

Washington University in St. Louis

Washington University Open Scholarship

Arts & Sciences Electronic Theses and
Dissertations

Arts & Sciences


Winter 12-15-2018

Modeling the Bioactivation and Subsequent Reactivity of Drugs

Tyler Brian Hughes

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

 Part of the [Other Chemistry Commons](#)

Recommended Citation

Hughes, Tyler Brian, "Modeling the Bioactivation and Subsequent Reactivity of Drugs" (2018). *Arts & Sciences Electronic Theses and Dissertations*. 1687.
https://openscholarship.wustl.edu/art_sci_etds/1687

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Computational and Systems Biology

Dissertation Examination Committee:

Jeremy Buhler, Chair

Mark Anastasio

Gregory Bowman

Gary Stormo

S. Joshua Swamidass

Modeling the Bioactivation and Subsequent Reactivity of Drugs

by

Tyler Brian Hughes

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2018
St. Louis, Missouri

© 2018, Tyler Brian Hughes

Contents

List of Figures	vi
List of Tables	ix
List of Abbreviations	x
Acknowledgements	xi
Abstract	xiii
1 Introduction	1
1.1 Idiosyncratic Adverse Drug Reactions	1
1.2 Metabolism and Bioactivation	2
1.3 Computational Methods of Predicting Reactivity and Metabolism	6
1.4 Overview of Chapters	9
2 Site of Reactivity Models Predict Molecular Reactivity of Diverse Chemicals with Glutathione	13
2.1 Introduction	13
2.2 Materials and Methods	15
2.2.1 Glutathione Site of Reactivity Training Data	15
2.2.2 External Reactivity Data	17
2.2.3 Comparison QSAR Models	17
2.2.4 Descriptors	18
2.2.5 Combined Atom- and Molecule-level Reactivity Model	19
2.3 Results and Discussion	20
2.3.1 Accuracy in Identifying Sites of Reactivity	22
2.3.2 Descriptors Driving Atom Reactivity Performance	23
2.3.3 Accuracy in Identifying Reactive Molecules	27
2.3.4 Descriptors Driving Molecule Reactivity Performance	29
2.3.5 Performance on External data sets	32
2.3.6 Performance on Drug-Metabolite Pairs	36
2.4 Non-Obvious Predictions	38

2.5	Conclusion	40
3	Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network	41
3.1	Introduction	41
3.2	Results and Discussion	43
3.2.1	Model Optimization	45
3.2.2	Accuracy at Predicting Sites of Reactivity	46
3.2.3	Accuracy at Identifying Reactive Molecules	49
3.2.4	Molecules Missed by Standard Screening Assays	51
3.3	Model Limitations	52
3.4	Conclusion	53
3.5	Methods	54
3.5.1	Site of Reactivity Training Data	54
3.5.2	External Site of Reactivity Test Data	55
3.5.3	Quantum Chemical Reactivity Indices	55
3.5.4	Topological Descriptors	56
3.5.5	Combined Atom- and Molecule-Level Reactivity Model	57
4	Modeling Epoxidation of Drug-like molecules with a Deep Machine Learning Network	60
4.1	Introduction	60
4.2	Methods	62
4.2.1	Epoxidation Training Data	62
4.2.2	Hydroxylation Negative Control Data	64
4.2.3	Descriptors	65
4.2.4	Combined Atom- and Molecule-Level Epoxidation Model	67
4.3	Results and Discussion	68
4.3.1	Accuracy in Identifying Sites of Epoxidation	70
4.3.2	Descriptors Driving Bond Epoxidation Score Performance	71
4.3.3	Accuracy in Identifying Aromatic and Double Bond Sites of Epoxidation	74
4.3.4	Distinguishing Epoxidation from Hydroxylation	76
4.3.5	π_p Orbital Occupancy and Epoxidation	78
4.3.6	Accuracy at Identifying Molecules that Undergo Epoxidation	78
4.3.7	Case Studies	81
4.4	Conclusion	86
5	Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism	87
5.1	Introduction	87
5.2	Materials and Methods	91
5.2.1	Quinone Formation Training Data	91

Contents

5.2.2	External Non-Quinone-Forming Test Sets	93
5.2.3	Topological Descriptors	94
5.2.4	Combined Pair- and Molecule-level Quinone Formation Model	94
5.2.5	Quantum Chemical Descriptors	95
5.3	Results and Discussion	97
5.3.1	Accuracy at Predicting Quinone-Forming Pairs	98
5.3.2	Accuracy at Identifying Quinone-Forming Molecules	103
5.3.3	Comparison to Structural Alerts	107
5.3.4	Case Studies	115
5.4	Model Limitations	121
5.5	Conclusions	122
6	The Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites	123
6.1	Introduction	123
6.2	Results and Discussion	125
6.2.1	Accurate Predictions of Metabolite Structures	127
6.2.2	Inferring Intermediate Metabolite Structures	129
6.2.3	Automatic Detection of Mislabeled Sites.	134
6.2.4	Accurate Generation of Quinone Structures.	135
6.2.5	Hypothesized Reactive Metabolites of Toxic Drugs.	135
6.2.6	Limitations	138
6.3	Conclusion	140
6.4	Methods	141
6.4.1	Phase 1 Metabolism Data	141
6.4.2	The Metabolic Forest	142
6.4.3	SMARTS Rules	142
6.4.4	Resonance Pair Rules	146
6.4.5	Resonance Structure Rules	148
6.4.6	Algorithm for Predicting Depth Two or Three Metabolite Structures	150
6.4.7	Comparison to RD-Metabolizer	150
7	Modeling the Bioactivation and Subsequent Reactivity of Drugs	152
7.1	Introduction	152
7.2	Methods	155
7.2.1	Bioactivation Training Data	155
7.2.2	External Non-Bioactivated Test Sets	156
7.2.3	DrugBank	156
7.2.4	Bioactivation Descriptors	157
7.2.5	Combined Path- and Molecule-level Bioactivation Model	157

Contents

7.3	Results and Discussion	162
7.3.1	Bioactivation Pathway Prediction Accuracy	162
7.3.2	Accuracy at Predicting Molecule Bioactivation	167
7.3.3	Toxicity Drivers Hypothesis Generation	171
7.4	Model Limitations	174
7.5	Conclusion	174
	Bibliography	175

List of Figures

1.1	Bioactivation occurs when metabolism produces reactive metabolites that covalently bind to biological macromolecules.	3
1.2	Reactive Metabolites can Induce IADRs.	5
1.3	Structural alerts are not specific enough.	7
1.4	Neural networks are highly customizable	10
1.5	Summary of Chapters.	12
2.1	Adverse drug reactions are often caused by reactive metabolites.	16
2.2	The structure of the reactivity model.	21
2.3	Atom reactivity scores accurately identified sites of reactivity.	24
2.4	The importance of specific descriptors to the atom reactivity model.	26
2.5	The reactivity model accurately identified reactive molecules.	30
2.6	Importance of descriptors to molecular reactivity score.	31
2.7	Molecule reactivity scores correlated with glutathione reactivity and toxicity of substituted quinones.	34
2.8	Reactivity scores correlated with the nucleophile reactivity of structurally diverse contact allergens.	35
2.9	Molecule reactivity scores distinguished drugs and their reactive metabolites.	37
2.10	Atom reactivity scores accurately identified non-obvious sites of reactivity.	39
3.1	Examples of the four types of sites of electrophilic reactivity modeled.	44
3.2	Atom reactivity scores (ARS) accurately identified sites of reactivity (SORs).	48
3.3	Molecule reactivity scores (MRS) accurately identified reactive molecules.	50
3.4	Molecule reactivity scores (MRS) predict some molecules significantly more reactive with biological macromolecules than nucleophilic traps used in standard screening assays.	52
3.5	The structure of the XenoSite reactivity model.	59
4.1	Adverse drug reactions are often caused by reactive metabolites.	63
4.2	In the database, each epoxidation reaction acting on a site of epoxidation (abbreviated SOE and circled) forms an epoxide, dihydrodiol, or a conjugate adjacent to a hydroxylation.	66
4.3	The structure of the epoxidation model.	69

List of Figures

4.4	Bond epoxidation scores accurately (BES) identify sites of epoxidation (SOEs).	72
4.5	The importance of specific descriptors to the bond epoxidation model.	74
4.6	Bond epoxidation scores (BES) accurately identified both aromatic and double bond sites of epoxidation.	75
4.7	Bond epoxidation scores (BES) distinguish sites of epoxidation (SOEs) from sites of hydroxylation (SOHs).	77
4.8	Bond epoxidation scores (BES) represent a well-scaled probability that a site will be epoxidized.	79
4.9	Molecule epoxidation scores accurately identify molecules subject to epoxidation.	82
4.10	MES offers a well-scaled probabilistic prediction of molecule epoxidation. . . .	83
4.11	The epoxidation model recognizes sites of epoxidation within drugs that can be modified to reduce toxicity	85
5.1	Examples of one- and two-step quinone formation: the formation of a quinone by metabolic oxidation.	89
5.2	Both quinone-forming and non-quinone-forming molecules can undergo comparable initial metabolic events.	90
5.3	The structure of the quinone formation model.	92
5.4	The pair-level model, computing pair quinone formation scores (PQS), accurately predicted quinone-forming pairs (QP).	100
5.5	The pair-level neural network computes the best scaled prediction of quinone-forming pairs, and corresponds closely with a probability.	102
5.6	Molecule quinone formation scores (MQS) accurately identified quinone-forming molecules.	104
5.7	Example pairs of molecules that do and do not form quinones are visualized. . .	105
5.8	The neural network produces a well-scaled probabilistic prediction of molecule quinone formation.	106
5.9	Pair-level scores predicted whether several structural alerts form quinones. . . .	109
5.10	AUC curves for all structural alerts scored by PQS.	110
5.11	Pair-level scores predicted whether catechols form quinones.	111
5.12	Pair-level scores predicted whether nitrogen-bound carbons <i>para</i> to a carbon with a hydrogen form quinones.	113
5.13	Pair-level scores predicted whether a quinone will form <i>para</i> to a phenol. . . .	114
5.14	Pair-level scores predicted whether quinones form at motifs that do not match structural alerts.	115
5.15	Molecule quinone formation scores accurately identified quinone-forming molecules.	117
5.16	Examples of drugs associated with idiosyncratic toxicity that received high molecule quinone formation scores (MQS), but are not known to form quinones in our data set.	118

List of Figures

5.17	The model reflects the impact of structural modification to prevent quinone formation.	120
6.1	The metabolic forest explores metabolic pathways by sequentially applying metabolism rules.	126
6.2	The metabolic forest took as input the reactant and product structures of AMD records.	131
6.3	Examples of successes and failures records.	132
6.4	Missing intermediate structures can be predicted from those records for which a depth two or depth three search found a metabolic pathway not found at depth one.	133
6.5	Across 20736 Phase 1 reactions, the depth one search found a reaction connecting reactant and product in 79.42% of cases, whereas the manually annotated sites only had an accuracy of 78.36%.	134
6.6	The metabolic forest accurately produced quinone structures.	136
6.7	Previously developed models of metabolism and reactivity are connected by the metabolic forest to generated bioactivation hypotheses for several IADR-associated drugs, whose IADR-drivers were never elucidated.	139
6.8	Raloxifene illustrates the quinone prediction method.	149
7.1	This study modeled four common pathways of bioactivation.	154
7.2	Bioactivation descriptors were computed using site-level metabolism predictions, metabolite structure predictions, and reactivity predictions.	158
7.3	The bioactivation model structure.	160
7.4	Example PBS visualization.	161
7.5	The model produced accurate cross-validated pathway bioactivation predictions.	163
7.6	Of all methods, the neural network computed the best well-scaled bioactivation pathways predictions	166
7.7	Examples of accurate accurate cross-validated bioactivation pathway predictions produced by the neural network.	167
7.8	Molecule bioactivation scores (MBS) accurately identified bioactivated molecules.	169
7.9	The probabilistic OR function produced the best-scaled MBS.	170
7.10	The ROC of the cross-validated molecule bioactivation scores was used to define to a score cutoff, which can be used to binarize predictions.	172
7.11	Examples of withdrawn drugs with high bioactivation predictions.	173

List of Tables

2.1	Quantum chemical reactivity descriptors.	19
2.2	Reactivity scores correlated more closely with trapping agent depletion than QSAR models.	35
3.1	Quantum Chemical Reactivity Descriptors.	56
4.1	Several bond descriptors were developed for the epoxidation model	67
5.3	Size of Rings with Atoms of Quinone Formation	101
5.4	Quinone Structural Alert Frequencies	108
5.5	Pair-level scores predicted whether structural alerts form quinones.	109
6.1	The performance of several methods at reproducing the structure(s) linking AMD record reactants and products.	128
6.2	The reaction rules and the number of records where that rule was used in the pathway between the reported substrate and metabolite structures.	130
6.3	The 24 rules used by the metabolic forest, their associated rule set, and their type.	143
6.4	Reaction SMARTS used by the metabolite prediction algorithm.	144
6.5	The transforms used by the resonance pair rules.	149
7.1	Bioactivation descriptors used to predict bioactivation pathways..	157
7.2	Molecule descriptors used to predict bioactivation pathways	159

List of Abbreviations

$\pi S(r)$	self-polarizability
AMD	accelrys metabolite database
AQS	atom quinone formation scores
ARS	atom reactivity scores
AUC	area under the receiver operating characteristic curve
DE(r)	electrophilic delocalizability
DILI	drug-induced liver injury
DN(r)	nucleophilic de-localizability
EHOMO	energy of the highest occupied molecular orbital
ELUMO	energy of the lowest unoccupied molecular orbital
GSH	glutathione
HOMO	highest occupied molecular orbital
IADRs	idiosyncratic adverse drug reactions
MBS	molecule bioactivation score
MQS	molecule quinone formation scores
MRS	molecule reactivity scores
NAPQI	n-acetyl-p-benzoquinone imine
PBS	pathway bioactivation score
PQS	pair quinone formation scores
QPs	quinone-forming pairs
QSAR	quantitative structural- activity relationship
ROC	receiver operating characteristic
SOMs	site of metabolism
SOR	site of reactivity

Acknowledgements

I was exceptionally fortunate to have done my Ph.D. training at Wash U., and fear that it may be impossible to properly thank all those that supported me, but I will do my best. I am forever indebted to Josh, a truly exceptional advisor. Thank you for your unfailing generosity, your trust, and for constantly pushing me to think deeper. I will forever be deeply inspired by your creativity, boundless work ethic, and your depth and breath of expertise. Lena Dang, I had so much fun working with you, and your friendship was one the brightest highlights of my time at Wash U. Your rarefied combination of extraordinary intelligence and serenity exceeds my own on both counts, but I learned so much by knowing you and watching you navigate so many challenges with incredible aplomb. Matt Matlock, thank you for so freely sharing your coding expertise and amazing suite of tools. I feel confident that many of the concepts you showed me will stay with me for my entire career.

Jed Zaretski, thank you for all your help when I was getting up to speed and knew next to nothing. Dan Lander, Mick Ward, Naomi Wilson, Justin Porter, Sarah Pyfrom, and Ayush Kumar, thank you for your assistance and warmth, and I wish I had gotten to know you all better. Jackie Payton, thank you for giving me the warmest welcome to Wash U. that I could ever have imagined. Jeremy Buhler, Gary Stormo, Mark Anastasio, Gregory Bowman, Cynthia Lo, and Kristen Naegle, thank you for serving on my committees and your guidance. Grover Miller, thank you for your sharing your boundless metabolism expertise, and helping me become a better writer. To anyone who I omitted, please accept my preemptive apologies and gratitude. And, Todd Druley, thank you for giving me the call that changed my life: letting me know I had been accepted to Wash U.

Much of my research was supported by grants from National Library Of Medicine of the National Institutes of Health, award numbers R01LM012222 and R01LM012482. I am also

List of Abbreviations

grateful for the computational facilities of the Washington University Center for High Performance Computing, which were partially funded by National Institutes of Health grants numbers 1S10RR022984-01A1 and 1S10OD018091-01. I also thank Washington University Center for Biological Systems Engineering and the Department of Immunology and Pathology at the Washington University School of Medicine a for their generous support of my work. The developers of the open-source chemoinformatics tools Open Babel and RDKit also have my gratitude.

Tyler Hughes

Washington University in St. Lous

August 2018

*To my Alpha Centauri, my wife, who I strive to be worthy of every day.
To my launchpad and ignition, my mother and father, who I can never repay.
To my fellow pioneer, my little sister, who will always have a place to stay.*

ABSTRACT OF THE DISSERTATION
Modeling the Bioactivation and Subsequent Reactivity of Drugs
by
Tyler Brian Hughes
Doctor of Philosophy in Biology and Biomedical Sciences
Computational and Systems Biology
Washington University in St. Louis, 2018
S. Joshua Swamidass, Advisor
Jeremy Buhler, Chair

Metabolism can convert drugs to harmful reactive metabolites that conjugate to DNA and off-target proteins. Reactive metabolites are a significant driver of both drug candidate attrition and withdrawal from the market of already approved drugs. Unfortunately, reactive metabolites are difficult to study *in vivo*, because they are transitory and generally do not circulate. Instead, this work computationally models both metabolism and reactivity. Using deep learning, predictive models were developed for the metabolic formation of quinones and epoxides, which together account for about half of known reactive metabolites. Additionally, an accurate model of DNA and protein reactivity was constructed, which predicts how likely a molecule is to be reactive, and therefore potentially toxic. To connect the metabolism and reactivity models, a system was developed for predicting the exact structures of quinones and epoxides. Finally, using the metabolite structure predictor as a stepping-stone, the quinone formation and epoxidation models were connected to the reactivity model to build an integrated bioactivation model of metabolism and reactivity.

Chapter 1

Introduction

1.1 Idiosyncratic Adverse Drug Reactions

Adverse drug reactions (ADRs) are a major current health problem. Various studies in different populations implicated ADRs in 6.5%^[1], 8.1%^[2], 8.7%^[3], and 12.8%^[4] of hospital admissions. Furthermore, toxicity problems are a major driver of drug candidate attrition.^[5, 6, 7, 8, 9, 10] Most ADRs can be directly traced to a drug's pharmacological action.^[11] However, some ADRs are the especially problematic idiosyncratic adverse drug reactions (IADRs), which are not readily traceable to a drug's pharmacology and pharmacokinetics, and can unexpectedly strike with severe symptoms in almost any organ.^[12]

IADRs are infrequent, occurring in 1 in 10000 to 1 in 100000 patients.^[13] Consequently, IADRs often do not manifest even in the largest phase of preclinical trials—phase III—which generally have a sample size of only around 3000^[14] patients. However, after drugs are approved and released to the market, previously undetected IADR drivers can become rapidly apparent, as the low incidence rate of IADRs is overcome by the size of the general patient population.^[15] Despite being rare, IADRs cause significant mortality and morbidity because their effects are often severe including blood disorders (such as hemolytic anemia, aplastic anemia, and agranulocytosis), skin diseases (toxic epidermal necrolysis and Stevens-Johnson syndrome), anaphylaxis, and, most commonly, liver disorders. IADRs in the form of drug-induced liver injury (DILI) are one of the most common reasons for drug withdrawal from the market,

and have been associated with over 1000 drugs.[16, 17, 18] IADRs necessitate around 15% of liver transplants in the United States, as well as 50% of acute liver failure cases.[19] Not only are IADRs a major cause of patient suffering, but also are a major drain on pharmaceutical companies. It takes about 1.9 billion dollars for a company to get a new drug approved, but if it becomes evident that the drug causes IADRs, it generally must be withdrawn from the market or labeled by the FDA with a black-box warning, severely curtailing its use and profitability.[20, 21]

Unfortunately, IADRs are challenging to reproduce and study *in vivo*, and almost all animal models have failed.[9, 22, 12, 23, 24] However, there is significant evidence that suggests many IADRs have a common mechanism: bioactivation.

1.2 Metabolism and Bioactivation

The body is constantly exposed to foreign chemical entities—xenobiotics—such as dietary toxins or environmental pollutants.[25] Even seemingly harmless xenobiotics can become harmful if permitted to accumulate to high enough levels.[26] Because xenobiotics are frequently lipophilic, they can traverse cell membranes and get stuck within this hydrophobic barrier.[27]

In response to this threat, organisms evolved xenobiotic-clearing defense mechanisms: metabolism. Generally, metabolism is beneficial, for it serves the critical function of making xenobiotics more hydrophilic and therefore easier to excrete. Drugs are also xenobiotics, and are subject to these same metabolic processes after ingestion. For example, the breast cancer tamoxifen is primarily eliminated from the body by hydroxylation and subsequent (Figure 1.1).

Unfortunately, metabolism sometimes instead produces reactive metabolites, which can be toxic. In the case of tamoxifen, it has a secondary metabolite route that leads to a reactive

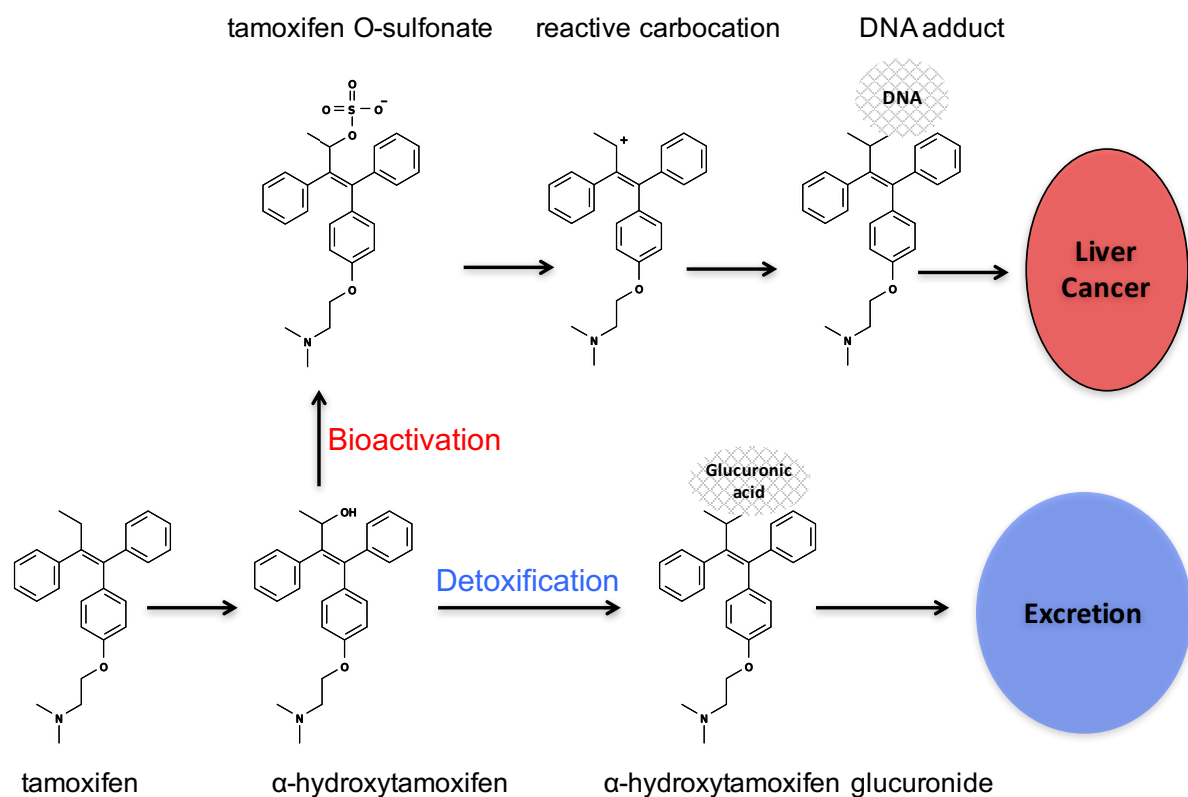


FIGURE 1.1: Bioactivation occurs when metabolism produce reactive metabolites that covalently bind to biological macromolecules. Generally, metabolism detoxifies drugs. For example, the typical metabolic pathway for the cancer drug tamoxifen is hydroxylation followed by glucuronidation. The glucuronide is very stable and facilitates excretion by increasing the polarity of tamoxifen. However, tamoxifen has a harmful secondary pathway: after the initial hydroxylation, it can undergo sulphonation instead of glucuronidation. This sulfate is unstable, and rapidly collapses to form a reactive metabolite, a carbocation. This carbocation binds to DNA, and can induce liver cancer. In rats, α -hydroxytamoxifen is primarily sulphonated, while in humans, glucuronidation is fortunately much more pronounced. Indeed, rats have a much higher chance of developing liver cancer after tamoxifen treatment than humans.[19, 28, 29, 30].

carbocation that binds to DNA and induces liver cancer. Fortunately, this metabolic pathway is infrequent in humans, or tamoxifen might not be a viable drug.

Many reactive metabolites are electrophiles, such as tamoxifen's reactive carbocation. Metabolism also produces another type of reactive metabolites: free radicals. Free radicals are potentially harmful as well. For example, free radicals can induce lipid peroxidation and damage cell membranes.[31] However, overall electrophiles are arguably the more pernicious type of reactive metabolites, because they have been linked to many IADRs.

Due to their electronic deficiency, electrophiles frequently conjugate to nucleophilic sites within biological macromolecules, including DNA and proteins. Conjugation to DNA can be mutagenic and potentially carcinogenic, as was mentioned for tamoxifen's carbocation metabolite. Conjugation of electrophiles to nucleophilic sites within proteins, such as cysteine or lysine residues, is also potentially harmful. These metabolite-protein adducts can induce toxic immune responses. For example, carbamazepine is an anti-epileptic with an IADR risk of about 1 out of every 5000 patients (Figure 1.2). This toxicity is believed to be often due to the formation of a reactive epoxide metabolite at a specific site, which then conjugates to nucleophilic sites within proteins

Indeed, there is a large body of evidence suggesting that many IADRs are immune-mediated. For example, most IADRs will reoccur much more quickly if a patient is retreated with the same drug for a second time, suggestive of immune system memory. [14, 24]. For instance, the IADR risk of carbamazepine (as well as lumiracoxib and ximelagatran), has been correlated with certain human leukocyte antigen alleles. These genes encode major histocompatibility common proteins, which among other functions help the immune system target pathogen-infected cells[32, 33]. This susceptibility of certain alleles to carbamazepine IADRs might explain why other drugs can cause some patients severe harm while leaving most untroubled. Despite exposure to likely the same metabolites, such rare variations in the immune system might cause unexpected reactions

One positive aspect of this bioactivation hypothesis is that it suggests small modifications

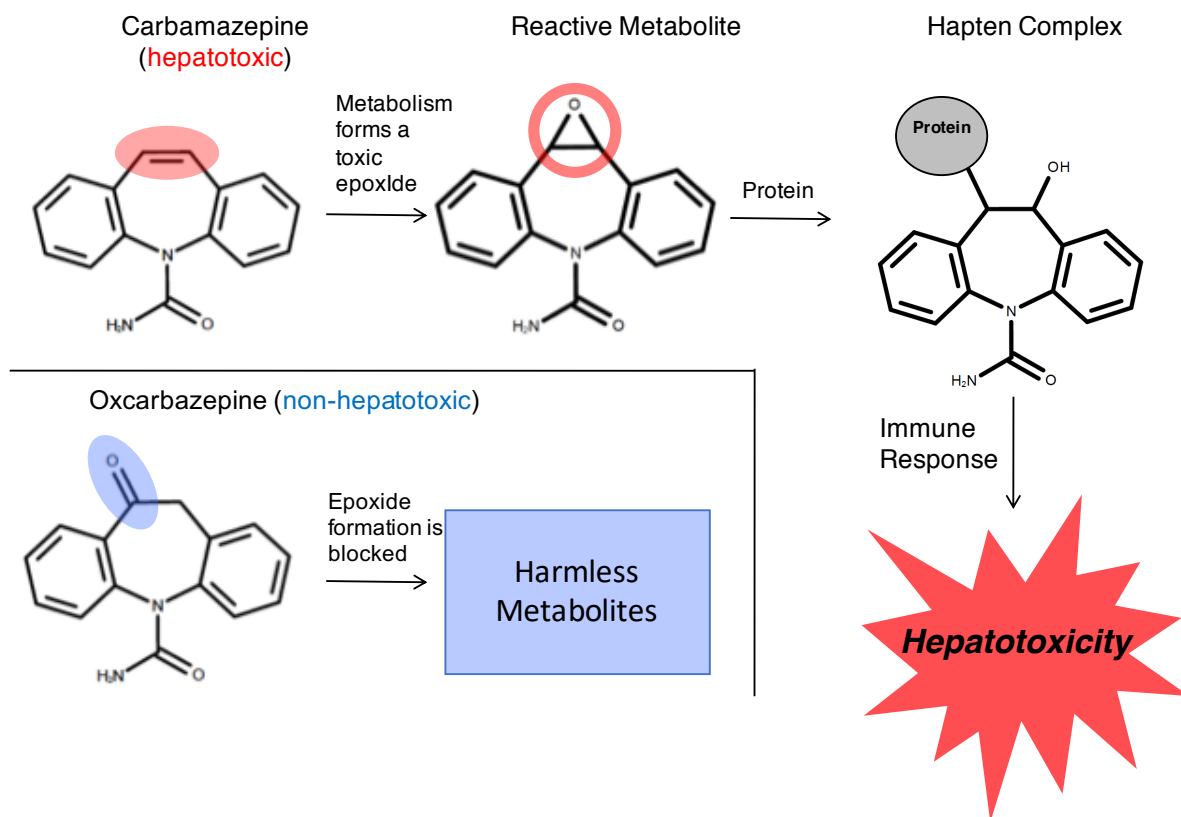


FIGURE 1.2: Reactive Metabolites can Induce IADR.

For example, the toxicity risk of carbamazepine—which causes IADRs in about 1 out of every 5000 patients—has been traced to the formation of a reactive epoxide metabolite. This metabolite conjugates to nucleophilic sites within proteins, and induces a toxic immune response in unlucky patients. In contrast, oxcarbazepine prevents epoxidation at that site, and has a lower risk of adverse effects.

can be made that may make drugs much safer. For example, carbamazepine has an analog, oxcarbazepine, which blocks carbamazepine's primary site of epoxidation with a ketone. This is likely the explanation for why oxcarbazepine has a much lower risk of causing IADRS. In fact, there are many well-studied cases of drug pairs that have only slightly different structures yet have significantly different IADR risk. For example, the hepatotoxic anxiolytic alpidem and the widely used, nontoxic zolpidem (Ambien), the hepatotoxic sudoxicam and the heavily prescribed meloxicam, and the occasionally several toxic nefazodone compared to the safe buspirone. The toxicity of each harmful drug within these pairs has been traced to a bioactivation event that is not observed in its safer analog. Highlighting these pairs is not to imply that each safe drug was designed in response to toxicity problems of its more dangerous analog. Oftentimes, these pairs are designed independently and only later linked or studied. Nevertheless, these comparisons demonstrate that understanding the specific bioactivation mechanism for a bioactivated molecule might enable design of safer analogs.

1.3 Computational Methods of Predicting Reactivity and Metabolism

While there is still much to be discovered about the exact mechanisms of most IADRs, in the absence of more specific understanding, drug developers are wary of advancing drug candidates that produce reactive metabolites (unless those reactive metabolites are necessary for a drug's mechanism of action). Unfortunately, accurately detecting reactive metabolites in the early stages of development can be expensive when considering the vast number of drug candidates present in the early stages of development, or molecules can be missed. Computational methods for predicting reactive metabolite formation have the potential to reduce the number of experiments needed. The most common computational tool for avoiding reactive metabolites are structural alerts. Structural alerts are simply certain motifs that are known to be often bioactivated. Using a data base of structural alerts, one can quickly identify potentially problematic molecules among a list of drug candidates. For example, one well-known structural alert is the

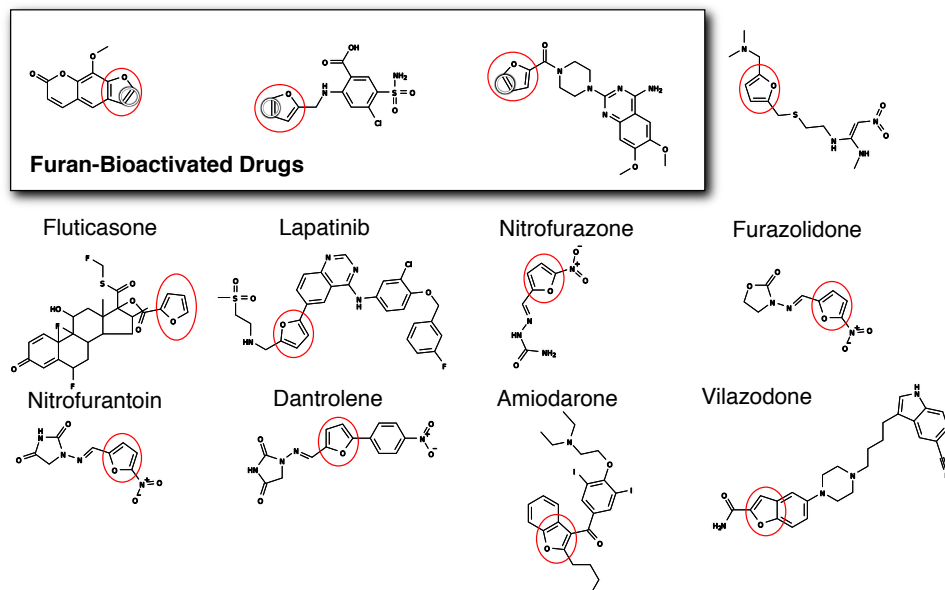


FIGURE 1.3: Structural alerts are not specific enough, because many safe drugs contain these motifs, which are not bioactivated due to specific molecule context. Therefore, avoiding any drug with a structural alert would result in many safe drugs being missed. In the case of furan, all 13 displayed drugs contain this structural alert, circled in red. However, in only 3 of these drugs is the structural alert actually bioactivated

furan group: an oxygen-containing aromatic ring with five-members that is often metabolized to an epoxide, a common type of reactive metabolite (Figure 1.3).

Structural alerts are a natural strategy for trying to avoid reactive metabolites, by simply trying to avoid in the future what has caused problems in the past. However, by sometimes confounding molecules with important differences that nevertheless have a common substructure, structural alerts also have several critical deficiencies that making them unsatisfactory in several ways. First, they are by definition entirely reliant on historical data, and consequently have no ability to forecast the properties of new drugs with novel structures. One can try to update a structural alert library based on new results, so at least not to remain stagnant, but they have no capacity to leverage understanding of already-known structural alerts to make new predictions. Additionally, some reactive metabolite-producing drugs do not contain structural alerts. For example, the hypertension drug ambrisentan does not contain a structural alert and has not been observed to produce reactive metabolites, yet caused several cases of severe hepatotoxicity, the

mechanism(s) of which were never uncovered.[34] Finally, not all structural-alert containing drugs are actually bioactivated.

Instead of structural alerts, more sophisticated data-driven methods seek to extrapolate underlying principles from current data in order to hopefully generalize to never-seen-before patterns. A common goal of these methods is to predict sites of metabolism: the specific atoms that are modified during reactions with cytochromes P450. One of the earlier and more successful of these was “SMARTCyp”, which began by performing high-quality quantum chemical computations that approximate cytochrome P450 activation energies. Using this simulation, the creator of SMARTCyp constructed a lookup table of “SMARTS” patterns (a system for denoting molecule substructures) that listed the activation energies of cytochromes P450 reactions with many different substructures. Because of the wide variety of substructures considered, this method offers higher-resolution predictions than structural alerts. For example, in some cases different manifestations of the same alert were scored, thereby recording different scores for motifs that would be considered identical by the structural alerts method. However, SMARTCyp only analyzes molecular substructures, and therefore might fail to differentiate molecules with common structures yet other important features outside of the matching area. Furthermore, if a particular motif has not been scored by SMART Cyp, there is no way to generalize from its previous computations to estimate a score for the new motif.

A more flexible approach, and in some ways a precursor to my own work, was taken by RS-Predictor (for RegioSelectivity).[35, 36, 37]. The method first calculated about 500 topological quantum chemical and topological descriptors, which are numerical representations of various properties of atoms and molecules. Next, support vector machines were used to find a function that maps between these values and experimentally-known sites of metabolism. Compared to SMARTCyp, RS-Predictor has the advantage of providing predictions for every atom within a test molecule. Consequently, even novel motifs are valid inputs for prediction, although there is no guarantee that the predictions will be accurate.

However, RS-Predictor has several disadvantages. The support vector machine algorithm

produces a complex function that obfuscates which features are important for its decision-making. Additionally, support vector machines do not produce probabilistic predictions, for their goal is merely to find a separating hyperplane between classes. In other words, a prediction for one site that is double that of another site does not imply that the model believes the first site is twice as likely to be metabolized. As a result, its predictions are fairly binary, and does not produce a reliable measure of confidence when making predictions. Without such scores, it is difficult to use its output to triage molecules for experimental validation. Moreover, the kernel it uses entails comparisons between every data point, so it does not scale well to large data sets. Finally, support vector machines by definition have to retain a subset of their input data—the support vectors—which can hamper model distribution if trained on proprietary structures.

A successor of RS-Predictor—XenoSite—improved on the previous model in several respects, by switching from support vector machines to neural networks. First, unlike support vector machines, neural network predictions are probabilistic. Secondly, neural networks extract parameters from the data during training, after which the original data can be discarded. Therefore, one can disseminate neural network models without risk of disclosing the underlying training structures. Additionally, neural networks scale more easily to larger data sets, making them easier to restrain as more data becomes available. Finally, these networks are highly modular, and can be customized for a variety of different tasks.

1.4 Overview of Chapters

As an example of this flexibility, consider the model structure utilized by the quinone formation model, detailed in Chapter 5 and Figure 1.4. This model was customized to output three different types of predictions, for atoms, atom pairs, and molecules. Such customizability is not as easily possible with the more rigid support vector machine framework. My dissertation harnesses the flexibility of descriptor-trained-neural networks to model different components of the bioactivation mechanism, and finally synthesize them into a single bioactivation model

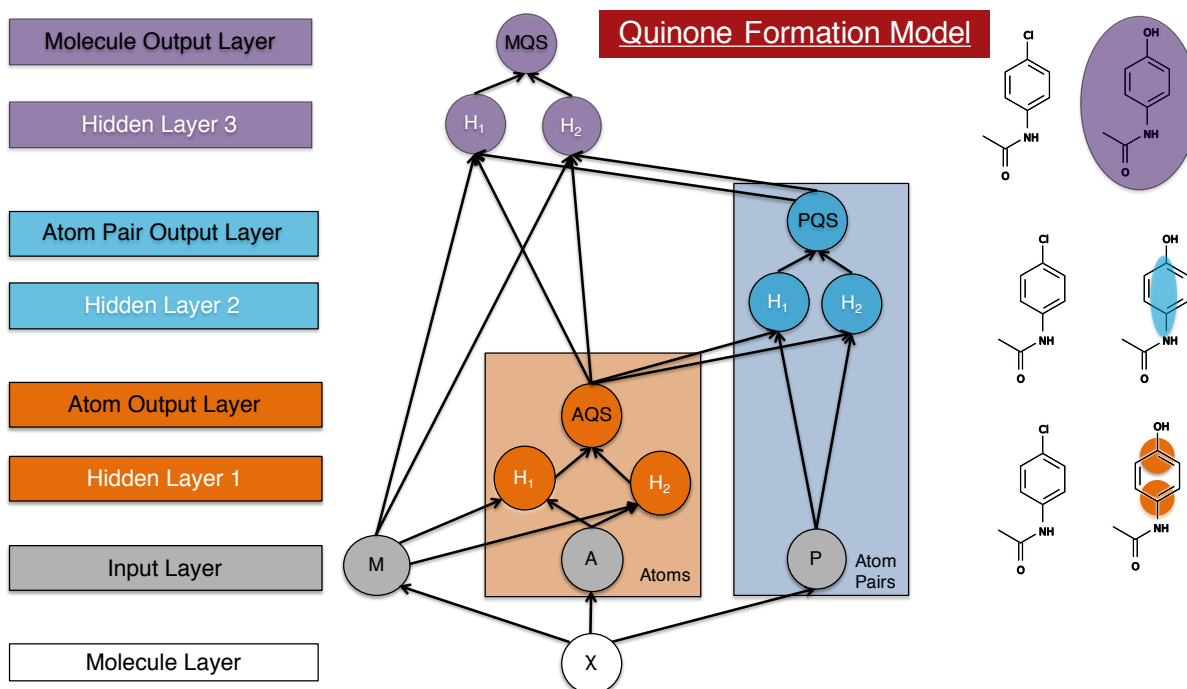


FIGURE 1.4: As an example of the customizability of neural networks, the structure of the quinone formation model is visualized here. By changing the input data, model structure, or input descriptors, similar models were constructed for epoxidation, reactivity, and bioactivation.

Chapter 1. Introduction

First, I extended much of the methodology and paradigm from site of metabolism to instead predict sites of reactivity. Chapter 2 details development of the first reactivity model I developed, for predicting reactivity to glutathione (GSH). GSH reactivity is important to predict because conjugation to GSH is frequently used in experimental assays to detect reactivity metabolites, as conjugation to GSH is an important endogenous defense mechanism for eliminating reactive metabolites. Previous studies of GSH reactivity had focused on small datasets of structurally similar molecules.[38, 39, 40] Consequently, these few reactivity indices had little applicability outside their training domain. Instead, in this study we accurately modeled reactivity across 1213 diverse molecules.

The success of GSH study was highly encouraging, because it was the first model of diverse chemical reactivity, and suggested that the similar approach used in both the original XenoSite study and the GSH paper might be further applicable to other important problems. Indeed we constructed a model that shared several features with the GSH reactivity model, but extended it to predicting reactivity to biological macromolecules, including DNA and protein (Chapter 3).

Next, I turned my attention to metabolism. Specifically, I built models that predict the formation of some of the most common types of reactive metabolites. First, in Chapter 4 I constructed a model of epoxidation, an example of which was seen in Figure [fig:carb_v_s_o_xcarb](#)). *Epoxides are about 10% of a*

Even more prevalent, however, are quinone species, which make up 40% of known reactive metabolites. In Chapter 5, I devise a method for accurately predicting quinone formation, on both the site and molecule level, and across multiple metabolic steps.

Up to this point, I had modeled both reactivity and metabolism, but had considered each in isolation from one another. The key missing link between these models were the actual metabolite structures. To fill this gap, I spent significant effort devising a comprehensive system of metabolite structure prediction, detailed in Chapter 6.

Finally, in Chapter 7 I tied together all my previous work by building a bioactivation model that predicts where epoxides or quinones might form on a given molecule, then enumerates the structures of these possible metabolites, and scores the reactivity of these structures. In this

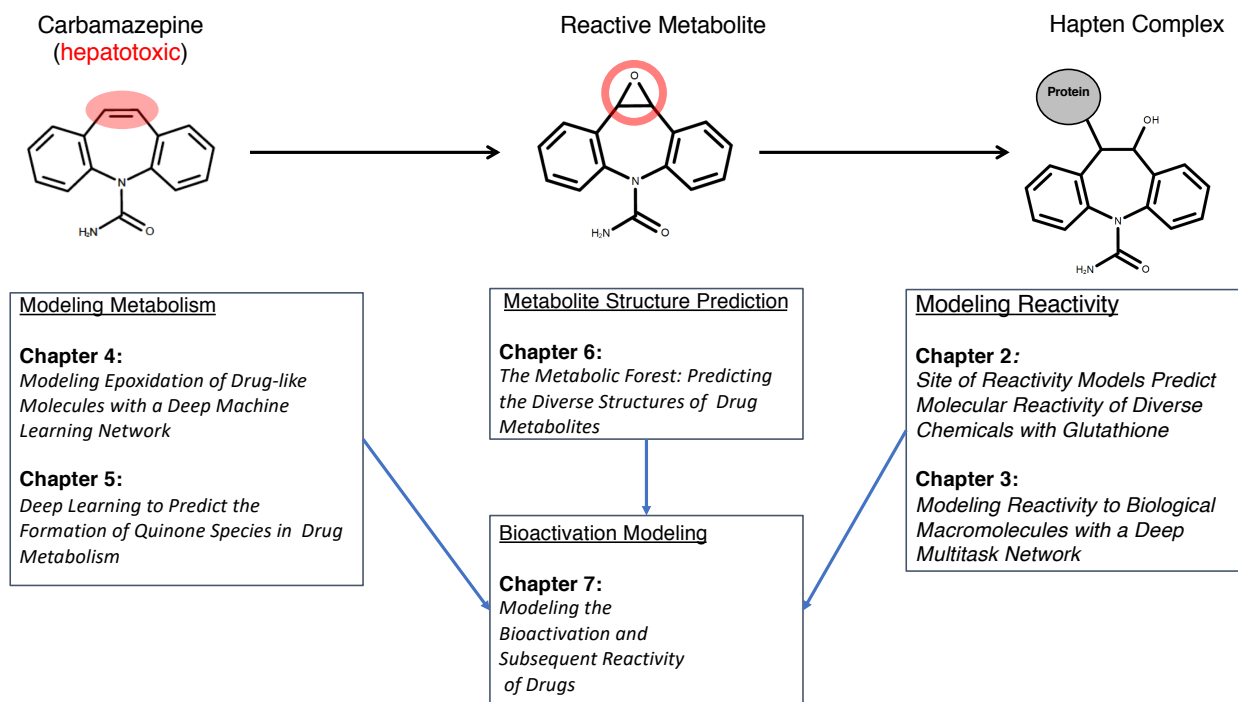


FIGURE 1.5: The chapters numbers are based on order of publication, with the final two chapters still pending publication. However, they are visually organized here to demonstrate the relationship between each study and the component of the bioactivation pathway to which each study relates.

manner, I was able to predict both how likely metabolites are to form, and to rank their potential reactivity.

These chapters are arranged virtually in Figure , to show how each address a different component of bioactivation.

Chapter 2

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

Reprinted (adapted) with permission from Hughes, T. B., Miller, G. P., and Swamidass, S. J. (2015). Site of Reactivity Models Predict Molecular Reactivity of Diverse Chemicals with Glutathione. *Chemical Research in Toxicology* 28(4), 797–809. Copyright (2015) American Chemical Society.

2.1 Introduction

It costs about one billion dollars to bring a single drug to market.[41, 42] Efforts to make the process more efficient may be necessary for survival of the pharmaceutical industry.[43, 44] An estimated 50% of drug leads fail due to low efficacy and 40% fail due to toxicity issues[10]; both of these issues are often associated with failure to properly predict the impact of drug metabolism.[45] 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.[19] DILI is responsible for 50% of acute liver failure cases, as well as 15% of liver transplants within the United States.[19] This drug toxicity often reflects formation of

electrophilic reactive metabolites, such as quinones or epoxides, which covalently bind to proteins or DNA.[19, 46] These metabolites typically bind nucleophilic sites in proteins, such as thiols,[47] and then cause adverse drug reactions by eliciting an immune response.[48, 19, 49]

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 exist. Several studies—by our group and others—have already demonstrated that computational models can predict how molecules are metabolized.[50, 51, 52, 36, 53, 54] 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,[55] 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 2.1).[56, 57, 58, 59] Experimental systems exploit this fact to monitor reactive metabolite formation by detecting metabolites conjugated to GSH.[60, 39, 56] 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 non-reactive molecules that contain the same substructure.[9] For example, all Michael acceptors are flagged as problematic even though many are not actually reactive.[61, 39] 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.[39, 56] Others have suggested that indices derived from quantum simulations may predict reactivity, but this idea has only been validated in limited studies that do not include GSH.[62, 63, 64, 65, 66] A computational way of predicting the reactivity of a large range of structurally diverse molecules would be a major advance in the field.

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 1,400 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.[67] We found success with this approach when predicting P450 metabolism.[50] When applied in this study, the strategy was capable of encoding non-linear 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.

2.2 Materials and Methods

2.2.1 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

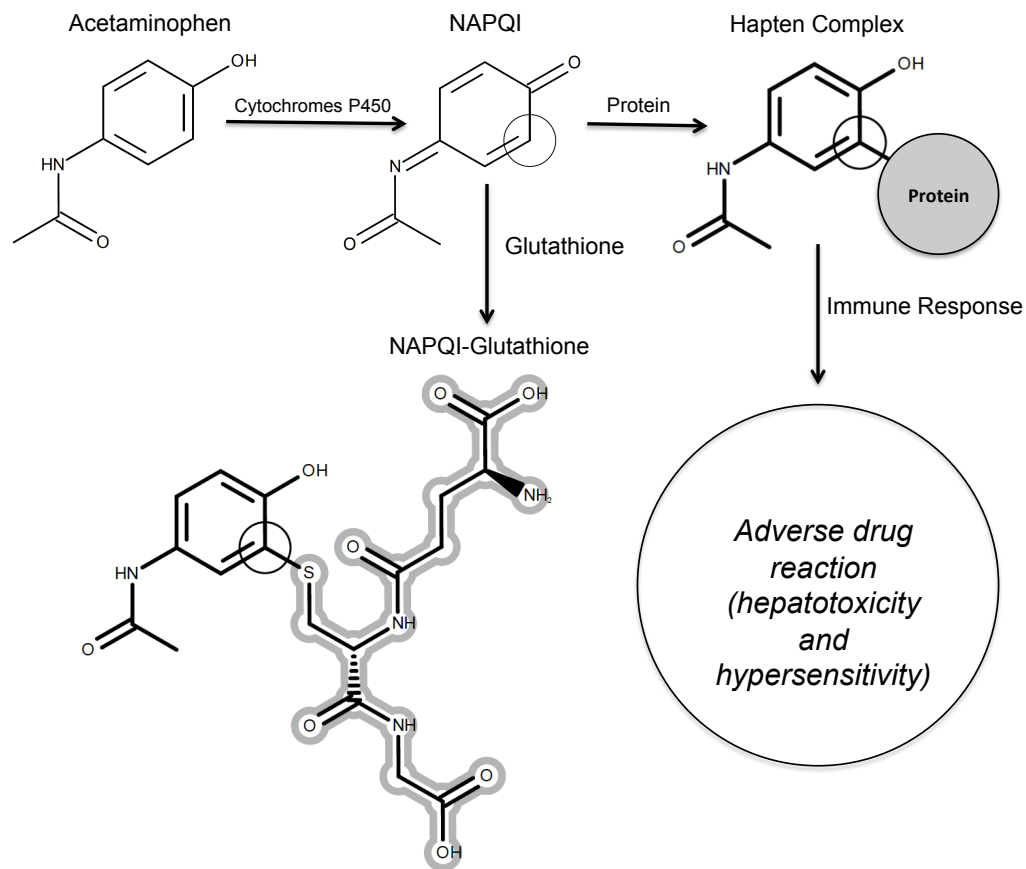


FIGURE 2.1: Adverse drug reactions are often caused by reactive metabolites. Acetaminophen is metabolized by Cytochromes P450 to *N*-acetyl-*p*-benzoquinone imine (NAPQI). NAPQI is electrophilically reactive and covalently binds to nucleophilic sites within proteins eliciting an immune response. Glutathione (GSH and outlined in grey) 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.

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 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 Supplementary Material, and this is enough information to reconstruct the database and replicate our results.

2.2.2 External Reactivity Data

We assessed the predictive ability of the reactivity model against two published external quantitative data sets. The first study focused on ten substituted *p*-benzoquinone compounds, measuring both the rate at which molecules react with GSH ($\log k_{\text{GSH}}$) and rat hepatocyte viability (the concentration at which 50% lethality is observed, $\log \text{LC}_{50}$).^[39] 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 hours or 15 minutes, respectively.^[60]

2.2.3 Comparison QSAR Models

As a baseline for comparison, QSAR models reported in the literature might be expected to predict GSH reactivity.^[39, 56] 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 ten *p*-benzoquinones.[39] 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:

$$\log \text{LC}_{50} = 24.54 + 17.7 \cdot E_{\text{LUMO}} + 3.36 \cdot (E_{\text{LUMO}})^2 \quad (2.1)$$

$$\log k_{\text{GSH}} = -18.38 - 16.78 \cdot E_{\text{LUMO}} - 3.19 \cdot (E_{\text{LUMO}})^2 \quad (2.2)$$

The authors excluded one molecule as an outlier when training Equation 2.1, while Equation 2.2 was based on the reactivity of all ten 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.

2.2.4 Descriptors

Our study utilized a total of 116 different descriptors, including both atom-level 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.[68] Descriptor subgroups include topological, quantum chemical reactivity, and molecule-level descriptors.[69] Table 2.1 provides a condensed summary of quantum chemical reactivity descriptors discussed in this paper; a comprehensive table is available in the Supplementary 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.[50] 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 ($\max [D^N(r)]$ and $\max [D^E(r)]$), and the maximum self-polarizability ($\max [\pi^S(r)]$). Moreover, $D^N(r)$, $D^E(r)$, and $\pi^S(r)$ have been proposed as atom-level reactivity indices that may predict sites of GSH reactivity.[70,

TABLE 2.1: Quantum chemical reactivity descriptors were generated at both the atom and molecule level.

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

[71, 72] These reactivity descriptors are calculated from self-consistent field computations using MOPAC, a semi-empirical quantum chemistry modeler, using the PM7 force field and an implicit solvent model.[73, 74]

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,[75] off-target prediction,[76, 77, 78] and high-throughput screening analysis[79, 80], their use to predict reactivity would be problematic. Instead of overall shape, it is specific substructures in 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.[81] 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.

2.2.5 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 2.2). 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.[82]. 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 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.

2.3 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

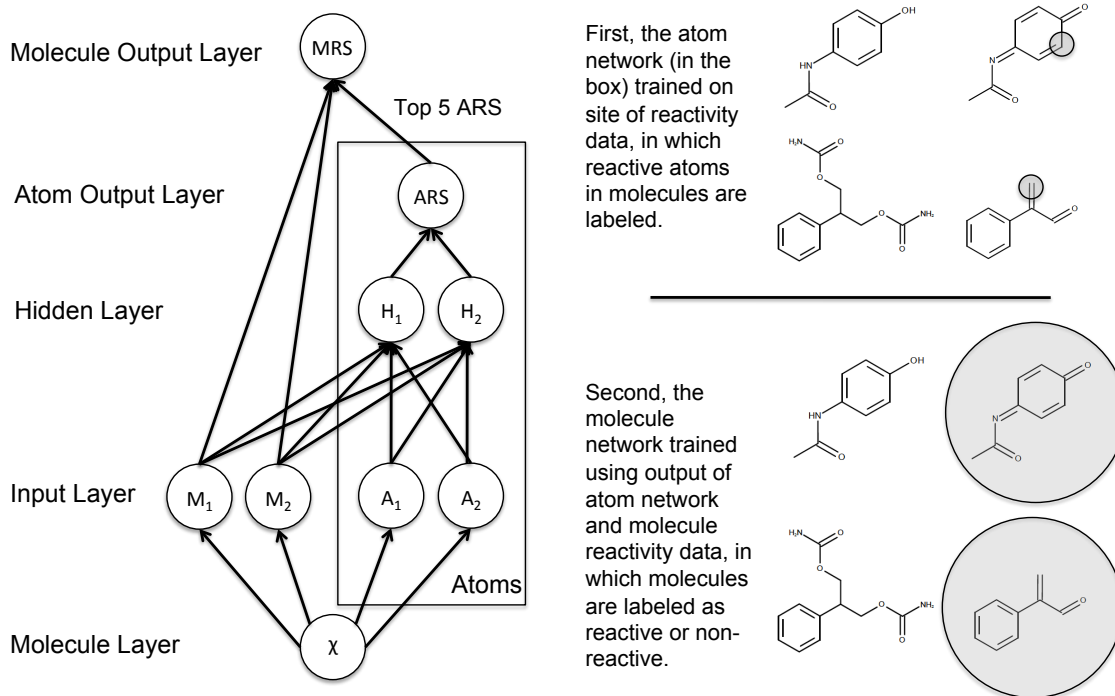


FIGURE 2.2: The structure of the reactivity model. This diagram shows how information flows through the model, which is composed of one input layer, one hidden layer and two output layers. This model computed a prediction for each test molecule and atom in the test molecule. Atom reactivity scores (ARS) 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 \mathcal{X} , 30 molecule-level and 86 atom-level descriptors were calculated (two input layer nodes for each category are displayed). The diagram only shows 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 \mathcal{X} , all 116 descriptors were fed into the ten hidden layer nodes (two are displayed), which generated an ARS. The molecule reactivity score (MRS) of \mathcal{X} was computed from the top five ARS corresponding to the scores of the five atoms predicted to be 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.

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 which descriptors the model relies upon. Fourth, we assessed the reactivity model using external quantitative data from the literature.

2.3.1 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 molecular 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 at a time. The remainder of the molecules was used to train a model, and made predictions on all 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.[62, 63, 64, 65, 66] 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 2.3).

Second, ARS accuracy was greater than any of the individual descriptors including those specifically designed to predict reactivity. For example, self-polarizability, $\pi^S(r)$ separated reactive and unreactive atoms with an accuracy of 75.3%. Likewise, the nucleophilic and electrophilic delocalizabilities ($D^N(r)$ and $D^E(r)$), also known as fukui reactivity, separated reactive and unreactive atoms, but 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 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 non-linear component in the reactivity model that was consistent with previous work on the best models predicting sites of metabolism.[83]

2.3.2 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.[84] First, a model was built using all 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

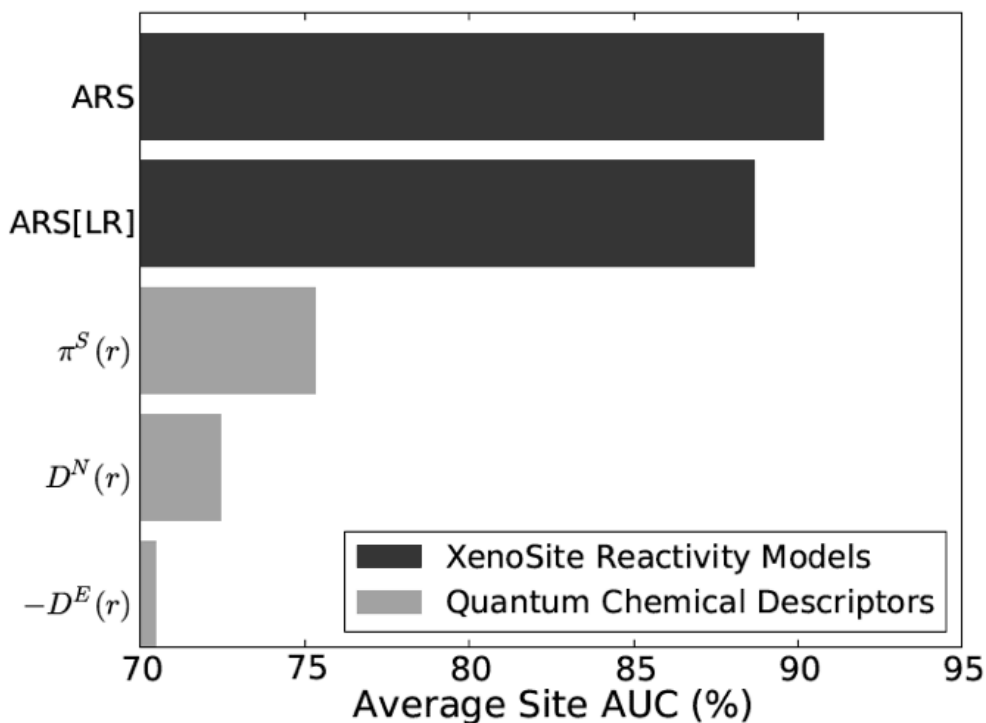


FIGURE 2.3: Atom reactivity scores accurately identified sites of reactivity. For each prediction method, average site AUC was computed for 1213 reactive molecules with their sites of conjugation to glutathione labeled. This metric reflected how often reactive atoms were ranked above unreactive atoms within reactive molecules. The cross-validated atom reactivity scores (ARS), generated by a neural network with ten 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.

groups of descriptors were included in the Supplementary Information, and only the most salient results were presented here. For each set of descriptors, we repeated the randomization process ten 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 4.5). 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 twelve 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.[54, 85, 86] 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.

