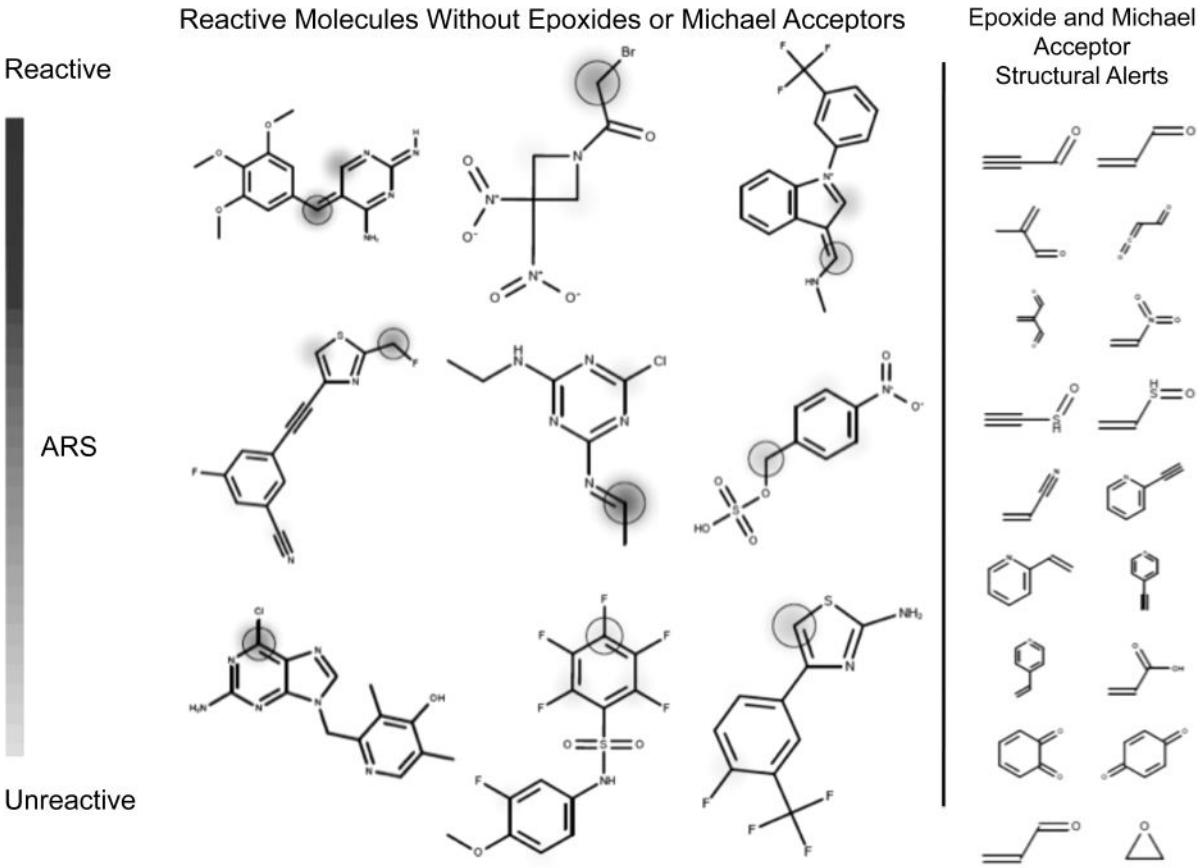


Figure 10.

Atom reactivity scores accurately identified nonobvious sites of reactivity The right panel displays 18 structural motifs known to be reactive.[26] The left panel displays reactive molecules drawn from our training data that do not contain these specific substructures.[25,55–62] The molecules are sorted by molecule reactivity scores, which range from 0.78 to 0.99. The shading intensity represents scaled atom reactivity scores (ARS), which range from 0 to 0.78. Our model accurately identified reactive molecules that do not match commonly used structural alerts.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Condensed List of Quantum Chemical Reactivity Descriptors[a]

atom-level descriptors	
$D^N(r)$	nucleophilic delocalizability
$D^E(r)$	electrophilic delocalizability
$\pi^S(r)$	self-polarizability

molecule-level descriptors	
E_{LUMO}	energy of the lowest unoccupied molecular orbital
E_{HOMO}	energy of the highest occupied molecular orbital
$\max[D^N(r)]$	maximum atom nucleophilic delocalizability
$\max[D^E(r)]$	maximum atom electrophilic delocalizability
$\max[\pi^S(r)]$	maximum atom self-polarizability

[a]Descriptors were generated at both the atom and molecule levels. A full list of descriptors is available in the Supplementary Information.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Reactivity Scores Correlated More Closely with Trapping Agent Depletion than QSAR Models[a]

peptide	QSAR reactivity model	MRS
GSH	0.30 (*p*-value = 0.0709)	**0.43** (*p*-value = 0.0064)
cysteine	**0.38** (*p*-value = 0.0210)	**0.56** (*p*-value = 0.0003)
lysine	0.26 (*p*-value = 0.1310)	**0.41** (*p*-value = 0.0126)
histidine	0.14 (*p*-value = 0.4295)	0.24 (*p*-value = 0.1504)

[a]The correlations between peptide reactivity and QSAR predictions (eq 2) are molecule reactivity scores (MRS) were calculated for 38 structurally diverse contact allergens.[23] Reactivity was measured by peptide depletion assays, with incubation times of 15 min for GSH or 24 h for peptides containing cysteine, lysine, or histidine. Across the results of all four assays, reactivity was more significantly correlated with MRS than by the QSAR model. Statistically significant correlations are in bold.