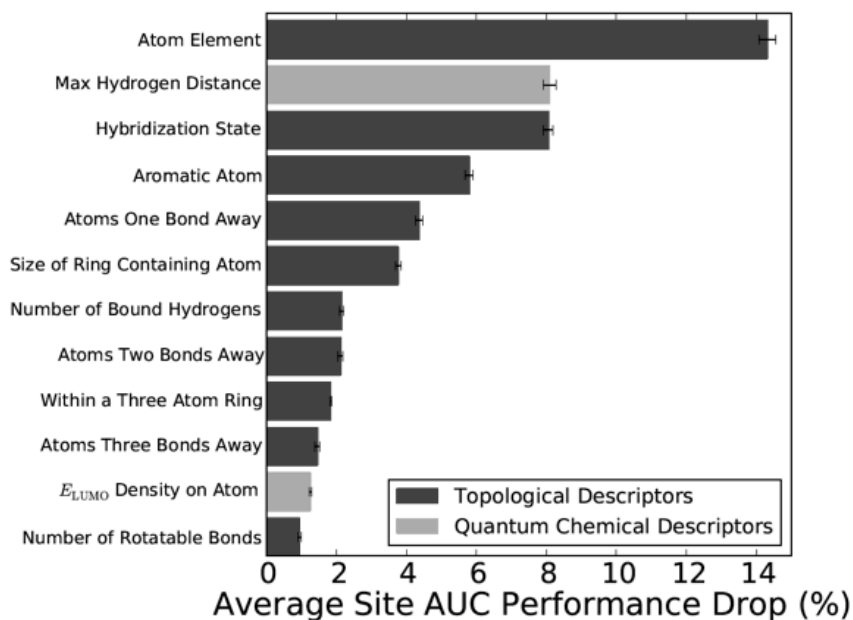


FIGURE 2.4: The importance of specific descriptors to the atom reactivity model. A permutation sensitivity analysis quantified the importance of descriptors for the final trained atom reactivity model. This listing indicates the twelve most important descriptors in decreasing order of importance from top to bottom. The graph shows the model performance drop associated after permuting the associated descriptor values, averaging over ten iterations. All top descriptors with the exception of two were topological; the remainder were derived from a quantum simulation of the molecule structure.

2.3.3 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.[39, 56] The authors of these studies were quick to point out that these QSAR models were trained on very small data sets of similar molecules, so that 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 contributed 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[39], and reactivity indices proposed in the literature[70, 71, 72], including individual molecule descriptors and the maximum of each atom descriptor associated with the molecule atoms.

From this assessment (Figure 2.5), 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-level 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 non-linearity to improve accuracy. Second, the model MRS outperformed two published QSAR models for reactivity which 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 non-reactive molecules had average scores of 0.85 and 0.65, respectively, which suggests that the model more confidently identified reactive molecules than non-reactive molecules. This could be a consequence of error in the training data arising from difficulty in accurately identifying non-reactive 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 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 bench top 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 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 still 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.

2.3.4 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 2.6). 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[87]. Lastly, the weights of the remaining reactivity indices

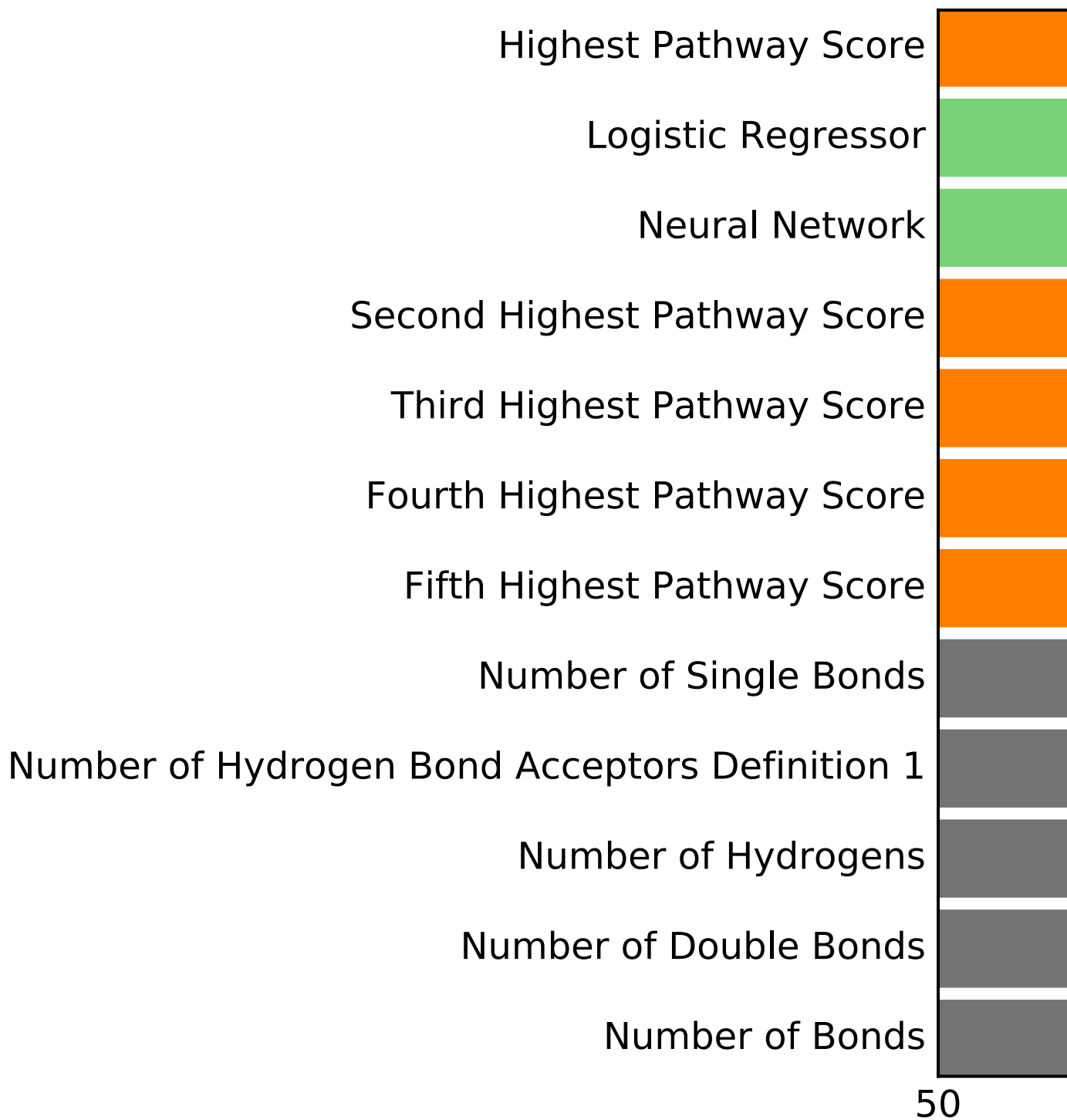


FIGURE 2.5: The reactivity model accurately identified reactive molecules. Several prediction methods were compared based on their ability to identify reactive

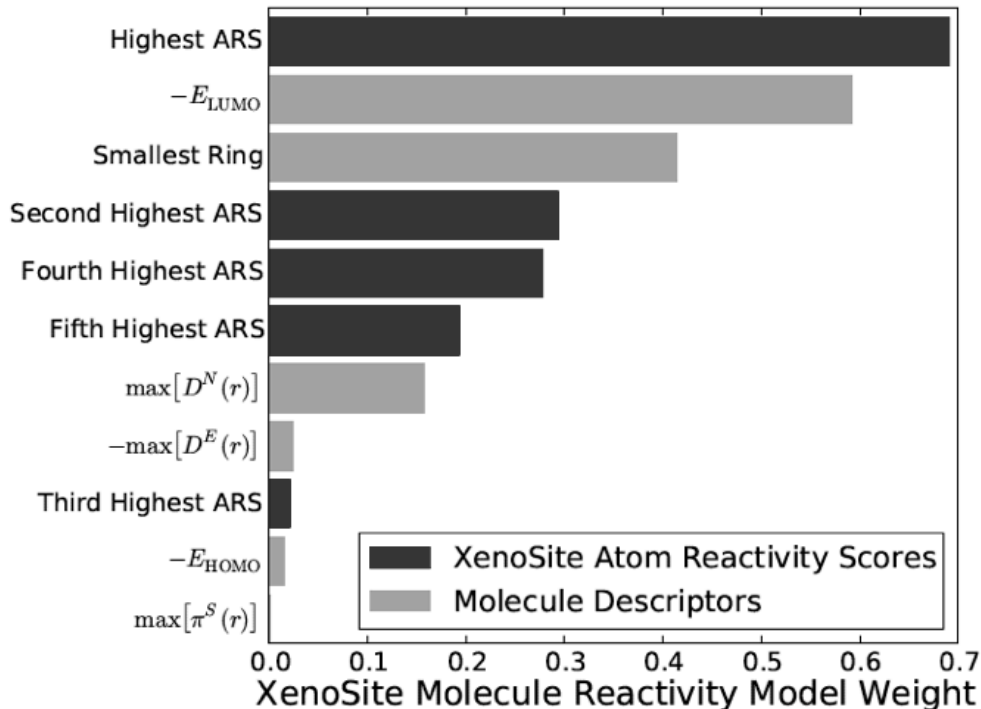


FIGURE 2.6: Importance of descriptors to molecular reactivity score. The weights of the final model nodes revealed the relative contribution of individual descriptors to model performance. The descriptors were normalized before training, so that the magnitude of the weights directly measured the importance 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.

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.

2.3.5 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 assure us that errors in training data are not so significant that they prevent the construction of a useful model.

The first external data set on ten 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, and so we determined correlations between MRS and either set of experimental values (Figure 2.7). 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 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.[60] 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 minutes (for GSH) or 24 hours (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 2.2 and Figure 2.8). MRS correlated with GSH depletion with a significant Pearson correlation ($R = 0.43$), whereas the QSAR model had a non-significant correlation ($R = 0.30$). Of the three 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 than traditional QSAR models, which are

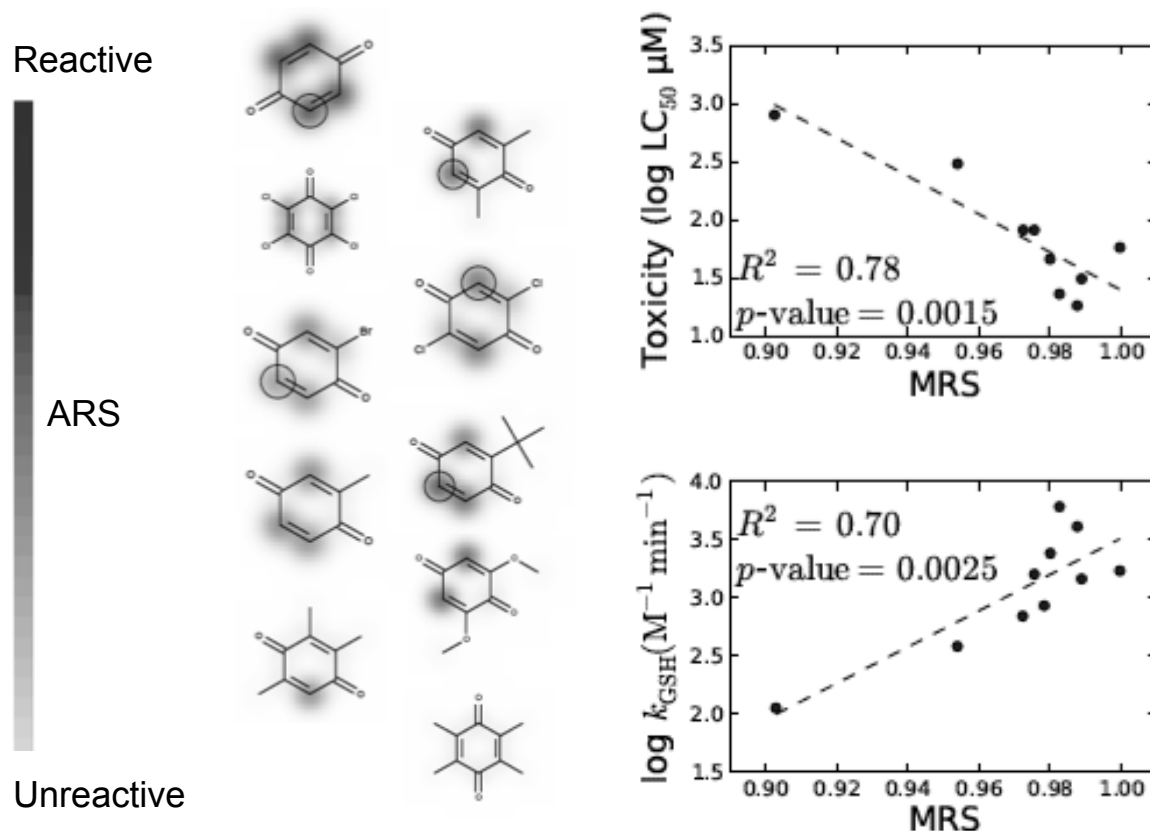


FIGURE 2.7: Molecule reactivity scores correlated with glutathione reactivity and toxicity of substituted quinones. The model molecule reactivity scores (MRS) correlated closely with hepatocyte toxicity (LC₅₀, top graph) and the rate of reactivity with GSH (k_{GSH} , bottom graph) of ten substituted *p*-benzoquinones.[39] The left panel illustrates all ten 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.

TABLE 2.2: Reactivity scores correlated more closely with trapping agent depletion than QSAR models. The correlations between peptide reactivity and QSAR predictions (Equation 2.2) are molecule reactivity scores (MRS) were calculated for 38 structurally diverse contact allergens.[60] Reactivity was measured by peptide depletion assays, with incubation times of 15 minutes for GSH or 24 hours 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.

Peptide	QSAR Reactivity Model	MRS
GSH	0.30 (<i>p</i> -value = 0.0709)	0.43 (<i>p</i> -value = 0.0064)
Cysteine	0.38 (<i>p</i> -value = 0.0210)	0.56 (<i>p</i> -value = 0.0003)
Lysine	0.26 (<i>p</i> -value = 0.1310)	0.41 (<i>p</i> -value = 0.0126)
Histidine	0.14 (<i>p</i> -value = 0.4295)	0.24 (<i>p</i> -value = 0.1504)

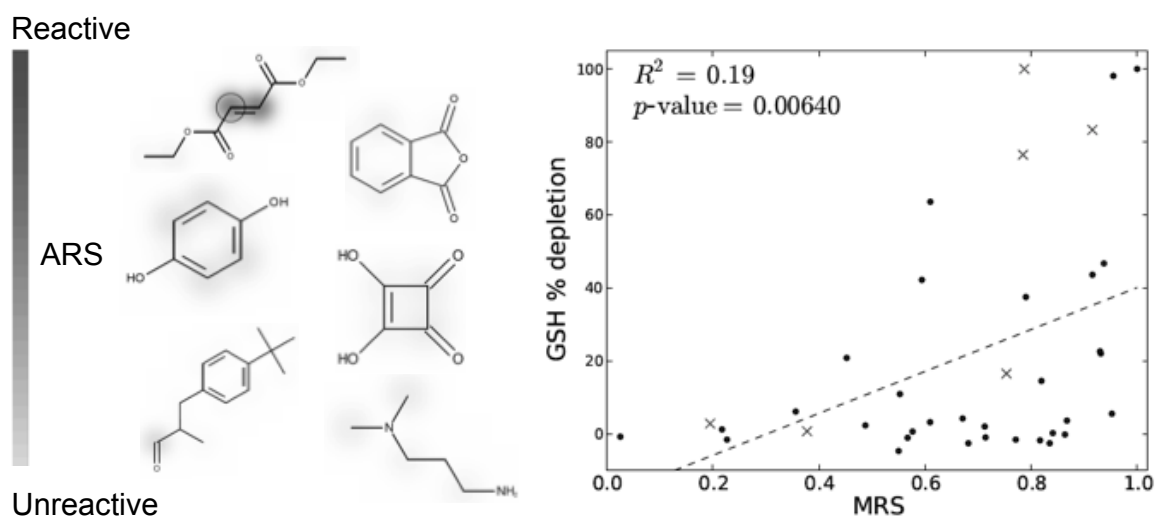


FIGURE 2.8: Reactivity scores correlated with the nucleophile reactivity of structurally diverse contact allergens. Model performance was assessed using an external data set with 38 molecules.[60] The y-axis is the percent depletion of GSH after 15 minute incubation with each molecule. The x-axis indicates 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 X in the right panel plot. The six corresponding X's are in the same horizontal order as visualized molecules.

MRS significantly correlated with GSH reactivity.

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.

2.3.6 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 non-reactive 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 paper, we consider two drugs, trimethoprim and felbamate, and their known reactive metabolites (Figure 2.9).

The metabolism of the antibacterial trimethoprim forms a reactive iminoquinone methide metabolite that likely contributes to trimethoprim hypersensitivity reactions.[88, 9] The reactive metabolite conjugates to GSH or *N*-acetyl cysteine in human and rat liver microsomes.[89] 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.[89] The anticonvulsant felbamate sometimes causes hepatotoxicity and aplastic anemia, and such adverse reactions have been traced to a reactive α,β -unsaturated aldehyde metabolite.[9] Our model assigned this reactive metabolite a very high MRS score of 0.96, whereas felbamate only scores 0.33. Moreover, the atom

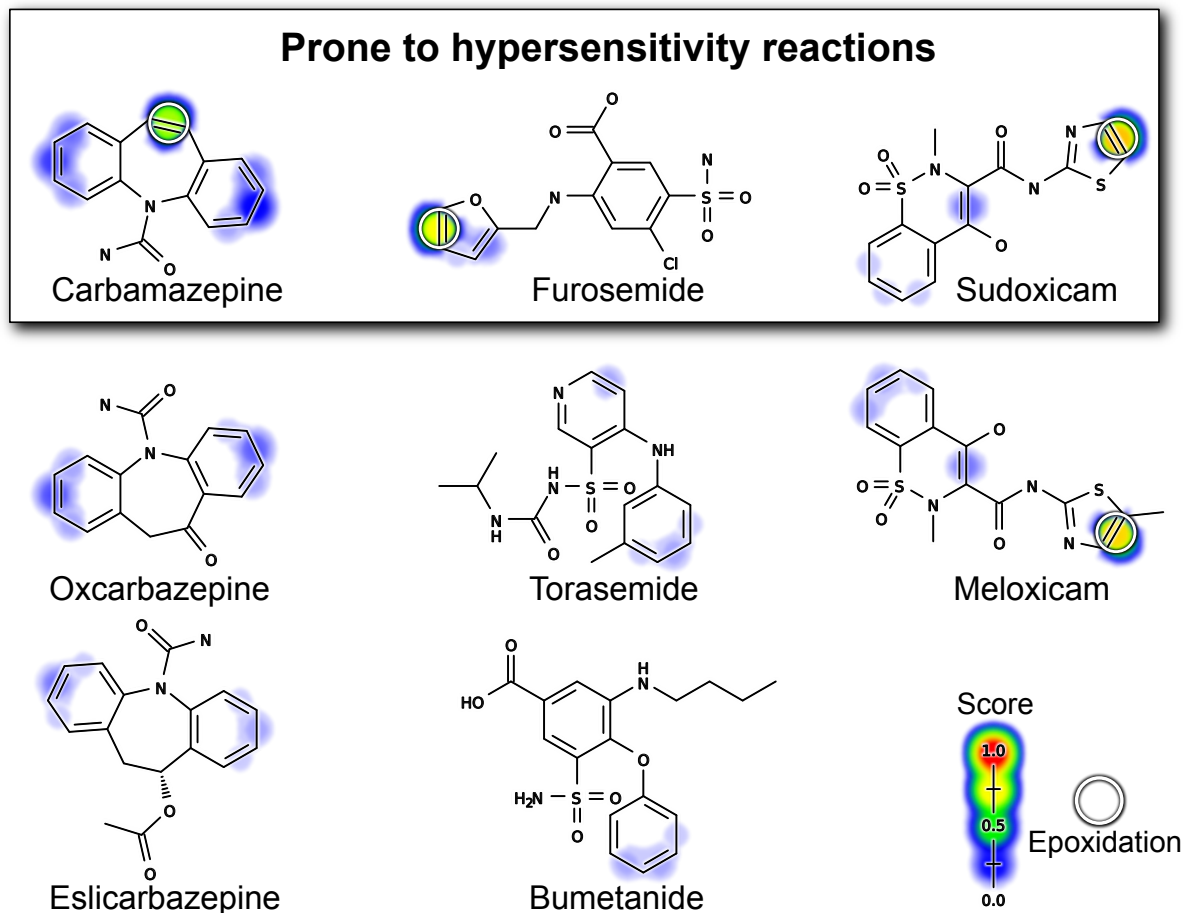


FIGURE 2.9: Molecule reactivity scores distinguished drugs and their reactive metabolites. For each molecule, the shading represents atom reactivity scores (ARS), which ranged from 0 to 0.768. The structures of trimethoprim and felbamate are shown, alongside their reactive metabolites and subsequent 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 non-reactive. And second, in the current form, it cannot predict how metabolism gives rise to reactive species.

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.

2.4 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.[61] 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 "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 2.10). Second, MRS could identify non-obvious reactive molecules, but 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

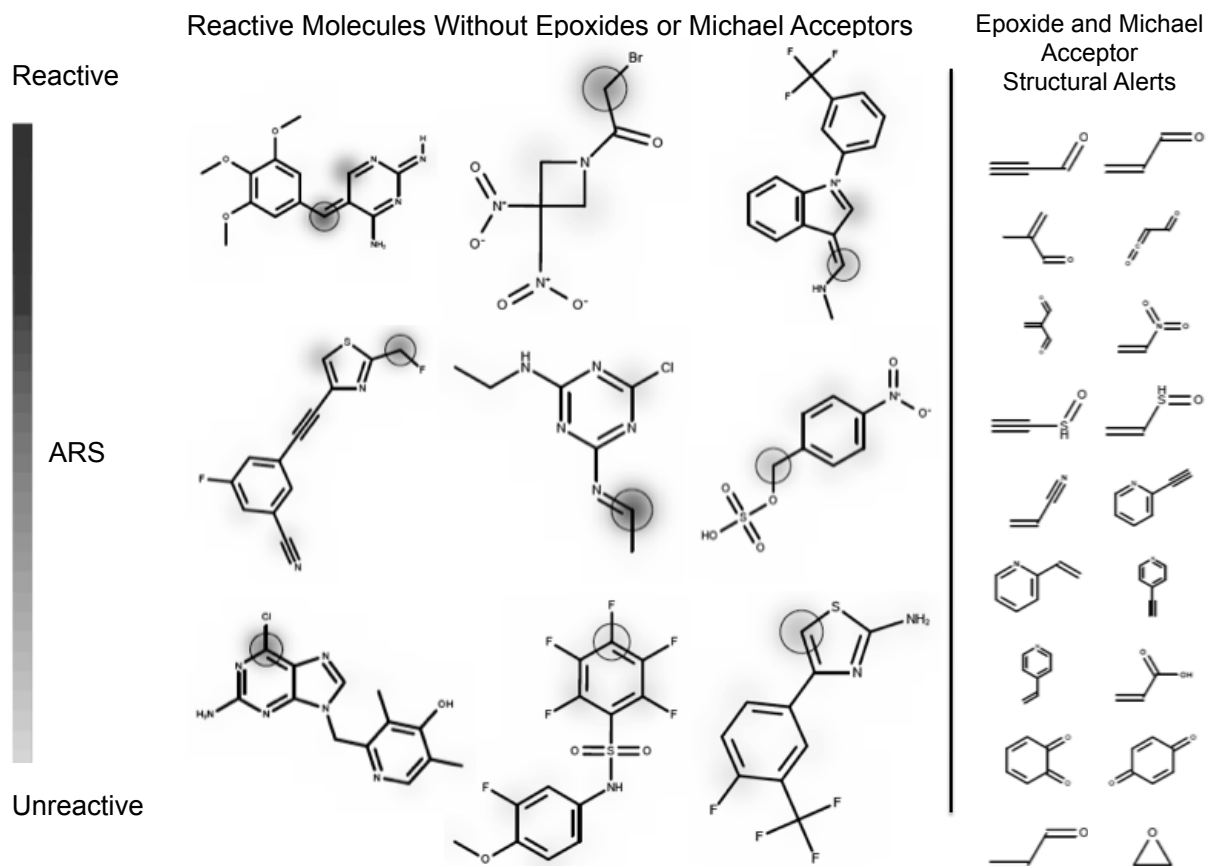


FIGURE 2.10: Atom reactivity scores accurately identified non-obvious sites of reactivity. The right panel displays 18 structural motifs known to be reactive. [61] The left panel displays reactive molecules drawn from our training data that do not contain these specific substructures.[9, 90, 91, 92, 93, 94, 95, 96, 97] 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.

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.

2.5 Conclusion

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 only trained on qualitative data—whether or not 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 only detect 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 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.

Chapter 3

Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network

This chapter is adapted from this published manuscript:

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 Central Science* 2(8), 529–537.

3.1 Introduction

Most small-molecule drug candidates fail before entering the market,^[8] frequently because of unexpected toxicity.^[10, 8] Often, toxicity is detected only late in drug development, because many types of toxicity, especially idiosyncratic adverse drug reactions (IADRs), are particularly hard to predict and detect.^[98, 6] Moreover, drug-induced liver injury (DILI) is the most frequent reason drugs are withdrawn from the market, and causes 50% of acute liver failure cases in the United States.^[19]

A common mechanism often underlies many types of drug toxicity, including both DILI and IADRs. Drugs are bioactivated by drug-metabolizing enzymes into reactive metabolites, which then conjugate to sites in proteins or DNA to form adducts. DNA adducts are often mutagenic, and may alter the reading and copying of genes and their regulatory elements, causing gene

Network

dysregulation and even triggering cancer.[99, 100, 19] Similarly, protein adducts can disrupt their normal biological functions and induce harmful immune responses.[48, 19, 49]

Metabolites are reactive because of their chemical properties, and are often generally classified as soft or hard electrophiles based on polarizability and their preferential reaction with targeted nucleophiles.[101] Soft electrophiles like epoxides or Michael acceptors have low electron density at multiple sites, while hard electrophiles such as carbocations or saturated aldehydes have a localized site with low electron density.[102, 31, 101] Soft electrophiles tend to react with soft nucleophiles, like cysteine residues within glutathione (GSH) or protein, whereas hard electrophiles generally react with hard nucleophiles, such as purine and pyrimidine bases in DNA or lysine and histidine residues within protein.[19, 102, 103, 101, 60, 104] Despite these general rules, it remains a challenge to predict the reactivity of small molecules and their likelihood for modifying DNA and proteins.

Conjugation between small molecules and nucleophilic GSH (soft) or cyanide (hard) is commonly used in screening studies to identify molecules capable of forming adducts. Detecting GSH or cyanide adducts is easier than detecting protein or DNA adducts, and serves as a proxy for reactivity to macromolecules in experimental studies [105, 106]. Moreover, GSH is a physiologically relevant trapping agent, which reaches millimolar levels in cells and protectively conjugates with many reactive molecules.[107]

However, GSH and cyanide are only imperfect proxies for protein and DNA. Proteins and DNA are structurally complex macromolecules and likely have correspondingly complex reactivities with a diverse set of soft and hard nucleophilic molecules. Therefore, we expect some molecules will react with protein or DNA, but not efficiently react with GSH or cyanide. These molecules are of special concern, because they do not react with GSH or cyanide, and consequently are likely to be missed by many standard reactivity experiments. Computational modeling could provide a complementary strategy for detecting molecules likely to be reactive, and therefore prone to causing DILI or IADRs, including those molecules missed by standard

screening assays. Others have proposed QSAR models to predict GSH reactivity, yet these models are of limited value, as they are focused on very limited structural groups.[40] In contrast, we recently published a computational model that predicted GSH reactivity toward a diverse set of chemicals at both the site and molecule level.[108]

Here, we aimed to extend our previously published GSH reactivity model to also predict reactivity with cyanide, protein and DNA. First, we extracted a structurally-diverse, literature-derived database of molecules known to bind DNA and protein, as well as the simple nucleophilic traps cyanide and GSH (Figure 1). Second, we labeled the site of conjugation on each molecule, known as its site of reactivity (SOR). Third, we used a deep convolution neural network to accurately predict these SORs in cross-validated experiments. Fourth, we transform SOR scores to accurate molecule-level electrophilic reactivity scores that accurately predict whether molecules will conjugate to DNA or protein. Fifth, we apply these molecule reactivity scores to calculate DNA and protein selectivity scores to estimate the fraction of molecules that are reactive to DNA and protein but not cyanide or GSH.

Of course, ultimately, the XenoSite reactivity model will be connected to models of drug metabolism to be most useful. Although out of the scope of this study, combined metabolism and reactivity models would be able to predict both bioactivation and subsequent toxicity of metabolites. Nonetheless, this study takes a significant step towards effectively managing the IADR and DILI risk of new medications with computational modeling.

3.2 Results and Discussion

In the initial section, we summarize a systematic effort to optimize the structure and training of the model, with the goal of choosing the best method for predicting reactivity. The following sections then investigate the performance of the final optimized model. First, we evaluated the model's cross-validated atom reactivity scores (ARS) by testing site of SOR classification performance within reactive molecules. Second, we compared ARS performance to that of

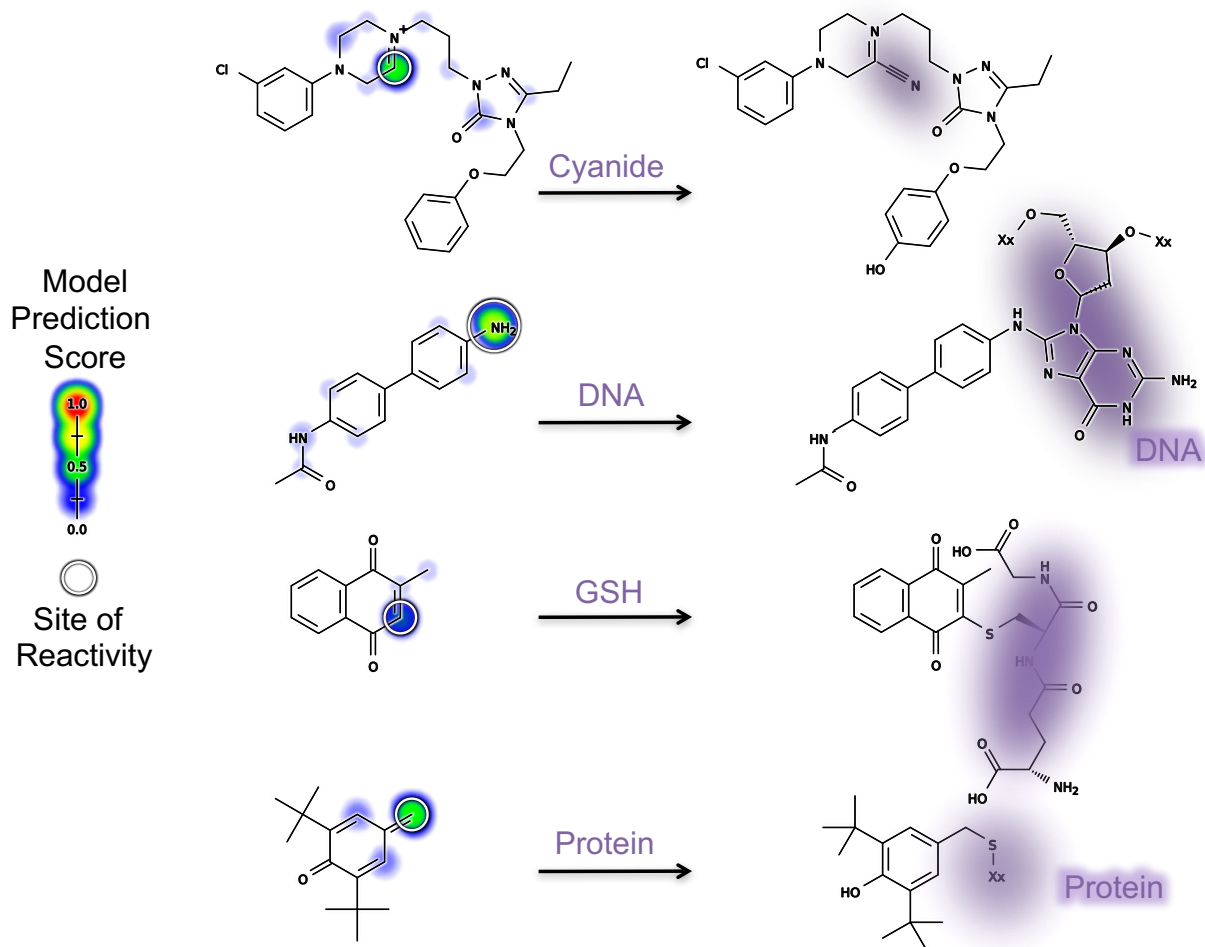


FIGURE 3.1: Examples of the four types of sites of electrophilic reactivity modeled. Predictions by the XenoSite reactivity model are indicated on the left with a colored shading gradient and white circle for each known site of reactivity. These predictions range from zero to one, indicating the probability that an atom is reactive with each of the four nucleophilic targets. From top to bottom, a cyanide conjugation of a nefazodone metabolite, [109] a DNA conjugation of *N*-acetylbenzidine, [110] a glutathione (GSH) conjugation of menadione,[111] and a protein conjugation of a butylated hydroxytoluene metabolite.[112] DNA and protein are inherently structurally diverse and thus cartoonized macromolecules are depicted, with the rest of each macromolecule represented by “Xx”.

atom-level quantum chemical reactivity indices. Third, we assessed ARS performance on an external test set. Fourth, we calculated the accuracy of the model’s cross-validated molecule reactivity scores (MRS) at predicting molecule reactivity. Fifth, we compared MRS performance to molecule-level quantum chemical reactivity indices. Sixth, we use the model to estimate the number of high throughput screening molecules that are reactive with macromolecules (DNA and protein), but are not flagged by small-molecule trapping agents (GSH and cyanide).

3.2.1 Model Optimization

Several experiments demonstrate how each of the innovations in our modeling approach improve performance. These experiments are briefly discussed here, and further details and data are available in Supporting Information. First, we hypothesized that jointly modeling several types of reactivity in a multitask learning model would improve predictions on the smaller data sets.[113, 114] Indeed, the multitask model outperformed the individual modeling approach at predicting cyanide and protein SOR (Figure S2). This is likely because the cyanide and protein reactivity tasks are the most difficult and, therefore, benefit most from integrated modeling. The cyanide dataset is difficult because it is small, and the protein dataset is both small and includes the most diverse mechanisms. The data reported herein reflects modeling all four types of reactivity together in a multi-task network, instead of building separate models for each task.

Second, a modular input layer was used to group related descriptors (such as the identities of all atoms a certain depth away), rather than a traditional three-layer neural network structure. We hypothesized that building explicit chemical knowledge into the model structure could reduce the total number of parameters in the model, thereby creating a simpler model with better generalizability. This possibility was inspired by several examples of modular neural networks in the literature.[115, 116, 117, 118, 119, 120] In fact, the modular structure did enable reduction of the total number of weights in the model by 50%, while retaining the same performance (Figure S3). This weight-reduced, modularly-structured model outperformed a traditionally-structured model with the same number of weights. Third, we found that including quantum

Network

chemical descriptors did not improve performance compared to a topological-descriptor-only model (Figure S4), so we did not include the quantum chemical descriptors in the final model. Construction of this topological-descriptor-only model was inspired by our previous model of GSH reactivity, which primarily relied upon topological descriptors rather than quantum chemical descriptors.[108] A common critique of neural networks is their opacity compared to more transparent methods with easily interpretable weights, like a logistic regressor. In response to this critique and to gain insight into the inner workings of our model, a permutation sensitivity analysis was performed (Figure S5). We have previously used this approach to expose the structure of similar models.[84, 108, 121] Details of the permutation sensitivity analysis are available in the Supporting Information. Fourth and finally, we found that including the negative epoxides in the training data substantially improved the model's ability to identify reactive epoxides (Figure S6).

3.2.2 Accuracy at Predicting Sites of Reactivity

A central objective of this study was to accurately predict SORs: the specific atom(s) within reactive molecules that covalently bind to nucleophilic sites within DNA and protein. The XenoSite reactivity model gives four reactivity scores to each atom (ARS) within a test molecule, each of which ranged from zero to one, and represented the probability that an atom is reactive with cyanide, DNA, GSH, or protein, respectively. Within a reactive molecule, a well-performing model should assign reactive atoms with higher scores than nonreactive atoms. The magnitude of the SOR values and identity of the target(s) of the reactive molecule sheds light on preferential chemistries leading to adduction and structures of the adducts. Such knowledge is relevant for reactive metabolite identification under experimental conditions and further study on the consequences of adducts on normal biological processes. Furthermore, SOR predictions indicate the probability of reactive hot spots leading to unfavorable adductions, and such knowledge could be leveraged to engineer rational modifications to reduce a molecule's reactivity, and potentially its toxicity.

Network

The accuracy of the ARS scores was assessed using ten replicates of ten-fold cross-validation, where groups of related molecules are held out in the same fold. Each replicate yielded very similar results, so for brevity we reported the results from the first one. Performance was quantified by two metrics. First, we calculated the average site AUC by computing the area under the receiver operating characteristic (ROC) curve (AUC) for each molecule and averaging the AUCs for each molecule in the data set.[108, 121] Second, we calculated the top-two metric, which is a standard for site of metabolism predictions. This approach considers a molecule correctly predicted if any of its SOR are predicted in the first or second rank positions.[121, 50, 122, 123, 52] Reactivity indices drawn from the quantum modeling literature are another method for predicting atom reactivity,[62, 63, 64, 65, 66] and thus, they provide an important point of comparison to our ARS for predicting SOR.

XenoSite's cross-validated ARS predicted reactive atoms with average site AUC accuracies of 96.6%, 89.8%, 92.8%, and 94.4%, and top-two accuracies of 83.9%, 80.6%, 80.9%, and 84.2%, for cyanide, DNA, GSH, and protein, respectively (Figure 3.2). These performances are very accurate, especially when compared to the current standard of atom-level reactivity indices (listed in Table 1) from quantum simulations.[62, 63, 64, 65, 66] Consistent with our previous model of GSH reactivity, ARS outperformed all reactivity indices tested across all four nucleophilic targets.[108] For example, for predicting DNA SOR, the best performing descriptor by both metrics was $\pi^S(r)$, with an average site AUC of 60.9% and a top-two accuracy of 27.6%, which are both significantly lower than ARS's corresponding performances of 89.8% and 80.6%. The machine learning approach we adopt here is strikingly more accurate than quantum chemical measures of reactivity.

As further evidence of generalizability, the model was applied to an external test of 14 molecules reactive with GSH that were collected from a newer version of the AMD than that used in training (Figure 3.2). The model predicted SOR on this test set with an average site AUC accuracy of 93.6%, matching the cross-validated performance of 92.8%. The top-two external test set performance was 64.3%, somewhat lower than the cross-validated performance

Network

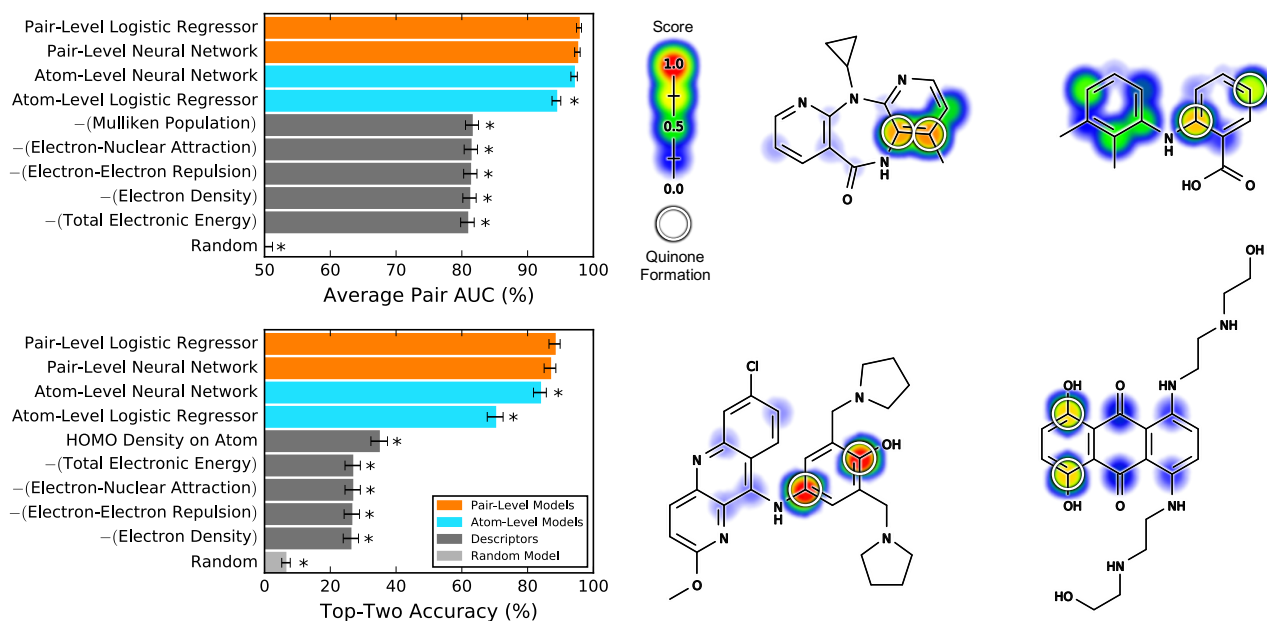


FIGURE 3.2: Atom reactivity scores (ARS) accurately identified sites of reactivity (SORs). The average site AUC (top left) and top-two (top center) metrics were computed for each of the four target nucleophiles and used to assess the cross-validated ARS for 1364 reactive molecules extracted from the Accelrys Metabolite Database (AMD) with their SORs labeled. ARS outperformed all quantum chemical descriptors tested for each target nucleophile. Bottom, from the latest AMD release, an external test of 14 molecules was extracted and SOR(s) labeled for GSH. Performance is equivalent to that of the cross-validated predictions as measured by both the average site AUC (bottom left) and the top-two (bottom center) metrics. Right, two example molecules from the external test set are visualized with their predictions, indicated by the colored shading. Each site of reactivity is labeled with a white circle. Top right, an antimalarial drug candidate metabolite,^[124] and bottom right, gefitinib.^[125]

of 80.9%. The lower performance is well explained by the high variance of the estimate, because there are only a few molecules in the set. Moreover, the performance of the model is still substantially better than the quantum chemical descriptors, which are indistinguishable from zero performance on the test molecules. Depictions of all 14 external test set molecules with their GSH ARS and SOR are available in the Supporting Information, alongside their most closely similar pair from the training set (Figure S7).

3.2.3 Accuracy at Identifying Reactive Molecules

Another central goal is to accurately discriminate between reactive and nonreactive molecules. XenoSite predicted four reactivity scores for each molecule (MRS), each of which ranged from zero to one, and represented the probability that the molecule is reactive with cyanide, DNA, GSH, or protein, respectively. The accuracies of MRS were measured by ten-fold cross-validation, with performance quantified by the area under the ROC curve (molecule AUC). The MRS scores were reasonably accurate in separating reactive and nonreactive molecules, with molecule AUCs of 90.3%, 78.7%, 77.7%, and 79.8% for cyanide, DNA, GSH, and protein, respectively (Figure 3.3). In contrast, the performances of the best reactivity indices using traditional quantum descriptors were much lower: 84.9% ($-E_{\text{HOMO}}$ for cyanide), 73.7% ($\max [D^N(r)]$ for DNA), 62.9% ($\pi^S(r)$ for GSH), and 65.1% ($\max [D^N(r)]$) for protein). For each nucleophilic target, MRS outperformed all quantum chemical descriptors.

The model's MRS scores were superior to other methods, yet still were lower accuracy than the ARS scores. This discrepancy was likely due to the presence of more noise in the molecule-level data than in the atom-level data. The ascertainment bias in the literature means that many of the negative molecules are actually reactive. When extracting nonreactive molecules, we assumed that molecules were not reactive to a nucleophile if they were not reported to be reactive to that nucleophile by any reactions in the AMD. This assumption was not solid evidence of molecules' nonreactivity, because many studies do not test for reactivity. By contrast, the atom-level reactivity data were all drawn from experiments that tested for reactivity, and thus were much less noisy with many fewer false negatives. A similar drop from atom- to molecule-level accuracy was observed in our previous studies with the literature derived data, including the GSH reactivity model that was a precursor to this work.[108, 121] In the future, we plan to improve the data by experimentally testing the reactivity of "nonreactive" molecules that nonetheless receive high MRS. We expect that some of these false positives will be shown to be in fact reactive, and that this effort will validate the models and improve the training data.

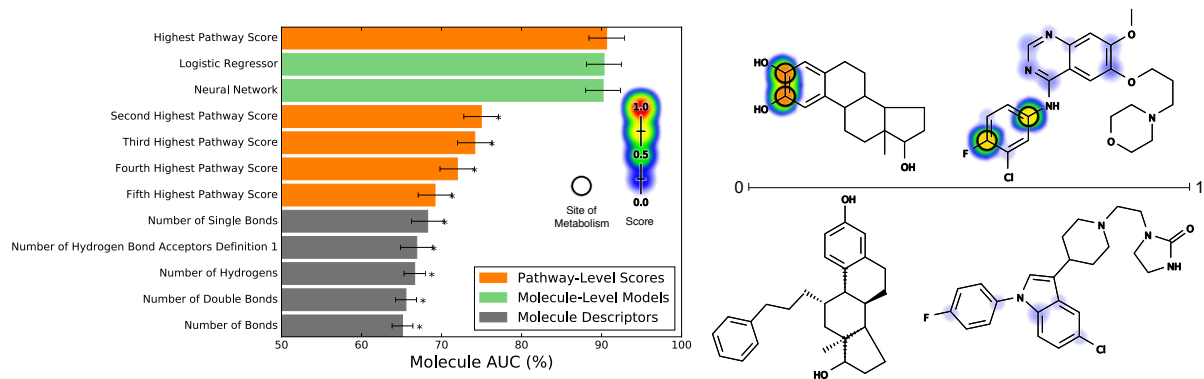


FIGURE 3.3: Molecule reactivity scores (MRS) accurately identified reactive molecules. A number of prediction methods were compared by their performance at identifying reactive molecules. Each nonreactive molecule is structurally similar to a reactive molecule. To measure performance, the area under the ROC curve was calculated (molecule AUC). For each target nucleophile, MRS outperformed all quantum chemical descriptors tested. Right, four example molecules are visualized with their protein and GSH reactivity predictions indicated by colored shading. From top to bottom, a TRPV1 antagonist metabolite,^[126] a nitroso metabolite of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (formed during cooking of meat),^[127] a nicotine iminium ion metabolite (reactive with cyanide),^[128] and a benoxaprofen acyl glucuronide metabolite (a non-steroidal anti-inflammatory drug withdrawn due to hepatotoxicity).^[129] Sites of reactivity are indicated by white circles.

3.2.4 Molecules Missed by Standard Screening Assays

Experimental studies with cyanide and GSH are traditional screening tools to trap hard and soft electrophilic reactive molecules, respectively, and thus provide potential insights on possible reactions between the reactive molecules and biological macromolecules.[105, 106] Importantly, cyanide and GSH possess only a single type of nucleophilic site, while biologically relevant macromolecules often contain both hard and soft nucleophiles with an array of different structures and hence different chemistries. Consequently, nucleophilic trapping assays using cyanide and GSH may not adequately reflect all possible reactions observed within biological macromolecules, and thus fail to detect potentially toxic electrophiles.

We estimated how frequently molecules will react with macromolecules (protein and DNA) but not with trapping agents (GSH and cyanide) commonly used in experimental screens. We calculated the probability that each molecule will form adducts to either DNA or protein, but neither cyanide nor GSH. The DNA probability was computed by taking the product of its probabilities of reacting with DNA (MRS_{DNA}) and not reacting with cyanide ($1 - MRS_{CN}$) or GSH ($1 - MRS_{GSH}$). The protein probability was computed by taking the product of its probabilities of reacting with protein (MRS_{PRO}) and not reacting with cyanide ($1 - MRS_{CN}$) or GSH ($1 - MRS_{GSH}$). Next, we summed up the probabilities for all molecules with respect to each macromolecule. These scores are well scaled probabilities, so this sum is a valid estimate of the total number of molecules selectively reacting with each macromolecule and not the traditional nucleophilic traps, and hence those that would be missed by standard screening approaches. For the entire data set of 2803 molecules, this approach yielded totals of 257 and 227 molecules predicted to be reactive only with DNA and protein, respectively, and not cyanide or GSH. This finding suggests that in our data set 9.2% of DNA reactive molecules and 8.1% of protein reactive molecules would be missed by standard screening assays with cyanide and GSH. Though not the majority of molecules, this result suggests there are shortcomings of using small-molecule trapping agents to infer reactivity with macromolecules. Computationally

Network

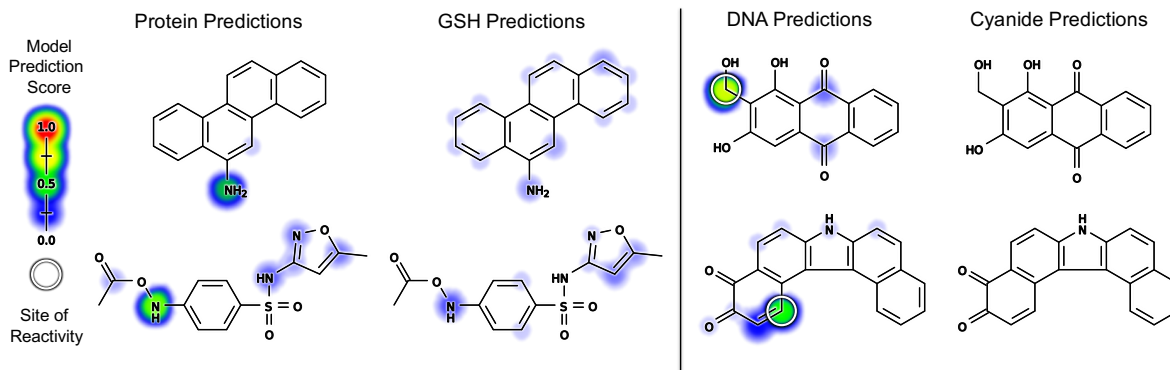


FIGURE 3.4: Molecule reactivity scores (MRS) predict some molecules significantly more reactive with biological macromolecules than nucleophilic traps used in standard screening assays. Discordant predictions between the protein and GSH models (left), and between the DNA and cyanide models (right) are visualized. As indicated both by the colored shading (corresponding to atom reactivity scores) and by the MRS listed below, for each molecule the biological macromolecule (protein or DNA) prediction is significantly greater than the corresponding standard trapping agent (GSH or cyanide) prediction. Experimentally-known sites of reactivity are labeled with white circles. Upper left pair: a metabolite of 6-nitrochrysene ($MRS_{\text{PRO}}: 0.42$, $MRS_{\text{GSH}}: 0.29$), an environmental pollutant.[130] Bottom left pair: *N*-acetoxy-sulfamethoxazole ($MRS_{\text{PRO}}: 0.51$, $MRS_{\text{GSH}}: 0.38$), a metabolite that may mediate the toxicity of the antibiotic sulfamethoxazole.[131] Upper right pair: a metabolite of lucidin-3-O-primeveroside ($MRS_{\text{DNA}}: 0.68$, $MRS_{\text{CN}}: 0.26$), a component of a food dye known to be carcinogenic in rats.[132] Bottom right pair: 7*H*-dibenzo(*c,g*)carbazole-3,4-dione ($MRS_{\text{DNA}}: 0.64$, $MRS_{\text{CN}}: 0.30$), an environmental multispecies carcinogen.[133]

modeling could fill a gap here, by identifying these problematic molecules when screening cannot (Figure 3.4). The experimental validation of specific missed reactive molecules and these estimates is outside the scope of this study, but is planned in our future work.

3.3 Model Limitations

For predicting reactive metabolites, this study's approach is limited by not modeling their metabolic activation, which may precede reactions with nucleophilic traps. Nevertheless, models for metabolic reactions are rapidly maturing as evidenced by our work [121, 50, 134, 135, 136] and others, [54, 137, 138] and it is conceivable that metabolism and reactivity models could be combined for a more biologically relevant prediction of risks posed by drugs and

other compounds. An additional caveat is that some reactions in the database are likely missing reactive intermediates. For example, acyl glucuronides are predicted somewhat reactive, yet are known not to be reactive themselves. Instead, only after acyl migration and ring opening do acyl glucuronides form short-lived, reactive open-chain aldehyde intermediates that can lead to covalent binding to molecules.[139] There are likely other uncertainties where intermediate reactive metabolites are not present in the database, as they may be so reactive that they are too short-lived to be observed experimentally. Missing intermediates is a limitation of the data, but is both a strength and weakness of our method. On the plus side, we implicitly model some types of rearrangements that yield reactive species, and in doing so, expand the utility of the model. Nevertheless, this process leads to motifs that are not reactive themselves being predicted reactive. Therefore, we are actually modeling a combination of intrinsic reactivity and potential for covalent binding. A final caveat is that the domain of applicability of XenoSite is likely limited to drug-like molecules. We do not expect the model to generalize well to molecules outside of this domain, such as to inorganic molecules. Precise assessment of the domain of applicability is a critical question for future work. We plan to directly test this with prospective experiments, which is much more convincing than computational strategies for assessing the applicability domain.

3.4 Conclusion

This study establishes that SOR modeling accurately predicts a key driver of drug toxicity: covalent binding to DNA and protein. The XenoSite reactivity model identifies SOR within reactive molecules with AUC performances of 96.6%, 89.8%, 92.8%, and 94.4%, for cyanide, DNA, GSH, and protein, respectively. Compared to building separate models, collectively modeling reactivity for all four nucleophilic targets enabled knowledge transfer between tasks, improving SOR predictions for cyanide and protein. For cyanide, DNA, GSH, and protein, the model separates reactive and nonreactive molecules with, respectively, 90.3%, 78.7%, 77.7%, and 79.8%

AUC. These predictions of molecular reactivity can highlight potentially toxic molecules that might be missed in the early stages of drug development. Ultimately, to accurately predict reactive metabolites, both metabolism and reactivity must be modeled. While there has been significant progress on metabolism modeling, [121, 50, 134, 135, 136, 54, 137, 138, 140] reactivity modeling has lagged far behind. By accurately modeling reactivity across a broad range of chemical space for biologically relevant macromolecules, such as DNA and protein, this study contributes a fundamental component of an integrative model of metabolism and reactivity.

3.5 Methods

3.5.1 Site of Reactivity Training Data

We assembled a training data set from the December 2014 release of the literature-derived Accelrys Metabolite Database (AMD). From 2489 total reactions, molecules were extracted based on reported reactions with cyanide, DNA, GSH, or protein. For each target nucleophile, the reactive atom(s) within each reactive molecule were marked based on the structures of the starting and product molecules, using an automated algorithm constructed using the RDKit python library.[141] Topologically equivalent atoms were found using the Pybel python library, and atoms equivalent to reactive atoms were themselves labeled as reactive.[82] The final data set included, in total, 1364 electrophilic molecules with their reactive atoms and SORs marked. This data set was composed of 51, 145, 1059, and 120 molecules known to be reactive with cyanide, DNA, GSH, or protein, respectively. For each of the four reactive nucleophilic targets, the atoms across the whole data set were labeled as reactive or nonreactive. Some atoms were marked reactive with more than one nucleophile.

Metabolically related but nonreactive molecules were extracted from the reaction network of each reactive molecule. Metabolic sibling and parent molecules were extracted from this network, while excluding molecules themselves known to be reactive. We also included 63 naturally occurring[142] and known nonreactive[143] epoxides as nonreactive molecules. While

epoxides are generally quite reactive, they can be stable in certain cases; however, a shortcoming of our previous method for predicting GSH reactivity was that it predicted all epoxides reactive.[108] These molecules were added to mitigate this shortcoming.

A total of 1439 molecules were labeled nonreactive, with 103, 248, 1269, and 255 labeled nonreactive to cyanide, DNA, GSH, and protein, respectively. At the molecule-level, for each of the four reactive nucleophilic targets, molecules across the entire data set were labeled as reactive or nonreactive. Some molecules were marked reactive or nonreactive with more than one nucleophile.

To enable replication of our work, we included all AMD molecule registry numbers in the Supporting Information, in addition to the SMILES strings of the nonreactive epoxides. Unfortunately, the AMD license precluded us from disseminating the rest of the molecule structures themselves.

3.5.2 External Site of Reactivity Test Data

After assembling the training data, a new version of the AMD was released (June 2015) with several new reactions, which was used as an external data set. We filtered out all molecules already labeled as reactive in the training data. This process left only 14 new molecules that reacted with GSH and no new molecules reacting with cyanide, DNA, or protein. These 14 molecules were completely withheld from model training and optimization decisions, and were only tested with the final model.

3.5.3 Quantum Chemical Reactivity Indices

Several quantum chemical descriptors were calculated from self-consistent field computations with MOPAC, a quantum chemistry package, using the semi-empirical method PM7 and the COSMO implicit solvent model.[73, 74] These included descriptors that have been previously proposed as reactivity indices, on both the atom- and molecule-level (Table 3.1).[70, 71, 72]

TABLE 3.1: Quantum Chemical Reactivity Descriptors.

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

Atom-level descriptors include the nucleophilic ($D^N(r)$) and electrophilic ($D^E(r)$) delocalizabilities, which are also known as fukui reactivity indices, as well as self-polarizability ($\pi^S(r)$). The maximum of each of these ($\max [D^N(r)]$, $\max [D^E(r)]$, and $\max [\pi^S(r)]$) were used as molecule descriptors, which also included the energies of the lowest unoccupied and highest occupied molecular orbitals (E_{LUMO} and E_{HOMO}). The performances of the final reactivity model—which only used topological descriptors—were compared to those of the quantum chemical indices.

3.5.4 Topological Descriptors

The reactivity model used 15 molecule-level and 194 atom-level topological descriptors, each of which describes a chemical property. An in-house python script calculated these descriptors from the structure of each molecule. This script derives these descriptors from several sources, such as each molecule’s connectivity distance matrix, periodic table properties, or motif patterns defined by Pybel.[82] The majority of our topological descriptors have been shown to be useful for the XenoSite metabolism, reactivity, and epoxidation models.[50, 108, 121] In this study, we used an expanded set of topological descriptors, which slightly improved performance in comparison to the previous set of topological descriptors (Figure S1). The full list of descriptors is available in the Supporting Information (Table S1 and Table S2), as well as a list of new

topological descriptors added for this project.

3.5.5 Combined Atom- and Molecule-Level Reactivity Model

We collectively modeled reactivity to cyanide, DNA, GSH, and protein at both the atom- and molecule-level by building a deep neural network using an in-house machine learning python library. The architecture included a molecule layer, an input layer, two hidden layers, and two output layers (Figure 3.5). The atom output layer computed atom-level SOR predictions against each nucleophile, while the molecule output layer computed molecule reactivity scores (MRS) for each nucleophile. The corresponding scores predicted the chances that a test molecule is reactive against that nucleophile.

We trained this network in two stages. First, the atom-level network was trained to produce atom reactivity scores (ARS). For this process, each atom was considered a possible SOR. A numerical vector was associated with each atom, which contained all of the descriptors for that atom. A binary target vector indicated if each atom was a SOR for each of the nucleophilic targets. For molecules of unknown reactivity against each nucleophilic target, the appropriate values were empty, neither one nor zero. Using gradient descent on the cross-entropy error, the weights of the network were trained such that SOR scored higher ARS than other atoms. Four ARS were produced for each atom, each ranging from zero to one, and representing the probability that the atom will be reactive with cyanide, DNA, GSH, or protein.

Second, the molecule-output nodes were trained to compute MRS. For this stage, each molecule was considered possibly reactive, and a numerical vector of descriptors was associated with each molecule. Descriptors included all molecule-level descriptors, as well as the top five ARS for each of the four nucleophilic targets, corresponding to the scores of the five atoms predicted within a molecule to be the most reactive to each nucleophile. For each molecule, a binary target vector (with unknown components empty) indicated if the molecule was reactive to each of the nucleophilic targets. For each nucleophile, a logistic regressor found a scoring

Network

function that gave reactive molecules high MRS and nonreactive molecules low MRS, ranging from zero to one.

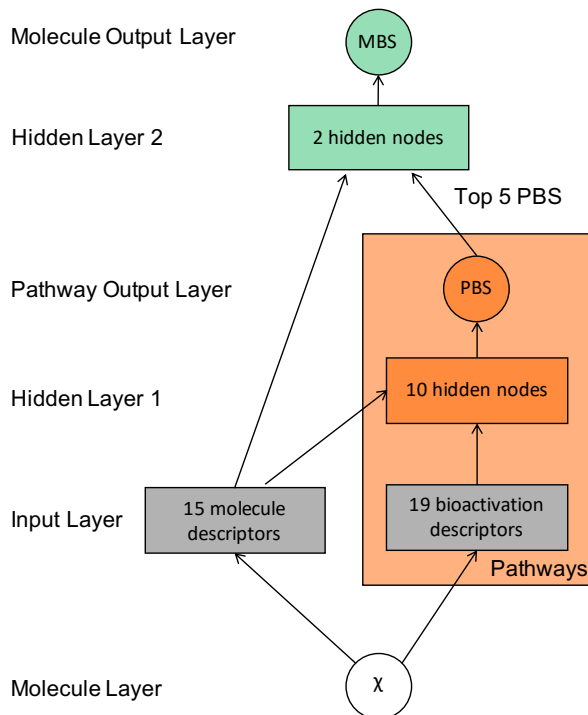


FIGURE 3.5: The structure of the XenoSite reactivity model. This diagram illustrates how information flowed through the model, which consisted of one input layer, two hidden layers, and two output layers. By simultaneously modeling all types of reactivity, the model was able to transfer knowledge between related tasks, thereby improving performance substantially over independent models. The model computed atom-level predictions for reactivity to each of four nucleophilic targets: cyanide (ARS_{CN}), DNA (ARS_{DNA}), GSH (ARS_{GSH}), and protein (ARS_{PRO}), collectively referred to as atom reactivity scores (ARS). Additionally, the model computed molecule reactivity scores (MRS): MRS_{CN} , MRS_{DNA} , MRS_{GSH} , and MRS_{PRO} , which predicted the chances of a molecule's reactivity to each of the four nucleophilic targets, respectively. From the structure of an input model χ , 15 molecule-level and 194 atom-level descriptors were calculated. Some chemically related descriptors, such as neighbor atom identities, were grouped in the first hidden layer (with 30 nodes). Grouped and ungrouped nodes were inputted into the second hidden layer (with 17 nodes), which outputted four atom-level scores. Finally, for each of the four nucleophilic targets, the respective MRS was computed from the top five ARS for each of the four nucleophilic targets, corresponding to the scores of the five atoms predicted within a molecule to be the most reactive to each nucleophile, as well as all molecule-level descriptors. The diagram is condensed and displays one representative molecule input node, five atom input nodes, and two nodes for each hidden layer. The molecule input node is a chemical structure; all other nodes are vectors of real numbers computed from nodes or layers from which there are incoming connections.

Chapter 4

Modeling Epoxidation of Drug-like molecules with a Deep Machine Learning Network

Reprinted (adapted) with permission from Hughes, T. B., Miller, G. P., and Swamidass, S. J. (2015). Modeling Epoxidation of Drug-like Molecules with a Deep Machine Learning Network. ACS Central Science 1(4), 168-180. The original article was published at <https://pubs.acs.org/doi/10.1021/acscentsci.5b00100>. Permission requests related to adapted or original material should be directed to the ACS.

4.1 Introduction

Drug discovery and development involves significant efforts to identify safe and efficacious drugs; nevertheless, unanticipated toxicity and adverse drug reactions do occur and cause approximately 40% of drug candidates to fail.[10] Frequently, these harmful outcomes are linked to the formation of electrophilic metabolites that covalently bind to proteins or DNA and, in some cases, elicit an immune response in susceptible patients [19, 46, 47, 48, 49] One of the most common types of reactive metabolites are epoxides, the subject of this study.

Epoxides are three membered cyclic ethers and are often highly reactive due to ring tension and polarized carbon-oxygen bonds.[144, 145, 146, 147, 148] Epoxides are formed by cytochromes P450 acting on aromatic or double bonds,[149, 150] and these epoxidation reactions comprise around 10%[151] to 15%[152] of all bioactivation reactions. Biological defense

Network

mechanisms to epoxides, include glutathione conjugation and cleavage by epoxide hydrolase, offer only partial protection.[148, 144, 153, 154] Glutathione can be depleted,[155, 156] and certain products of glutathione conjugation[154] and epoxide hydrolase [157, 158] are themselves toxic.

Epoxide metabolites often drive toxicity for drugs, and accurate strategies for anticipating the formation of epoxides are critical in drug development. Knowledge of epoxide formation aids assessment of drug candidates. Furthermore, the identity of the specific bond in a molecule undergoing epoxidation, its site of epoxidation (SOE), could enable rational modification of the molecule to reduce risk of reactive metabolite formation. An example of how this knowledge can lead to drugs with improved safety is illustrated by carbamazepine (Figure 4.1). The metabolism of this anti-epileptic drug forms carbamazepine-10,11-epoxide. Carbamazepine metabolism can also form an iminoquinone,[159] but the epoxide's formation is the focus of this study and more correlated with adverse reactions.[160, 161, 162] The molecular mechanism for this response involves reactions between the epoxide and proteins to form adducts [163]. However, the epoxide formation can be blocked by modifying carbamazepine's SOE. For example, oxcarbazepine[160] or eslicarbazepine are analogues of carbamazepine that are no longer epoxidized.[162] While oxcarbazepine and eslicarbazepine were not prospectively designed in order to reduce epoxide formation, they demonstrate how small molecular changes can significantly impact toxicity caused by epoxide metabolites. These analogues retain the same mechanism of action as carbamazepine, yet have a lower incidence of adverse effects because they prevent the formation of epoxides.[164, 162]

A number of studies, including those by our group, have established that computational methods can predict the sites at which molecules are metabolized.[50, 51, 52, 36, 53, 54] A shortcoming of those approaches has been the lack of predictions for the actual metabolites generated by those reactions. Cytochromes P450 catalyze many different types of oxidative reactions, including commonly observed hydroxylations.[165, 149, 52] While several cytochromes P450 site of metabolism models are reported in the literature, to the best of our knowledge, none

of those models specifically identify SOEs in molecules. Instead, all existing methods only report which atoms undergo oxidation, without distinguishing the specific type of reaction—such as epoxidation or hydroxylation—or the resulting modification to the structure.

In this study, we construct an epoxidation model—based on the structural data of several hundred diverse molecules—that is successful at three key objectives. First, the model accurately predicts SOE within epoxidized molecules; these SOE predictions can be used to direct structural modifications to drug candidates. Second, the model distinguishes SOE from sites of sp^2 hydroxylation (SOH), a key negative control. Both SOEs and SOHs are oxidized by P450s, and we expect a useful model to correctly identify which of these oxidations give rise to epoxides. In contrast, commonly reported P450 site of metabolism models will not distinguish these two cases and report both as sites of metabolism. Third, the model identifies which molecules are metabolized into epoxides, separating these molecules from closely related molecules that are not epoxidized. This enables rapid screening of drug candidates for molecules that are potentially toxic due to epoxidation.

4.2 Methods

4.2.1 Epoxidation Training Data

We mined a large, chemically diverse training data set from the Accelrys Metabolite Database (AMD), which includes a collection of metabolic reactions drawn from the literature.[168] A total of 702 reactions were extracted, each of which takes place in humans, human cells, or human microsomes, and is classified as epoxidation. Due to the short half-life of many epoxides, however, some product molecules do not explicitly contain an epoxide. Instead, an epoxidation product may be a dihydrodiol or a DNA, glutathione, or protein conjugate (Figure 4.2).[169, 170] An automated labeling algorithm used these motifs to label SOEs on the starting molecule of each reaction.

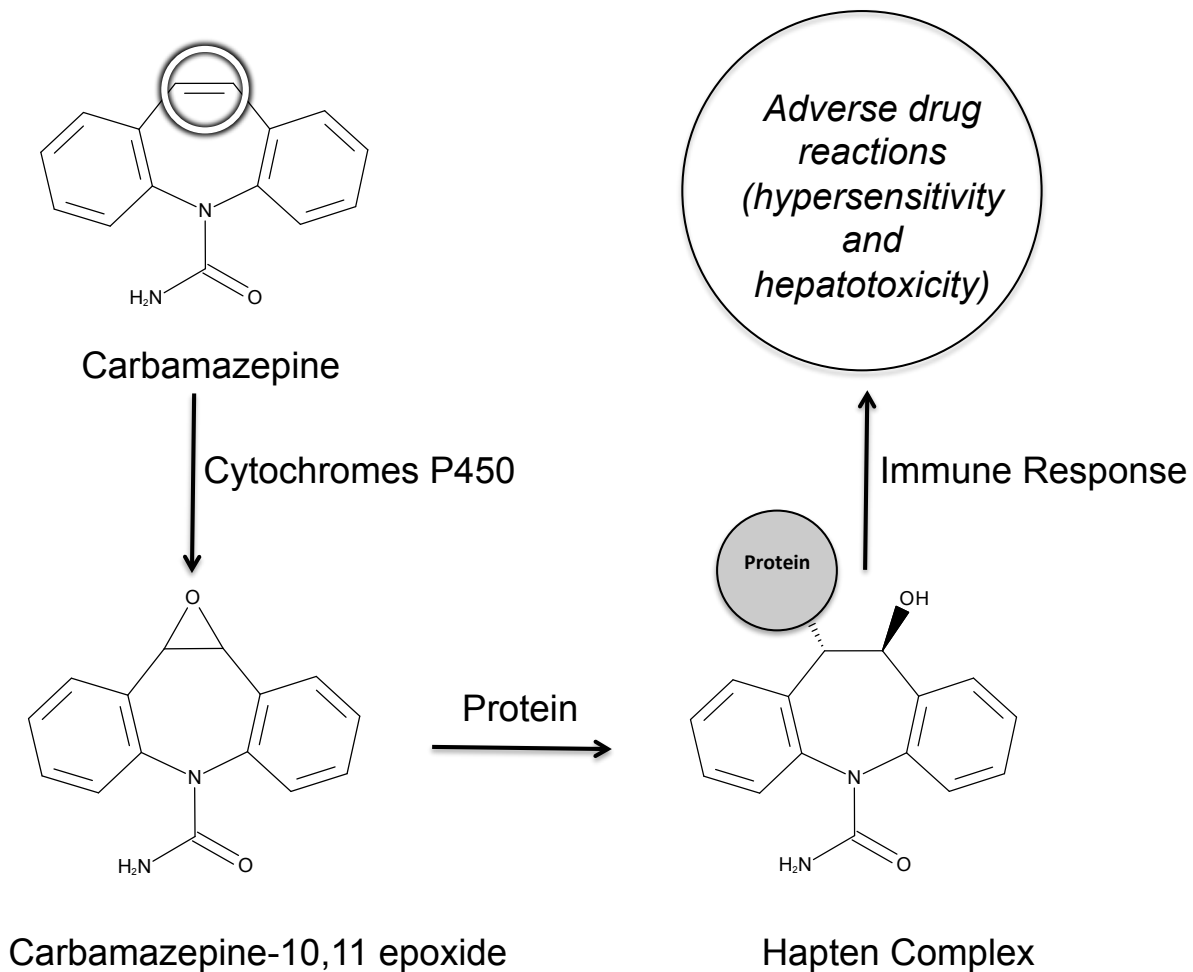


FIGURE 4.1: Adverse drug reactions are often caused by reactive metabolites. For example, carbamazepine is metabolized by cytochromes P450 to carbamazepine-10,11-epoxide. Carbamazepine metabolism can also form an iminoquinone [159], but the epoxide's formation is the focus of this study and more correlated with adverse reactions [160, 161, 162]. The epoxide is electrophilically reactive and covalently binds to nucleophilic sites within proteins. The resulting adduct serves as a hapten complex and elicits an immune response. This mechanism is thought to be responsible for many carbamazepine adverse reactions [166, 167]. This site of epoxidation is circled on carbamazepine.

In this study, we defined a SOE as the bond between the two carbons to which an epoxide forms and identified these bonds in depictions by circling the two adjoining atoms. When bonds were topologically equivalent to observed SOEs, as identified using the Pybel python library, they were themselves labeled as SOE.[82] Duplicate starting molecules were identified by canonical SMILES, and merged into a single training example with all observed SOEs labeled. The final data set included 389 epoxidized molecules, each with its SOEs labeled. These epoxidized molecules included 411 aromatic bond SOEs and 168 double bond SOE. Additionally, 20 single bond SOEs were included; the labeling of single bonds as SOEs is likely due to rearrangements or intermediates—absent from the database—allowing epoxidation to occur at an aromatic or double bond.

We also identified structurally similar but non-epoxidized molecules. These target compounds were mined from the reaction network for each previously identified epoxidized molecule. This strategy ensured the inclusion of the metabolic parent and sibling molecules so that a robust distinction between molecules undergoing epoxidation and those that are not became possible. After excluding molecules already classified as epoxidized, the remaining 135 molecules were marked non-epoxidized. Each one was metabolically studied and chemically similar to an epoxidized molecule in the data set.

Our license for the AMD data did not allow us to disclose the structures of the full data set. However, all molecule registry numbers are included in the Supporting Information, and this is sufficient data to rebuild the database and reproduce our results.

4.2.2 Hydroxylation Negative Control Data

As discussed in the Introduction section, sp^2 sites can be either epoxidized or hydroxylated. An epoxidation model must be validated using hydroxylation data as a negative control to distinguish the epoxidation model from a general oxidation model. An epoxidation model should rank SOEs above SOHs, whereas an oxidation model would rank them approximately equally. For use as negative controls, we also extracted SOHs from the AMD. Both SOHs and SOEs are

acted on by cytochromes P450, but the epoxides formed from SOEs are more likely to be toxic. To build a hydroxylation test data set, 3000 human hydroxylation reactions were randomly sampled from the AMD. We filtered out sp^3 hydroxylations, and any SOHs that included non-carbon atoms, both of which are easily distinguishable from epoxidations. After these filtrations, 1105 hydroxylations remained. Duplicate starting molecules were identified by canonical SMILES, and merged by labeling all known SOHs for each molecule. This final data set included 811 molecules, each with atoms labeled if they are SOHs, and bonds adjacent to these atoms labeled as bonds of hydroxylation.

4.2.3 Descriptors

Our approach used information encoded in descriptors for each bond to assess its susceptibility to epoxidation. Each bond was associated with a total of 214 numerical descriptors, including atom-level, bond-level, and molecule-level descriptors. Descriptors were calculated by in-house software that took as input SDF files with explicit hydrogens and 3D coordinates created by Open Babel.[68] The majority of our descriptors were atom-level descriptors previously developed for the XenoSite metabolism model[50] and the XenoSite reactivity model.[108] Each bond contained 89 descriptors from its “left” atom and its “right” atom. To prevent representation bias due to atom ordering, left and right atom assignment was randomized on a bond-by-bond basis. Twenty-three molecule-level descriptors, reported in our prior work, were also computed and used by the network to make predictions.

We supplemented these atom and molecule descriptors with bond descriptors developed specifically to capture the chemical properties of bonds. These 13 new bond descriptors are summarized in Table 4.1. There were two types of bond descriptors. First, topological bond descriptors summarized information from the molecular 2D structure. Second, quantum chemical descriptors were calculated from self-consistent field computations by MOPAC, a semi-empirical quantum chemistry modeler, utilizing an implicit solvent model and the PM7 force field.[73, 74]

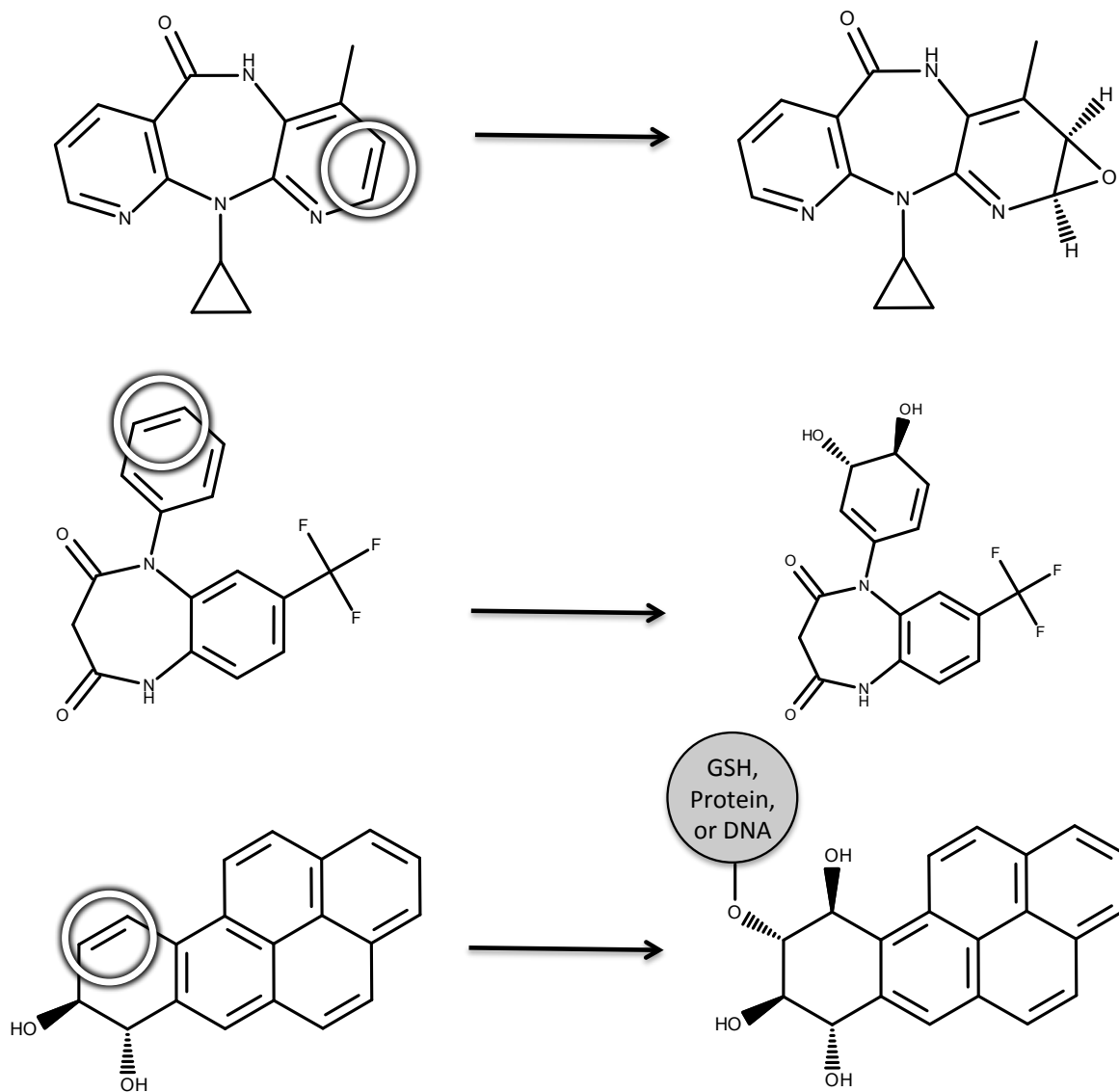


FIGURE 4.2: In the database, each epoxidation reaction acting on a site of epoxidation (abbreviated SOE and circled) forms an epoxide, dihydrodiol, or a conjugate adjacent to a hydroxylation. For example, the epoxidation reaction of nevirapine forms an epoxide (top),^[171] of N-desmethyl triflurazepam forms a dihydrodiol (middle),^[172] and of benzo(a)pyrene forms a DNA conjugate adjacent to a hydroxylation (bottom).^[170] The first case explicitly records the epoxide, while the other two record a tell-tale signature of a transient, reactive epoxide that is not directly observed.^[169, 170] A total of 702 human epoxidation reactions were identified in the Accelrys Metabolite Database. An automated labeling algorithm labeled SOEs on the starting molecule of each reaction based on these motifs.

Topological Bond Descriptors	
Single Bond	binary value indicating whether bond is a single bond
Aromatic Bond	binary value indicating whether bond is an aromatic bond
Double Bond	binary value indicating whether bond is a double bond
Conjugated Bond	binary value indicating whether bond is conjugated
Triple Bond	binary value indicating whether bond is a triple bond
Topologically Equivalent	number of topologically equivalent bonds in the same molecule
Quantum Chemical Bond Descriptors	
Bond Length	distance between the two atoms of a bond
σ_s occupancy	interaction of s-orbital σ bond electrons
σ_p occupancy	interaction of p-orbital σ bond electrons
π_p occupancy	interaction of p-orbital π bond electrons
σ_s - σ_p occupancy	interaction of s-orbital σ bond and p-orbital σ bond electrons
σ_s - π_p occupancy	interaction of s-orbital σ bond and p-orbital π bond electrons
σ_p - π_p occupancy	interaction of p-orbital σ bond and p-orbital π bond electrons

TABLE 4.1: .

Several bond descriptors were developed for the epoxidation model, using both topological and quantum chemical information.

In total, 214 numbers were used to describe each bond; 89 atom descriptors for the “left” atom, 89 for the “right” atom, 23 molecule descriptors, and 13 bond specific descriptors.

4.2.4 Combined Atom- and Molecule-Level Epoxidation Model

We built a model for bond and molecule epoxidation using a deep neural network with one input layer, two hidden layers and two output layers (Figure 4.3). The top-level output layer computed molecule-level predictions called the molecule epoxidation scores (MES); the next output layer computed bond-level predictions called the bond epoxidation scores (BES). Here, the term “deep network” does not mean a deep autoencoder network as is being increasingly used.[173] Instead, we mean a deep convolution network, with many more layers than a standard network and extensive weight sharing between replicates of the BES network.[174] This network was trained in two stages.

First, we trained the bond-level network to compute accurate BES values. In this training,

Network

each bond within a molecule was considered a possible SOE. Each bond had a vector of numbers (descriptors), with each entry of the vector describing a chemical property of that bond. The data set was a matrix, structured as one column per descriptor, and one row per bond. A final binary target vector labeled experimentally-observed SOEs with a 1. The weights of the network were trained using gradient descent on the cross-entropy error, so that SOEs scored higher BES than other bonds. These BES ranged from zero to one, representing the probability that a bond was a SOE.

Second, the molecule-level output layer was trained to compute MES values. Several versions of this output layer were considered, including another multilayer neural network, a logistic regressor, and a max layer that computed the MES as the maximum BES observed in the molecule. The logistic regressor and neural network took as input the top five BES, as well as all molecule-level descriptors. As we will see, both the neural network and the logistic regressor offer better scaled predictions with higher classification performances than the simpler max layer.

4.3 Results and Discussion

The following sections study the classification performance and inner workings of the epoxidation model. First, we evaluated the ability of BES to predict the SOE of epoxidized molecules. Second, we considered the credibility of the model by analyzing which descriptors are most important to the model's performance. Third, we increased resolution on the quality of the model predictions by calculating classification performance on aromatic and double bonds individually. Fourth, we asked whether BES distinguish SOEs from sites of sp^2 hydroxylation, because both epoxidation and sp^2 hydroxylation are catalyzed by P450s, but have significantly different implications for toxicity. Fifth, we tested how well MES separated epoxidized and non-epoxidized molecules. Finally, we studied how the model could direct drug modifications to reduce toxicity of known drugs.

Network

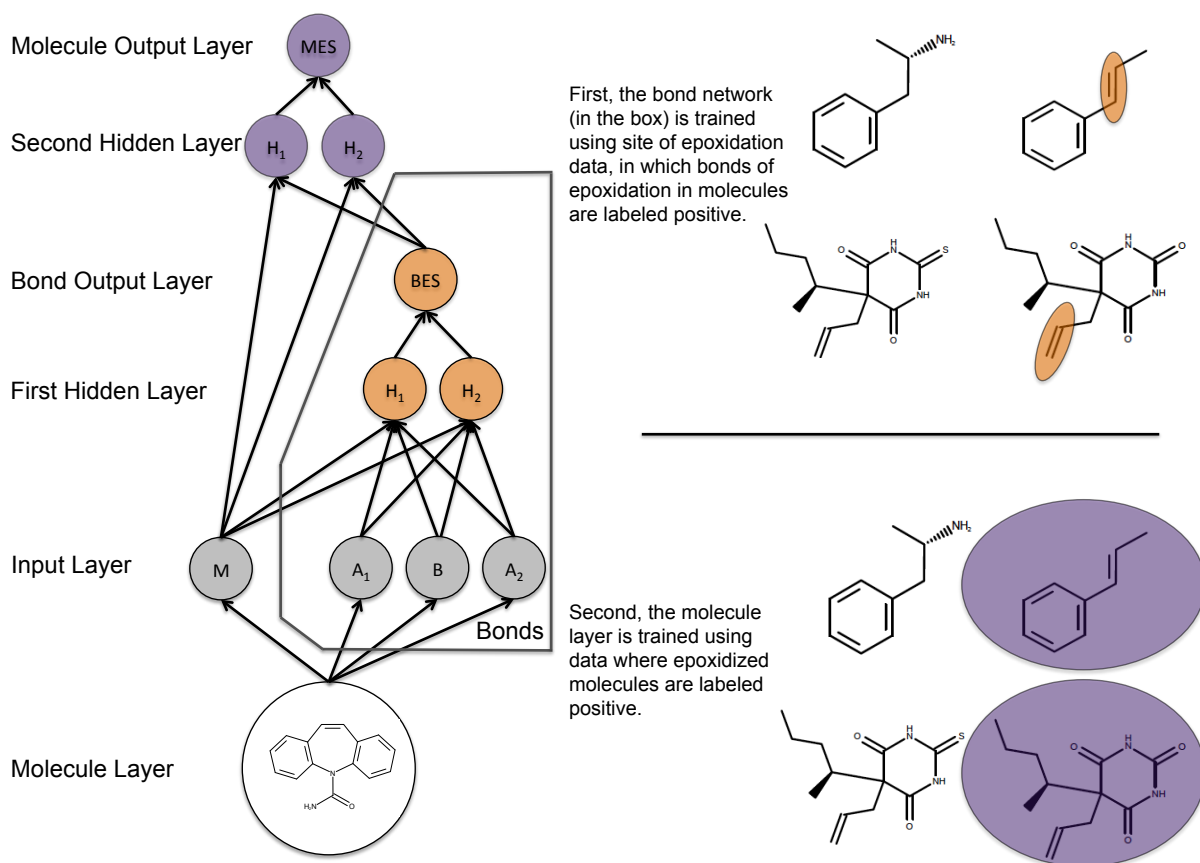


FIGURE 4.3: The structure of the epoxidation model. This diagram shows how information flowed through the model, which was composed of one input layer, two hidden layers and two output layers. This model computed a molecule-level prediction for each test molecule as well as predictions for each bond within that test molecule. From the 3D structure of an input molecule, 23 molecule-level and 191 bond-associated descriptors were calculated. These inputs nodes are inputted into the first hidden layer (with 10 nodes), which outputs a bond epoxidation score (BES) for each bond in the molecule. The BES quantifies the probability that the bond is a site of epoxidation. The top five BES, and all molecule-level descriptors, flow into the second hidden layer (with 10 nodes), which outputs a single molecule epoxidation score (MES) for the input molecule, reflecting the probability that the molecule will be epoxidized. For conciseness, the diagram is abbreviated and only shows two nodes for each hidden layer, one molecule input node, two atom input nodes (for each atom associated with the bond) and one bond input node. The actual model had several additional nodes in the input and hidden layers.

4.3.1 Accuracy in Identifying Sites of Epoxidation

An important goal for designing drugs less prone to metabolic activation is to accurately identify the site (bond) within a molecule that undergoes epoxidation. In our study, SOE predictions gave a specific hypothesis about the mechanism of a molecule's toxicity. Furthermore, knowledge of the SOE lays a strong foundation for guiding the modification of a molecule to make it less susceptible to epoxidation and thus less likely to cause protein and DNA adducts that lead to toxic effects. There are currently no other published computational methods that specifically predict SOEs among a diverse set of molecules.

The trained model predicts SOEs by computing a BES for each bond in a test molecule. These scores ranged between zero and one and reflected the probability that an epoxide will form on the two atoms within the corresponding bond. If accurate, BES should discriminate between SOEs and all other bonds within epoxidized molecules.

We assessed the generalization performance of our model using a cross-validation protocol. In this procedure, we separated molecules into metabolically related groups that represented metabolic networks in the database. Each group was comprised of epoxidized molecules and all parent and sibling molecules of those epoxidized molecules. One by one, each group of molecules derived from these networks was withheld from the training set. The rest of the molecules was used to train a model and make predictions on all the molecules present in the group left out of the training process. In each cross-validation fold, the model predictions for test molecules then did not depend on training data from identical or closely related molecules and thus provided a rigorous evaluation of the model. In this way, BES predictions were made on all molecules in the training data.

We used two metrics to quantitatively measure the classification performance of the cross-validated BES. First, we computed the "average site AUC" by calculating the area under the ROC curve (AUC) for each molecule and quantified the whole data set performance by averaging the AUCs for each molecule in the data set. Second, we used the "top-two" metric, which is often used in site of metabolism prediction.[50, 122, 123] By this metric, a molecule

was considered correctly predicted if any of its observed SOEs were predicted in the first- or second-rank position by a given model. Both metrics measure the separation of known SOEs from all other bonds within each molecule known to undergo epoxidation.

The BES reported by the neural network model accurately identified SOEs with an average site AUC performance of 94.9% and a top-two performance of 83.0% (Figure 4.4). The neural network outperformed a simpler logistic regressor model (BES[LR] in the figure), which had an average site AUC performance of 93.7% and a top-two performance of 80.5%. The neural network was significantly more accurate than the logistic regressor, reducing the error by 19.0% (average site AUC) and 12.8% (top-two). This improvement is significant according to a paired *t*-test, with *p*-values of 0.000454 (average site AUC) and 0.0328 (top-two).^[175] This improvement indicated non-linearity in the epoxidation data that cannot be taken into account by a logistic regressor. This finding justified the use of the more complex neural network, and was consistent with a previous study on site of metabolism prediction ^[83], as well as our previous work on sites of glutathione reactivity. ^[108]

This model for epoxidation is the first of its kind, and thus there are no other published models by which performance can be compared. Instead, we tested the performance of each raw descriptor to provide a baseline for comparison. Each descriptor was treated as a very simple model limited to a single chemical attribute to predict SOE. The best performing descriptor was π_p occupancy; however, this descriptor significantly underperformed our model, with accuracies of 90.8% (average site AUC) and 72.8% (top-two). Using machine learning to collectively consider many chemical attributes classified SOEs more accurately than any attribute considered in isolation.

4.3.2 Descriptors Driving Bond Epoxidation Score Performance

We identified which descriptors the model relied upon to further assess the sensibility of the model. The contribution of individual descriptors for identifying SOEs was measurable with a permutation sensitivity analysis. ^[84, 108] First, a baseline model was built using the entire

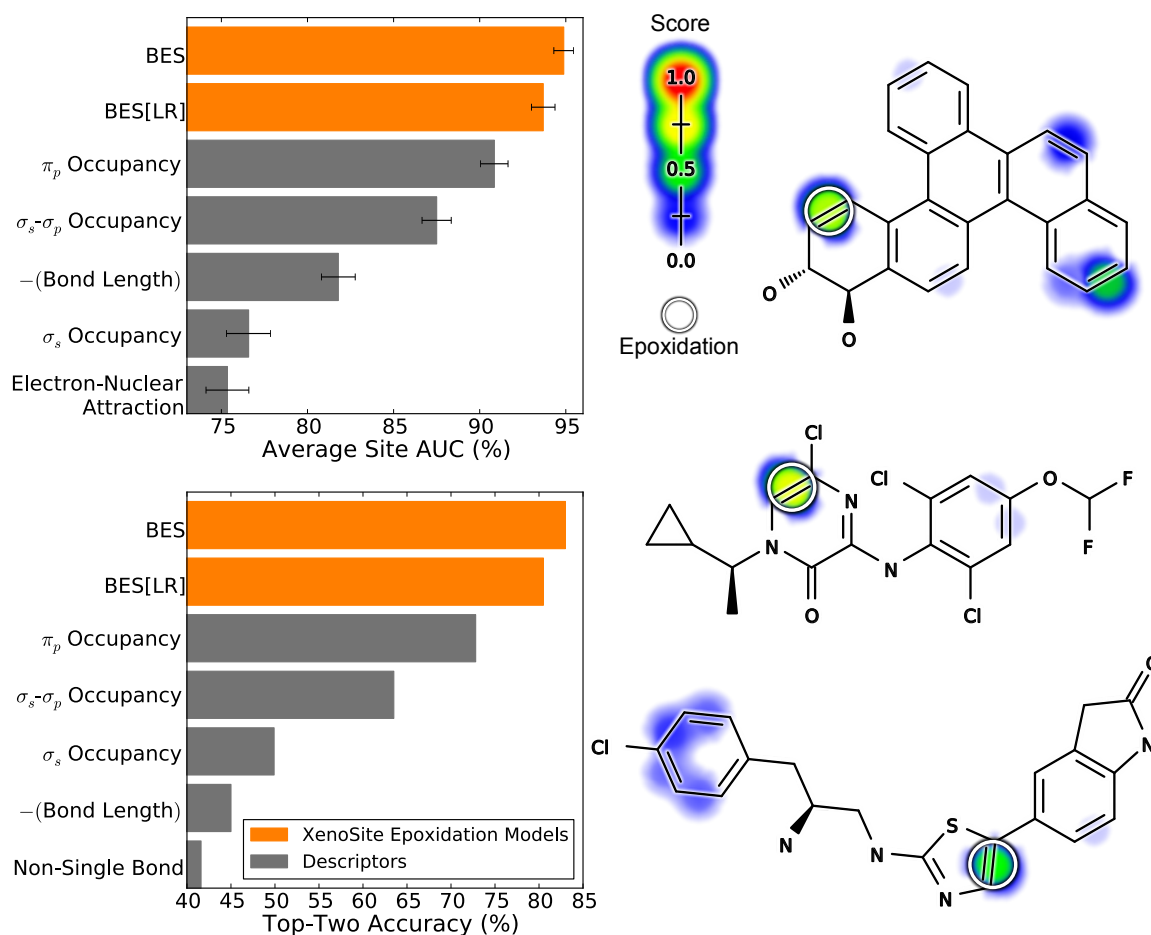


FIGURE 4.4: Bond epoxidation scores accurately (BES) identify sites of epoxidation (SOEs). Top left, for each prediction method, average site AUC was computed for 389 molecules extracted from the Accelrys Metabolite Database with their SOEs labeled. This metric reflected how often SOEs were ranked above other sites within these molecules. Bottom left, top-two classification performance was computed, by which a molecule was considered correctly predicted if any of its observed SOEs were predicted in the first- or second-rank position. By both metrics, the cross-validated predictions generated by a neural network (BES) outperformed the predictions of a logistic regressor (BES[LR]). The classification performance of BES also exceeded that of all raw descriptors, the five best of which are included in each panel. Right, examples from the data set are visualized with their predictions^[176, 177, 178] In the bar graph axis, the two-center electron-nuclear attraction energy is abbreviated as electron-nuclear attraction. For each molecule, the colored shading represents BES, which range from 0 to 0.73. Each experimentally-observed SOEs is circled.

Network

training data set, and its performance was calculated on this training data. The average site AUC performance was used for the sensitivity analysis, because it most closely measures performance in the intended use case. It quantifies how accurately the model identifies the correct SOEs within epoxidized test molecules, relative to all other potential sites, on a molecule-by-molecule basis. Reassuringly, very similar results from the sensitivity analysis are obtained using other metrics (data not shown). Next, the influence of individual descriptors, as well as groups of descriptors, was measured by recording the drop in the model's performance on the training data when the descriptor values were shuffled randomly. For each descriptor set, the shuffling procedure was performed 10 times, and the mean performance drop reported. Descriptors more heavily relied upon by the model were associated with higher performance drops.

As seen in Figure 4.5, the model primarily relied on quantum chemical bond descriptors. Shuffling all quantum chemical bond descriptors (listed in Table 4.1) as a group resulted in a performance drop of 10.3%. The most important individual descriptor was π_p occupancy; shuffling of this descriptor was associated with a performance drop of 4.8%. This observation was consistent with π_p occupancy predicting SOEs reasonably well by itself, with the best performance among all lone descriptors (Figure 4.4). Heavy reliance on π_p occupancy by the model is logical given its role in epoxidation. In fact, a π -complex is the initial intermediate formed during epoxidation by cytochromes P450.[169, 179, 180] Consistent with this mechanism, this sensitivity analysis is consistent with critical importance of π_p in forming the intermediate π -complex. While reasonable, π_p occupancy has never been proposed as a way to identify SOE.

The second most important descriptor was SMARTCyp reactivity, with a performance drop of 2.5%. The relevance of SMARTCyp reactivity is readily understandable, because it predicts the sites of cytochromes P450 metabolism of drug-like molecules.[52] The remaining most important individual descriptors were topological. Previous studies by our group found topological descriptors to be important for many different types of chemical modeling.[108] Topology encompasses fundamental information, such as atom element identity or bond type, which has

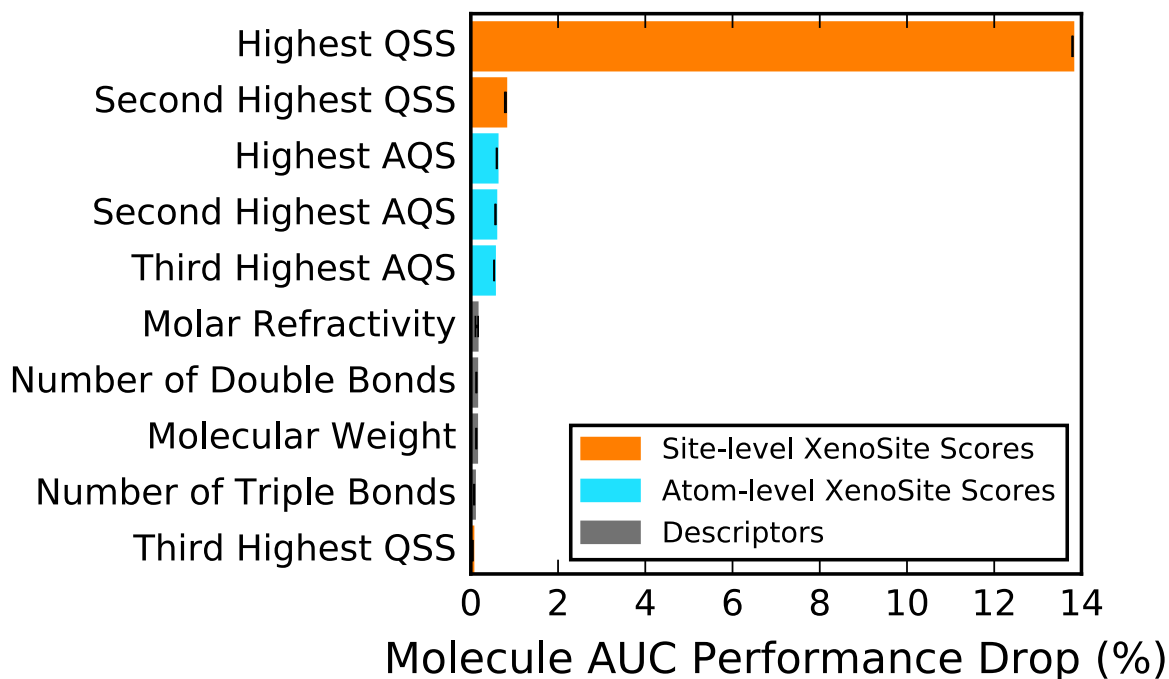


FIGURE 4.5: The importance of specific descriptors to the bond epoxidation model. A permutation sensitivity analysis quantified the importance of descriptors for the final trained site of epoxidation model. Left, the 10 most important individual descriptors in decreasing order of importance from top to bottom. Right, the importance of four broad descriptor categories. The graph shows the model performance drop associated after permuting the associated descriptor values, averaging over 10 iterations.

been useful for finding many different types of patterns. Overall, the results of sensitivity analysis indicated that the model logically relied upon descriptors relevant to epoxidation.

4.3.3 Accuracy in Identifying Aromatic and Double Bond Sites of Epoxidation

Ideally, the model should be able to distinguish SOEs from all other bonds across the entire dataset. This is not assessed by the average site AUC and top-two metrics used in prior sections, which only compare BES predictions on a molecule-by-molecule basis. In contrast, global AUC, computed across all atoms in the dataset does measure this behavior. The model's BES is very accurate across the whole dataset, with a global AUC of 95.6%. The logistic regressor is

Network

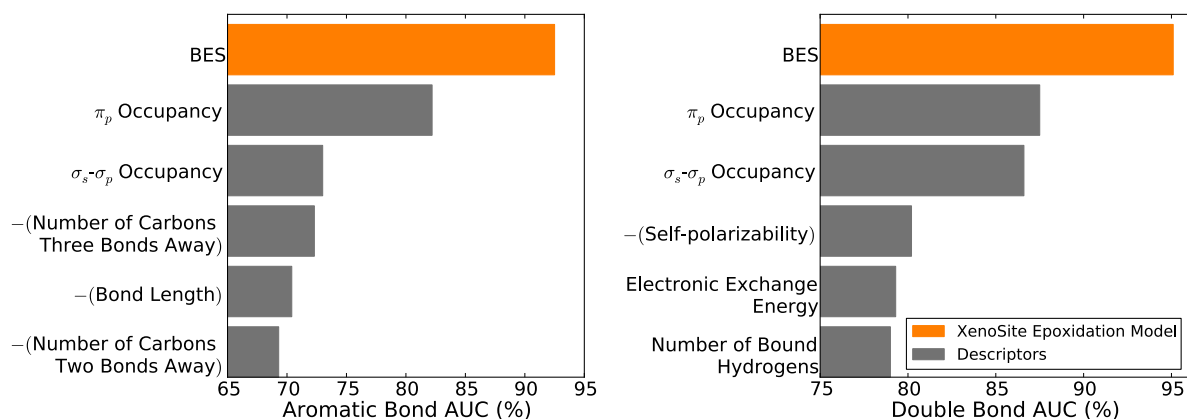


FIGURE 4.6: Bond epoxidation scores (BES) accurately identified both aromatic and double bond sites of epoxidation. Across the 389 molecules that underwent epoxidation, the model accurately separated epoxidized and non-epoxidized aromatic bonds (left) and double bonds (right). Using cross-validated scores, classification performance was quantified by computing the AUC of the model on either the aromatic or the double bonds in the full data set. The AUC of the model was compared with similarly computed AUCs for individual descriptors. In both cases, the model BES outperformed all individual descriptors.

slightly less accurate with a global AUC of 94.5%, but this performance drop is significant with a p-value of 10^{-8} computed with a paired *t*-test [175]. Similarly, the best performing descriptor is the π_p occupancy with a global AUC of 88.4%, which is also a significant performance drop from the BES with a p-value approaching zero.

We further assessed the model's performance by ensuring it was able to distinguish SOEs from either aromatic and double bonds (Figure 4.6). These tests excluded (for example) single bonds, which are very rarely epoxidized and might artificially inflate performance if included in performance calculations. An aromatic bond AUC was computed by first extracting all aromatic bonds within epoxidized molecules, and then calculating AUC. A double bond AUC was calculated similarly. Encouragingly, BES were very accurate in identifying both epoxidized aromatic bonds (92.5%) and epoxidized double bonds (95.1%) and also substantially outperformed all individual descriptors.

4.3.4 Distinguishing Epoxidation from Hydroxylation

Another key task was to accurately SOEs from SOHs, because epoxidation and hydroxylation may have significantly different implications for toxicity and downstream metabolism. Generally, SOEs are not obviously distinguishable from sites of sp^2 hydroxylation, because either epoxidation or hydroxylation may occur at sp^2 atoms. While several studies have already demonstrated that computational models can predict the sites where molecules are oxidized[50, 51, 52, 36, 53, 54], they do not predict if the oxidation is an epoxidation or a hydroxylation.

For our study, we tested whether BES distinguished SOEs from SOHs. We initially built a hydroxylation data set of 3000 hydroxylation reactions that were randomly sampled from the AMD resource, as described in the Materials and Methods. This final data set included 811 molecules, in which atoms were marked if they are sites of sp^2 hydroxylation.

In this study, a SOE was defined as a *bond* between the two carbons of the final epoxide, whereas a SOH is usually defined as the single *atom* targeted for hydroxylation. However, our model only makes predictions on bonds. So, for validation purposes, we labeled the *bonds* connecting to hydroxylated atoms as SOHs and asked whether these sites receive lower scores than SOEs. Only bonds between two sp^2 carbon atoms were included. Each of the 811 molecules in the hydroxylation data set were tested by our model, and the predictions for each bond of hydroxylation extracted. As previously explained, the hydroxylation reactions were sampled randomly from our database. Therefore, molecules subject to both hydroxylation and epoxidation data sets were included. Cross-validated predictions were used for molecules that were also part of the training set. Within these molecules, it was possible for the same site to be subject to both epoxidation and hydroxylation. These sites were labeled as SOEs.

We investigated whether these SOEs were distinguishable from SOHs. Encouragingly, BES separated SOEs from SOHs with an AUC of 83.3% (Figure 4.7). In contrast, the best performing raw descriptor among all tested was π_p occupancy, with an AUC of only 77.0%. This is a critical result because it demonstrates that the model can distinguish identify SOEs from bonds that are also acted on by P450s, but not epoxidized.

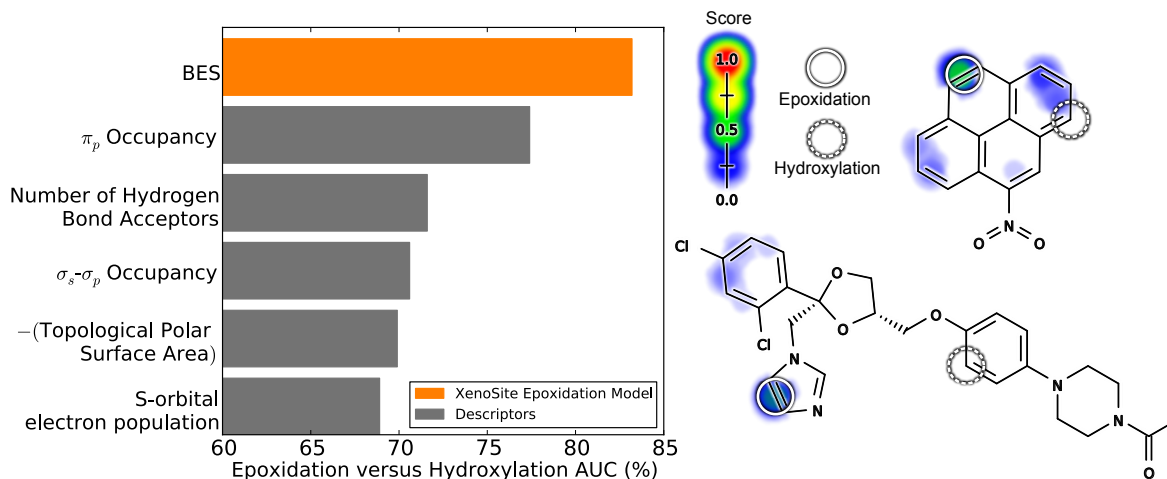


FIGURE 4.7: Bond epoxidation scores (BES) distinguish sites of epoxidation (SOEs) from sites of hydroxylation (SOHs). Top, each prediction method was assessed by its ability to separate SOEs from SOHs. The cross-validated scores on the SOEs of 389 epoxidized training molecules were compared with the SOH scores on 811 test molecules with their sites of sp^2 hydroxylation labeled. The scores for each SOE and SOH were extracted and performance quantified by computing the AUC. The classification performance of the model was then compared with similarly computed AUCs for individual descriptors. The model's BES outperformed all individual descriptors. Right, from top to bottom are 1-nitropyrene[181] and ketoconazole[182], example molecules subject to both epoxidation and hydroxylation. Each SOE is indicated by solid circles, and SOHs are indicated by dashed circles. The colored shading indicates BES (which range from 0 to 0.45).

4.3.5 π_p Orbital Occupancy and Epoxidation

One striking result from these experiments is the consistently high importance of π_p occupancy in identifying SOEs. Although, it has been known for a long time that a π -complex is the initial intermediate formed during epoxidation by cytochromes P450,[169, 179, 180] no published literature has suggested π_p occupancy is a marker for SOEs or quantitatively assessed its ability to identify SOEs.

To further investigate this observation, which may provide mechanistic clues, we studied the distribution of π_p occupancy and BES as a function of epoxidation and bond type (Figure 4.8). From these distributions, it seems immediately clear that SOEs have higher π_p occupancy than non-epoxidized sites. However, π_p occupancy is also strongly correlated with the type of bond; and the optimal cutoff between SOEs and non-epoxidized sites is different for double and aromatic bonds. This result suggests that π_p occupancy may not be the direct driver of the π -intermediate's formation. Instead, π_p occupancy may be a proxy for another factor that we do not directly capture in other descriptors. One possible factor may be the ability of neighboring groups to donate π_p electrons, but directly testing this hypothesis is beyond the immediate scope of this study and will be left for future work.

These distributions also highlight another key feature of our approach; the model's output is well-scaled and can be interpreted as a probability. In other words, bonds with a BES score of 0.8 have approximately a 80% chance of being epoxidized. In contrast, π_p occupancy, though predictive, is not scaled to be a SOE probability.

4.3.6 Accuracy at Identifying Molecules that Undergo Epoxidation

We also assessed the ability of our model to separate epoxidized from non-epoxidized molecules. With high enough classification performance, our model might be a useful tool to rapidly screen drug candidates for potentially problematic molecules.[144, 145, 146, 147, 148]

In this assessment, we trained the model for epoxidation to distinguish between those molecules that underwent epoxidation and those that did not. We included in our training data set molecules

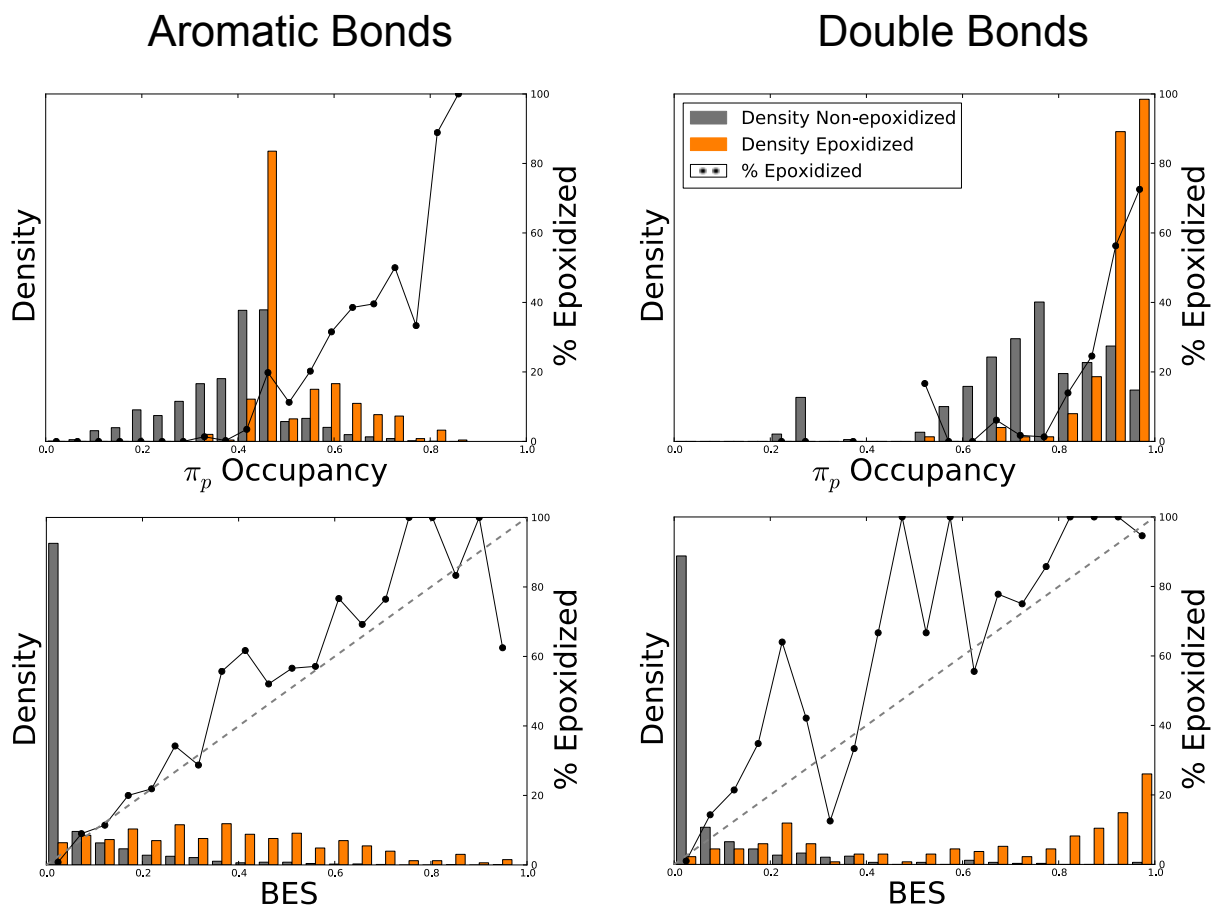


FIGURE 4.8: Bond epoxidation scores (BES) represent a well-scaled probability that a site will be epoxidized. Across the 389 molecules that underwent epoxidation, the normalized distribution of BES (bottom) and π_p occupancy (top) across both aromatic bonds (left) and double bonds (right) are displayed for all epoxidized and non-epoxidized sites, indicated by the shaded bars. The solid lines represents the percentage of bonds that are epoxidized (using non-normalized frequencies). The diagonal dashed lines on the bottom plots indicates a hypothetical perfectly scaled prediction. This demonstrates that BES is much better scaled than π_p occupancy.

that are structurally closely related to epoxidized molecules, but are not themselves epoxidized in our database. After training the model on the SOE level, we tested several methods of separating epoxidized and non-epoxidized molecules (Figure 4.9). In this case, classification performance was quantified by measuring the AUC across the entire data set.

The simplest method for predicting molecule epoxidation was to take the cross-validated maximum BES score within each molecule. Across the entire data set, this approach yielded MES that separate epoxidized and non-epoxidized molecules with an AUC of 78.6%. The addition of a training step to input the top five BES and molecule-level descriptors into a logistic regressor or neural network slightly improved classification performance. The cross-validated scores outputted by a logistic regressor (MES[LR] in the figure) had a higher AUC of 78.9% and those of the neural network (MES[NN]) had an AUC of 79.3%. A false positive rate paired *t*-test^[175] indicated that MES[NN] was not significantly better than max [BES] (*p*-value 0.14) or MES[LR] (*p*-value 0.19).

However, MES[NN] provided a better scaled prediction than either max [BES] or MES[LR], as demonstrated by the reliability plots in Figure 4.10. The neural network closely approximated a perfectly well-scaled prediction, with an R^2 value of 0.971, compared to 0.956 for the logistic regressor or 0.889 for max [BES]. The neural network's reliability plot is superior to that of the logistic regressor, not only due to the higher R^2 value, but also because it assigns significantly more non-epoxidized molecules low scores, and epoxidized molecules high scores, evidenced by the relative densities in Figure 4.10.

Nevertheless, choosing between the logistic regressor and neural network is debatable. The logistic regressor offers a simpler model structure, whereas the neural network provides a slightly higher classification performance and better scaled prediction. Going forward, we decided to use the neural network, but we believe that the logistic regressor could also be used with similar results. For the rest of the study, we define MES to mean MES[NN].

The significantly lower AUC of the molecule-level MES compared to the site-level BES was a consequence of the lower quality of the molecule-level data, which included "non-epoxidized"

molecules. This was based on our assumption that molecules were non-epoxidized if they were not subject to any epoxidation reaction in our literature-derived database. While necessary, this assumption was not strong evidence that molecules were not subject to epoxidation, because not all studies look for epoxidation products. As a consequence, some epoxidizable molecules were incorrectly labeled as non-epoxidized in our data. In contrast, our site-level epoxidation data is much less noisy, because it is drawn from experiments detecting epoxidation, and this is reflected in the higher site-level performance.

Nevertheless, MES separated epoxidized and non-epoxidized molecules with 79.3% AUC. This result is consistent with our presumption that the most of the molecules labeled as non-epoxidized, are truly not epoxidized. If epoxidized and non-epoxidized molecules were drawn from the same chemical distribution, it would not be possible to separate them with any accuracy. Furthermore, MES outperformed all molecule-level descriptors in terms of classification performance. This result demonstrated that our model offers an informative prediction on the molecule level. The best performing descriptor was the negative of the total number of single bonds in a molecule, yet its AUC was only 72.3%, considerably worse than MES. In contrast to site-level epoxidation, for which π_p occupancy was quite predictive (Figure 4.4), maximum π_p occupancy predicts molecule epoxidation with only 57.7% AUC. The model MES much more accurately predicted which molecules will be epoxidized than any single chemical descriptor.

4.3.7 Case Studies

Knowledge of the SOE of a drug or drug candidate can direct rational drug design to avoid the formation of reactive metabolites and reduce the risk of adverse drug reactions. Case studies provide excellent examples of how our model could enable the development of safer drugs (Figure 4.11).

Carbamazepine is an effective drug to treat epilepsy; however, it can cause severe adverse reactions mediated by reactive metabolites. Carbamazepine metabolism can form several reactive metabolites, including an iminoquinone,[159] but the epoxide's formation is the focus

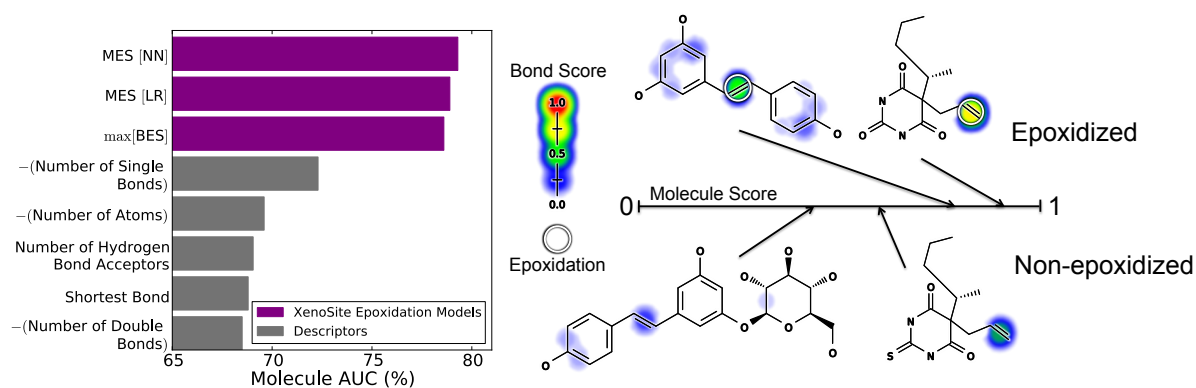


FIGURE 4.9: Molecule epoxidation scores accurately identify molecules subject to epoxidation. Left, several prediction methods were compared by their ability to identify molecules that underwent epoxidation. The data set included 524 molecules, 389 of which were epoxidized and 135 structurally similar but not epoxidized molecules. Model performance was measured by computing the AUC across epoxidized and non-epoxidized molecules (Molecule AUC), using cross-validated scores. By this metric, the best approach inputted the top five bond epoxidation scores (BES) and all molecule-level descriptors into a neural network (MES[NN]). This slightly outperformed the simpler methods of using a logistic regressor (MES[LR]) or merely taking the maximum bond epoxidation score (max[BES]). While this improvement is not statistically significant, on the basis of the reliability plots in Figure 4.10, the neural network (MES[NN]) was chosen to calculate molecule epoxidation scores (MES) for this study. Right, example pairs of epoxidized and closely related non-epoxidized molecules are visualized. From left to right, top to bottom: resveratrol (MES: 0.79)[183], quinalbarbitone (MES: 0.88)[184], glucuronidated resveratrol (MES: 0.37)[185], and thiopental (MES: 0.60).[186] Each experimentally-observed site of epoxidation is circled. For each molecule, the colored shading represents BES, which range from 0 to 0.76.

Network

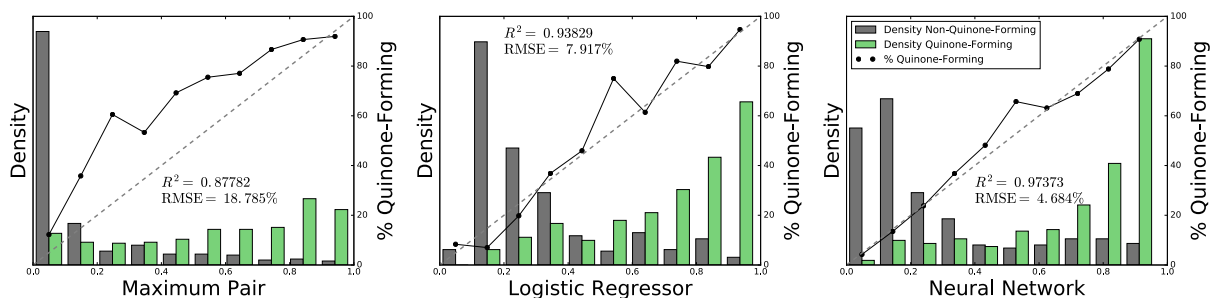


FIGURE 4.10: MES[NN] offers a well-scaled probabilistic prediction of molecule epoxidation. The bar graphs plot the normalized distributions of max [BES], MES[LR], and MES[NN] across 525 epoxidized and non-epoxidized molecules. The solid lines plot the percentage of molecules that are epoxidized (using non-normalized frequencies) in each bin. The diagonal dashed lines indicate a hypothetical perfectly scaled prediction. MES[NN] offers the best scaled prediction of the three methods, with a strong correlation to a perfectly scaled prediction. This means that the MES[NN] is interpretable as the probability that a molecule is epoxidized.

of this study and more correlated with adverse reactions.[160, 161, 162] Analogues of carbamazepine that block the epoxidation have a lower incidence of adverse effects. Replacement of the problematic double bond with a ketone yielded oxcarbazepine, which lacks the metabolic activation to an epoxide and adverse events, yet retains similar efficacy.[164] Similarly, eslicarbazepine does not contain the problematic double bond, is no longer epoxidized at this position, and also has a lower incidence of adverse reactions. [187, 162] The model correctly identifies carbamazepine's SOE. Furthermore, the model correctly identified two carbamazepine analogues as less likely to be epoxidized: oxcarbazepine (MES: 0.38) and eslicarbazepine (MES: 0.20) compared with carbamazepine (MES: 0.88).

Furosemide is a commonly prescribed diuretic, but is prone to hypersensitivity reactions and hepatotoxicity due to the epoxidation of its furan ring.[9] The model correctly identifies this as a SOE. There are no close analogues of furosemide on the market. However, there are three other drugs in the same class that containing the same sulfamyl-based active scaffold: piretanide, bumetanide, and torasemide. None of these drugs contain the problematic furan, are all predicted not to form epoxides (MES: 0.21, 0.19, and 0.21, respectively, compared with 0.94 for furosemide), and all are less prone to hypersensitivity driven reactions than furosemide.[9]

The case of hepatotoxic sudoxicam and its non-hepatotoxic analogue, meloxicam, is a more complicated.[9] Sudoxicam is a NSAID that was withdrawn from testing due to hepatotoxicity caused by epoxidation of its thiazole ring; the unstable epoxide causes ring scission and formation of a reactive acylthiourea metabolite.[9, 158] This reaction pathway is suppressed by the addition of a single methyl group to sudoxicam's SOE. The resulting drug meloxicam is less prone to epoxidation, although the epoxide still forms.[158] Instead, meloxicam is primarily hydroxylated at the added carbon.[158] As a result, the reactive acylthiourea urea metabolite forms less often, and consequently meloxicam is not hepatotoxic, despite being prescribed at a similar dose to the hepatotoxic sudoxicam.[9, 158]

The model correctly predicts the SOEs of both sudoxicam and meloxicam, and assigns them high MES of 0.95 and 0.96, respectively. However, the model does not identify meloxicam as the less toxic molecule. This is exactly what we should expect, because both molecules are epoxidized by P450s.[158] Meloxicam's modification introduces an alternative hydroxylation pathway that reduces the amount of epoxide formed, and this change is responsible for its reduced toxicity. This highlights the limitations of considering the epoxidation pathway in isolation. A better risk assessment might combine epoxidation predictions with more comprehensive models of metabolism to predict if epoxides are a major metabolite. Building this system is exactly our long-range goal, but beyond the scope of the current study.

Nevertheless, our findings provide a critical step in the right direction: the first reported model that predicts the formation of reactive epoxides from drug candidates and the accurate identification of the specific epoxidized bonds. As clear in both cases, carbamazepine and sudoxicam, the model can be used to identify SOEs that can then be modified to make drugs safer.

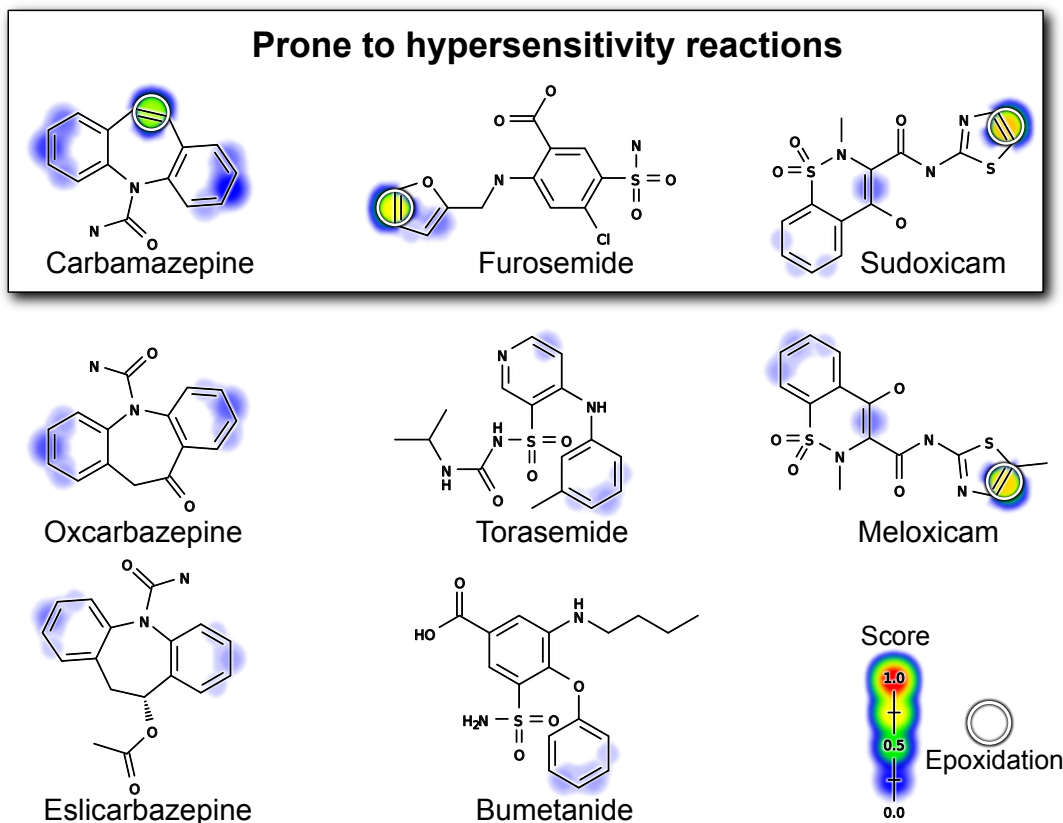


FIGURE 4.11: .

The epoxidation model recognizes sites of epoxidation within drugs that can be modified to reduce toxicity. The figure includes three groups of closely related drugs shaded by their BES scores; the top three are prone to hypersensitivity reactions while their analogues are not. The top three molecules and meloxicam are epoxidized and their sites of epoxidation are circled.[160, 161, 162, 188, 189, 158] The model's BES correctly identifies the SOEs in these molecules. The model's MES correctly identifies these molecules as epoxidized, with higher scores than the non-epoxidized molecules. For the top three molecules, epoxidation is the primary mechanism of their hypersensitivity.[9] Encouragingly, the two analogues of carbamazepine are correctly identified as non-epoxidized, and therefore non-hepatotoxic. This demonstrates how the model could be used to identify less toxic analogues. Furosemide does not have a close analogue on the market, but the model correctly identifies the furan ring as problematic. The other diuretics with the same active scaffold, but without this furan, are less toxic.[9] Identifying meloxicam as less toxic is a more difficult task and would require more comprehensive metabolism modeling. Meloxicam is a safer analogue of sudoxicam because an alternate hydroxylation pathway is introduced by the modification that out competes the epoxidation pathway.[158]

4.4 Conclusion

This study establishes a new system to predicting the formation of reactive epoxide metabolites. The epoxidation model—trained on SOE data—identifies with 94.9% AUC performance the SOEs within epoxidized molecules. The model also classifies epoxidized and non-epoxidized molecules with 79.3% AUC . This method needs to be combined with additional tools to be useful for predicting the toxicity of drugs. For example, while this model predicts the formation of epoxides, it does not score the reactivity of these epoxides. Epoxide reactivity can vary widely, with half-lives ranging from one second to several hours,[169] and this variation may have significant implications for toxicity. To address this, we plan to combine this epoxidation model with a model of reactivity already developed.[108] Furthermore, we will expand to model quinone formation, another motif of potentially high reactivity that frequently causes adverse drug reactions.[103, 190, 152] Ultimately, we envision a powerful model for predicting adverse drug reactions that integrates metabolism models, reactivity models, and dosage information. By accurately modeling epoxidation, this study provides a key piece of this ultimate goal.

Chapter 5

Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism

Reprinted (adapted) with permission from Hughes, T. B. and Swamidass, S. J. (2017). Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism. *Chemical Research in Toxicology* 30(2), 642–656. Copyright (2017) American Chemical Society.

5.1 Introduction

Unanticipated toxicity frequently causes drug candidates to be discontinued late in development.[8] Furthermore, idiosyncratic adverse drug reactions (IADRs) are often detected only after approval, causing significant morbidity and mortality.[98, 6] Around 75% of IADR cases result in liver transplants or death.[14, 191] There is considerable evidence that many varieties of drug toxicity—including IADRs—may be driven by a common mechanism: drug-metabolizing enzymes bioactivate drugs into electrophilically reactive metabolites, which covalently bind to sites within DNA or proteins.[31, 192, 193, 24] Frequently, DNA adducts are mutagenic, and may interfere with transcription, replication, or regulation, causing gene dysfunction or initiating cancer.[99, 100, 19] Likewise, protein adducts can disturb important biological activities or engender adverse immune responses.[48, 19, 49] This study focuses on quinones, a major

class of Michael acceptors, which are soft electrophiles[61] and the most common type of reactive metabolites.[152] Quinone species, such as quinone-imines, quinone-methides, or imine-methides, are commonly highly electrophilically reactive.[194] Over 40% of all known reactive metabolites are quinones.[152] Quinone formation—the formation of quinones by metabolic oxidation—is performed by cytochromes P450[9, 195, 196] and peroxidases.[24, 197, 198, 199] Quinone formation generally occurs in one or two metabolic steps, where the final step is usually the co-reduction of a cytochrome P450's catalytic iron by the substrate, thereby dehydrogenating and oxidizing the substrate into a quinone species.[200] For example, in a single step both cytochromes P450[81] and peroxidases[201] convert acetaminophen to the toxic metabolite *N*-acetyl-*p*-benzoquinone imine, which covalently binds to proteins (Figure 5.1). In contrast, lumiracoxib requires two-steps to form a quinone. First, cytochromes P450 hydroxylate lumiracoxib, the product of which then forms a quinone after further metabolism by cytochromes P450 or peroxidases.[9, 202, 203]

Detecting or anticipating quinone formation is critical for avoiding drug candidates early in drug development that can form reactive metabolites. Experimental methods for detecting quinone formation—most commonly incubation with glutathione—are well developed.[204] However, these techniques require time and resources, which can be significant, especially when considering thousands of candidates during the initial screening phase of drug development. Moreover, some methods detect metabolic events that are inconclusive for quinone formation—such as aromatic hydroxylation—because the product may not form a subsequent quinone in the *in vitro* screens, which are often tuned to only generate one-step metabolites. For example, buspirone undergoes aromatic hydroxylation but does not form a quinone (Figure 5.2). In contrast, after hydroxylation at a structurally similar motif, the analog nefazodone does proceed to form a hepatotoxic quinone.[205, 206] An accurate computational method for predicting both one- and two-step quinone formation pathways would be complementary to experimental screens, and might identify otherwise undetected quinone formation risk.

Several studies demonstrated that computational methods can predict sites of metabolism.[50,

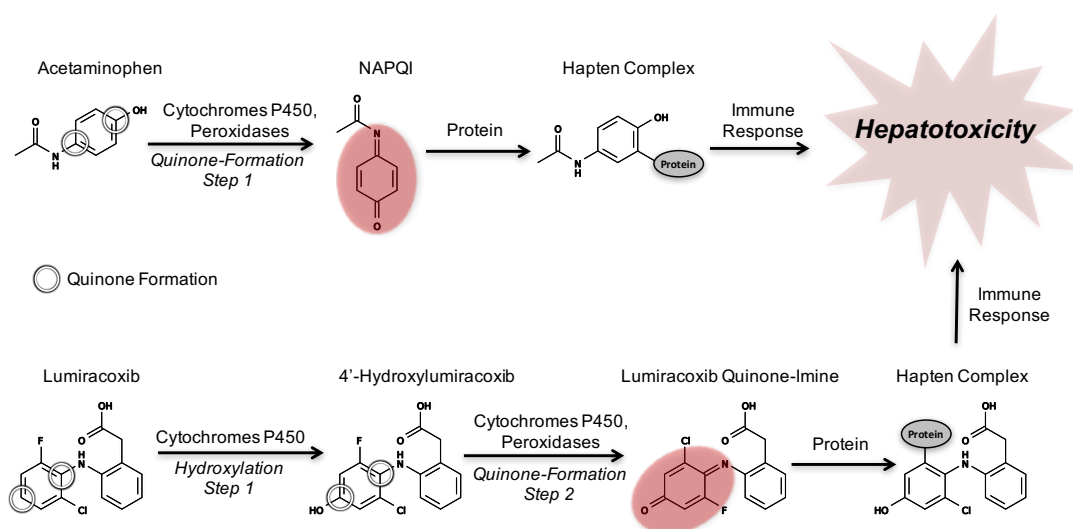


FIGURE 5.1: Examples of one- and two-step quinone formation: the formation of a quinone by metabolic oxidation. The quinone (indicated by red ovals) metabolites of acetaminophen and lumiracoxib cause hepatotoxicity due to formation of a hapten complex with off-target proteins that triggers toxic immune responses [9, 202, 203]. Acetaminophen's conversion to the reactive quinone *N*-acetyl-*p*-benzoquinone imine (NAPQI) illustrates one-step quinone formation, and lumiracoxib's conversion to lumiracoxib quinone-imine via the intermediate 4'-hydroxylumiracoxib demonstrates two-step quinone formation [9, 202, 203]. The quinone formation model encompasses both one- and two-step quinone formation.

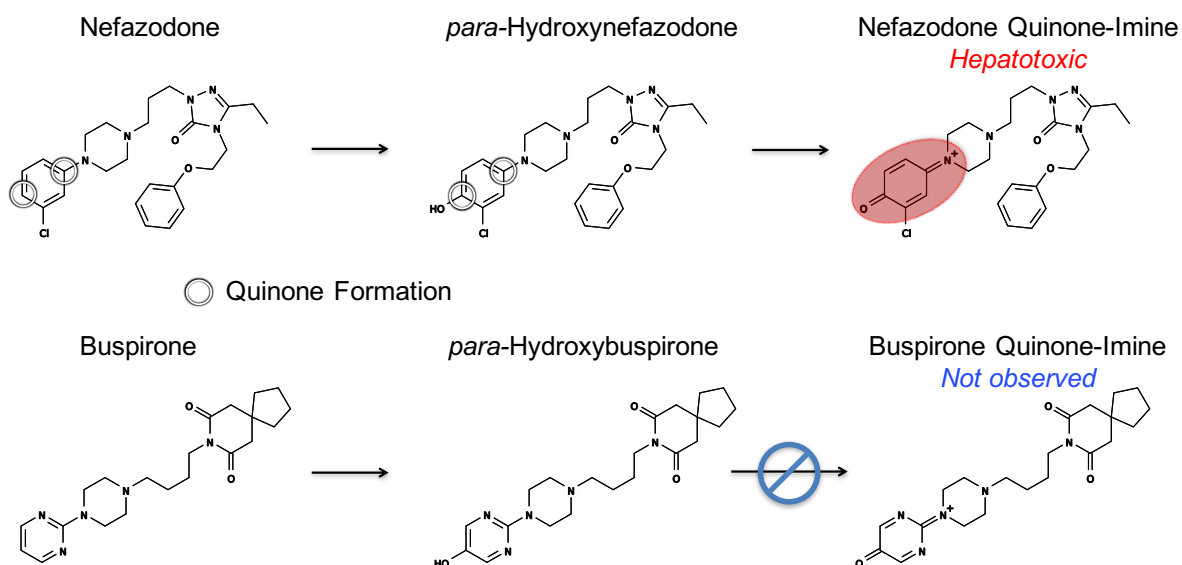


FIGURE 5.2: Both quinone-forming and non-quinone-forming molecules can undergo comparable initial metabolic events. Nefazodone first undergoes aromatic hydroxylation to form hydroxy-nefazodone, which is then oxidized to form the reactive nefazodone quinone, thereby causing hepatotoxicity in some patients [205, 206]. In contrast, the analog buspirone does not form a quinone, despite undergoing a comparable initial hydroxylation event [206]. A key goal of our model was to accurately identify molecules that form quinones.

51, 52, 36, 53, 54, 207] However, these methods do not predict the metabolic structures produced by these reactions, and 450s can alter the same site in different ways, each with different consequences to toxicity.[165, 149] In contrast, we recently published a computational model that went beyond site of metabolism prediction to specifically predict the formation of epoxides—the second most common type of reactive metabolites after Michael acceptors such as quinones—which demonstrated the feasibility and utility of modeling the formation of specific reactive metabolites.[121]

In this study, we built the quinone formation model (Figure 5.3), which succeeded at two crucial tasks. First, the model accurately predicted the specific atom pairs within molecules that form quinones: their quinone-forming pairs (QPs). Knowledge of QPs can guide structural modifications to reduce the chances of reactive metabolite formation. Second, the model distinguished quinone-forming and non-quinone-forming molecules. Molecule quinone formation predictions can be used to quickly screen for molecules likely to form reactive metabolites, a key toxicity risk.

5.2 Materials and Methods

5.2.1 Quinone Formation Training Data

From the literature-derived Accelrys Metabolite Database (AMD), we extracted an extensive data set of chemically diverse metabolic reactions that form quinones. Overall, 576 reactions were extracted, each observed in humans, human cells, or human microsomes. These reactions consisted of 377 single-step quinone formation reactions and 199 two-step quinone formation reactions. An automated algorithm used the structure of each quinone product to label the QP on its metabolic parent, for one-step quinone formation, and its metabolic grandparent, for two-step quinone formation. We use “metabolic parent” to refer to the starting molecule of a one-step quinone formation reaction, and “metabolic grandparent” to refer to the starting molecule of a two-step quinone formation reaction. Metabolic parents form quinones directly, whereas

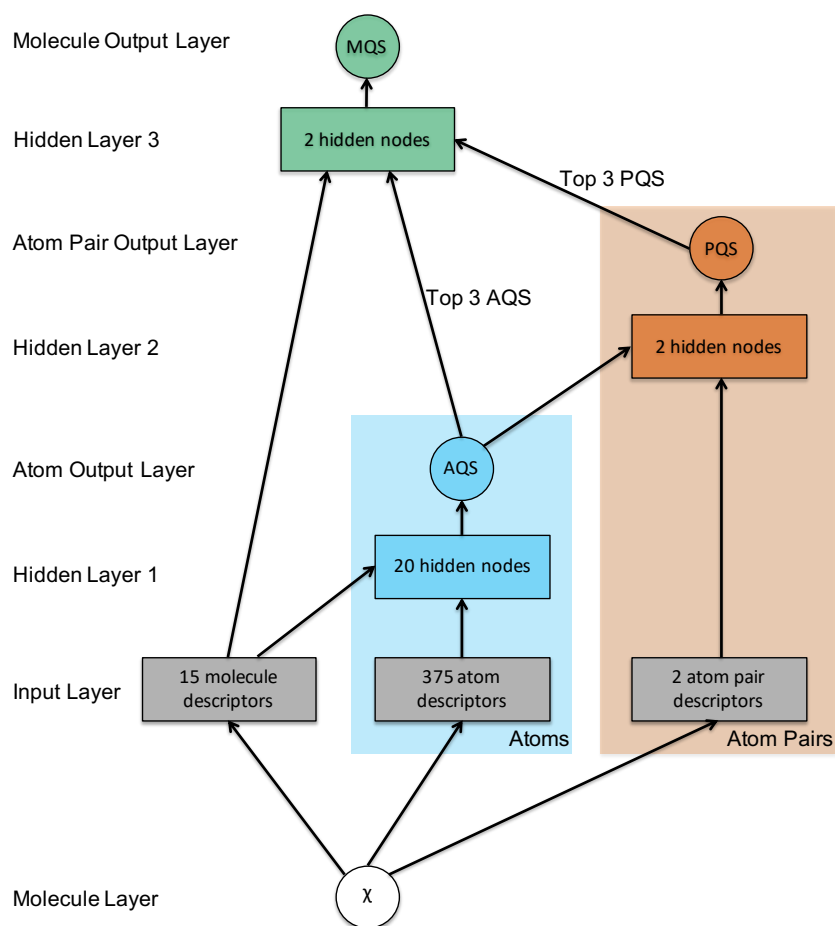


FIGURE 5.3: The network depicts how information propagated through the model, which contained one molecule layer, one input layer, three hidden layers, and three output layers. From the structure of an input molecule, several molecule-level, atom-level, and pair-level descriptors were calculated at the input layer. The molecule- and atom-level descriptors flowed into the first hidden layer, which outputted an atom quinone formation score (AQS) for each atom in the input molecule. In this study, we defined a quinone-forming pair (QP) as the exact pair of ring carbons that forms a quinone. To predict QP, the atom-level scores of both atoms and the pair-level descriptors were presented to the second hidden layer, which computed an pair-level quinone formation score (PQS). Finally, the top three atom-level scores, the top three pair-level scores, and all molecule-level descriptors were submitted to the third hidden layer, which calculated a molecule quinone formation score (MQS). On the right, atom-level data is illustrated on bottom (with atoms of quinone formation circled), pair-level data illustrated in the middle (with a pair of atoms circled), and molecule-level data is illustrated on top (with the quinone-forming molecule circled). The molecule input node is a chemical structure, and all other circles are predictions ranging from 0 to 1. Blocks are vectors of real numbers.

metabolic grandparents form a quinone through an intermediate structure. Both metabolic parents and grandparents were included in the training data. The results of the labeling algorithm were manually inspected and corrected as necessary.

For this study, we defined a QP as the pair of ring carbons that forms a quinone. In depictions, the atoms of these QPs are indicated by circles. When a pair of atoms was topologically equivalent to an observed QP, it was itself labeled as a QP. Duplicate starting molecules were merged into a lone training molecule with all observed QPs marked. The final data set included 359 quinone-forming molecules, each with its QPs labeled. Primarily, these QPs were found in six-membered rings (Table 5.3). Excluding atoms shared between fused rings of different sizes—such as the fused six- and seven-member rings of nevirapine in Figure 4—90.2% of the QP atoms were in six-membered rings, compared to 8.6% and 1.2% for five- and seven-membered rings, respectively.

Predicting which molecules will form quinones is a key goal of a useful model. Therefore, the training data set also included molecules that not form quinones. To assemble these molecules, first all metabolically studied molecules in human-relevant experiments were extracted from the AMD. Second, all molecules that contained a six-membered aromatic ring were selected. Third, all molecules that formed quinones were filtered out, including quinones that formed in multiple steps. This procedure yielded a pool of 11884 molecules that have been studied in humans, human cells, or human microsomes, but have not been reported to form quinones. From this pool, 359 (the same number of quinone-forming molecules) randomly selected non-quinone-forming molecules were added to the training set, for a total of 718 training molecules.

5.2.2 External Non-Quinone-Forming Test Sets

After removing the 359 non-quinone-forming molecules added to the training data, 11525 non-quinone-forming molecules remained. From this pool, external test sets were similarly extracted by randomly selecting 359 molecules for each test set. We repeated this process 20 times,

thereby constructing 20 external test sets with 359 molecules apiece. All of these molecules were withheld from model training, and were only evaluated by the final model.

5.2.3 Topological Descriptors

To predict quinone formation, the first step of our method assigned each atom a vector of numbers—topological descriptors—which described chemical properties of that atom. Each atom was assigned 390 numbers, including 375 atom descriptors and 15 molecule descriptors. These descriptors were computed by in-house python software using the structure of each molecule as input.[82] We used topological descriptors, previous versions of which have been used for models of metabolism,[136, 50] reactivity,[208, 108] and epoxidation.[121] For this study, we started with our most recent set of topological descriptors,[208] which we supplemented with several new atom descriptors for *ortho*, *meta*, and *para* substituents. A second descriptor generation step assigned two descriptors to each pair of atoms: the connectivity distance between the atoms, and whether this distance was odd or even.

5.2.4 Combined Pair- and Molecule-level Quinone Formation Model

We built a model using a deep convolutional neural network with one molecule layer, one input layer, three hidden layers, and three output layers (Figure 5.3). The top output layer computed molecule quinone formation scores (MQS), the middle output layer computed pair quinone formation scores (PQS), and the bottom output layer computed atom quinone formation scores (AQS). Respectively, the molecule-, pair-, and atom-level scores represented the probability that a molecule, pair, or atom formed a quinone, each score ranging from zero to one. The network was trained in three stages.

First, the atom-level network was trained to produce accurate atom-level scores. In this training step, each ring carbon within a molecule was considered a possible atom of quinone formation. Each atom was assigned a vector of numbers that described chemical properties of that atom. The data set was a matrix, with one row per atom, and one column per descriptor.

Experimentally observed atoms of quinone formation were labeled with a 1 in a final binary target vector. Using gradient descent on the cross-entropy error, we trained the network weights so that atoms of quinone formation received higher scores than other atoms.

Second, the pair-level network was trained to compute scores for each pair of atoms. In this training step, each row of the data matrix was a pair of ring carbons, and each column was a descriptor. Descriptors included the atom-level score of each atom and both pair-level descriptors. For each atom pair, the two associated atom-level score were sorted by magnitude and transformed by the logit function into a log-odds. The pair-level network was trained to assign experimentally observed QP with higher scores than all other atom pairs.

Third, the molecule-level network was trained. Each row of this data matrix was a molecule, and each column was a descriptor. Descriptors included the top three atom-level scores, the top three pair-level scores, and all molecule-level descriptors. The weights of the network were trained to assign quinone-forming molecules with higher scores than non-quinone-forming molecules.

The final model included hidden layers at all three stages. Several alternative models were considered, including a logistic regressor at all three stages, an AND function that mapped the atom-level to the pair-level, and a max layer that used a molecule's maximum PQS as its molecule-level score. For all three stages, adding a hidden layer offered either higher classification performance or a better scaled prediction than alternative methods.

5.2.5 Quantum Chemical Descriptors

Currently, there are no other methods for explicitly predicting quinone formation. However, there are several computational tools available for predicting sites of metabolism, which frequently use quantum chemical calculations that correlate to the activation energies of cytochromes P450.[52, 209, 137, 35, 36, 37] Due to the absence of other methods for predicting quinone formation, we computed several quantum chemical descriptors to establish baseline performances (Tables 5.1 and 5.2).

Table 5.1: Atom-level quantum chemical reactivity indices.

Label	Definition
$\pi^S(r)$	self-polarizability
CHARGE	partial charge on atom
$D^N(r)$	nucleophilic delocalizability
$D^E(r)$	electrophilic delocalizability
D_{LUMO}	E_{LUMO} density on atom
$D_{\text{LUMO}+1}$	$E_{\text{LUMO}+1}$ density on atom
D_{HOMO}	E_{HOMO} density on atom
$D_{\text{HOMO}-1}$	$E_{\text{HOMO}-1}$ density on atom
ED	electron density on atom
ERE	Electronic resonance energy
EEE	Electronic exchange energy
EERE	one-center electron-electron repulsion energy
ENAE	one-center electron-nuclear attraction energy
EE	EERE + ENAE
FarthestBondedHydrogen	distance to the farthest hydrogen bound to atom
FarthestBondedHydrogenIndicator	indicates whether or not an atom is bound to a hydrogen
NNRE	nuclear-nuclear repulsion energy
C	coulomb interaction energy
MCHARGE	mulliken charge
MPOP	mulliken population
POP _S	S-orbital electron population
POP _P	P-orbital electron population
POP _D	D-orbital electron population

Table 5.2: Molecule-level quantum chemical reactivity indices.

Label	Definition
η	electron hardness: $[E_{\text{HOMO}} - E_{\text{LUMO}}] / 2$
μ	chemical potential: $[E_{\text{HOMO}} + E_{\text{LUMO}}] / 2$
ω	electrophilic index: $\mu / 2\eta$
ΔH_f°	heat of formation (kcal/mol)
AREA	squared area (angstrom)
VOLUME	cubic volume (angstrom)
DIP	dipole moment (debye)
EE	electronic energy (eV)
E_{LUMO}	energy of the lowest unoccupied molecular orbital
$E_{\text{LUMO}+1}$	energy of the second lowest unoccupied molecular orbital
E_{HOMO}	energy of the highest occupied molecular orbital

$E_{\text{HOMO}-1}$	energy of the second highest unoccupied molecular orbital
IP	ionization potential
$\max [D^N(r)]$	maximum atom nucleophilic delocalizability
$\max [D^E(r)]$	maximum atom electrophilic delocalizability
$\max [\pi^S(r)]$	maximum atom self-polarizability
NE	nuclear energy (eV)
num_α_E	number of α electrons
num_β_E	number of β electrons
S	electron softness: $1/\eta$

These descriptors were calculated by MOPAC, a quantum chemistry package that performs self-consistent field computations, with the COSMO implicit solvent model and the semi-empirical method PM7.[73, 74] On both the pair- and molecule-level, we compared the performances of quantum chemical descriptors to those of the quinone formation predictions from the model, which only used topological descriptors. Atom-level quantum chemical descriptors were mapped to the pair-level by multiplying the descriptor values of both atoms together.

5.3 Results and Discussion

The following sections analyze the performance and applications of the quinone formation model. First, we quantified the power of pair quinone formation scores (PQS) to predict quinone-formation atom pairs within quinone-forming molecules. Second, we evaluated the ability of molecule quinone formation scores (MQS) to separate quinone-forming and non-quinone-forming molecules. Third, we used the molecule-level score to identify drugs that are likely to form quinones, but that have not previously been reported to do so. Fourth, we investigated whether MQS can separate drugs containing the same structural alert. Fifth, we screened for drugs that are predicted to form quinones, but are not currently known to do so. Finally, we studied how the model can potentially direct rational drug modifications to prevent quinone formation.

5.3.1 Accuracy at Predicting Quinone-Forming Pairs

To minimize reaction metabolite formation during drug design, identifying a molecule's QP is critical. Additionally, knowledge of the QP—the exact pair of carbons involved in quinone formation—provides a precise hypothesis about the source of a molecule's toxicity. Information about QPs also suggests where a molecule could be modified to prevent quinone formation, thereby minimizing a key toxicity risk. The quinone formation model is the first published computational method that specifically predicts QPs across a wide range of molecules.

To predict QPs, the model was trained in two stages. First, the model computed an atom quinone formation score (AQS) for each ring carbon within a test molecule. The atom-level scores ranged from zero to one, reflecting the probability that an atom was involved in quinone formation. The weights of the network were trained such that experimentally-known atoms of quinone formation receive high scores, and all other atoms received low scores.

The atom-level scores were intended to predict whether a single atom is part of a quinone formation reaction. However, because quinone formation always operates on a pair of two atoms, the atom-level scores did not directly predict quinone formation. Therefore, a second training step maps the atom score to the pair level, producing a pair level score. Within each molecule, this step considered all pairs of ring carbons as possible QPs. For each pair of atoms, this training step took as input two atom level score (corresponding to each atom), as well as pair-level descriptors. The weights of this stage of the network were trained to assign QPs with high pair-level scores, and all other pairs with low scores.

For both the atom-level and pair-level training steps, 10-fold cross-validation was used, a standard procedure in machine learning for estimating generalization accuracy. Within quinone-forming molecules, accurate predictions should differentiate QP from all other pairs. We used two metrics to quantify the accuracy of cross-validated atom- and pair-level scores. First, we calculated the "average pair AUC" by computing the area under the receiver operating characteristic curve (AUC) for each molecule and averaging the AUCs across all molecules to measure performance for the entire data set.[108, 121, 208] Second, we calculated "top-two"

performance—which considers a molecule correctly predicted if any of its QPs are predicted in the first or second rank positions—a standard metric for site of metabolism predictions.[121, 50, 122, 123, 52]

To assess the performance of the atom level model, we translated the atom-level scores to the pair-level by multiplying the scores of both atoms together, thereby calculating the probability that both atoms were involved in quinone formation. By this approach, the atom-level scores computed by a neural network yielded an average pair AUC performance of 97.1%, and a top-two performance of 83.8% (Figure 5.4). In contrast, a logistic regressor computed significantly less accurate predictions, with performances of 94.4% (average pair AUC) and 70.2% (top-two). Based on this result, we selected the neural network for the atom-level training stage.

For the second training stage—on the pair-level—we similarly tested both a neural network and a logistic regressor. The neural network predicted QP with performances of 97.6% and 86.9%, for average pair AUC and top-two, respectively. These did not exceed the performances of the logistic regressor: 97.8% (average pair AUC) and 88.3% (top-two). However, both the pair-level neural network and the pair-level logistic regressor significantly improved performance compared to only training on the atom-level, especially by the top-two metric. For example, comparing the top-two performances of the pair-level logistic regressor and the atom-level neural network by a paired *t*-test[175] yielded a *p*-value of 0.037. Consequently, we retained the pair-level training and model.

However, at the pair level, choosing between the neural network and the logistic regressor was more difficult, because they had statistically indistinguishable classification performances. Paired *t*-tests[175] indicated that the neural network and the logistic regressor were statistically equivalent for both average pair AUC (*p*-value 0.339) and top-two (*p*-value 0.286). To break the tie, we constructed reliability plots, which quantify how well predictions correlate to probabilities (Figure 5.5).[215, 121] These plots revealed that the neural network is a better scaled prediction than the logistic regressor, with the lowest root-mean-square error to a perfectly scaled prediction. Well-scaled predictions are interpretable as probabilities, so we decided to use the

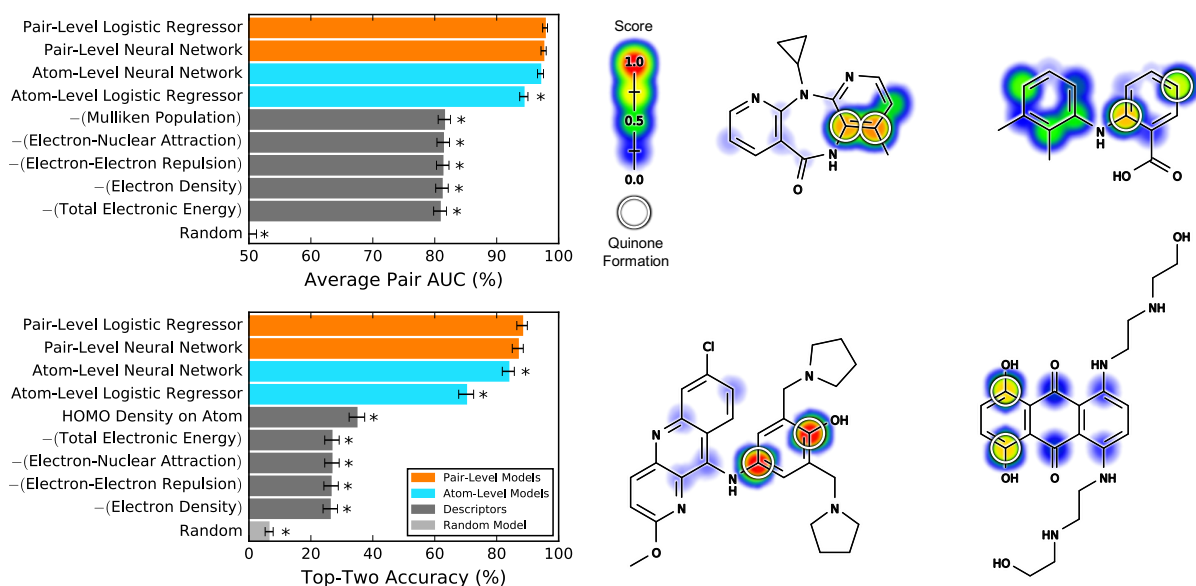


FIGURE 5.4: The pair-level model, computing pair quinone formation scores (PQS), accurately predicted quinone-forming pairs (QP). Top left, average pair AUC was calculated for 359 molecules extracted from the Accelrys Metabolite Database and labeled with their QPs. Within these quinone-forming molecules, the average pair AUC metric quantifies how frequently QP were ranked higher than other atom pairs.[108, 121, 208] Bottom left, the top-two metric was calculated, which counts a molecule correctly predicted if any of its QP received the first- or second-highest prediction.[121, 50, 122, 123, 52] By both metrics, the cross-validated pair-level scores outperformed atom quinone formation scores (AQS), computed by a control model that did not include training on the pair-level, instead mapping the atom-level to the pair-level by multiplying the scores of both atoms together. The pair-level model performance also exceeded that of quantum chemical descriptors, the five best of which are included in each panel. The baseline of each metric—indicated by the Random Model—was computed by calculating performance over randomized predictions. For each metric, asterisks indicate results significantly worse than the best-performing method, as determined by a paired *t*-test[175]. Examples from the data set are visualized with their scores (ranging from 0 to 0.98) indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including the atom. Top center, nevirapine,[210, 211] top right, mefenamic acid,[212] bottom center, pyronaridine,[213] and bottom right, mitoxantrone.[214] The atoms of experimentally-observed QPs are circled.

RingSize	% Total	% Single-ring
3	2.6	0.0
5	8.6	8.6
6	85.6	90.2
7	2.5	1.2
8	0.7	0.0

TABLE 5.3: Size of Rings with Atoms of Quinone Formation

neural network. Nevertheless, the choice between the neural network and the logistic regressor is arguable, because the logistic regressor is a simpler model structure. While we settled on the neural network for this component of the model, we believe a logistic regressor could likely be utilized with comparable results.

The majority of the QPs in the data were found in six-membered rings (Table 5.3), which raised concerns that we would not accurately predict quinone formation on five-membered rings due to insufficient data. With this in mind, we investigated whether QP in five-membered rings were also accurately predicted by calculating accuracy only on those molecules. Across the 36 molecules where a five-membered ring formed a quinone, pair-level scores predicted QPs with performances of 96.7% and 86.1%, for average pair AUC and top-two, respectively. These performances matched the accuracies across the whole data set of 97.6% (average pair AUC) and 86.9% (top-two). We accurately predicted QP in five-membered rings despite limited training data, suggesting that the underlying principles that guide quinone formation on six-membered rings also applied to quinone formation on five-membered rings.

Currently, there are no other models published for predicting quinone formation to which this model can be compared. Instead, to provide a baseline we calculated several quantum chemical descriptors (Tables 5.1 and 5.2), which correlate to the activation energies of cytochromes P450 and are often used by sites of metabolism predictors.[52, 209, 137, 35, 36, 37] Each quantum chemical descriptor was considered as a model to predict QP. To map each descriptor to the pair-level, the descriptor value of both atoms was multiplied together. None of the quantum chemical descriptors approached the accuracy of the pair-level score. For example,

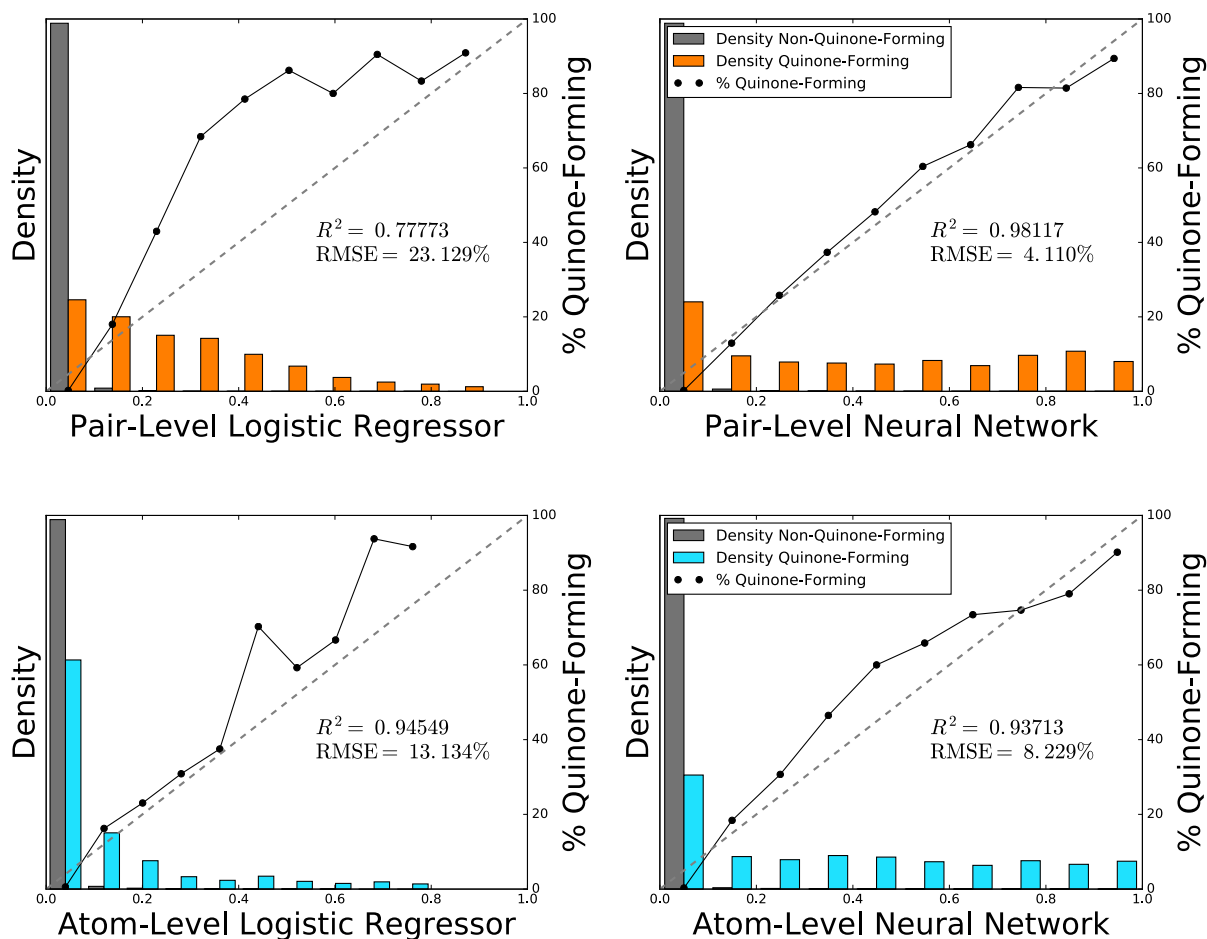


FIGURE 5.5: The pair-level neural network computes the best scaled prediction of quinone-forming pairs, and corresponds closely with a probability. Across 78341 atom pairs within 718 quinone-forming and non-quinone-forming molecules, the bar graphs plots the normalized distributions of the pair-level logistic regressor, the pair-level neural network, the atom-level logistic regressor, and the atom-level neural network. Using non-normalized frequencies, the solid lines in each bin plot the percentage of atom pairs that form quinones. Atom-level scores were mapped to the pair-level by multiplying the scores of both atoms together. On the pair-level, the logistic regressor and the neural network used the outputs of the atom-level neural network as inputs, along with pair-level descriptors. Each diagonal dashed line indicates a hypothetical perfectly scaled prediction. Of the four methods, the pair-level neural network provided the best scaled prediction, with the lowest root-mean-square error of a perfectly scaled prediction, and the highest correlation to the best-fit line. Therefore, the pair-level neural network most accurately reflected the probability that an atom pair will form a quinone.

the best descriptor by the top-two metric was the density of the highest occupied molecular orbital (HOMO) on the test atom, which only had a top-two performance of 34.8%, far below the pair-level model's performance of 86.9%. By considering many chemical attributes together using machine learning, the model predicted QPs more accurately than any quantum chemical descriptor.

5.3.2 Accuracy at Identifying Quinone-Forming Molecules

We also evaluated how well the model separated quinone-forming from non-quinone-forming molecules. A model that accurately predicts molecule quinone formation could be used to quickly screen drug candidates for a key toxicity risk.

In this assessment, we trained the model to differentiate molecules that formed quinones from those that did not. In addition to quinone-forming molecules, the training data contained molecules that could form quinones—defined as containing a six-membered aromatic ring—and that were metabolically studied in human-relevant experiments in the Accelrys Metabolite Database (AMD), but did not form quinones. Following pair-level training, we investigated several methods of discriminating between quinone-forming and non-quinone-forming molecules (Figure 5.6). To measure performance at this objective, we calculated the AUC across the full data set (molecule AUC).

First, we tried the most straightforward method of predicting molecule quinone formation: taking each molecule's maximum cross-validated pair-level scores. This procedure distinguished quinone-forming and non-quinone-forming molecules with a molecule AUC of 88.5%. Second, we added a training step that submitted the top three pair-level scores, the top three atom-level scores, and all molecule descriptors to a neural network or a logistic regressor. Using 10-fold cross-validation, neither the neural network nor the logistic regressor outperformed the maximum pair-level score, with molecule AUCs of 88.2% (p -value 0.16) and 88.4% (p -value 0.31), respectively, with significance evaluated by a false positive rate paired t -test.[175]

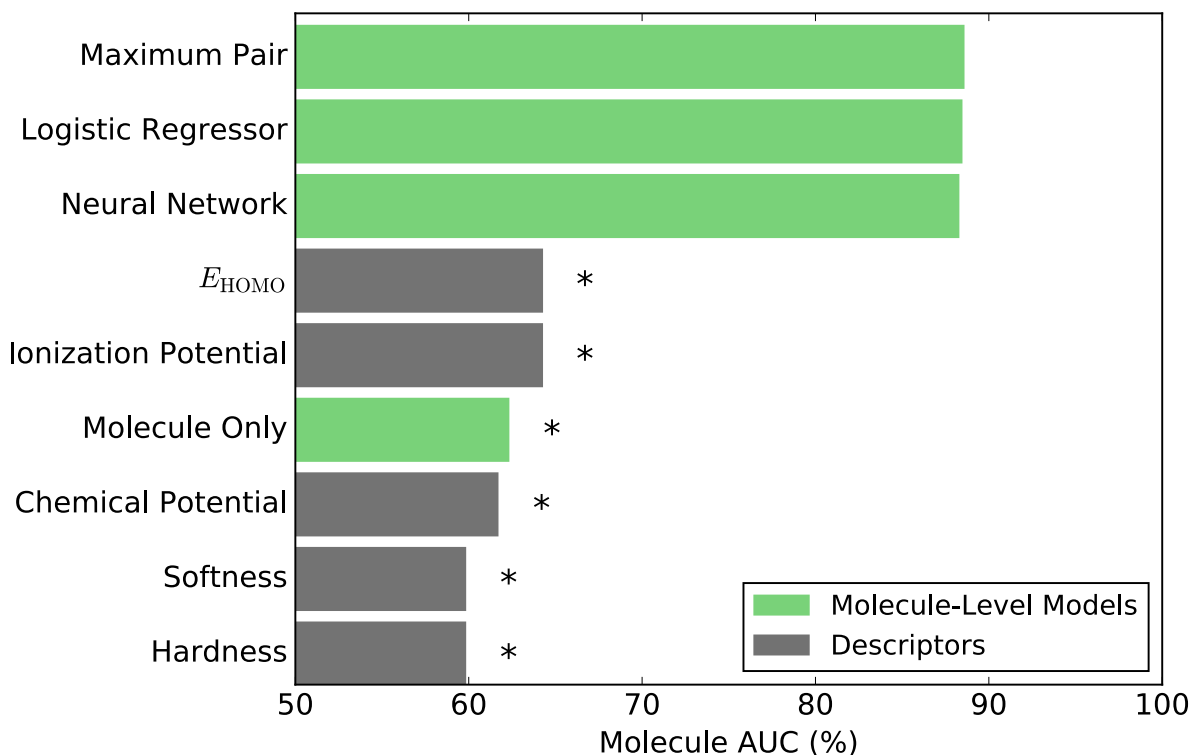


FIGURE 5.6: Molecule quinone formation scores (MQS) accurately identified quinone-forming molecules. The ability of several methods to identify molecules that form quinones was computed, including taking each molecule's maximum pair quinone formation score (PQS), training a logistic regressor or a neural network on the top three pair-level scores, the top three atom-level scores, and all molecule descriptors, and training a control model in the form of a neural network that only used as input molecule descriptors. The data set set included 359 quinone-forming molecules and 359 non-quinone-forming molecules. To quantify performance, the AUC was computed across quinone-forming and non-quinone-forming molecules (Molecule AUC), using cross-validated predictions. Asterisks indicate results significantly worse than the best-performing method, as determined by a false positive rate paired t -test.[175]

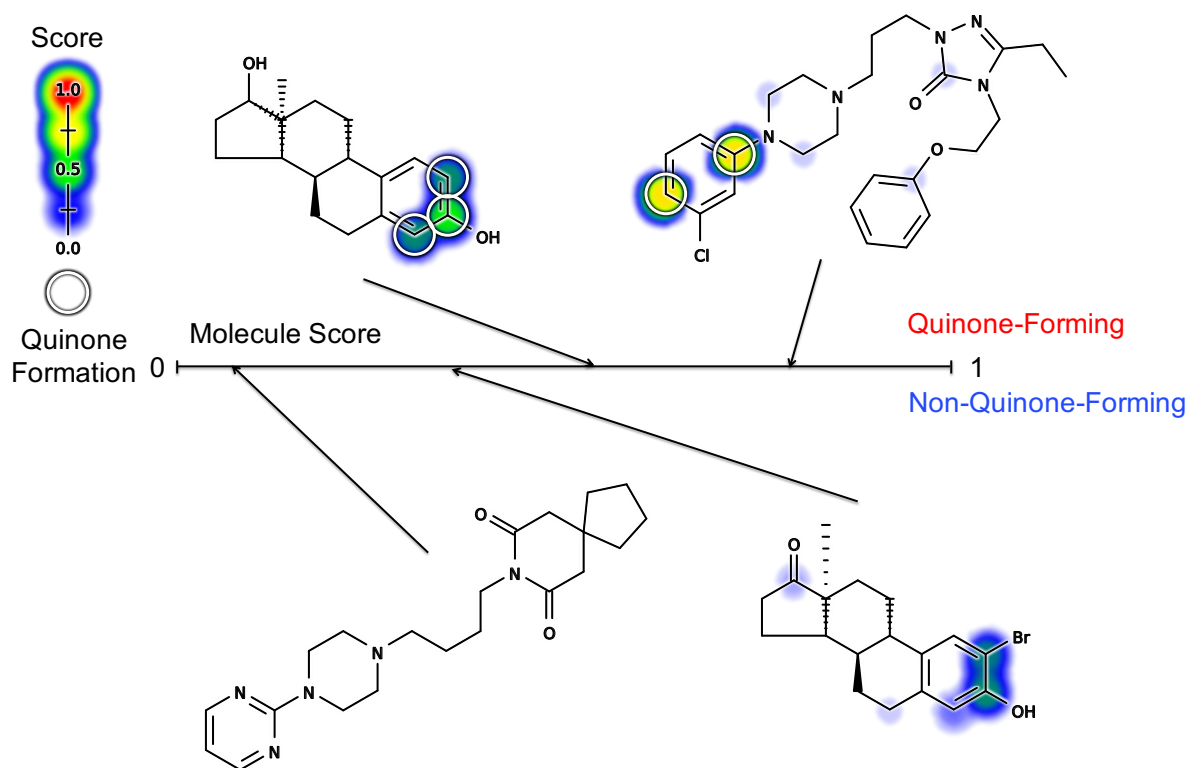


FIGURE 5.7: Example pairs of molecules that do and do not form quinones are visualized. Cross-validated scores (ranging from 0 to 0.78) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including the atom. The number line denotes each molecule's cross-validated score, computed by the neural network. From left to right, top to bottom: estradiol (MQS: 0.60)[216], nefazodone (MQS: 0.83)[205, 206], buspirone (MQS: 0.09)[205, 206], and a 2 α -substituted androstenedione analog (MQS: 0.42).[217] The atoms of experimentally-observed quinone-forming pairs are circled.

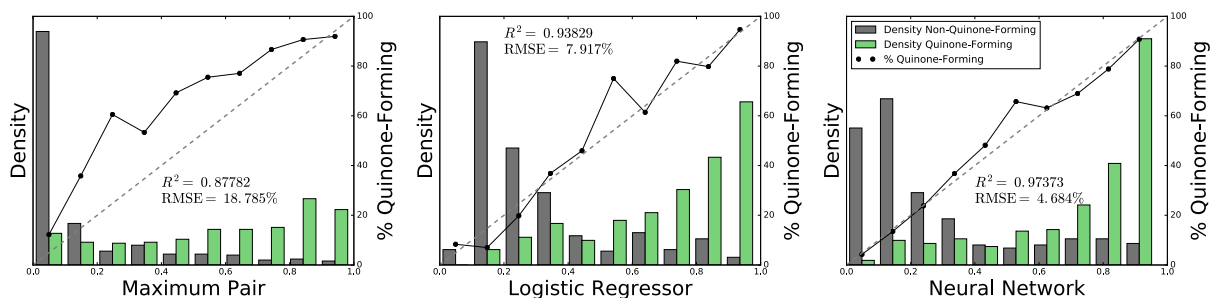


FIGURE 5.8: The neural network produces a well-scaled probabilistic prediction of molecule quinone formation. Across 718 quinone-forming and non-quinone-forming molecules, the bar graphs plot the normalized distributions of the maximum pair-level score, the molecule-level logistic regressor, and the molecule-level neural network. Using non-normalized frequencies, the solid lines in each bin plot the percentage of molecules that form quinones. Each diagonal dashed line indicates a hypothetical perfectly scaled prediction. The neural network provided the best scaled prediction, with the lowest root-mean-square error (RMSE) of a perfectly scaled prediction, and the highest correlation to the best-fit line. Therefore, the neural network most accurately reflected the probability that a molecule will form a quinone.

However, the neural network supplied a better scaled prediction than both the logistic regressor and the maximum pair quinone formation score, as indicated by the reliability plots of Figure 5.8. Furthermore, compared to both other methods, the neural network predicted more quinone-forming molecules with high scores, and non-quinone-forming molecules with low scores. On the other hand, both the logistic regressor or the maximum PQS were simpler models with equivalent molecule AUC performance. Still, because of its well-scaled predictions, we decided to use the neural network, but we expect that the other two methods would provide similar results.

Compared to the performance of pair-level model's performance in identifying QPs (97.6% AUC), the molecule-level model had a considerably lower AUC of 88.2%. This drop in accuracy might be because the molecule-level data contains more noise than the pair-level data. When selecting "non-quinone-forming" molecules, we inferred that a molecule did not form a quinone if it was not reported to do so in the AMD. This supposition, while obligatory for molecule-level training, was not a robust indication that molecules do not form quinones, because not every study measures quinone products. As a result, our data set construction procedure imprecisely

labeled some quinone-forming molecules as non-quinone-forming. In contrast, the pair-level quinone formation data was much cleaner, because it was extracted from experiments capable of reporting quinone formation. Our previous studies with the AMD have observed comparable drops in site-level (equivalent to the pair-level in this study) to molecule-level accuracy.[121, 208, 108]

Even so, the molecule-level scored classified quinone-forming molecules with 88.2% AUC, greatly exceeding that of any comparison quantum chemical descriptor and demonstrating that the model produces illuminating molecule-level predictions. The best quantum chemical descriptor was the energy of the highest occupied molecular orbital (E_{HOMO}), which only had an AUC of 64.2%. The relatively high performance of the molecule-level model supports our assumption that many of the molecules labeled as non-quinone-forming indeed do not form quinones. If the quinone-forming and non-quinone-forming molecules were actually extracted from identical chemical populations, separating them in our cross-validated experiments—which simulated performance on external data—would be highly unlikely.

To further confirm that molecule-level performance was not a consequence of overfitting, we evaluated the model on several external test sets of non-quinone-forming molecules. Each external test set was predicted by the final trained model. Separation between each external non-quinone-forming test set and the quinone-forming training molecules was measured by the molecule AUC. Over the 20 test sets, the AUC was $85.9\% \pm 1.2\%$, which is comparable to the previously reported cross-validated AUC of 88.2% computed by training on both quinone-forming and non-quinone-forming molecules. The model successfully generalized to new data, assigning non-quinone-forming molecules it had never seen before with much lower scores than quinone-forming training molecules.

5.3.3 Comparison to Structural Alerts

We compared the model to “structural alerts,” which are motifs known to commonly form reactive metabolites. A library of structural alerts can be easily used to flag potentially problematic

Motif	SMARTS	Quinone-Forming Pairs	Total Pairs
Nitrogen <i>ortho</i> to Oxygen	<chem>[c1([#7])c([#8])cccc1],[c1([#8])c([#7])cccc1]</chem>	3	34
Oxygen <i>para</i> to Hydrogen	<chem>[\$c1([#8])ccc([#1])cc1],[c1([#1])ccc([#8])cc1]</chem>	8	284
Nitrogen <i>ortho</i> to Carbon	<chem>[\$c1([#7])c([#6])cccc1],[c1([#6])c([#7])cccc1]</chem>	6	210
Nitrogen <i>para</i> to Nitrogen	<chem>[\$c1([#7])ccc([#7])cc1],[c1([#7])ccc([#7])cc1]</chem>	15	39
Oxygen <i>ortho</i> to Hydrogen	<chem>[\$c1([#8])c([#1])cccc1],[c1([#1])c([#8])cccc1]</chem>	57	1038
Oxygen <i>ortho</i> to Carbon	<chem>[\$c1([#8])c([#6])cccc1],[c1([#6])c([#8])cccc1]</chem>	6	344
<i>para</i> to Phenol	<chem>[\$c1ccc([OH1])cc1],[c([OH1])1cccc1]</chem>	108	433
Oxygen <i>para</i> to Carbon	<chem>[\$c1([#8])ccc([#6])cc1],[c1([#6])ccc([#8])cc1]</chem>	33	345
Nitrogen <i>para</i> to Carbon	<chem>[\$c1([#7])ccc([#6])cc1],[c1([#6])ccc([#7])cc1]</chem>	6	50
Hydroquinone	<chem>[\$c1([#8][#1])ccc([#8][#1])cc1],[c1([#8][#1])ccc([#8][#1])cc1]</chem>	27	27
<i>ortho</i> to Phenol	<chem>[\$c1c([OH1])cccc1],[c([OH1])1cccc1]</chem>	79	892
Nitrogen <i>ortho</i> to Hydrogen	<chem>[\$c1([#7])c([#1])cccc1],[c1([#1])c([#7])cccc1]</chem>	12	495
Nitrogen <i>para</i> to Oxygen	<chem>[\$c1([#7])ccc([#8])cc1],[c1([#8])ccc([#7])cc1]</chem>	41	81
Oxygen <i>para</i> to Oxygen	<chem>[\$c1([#8])ccc([#8])cc1],[c1([#8])ccc([#8])cc1]</chem>	33	57
Oxygen <i>ortho</i> to Oxygen	<chem>[\$c1([#8])c([#8])cccc1],[c1([#8])c([#8])cccc1]</chem>	51	196
Nitrogen <i>ortho</i> to Nitrogen	<chem>[\$c1([#7])c([#7])cccc1],[c1([#7])c([#7])cccc1]</chem>	1	24
Nitrogen <i>para</i> to Hydrogen	<chem>[\$c1([#7])ccc([#1])cc1],[c1([#1])ccc([#7])cc1]</chem>	34	174
Halogen <i>para</i> to Carbon or Nitrogen or Oxygen	<chem>[\$c1([F,Cl,Br,I])ccc([#6,#7,#8])cc1],[c1([#6,#7,#8])ccc([F,Cl,Br,I])cc1]</chem>	109	109
Catechol	<chem>[\$c1([#8][#1])c([#8][#1])cccc1],[c1([#8][#1])c([#8][#1])cccc1]</chem>	75	75
Not Matched		417	74815

TABLE 5.4:

We analyzed the frequencies of quinone structural alerts using SMARTS strings (Table 5.4). We matched each training molecule against each SMARTS string. If there was a match, we extracted all matching atoms. For each pair of matching atoms, we retained only those pairs where both atoms were in the same ring and were the correct distance away from one another. In the table, the total number of pairs that form quinones are listed in the “Quinone-Forming Pairs” column, and the total number of pairs recorded in the “Total Pairs” column.

molecules during drug development.^[9, 218] This is a widely used approach, but unfortunately structural alerts for quinone formation—such as anilines or phenols—are also found in many safe drugs, because they are not bioactivated due to specific molecular context. We would hope that the model could identify which structural alerts are bioactivated to quinones and which are not.

To systematically evaluate whether the model distinguishes structural alerts, we focused on several motifs known to commonly form quinones. The exact patterns used are listed in Table 5.4. We included both motifs that form quinones in one step, such as hydroquinones and catechols, and motifs that form quinones in two steps, such as an oxygen or nitrogen *ortho* or *para* to a hydrogen. For each motif, we extracted all atom pairs from the training data set that matched the structure, and recorded whether this pair formed a quinone. Next, we assigned each pair its cross-validated pair-level score, and calculated the AUC across all pairs of that substructure, the pair AUC (Figure 5.9 and Table 1.4). For each structural alert, the model predicted quinone formation with performances significantly higher than random.

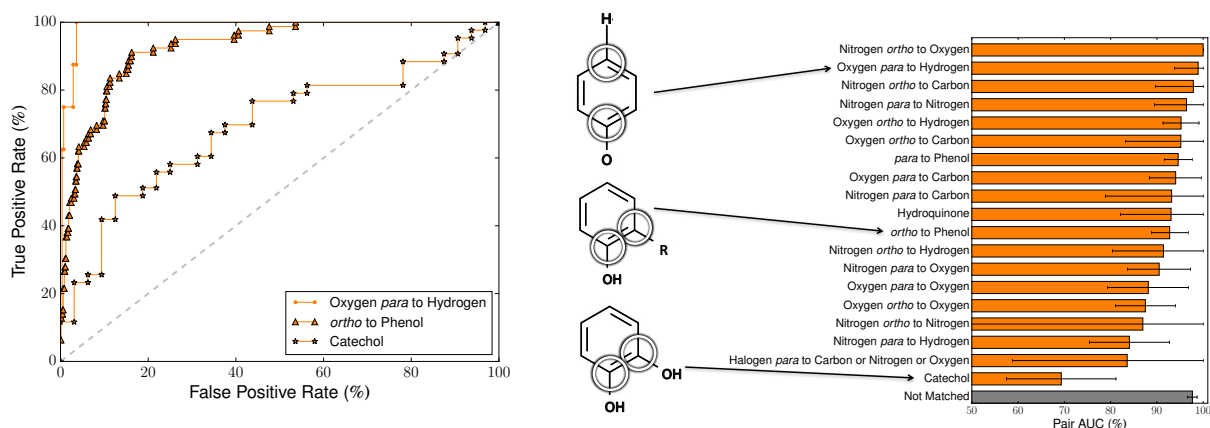


FIGURE 5.9: Pair-level scores predicted whether several structural alerts form quinones. We tested several motifs that are structural alerts for both one- and two-step quinone formation. For each motif, across the 718 training molecules, the cross-validated scores of all atoms pairs matching that motif were extracted. Left, the receiver operating characteristic (ROC) curve is displayed across all the atom pairs of three example motifs (visualized in the center). The graph displays example ROC curves for the second best, typical, and worst performing alerts. The diagonal dashed line indicates baseline performance. The ROC curves of all the motifs are displayed in Figure 5.10. In the depiction of the *ortho* to phenol motif, the “R” represents any atom, including hydrogens. Right, the area under the ROC curve (AUC) was calculated across all the atom pairs of each motif. The error bars represent 95% two-sided confidence intervals.[219]

Motif	Category	Quinone-Forming Pairs	Non-Quinone-Forming Pairs	Pair AUC (%)
Nitrogen <i>ortho</i> to Oxygen	One Step	3	31	100.0
Oxygen <i>para</i> to Hydrogen	Two Step	8	276	98.9
Nitrogen <i>ortho</i> to Carbon	One Step	6	204	97.9
Nitrogen <i>para</i> to Nitrogen	One Step	15	24	96.4
Oxygen <i>ortho</i> to Hydrogen	Two Step	57	981	95.2
Oxygen <i>ortho</i> to Carbon	One Step	6	338	95.2
<i>para</i> to Phenol	One and Two Step	108	325	94.6
Oxygen <i>para</i> to Carbon	One Step	33	312	94.1
Nitrogen <i>para</i> to Carbon	One Step	6	44	93.2
Hydroquinone	One Step	24	3	93.1
<i>ortho</i> to Phenol	One and Two Step	79	813	92.8
Nitrogen <i>ortho</i> to Hydrogen	Two Step	12	483	91.4
Nitrogen <i>para</i> to Oxygen	One Step	41	40	90.5
Oxygen <i>para</i> to Oxygen	One Step	33	24	88.1
Oxygen <i>ortho</i> to Oxygen	One Step	51	145	87.5
Nitrogen <i>ortho</i> to Nitrogen	One Step	1	23	87.0
Nitrogen <i>para</i> to Hydrogen	Two Step	34	140	84.1
Halogen <i>para</i> to Carbon or Nitrogen or Oxygen	Two Step	4	105	83.6
Catechol	One Step	43	32	69.3
Not Matched		417	74398	97.7

TABLE 5.5: P

air-level scores predicted whether structural alerts form quinones.

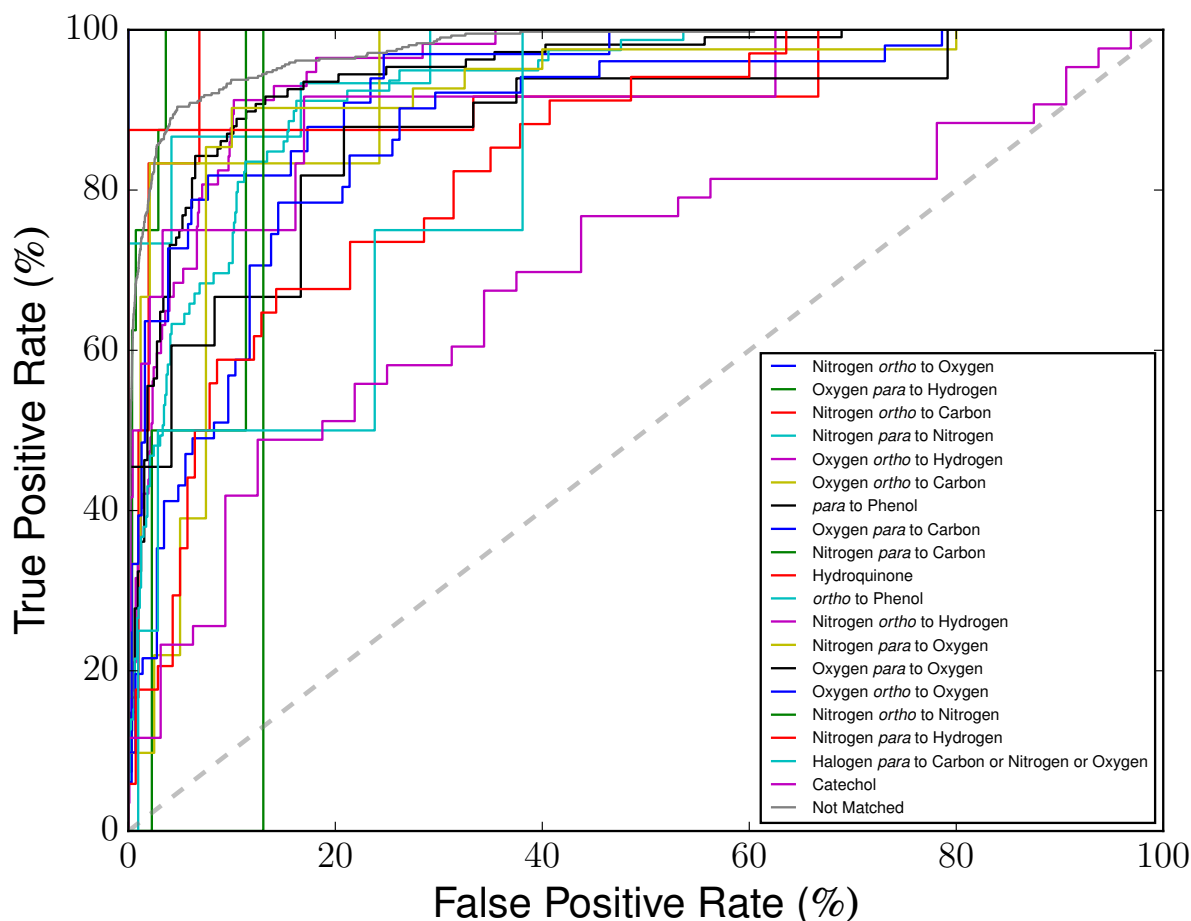


FIGURE 5.10: Pair-level scores predicted whether several structural alerts form quinones. We tested several motifs (Table 5.4) that are structural alerts for both one- and two-step quinone formation. For each motif, across the 718 training molecules, the cross-validated scores of all atoms pairs matching that motif were extracted. The receiver operating characteristic (ROC) curve is displayed across all the atom pairs of each motif. The diagonal dashed line indicates baseline performance: 50% area under the ROC curve (AUC).

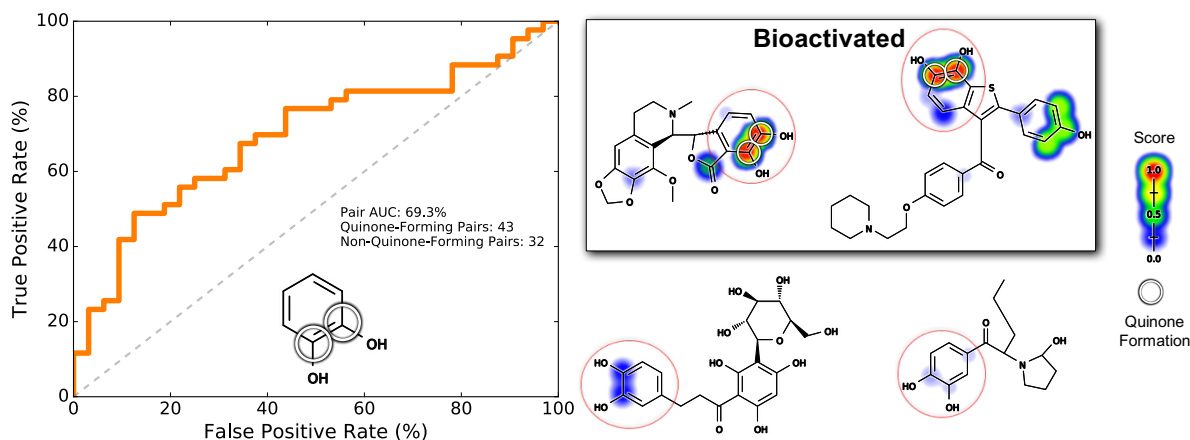


FIGURE 5.11: Pair-level scores predicted whether catechols form quinones. Of all structural alerts tested, the model performed the worst on catechols (Figure 5.9). Nevertheless, the model distinguished catechols with a statistically significant performance. This motif (indicated by the red circles) requires a single step to form a quinone. Across the 718 training molecules, the cross-validated scores of all atoms pairs matching this motif were extracted. Left, the receiver operating characteristic (ROC) curve is displayed across all these atom pairs. The diagonal dashed line indicates baseline performance: 50% area under the ROC curve (AUC). Four example molecules are visualized. Scores (ranging from 0 to 0.97) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including the atom. The atoms of experimentally-observed quinone-forming pairs are circled. Top center, a noscapine metabolite[221] and top right, a raloxifene metabolite[222], both of which are correctly predicted to form quinones at the catechol. Bottom center, aspalathin[223] and bottom right, a methylenedioxypyrovalerone metabolite[224] neither of which form quinones at the catechol-*para*-to-hydrogen motifs, which are correctly assigned low scores.

For example, the catechol motif commonly forms an *ortho* quinone in a single step (Figure 5.11).[9, 220, 218] Catechols might be considered a motif that “obviously” forms quinones. However, of the 75 catechols in our data set, only 43 actually are known to form quinones.

The quinone formation model’s pair-level scores separated quinone-forming and non-quinone-forming catechols with a pair AUC of 69.3%. This was the lowest performance of all the structural alerts we tested, and there is certainly room for improvement. Nevertheless, as indicated by the 95% confidence intervals[219] in Figure 5.11, we separated catechols with a performance significantly higher than baseline (AUC 50%). We find this an encouraging result, because

the method helps identify which catechols might not actually form quinones, unlike the structural alerts approach that considers all catechols a bioactivation risk without evaluating specific molecular context.

We also found that pair-level scores are informative for two-step quinone formation structural alerts. For example, for the motif of a nitrogen *para* to a hydrogen to form a quinone, a hydroxylation must first occur at the unsubstituted carbon, followed by quinone formation. The nitrogen-*para*-to-hydrogen motif can be found in some anilines or anilides, which are well known structural alerts.[9, 22, 218] The entire training data set contained 174 total nitrogen-*para*-to-hydrogen atom pairs, of which 34 are actually known to form quinones. Encouragingly, the quinone formation model's pair-level scores accurately separated quinone-forming and non-quinone-forming nitrogen-*para*-to-hydrogen groups with a pair AUC of 84.1% (Figure 5.12).

The phenol motif is a broad structural alert found in a wide variety of drugs.[9, 218] In contrast to catechols (which require one step to form a quinone), and nitrogen-*para*-to-hydrogen (which require two steps), phenols can form quinones in one or two steps, depending on ring substituents. For example, phenols can form a quinone-methide in a single step if there is a *para* carbon, or form a quinone in two steps via intermediate aromatic hydroxylation at the *para* position. Across our training data set, there are 433 carbons *para* to a phenol. For each atom, we extracted the atom-pair-level score of the carbon and the ipso-carbon of the corresponding phenol. Of these atom pairs, 108 form a *para* quinone.

The model accurately distinguished whether a quinone will form *para* to a phenol, with a pair AUC of 94.6% (Figure 5.13). Both one- and two-step quinone formation at the *para* position to a phenol were accurately predicted. For example, both oxymetazoline and an analog of androstenedione contain a phenol-*para*-to-carbon motif, but only in oxymetazoline does the phenol-*para*-to-carbon motif form a quinone-methide.[228, 229] The model's pair-level scores cleanly distinguished the phenol-*para*-to-carbon in both molecules. Similarly, both propofol and doxycycline contain a phenol-*para*-to-hydrogen, which can form a quinone in two

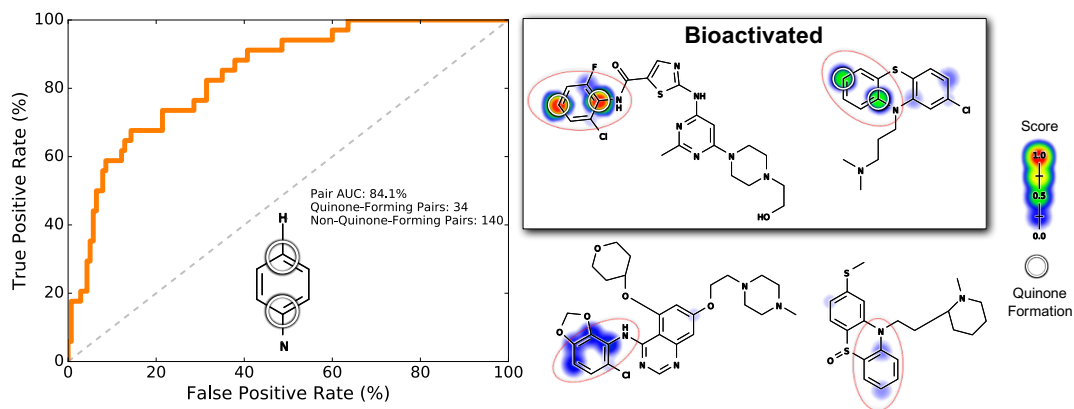


FIGURE 5.12: Pair-level scores predicted whether nitrogen-bound carbons *para* to a carbon with a hydrogen form quinones. This motif (indicated by the red circles) requires aromatic hydroxylation to form a quinone. Across the 718 training molecules, the cross-validated scores of all atoms pairs matching this motif were extracted. Left, the receiver operating characteristic (ROC) curve is displayed across all these atom pairs. The diagonal dashed line indicates baseline performance: 50% area under the ROC curve (AUC). Four example molecules are visualized. Scores (ranging from 0 to 0.96) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including that atom. The atoms of experimentally-observed quinone-forming pairs are circled. Top center, a dasatinib analog^[225] and top right, chlorpromazine^[226], both of which are correctly predicted to form quinones at the nitrogen-*para*-to-hydrogen motif. Bottom center, saracatinib^[225] and bottom right, a thioridazine metabolite^[227] neither of which form quinones at the nitrogen-*para*-to-hydrogen motifs, and are correctly assigned low scores.

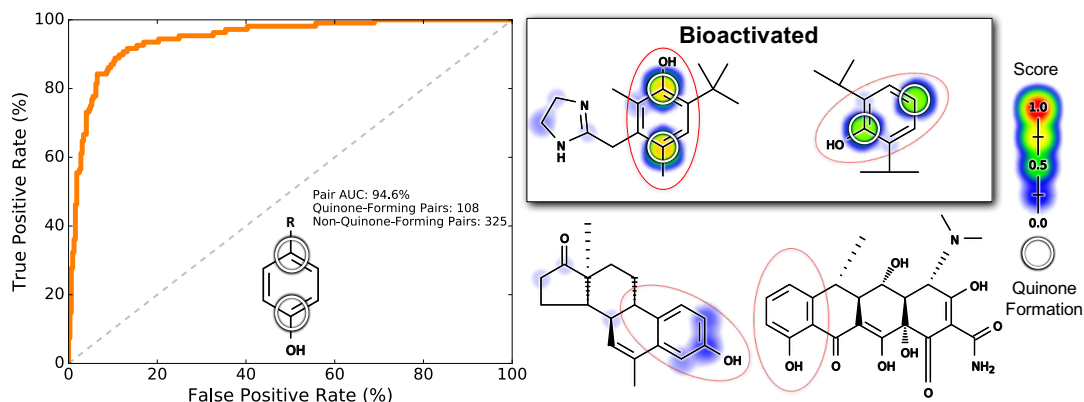


FIGURE 5.13: Pair-level scores predicted whether a quinone will form *para* to a phenol. This motif (indicated by the red circles) may require one or two metabolic steps to form a quinone. In the depiction of the phenol motif on the left, the “R” represents any atom, including hydrogens. Across the 718 training molecules, the cross-validated scores of all atoms pairs matching this motif were extracted. Left, the receiver operating characteristic (ROC) curve is displayed across all these atom pairs. The diagonal dashed line indicates baseline performance: 50% area under the ROC curve (AUC). Four example molecules are visualized. Scores (ranging from 0 to 0.77) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including that atom. The atoms of experimentally-observed quinone-forming pairs are circled. Top center, oxymetazoline[228] and top right, propofol[230], both of which are correctly predicted to form quinones at the *para*-to-phenol motif. Bottom center, an androstenedione analog[229] and bottom right, doxycycline[231] neither of which form quinones at their *para*-to-phenol motifs, and are correctly assigned low scores.

steps with an intermediate hydroxylation. Propofol’s phenol-*para*-to-hydrogen received a much higher pair-level score than doxycycline, consistent with only propofol being known to form a quinone.[230, 231]

These results demonstrate that the model has better specificity than structural alerts. Specific substructures can determine the parts of molecules that have the capacity to form quinones, but they do not assess the likelihood of the quinone actually forming. The quinone formation model, however, can stratify molecules by their propensity of forming quinones, identifying which alerts are bioactivated, and which are not.

Structural alerts are retrospective in nature. Consequently, they cannot detect quinone formation for motifs that have not previously been observed to form quinones. As a result,

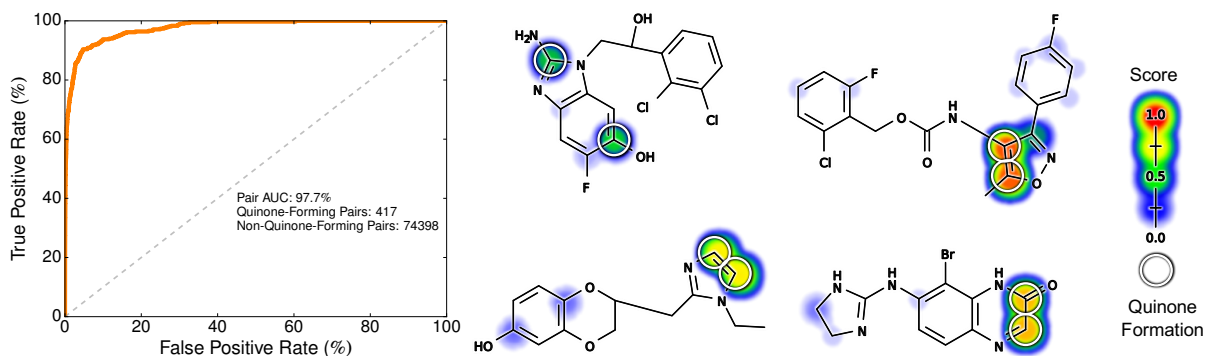


FIGURE 5.14: Pair-level scores predicted whether quinones form at motifs that do not match structural alerts. Across the 718 training molecules, the cross-validated scores of all atoms pairs that did not match any structural alert were extracted. Left, the receiver operating characteristic (ROC) curve is displayed across all these atom pairs. The diagonal dashed line indicates baseline performance: 50% area under the ROC curve (AUC). Four example molecules are visualized. Top center, an antimalarial drug candidate.[124] Top right, a voltage-gated sodium channel inhibitor drug candidate.[232] Bottom center, an imiloxan metabolite.[233] Bottom right, a brimonidine metabolite.[234] The atoms of experimentally-observed quinone forming pairs are circled. Scores (ranging from 0 to 0.92) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including the atom.

molecules that form quinones at atypical sites will likely be missed by structural alerts. In contrast, our model can potentially detect these unusual cases, because it is built on a diverse data set that includes a wide variety of typical and atypical sites of quinone formation. To test this, we investigated the accuracy of the model at predicting quinone formation at sites that do not match structural alerts. After filtering out all atom pairs that matched any of the quinone structural alerts used (Table 1.4), we calculated the pair AUC over all remaining atom pairs. The model accurately predicted quinone formation at these non-obvious sites, with a pair AUC of 97.7% (Figure 5.14). This indicates that the model is able to accurately predict quinone formation even for atom pairs that would not generally be considered to be at risk.

5.3.4 Case Studies

We considered several case studies to investigate the model's utility. One potential application is to screen for molecules that are likely to form quinones, but that are currently not known

to do so. To explore this application, we predicted the quinone formation of drugs that were not reported to form quinones in the AMD. After downloading a database of FDA-approved and withdrawn drugs, the molecules present in the training data set were removed.^[235] Next, each drug was submitted to the trained quinone formation model, thereby assigning a molecule quinone formation score to each drug.

To define a cutoff above which a drug was considered predicted to form a quinone, we used the optimal point on the cross-validated molecule ROC curve from the training data (Figure 5.15).

This optimal point—corresponding to a molecule score of 0.515—is the cutoff that offers the best trade-off between sensitivity and specificity. We selected all drugs that were assigned a higher score than the optimal cutoff. This revealed 288 drugs that are predicted to form quinones, but that are currently thought to be non-quinone-forming across all the relevant literature in the AMD. All 288 drugs and their molecule quinone formation scores are available in the Supporting Information. Here, we highlight three example drugs (Figure 5.16). Each drug carries a risk of idiosyncratic toxicity, the causes of which are currently obscure.

First, the antiarrhythmic flecainide is associated with idiosyncratic hepatotoxicity, the origin of which is currently unknown.^[239, 9, 22] There are no known reactive metabolites of flecainide, and previous studies have suggested that flecainide is harmlessly metabolized to a phenol via oxygen-dealkylation, followed by glucuronidation.^[240, 9, 22] However, our model suggests that the phenol metabolite may actually form a quinone.

Second, the antihistamine thenalidine is associated with idiosyncratic cases of agranulocytosis, a potentially deadly condition that entails a severely lowered blood white cell count that leaves a patient vulnerable to critical infection.^[241, 242, 243, 244] Indeed, three years after thenalidine was introduced to the U.S. market, it was withdrawn in 1958 due to fatal cases of agranulocytosis.^[245, 246, 247, 248] The mechanism of thenalidine's idiosyncratic toxicity was never uncovered, and it has no known reactive metabolites. An obvious hypothesis

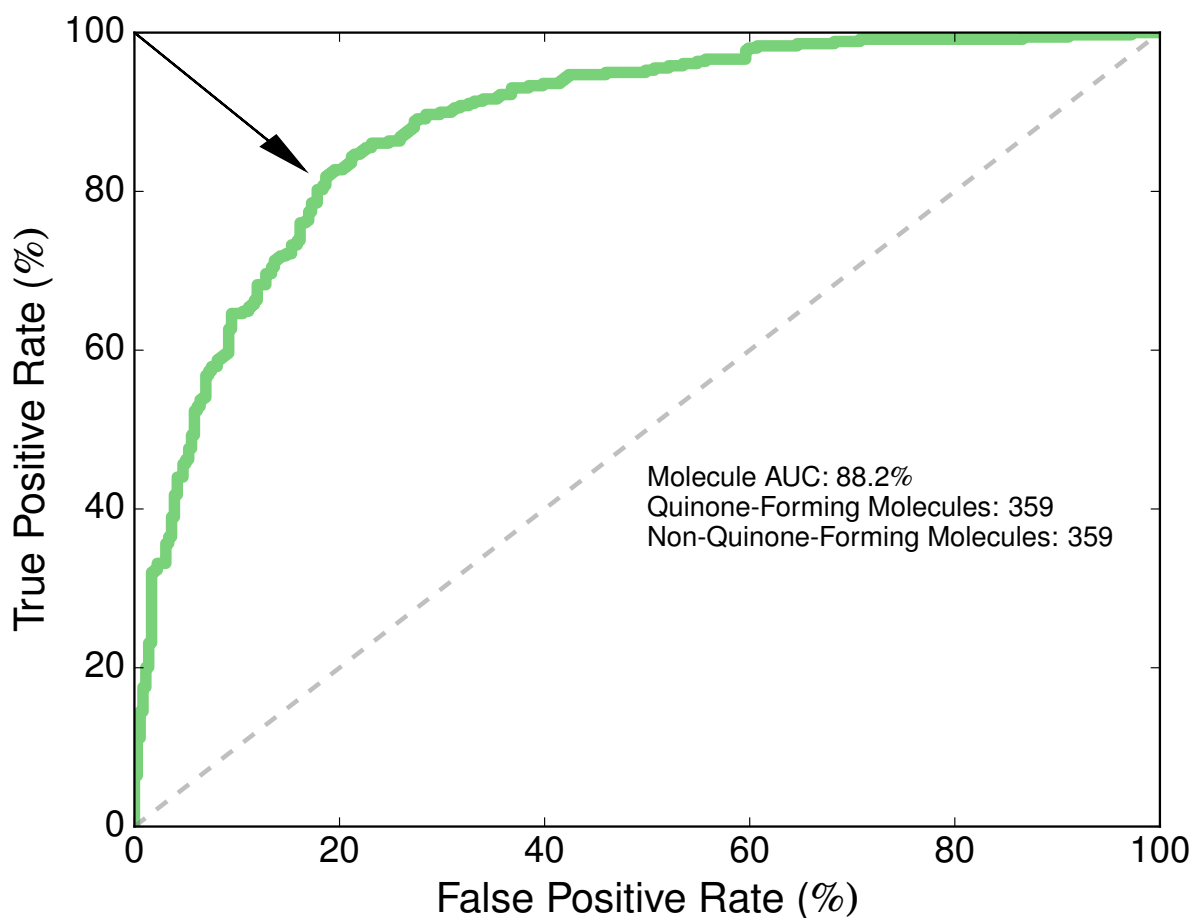


FIGURE 5.15: Molecule quinone formation scores accurately identified quinone-forming molecules. Each molecule was assigned its cross-validated score, computed by a neural network that took as input the top three pair-level scores, the top three atom-level scores, and all molecule descriptors (Figure 5.3). The receiver operating characteristic (ROC) curve is displayed across all molecules. The diagonal dashed line indicates baseline performance: 50% area under the ROC curve (AUC). The arrow points to the location on the ROC curve that offers the optimal tradeoff between sensitivity (the true positive rate) and specificity (one minus the false positive rate). This location corresponds to a cutoff molecule quinone formation score of 0.515.

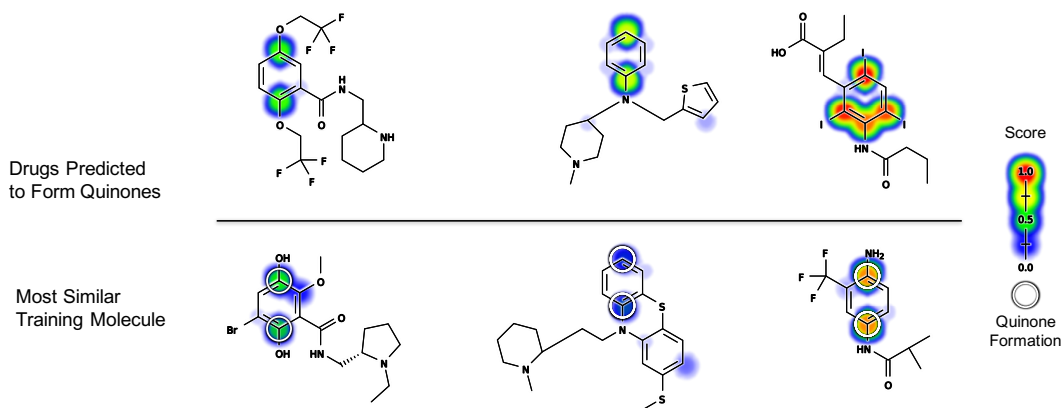


FIGURE 5.16: Examples of drugs associated with idiosyncratic toxicity that received high molecule quinone formation scores (MQS), but are not known to form quinones in our data set. Each drug predicted to form a quinone is visualized above its most similar quinone-forming molecule from the training data. Similarity was quantified using path based fingerprints and MinMax similarity, a modification of Tanimoto appropriate for count fingerprints.[236] Top, flecainide (MQS: 0.81), thenalidine (MQS: 0.79), and bunamiodyl (MQS: 0.96). Bottom, a remoxipride metabolite (MQS: 0.58),[237] thioridazine (MQS 0.36),[226] and a flutamide metabolite (MQS 0.88).[238] The atoms of experimentally-observed quinone forming pairs are circled. Scores (ranging from 0 to 0.99) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including the atom. For the training examples, cross-validated scores are reported.

is that thenalidine's thiophene ring—a well-known structural alert—could undergo bioactivation through epoxidation or sulfur-oxidation.[249, 9] However, the quinone model's predictions yield a second, more surprising hypothesis: thenalidine may undergo hydroxylation at the *para* position on its aniline ring, followed by formation of a potentially highly reactive quinone imine.

Third, bunamiodyl carries a risk of severe nephropathy.[245, 250] It was introduced in 1958 as a cholecystographic contrast medium, but was withdrawn in 1964 after around one hundred patients died due to renal failure.[251, 248, 245, 250] While bunamiodyl is not known to produce reactive metabolites, our model predicts that it undergoes dehalogenation followed by formation of an *ortho* or *para* quinone imine.

As seen for flecainide, thenalidine, and bunamiodyl, the model can yield an explicit, testable hypothesis about the mechanisms of a molecule's toxicity. Experimentally validating these quinone formation predictions is beyond the scope of this study, but is planned for future work.

In another application, we see that the model's scores accurately reflect the impact of rational drug modifications. Often, small modification to a molecule can prevent reactive metabolite formation while retaining drug efficacy. For example, the antimalarial drug amodiaquine forms a reactive quinone imine metabolite that causes the hepatotoxicity and agranulocytosis sometimes associated with amodiaquine treatment, severely limiting its use.[9, 252, 253, 254] Amodiaquine has an analog that retains pharmacologic efficacy while preventing reactive metabolite formation.[9, 255, 256] The quinone formation model detects the effect of this subtle change, cleanly distinguishing amodiaquine from its safe analog (Figure 5.17). The quinone formation model also correctly predicts amodiaquine's QP, demonstrating how the model could help guide rational drug modifications that require specific knowledge about where on a molecule a reactive metabolite forms.

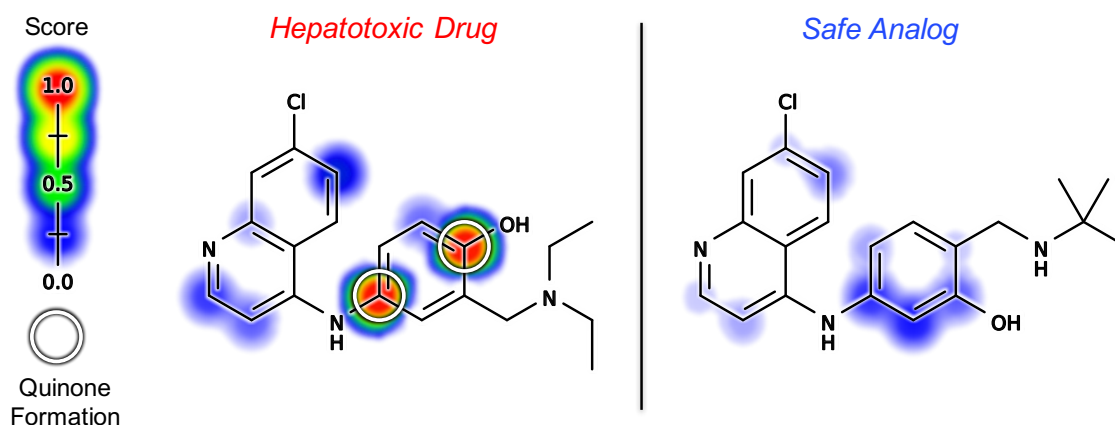


FIGURE 5.17: The model reflects the impact of structural modification to prevent quinone formation. The hepatotoxicity and agranulocytosis that have significantly curtailed the use of the antimalarial drug amodiaquine (left) have been traced to the formation of a reactive quinone imine metabolite. [9, 252, 253, 254] An amodiaquine analog (right) has been identified that retains pharmacologic efficacy while preventing reactive metabolite formation.[9, 255, 256] Molecule quinone formation scores (MQS) cleanly distinguished amodiaquine (MQS: 0.93) from this safe analog (MQS: 0.18). Additionally, pair-level quinone formation scores (PQS) correctly predicted amodiaquine's quinone formation pair, the atoms of which are circled. Scores (ranging from 0 to 0.98) are indicated by the colored shading, assigning each atom the probabilistic OR of all pair-level quinone formation scores including the atom. Cross-validated scores are reported for amodiaquine.

5.4 Model Limitations

Quinone formation is only one piece of the toxicity puzzle, and this model is only one-step towards effective toxicity management. For example, while the model predicts quinone formation, it does not predict the reactivity of those quinones. Subtle changes in ring substituents can have a large impact on quinone reactivity and toxicity.[40] Consequently, in future work we plan to incorporate the quinone formation model with already-developed reactivity models.[108, 208]

Moreover, the quinone formation model does not correct for alternate metabolic pathways that may detoxify molecules before they can form quinones. Due to these metabolic alternatives, quinone formation is sometimes observed *in vitro*, but does not occur *in vivo*.^[9] Furthermore, after quinones form, they can be further metabolized by reduction, which for example can produce hydroquinones that can then be conjugated and eliminated.^[257] In the long run, integration of the quinone formation model with models of reduction and detoxification, such as uridine 5'-diphospho-glucuronosyltransferase conjugation^[140], could offer more nuanced predictions by weighing quinone formation against other metabolic routes.

Likewise, quinone formation that occurs in several metabolic steps may not be detected by the current model, which focused on one- and two-step quinone formation. As we continue to develop more complex systems of metabolite structure prediction, we plan to expand the quinone formation model by explicitly modeling intermediate structures, potentially extending its utility beyond two metabolic steps. Additionally, the current method only uses topological descriptors. As evidenced by the quinone formation model's accuracy, these descriptors are tightly correlated with electronic structure, but performance on rare subclasses or new molecules could be affected by not explicitly including quantum chemical descriptors. In the future, we will consider expanding our method to include quantum chemistry, such as descriptors relating to the half-reactions between the substrate and Compound I within cytochromes P450.

Finally, there is no guarantee that the model's applicability domain extends beyond its training domain of metabolically studied drug-like molecules. New areas of chemical space—possibly only currently explored in proprietary data within pharmaceutical companies—may

not be well suited to the current topological-descriptor-based model trained on literature-derived data. However, this study's methodology could be easily applied to new data, thereby expanding the utility of the model.

5.5 Conclusions

This study constructed a novel method that forecasts the formation of reactive quinone metabolites. The QP-trained quinone formation model predicted with 97.6% AUC the QPs within quinone-forming molecules. These pair-level predictions suggest where drug candidates could be modified to make them safer. The model also separated quinone-forming and non-quinone-forming molecules with 88.2% AUC. Furthermore, the model distinguished molecules containing the same quinone-formation structural alert, and detected the impact of rational drug modification to prevent quinone formation. Molecule-level quinone formation predictions can be used to flag problematic molecules in the early stages of drug development. However, eventually both reactivity and metabolism must be modeled to accurately predict reactive metabolite formation. While we have recently made progress on modeling the reactivity of diverse chemicals [108, 208], previous studies of metabolism have primarily focused on predicting sites of metabolism, rather than actual metabolite structures.[50, 134, 135, 136, 54, 137, 138, 140] This study explicitly predicted the formation of quinones—one of the most common types of reactive metabolites—thereby supplying an essential piece of a unified model of reactivity and metabolism.

Chapter 6

The Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites

6.1 Introduction

Safety problems are one of the primary causes of drug candidate attrition.[5, 6, 7] Furthermore, idiosyncratic adverse drug reactions (IADRs) frequently only arise after approval, incurring significant resources.[9, 12] Many IADRs present as drug-induced liver injury, the leading driver of drug withdrawal from the market.[258, 98, 259, 260, 261] Although IADRs are poorly understood, many are linked to bioactivation: enzymatic conversion of drugs to electrophilically reactive metabolites.[193, 262, 192, 24, 88] Reactive metabolites covalently bind to nucleophilic sites within biological macromolecules, including DNA[263, 132, 264] and off-target proteins.[265, 266, 267] Conjugation to DNA is frequently mutagenic,[110, 268] and conjugation to proteins can disrupt their functions[269, 104] or incite adverse immune responses leading to IADRs.[31, 19, 270, 271, 18]

To minimize toxicity risk, pharmaceutical companies strive to detect reactive metabolites early in the development process.[15, 261, 272, 273] If bioactivation is observed for a certain drug candidate, small structural differences may avoid reactive metabolite formation without significantly affecting its pharmacophore.[190, 9] To detect reactive metabolites, mass spectrometry is frequently used.[204, 109] However, these assays have a significant caveat: they

only measure binding to simple traps such as glutathione or cyanide, with a single nucleophilic site.[105, 106] Binding to these traps may not accurately predict binding to DNA or protein, which are much larger and contain a diverse array of nucleophilic sites.[103] Instead, protein binding can be quantified in experiments with radiolabeled compounds, but these are generally not used until the late stage of preclinical development because of their expense to synthesize.[190, 274] Ideally, computational approaches could facilitate reactive metabolite detection by rapidly flagging drug candidates susceptible to bioactivation, thereby reducing both the total number of needed experiments and the chances of missing potentially toxic molecules.

Currently, the dominant computational aid for avoiding bioactivation is the structural alert approach.[218, 9, 46, 275] Using a data base of structural alerts, which are simply motifs such as phenols or furans that are often bioactivated, molecules containing those substructures can easily be flagged.[276] However, structural alerts have several shortcomings: they are often not actually bioactivated due to specific molecular context, they do not match all drugs that produce reactive metabolites, and they are purely retrospective. Instead, several alternative, more flexible approaches have the potential to learn concepts from the data that generalize to never-before-seen structures. [135, 35, 36, 37, 52, 209, 137, 50, 136, 54, 134, 140, 277, 278, 138, 121] These methods focus on predicting sites of metabolism (SOMs): the specific atom(s) that are metabolically modified by enzymes. A molecule's SOMs can be used to infer the structures of its metabolites or suggest where a molecule might be rationally redesigned.

However, SOM predictors also have limitations. First, by exclusively focusing on metabolism, they omit consideration of reactivity and therefore miss the second half of the bioactivation mechanism. As a result, benign metabolic steps like hydroxylation are indistinguishable from bioactivation events. Secondly, even highly reliable SOM predictions are limited to a single metabolic step, whereas in reality many molecules are subject to sequential metabolism. For example, one systematic review of experimental metabolism literature reported that only 45.6% of all metabolites were formed by a single step, compared to 32.0% and 26.5% of metabolites that were generated by two- or three-or-more-steps, respectively. [152] Intriguingly, both shortcomings

of the SOM paradigm are solvable by the focus of this study: automated metabolite structure generation.

In this work, we built the metabolic forest: collection of metabolism rules that rapidly enumerates metabolite structures for one or several sequential steps (Figure 6.1). To rigorously evaluate its ability to reproduce experimentally-observed structures, we quantified performance across a large, literature derived data set of tens of thousands of reactions. By performing a systematic search between reported substrates and products, the metabolic forest automatically label SOMs (and reveals manual annotation errors), and suggests missing intermediates structures. The metabolic forest predicts many different types of metabolism, including quinone formation. Although their formation is nontrivial to programmatically represent, quinone species are especially important to include in a useful method, because they compose over 40% of known reactive metabolites.[152] Using the quinone and epoxide structures generated by the metabolic forest, we link previously developed metabolism[121, 279] and reactivity[208] models to generate bioactivation hypotheses.

6.2 Results and Discussion

We explored several applications of the metabolic forest. First, we quantified performance at accurately reproducing metabolite structures across a large, literature derived data set of Phase 1 reactions. Second, using these paths, we automatically labeled SOMs, thereby also fixing any mistakes in the original, manual labels. Third, we measured the intermediate metabolites generated when finding paths between substrates and metabolites. Fourth, we quantified how well the metabolite predictor produced quinone structures, a reaction type that is both especially challenging to represent and especially important for anticipating toxicity. Fifth, we hypothesized the toxicity drivers of several drugs with IADRs of unknown etiology.

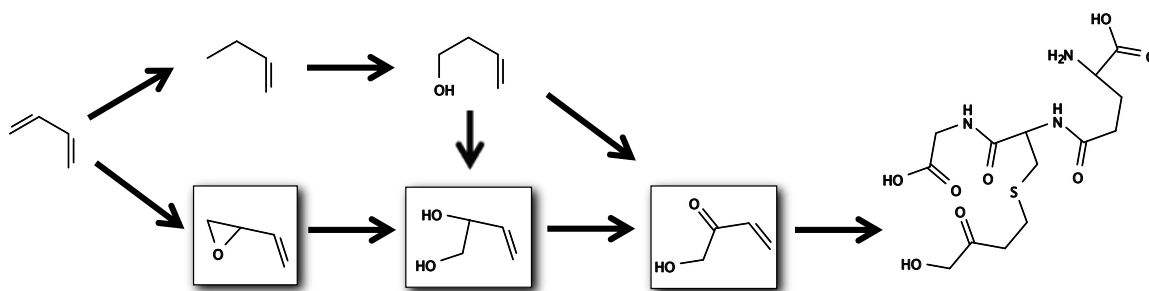


FIGURE 6.1: The metabolic forest explores metabolic pathways by sequentially applying metabolism rules. These pathways can reproduce experimentally-observed structures, or enumerate new structures. Oftentimes, there are several possible pathways between substrates and metabolites. For example, butadiene, a carcinogenic industrial chemical,[88, 280, 281, 282] has several known metabolites.[282] The bottom pathway demonstrates an experimentally elucidated pathway. First, cytochromes P450 and myeloperoxidase epoxidize 1,3-butadiene [283, 284, 285] to form butadiene monoxide. Second, epoxide hydrolase forms 3-butene-1,2-diol.[286] Third, cytochromes P450 form hydroxymethylvinyl ketone, a highly reactive Michael acceptor.[287] Fourth, glutathione (GSH) conjugates to the reactive metabolite.[287] Top, two alternative pathways are shown. The middle row shows an alternative pathway with an equivalent number of metabolic steps, with two different intermediates than the experimentally-known pathway. The top row shows yet another alternative pathway, with several additional intermediates.

6.2.1 Accurate Predictions of Metabolite Structures

The main objective of this study was to develop a system for accurately predicting metabolite structures. With this in mind, a critical metric was the percentage of AMD records for which the metabolite predictor generated the exact experimentally-observed metabolite. For any record that adheres to known metabolism patterns, a useful method should find a series of transformations linking substrate and product.

For each of the 20736 records in the data set, we submitted the substrate and product structures to a breadth-first algorithm that considered all possible combinations of the encoded rules. Each search continued until either a pathway to the exact product structure was found, the user-specified depth-limit was exceeded, or a maximum time limit reached. We performed several experiments to measure the ability of the metabolite predictor to accurately predict metabolite structures (Table 6.1). First, we ran an annotated search, whereby we generated all metabolite structures for all of a molecule's manually-labeled SOMs. This resulted in an overall performance of 78.36%. Second, we ran a depth one search (Algorithm 3). This improved over the known site search, with a top performance of 79.42%. Deepening the search to depths two or three depth increased performance to 88.43% and 88.77%, respectively. There was also a significant improvement between the depth one search and the depth two search. Previous experience with the AMD has suggested that many reactions reported in a single record may in fact be two or more metabolic steps, and this is confirmed by the superior performance of the deeper search.

Due to the combinatorial explosion of metabolite structures, we found it necessary to define a time cutoff for each breadth-first search, at which point the search terminated. Without this limit, the total number of metabolites grows rapidly, this is especially problematic for large molecules. We tried several time cutoffs, including 50, 125, 250, 500, and 750 seconds for each reaction. Unsurprisingly, for searches of depth one, the time cutoff had very little effect on performance. This is as expected, because all depth-one possibilities were explored within the allotted time window. In contrast, searches of greater depth often required a longer run time to find a valid solution (Figure 6.2).

TABLE 6.1: The performance of several methods at reproducing the structure(s) linking AMD record reactants and products.

reaction type	depth three	depth two	depth one	annotated site	RD-Metabolizer
overall	88.77	88.43	79.42	78.36	12.30
C-hydroxylation	91.29	91.07	84.71	84.24	33.50
hydrolysis	86.73	86.41	75.53	75.70	4.44
N-dealkylation	90.09	89.91	81.42	81.17	0.11
reduction	82.71	82.33	70.15	68.59	0.00
aliphatic hydroxylation	92.69	92.45	86.60	86.07	0.29
aromatic hydroxylation	92.11	92.06	86.85	86.55	84.88
O-dealkylation	91.57	91.4	86.15	85.76	9.32
C-oxidation	83.39	82.61	56.69	55.07	13.83
hydrogenation	84.71	84.32	66.87	66.72	0.15
N-demethylation	88.66	88.42	80.11	79.88	0.08
O-demethylation	93.89	93.72	90.58	90.31	0.52
dehydrogenation	86.93	86.04	68.89	66.31	7.47
epoxidation	79.19	78.51	70.54	70.54	31.22
oxidation	77.41	76.24	54.08	48.69	5.69
S-oxidation	85.9	85.32	77.24	73.86	0.00
N-oxidation	88.71	87.62	80.88	80.88	10.82
ring opening	55.25	53.22	19.49	16.10	0.34
dehalogenation	62.6	58.33	43.09	29.88	3.25
dearomatization	65.29	65.08	51.45	46.49	0.62
N-deacylation	92.26	92.26	80.87	80.18	3.87
O-deacylation	97.65	97.65	90.85	90.85	12.68
aromatization	62.34	59.48	40.26	26.23	13.51
oxidative n-dealkylation	91.29	91.29	86.19	85.59	0.00
chain shortening	67.2	67.2	12.54	7.40	13.50
N-reduction	78.35	78.35	69.76	69.42	0.00
oxidative deamination	88.65	87.23	74.47	74.11	1.06
glutathionation	34.53	34.53	27.35	0.00	0.00
optical resolution	94.44	94.44	91.67	91.67	17.59
tautomerization	56.13	52.36	8.02	6.60	0.00

Examples of records for which a path were successfully elucidated are shown in Figure 6.3. For example, for zotepine, it was straightforward to find a one-step sulfur oxidation linking substrate and its metabolite. Also highlighted in Figure 6.3 are examples of records for which a valid path was not found. For example, for a capsaicin metabolite, glutathione attaches to an unsubstituted carbon on a phenyl ring, and a hydroxyl is simultaneously added at a different carbon *meta* to the glutathionation site. Some steps seem to be missing, as the simultaneous glutathionation and *meta* hydroxylation is not a known metabolic pathway. Although we might hope to capture this metabolite with a deeper or faster search by combining metabolism rules, encoding a rule for such an idiosyncratic case would suggest mere memorization of the data, rather than generation of metabolically-logical structures.

We quantified the usage of each rule for the depth three, depth two, and depth one searches (Table 6.2). Additionally, we measured how often any Phase I rule was used (6.3). Unsurprisingly, rules representing very common reaction types, like dealkylation or hydroxylation, were used much more heavily than more specialized rules like azo splitting.

6.2.2 Inferring Intermediate Metabolite Structures

Many AMD records have missing intermediates. Such records are often immediately obvious during manual inspection, for they often entail multiple transformations at different parts of the molecule, such as two hydroxylations. Another frequent variety is sequential transformations at a single site being conflated into a single reaction. For example, some halogenated aromatic rings are reported to immediately form quinones, without reporting the intermediate dehalogenation event that needs to take place before the two-electron oxidation to a quinone. As previously discussed, we implemented a depth-first search algorithm to infer intermediate metabolite structures. To quantify the number of missing metabolites discovered, we subtract the number of metabolites found by the depth one search from the number of metabolites found by the depth two search. This resulted in 83 metabolites, which can be interpreted as missing metabolites in AMD records.

TABLE 6.2: The reaction rules and the number of records where that rule was used in the pathway between the reported substrate and metabolite structures.

rule	depth three	depth two	depth one
Phase 1	19457	19281	16137
dealkylation	5379	5330	4292
hydroxylation	4850	4841	4297
hydrolysis	3771	3734	3136
oxygen reduction	1258	1252	1061
dehydrogenation	908	879	644
hydrogenation	848	829	446
nitrogen oxidation	670	659	595
epoxidation	644	635	522
sulfur oxidation	552	549	500
nitrogen reduction	269	266	210
oxidative dehalogenation	202	191	143
tautomerization	222	206	60
dephosphorylation	135	135	125
quinone formation	152	150	80
dehydration	157	148	54
benzodioxole reduction	89	89	86
reductive dehalogenation	100	96	64
glutathionation	109	92	58
epoxide opening	88	85	1
sulfur reduction	61	61	47
sulfation	52	51	21
azo splitting	27	26	20
acetylation	30	22	0
glucuronidation	5	5	0

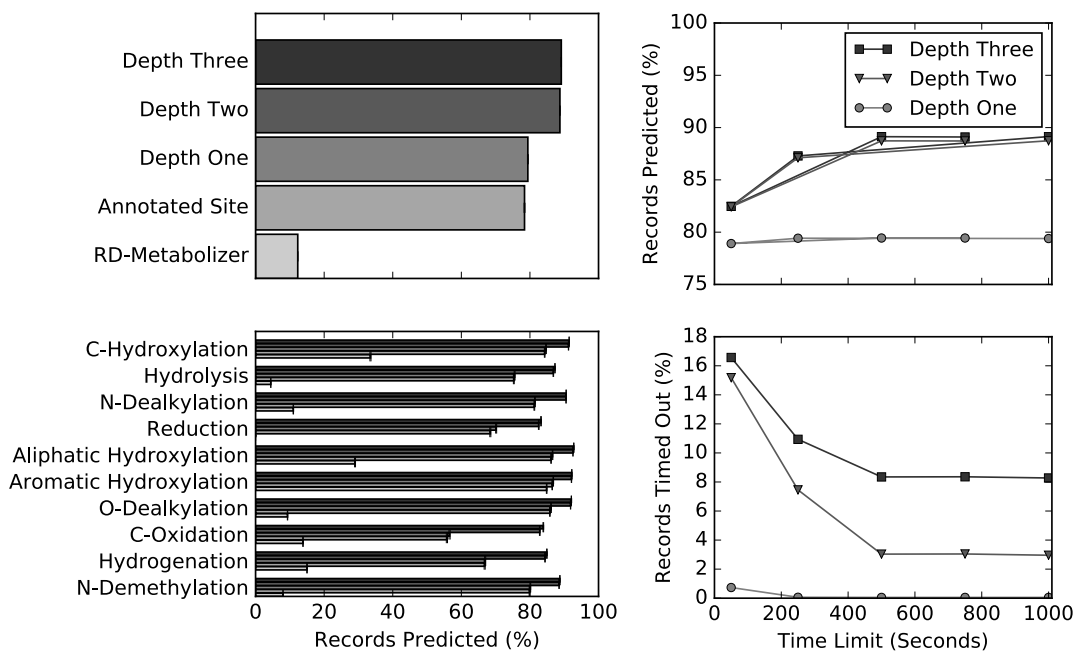


FIGURE 6.2: The metabolic forest took as input the reactant and product structures of AMD records. For each record, the algorithm performed a breadth-first search to link the reactant and product using a collection of metabolic rules. These searches continued until a pathway to the exact product structure is found, the depth-limited was reached, or a maximum time limit was exceeded. Top left, the overall percentage of records for which a valid pathway was found, using a time limit of 1000 seconds. In the Annotated Site comparison experiment a depth one search was restricted to using sites of metabolism matching those that were manually annotated during construction of the data set. Another comparison method was performed using the rules published by RD-Metabolizer.[288] Bottom left, accuracy is broken down for the ten most frequent reaction types of the data set, as recorded in the AMD. Detailed results are reported in Table 6.1. Top right, percent error as a function of the maximum time limit allowed per record. The depth one search requires little running time to exhaust all possibilities. In contrast, performances of the depth two and depth three search are improved by allowing for a longer search. Bottom right, the corresponding number of records that timed out for each depth and time limit.

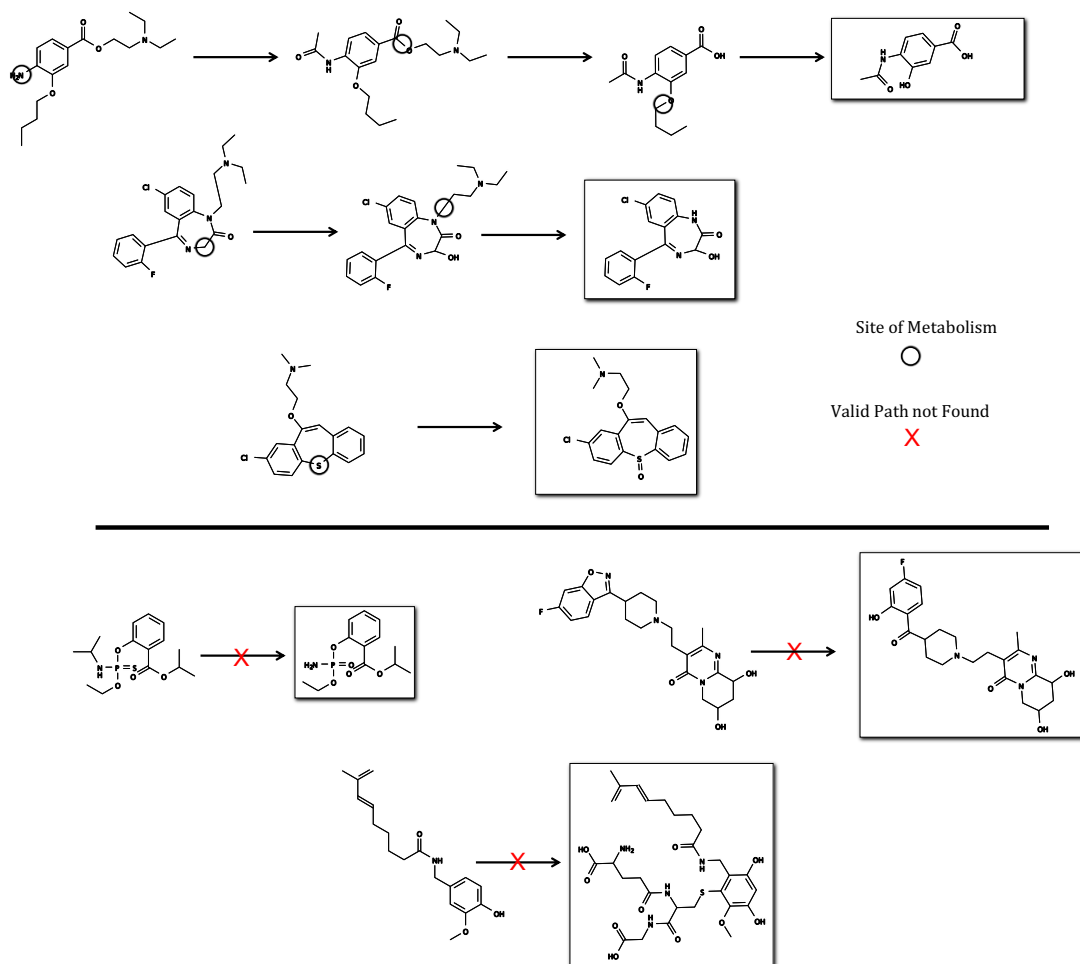


FIGURE 6.3: Examples of successes and failures records. Top panel, example of reaction records from the literature for which a transformation pathway was found by the metabolic forest: top row, a three-step reaction pathway of benoxinate,[289] second row, a two-step reaction pathway of flurazepam,[290] and bottom row, a one-step pathway from zotepine.[291] Lower panel, examples of records for which no transformation pathway was found. Top left, a dealkylation and desulfuration of isofenphos, an insecticide, [292] Bottom left, a ring opening reaction of a metabolite of oxybuprocaine, a local anesthetic.[293] Top row, right, a simultaneous glutathionation and hydroxylation reaction of a metabolite of capsaicin, the primary irritant of chili peppers,[294] Experimentally-observed metabolites are enclosed within boxes. Inferred sites of metabolism are circled. Red crosses indicate records for which no pathway was found to link substrate and metabolite.

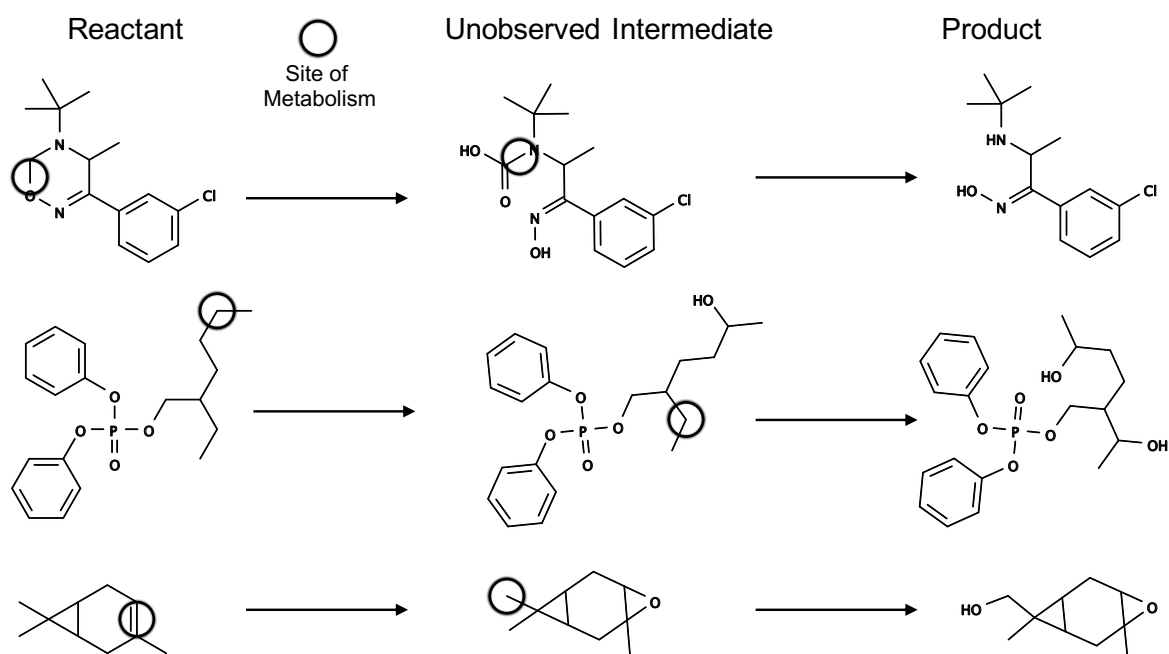


FIGURE 6.4: Missing intermediate structures can be predicted from those records for which a depth two or depth three search found a metabolic pathway not found at depth one. Three examples are visualized. Each substrate-metabolite pair could only be linked by searching a depth greater than one. Such intermediates can be short-lived and consequently difficult to observe experimentally, yet may have important biological consequences. Top row, a candidate prodrug of the antidepressant bupropion.[295] Middle row, the plasticizer additive and flame retardant 2-ethylhexyl diphenyl phosphate.[296] Bottom row, Δ^3 -carene, a natural monoterpene.[297]

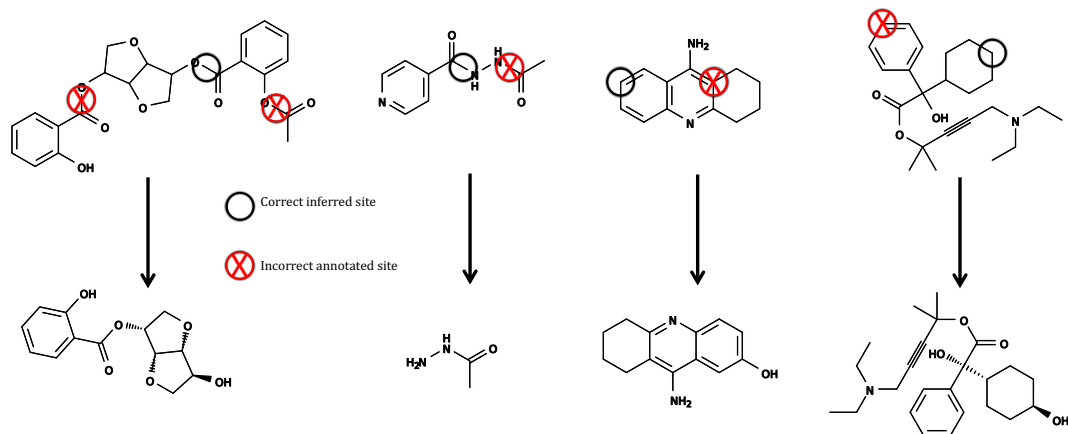


FIGURE 6.5: Across 20736 Phase 1 reactions, the depth one search found a reaction connecting reactant and product in 79.42% of cases, whereas the manually annotated sites only had an accuracy of 78.36%. By examining reactions solved by the depth one search but not the annotated sites, labeling mistakes were revealed. Three examples are shown, where the incorrect, manually annotated sites are circled in red and crossed, and the correct sites uncovered by the depth one search are circled in black. From left to right, acetylisoniazid[298], tacrine [299], and (NS-21), a former drug candidate for bladder disorders[300].

6.2.3 Automatic Detection of Mislabeled Sites.

Comparing the annotated site search to the depth one search revealed cases where SOMs were mislabeled (Figure 6.5). Across 20736 Phase 1 reaction records, the depth one search found a reaction connecting reactant and product in 79.42% of cases,. In contrast, an annotated site search that limited rules to generating structures by modifying at least one of the manually-annotated SOMs only had an accuracy of 78.36%. The comparable accuracy of the 79.42% and 78.36% implies that the metabolic forest produces SOMs that are of similar validity as those produced by manual labor. Furthermore, mistakes in the human annotations are revealed by examining cases where the depth one search found a solution, but the annotated-site-limited search failed.

6.2.4 Accurate Generation of Quinone Structures.

Accurate quinone structure generation is an especially important task for a metabolite structure predictor. Quinone species, including quinone-imines, quinone-methides, and imine-methides, represent over 40% of reactive metabolites.[152] Many drugs are vulnerable to quinone formation due to the ubiquity of the phenyl ring in drug design. Due to their abundance, quinone formation in drug metabolism has been extensively studied experimentally.[257, 301, 254, 302, 197] Recently, we published the first study that explicitly predicted quinone formation.[279] However, that model had the limitation—common to most previous metabolism models—of only making predictions on the input molecule, and not producing actual metabolite structures.

For the present study, we designed a specialized rule for quinone structure prediction. To evaluate performance at predicting quinone structures, we used the exact data set from our previous quinone study.[279] Quinone formation reactions are challenging to encode because they can involve a variety of substituents and by definition entail a loss of aromaticity. Nevertheless, we accurately modeled quinone formation using a combination of SMARTS reactions rules and more fine-coded chemical programming (Algorithm 1). To our knowledge, this is the first published algorithm for predicting the structure of quinones. It is generalizable beyond drug-like molecules to poly aromatic hydrocarbons, which represent a challenge in that they have many molecules with many paths through those molecules. Nevertheless, the algorithm finds quinones all the way across the aromatic system (Figure 6.8). Across the 576 quinone formations reactions, the metabolic forest finds a formation pathway 91.84%, 91.84%, and 76.22% of the time, for depth three, two, and one searches, respectively (Figure 6.6).

6.2.5 Hypothesized Reactive Metabolites of Toxic Drugs.

Building a bioactivation model by linking models of metabolism[121, 208, 308] and reactivity[108, 279] was a primary motivation for constructing an accurate method of metabolite structure generation. The full construction and analysis of this unified bioactivation model is beyond the scope of this study. Nevertheless, to demonstrate the value of the metabolic forest, we built

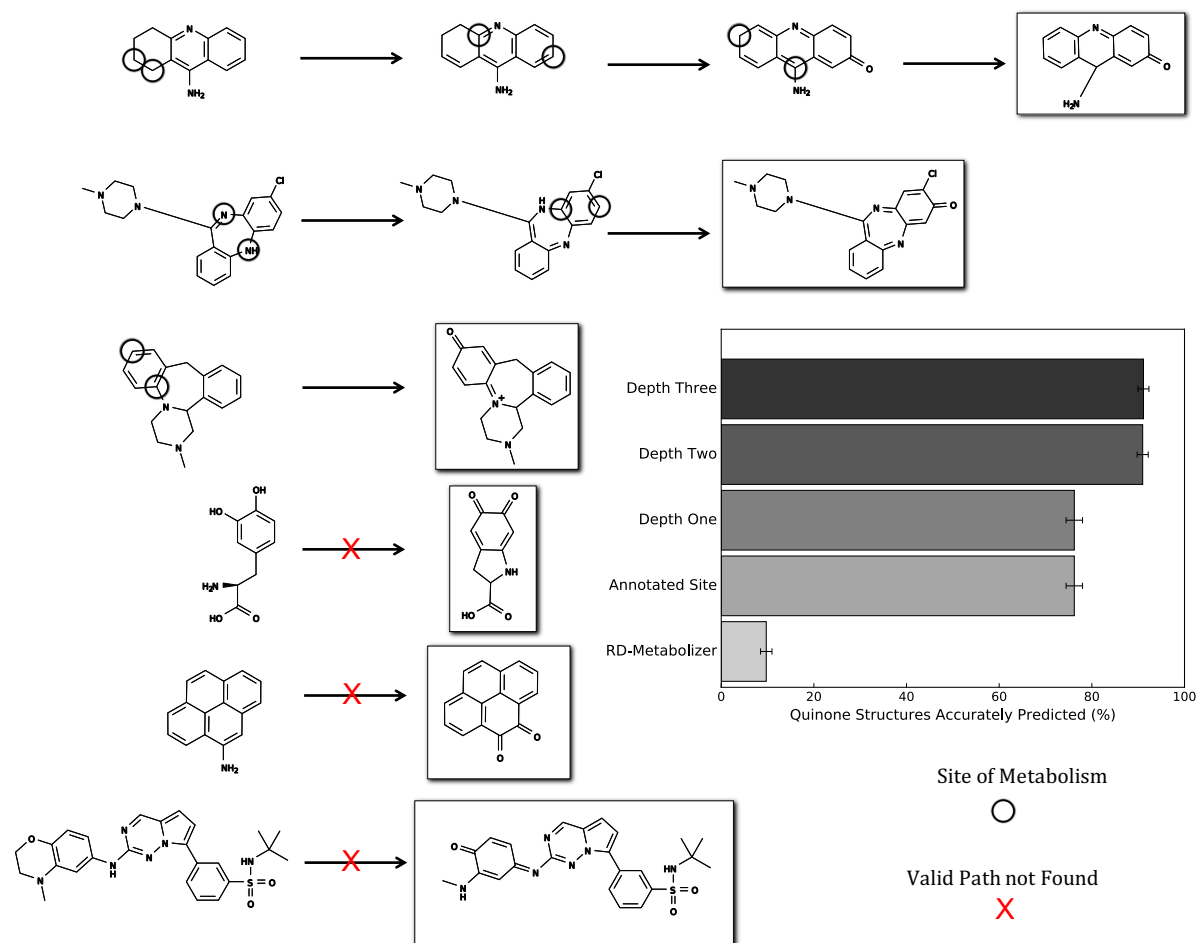


FIGURE 6.6: The metabolic forest accurately produced quinone structures. Across a data set of 576 quinone formations,[279] the metabolic forest found a formation pathway 91.84%, 91.84%, and 76.22% of the time, for depth three, two, and one searches, respectively. Bottom half, examples of experimentally-observed quinones successfully reproduced by depth one, depth two, and depth three searches. From top to bottom, the starting drugs are tacrine, a cholinesterase inhibitor[303, 304] clozapine, an atypical antipsychotic[106] and mianserin, an antidepressant.[305] Third to bottom, dihydroxyphenylalanine.[306] Second to bottom row, a quinone is formed from 4-nitropyrene, an urban air pollutant from diesel engines.[181, 263] Bottom, a kinase inhibitor drug candidate.[307]

a prototype implementation. In this system, we used the explicit metabolite structures generated by the metabolic forest to unite three, heretofore incompatible, models: the previously discussed quinone-formation model[279], a model of epoxidation[121], and a model of reactivity.[208]

The quinone formation and epoxidation models were originally conceived as specialized metabolism studies, with their respective focuses chosen because quinones and epoxides together represent around 50% of all known reactive metabolites.[152, 151] However, classifying quinones and epoxides as either reactive or nonreactive is a generalization, because reactivity is actually a continuum. For example, there are naturally-occurring epoxides[142] and other small-molecule-containing epoxides that are known to be nonreactive[143] In these cases, the presence of electron-donating groups on the carbons in an epoxide often stabilized the motif, by reducing the electron deficiency. Similarly, nearly identical quinones can vary widely in their reactivity and toxicity.[40] Consequently, evaluating the reactivity of possible quinones and epoxides enables a ranking of possible structures that may drive toxicity.

The quinone model predicts how likely quinones are to form at both the site- and molecule-level. Similarly, the epoxidation model yields probabilistic scores for epoxide formation, also at the site- and molecule level. The reactivity model predicts reactivity to biological macromolecules, including DNA and protein, as well as glutathione (GSH) and cyanide, which are frequently used experimentally to detect reactive molecules.[105, 106] To build our bioactivation model, we constructed a pipeline where, for a given input molecule, all the structures of all possible epoxides and quinones were generated, and these structures submitted to the reactivity model. By multiplying the probabilistic metabolism scores from the epoxidation and quinone models by the reactivity scores, metabolites can be identified that are both likely to form and likely to be reactive. In the future, we plan to quantitatively use these bioactivation scores to predict reactive metabolite formation, and to systematically evaluate its performance at this task. Here, we highlight a few case studies where the metabolite predictor has generated hypotheses about reactive metabolites that may be responsible for IADRs (Figure 6.7).

The chemotherapy drug imatinib has been implicated in several cases of idiosyncratic, severe hepatotoxicity.[9, 309]. However, no reactive metabolite has been detected.[310] Our models predict formation of a highly reactive quinone-methide. Our quinone formation model [279] predicted formation of a quinone-imide that is also predicted to be highly reactive by our reactivity model.[208]. The potent non-opioid analgesic metamizole was withdrawn from the market in several countries, including the United States, due to cases of agranulocytosis.[311, 312, 313, 314, 315] The mechanism of this toxicity has not been elucidated, although evidence has been found for an immunoallergic origin.[316] A cation radical has been observed in incubations with myeloperoxidase,[317] but without evidence of its formation *in vivo* the culprit of metamizole remains an open question. Our bioactivation model presents an alternative hypothesis: the formation of a reactive quinone on metamizole's pyrazolone motif. Famotidine, a histamine H_2 receptor antagonist, has been associated with unexplained idiosyncratic reactions, including toxic epidermal necrolysis. [318] A reactive epoxide forming on the thiazole within famotidine is a possible explanation suggested by epoxidation and reactivity scores,

6.2.6 Limitations

The metabolic forest accurately produced the metabolite structures across a large, literature-derived data set of 20736 Phase 1 reaction records. Although diverse, this data may have biases that limit our results. For example, many short-lived intermediates are difficult or impossible to experimentally detect. As a result, such transitory molecules are likely under-reported in metabolic studies. Due to our focus on accurately reproducing literature-derived data, our current tool may not be as well tuned to these underreported intermediates.

Secondly, the search algorithm naively considers all possible combinations of rules, without regard to biological patterns that make some combinations more likely than others. For example, the current method blindly mixes Phase 1 and Phase 2 reactions, ignoring the well-known paradigm of redox reactions often introducing functional groups for subsequent conjugation

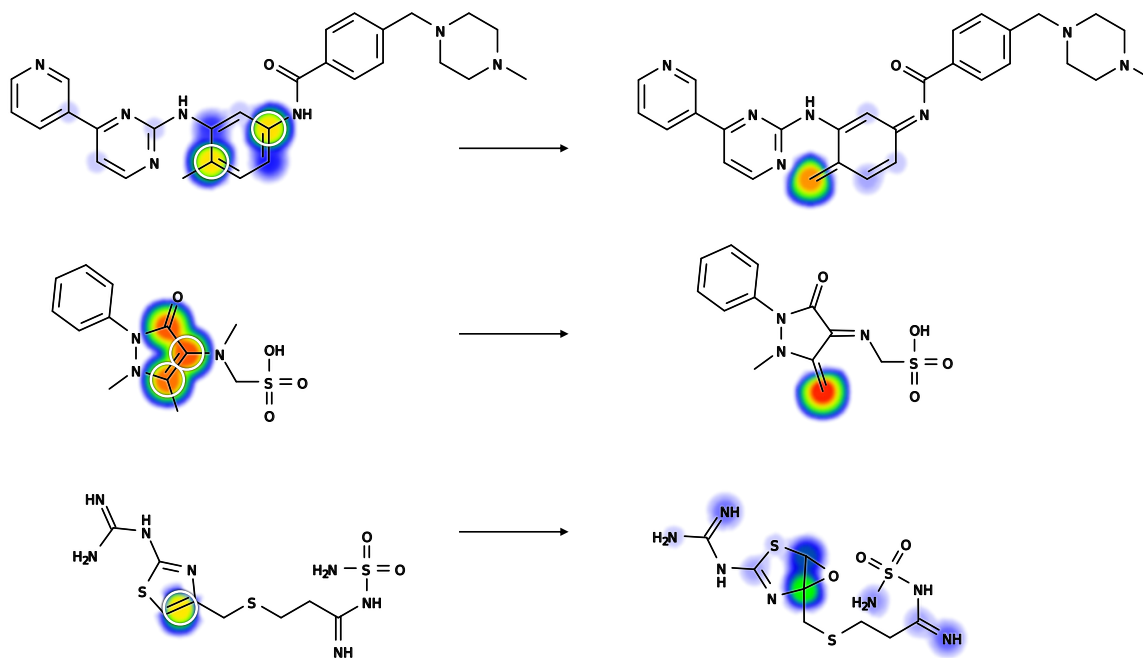


FIGURE 6.7: Previously developed models of metabolism and reactivity are connected by the metabolic forest to generated bioactivation hypotheses for several IADR-associated drugs, whose IADR-drivers were never elucidated. Top, the chemotherapy drug imatinib has been implicated in several cases of idiosyncratic, severe hepatotoxicity.[9, 309]. However, no reactive metabolite has been detected.[310] The potent non-opioid analgesic metamizole was withdrawn from the market in several countries, including the United States, due to cases of agranulocytosis.[311, 312, 313, 314, 315] The mechanism of this toxicity has not been elucidated, although evidence has been found for an immunoallergic origin,[316] suggestive of a bioactivation mechanism. A cation radical has been observed in incubations with myeloperoxidase,[317] but without evidence of its formation *in vivo* the culprit of metamizole remains an open question. Famotidine, an histamine H_2 receptor antagonist, has been associated with idiosyncratic reactions, including toxic epidermal necrolysis. [318] The colored shading represents the quinone formation score[279] for imatinib and metamizole, the epoxidation score[121] for famotidine, and the reactivity scores[208] of all three predicted metabolites.

reactions.[152] Additionally, the search does not take into account the connectivity distance between the sites modified by subsequent metabolic steps. It would be logical to start the search at each depth near the site mostly recently modified, to reduce the number of false paths considered for reactions that are highly correlated, such as hydroxylation followed by quinone formation.[88]

A third possible caveat is the risk of incorrect metabolite structures in the data misinforming the design of the metabolic rules. The potential impact of this possibility on our reported accuracies seems low, because the large data set size of 20736 records made the effect of any single record negligible. Furthermore, when constructing the metabolic forest, only a small fraction of the data was inspected. It seems more likely that incorrect or unorthodox metabolite structures resulted in an accuracy underestimation. For example, one poorly performing reaction type were glutathionation reactions, because glutathione was reported as several slightly different forms in the data and therefore did not match any product structure. Explicitly encoding all these glutathione forms would be undesirable, because it would unnecessarily expand the total number of possible structures at each depth, and place undue emphasis on a single reaction type.

6.3 Conclusion

This study established a validated, accurate tool for predicting metabolic structures across one or several metabolic steps. Our method combined simple rules encoded by reaction SMARTS with novel algorithms for complex, resonance-structure-based transformations, including quinone formation, hydrogenation, dehydrogenation, and tautomerization. We validated the metabolite structure predictor on a diverse collection of 20736 records from a literature-derived database. Beginning with the substrate of each record, a breadth-first search successfully found a transformation resulting in the exact experimentally-observed product 79.42%, 88.43%, and 88.77% of the time when generating a metabolite tree limited to one, two, or three successive rules,

respectively. We also validated our method on a previously published data set of 576 quinone formations,[279] producing the correct quinone structure with accuracies of 76.22%, 91.84%, and 91.84% with searches of depths one, two, and three, respectively. Our tool can also be used to infer missing intermediate structures, and to automatically label sites of metabolism. Most importantly, the metabolite predictor enables integration of metabolism and reactivity models to construct a bioactivation model. Until now, metabolite structures were the missing link in that endeavor. We constructed a prototype model that predicted novel putative reactive metabolites for the drugs imatinib, metamizole, and famotidine, each of which is associated with IADRs, the causes of which have never been elucidated. These reactive metabolites are specific, testable hypotheses about the mechanism of their parent drug's idiosyncratic toxicity. We plan to comprehensively model bioactivation using the metabolite predictor, and expect that accurate enumeration of possible metabolite structures will become a cornerstone of many other future investigations.

6.4 Methods

6.4.1 Phase 1 Metabolism Data

We measured the performance of the metabolite predictor on the Phase 1 metabolism data set, detailed in our simultaneous study.[319] In short, this data consisted of 20736 *in vitro* and *in vivo* Phase I human records from the literature-derived Accelrys Metabolite Database (AMD). We identified five categories of Phase 1 metabolism: stable oxygenation, unstable oxygenation, dehydrogenation, hydrolysis, and reduction, and manually labeled 10280, 5811, 2794, 3869, and 1590 sites of metabolism. Do their complex, often multi-step mechanisms, quinone formations do not fit well into the criteria used to extract and label the Phase 1 data set. Instead, to quantify performance at predicting quinone structures, we used a data set from one of our previous studies consisting of 576 quinone formations extracted from the AMD.[279]

6.4.2 The Metabolic Forest

We built a metabolite structure predictor that rapidly enumerates trees of metabolic pathways. This algorithm—codenamed the metabolic forest—was built in python using the 2017.09.01 release of RDKit, an open-source cheminformatics package.[141] The metabolic forest included 24 reaction rules (Table 6.3). Each rule belonged to a rule set, including 1) each of the five broad classes of Phase 1 metabolism labeled in our simultaneous study, [319] 2) conjugation, 3) quinone formation, and 4) tautomerization.

For Phase 1 metabolism, the stable oxygenation ruleset included the epoxidation, hydroxylation, nitrogen oxidation, and sulfur oxidation rules, the unstable oxygenation ruleset included the dealkylation and oxidation dehalogenation rules, the dehydrogenation ruleset included a lone dehydrogenation rule, the hydrolysis ruleset included the dephosphorylation, epoxide opening, carbonyl cleavage, and azo splitting rules, and the reduction ruleset included the benzodioxole reduction, dehydration, hydrogenation, nitrogen reduction, sulfur reduction, oxygen reduction, and reductive dehalogenation rules. The conjugation ruleset included four rules specifying the reactions acetylation, glucuronidation, glutathionation, and sulfation. The quinone formation ruleset included a single quinone formation rule that both modeled the two-electron oxidation that directly forms quinones and several types of reactions that often set the stage for that oxidation, such as aromatic hydroxylation. Similarly, the tautomerization ruleset had a single eponymous rule. Tautomerization, although not generally regarded as a type of metabolism, nevertheless plays a role in known metabolic pathways of drugs like clopidogrel [320] and ranitidine.[321]

Programmatically, these rules fell into three archetypes: SMARTS rules, resonance pair rules, and resonance structure rules, detailed in the following sections.

6.4.3 SMARTS Rules

Many of the reaction rules used the Reaction SMARTS syntax provided by the open-source cheminformatics library RDKit (Table 6.4).[141] Reaction SMARTS syntax is derived from

TABLE 6.3: The 24 rules used by the metabolic forest, their associated rule set, and their type.

rule	ruleset	type
dehydrogenation	dehydrogenation	
hydrogenation	reduction	resonance pair rule
quinone formation	quinone formation	
epoxidation	stable oxygenation	resonance structure rule
tautomerization	tautomerization	
acetylation	conjugation	
azo splitting	hydrolysis	
benzodioxole reduction	reduction	
dealkylation	unstable oxygenation	
dehydration	reduction	
dephosphorylation	hydrolysis	
epoxide opening	stable oxygenation	
glucuronidation	conjugation	
glutathionation	conjugation	
hydrolysis	carbonyl cleavage	smarts reaction rule
hydroxylation	stable oxygenation	
nitrogen oxidation	stable oxidation	
nitrogen reduction	reduction	
oxidative dehalogenation	unstable oxygenation	
oxygen reduction	reduction	
reductive dehalogenation	reductive dehalogenation	
sulfation	conjugation	
sulfur oxidation	stable oxygenation	
sulfur reduction	reduction	

SMARTS patterns, and has similarities to the alternative reaction languages “SMIRKS”[322, 323] and “SMILES”.[324] Reaction SMARTS are a compact method for encoding transformations that occur on a small number of localized atoms. For example, the string: “[#6h:1]>[*:1]O” expresses a hydroxylation reaction where the oxygen (“O”) is connected with a single bond to a carbon with at least one hydrogen (“[#6h:1]”). The oxygen will be protonated during a sanitization step before outputting the final structure. The exact Reaction SMARTS used for each rule are listed in Table 6.4.

TABLE 6.4: Reaction SMARTS used by the metabolite prediction algorithm.

Rule	Reaction SMARTS
acetylation	[#7,#8,#16;h:1]>[*:1][#6](=[#8])[#6]
azo splitting	[#7:1]=[#7:2]>[*:1].[*:2]
benzodioxole reduction	[#6R:1]-[#8R:2]-[#6H2R:3]-[#8R:4]-[#6R:5]>[*:1]-[*:2].[*:3].[*:4]-[*:5]) [#6H3:1][#7,#8H0,#16:2]>[*:2].[*:1](=O)O [#6H3:1][#7,#8H0,#16:2]>[*:2].[*:1]=O [#6H3:1][#7,#8H0,#16:2]>[*:2].[*:1]-O [#6H2:1][#7,#8H0,#16:2]>[*:2].[*:1](=O)O [#6H2:1][#7,#8H0,#16:2]>[*:2].[*:1]=O [#6H2:1][#7,#8H0,#16:2]>[*:2].[*:1]-O
dealkylation	[#6H1:1][#7,#8H0,#16:2]>[*:2].[*:1]=O [#6H1:1][#7,#8H0,#16:2]>[*:2].[*:1]-O [#6H0:1][#7,#8H0,#16:2]>[*:2].[*:1]-O [#6:1][#6:2]>(O-[*:1].[*:2]) [#6h:1][#6:2]>(O-[*:1].[*:2]) [#6h:1][#6:2]>(O=[*:1].[*:2]) [#8H1:3]-[#6:1]-[#7,#8,#16:2]>[*:3]=[*:1].[*:2]) [#6,#7:1]-[#8H1:2]>[*:1].[*:2]
dehydration	[#6:3]-[#6:1]-[#8H1:2]>[*:3]=[*:1].[*:2] [#6,#7:1]=[#8:2]>[*:1].[*:2]
dehydrogenation	[#16v4:1]-[Oh:2]>[*:1]=[*:2] [#6h:1]-[#6D1H3,#6D2H2,#6D3H1,#7D2H1,#7D1H2,#7D3,#8H1:2]>[*:1]=[*:2]

Chapter 6. The Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites

Rule	SMARTS
dephosphorylation	<chem>[#8:1]=[#15:2]([#8:3])([#8:4])[#8:5][#6:6]>[*:1]=[*:2]([*:3])([*:4])[*:5].O[*:6]</chem>
epoxidation	<chem>[#6:1]=[#6,#7:2]>[*:1]1-[*:2][O]1</chem>
epoxide opening	<chem>[#6:1]1[#8:2][#6:3]1>([*:2][*:3][*:1])</chem> <chem>[#6:1]1[#8:2][#6:3]1>([*:2][*:3][*:1]O)</chem>
glucuronidation	<chem>[#6:1][#6:2](=[O,N,P,S:3])[#8:4]> O1C(C(=O)O)C(O)C(O)C(O)C([*:4][*:2](=[#8:3])[*:1])1</chem> <chem>[#8H1:1][#6:2]>O1C(C(=O)O)C(O)C(O)C(O)C([*:1][*:2])1</chem>
glutathionation	<chem>[#6:1][Cl:2]>C(CC(=O)N[C@@H](CS([*:1][*:3][*:2]))C(=O)NCC(=O)O)[C@@H](C(=O)O)N</chem> <chem>[#16h1:1]>C(CC(=O)N[C@@H](CS([*:1]))C(=O)NCC(=O)O)[C@@H](C(=O)O)N</chem>
hydrogenation	<chem>[#6:1]#[#6:2]>[*:1]=[*:2]</chem> <chem>[#6:1]=[#6:2]>[*:1]-[*:2]</chem>
hydrolysis	<chem>[#8,#16:1]=[#6:2]-[#7,#8,#16:3]>([*:1]=[*:2](O).[*:3])</chem> <chem>[#8,#16:1]=[#6:2]-[#7,#8,#16:3]>([*:1]=[*:2].[*:3])</chem>
hydroxylation	<chem>[#6h:1]>[*:1]O</chem> <chem>[#6h2:1]>[*:1]=O</chem>
nitrogen oxidation	<chem>[#7v3:1]>[*:1]O</chem> <chem>[#7v3:1]>[*:1]=O</chem>
nitrogen reduction	<chem>[#7:1](=[#8:2])-[#8:3]>[*:1].[*:2].[*:3]</chem> <chem>[#7:1]([#8:2]) [#8:3]>[*:1].[*:2].[*:3]</chem> <chem>[#8:2]=[#7:1]-[#8:3]>[*:1].[*:2].[*:3]</chem> <chem>[#8:2]=[#7:1]-[#8-1:3]>[*:1]=[*:2].[*:3]</chem> <chem>[#7D2:1]=[#8:2]>[*:1]</chem> <chem>[#7:1]-[#8:2]>([*:1].[*:2])</chem> <chem>[#9,#17,#35,#53,#85:1]-[#6:2]>[*:1].[*:2]O</chem> <chem>[#9,#17,#35,#53,#85:1]-[#6h1:2]>[*:1].[*:2]=O</chem>
oxidative dehalogenation	<chem>[#9,#17,#35,#53,#85:1]-[#6:2][#6H1:3]>[*:2](O)[*:3]-[*:1]</chem> <chem>[#9,#17,#35,#53,#85:1]-[#6:2]-[#9,#17,#35,#53,#85:3]>[*:1].[*:2](O)=O.[*:3]</chem> <chem>[#9,#17,#35,#53,#85:1]-[#6:2]-[#9,#17,#35,#53,#85:3]>[*:1].[*:2](O)O.[*:3]</chem>
oxygen reduction	<chem>[#8:1]=[#6,#7:2]>[*:1]-[*:2]</chem> <chem>[#8:1]-[#8:2]>[*:1].[*:2]</chem>

Rule	SMARTS
reductive dehalogenation	<chem>[#9,#17,#35,#53,#85:1]-[#6:2]>>[*:1].[*:2]</chem> <chem>[#9,#17,#35,#53,#85:1]-[#6:2]-[#6:3]>>[*:1].[*:2]=[*:3]</chem>
sulfation	<chem>[#6:1][#8:2]>>[*:1][*:2]S(=O)(=O)O</chem> <chem>[#6:1]1=[#6:2][#6:3]2[#8:7][#6:4]2[#6:5]=[#6:6]1>>[*:1]1=[*:2][*:3]=[*:4](-S(C)(=O)(=O))[*:5]=[*:6]</chem>
sulfur oxidation	<chem>[#16;v2,v4:1]>>[*:1]O</chem> <chem>[#16;v2,v4:1]>>[*:1]=O</chem>
sulfur reduction	<chem>[#16:1]=[#8:2]>>[*:1].[*:2]</chem> <chem>[#16:1]-[#16:2]>>[*:1].[*:2]</chem>

6.4.4 Resonance Pair Rules

While Reaction SMARTS work well for encoding transformations that occur on small groups of neighboring atoms, they do not extend well to reactions that can occur at distal sites on a molecule due to resonance structures. For example, quinone formation reactions are challenging to encode because they can involve several atoms across one or more rings, and by definition entail a loss of aromaticity. Consequently, there is no way to write a well-generalized SMARTS rule for quinone formation, because bond and atom attributes shift simultaneously. A SMARTS rule matching a certain quinone formation may not generalize to even slightly different ring patterns, despite identical underlying principles.

Instead, we designed resonance reaction rules that share a common problem: finding a path across the conjugated or aromatic systems within a molecule, while also efficiently iterating through the resonance structures in that molecule. This is more complex than one might initially imagine, because input molecules may have many possible resonance structures, and many possible paths through those resonance structures. Several implementations were ultimately abandoned to combinatorial explosions on certain problematic types of molecules. Ultimately, we found an efficient implementation that enables rapid computation of resonance-based metabolites, even across large aromatic systems like polycyclic aromatic hydrocarbons.

Fundamentally, this procedure considers various resonance structures in turn. However, rather than naively consider all resonance structures for an input molecule, only potentially useful resonance structures are produced. This is possible due to the insight that conjugated-system-based metabolic events can only occur across a single conjugated system, so the various conformations of any other conjugated system are not relevant. To efficiently make resonance structures, a molecule is first fragmented into its constitutive conjugated systems. Next, for each conjugated system fragment, all resonance structures are generated for the fragment, and then reattached to the rest of the molecule. As a result, the total number of resonance structures produced by a molecule is a linear combination of the number of resonance structures in each of its conjugated systems, rather than a multiple. This significantly cuts down on the total number of resonance structures produced, but does not forgo any resonance structure that might be relevant for a given transformation, which can only occur across a single conjugated system.

For the resonance pair rules, pairs of atoms within each resonance structure were considered (Algorithm 1). Only pairs were considered that matched all of the following criteria: 1) both within the same conjugated system, 2) at least one connecting path that alternated double and single bonds and 3) both matching any of several predefined endpoint structures. Each of these endpoint structures had a corresponding list of possible structural modifications. For each valid pair, each possible combination of modifications was performed. For each resulting structure, for each connecting path of alternating single and double bonds between the two atoms, a final structure was generated by swapping the single and double bonds.

The exact specifications of the three resonance pair rules, dehydrogenation, hydrogenation, and quinone formation rules, are detailed in Table 6.5. Dehydrogenation and hydrogenation considered conjugated systems, as previously described. However, due to its mechanism, the quinone formation rule only needed to consider aromatic systems (Figure 6.8).

Algorithm 1 The resonance pair algorithm. Algorithm takes inputs (1) the molecule, (2) end-point patterns as SMARTS strings, and (3) sets of atoms that are grouped by connected system, aromatic or conjugate as needed determined for the rule.

```

1: procedure RESONANCEPAIRALGORITHM(Molecule,Transforms,Systems)
2:    $T \leftarrow$  all matches in Molecule to Transforms
3:   for  $S$  in Systems do
4:     for KekuleForm ( $K$ ) in  $S$  do
5:       for PairedMatches ( $A_1, A_2$ ) in both  $S$  and  $T$  do
6:          $P \leftarrow$  path of alternating single/double bonds between  $A_1$  and  $A_2$ , with start and end double bond
7:          $M \leftarrow$  transform  $S$  in Molecule to match  $K$ 
8:          $M \leftarrow$  apply transform from  $A_1$  to  $M$ 
9:          $M \leftarrow$  apply transform from  $A_2$  to  $M$ 
10:         $M \leftarrow$  swap single and double bonds along  $P$  in  $M$ 
11:        yield  $M$ 

```

6.4.5 Resonance Structure Rules

The tautomerization and epoxidation rules also used resonance structures, but are distinct from the resonance pair rules because they do not perform combinations of modifications on pairs of atoms. Instead, the tautomerization rule does not change the overall number of double bonds, single bonds, or hydrogens (Algorithm 2). The rule works by iterating through resonance structures and enumerating all atoms one-bond away from each conjugated system. Next, for each of these neighboring atoms, all paths were considered between the neighbor and each atom within the system, where the number of double and single bonds was the same. For each path, all double and single bonds were swapped to emit a tautomer.

The epoxidation rule used resonance structures and a pattern listed in Table 6.4 to replace double bonds between carbons with epoxides. By first converting all aromatic bonds in the input molecule to single or double bonds (known as kekulization), and then considering all double bonds within each resonance structures, all possible epoxides are generated. Instead, only using a single kekulized form of the input molecule could miss some epoxides for those sites assigned to be single rather than double bonds.

TABLE 6.5: The transforms used by the resonance pair rules.

Rule	Endpoint (Transforms)	System Type
dehydrogenation	[#6h:1][#6D1H3,#6D2H2,#6D3H1,#7D2H1,#7D1H2,#8H:2] ('single2double')	conjugated
hydrogenation	[#6h:1][#7D3:2] ('single2double', 'addPlus1')	conjugated
	[*:1]	
	[#6R:1][#6D1H3,#6D2H2,#6D3H1,#7D2H1,#7D1H2,#8H:2] ('single2double')	
	[#6D2H1:1] ('addO')	
quinone formation	[#6HOR:1]-[F,Cl,Br,I:2] ('replaceHalogenWithO', 'single2double')	aromatic
	[#6HOR:1][#7D3:2] ('single2double', 'addPlus1')	
	[#6R:1][#7,#8:2][#6:3] ('dealk', 'single2double')	

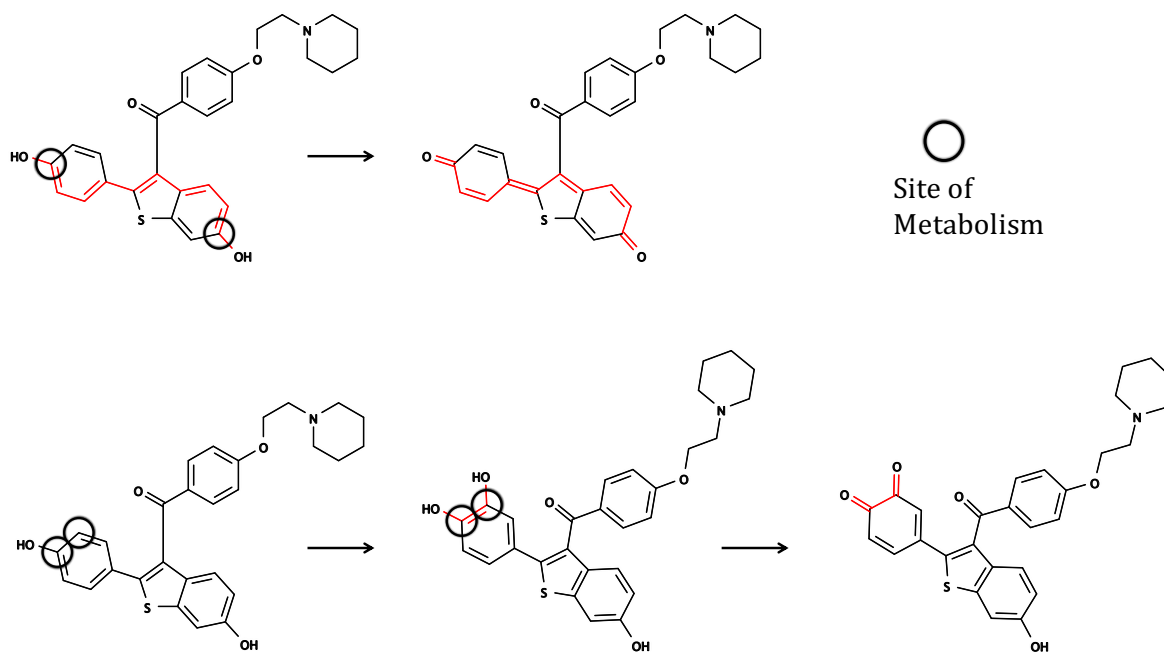


FIGURE 6.8: Raloxifene illustrates the quinone prediction method. Various resonance structures are considered in turn. For each structure, all paths between each pair of atoms are examined. Each atom pair is tested for compatibility with several predefined endpoint structures. A valid path is indicated by red bonds on the upper left. For this path, both endpoints are hydroxyls, in purple, which form a long range quinone in a one step, two-electron oxidation when the single and double bonds swap as indicated by the arrows. Two-step quinone formations are also predicted by our method, such as that formed by hydroxylation followed by oxidation to form the lower right structure.

Algorithm 2 The tautomerization algorithm.

```
1: procedure TAUTOMERIZATIONRULE(Molecule)
2:   for  $S$  in ConjugatedSystems do
3:     for KekuleForm ( $K$ ) in  $S$  do
4:       for Neighbor ( $N$ ) in Molecule not in  $S$  do
5:         for Atom ( $A$ ) in  $S$  do
6:            $P \leftarrow$  path of alternating single/double bonds between  $A$  and  $N$ , with equal number of
           single and double bonds
7:            $M \leftarrow$  swap single and double bonds along  $P$  in  $M$ 
8:           yield  $M$ 
```

6.4.6 Algorithm for Predicting Depth Two or Three Metabolite Structures

To search for multiple-step metabolites, a breadth first search was performed, with depth capped at two or three. In this algorithm, the search continued until a metabolic path was found between substrate and metabolite. For example, for a depth two search, each depth one metabolite was first consider in turn. If any of the depth one metabolites matched the experimentally known metabolite, the search terminated. Otherwise, all the depth one metabolites were considered, one at a time. For each depth one metabolite, all its metabolites were generated (equivalent to depth two metabolites relative to the starting reactant). If any of those depth two metabolites matched the experimentally known product, the search terminated. This process continued until either a match was found or the specified depth limit was exceeded. This search is only limited by the size of the input molecule, the quality of the rule sets, and the computational power available. In this study, we at most ran a depth three search to find paths between substrates and metabolites, by deeper searches are readily possible. This algorithm can also be modified to only take as input a substrate structure, and generate all possible metabolites to a given depth.

6.4.7 Comparison to RD-Metabolizer

A recent study included metabolite structure generation.[325] This work used Reaction SMARTS, and in total 20 are included in Tables 1 and 3 in the the study. To construct a comparison rule set, we transcribed all of these patterns and evaluated them using the same criteria as our method: ability to exactly reproduce metabolite structures. Although these were all the patterns that

Algorithm 3 The breadth-first search algorithm. Algorithm takes inputs (1) the substrate, (2) the metabolite, (3) a ruleset(a collection of metabolic rules) and (4) a depth limit.

```
1: procedure BFSMETABOLITE(Substrate, Metabolite, RuleSet, Depth)
2:   Intermediates(I) = [] ▷ Empty list
3:   for Rule(R) in RuleSet do
4:     for Transforms in R do
5:       T ← all matches in Substrate to Transforms
6:       for Site(S) in T do
7:         M ← apply T at S on Substrate
8:         if M == Metabolite then return M
9:         else
10:          I ← append M to I
11:   if Depth >= 1 then
12:     for M in I do
13:       BFSMetabolite(M, Metabolite, RuleSet, Depth - 1)
```

seem to be available from this work, they are labeled as examples and therefore our evaluation of this ruleset might not reflect its true ability. As previously described, Reaction SMARTS are also part of the metabolic forest, and we have comprehensively listed these in [Table 6.4](#).

Chapter 7

Modeling the Bioactivation and Subsequent Reactivity of Drugs

7.1 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%^[326] Extrapolating from those results, the FDA estimated that annual ADRs in the U.S. cause over 2,216,000 hospitalizations and more than 106,00 deaths.^[327] Furthermore, predicated on the meta-analysis's accuracy, the FDA proposed that ADRs are the 4th leading cause of death in the U.S., exceeding automobile deaths, diabetes, AIDS, pulmonary disease, and pneumonia.^[327] Some ADRs are traceable to the pharmacological effects of certain drugs, and this mechanistic understanding can inform efforts to reduce risk.^[11] However, a subset of ADRs—idiosyncratic adverse drug reactions (IADRs)—have 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.^{[14, 328, 329, 330, 105, 331, 317, 241,}

246, 243, 247] 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 10000 to 1 in 100000 patients.[13] 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 3000 patients.[14] After approval and market release, however, exposure to much large patient populations can reveal a drug's hidden risk. Indeed, already-approved drugs are mostly commonly withdrawn from the market due to intolerable numbers of IADR cases.[258, 98, 259, 260, 261] Even if a drug is not withdrawn, it may be labeled with a "black-box" warning by the FDA, significantly curtailing its profitability.[332]

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 unlucky patients. However, a growing body of evidence suggests that many IADRs are induced by a specific mechanism: bioactivation, the focus of this study (Figure 7.1).

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, 28, 29], and metabolite-protein adducts can disrupt protein function or trigger toxic immune responses. In fact, many IADRs have been linked to overzealous autoimmune attacks set off by the production and covalently 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. These assays may not accurately reflect endogenous metabolism, consume time and resources, and require physical synthesis of each compound under consideration.

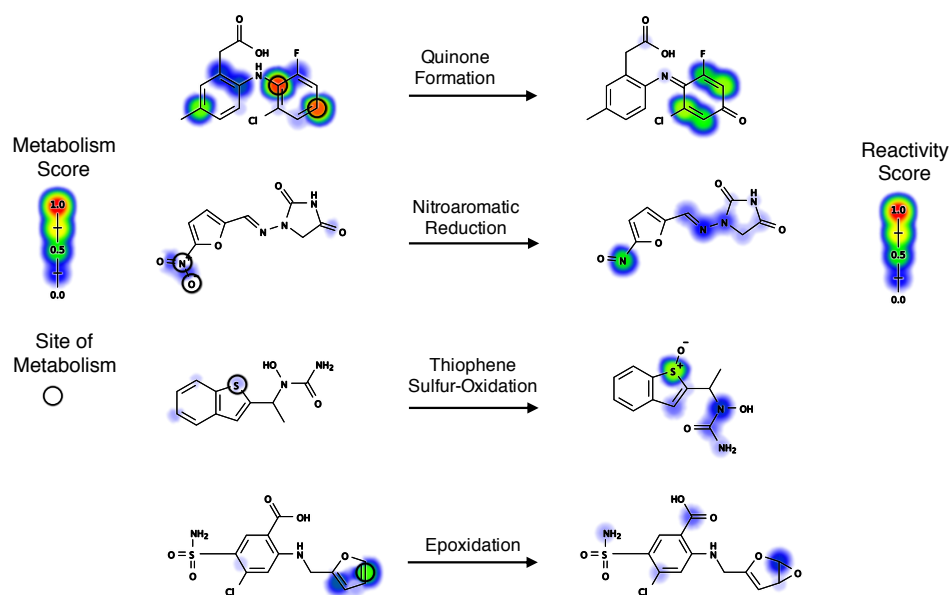


FIGURE 7.1: This study modeled four common pathways of bioactivation: quinone formation, nitroaromatic reduction, thiophene sulfur-oxidation, and epoxidation. Top row, lumiracoxib, a cyclooxygenase-2 selective inhibitor, was withdrawn from several countries after several cases of severe liver damage.[202, 203] This toxicity was traced to the formation of reactive quinone-imine metabolite that conjugates to off-target proteins, inducing deleterious immune responses.[202, 203] Second row, nitrofurantoin, an antibiotic, carries a risk of acute liver failure,[333], which is thought to be caused by reduction of nitrofurantoin's nitroaromatic group to a reactive nitroso.[334] 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 S-oxide.[335] Bottom row, furosemide, a diuretic, confers a risk of idiosyncratic hepatitis due to production of a reactive epoxide metabolite.[9, 336, 337, 338, 339]

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. 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 7.1). These pathways were chosen because we have well-developed models for these metabolic routes, including an accurate model of quinone formation[279], epoxidation[121], and a phase I metabolism model that includes predictions for nitroaromatic reduction and thiophene sulfur-oxidation.[319]

7.2 Methods

7.2.1 Bioactivation Training Data

We assembled a heterogeneous data set of bioactivation reactions from the literature-derived Accelrys Metabolite Database. Each reaction took place in humans, human cells, or human liver microsomes. Four types of bioactivation reactions were extracted: quinone formation, epoxidation, nitroaromatic reduction, and thiophene S-oxidation. Each of these pathways are well-known bioactivation mechanisms, and in previous work we have built models that predict whether molecules will subject to each type of metabolism. Quinone formation[279] and epoxidation[121] were modeled independently, and nitroaromatic reduction and thiophene S-oxidation were included in a model of diverse Phase 1 reactions.[319] In total, we extracted from the experimental data 210 quinone formations, 174 epoxidations, 4 nitroaromatic reductions, and 10 thiophene S-oxidations. For each molecule, we enumerated all possible metabolites for each of these 4 pathways, producing 6594 quinone formations, 6077 epoxidations, 48 nitroaromatic reductions, and 10 thiophene S-oxidations. We then merged all duplicate molecules into a single representation per molecule, with all of its experimentally-known bioactivation pathways

labeled. Finally, we filtered out any molecules that did not have at least one experimentally-not-observed bioactivation pathway, because such cases would not be a good test of the model's ability to predict the correct pathway within bioactivated molecules. This procedure produced a total of 340 bioactivated training molecules.

We also selected molecules that are not known to be bioactivated. To select these molecules, we started with the same set of reactions from the Accelrys Metabolite Database, where each reaction took place in humans, human cells, or human liver microsomes. We then filtered out all molecules that were intrinsically reactive: those that are known to directly conjugate to protein or glutathione. We also filtered out all bioactivated molecules. Finally, we selected all molecules that had at least one possible bioactivation pathway among the 4 mechanisms considered in this study. This left us with a large pool of negative molecules. From this pool, we randomly selected an equivalent number as the total number of bioactivated molecules, forming a final data set of 680 molecules.

7.2.2 External Non-Bioactivated Test Sets

A total of 12882 non-bioactivated molecules remained after the training molecules were removed. From these, we extracted 38 negative external test sets by randomly shuffling the list of negatives and splitting it into batches of 340 molecules. We chose this batch size to equal the number of negatives included in the training data. None of these molecules were considered during training, and only tested by the final model.

7.2.3 DrugBank

To provide an external data for investigating the model's predictions, the withdrawn drug list from the July 3rd, 2018 release of DrugBank was downloaded. This file included the structures of 221 drugs. We removed any drug that was also present in our training data. Afterwards, 208 molecules remained in our training data.

TABLE 7.1: Bioactivation descriptors used to predict bioactivation pathways.

Descriptor
Formation Score
GSH Atom Reactivity Delta
GSH Bioactivation Score
Protein Atom Reactivity Delta
Protein Bioactivation Score

7.2.4 Bioactivation Descriptors

For this study, we synthesized several previous models of metabolism and reactivity to design specific bioactivation descriptors that were inputted to a neural network (Figure 7.2). Using previously designed models for quinone formation, epoxidation, nitroaromatic reduction, and thiophene S-oxidation, formation scores were computed for each possible transformation for a given input molecule. Next, the actual structures of each of these possible metabolites was generated using an in-house metabolite structure predictor.[340] After enumeration of the metabolite structures, the atom-level reactivity GSH- and protein-reactivity scores were computed for both the substrate and the product molecule. Next, the atom-level reactivity deltas between the metabolite and substrate were 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 was computed by multiplying each reactivity delta by the corresponding formation score.

We also computed simple molecule descriptors, such as molecule weight or the total number of atoms. Overall, the descriptor generation produced 20 descriptors for each possible metabolite, including 5 bioactivation pathway descriptors (Table 7.1) and 15 molecule descriptors (Table 7.2).

7.2.5 Combined Path- and Molecule-level Bioactivation Model

The 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.

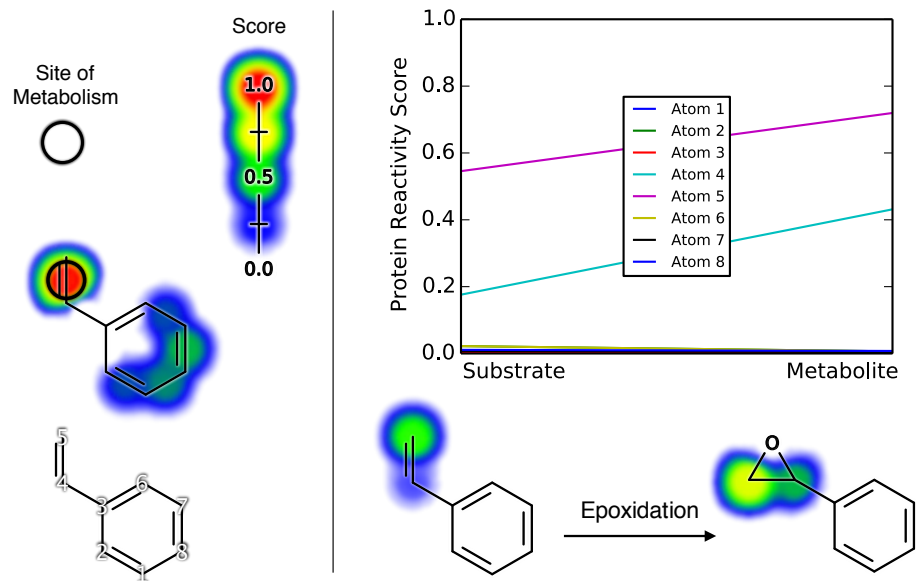


FIGURE 7.2: Bioactivation descriptors were computed using site-level metabolism predictions, metabolite structure predictions, and reactivity predictions. Left, the predictions generated by a previously developed epoxidation model[121] are visualized on styrene. The colored shading indicates site-level epoxidation scores, which reflect the probability that an epoxide will form at each possible location within styrene. Bottom right, using a previously developed metabolite structure generator,[340] the exact epoxide structure corresponding to the highest site-level epoxidation score was generated. Next, we applied our previously published reactivity model [208], which has prove useful in other studies as well [341, 342], 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. As shown in the chart, atoms 4 and 5 are both predicted to be more reactive in the substrate than the metabolite, while the rest of the atoms are unchanged.

TABLE 7.2: Molecule descriptors used to predict bioactivation pathways

Descriptor
Molar Refractivity
Molecular Weight
Number of Aromatic Bonds
Number of Bonds
Number of Double Bonds
Number of Heavy Atoms
Number of Hydrogen Bond Acceptors Definition 1
Number of Hydrogen Bond Acceptors Definition 2
Number of Hydrogen Bond Donors
Number of Hydrogens
Number of Rings
Number of Single Bonds
Number of Triple Bonds
Octanol/Water Partition Coefficient
Topological Polar Surface Area

In this, the descriptors compose our features, and our targets are a binary column indicating whether a pathway was experimentally observed.

The bioactivation model is a convolutional neural network, with one molecule layer, one input layer, one hidden layer, and two output layers (Figure 7.3). The first output layer calculates path bioactivation scores (PBS), and the second output layer computes a single molecule bioactivation score (MBS) for each input molecule. Each PBS reflects the probability of a specific bioactivation event at a specific sites within a molecule. Using the probabilistic OR function, the MBS is computed by calculating the probability of bioactivation at any site by any of the pathways considered in this study: quinone formation, epoxidation, nitroaromatic reduction, and thiophene S-oxidation.

Our procedure can generate a several different PBS, which may share some atoms in common. In order to aid interpretability, we devised a method for visualizing PBS for a given input molecule (Figure 7.4).

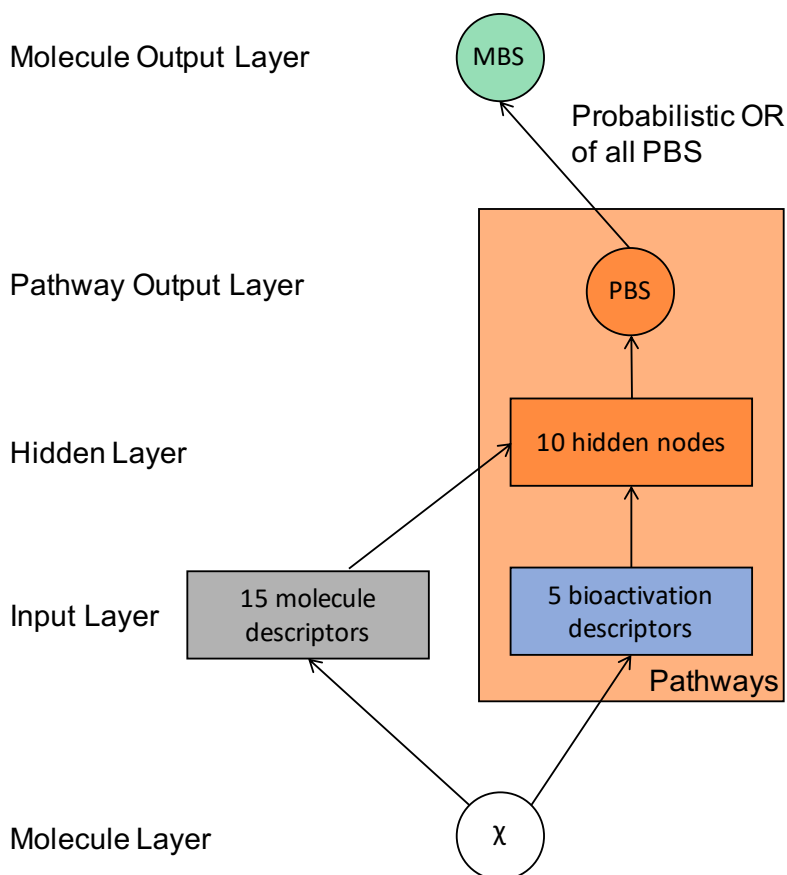


FIGURE 7.3: The diagram demonstrates the flow of data through the model. The model consists of one molecule layer, one input later, one hidden layers, and two output layers. First, several descriptors are calculated from an input molecule's structure. These descriptors are submitted to the hidden layer, which computes pathway bioactivation scores (PBS) for the input molecule. Each PBS 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 bioactivations, including all possible quinone formations, epoxidations, nitroaromatic reductions, and thiophene S-oxidations. Next, a probabilistic OR function computes a molecule bioactivation score (MBS), which also ranges from from zero to one and represents the probability of a molecule undergoing bioactivation by any of the pathways considered. 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.

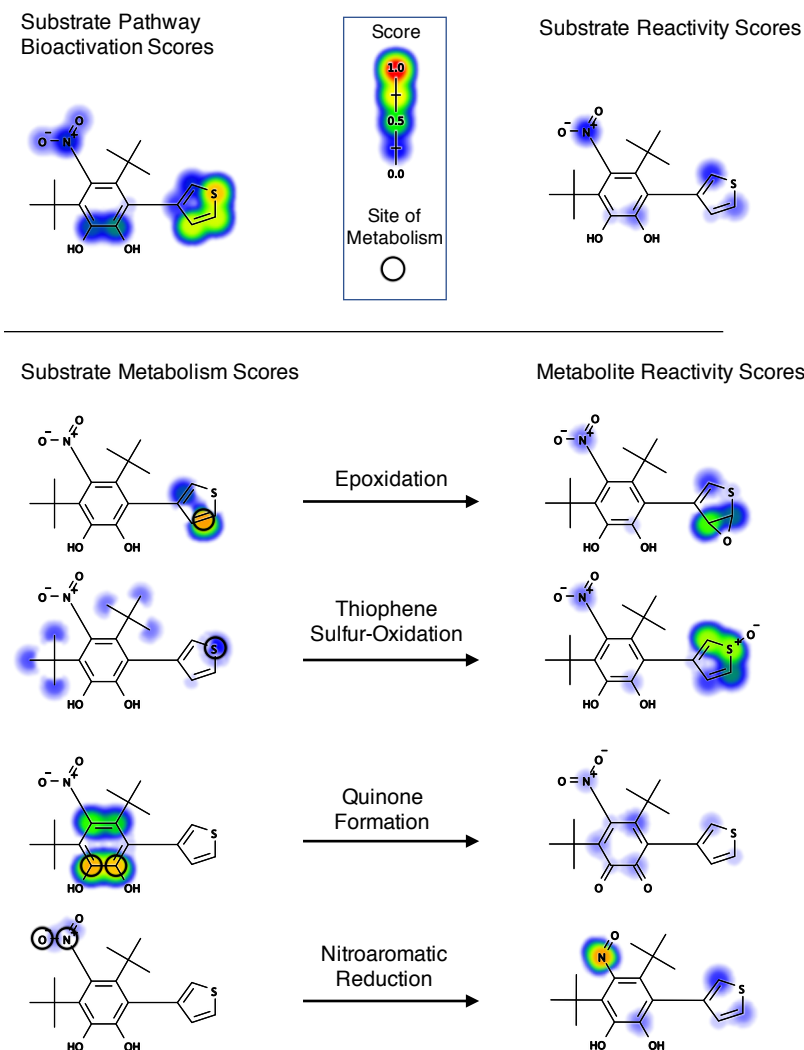


FIGURE 7.4: Epoxidations and nitroaromatic reductions entail bond cleavage. However, quinones take place on atom pairs, and thiophene sulfur-oxidation occurs on a atom, the sulfur. Consequently, while each of these predictions types can be visualized on separate structures (Figure 7.1), this paradigm does not make for easily-understandable predictions. Instead, we first mapped all PBS to the atom level. This is straightforward: for a quinone pair prediction, each atom 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 moved closer to an interpretable result, in the same, the same atom can be subject to multiple possible bioactivation pathways. For example, the atoms making up sites of aromatic epoxidation are often themselves possible sites of quinone formation. Therefore, in the final step prior to visualization we computed the scores on each atom by using the probabilistic OR function across all predictions that included the current atom under consideration.

7.3 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 (PBS). Using several metrics, we quantify how well PBS 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 (MBS). Using similar methodologies as the pathway-level analysis, we measure the performance of MBS 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.

7.3.1 Bioactivation Pathway Prediction Accuracy

For bioactivated molecules, knowledge of their specific bioactivation pathway (s) yields potentially fundamental insights about their possible toxicity mechanism (s). 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.

To train the model, we use gradient descent on the cross-entropy error. In this procedure, the model's weights are gradually adjusted to assign high PBS for experimentally known bioactivation pathways, and low PBS for all other pathways. We used cross-validation to produce PBS for the entire data set, a standard practice in machine learning for simulating performance on external data. In this procedure, any metabolically-related molecules are withheld together, and the model trained on the remaining data. Next, the trained model predicts the PBS of the withheld molecule (s). In total, there are groups of related molecules, so the cross-validation procedure entails training individual models. This process insures that each molecule's predictions are computed by a model that does not contain information about that molecule. We use several metrics to assess the model's performance at predicting bioactivation pathways.

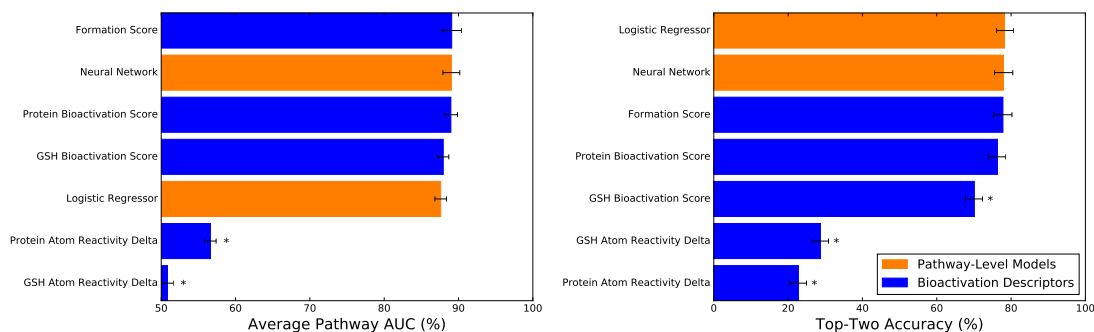


FIGURE 7.5: The model produced accurate cross-validated bioactivation pathway predictions. Right, the top-two metric was computed across 340 bioactivated molecules. The top-two metric reflects the percentage of correctly predicted molecules, where a correct prediction is defined as molecule for which any of its bioactivated pathways received the highest or second-highest score for all possible pathways within that molecule. Left, across the same 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 statistically significant worse than the highest-scoring method, using a paired *t*-test.[175]

First, we compute the “top-two” performance, a commonly used metric in site of metabolism prediction studies (Figurefig:pathtoptwoandaveragepathauc). [121, 50, 122, 123, 52] This metric counts a molecule as accurately predicted only if any of its bioactivated pathways receive the highest or second-highest PBS for the entire molecule. The total number of correct predictions are divided by the total number of bioactivated molecules and multiplied by 100 to produce the percentage of correct predictions. Second, we computed “average path AUC” for calculating the area under the receiver operating characteristic curve (AUC) for each bioactivated molecule, followed by averaging these AUCs. We have used this metric in several precious studies because it measures how accurately predictions are ranked on intra-molecule basis, which seems important for a hypothetical user trying to interpret the results.[279, 108, 121, 208]

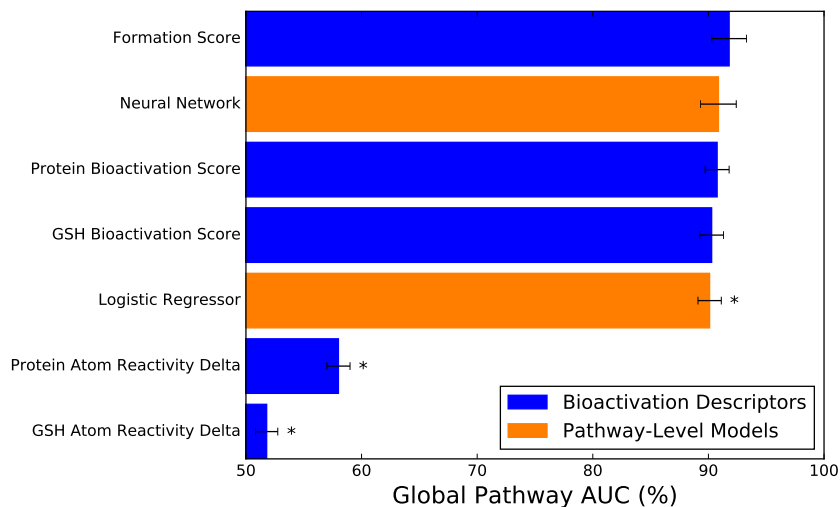
For the purpose of choosing the best model, the results of both the top-two and average site AUC metrics were inconclusive. By the top-two metric,the logistic regressor had a performance of 78.24%, equivalent to the neural network’s performance of 77.94%. The two models also produced equivalent average pathway AUCs: 88.97%and 87.60%for the neural network and

logistic regressor, respectively.

Intriguingly, for both metrics, three of the bioactivation scores produced equivalent performances. These performances are calculated simply by treating a descriptor as a model prediction, without any training. In previous studies, we frequently compare our results to individual descriptors, and we have always found that machine learning outperforms the naive descriptor approach. In the past, this has seemed logical, because machine learning algorithms can consider many different chemical attributes in concert, and learn complex functions based on these features. This is the first study in which any descriptor had such a strong performance compared to the modeling approach, let alone three descriptors.

It is noteworthy that the two descriptors that have much less predictive value are the two reactivity deltas. This suggests that merely calculating the reactivity of a potential product is not very informative for predicting bioactivation, because that metabolite may be very unlikely to occur. Indeed, 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. From these results, it is not clear whether the reactivity component is informative to the model, because these scores do not outperform the much simpler formation score.

To attempt to differentiate the two models and three the highly predictive bioactivation scores, we computed a third metric, the global pathway AUC. Unlike the previous metrics, this measure does not consider molecule identify, and merely computed an AUC for all pathways within positives for the data set. Despite this distinction, the global pathway AUC showed a very similar pattern as the previous two metrics. The neural network had score had a global pathway AUC of 90.87%, which as measured by a false positive rate paired t-test[175] was statistically equivalent to the performances of the same three descriptors formation score, protein bioactivation score, and gsh bioactivation score. The only difference between these results and the previous metrics is that the logistic regressor had a performance of 90.1, which was



caption[The global path AUC was computed for several methods across all bioactivated molecules] The global path AUC was computed for several methods across all bioactivated molecules. Asterisks denote performances that were statistically significant worse than the highest-scoring method, using a false-positive-rate paired *t*-test.[175]

low enough to result in a borderline significant p-value of 0.00364 when compared to the best performing method : the raw formation score descriptor value.

We finally observed some distinction by computing reliability plots for each of the top performing descriptors and models (Figure 7.6). In this measure, for each method, scores are assigned to 1 of 10 bins, with each bin representing a score of width 0.1 We then computed the percentage of bioactivations in each bin, after normalizing the frequencies bioactivated and non-bioactivated examples, and compute the root mean square error (RMSE) between each bin percentage and the bidpoint percentage value of that bin. The normalized bin percentages are plotted on the right Y-axis of each plot. A perfectly scaled prediction would have an RMSE of 0 by this procedure. To assess performance, we also compute R^2 pf the best fit line, which should be 1 for a model that produces scores that bin into a perfectly straight line.

With an RMSE of only 5.206% and an R^2) of 0.964, the neural network outperformed all other methods by both counts. The formation score also has a high R^2) of 0.958, but its slope does not match a scaled probability as well as the neural network, and it only has an RMSE of 18.595% compared the diagonal identify line representing a perfect probability. Based on this result, we chose the use the neural network going forward, but do not claim that this

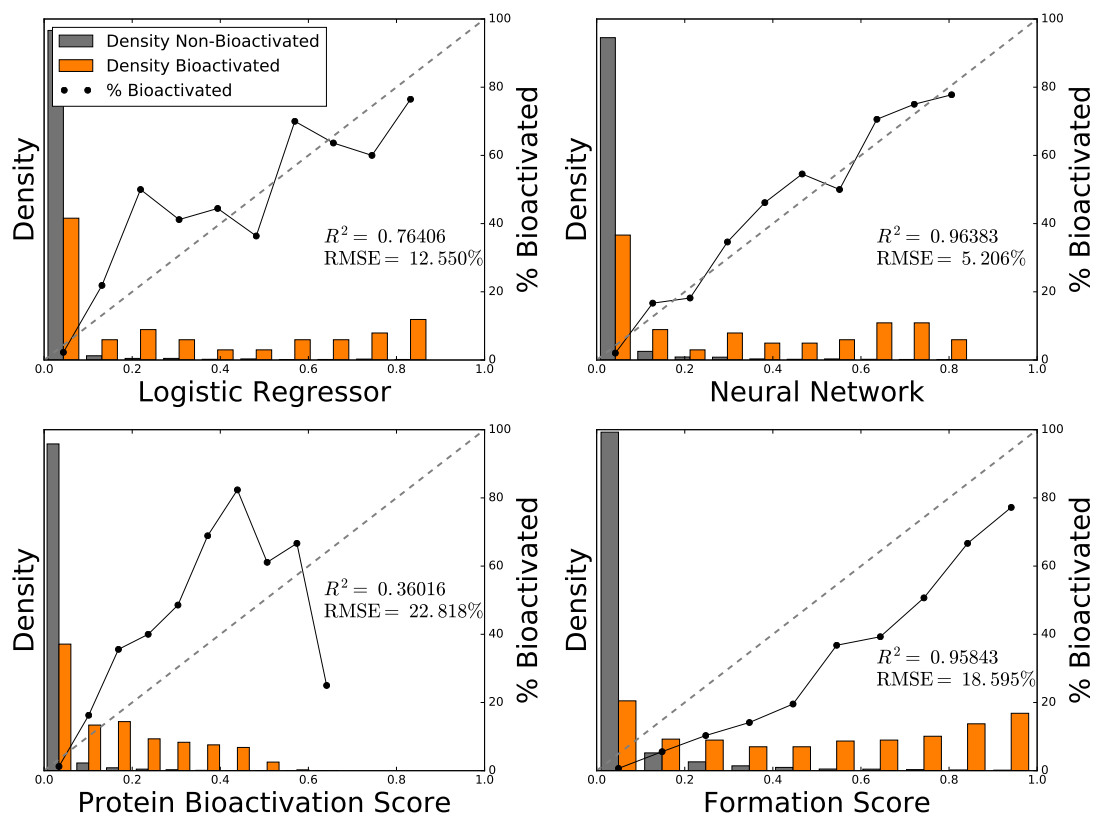


FIGURE 7.6: .

. 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 398 pathways within 340 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 root-mean-square error (RMSE) compared to a perfectly scaled prediction.

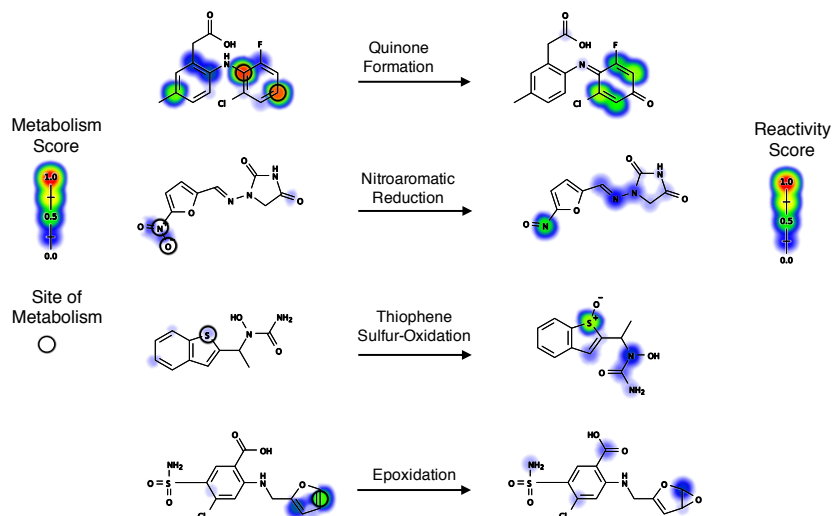


FIGURE 7.7: From top to bottom, ethacrynic acid, a diuretic, mianserin, an antidepressant, tolmetin, a nonsteroidal anti-inflammatory, and Imipramine, an antidepressant. For each drug, the experimentally known site of metabolism is circled in black, and the observed reactive metabolite displayed. tolmetin is known to produce two different reactive metabolites: an epoxide and a quinone, both of which are predicted and visualized. Notice that tolmetin's plane of symmetry means that the highly predicted sites on the other side of the molecule merely indicate the same quinone pathway.

is objectively the best option. Some might prefer the simpler logistic regressor model, for example, or the even simpler formation score descriptor.

We chose the neural network's scores because we felt that their probabilistic nature might be helpful for the next objective: predicting molecule bioactivation. Before pivoting to molecule-level bioactivation predictions, however, considered a few examples of bioactivation pathways correctly predicted by the neural network (Figure 7.7). Four drugs are visualized with their cross-validated PBS produced by the neural network, with the magnitude of the PBS indicated by the color shading gradient. For each drug, the experimentally known site of bioactivation is circled, and the observed reactive metabolite displayed

7.3.2 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, or simply set aside if there are acceptable alternatives. To enable the bioactivation model to make these molecule-level bioactivation predictions, we included in the training data molecules that are not bioactivated, despite having structures capable of forming quinones, epoxides, thiophene S-oxides, or aromatic nitrosos.

After the first-stage of training, which produced accurate pathway-level predictions, we performed a second training stage to tune the model to distinguish between bioactivated and non-bioactivated molecules. Performance at this objective was assessed by computing the “molecule AUC”: the AUC over all molecules in the training data set. We evaluated several models using this method.

The most intuitive approach is to simply assign each molecule its maximum cross-validated pathway score, which resulted in a molecule AUC of (Figure 7.8). This method produced a molecule AUC of 80.80%. A more complicated formula that still entails no training is to calculate the probabilistic OR over all PBS for a given input molecule. This is computed by subtracting each PBS from 1, multiplying all those decimals together, and subtracting that product from 1. Essentially, this operation calculates the the the joint probability of none of the bioactivation pathways occurring, and then takes the complement of that probability to produce a probability of at least one bioactivation taking place on a single molecule. The probabilistic OR had a molecule AUC of 82.37%.

Next, we explored the benefits of adding a second training step. For this training step, the descriptors for each molecule were its top twelve the probabilistic OR of all of its PBS, as well as all molecule topological descriptors, for a total of 21 descriptors. We used this descriptor set as the input for a second training stage, and tested both a logistic regressor and a neural network with ten hidden nodes. We used the same cross-validation folds as the pathway-level training, where any similar molecules were held out at the same time. The cross-validated scores produced by the logistic regressor had a molecule AUC of 81.69%, and the neural network had a molecule AUC of 81.42%.

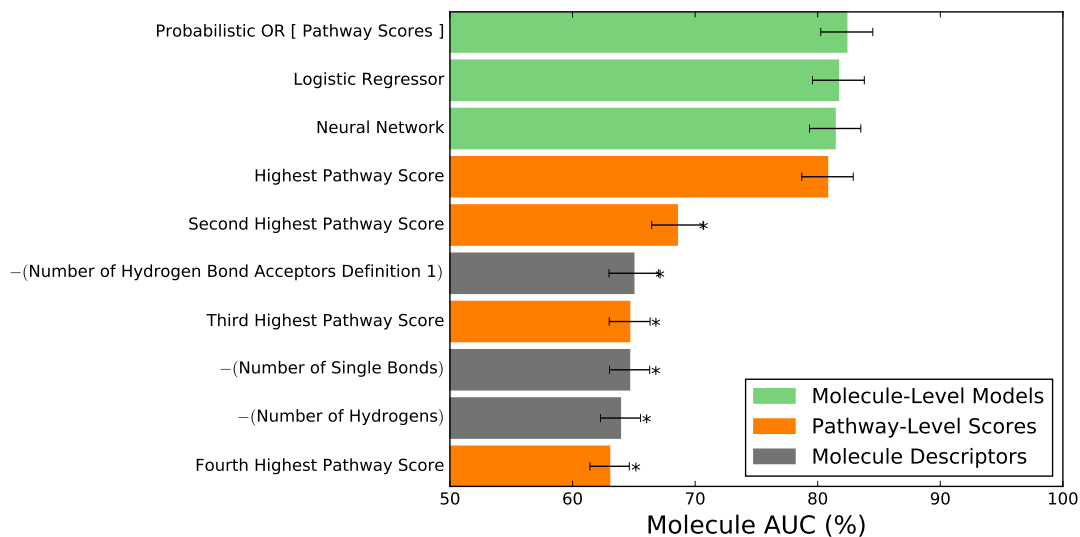


FIGURE 7.8: Molecule bioactivation scores (MBS) accurately identified bioactivated molecules

All four methods: the highest PBS, the probabilistic OR of all PBS, the neural network, and the logistic regressor, had statistically equivalent results when evaluated by a false positive rate paired t-test.[175] However, the probabilistic OR function outperformed all of the other three methods, with an R^2 of 0.97508 and an RMSE of 6.601. This might seem surprising, because generally the predictions from trained models are better calibrated than raw descriptors. However, in this case, the probabilistic OR does not merely represent a single chemical property, but instead by definition computes a probability that applies to the entire molecule. When considered in the light of simply being computed by the rules of probability, it is perhaps not so surprising that it performed better than other methods. Based on this result, we used the probabilistic OR to compute MBS for the remainder of this study.

As an external tests, we used five of the previously discussed batches of non-bioactivated molecules. Each batch consisted of 340 unique molecules that were not present in the training data. The PBS for each molecules in each external test set was predicted by the final trained model. For each test set, we measured the separation between the MBS of the test molecules and the MBS of the bioactivated molecules, computed using the cross-validated PBS.

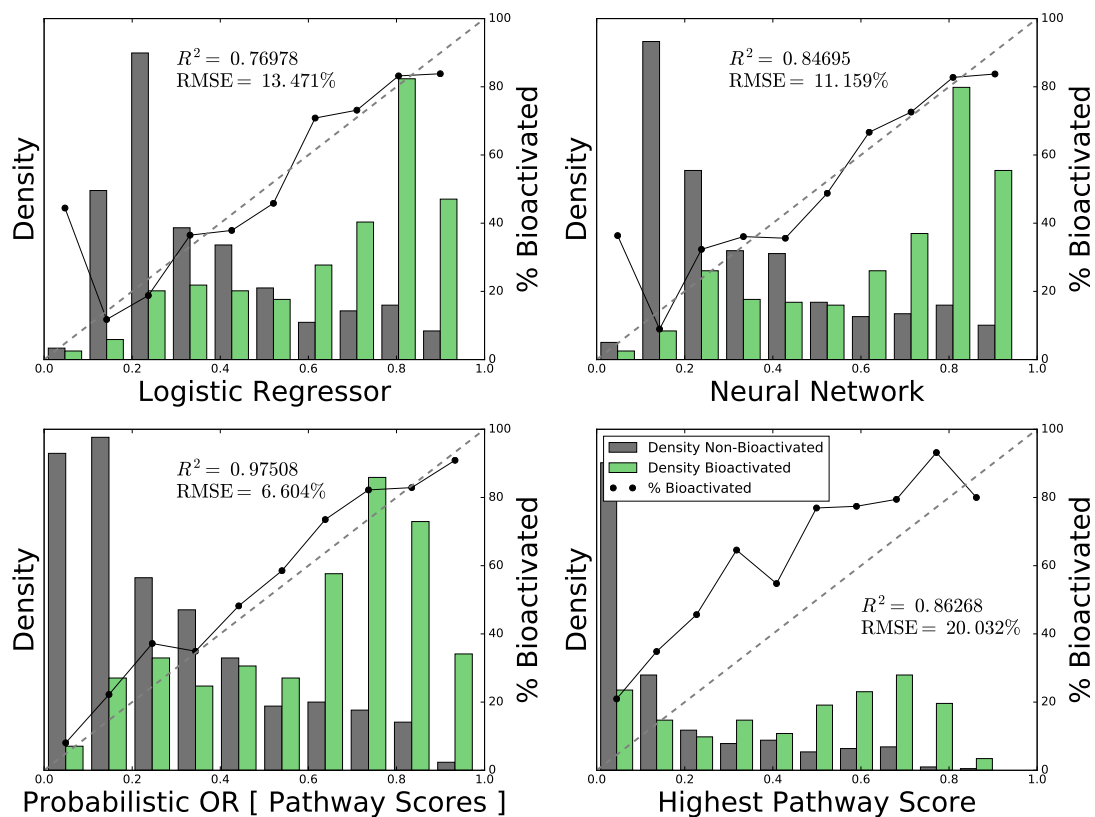


FIGURE 7.9: Of all methods, The probabilistic OR function produced the best-scaled MBS, with predictions that were highly probabilistic. The bar graphs plot the normalized distributions for each method across 398 pathways within 340 bioactivated molecules. The solid lines plot in each bin the percentage 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 probabilistic OR function produced the best scaled prediction of all methods, indicated by the highest correlation to the best fit-line, and the lowest root-mean-square error (RMSE) compared to a perfectly scaled prediction

Over the five external test sets, the molecule AUC was 79.92 ± 2.07 , which is equivalent to the accuracy of the MBS on the full cross validated training set 82.37 ± 1.60 . The model successfully generalized to new data, assigning non-bioactivated molecules it had never seen before with much lower scores than bioactivated molecules.

7.3.3 Toxicity Drivers Hypothesis Generation

We used the bioactivation model to enumerate hypotheses for the toxicity drivers of withdrawn drugs, after downloading a list of withdrawn drugs from DrugBank and removing any molecules found in our training data. Each of these 208 molecules were submitted to our final trained model, thereby producing a MBS for each molecule.

The model outputs probabilistic MBS, 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 7.10). While several methods for defining this cutoff are possible, in this study we computed the cutoff that optimizes both sensitivity and specificity.[343] This analysis resulted in a *MBS* threshold of 0.48. Using this threshold, we predicted that among the dataset of withdrawn drugs, 43 are bioactivated.

By providing hypotheses for the toxicity mechanisms of these drugs, the model provides a starting point for experimentalists seeking to unravel the etiology of certain IADRs. Perhaps especially interesting to investigate are drugs predicted to be bioactivated, but that are not currently known to do so. In terms of space, this is not an appropriate place list all the drugs potentially worthy of investigation, but I will list a few interesting examples (Figure 7.11).

For example, pipamazine, an antiemetic[344] was withdrawn from the U.S. in 1969 due to rare, severe cases of hepatotoxicity.[98, 248] The mechanism of these IADRs were not elucidated, and no reactive metabolites of pipamazine have been experimentally detected. Nevertheless, given that use of pipamazine was curtailed so long ago, there may have been little incentive

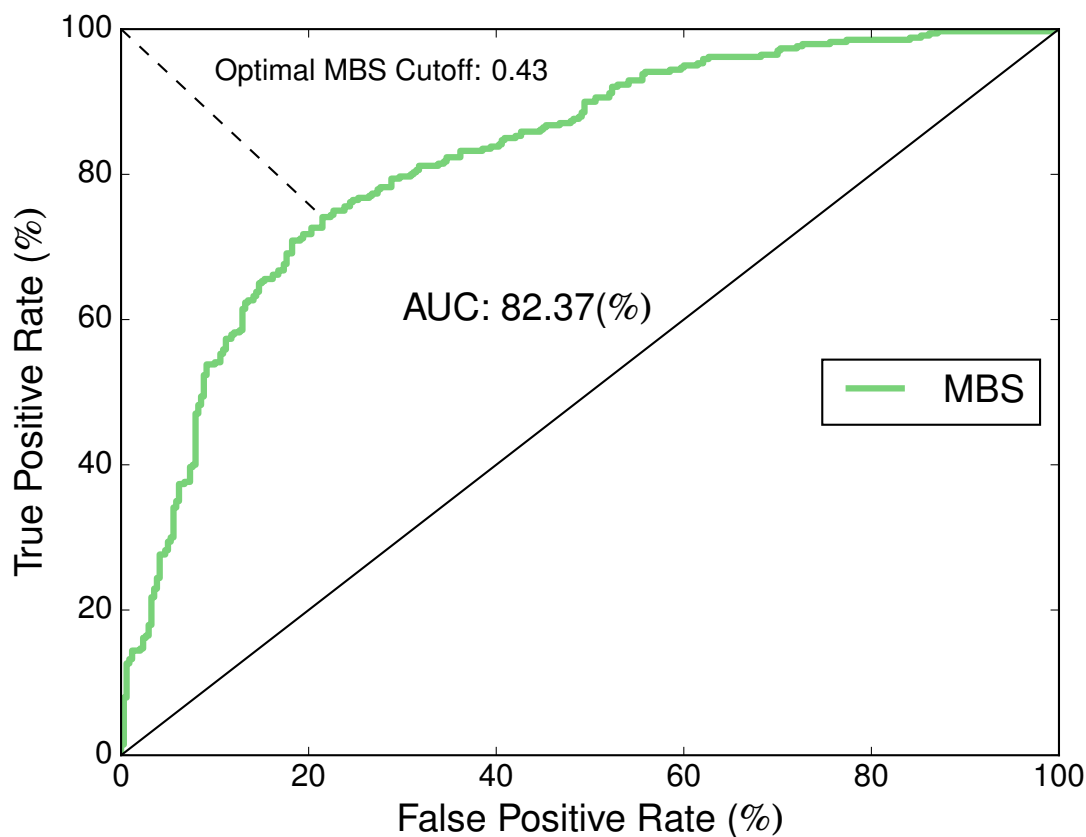


FIGURE 7.10: The ROC of the cross-validated molecule bioactivation scores was used to define to a score cutoff, which can be used to binarize predictions. Several methods for defining this cutoff are possible. In this study, we found the cutoff by computing the score on the ROC curve that is closet to 100% true positive rate and 0% false positive rate, thereby representing an optimal threshold for both sensitivity (the true positive rate) and specificity ($1 - \text{false positive rate}$).^[343] Based on whether sensitivity or specificity or more important for a given purpose, other methods of defining an optimal cutoff are certainly possible.

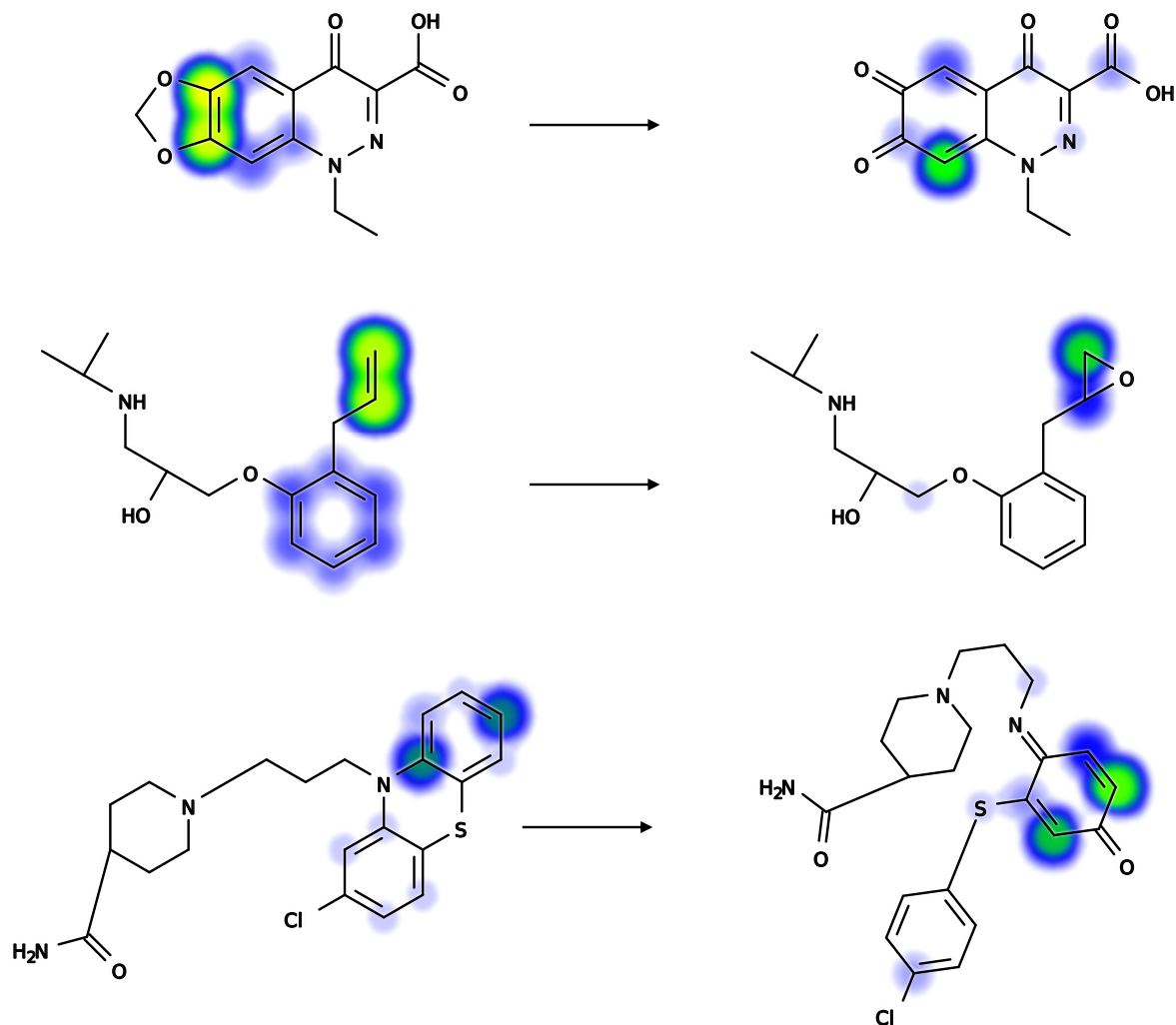


FIGURE 7.11: Examples of withdrawn drugs with high bioactivation predictions. We submitted the structures of 221 withdrawn drugs to the bioactivation model, the majority of which were downloaded directly from DrugBank, and supplemented by literature review. In order to generate toxicity mechanism hypotheses 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 top to bottom: thenalidine, cinoxacin, clprenolol, and pipamazine.

to study it with assays or technologies not available in the 1960s. The high score the bioactivation model assigns to pipamazine, suggests a possible explanation for the IADRs associated with pipamazine, with the top bioactivation pathway prediction consisting of a simultaneous quinone formation and dealkylation reaction.

7.4 Model Limitations

One limitation of our approach is that there are other bioactivation pathways beyond those included in this study. However, the approach demonstrated here is easily extendable to additional pathways. 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 include multiple steps, we plan to extend the ability of the model to pick up multi-step bioactivations. 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. We ultimately envision a broader toxicity model, where bioactivation predictions are combined with other important factors like daily dose, in order to build a model that explicitly predicts the toxicity risk of a certain molecule.

7.5 Conclusion

In this study, we built a bioactivation model that uses metabolism and reactivity predictions to predict whether a molecule will be metabolized into a metabolite that binds to protein or GSH. The model predicted with 89.04% AUC the bioactivation pathway within bioactivated molecules. These path-level predictions make a specific hypothesis about the mechanism of a molecule's toxicity. Furthermore, with 82.37% AUC the model predicted the bioactivation of molecules. Molecule-level predictions can be used to rapidly screen large number of molecules for the key toxicity risk of bioactivation. This model synthesizes several previous studies of both metabolism[121, 279, 319] and reactivity[108, 208] into a single model that weighs the two

Chapter 7. Modeling the Bioactivation and Subsequent Reactivity of Drugs

components of bioactivation. Future work will expand the power of this model by enumerating additional metabolic reactions. Furthermore, the model will be improved by considering multiple metabolic steps that precede reactive metabolite formation.

Bibliography

- [1] Munir Pirmohamed et al. “Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients”. In: *BMJ [Br. Med. J.]* 329.7456 (2004), pp. 15–19.
- [2] Sze Ling Chan et al. “Prevalence and characteristics of adverse drug reactions at admission to hospital: a prospective observational study”. In: *Br. J. Clin. Pharmacol.* 82.6 (2016), pp. 1636–1646.
- [3] TJ Oscanoa, F Lizaraso, and Alfonso Carvajal. “Hospital admissions due to adverse drug reactions in the elderly. A meta-analysis”. In: *Eur. J. Clin. Pharmacol.* 73.6 (2017), pp. 759–770.
- [4] Alexandra Alexopoulou et al. “Adverse drug reactions as a cause of hospital admissions: a 6-month experience in a single center in Greece”. In: *Eur. J. Intern. Med.* 19.7 (2008), pp. 505–510.
- [5] Ruth A Roberts et al. “Reducing attrition in drug development: smart loading preclinical safety assessment”. In: *Drug Discovery Today* 19.3 (2014), pp. 341–347.
- [6] Ismail Kola and John Landis. “Can the pharmaceutical industry reduce attrition rates?” In: *Nat. Rev. Drug Discovery* 3.8 (2004), pp. 711–716.
- [7] Trial Watch Phase II. “Phase III attrition rates 2011-2012 Arrowsmith, John; Miller, Philip”. In: *Nat. Rev. Drug Discovery* 12.8 (2013), p. 569.
- [8] James M McKim Jr. “Building a tiered approach to in vitro predictive toxicity screening: a focus on assays with in vivo relevance”. In: *Comb. Chem. High Throughput Screening* 13.2 (2010), p. 188.
- [9] Antonia F Stepan et al. “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”. In: *Chem. Res. Toxicol.* 24.9 (2011), pp. 1345–1410.
- [10] Joseph A DiMasi. “Success rates for new drugs entering clinical testing in the United States”. In: *Clin. Pharmacol. Ther.* 58.1 (1995), pp. 1–14.
- [11] I Ralph Edwards and Jeffrey K Aronson. “Adverse drug reactions: definitions, diagnosis, and management”. In: *The Lancet* 356.9237 (2000), pp. 1255–1259.

BIBLIOGRAPHY

- [12] Jack Uetrecht and Dean J Naisbitt. “Idiosyncratic adverse drug reactions: current concepts”. In: *Pharmacological reviews* 65.2 (2013), pp. 779–808.
- [13] Jonathon N Bauman et al. “Can in vitro metabolism-dependent covalent binding data distinguish hepatotoxic from nonhepatotoxic drugs? An analysis using human hepatocytes and liver S-9 fraction”. In: *Chemical research in toxicology* 22.2 (2009), pp. 332–340.
- [14] William M Lee. “Drug-induced hepatotoxicity”. In: *N. Engl. J. Med.* 349.5 (2003), pp. 474–485.
- [15] Andreas Brink et al. “Minimizing the risk of chemically reactive metabolite formation of new drug candidates: implications for preclinical drug design”. In: *Drug Discovery Today* 22.5 (2017), pp. 751–756.
- [16] Samy Babai, Laurent Auclert, and Hervé Le-Louët. “Safety data and withdrawal of hepatotoxic drugs Hepatotoxic drugs withdrawn”. In: *Thérapie* (2018).
- [17] S Schadt et al. “Minimizing DILI risk in drug discovery—a screening tool for drug candidates”. In: *Toxicology in Vitro* 30.1 (2015), pp. 429–437.
- [18] Tamara Alempijevic, Simon Zec, and Tomica Milosavljevic. “Drug-induced liver injury: Do we know everything?” In: *World journal of hepatology* 9.10 (2017), p. 491.
- [19] A Srivastava et al. “Role of reactive metabolites in drug-induced hepatotoxicity”. In: *Adverse Drug Reactions*. Springer: Berlin, 2010, pp. 165–194.
- [20] Michael T Solotke et al. “New and incremental FDA black box warnings from 2008 to 2015”. In: *Expert opinion on drug safety* 17.2 (2018), pp. 117–123.
- [21] Food, Drug Administration, et al. “Guidance for industry: warnings and precautions, contraindications, and boxed warning sections of labeling for human prescription drug and biological products—content and format”. In: *Food and Drug Administration* (2011).
- [22] Amit S Kalgutkar and Deepak Dalvie. “Predicting Toxicities of Reactive Metabolite-Positive Drug Candidates”. In: *Annu. Rev. Pharmacol. Toxicol.* 55 (2015), pp. 35–54.
- [23] Neil Kaplowitz. “Idiosyncratic drug hepatotoxicity”. In: *Nat. Rev. Drug Discovery* 4.6 (2005), p. 489.
- [24] Jacintha M Shenton, Jie Chen, and Jack P Uetrecht. “Animal models of idiosyncratic drug reactions”. In: *Chem.-Biol. Interact.* 150.1 (2004), pp. 53–70.
- [25] Matti Lang and Olavi Pelkonen. “Metabolism of xenobiotics and chemical carcinogenesis.” In: *IARC scientific publications* 148 (1999), pp. 13–22.
- [26] John R Bend, Cosette J Serabjit-Singh, and Richard M Philpot. “The pulmonary uptake, accumulation, and metabolism of xenobiotics”. In: *Annual review of pharmacology and toxicology* 25.1 (1985), pp. 97–125.

BIBLIOGRAPHY

- [27] Daniel W Nebert and Timothy P Dalton. “The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis”. In: *Nature Reviews Cancer* 6.12 (2006), p. 947.
- [28] Elizabeth A Martin et al. “³²P-postlabelled DNA adducts in liver obtained from women treated with tamoxifen”. In: *Carcinogenesis* 16.7 (1995), pp. 1651–1654.
- [29] David J Boocock et al. “ α -Hydroxytamoxifen, a genotoxic metabolite of tamoxifen in the rat: identification and quantification in vivo and in vitro”. In: *Carcinogenesis* 20.1 (1999), pp. 153–160.
- [30] David J Boocock et al. “Major inter-species differences in the rates of O-sulphonation and O-glucuronylation of α -hydroxytamoxifen in vitro: a metabolic disparity protecting human liver from the formation of tamoxifen–DNA adducts”. In: *Carcinogenesis* 21.10 (2000), pp. 1851–1858.
- [31] Andrew V. Stachulski et al. “The Generation, Detection, and Effects of Reactive Drug Metabolites”. In: *Med. Res. Rev.* 33.5 (2013), pp. 985–1080. ISSN: 1098-1128. DOI: [10.1002/med.21273](https://doi.org/10.1002/med.21273). URL: <http://dx.doi.org/10.1002/med.21273>.
- [32] Ronald N Germain. “MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation”. In: *Cell* 76.2 (1994), pp. 287–299.
- [33] Florian Bihl et al. “Impact of HLA-B alleles, epitope binding affinity, functional avidity, and viral coinfection on the immunodominance of virus-specific CTL responses”. In: *The Journal of Immunology* 176.7 (2006), pp. 4094–4101.
- [34] J Craig Hartman et al. “Evaluation of the endothelin receptor antagonists ambrisentan, darusentan, bosentan, and sitaxsentan as substrates and inhibitors of hepatobiliary transporters in sandwich-cultured human hepatocytes This article is one of a selection of papers published in the two-part special issue entitled 20 Years of Endothelin Research.” In: *Canadian journal of physiology and pharmacology* 88.6 (2010), pp. 682–691.
- [35] Jed Zaretski et al. “RS-Predictor: A new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4”. In: *J. Chem. Inf. Model.* 51.7 (2011), pp. 1667–1689.
- [36] Jed Zaretski et al. “RS-Predictor Models Augmented with SMARTCyp Reactivities: Robust Metabolic Regioselectivity Predictions for Nine CYP Isozymes”. In: *J. Chem. Inf. Model.* 52.6 (2012), pp. 1637–1659.
- [37] Jed Zaretski et al. “RS-WebPredictor: a server for predicting CYP-mediated sites of metabolism on drug-like molecules”. In: *Bioinformatics* 29.4 (2013), pp. 497–498.

BIBLIOGRAPHY

- [38] Katie Chan, Raymond Poon, and Peter J O'Brien. "Application of structure–activity relationships to investigate the molecular mechanisms of hepatocyte toxicity and electrophilic reactivity of α , β -unsaturated aldehydes". In: *J. Appl. Toxicol.* 28.8 (2008), pp. 1027–1039.
- [39] Katie Chan and Peter J O'Brien. "Structure–activity relationships for hepatocyte toxicity and electrophilic reactivity of α , β -unsaturated esters, acrylates and methacrylates". In: *J. Appl. Toxicol.* 28.8 (2008), pp. 1004–1015.
- [40] Katie Chan, Neil Jensen, and Peter J O'Brien. "Structure–activity relationships for thiol reactivity and rat or human hepatocyte toxicity induced by substituted p-benzoquinone compounds". In: *J. Appl. Toxicol.* 28.5 (2008), pp. 608–620.
- [41] JP Hughes et al. "Principles of early drug discovery". In: *Br. J. Pharmacol.* 162.6 (2011), pp. 1239–1249.
- [42] Christopher Paul Adams and Van Vu Brantner. "Spending on new drug development". In: *Health Econ.* 19.2 (2010), pp. 130–141.
- [43] David W Borhani and David E Shaw. "The future of molecular dynamics simulations in drug discovery". In: *J. Comput.-Aided Mol. Des.* 26.1 (2012), pp. 15–26.
- [44] Mark Kessel. "The problems with today's pharmaceutical business—an outsider's view". In: *Nat. Biotechnol.* 29.1 (2011), pp. 27–33.
- [45] Isa ELM Kuppens et al. "Modulation of oral drug bioavailability: from preclinical mechanism to therapeutic application". In: *Cancer investigation* 23.5 (2005), pp. 443–464.
- [46] Amit S Kalgutkar and Mary T Didiuk. "Structural alerts, reactive metabolites, and protein covalent binding: how reliable are these attributes as predictors of drug toxicity?" In: *Chemistry & biodiversity* 6.11 (2009), pp. 2115–2137.
- [47] Zeguang Wu et al. "Acute liver failure: mechanisms of immune-mediated liver injury". In: *Liver International* 30.6 (2010), pp. 782–794.
- [48] Sandra R Knowles, Jack Uetrecht, and Neil H Shear. "Idiosyncratic drug reactions: the reactive metabolite syndromes". In: *Lancet* 356.9241 (2000), pp. 1587–1591.
- [49] Kosuke Numata et al. "Overexpression of suppressor of cytokine signaling-3 in T cells exacerbates acetaminophen-induced hepatotoxicity". In: *J. Immunol.* 178.6 (2007), pp. 3777–3785.
- [50] Jed Zaretski, Matthew Matlock, and S Joshua Swamidass. "XenoSite: Accurately predicting CYP-mediated sites of metabolism with neural networks". In: *J. Chem. Inf. Model.* 53.12 (2013), pp. 3373–3383.
- [51] Valérie Campagna-Slater et al. "Development of a Computational Tool to Rival Experts in the Prediction of Sites of Metabolism of Xenobiotics by P450s". In: *J. Chem. Inf. Model.* (2012). DOI: [10.1021/ci3003073](https://doi.org/10.1021/ci3003073).

BIBLIOGRAPHY

- [52] Patrik Rydberg et al. "SMARTCyp: A 2D method for prediction of cytochrome P450-mediated drug metabolism". In: *ACS Med. Chem. Lett.* 1.3 (2010), pp. 96–100.
- [53] Jeffrey P Jones, Michael Mysinger, and Kenneth Ray Korzekwa. "Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction". In: *Drug Metab. Dispos.* 30.1 (2002), pp. 7–12.
- [54] Doo Nam Kim et al. "EaMEAD: Activation energy prediction of cytochrome P450 mediated metabolism with effective atomic descriptors". In: *J. Chem. Inf. Model.* 49.7 (2009), pp. 1643–1654.
- [55] Ke Yu et al. "High Daily Dose and Being a Substrate of Cytochrome P450 Enzymes Are Two Important Predictors of Drug-Induced Liver Injury". In: *Drug Metab. Dispos.* 42.4 (2014), pp. 744–750.
- [56] David Asturiol and Andrew Worth. *The Use of Chemical Reactivity Assays in Toxicity Prediction*. 2011.
- [57] K Fujimoto et al. "In vitro cytotoxicity assay to evaluate the toxicity of an electrophilic reactive metabolite using glutathione-depleted rat primary cultured hepatocytes." In: *Chem.-Biol. Interact.* 188.3 (2010), pp. 404–411.
- [58] Laura P James, Philip R Mayeux, and Jack A Hinson. "Acetaminophen-induced hepatotoxicity". In: *Drug Metab. Dispos.* 31.12 (2003), pp. 1499–1506.
- [59] JR Mitchell et al. "Acetaminophen-induced hepatic necrosis. IV. Protective role of glutathione". In: *J. Pharmacol. Exp. Ther.* 187.1 (1973), pp. 211–217.
- [60] G Frank Gerberick et al. "Development of a peptide reactivity assay for screening contact allergens". In: *Toxicol. Sci.* 81.2 (2004), pp. 332–343.
- [61] T Wayne Schultz et al. "Verification of the structural alerts for Michael acceptors". In: *Chem. Res. Toxicol.* 20.9 (2007), pp. 1359–1363.
- [62] Patrick Bultinck et al. "Influence of electron correlation and degeneracy on the Fukui matrix and extension of frontier molecular orbital theory to correlated quantum chemical methods". In: *Phys. Chem. Chem. Phys.* 14.7 (2012), pp. 2408–2416.
- [63] Kenichi Fukui, Teijiro Yonezawa, and Haruo Shingu. "A molecular orbital theory of reactivity in aromatic hydrocarbons". In: *J. Chem. Phys.* 20.4 (1952), pp. 722–725.
- [64] RK Roy et al. "Local softness and hardness based reactivity descriptors for predicting intra-and intermolecular reactivity sequences: carbonyl compounds". In: *J. Phys. Chem. A* 102.21 (1998), pp. 3746–3755.
- [65] Christophe Morell, André Grand, and Alejandro Toro-Labbé. "New dual descriptor for chemical reactivity". In: *J. Phys. Chem. A* 109.1 (2005), pp. 205–212.

BIBLIOGRAPHY

- [66] Pratim Kumar Chattaraj, Buddhadev Maiti, and Utpal Sarkar. "Philocity: a unified treatment of chemical reactivity and selectivity". In: *J. Phys. Chem. A* 107.25 (2003), pp. 4973–4975.
- [67] P. Baldi and S. Brunak. *Bioinformatics: the machine learning approach*. The MIT Press, 2001.
- [68] Noel M OLBoyle et al. "Open Babel: An open chemical toolbox". In: *J Cheminf* 3 (2011), p. 33.
- [69] Patrik Rydberg et al. "The contribution of atom accessibility to site of metabolism models for cytochromes P450". In: *Molecular pharmaceutics* 10.4 (2013), pp. 1216–1223.
- [70] Mati Karelson, Victor S Lobanov, and Alan R Katritzky. "Quantum-chemical descriptors in QSAR/QSPR studies". In: *Chem. Rev.* 96.3 (1996), pp. 1027–1044.
- [71] Mark TD Cronin and Judith C Madden. *In Silico Toxicology: Principles and Applications*. 7. Royal Society of Chemistry: Liverpool, 2010.
- [72] Johannes AH Schwöbel et al. "Measurement and estimation of electrophilic reactivity for predictive toxicology". In: *Chem. Rev.* 111.4 (2011), pp. 2562–2596.
- [73] James JP Stewart. "MOPAC: a semiempirical molecular orbital program". In: *J. Comput.-Aided Mol. Des.* 4.1 (1990), pp. 1–103.
- [74] Jiří Hostaš, Jan Řezáč, and Pavel Hobza. "On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions". In: *Chem. Phys. Lett.* 568 (2013), pp. 161–166.
- [75] R.J. Nasr, S.J. Swamidass, and P.F. Baldi. "Large scale study of multiple-molecule queries". In: *Journal of Cheminformatics* 1.1 (2009). PMID: 20298525, p. 7. ISSN: 1758-2946.
- [76] G. Schneider, Y. Tanrikulu, and P. Schneider. "Self-organizing molecular fingerprints: a ligand-based view on drug-like chemical space and off-target prediction". In: *Future* 1.1 (2009). DOI: 10.4155/fmc.09.11, pp. 213–218.
- [77] Qian Zhu et al. "WENDI: A tool for finding non-obvious relationships between compounds and biological properties, genes, diseases and scholarly publications". In: *Journal of Cheminformatics* 2.1 (2010). PMID: 20727184, p. 6. ISSN: 1758-2946. DOI: 10.1186/1758-2946-2-6. URL: <http://www.jcheminf.com/content/2/1/6>.
- [78] M.J. Keiser et al. "Relating protein pharmacology by ligand chemistry". In: *Nature Biotechnology* 25.2 (2007). PMID: 17287757, pp. 197–206.
- [79] M. Glick et al. "Enrichment of extremely noisy high-throughput screening data using a naive Bayes classifier". In: *Journal of Biomolecular Screening* 9.1 (2004). PMID: 15006146, p. 32. ISSN: 1087-0571.

BIBLIOGRAPHY

- [80] T.J. Crisman et al. ““Plate Cherry Picking”: A Novel Semi-Sequential Screening Paradigm for Cheaper, Faster, Information-Rich Compound Selection”. In: *Journal of Biomolecular Screening* 12.3 (2007), p. 320. ISSN: 1087-0571.
- [81] E Albano et al. “Mechanisms of N-acetyl-p-benzoquinone imine cytotoxicity.” In: *Mol. Pharmacol.* 28.3 (1985), pp. 306–311.
- [82] Noel M O’Boyle, Chris Morley, and Geoffrey R Hutchison. “Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit”. In: *Chem. Cent. J.* 2.5 (2008), pp. 1–5.
- [83] Michael J Sorich et al. “Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms”. In: *Journal of Chemical Information and Computer Sciences* 43.6 (2003), pp. 2019–2024.
- [84] Andrew Hunter et al. “Application of neural networks and sensitivity analysis to improved prediction of trauma survival”. In: *Comp. Meth. Progr. Biom.* 62.1 (2000), pp. 11–19.
- [85] JACEK Ziolkowski. “New relation between ionic radii, bond length, and bond strength”. In: *J. Solid State Chem* 57.3 (1985), pp. 269–90.
- [86] ID t Brown and RD Shannon. “Empirical bond-strength-bond-length curves for oxides”. In: *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 29.3 (1973), pp. 266–282.
- [87] Diana Thaens et al. “Chemoassay Screening of DNA-Reactive Mutagenicity with 4-(4-Nitrobenzyl) pyridine—Application to Epoxides, Oxetanes, and Sulfur Heterocycles”. In: *Chem. Res. Toxicol.* 25.10 (2012), pp. 2092–2102.
- [88] Amit S Kalgutkar et al. “A comprehensive listing of bioactivation pathways of organic functional groups”. In: *Curr. Drug Metab.* 6.3 (2005), pp. 161–225.
- [89] Micaela C Damsten et al. “Trimethoprim: novel reactive intermediates and bioactivation pathways by cytochrome p450s”. In: *Chem. Res. Toxicol.* 21.11 (2008), pp. 2181–2187.
- [90] Jan Scicinski et al. “Preclinical evaluation of the metabolism and disposition of RRx-001, a novel investigative anticancer agent”. In: *Drug Metab. Dispos.* 40.9 (2012), pp. 1810–1816.
- [91] Hao Sun and Garold S Yost. “Metabolic activation of a novel 3-substituted indole-containing TNF- α inhibitor: dehydrogenation and inactivation of CYP3A4”. In: *Chem. Res. Toxicol.* 21.2 (2007), pp. 374–385.
- [92] H Umesh Shetty et al. “Radiodefluorination of 3-fluoro-5-(2-(2-[¹⁸F](fluoromethyl)-thiazol-4-yl) ethynyl) benzonitrile ([¹⁸F] SP203), a radioligand for imaging brain metabotropic glutamate subtype-5 receptors with positron emission tomography, occurs by glutathionylation in rat brain”. In: *J. Pharmacol. Exp. Ther.* 327.3 (2008), pp. 727–735.

BIBLIOGRAPHY

- [93] André LeBlanc and Lekha Sleno. "Atrazine metabolite screening in human microsomes: detection of novel reactive metabolites and glutathione adducts by LC-MS". In: *Chem. Res. Toxicol.* 24.3 (2011), pp. 329–339.
- [94] Lin Xu et al. "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". In: *Drug Metab. Dispos.* 40.4 (2012), pp. 680–693.
- [95] JD deBethizy and DE Rickert. "Metabolism of nitrotoluenes by freshly isolated Fischer 344 rat hepatocytes." In: *Drug metabolism and disposition: the biological fate of chemicals* 12.1 (1984), p. 45.
- [96] Walter P Frankmoelle et al. "Glutathione S-transferase metabolism of the antineoplastic pentafluorophenylsulfonamide in tissue culture and mice". In: *Drug Metab. Dispos.* 28.8 (2000), pp. 951–958.
- [97] Amit S Kalgutkar et al. "A rational chemical intervention strategy to circumvent bioactivation liabilities associated with a nonpeptidyl thrombopoietin receptor agonist containing a 2-amino-4-arylthiazole motif". In: *Chem. Res. Toxicol.* 20.12 (2007), pp. 1954–1965.
- [98] F Peter Guengerich. "Mechanisms of drug toxicity and relevance to pharmaceutical development". In: *Drug Metab. Pharmacokinet.* 26.1 (2011), pp. 3–14.
- [99] Werner K Lutz. "In vivo covalent binding of organic chemicals to DNA as a quantitative indicator in the process of chemical carcinogenesis". In: *Mutat. Res., Rev. Genet. Toxicol.* 65.4 (1979), pp. 289–356.
- [100] Harry V Gelboin. "Benzo [alpha] pyrene metabolism, activation and carcinogenesis: role and regulation of mixed-function oxidases and related enzymes." In: *Physiol. Rev.* 60.4 (1980), pp. 1107–1166.
- [101] Richard M LoPachin and Terrence Gavin. "Molecular mechanisms of aldehyde toxicity: a chemical perspective". In: *Chem. Res. Toxicol.* 27.7 (2014), pp. 1081–1091.
- [102] B Kevin Park et al. "The role of metabolic activation in drug-induced hepatotoxicity". In: *Annu. Rev. Pharmacol. Toxicol.* 45 (2005), pp. 177–202.
- [103] Sabry M Attia. "Deleterious effects of reactive metabolites". In: *Oxid. Med. Cell. Longevity* 3.4 (2010), pp. 238–253.
- [104] Michelle K Dennehy et al. "Cytosolic and nuclear protein targets of thiol-reactive electrophiles". In: *Chem. Res. Toxicol.* 19.1 (2006), pp. 20–29.
- [105] Shuguang Ma and Raju Subramanian. "Detecting and characterizing reactive metabolites by liquid chromatography/tandem mass spectrometry". In: *J. Mass Spectrom.* 41.9 (2006), pp. 1121–1139.

BIBLIOGRAPHY

- [106] Georgina Meneses-Lorente et al. “A quantitative high-throughput trapping assay as a measurement of potential for bioactivation”. In: *Anal. Biochem.* 351.2 (2006), pp. 266–272.
- [107] Davide Montero et al. “Intracellular glutathione pools are heterogeneously concentrated”. In: *Redox Biol.* 1.1 (2013), pp. 508–513.
- [108] Tyler B Hughes, Grover P Miller, and S Joshua Swamidass. “Site of Reactivity Models Predict Molecular Reactivity of Diverse Chemicals with Glutathione”. In: *Chem. Res. Toxicol.* 28.4 (2015), pp. 797–809.
- [109] Dayana Argoti et al. “Cyanide trapping of iminium ion reactive intermediates followed by detection and structure identification using liquid chromatography-tandem mass spectrometry (LC-MS/MS)”. In: *Chem. Res. Toxicol.* 18.10 (2005), pp. 1537–1544.
- [110] DM DeMarini et al. “Urinary mutagenicity as a biomarker in workers exposed to benzidine: correlation with urinary metabolites and urothelial DNA adducts.” In: *Carcinogenesis* 18.5 (1997), pp. 981–988.
- [111] Amy F Loughlin et al. “An ion exchange liquid chromatography/mass spectrometry method for the determination of reduced and oxidized glutathione and glutathione conjugates in hepatocytes”. In: *J. Pharm. Biomed. Anal.* 26.1 (2001), pp. 131–142.
- [112] Michael D Mitchell et al. “Peptide-based in vitro assay for the detection of reactive metabolites”. In: *Chem. Res. Toxicol.* 21.4 (2008), pp. 859–868.
- [113] Ronan Collobert and Jason Weston. “A unified architecture for natural language processing: Deep neural networks with multitask learning”. In: *Proceedings of the 25th international conference on Machine learning*. ACM. 2008, pp. 160–167.
- [114] Sinno Jialin Pan and Qiang Yang. “A survey on transfer learning”. In: *Knowledge and Data Engineering, IEEE Transactions on* 22.10 (2010), pp. 1345–1359.
- [115] Farooq Azam. “Biologically inspired modular neural networks”. PhD thesis. Citeseer, 2000.
- [116] Rangachari Anand et al. “Efficient classification for multiclass problems using modular neural networks”. In: *Neural Networks, IEEE Transactions on* 6.1 (1995), pp. 117–124.
- [117] Daniel N Osherson, Scott Weinstein, and Michael Stob. “Modular learning”. In: *Computational neuroscience*. MIT Press: Cambridge, MA, 1993, pp. 369–377.
- [118] Gasser Auda and Mohamed Kamel. “Modular neural network classifiers: A comparative study”. In: *J. Intell. Rob. Syst.* 21.2 (Feb. 1998), pp. 117–129.
- [119] Bao-Liang Lu and Masami Ito. “Task decomposition based on class relations: a modular neural network architecture for pattern classification”. In: *Biological and Artificial Computation: From Neuroscience to Technology*. Springer: Berlin, 1997, pp. 330–339.

BIBLIOGRAPHY

- [120] Konstantin V Balakin et al. “Quantitative structure-metabolism relationship modeling of metabolic N-dealkylation reaction rates”. In: *Drug Metab. Dispos.* 32.10 (2004), pp. 1111–1120.
- [121] Tyler B Hughes, Grover P Miller, and S Joshua Swamidass. “Modeling Epoxidation of Drug-like molecules with a Deep Machine Learning Network”. In: *ACS Cent. Sci.* 1.4 (2015), pp. 168–180.
- [122] Gabriele Cruciani et al. “Exposition and reactivity optimization to predict sites of metabolism in chemicals”. In: *Drug Discovery Today: Technol.* 10.1 (2013), e155–e165.
- [123] Ismael Zamora, Lovisa Afzelius, and Gabriele Cruciani. “Predicting drug metabolism: a site of metabolism prediction tool applied to the cytochrome P450 2C9”. In: *J. Med. Chem.* 46.12 (2003), pp. 2313–2324.
- [124] Abhishek Srivastava et al. “Identification and mitigation of a reactive metabolite liability associated with aminoimidazoles”. In: *Chem. Res. Toxicol.* 27.9 (2014), pp. 1586–1597.
- [125] Prashant Surve et al. “Metabolite characterization of anti-cancer agent gefitinib in human hepatocytes”. In: *Drug Metab. Lett.* 7.2 (2013), pp. 126–136.
- [126] Johan Bylund et al. “Metabolic profiling of TRPV1 antagonists of the benzothiazole amide series: implications for in vitro genotoxicity assessment”. In: *Xenobiotica* 43.2 (2013), pp. 201–210.
- [127] Lijuan Peng and Robert J Turesky. “Mass spectrometric characterization of 2-amino-1-methyl-6-phenylimidazo [4, 5-b] pyridine N-oxidized metabolites bound at Cys34 of human serum albumin”. In: *Chem. Res. Toxicol.* 24.11 (2011), pp. 2004–2017.
- [128] G Aislaitner et al. “Metabolism of (-)-(S)-nicotine in the isolated perfused rabbit lung”. In: *Eur. J. Drug Metab. Pharmacokinet.* 22.4 (1997), pp. 395–402.
- [129] Jennifer Q Dong, Jianhua Liu, and Philip C Smith. “Role of benoxaprofen and flunoxaprofen acyl glucuronides in covalent binding to rat plasma and liver proteins in vivo”. In: *Biochem. Pharmacol.* 70.6 (2005), pp. 937–948.
- [130] Eddie E Li et al. “Relationships of DNA adduct formation, K-ras activating mutations and tumorigenic activities of 6-nitrochrysene and its metabolites in the lungs of CD-1 mice”. In: *Carcinogenesis* 15.7 (1994), pp. 1377–1385.
- [131] H Nakamura et al. “In vitro formation, disposition and toxicity of N-acetoxy-sulfamethoxazole, a potential mediator of sulfamethoxazole toxicity.” In: *J. Pharmacol. Exp. Ther.* 274.3 (1995), pp. 1099–1104.
- [132] Yuji Ishii et al. “Determination of lucidin-specific DNA adducts by liquid chromatography with tandem mass spectrometry in the livers and kidneys of rats given lucidin-3-O-primeveroside”. In: *Chem. Res. Toxicol.* 25.5 (2012), pp. 1112–1118.

BIBLIOGRAPHY

- [133] Weiling Xue et al. "A metabolic activation mechanism of 7 H-dibenzo [c, g] carbazole via O-quinone. Part 2: Covalent adducts of 7 H-dibenzo [c, g] carbazole-3, 4-dione with nucleic acid bases and nucleosides". In: *Chem. Res. Toxicol.* 15.7 (2002), pp. 915–921.
- [134] Jed Zaretski, Kevin M Boehm, and S Joshua Swamidass. "Improved Prediction of CYP-Mediated Metabolism with Chemical Fingerprints". In: *J. Chem. Inf. Model.* 55.5 (2015), pp. 972–982.
- [135] Jed Zaretski et al. "Extending P450 Site-of-Metabolism Models with Region-Resolution Data". In: *Bioinformatics* 31.12 (2015), pp. 1966–1973. DOI: [10.1093/bioinformatics/btv100](https://doi.org/10.1093/bioinformatics/btv100). eprint: <http://bioinformatics.oxfordjournals.org/content/early/2015/02/18/bioinformatics.btv100.full.pdf+html>. URL: <http://bioinformatics.oxfordjournals.org/content/early/2015/02/18/bioinformatics.btv100.abstract>.
- [136] Matthew K Matlock, Tyler B Hughes, and S Joshua Swamidass. "XenoSite server: a web-available site of metabolism prediction tool". In: *Bioinformatics* 31.7 (2015), pp. 1136–1137.
- [137] Patrik Rydberg and Lars Olsen. "Predicting drug metabolism by cytochrome P450 2C9: comparison with the 2D6 and 3A4 isoforms". In: *ChemMedChem* 7.7 (2012), pp. 1202–1209.
- [138] Anastasia Rudik et al. "SOMP: web-server for in silico prediction of sites of metabolism for drug-like compounds". In: *Bioinformatics* 31 (2015), pp. 2046–2048.
- [139] Sophie L Regan et al. "Acyl glucuronides: the good, the bad and the ugly". In: *Bio pharm. Drug Dispos.* 31.7 (2010), pp. 367–395.
- [140] Na Le Dang et al. "A simple model predicts UGT-mediated metabolism". In: *Bioinformatics* 32.20 (2016), pp. 3183–3189.
- [141] Greg Landrum. "RDKit: Open-source cheminformatics". <http://www.rdkit.org>.
- [142] José Marco-Contelles, María Teresa Molina, and Shazia Anjum. "Naturally occurring cyclohexane epoxides: sources, biological activities, and synthesis". In: *Chem. Rev.* 104.6 (2004), pp. 2857–2900.
- [143] Tamara Delaine et al. "Epoxyalcohols: Bioactivation and Conjugation Required for Skin Sensitization". In: *Chem. Res. Toxicol.* 27.10 (2014), pp. 1860–1870.
- [144] Christophe Morisseau. "Role of epoxide hydrolases in lipid metabolism". In: *Biochimie* 95.1 (2013), pp. 91–95.
- [145] R-E_ Parker and NS Isaacs. "Mechanisms of epoxide reactions". In: *Chem. Rev.* 59.4 (1959), pp. 737–799.
- [146] Wei Tang. "Drug metabolite profiling and elucidation of drug-induced hepatotoxicity". In: (2007).

BIBLIOGRAPHY

- [147] David E Amacher. “The primary role of hepatic metabolism in idiosyncratic drug-induced liver injury”. In: *Expert Opin. Drug Metab. Toxicol.* 8.3 (2012), pp. 335–347.
- [148] Adrian J Fretland and Curtis J Omiecinski. “Epoxide hydrolases: biochemistry and molecular biology”. In: *Chem.-Biol. Interact.* 129.1 (2000), pp. 41–59.
- [149] David C Lamb, Michael R Waterman, and Bin Zhao. “Streptomyces cytochromes P450: applications in drug metabolism”. In: *Expert Opin. Drug Metab. Toxicol.* 9.10 (2013), pp. 1279–1294.
- [150] Patrik Rydberg et al. “Trends in predicted chemoselectivity of cytochrome P450 oxidation: B3LYP barrier heights for epoxidation and hydroxylation reactions”. In: *Journal of Molecular Graphics and Modelling* 52 (2014), pp. 30–35.
- [151] Slobodan Rendic and F Peter Guengerich. “Contributions of human enzymes in carcinogen metabolism”. In: *Chem. Res. Toxicol.* 25.7 (2012), pp. 1316–1383.
- [152] Bernard Testa, Alessandro Pedretti, and Giulio Vistoli. “Reactions and enzymes in the metabolism of drugs and other xenobiotics”. In: *Drug Discovery Today* 17.11 (2012), pp. 549–560.
- [153] Janeric Seidegård and Gunilla Ekström. “The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics.” In: *Environ. Health Perspect.* 105.Suppl 4 (1997), p. 791.
- [154] Sanjay K Srivastava et al. “Role of glutathione conjugate efflux in cellular protection against benzo [a] pyrene-7, 8-diol-9, 10-epoxide-induced DNA damage”. In: *Molecular carcinogenesis* 33.3 (2002), pp. 156–162.
- [155] Karima Begriche et al. “Drug-induced toxicity on mitochondria and lipid metabolism: mechanistic diversity and deleterious consequences for the liver”. In: *Journal of hepatology* 54.4 (2011), pp. 773–794.
- [156] Jose C Fernandez-Checa and Neil Kaplowitz. “Hepatic mitochondrial glutathione: transport and role in disease and toxicity”. In: *Toxicology and applied pharmacology* 204.3 (2005), pp. 263–273.
- [157] Mehran F Moghaddam et al. “Bioactivation of leukotoxins to their toxic diols by epoxide hydrolase”. In: *Nature medicine* 3.5 (1997), pp. 562–566.
- [158] R Scott Obach et al. “In vitro metabolism and covalent binding of enol-carboxamide derivatives and anti-inflammatory agents sudoxicam and meloxicam: insights into the hepatotoxicity of sudoxicam”. In: *Chem. Res. Toxicol.* 21.9 (2008), pp. 1890–1899.
- [159] Robin E Pearce, Jack P Uetrecht, and J Steven Leeder. “Pathways of carbamazepine bioactivation in vitro: II. The role of human cytochrome P450 enzymes in the formation of 2-hydroxyiminostilbene”. In: *Drug metabolism and disposition* 33.12 (2005), pp. 1819–1826.

BIBLIOGRAPHY

- [160] Michelle M Kalis and Nancy A Huff. "Oxcarbazepine, an antiepileptic agent". In: *Clinical therapeutics* 23.5 (2001), pp. 680–700.
- [161] Jennifer L Harder et al. "Carbamazepine and the active epoxide metabolite are effectively cleared by hemodialysis followed by continuous venovenous hemodialysis in an acute overdose". In: *Hemodialysis International* 15.3 (2011), pp. 412–415.
- [162] S Bhargavdilip, K Kishore, C Ramasamy, et al. "Comparision of efficacy and safety of carbamazepine and eslicarbazepine in adult partial and generalized seizures". In: *Asian Journal of Pharmaceutical and Clinical Research* 7.2 (2014).
- [163] Vincent Yip et al. "Covalent adduction of carbamazepine 10, 11-epoxide with human serum albumin and glutathione S-transferase pi: implications for carbamazepine hypersensitivity". In: *The Lancet* 383 (2014), S114.
- [164] Mogens Dam et al. "A double-blind study comparing oxcarbazepine and carbamazepine in patients with newly diagnosed, previously untreated epilepsy". In: *Epilepsy research* 3.1 (1989), pp. 70–76.
- [165] F Peter Guengerich. "Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity". In: *Chem. Res. Toxicol.* 14.6 (2001), pp. 611–650.
- [166] D Galeone, G Lamontanara, and D Torelli. "Acute hepatitis in a patient treated with carbamazepine". In: *Journal of neurology* 232.5 (1985), pp. 301–303.
- [167] Victoria J Green et al. "Genetic analysis of microsomal epoxide hydrolase in patients with carbamazepine hypersensitivity". In: *Biochem. Pharmacol.* 50.9 (1995), pp. 1353–1359.
- [168] BIOVIA. "Accelrys Metabolite". <http://accelrys.com/products/databases/bioactivity/metabolite.html>. Mar. 2015.
- [169] F Peter Guengerich. "Cytochrome P450 oxidations in the generation of reactive electrophiles: epoxidation and related reactions". In: *Arch. Biochem. Biophys.* 409.1 (2003), pp. 59–71.
- [170] Margarita Rojas et al. "Anti-benzo [a] pyrene diolepoxide—DNA adduct levels in peripheral mononuclear cells from coke oven workers and the enhancing effect of smoking". In: *Carcinogenesis* 16.6 (1995), pp. 1373–1376.
- [171] Abhishek Srivastava et al. "Quantifying the metabolic activation of nevirapine in patients by integrated applications of NMR and mass spectrometries". In: *Drug Metab. Dispos.* 38.1 (2010), pp. 122–132.
- [172] KB Alton et al. "Biotransformation of a 1, 5-benzodiazepine, triflubazam, by man". In: *Drug Metabolism and Disposition* 3.5 (1975), pp. 352–360.

BIBLIOGRAPHY

- [173] Davide Chicco, Peter Sadowski, and Pierre Baldi. “Deep autoencoder neural networks for gene ontology annotation predictions”. In: *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. ACM. 2014, pp. 533–540.
- [174] Tara N Sainath et al. “Deep convolutional neural networks for LVCSR”. In: *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE. 2013, pp. 8614–8618.
- [175] S Joshua Swamidass et al. “A CROC stronger than ROC: measuring, visualizing and optimizing early retrieval”. In: *Bioinformatics* 26.10 (2010), pp. 1348–1356.
- [176] Brinda Mahadevan et al. “Investigation of the genotoxicity of dibenzo [c, p] chrysene in human carcinoma MCF-7 cells in culture”. In: *Chemico-biological interactions* 164.3 (2006), pp. 181–191.
- [177] Xiaoliang Zhuo et al. “Comparative biotransformation of pyrazinone-containing corticotropin-releasing factor receptor-1 antagonists: minimizing the reactive metabolite formation”. In: *Drug Metabolism and Disposition* 38.1 (2010), pp. 5–15.
- [178] Raju Subramanian et al. “Cytochrome P450-mediated epoxidation of 2-aminothiazole-based AKT inhibitors: identification of novel GSH adducts and reduction of metabolic activation through structural changes guided by in silico and in vitro screening”. In: *Chem. Res. Toxicol.* 23.3 (2010), pp. 653–663.
- [179] Drazen Ostovic and Thomas C Bruice. “Intermediates in the epoxidation of alkenes by cytochrome P-450 models. 5. Epoxidation of alkenes catalyzed by a sterically hindered (meso-tetrakis (2, 6-dibromophenyl) porphinato) iron (III) chloride”. In: *Journal of the American Chemical Society* 111.17 (1989), pp. 6511–6517.
- [180] Masanori Sono et al. “Heme-containing oxygenases”. In: *Chem. Rev.* 96.7 (1996), pp. 2841–2888.
- [181] Yuan-Wan Sun et al. “Human cytochromes P450 1A1 and 1B1 catalyze ring oxidation but not nitroreduction of environmental pollutant mononitropyrene isomers in primary cultures of human breast cells and cultured MCF-10A and MCF-7 cell lines”. In: *Chem. Res. Toxicol.* 17.8 (2004), pp. 1077–1085.
- [182] William L Fitch et al. “Revisiting the metabolism of ketoconazole using accurate mass”. In: *Drug metabolism letters* 3.3 (2009), pp. 191–198.
- [183] RC Steenwyk and B Tan. “In vitro evidence for the formation of reactive intermediates of resveratrol in human liver microsomes”. In: *Xenobiotica* 40.1 (2010), pp. 62–71.
- [184] JNT Gilbert et al. “Metabolism of quinalbarbitone”. In: *Journal of Pharmacy and Pharmacology* 27.5 (1975), pp. 343–347.

BIBLIOGRAPHY

- [185] Donggeng Wang et al. “Investigation of piceid metabolites in rat by liquid chromatography tandem mass spectrometry”. In: *Journal of Chromatography B* 879.1 (2011), pp. 69–74.
- [186] Akira Namera et al. “Automated preparation and analysis of barbiturates in human urine using the combined system of PrepStation and gas chromatography–mass spectrometry”. In: *Journal of Chromatography B: Biomedical Sciences and Applications* 706.2 (1998), pp. 253–259.
- [187] Martin E Brown and Rif S El-Mallakh. “Role of eslicarbazepine in the treatment of epilepsy in adult patients with partial-onset seizures”. In: *Therapeutics and clinical risk management* 6 (2010), p. 103.
- [188] Jerry R Mitchell, Wayne R Snodgrass, and James R Gillette. “The role of biotransformation in chemical-induced liver injury.” In: *Environmental health perspectives* 15 (1976), p. 27.
- [189] JR Mitchell, WZ Potter, and DJ Jollow. “Furosemide-induced hepatic and renal tubular necrosis. I. Effects of treatments which alter drug-metabolizing enzymes”. In: *Federation Proceedings*. Vol. 32. 3. FEDERATION AMER SOC EXP BIOL 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998. 1973, p. 305.
- [190] B Kevin Park et al. “Managing the challenge of chemically reactive metabolites in drug development”. In: *Nature Reviews Drug Discovery* 10.4 (2011), pp. 292–306.
- [191] George Ostapowicz et al. “Results of a prospective study of acute liver failure at 17 tertiary care centers in the United States”. In: *Ann. Intern. Med.* 137.12 (2002), pp. 947–954.
- [192] Werner J Pichler, Dean J Naisbitt, and B Kevin Park. “Immune pathomechanism of drug hypersensitivity reactions”. In: *J. Allergy Clin. Immunol.* 127.3 (2011), S74–S81.
- [193] David H Adams et al. “Mechanisms of immune-mediated liver injury”. In: *Toxicol. Sci.* 115.2 (2010), pp. 307–321.
- [194] JA Schwöbel et al. “Prediction of michael-type acceptor reactivity toward glutathione.” In: *Chem. Res. Toxicol.* 23.10 (2010), pp. 1576–1585.
- [195] Terrence J Monks and Douglas C Jones. “The metabolism and toxicity of quinones, quinonimines, quinone methides, and quinone-thioethers”. In: *Curr. Drug Metab.* 3.4 (2002), pp. 425–438.
- [196] David C Dahlin et al. “N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen”. In: *Proc. Natl. Acad. Sci. U. S. A.* 81.5 (1984), pp. 1327–1331.
- [197] Hanem M Awad et al. “Peroxidase-catalyzed formation of quercetin quinone methide–glutathione adducts”. In: *Arch. Biochem. Biophys.* 378.2 (2000), pp. 224–233.

BIBLIOGRAPHY

- [198] Gerard Meunier and B Meunier. "Peroxidase-catalyzed O-demethylation reactions. Quinone-imine formation from 9-methoxyellipticine derivatives." In: *J. Biol. Chem.* 260.19 (1985), pp. 10576–10582.
- [199] David Basketter et al. "The chemistry of contact allergy: why is a molecule allergenic?" In: *Contact Dermatitis* 32.2 (1995), pp. 65–73.
- [200] P Anzenbacher and E Anzenbacherova. "Cytochromes P450 and metabolism of xenobiotics". In: *Cell. Mol. Life Sci.* 58.5-6 (2001), pp. 737–747.
- [201] David Ross and Peter Moldeus. "Generation of reactive species and fate of thiols during peroxidase-catalyzed metabolic activation of aromatic amines and phenols." In: *Environ. Health Perspect.* 64 (1985), p. 253.
- [202] Ping Kang et al. "Bioactivation of lumiracoxib by peroxidases and human liver microsomes: identification of multiple quinone imine intermediates and GSH adducts". In: *Chem. Res. Toxicol.* 22.1 (2008), pp. 106–117.
- [203] Ying Li et al. "In vitro metabolic activation of lumiracoxib in rat and human liver preparations". In: *Drug Metab. Dispos.* 36.2 (2008), pp. 469–473.
- [204] Mingshe Zhu et al. "Detection and structural characterization of glutathione-trapped reactive metabolites using liquid chromatography-high-resolution mass spectrometry and mass defect filtering". In: *Anal. Chem.* 79.21 (2007), pp. 8333–8341.
- [205] Jonathan N Bauman et al. "Comparison of the bioactivation potential of the antidepressant and hepatotoxin nefazodone with aripiprazole, a structural analog and marketed drug". In: *Drug Metab. Dispos.* 36.6 (2008), pp. 1016–1029.
- [206] Amit S Kalgutkar et al. "Bioactivation of the nontricyclic antidepressant nefazodone to a reactive quinone-imine species in human liver microsomes and recombinant cytochrome P450 3A4". In: *Drug Metab. Dispos.* 33.2 (2005), pp. 243–253.
- [207] Vaibhav A Dixit and Shirish Deshpande. "Advances in Computational Prediction of Regioselective and Isoform-Specific Drug Metabolism Catalyzed by CYP450s." In: *ChemistrySelect* 1.20 (2016), pp. 6571–6597.
- [208] Tyler B Hughes et al. "Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network". In: *ACS Cent. Sci.* 2.8 (2016), pp. 529–537.
- [209] Patrik Rydberg, David E Gloriam, and Lars Olsen. "The SMARTCyp cytochrome P450 metabolism prediction server". In: *Bioinformatics* 26.23 (2010), pp. 2988–2989.
- [210] Amy M Sharma et al. "Bioactivation of nevirapine to a reactive quinone methide: implications for liver injury". In: *Chem. Res. Toxicol.* 25.8 (2012), pp. 1708–1719.
- [211] Bo Wen, Yuan Chen, and William L Fitch. "Metabolic activation of nevirapine in human liver microsomes: dehydrogenation and inactivation of cytochrome P450 3A4". In: *Drug Metab. Dispos.* 37.7 (2009), pp. 1557–1562.

BIBLIOGRAPHY

- [212] Joanna Zheng et al. “Screening and identification of GSH-trapped reactive metabolites using hybrid triple quadrupole linear ion trap mass spectrometry”. In: *Chem. Res. Toxicol.* 20.5 (2007), pp. 757–766.
- [213] Jaeick Lee et al. “In vitro and in vivo metabolism of pyronaridine characterized by low-energy collision-induced dissociation mass spectrometry with electrospray ionization”. In: *J. Mass Spectrom.* 39.9 (2004), pp. 1036–1043.
- [214] Emanuela Salvatorelli et al. “The novel anthracenedione, pixantrone, lacks redox activity and inhibits doxorubicinol formation in human myocardium: insight to explain the cardiac safety of pixantrone in doxorubicin-treated patients”. In: *J. Pharmacol. Exp. Ther.* 344.2 (2013), pp. 467–478.
- [215] Jerome P Charba and William H Klein. “Skill in precipitation forecasting in the National Weather Service”. In: *Bull. Am. Meteorol. Soc.* 61.12 (1980), pp. 1546–1555.
- [216] W Yue et al. “Genotoxic metabolites of estradiol in breast: potential mechanism of estradiol induced carcinogenesis”. In: *J. Steroid Biochem. Mol. Biol.* 86.3 (2003), pp. 477–486.
- [217] Mitsuteru Numazawa et al. “Structure–activity relationships of 2 α -substituted androstenedione analogs as aromatase inhibitors and their aromatization reactions”. In: *J. Steroid Biochem. Mol. Biol.* 97.4 (2005), pp. 353–359.
- [218] AS Kalgutkar. “Should the Incorporation of Structural Alerts be Restricted in Drug Design? An Analysis of Structure-Toxicity Trends with Aniline-Based Drugs”. In: *Curr. Med. Chem.* 22.4 (2015), pp. 438–464.
- [219] Corinna Cortes and Mehryar Mohri. “Confidence Intervals for the Area Under the ROC Curve”. In: *Advances in Neural Information Processing Systems*. 2004, pp. 305–312.
- [220] PJ O’Brien. “Molecular mechanisms of quinone cytotoxicity”. In: *Chem.-Biol. Interact.* 80.1 (1991), pp. 1–41.
- [221] Zhong-Ze Fang and Ling Yang. “Characterization of in vitro bioactivation of noscapine in Human Liver Microsomes”. In: *Drug Metab. Rev.* Vol. 42. INFORMA HEALTH-CARE TELEPHONE HOUSE 69-77 PAUL STREET, LONDON EC2A 4LQ, ENGLAND. 2010, pp. 79–79.
- [222] Linning Yu et al. “Oxidation of raloxifene to quinoids: potential toxic pathways via a diquinone methide and o-quinones”. In: *Chem. Res. Toxicol.* 17.7 (2004), pp. 879–888.
- [223] Fraser L Courts and Gary Williamson. “The C-glycosyl flavonoid, aspalathin, is absorbed, methylated and glucuronidated intact in humans”. In: *Mol. Nutr. Food Res.* 53.9 (2009), pp. 1104–1111.

BIBLIOGRAPHY

- [224] Markus R Meyer et al. “Studies on the metabolism of the α -pyrrolidinophenone designer drug methylenedioxy-pyrovalerone (MDPV) in rat and human urine and human liver microsomes using GC–MS and LC–high-resolution MS and its detectability in urine by GC–MS”. In: *J. Mass Spectrom.* 45.12 (2010), pp. 1426–1442.
- [225] Xiaohai Li et al. “Characterization of dasatinib and its structural analogs as CYP3A4 mechanism-based inactivators and the proposed bioactivation pathways”. In: *Drug Metab. Dispos.* 37.6 (2009), pp. 1242–1250.
- [226] Bo Wen and Mingyan Zhou. “Metabolic activation of the phenothiazine antipsychotics chlorpromazine and thioridazine to electrophilic iminoquinone species in human liver microsomes and recombinant P450s”. In: *Chem.-Biol. Interact.* 181.2 (2009), pp. 220–226.
- [227] G Lin et al. “Metabolism of piperidine-type phenothiazine antipsychotic agents. IV. Thioridazine in dog, man and rat”. In: *Xenobiotica* 23.10 (1993), pp. 1059–1074.
- [228] Mukesh K Mahajan et al. “In vitro metabolism of oxymetazoline: evidence for bioactivation to a reactive metabolite”. In: *Drug Metab. Dispos.* 39.4 (2011), pp. 693–702.
- [229] Mitsuteru Numazawa and Akiko Yoshimura. “Biological aromatization of Δ 4, 6-and Δ 1, 4, 6-androgens and their 6-alkyl analogs, potent inhibitors of aromatase”. In: *J. Steroid Biochem. Mol. Biol.* 70.4 (1999), pp. 189–196.
- [230] P Favetta et al. “High-performance liquid chromatographic assay to detect hydroxylate and conjugate metabolites of propofol in human urine”. In: *J. Chromatogr. B: Biomed. Sci. Appl.* 742.1 (2000), pp. 25–35.
- [231] HJCF Nelis and AP De Leenheer. “Evidence for metabolic inertness of doxycycline”. In: *J. Pharm. Sci.* 70.2 (1981), pp. 226–228.
- [232] Johan Bylund et al. “Novel bioactivation mechanism of reactive metabolite formation from phenyl methyl-isoxazoles”. In: *Drug Metab. Dispos.* 40.11 (2012), pp. 2185–2191.
- [233] William L Fitch et al. “Application of modern drug metabolism structure determination tools and assays to the in vitro metabolism of imiloxan”. In: *Drug Metab. Lett.* 4.2 (2010), pp. 77–87.
- [234] J Ni et al. “Characterization of benzimidazole and other oxidative and conjugative metabolites of brimonidine in vitro and in rats in vivo using on-line H/D exchange LC-MS/MS and stable-isotope tracer techniques”. In: *Xenobiotica* 37.2 (2007), pp. 205–220.
- [235] David S Wishart et al. “DrugBank: a knowledgebase for drugs, drug actions and drug targets”. In: *Nucleic Acids Res.* 36.suppl 1 (2008), pp. D901–D906.
- [236] Liva Ralaivola et al. “Graph kernels for chemical informatics”. In: *Neural Networks* 18.8 (2005), pp. 1093–1110.

BIBLIOGRAPHY

- [237] John CL Erve et al. “Characterization of glutathione conjugates of the remoxipride hydroquinone metabolite NCQ-344 formed in vitro and detection following oxidation by human neutrophils”. In: *Chem. Res. Toxicol.* 17.4 (2004), pp. 564–571.
- [238] Fengping Li et al. “In vitro nimesulide studies toward understanding idiosyncratic hepatotoxicity: diiminoquinone formation and conjugation”. In: *Chem. Res. Toxicol.* 22.1 (2008), pp. 72–80.
- [239] G Hopmann and T Surmann. “Cholestatic jaundice during flecainide therapy”. In: *Dtsch. Med. Wochenschr.* 109.48 (1984), p. 1863.
- [240] RL McQuinn et al. “Biotransformation and elimination of ¹⁴C-flecainide acetate in humans.” In: *Drug Metab. Dispos.* 12.4 (1984), pp. 414–420.
- [241] Emmanuel Andrès et al. “Recognition and management of drug-induced blood cytopenias: the example of drug-induced acute neutropenia and agranulocytosis”. In: *Expert Opin. Drug Saf.* 7.4 (2008), pp. 481–489.
- [242] IR Edwards et al. “Quality criteria for early signals of possible adverse drug reactions”. In: *The Lancet* 336.8708 (1990), pp. 156–158.
- [243] J Stevenson and AC Kennedy. “A fatal case of agranulocytosis due to thenalidine tartrate complicated by acute renal failure and mycelial abscesses of brain.” In: *Scott. Med. J.* 6.11 (1961), pp. 522–5.
- [244] Per Arneborn and Jan Palmblad. “Drug-Induced Neutropenia—A Survey for Stockholm 1973–1978”. In: *Acta Med. Scand.* 212.5 (1982), pp. 289–292.
- [245] Myles DB Stephens. *The dawn of drug safety*. George Mann, 2010.
- [246] Donald A Adams and Seymour Perry. “Agranulocytosis associated with thenalidine (sandostene) tartrate therapy: report of three cases”. In: *JAMA, J. Am. Med. Assoc.* 167.10 (1958), pp. 1207–1210.
- [247] B Zolov. “Agranulocytosis resulting from sandostene.” In: *The Journal of the Maine Medical Association* 49.9 (1958), pp. 335–337.
- [248] Benson Ninan and Albert I Wertheimer. “Withdrawing Drugs in the US Versus Other Countries”. In: *Innovations in Pharmacy* 3.3 (2012), p. 6.
- [249] Darja Gramec, Lucija Peterlin Mašič, and Marija Sollner Dolenc. “Bioactivation potential of thiophene-containing drugs”. In: *Chem. Res. Toxicol.* 27.8 (2014), pp. 1344–1358.
- [250] Hilmi Orhan. “Extrahepatic targets and cellular reactivity of drug metabolites”. In: *Curr. Med. Chem.* 22.4 (2015), pp. 408–437.
- [251] George E Schreiner. “Nephrotoxicity and Diagnostic Agents”. In: *JAMA, J. Am. Med. Assoc.* 196.5 (1966), pp. 413–415.

BIBLIOGRAPHY

- [252] Frederick C Churchill et al. "Amodiaquine as a prodrug: importance of metabolite (s) in the antimalarial effect of amodiaquine in humans". In: *Life Sci.* 36.1 (1985), pp. 53–62.
- [253] JL Maggs et al. "Drug-protein conjugates—XIV: mechanisms of formation of protein-arylated intermediates from amodiaquine, a myelotoxin and hepatotoxin in man". In: *Biochem. Pharmacol.* 37.2 (1988), pp. 303–311.
- [254] GAMBK Christie, AM Breckenridge, and BK Park. "Drug-protein conjugates—XVIII: Detection of antibodies towards the antimalarial amodiaquine and its quinone imine metabolite in man and the rat". In: *Biochem. Pharmacol.* 38.9 (1989), pp. 1451–1458.
- [255] Paul M O'Neill et al. "The effect of fluorine substitution on the metabolism and anti-malarial activity of amodiaquine". In: *J. Med. Chem.* 37.9 (1994), pp. 1362–1370.
- [256] Paul M O'Neill et al. "Synthesis, antimalarial activity, and preclinical pharmacology of a novel series of 4-fluoro and 4-chloro analogues of amodiaquine. Identification of a suitable "back-up" compound for N-tert-butyl isoquine". In: *J. Med. Chem.* 52.7 (2009), pp. 1828–1844.
- [257] Terrence J Monks et al. "Quinone chemistry and toxicity". In: *Toxicol. Appl. Pharmacol.* 112.1 (1992), pp. 2–16.
- [258] Eileen EN Almario et al. "Drug-Induced Liver Injury". In: *BioMed Research International* 2017 (2017).
- [259] Neil Kaplowitz. "Avoiding idiosyncratic DILI: two is better than one". In: *Hepatology* 58.1 (2013), pp. 15–17.
- [260] Xiaomin Deng et al. "Inflammatory stress and idiosyncratic hepatotoxicity: hints from animal models". In: *Pharmacological reviews* 61.3 (2009), pp. 262–282.
- [261] Graham F Smith. "Designing drugs to avoid toxicity". In: *Progress in medicinal chemistry*. Vol. 50. Elsevier, 2011, pp. 1–47.
- [262] DP Williams et al. "Are chemically reactive metabolites responsible for adverse reactions to drugs?" In: *Curr. Drug Metab.* 3.4 (2002), pp. 351–366.
- [263] J Gallagher et al. "Formation of DNA adducts in rat lung following chronic inhalation of diesel emissions, carbon black and titanium dioxide particles". In: *Carcinogenesis* 15.7 (1994), pp. 1291–1299.
- [264] Hester van der Woude et al. "Formation of transient covalent protein and DNA adducts by quercetin in cells with and without oxidative enzyme activity". In: *Chem. Res. Toxicol.* 18.12 (2005), pp. 1907–1916.
- [265] JH Lillibridge et al. "Protein-reactive metabolites of carbamazepine in mouse liver microsomes." In: *Drug Metab. Dispos.* 24.5 (1996), pp. 509–514.

BIBLIOGRAPHY

- [266] Hideo Takakusa et al. “Covalent binding and tissue distribution/retention assessment of drugs associated with idiosyncratic drug toxicity”. In: *Drug Metab. Dispos.* 36.9 (2008), pp. 1770–1779.
- [267] David C Evans et al. “Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development”. In: *Chem. Res. Toxicol.* 17.1 (2004), pp. 3–16.
- [268] Vandana Purohit and Ashis K Basu. “Mutagenicity of nitroaromatic compounds”. In: *Chem. Res. Toxicol.* 13.8 (2000), pp. 673–692.
- [269] Keisuke Ikehata et al. “Protein targets of reactive metabolites of thiobenzamide in rat liver in vivo”. In: *Chem. Res. Toxicol.* 21.7 (2008), pp. 1432–1442.
- [270] Zhang-Xu Liu and Neil Kaplowitz. “Immune-mediated drug-induced liver disease”. In: *Clinics in liver disease* 6.3 (2002), pp. 755–774.
- [271] Dean J Naisbitt et al. “Reactive metabolites and their role in drug reactions”. In: *Current opinion in allergy and clinical immunology* 1.4 (2001), pp. 317–325.
- [272] Sabine Mazerbourg et al. “Reprofiling of troglitazone towards more active and less toxic derivatives: a new hope for cancer treatment?” In: *Current topics in medicinal chemistry* 16.19 (2016), pp. 2115–2124.
- [273] Jack Uetrecht. “Screening for the potential of a drug candidate to cause idiosyncratic drug reactions”. In: *Drug Discovery Today* 8.18 (2003), pp. 832–837.
- [274] SH Day et al. “A semi-automated method for measuring the potential for protein covalent binding in drug discovery”. In: *J. Pharmacol. Toxicol. Methods* 52.2 (2005), pp. 278–285.
- [275] Iurii Sushko et al. “ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions”. In: *J. Chem. Inf. Model.* 52.8 (2012), pp. 2310–2316.
- [276] P J Edwards and C Sturino. “Managing the liabilities arising from structural alerts: a safe philosophy for medicinal chemists”. In: *Current medicinal chemistry* 18.20 (2011), pp. 3116–3135.
- [277] Suresh B Singh et al. “A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules”. In: *Journal of medicinal chemistry* 46.8 (2003), pp. 1330–1336.
- [278] Jianlong Peng et al. “In silico site of metabolism prediction for human UGT-catalyzed reactions”. In: *Bioinformatics* 30.3 (2014), pp. 398–405.
- [279] Tyler B Hughes and S Joshua Swamidass. “Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism”. In: *Chem. Res. Toxicol.* 30.2 (2017), pp. 642–656.

BIBLIOGRAPHY

- [280] Ronald L Melnick and Michael C Kohn. “Mechanistic data indicate that 1, 3-butadiene is a human carcinogen”. In: *Carcinogenesis* 16.2 (1995), pp. 157–163.
- [281] JA Bond et al. “Research strategy for assessing target tissue dosimetry of 1, 3-butadiene in laboratory animals and humans.” In: *IARC Sci. Publ.* 127 (1993), pp. 45–55.
- [282] U.S. Environmental Protection Agency. *Health assessment of 1,3-butadiene*. Tech. rep. National Center for Environmental Assessment, Washington, DC: EPA/600/P-98/001F, 2002.
- [283] György A Csanády, F Peter Guengerich, and James A Bond. “Comparison of the biotransformation of 1, 3-butadiene and its metabolite, butadiene monoepoxide, by hepatic and pulmonary tissues from humans, rats and mice”. In: *Carcinogenesis* 13.7 (1992), pp. 1143–1153.
- [284] Renee J Duescher and Adnan A Elfarrar. “Human liver microsomes are efficient catalysts of 1, 3-butadiene oxidation: evidence for major roles by cytochromes P450 2A6 and 2E1”. In: *Arch. Biochem. Biophys.* 311.2 (1994), pp. 342–349.
- [285] RJ Duescher and AA Elfarrar. “1, 3-Butadiene oxidation by human myeloperoxidase. Role of chloride ion in catalysis of divergent pathways.” In: *J. Biol. Chem.* 267.28 (1992), pp. 19859–19865.
- [286] Renee J Krause, Jane E Sharer, and Adnan A Elfarrar. “Epoxide hydrolase-dependent metabolism of butadiene monoxide to 3-butene-1, 2-diol in mouse, rat, and human liver”. In: *Drug Metab. Dispos.* 25.8 (1997), pp. 1013–1015.
- [287] Renee J Krause, Raymond A Kemper, and Adnan A Elfarrar. “Hydroxymethylvinyl ketone: a reactive Michael acceptor formed by the oxidation of 3-butene-1, 2-diol by cDNA-expressed human cytochrome P450s and mouse, rat, and human liver microsomes”. In: *Chem. Res. Toxicol.* 14.12 (2001), pp. 1590–1595.
- [288] Jiajia Meng et al. “RD-Metabolizer: an integrated and reaction types extensive approach to predict metabolic sites and metabolites of drug-like molecules”. In: *Chem. Cent. J.* 11.1 (2017), p. 65.
- [289] Fumiyo Kasuya, Kazuo Igarashi, and Miyoshi Fukui. “Metabolism of benoxinate in humans”. In: *J. Pharm. Sci.* 76.4 (1987), pp. 303–305.
- [290] Robert Meatherall. “Benzodiazepine screening using EMIT II® and TDx®: urine hydrolysis pretreatment required”. In: *J. Anal. Toxicol.* 18.7 (1994), pp. 385–390.
- [291] T Shiraga et al. “Identification of cytochrome P450 enzymes involved in the metabolism of zotepine, an antipsychotic drug, in human liver microsomes”. In: *Xenobiotica* 29.3 (1999), pp. 217–229.

BIBLIOGRAPHY

- [292] JB Knaak et al. “Development of in Vitro V_{max} and K_m Values for the Metabolism of Isofenphos by P-450 Liver Enzymes in Animals and Human”. In: *Toxicol. Appl. Pharmacol.* 120.1 (1993), pp. 106–113.
- [293] Willem Meuldermans et al. “The metabolism and excretion of risperidone after oral administration in rats and dogs.” In: *Drug Metab. Dispos.* 22.1 (1994), pp. 129–138.
- [294] Christopher A Reilly et al. “Reactive intermediates produced from the metabolism of the vanilloid ring of capsaicinoids by P450 enzymes”. In: *Chemical research in toxicology* 26.1 (2012), pp. 55–66.
- [295] Paul Matthew O’Byrne et al. “Synthesis, screening and pharmacokinetic evaluation of potential prodrugs of bupropion. Part one: in vitro development”. In: *Pharmaceuticals* 7.5 (2014), pp. 595–620.
- [296] Ana Ballesteros-Gómez et al. “In vitro metabolism of 2-ethylhexyldiphenyl phosphate (EHDPPH) by human liver microsomes”. In: *Toxicol. Lett.* 232.1 (2015), pp. 203–212.
- [297] Lukas Schmidt, Vladimir N Belov, and Thomas Göen. “Human metabolism of Δ 3-carene and renal elimination of Δ 3-carene-10-carboxylic acid (chaminic acid) after oral administration”. In: *Arch. Toxicol.* 89.3 (2015), pp. 381–392.
- [298] Feng Li et al. “Metabolomic analysis reveals novel isoniazid metabolites and hydrazones in human urine”. In: *Drug Metab. Pharmacokinet.* 26.6 (2011), pp. 569–576.
- [299] Lygia Azevedo Marques et al. “Production and on-line acetylcholinesterase bioactivity profiling of chemical and biological degradation products of tacrine”. In: *J. Pharm. Biomed. Anal.* 53.3 (2010), pp. 609–616.
- [300] Akio Nakamura. “Cytochrome P450 isoforms responsible for the N-deethylation and cyclohexane-hydroxylation of NS-21”. In: *Xenobiotica* 29.3 (1999), pp. 243–252.
- [301] Rui Xiong, David Siegel, and David Ross. “Quinone-induced protein handling changes: implications for major protein handling systems in quinone-mediated toxicity”. In: *Toxicology and applied pharmacology* 280.2 (2014), pp. 285–295.
- [302] Yui Yamamoto et al. “Formation of a novel quinone epoxide metabolite of troglitazone with cytotoxic to HepG2 cells”. In: *Drug Metab. Dispos.* 30.2 (2002), pp. 155–160.
- [303] V Spaldin et al. “The effect of enzyme inhibition on the metabolism and activation of tacrine by human liver microsomes.” In: *Br. J. Clin. Pharmacol.* 38.1 (1994), pp. 15–22.
- [304] S Madden et al. “Species variation in the bioactivation of tacrine by hepatic microsomes”. In: *Xenobiotica* 25.1 (1995), pp. 103–116.
- [305] Bo Wen and William L Fitch. “Screening and characterization of reactive metabolites using glutathione ethyl ester in combination with Q-trap mass spectrometry”. In: *J. Mass Spectrom.* 44.1 (2009), pp. 90–100.

BIBLIOGRAPHY

- [306] Sang Yoon Choi et al. “Benzylamide derivative compound attenuates the ultraviolet B-induced hyperpigmentation in the brownish guinea pig skin”. In: *Biochem. Pharmacol.* 67.4 (2004), pp. 707–715.
- [307] Kevin J Wells-Knecht et al. “2, 7-Disubstituted-pyrrolotriazine kinase inhibitors with an unusually high degree of reactive metabolite formation”. In: *Chem. Res. Toxicol.* 24.11 (2011), pp. 1994–2003.
- [308] Na Le Dang et al. “A Computational Approach to Structural Alerts: Furans, Phenols, Nitroaromatics, and Thiophenes”. In: *Chem. Res. Toxicol.* DOI: 10.1021/acs.chemrestox.6b00336 (2017).
- [309] J Ramanarayanan and GS Krishnan. “hepatotoxicity and EGFR inhibition.” In: *Clinical advances in hematology & oncology: H&O* 6.3 (2008), pp. 200–201.
- [310] Hans-Peter Gschwind et al. “Metabolism and disposition of imatinib mesylate in healthy volunteers”. In: *Drug Metab. Dispos.* 33.10 (2005), pp. 1503–1512.
- [311] Christine M Baumgartner et al. “Cardiovascular effects of dipyron and propofol on hemodynamic function in rabbits”. In: *Am. J. Vet. Res.* 70.11 (2009), pp. 1407–1415.
- [312] Stephan A Schug and Andreas Manopas. “Update on the role of non-opioids for post-operative pain treatment”. In: *Best Pract. Res., Clin. Anaesthesiol.* 21.1 (2007), pp. 15–30.
- [313] Julia C Wessel et al. “Characterization of oxalic acid derivatives as new metabolites of metamizol (dipyron) in incubated hen’s egg and human”. In: *Eur. J. Pharm. Sci.* 28.1-2 (2006), pp. 15–25.
- [314] GW Basak, J Drozd-Sokołowska, and W Wiktor-Jedrzejczak. “Update on the incidence of metamizole sodium-induced blood dyscrasias in Poland”. In: *J. Int. Med. Res.* 38.4 (2010), pp. 1374–1380.
- [315] A Jasińska, T Maślanka, and JJ Jaroszewski. “Pharmacological characteristics of metamizole”. In: *Pol. J. Vet. Sci.* 17.1 (2014), pp. 207–214.
- [316] JM García-Martínez et al. “Effect of metamizol on promyelocytic and terminally differentiated granulocytic cells. Comparative analysis with acetylsalicylic acid and diclofenac.” In: *Biochem. Pharmacol.* 65.2 (2003), pp. 209–217.
- [317] Jack P Uetrecht et al. “Oxidation of aminopyrine by hypochlorite to a reactive dication: possible implications for aminopyrine-induced agranulocytosis”. In: *Chem. Res. Toxicol.* 8.2 (1995), pp. 226–233.
- [318] Noah Scheinfeld et al. “Acute generalized exanthematous pustulosis resembling toxic epidermal necrolysis caused by famotidine”. In: *ACTA DERMATOVENEREOLÓGICA-STOCKHOLM-* 83.1 (2003), pp. 76–76.

BIBLIOGRAPHY

- [319] Na Le Dang et al. “The Metabolic Rainbow: Deep Learning Phase 1 Metabolism in Five Colors”. In: *Under Review* (2018).
- [320] Paul R Ortiz de Montellano. “Cytochrome P450-activated prodrugs”. In: *Future Med. Chem.* 5.2 (2013), pp. 213–228.
- [321] Abdul W Basit and Larry F Lacey. “Colonic metabolism of ranitidine: implications for its delivery and absorption”. In: *Int. J. Pharm* 227.1-2 (2001), pp. 157–165.
- [322] Laura Guasch, Markus Sitzmann, and Marc C Nicklaus. “Enumeration of Ring–Chain Tautomers Based on SMIRKS Rules”. In: *J. Chem. Inf. Model.* 54.9 (2014), pp. 2423–2432.
- [323] Wei Deng, Gisbert Schneider, and W Venus So. “Mapping chemical structures to Markush structures using SMIRKS”. In: *Mol. Inf.* 30.8 (2011), pp. 665–671.
- [324] David Weininger. “SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules”. In: *J. Chem. Inf. Model.* 28.1 (1988), pp. 31–36.
- [325] Anastasia V Rudik et al. “MetaTox: Web Application for Predicting Structure and Toxicity of Xenobiotics’ Metabolites”. In: *J. Chem. Inf. Model.* (2017).
- [326] Jason Lazarou, Bruce H Pomeranz, and Paul N Corey. “Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies”. In: *JAMA, J. Am. Med. Assoc.* 279.15 (1998), pp. 1200–1205.
- [327] U.S. Food and Drug Administration. “Preventable Adverse Drug Reactions: A Focus on Drug Interactions”. <https://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources>
- [328] Neil Kaplowitz. “Drug-induced liver disorders”. In: *Drug safety* 24.7 (2001), pp. 483–490.
- [329] Robert J Temple and Martin H Himmel. “Safety of newly approved drugs: implications for prescribing”. In: *JAMA, J. Am. Med. Assoc.* 287.17 (2002), pp. 2273–2275.
- [330] Mark W Russo et al. “Liver transplantation for acute liver failure from drug induced liver injury in the United States”. In: *Liver Transplantation* 10.8 (2004), pp. 1018–1023.
- [331] Masahiro Tohkin et al. “Prediction of severe adverse drug reactions using pharmacogenetic biomarkers”. In: *Drug metabolism and pharmacokinetics* 25.2 (2010), pp. 122–133.
- [332] Cassie Frank et al. “Era of faster FDA drug approval has also seen increased black-box warnings and market withdrawals”. In: *Health Affairs* 33.8 (2014), pp. 1453–1459.
- [333] Shannan Tujios and Robert J Fontana. “Mechanisms of drug-induced liver injury: from bedside to bench”. In: *Nature reviews Gastroenterology & hepatology* 8.4 (2011), p. 202.

BIBLIOGRAPHY

- [334] Luis J Núñez-Vergara et al. “Electrochemical, UV-Visible and EPR studies on nitrofurantoin: Nitro radical anion generation and its interaction with glutathione”. In: *Free radical research* 32.5 (2000), pp. 399–409.
- [335] Elizabeth M Joshi et al. “In vitro metabolism of 2-acetylbenzothiophene: relevance to zileuton hepatotoxicity”. In: *Chem. Res. Toxicol.* 17.2 (2004), pp. 137–143.
- [336] Noriko Masubuchi, Chie Makino, and Nobuyuki Murayama. “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”. In: *Chem. Res. Toxicol.* 20.3 (2007), pp. 455–464.
- [337] PETER J WIRTH, CAROL J BETTIS, and WENDEL L NELSON. “Microsomal metabolism of furosemide evidence for the nature of the reactive intermediate involved in covalent binding”. In: *Mol. Pharmacol.* 12.5 (1976), pp. 759–768.
- [338] Qiang Qu et al. “Nrf2 protects against furosemide-induced hepatotoxicity”. In: *Toxicology* 324 (2014), pp. 35–42.
- [339] Randolph J McMurtry and Jerry R Mitchell. “Renal and hepatic necrosis after metabolic activation of 2-substituted furans and thiophenes, including furosemide and cephaloridine”. In: *Toxicology and applied pharmacology* 42.2 (1977), pp. 285–300.
- [340] Tyler B Hughes et al. “The Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites”. In: *Under Review* (2018).
- [341] Matthew K Matlock et al. “Modeling Small-Molecule Reactivity Identifies Promiscuous Bioactive Compounds”. In: *J. Chem. Inf. Model.* 58.8 (2018), pp. 1483–1500.
- [342] Na Le Dang et al. “Computationally Assessing the Bioactivation of Drugs by N-Dealkylation”. In: *Chem. Res. Toxicol.* 31.2 (2018), pp. 68–80.
- [343] Rajeev Kumar and Abhaya Indrayan. “Receiver operating characteristic (ROC) curve for medical researchers”. In: *Indian pediatrics* 48.4 (2011), pp. 277–287.
- [344] Ellen Blatchford. “Studies of anti-emetic drugs: A comparative study of cyclizine (Marzine®), pipamazine (Mornidine®), trimethobenzamide (Tigan®), and hyoscine”. In: *Canadian Anaesthetists’ Society Journal* 8.2 (1961), pp. 159–165.