Arts & Sciences Electronic Theses and Dissertations

Arts & Sciences

Spring 5-15-2021

# Modeling Small Molecule Metabolism in Human Liver Microsome to Better Predict Toxicity Risk

Na Le Dang
*Washington University in St. Louis*

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

Part of the Chemistry Commons

# Washington University in St. Louis

Division of Biology and Biomedical Sciences
Computational System Biology

Dissertation Examination Committee:
S. Joshua Swamidass, Chair
Michael Brent, Co-Chair
Mark Anastasio
Greg Bowman
Obi Griffith
Kristen Naegle

Modeling Small Molecule Metabolism in Human Liver Microsome to Better Predict Toxicity Risk
by
Na Le Dang

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

May 2021
Saint Louis, Missouri

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

I am grateful to so many people for their constant support and encouragement throughout my higher education journey in the U.S. I must first thank my undergraduate mentors who introduced me to research and motivated me to apply M.D./Ph.D. programs. Without them, I would not have discovered a career path that is as both challenged and rewarding as a medical scientist.

Thank you to my graduate mentor, Dr. S. Joshua Swamidass, who always believes in me and encourages me to do my best to grow professionally and personally. I am always in awe of the depth and breath of your knowledge and your unbound wealth of ideas. I am a better scientist and communicator because you have taken all the time to teach me how to ask scientific questions, conduct experiments to find out the answers, and practice presenting my research. I am in debt for all of your insights into career development and for your advocacy of women in science. Because of your trust, I had the flexibility to fulfill my roles both as a graduate student and a mother. I would also like to thank the members of my thesis committee: Drs. Mark Anastasio, Michael Brent, Greg Bowman, Obi Griffith, and Kristen Naegle. I value their scientific critique, advice, and guidance throughout the years.

My training would not have been the same without the members of the Swamidass Lab during my time. A special thank you to Tyler Hughes and Matthew Matlock for their scientific capacity, inspiring hard-work, and sincere friendship. I cannot imagine anyone better than you guys to

shared the past 4 years+ of graduate school. I have learned so much from you both inside and outside of lab and will carry those lessons with me through the rest of my career. To other current and past members of the Swamidass lab: Jed Zaretzki, Michael Ward, Jon Marsh, Jack Elliot-Higgins, Noah Flint, Rohit Farmer, and Argha Datta, thank you for the scientific and career development insights and friendship.

To my wonderful MSTP classmates, I hope that the next stage of my life brings me friends and colleagues as amazing as you all. Thank you for all the treasured moments we shared.

Last but not least, thank you my husband, Dang Du. Thank you for having been my listening ear, my stoic shoulder, and my root. Without you, none of this is possible .

<div align="right">

Na Le Dang

Washington University, May 2021

</div>

# Abstract of the dissertation

Modeling Small Molecule Metabolism in Human Liver Microsome to Better Predict Toxicity Risk

by

Na Le Dang

Doctor of Philosophy in Computational System Biology

Division of Biology and Biomedical Sciences

Washington University in St. Louis 2021

Thesis advisor: S. Joshua Swamidass, Chair

Adverse drug reactions (ADRs) are a serious problem with increasing morbidity, mortality, and health care costs worldwide. In the U.S., ADRs are responsible for more than 50% of acute liver failure cases and are the fourth most common cause of death, costing 100,000 lives annually.

Idiosyncratic adverse drug reactions (IADRs) are immune-mediated hypersensitivity ADRs that are difficult to foresee during drug development. IADRs are often caused by reactive metabolites produced during drug metabolism. These reactive metabolites covalently attach to cellular components, and the resulting conjugates may provoke toxic immune response. Because reactive metabolites are short-lived, they can be difficult to detect. Tools to reliably predict whether a compound forms reactive metabolites would enable us to avoid drug candidates prone to causing IADRs and make new medicines safer. Unfortunately, due to inadequate modeling of metabolism, current experimental and computational approaches do not reliably identify drug candidates that form reactive metabolites.

Bioactivation pathways leading to reactive metabolite formations often are composed of multiple steps. To accurately predict reactive metabolite formation, we must explicitly model metabolic steps of bioactivation pathways. Therefore, we built models to predict specific metabolic transformations such as hydroxylation, epoxidation, dehydrogenation, quinonation, hydrolysis, reduction, glucuronidation, sulfuration, acetylation, and methylation. Using machine learning and literature-derived data, we trained models that can predict both the likelihood that a

molecules undergoes a certain chemical transformation and the specific site(s) within the molecule where this transformation happens. Together, our metabolism models cover $\sim 95\%$ of enzymatically-driven chemical reactions in human. Our models achieve high area under the receiver operating characteristic curve scores (AUCs) of $\sim 90\%$ in cross-validated tests.

Our mechanistic approach outperformed structural alerts—a common tool used to screen out candidate compounds during drug development. Structural alerts are chemical moieties that were frequently observed to give rise to reactive metabolite upon bioactivation. However, many safe drugs also contain structural alerts which are not bioactivated and, conversely, many toxic drugs contain no structural alert. We combined models of metabolism, metabolite structure prediction, and reactivity to offer a better prediction of reactive metabolite formation in the context of structural alerts. Based on the known bioactivation pathway(s) of each structural alert, appropriate metabolism models were applied to evaluate whether drugs containing the structural alert actually form reactive metabolites. Our study focused on the furan, phenol, nitroaromatic, and thiophene alerts. Specifically, we used models of epoxidation, quinone formation, reduction, and sulfur-oxidation to predict the bioactivation of furan-, phenol-, nitroaromatic-, and thiophene-containing drugs. Our models separated bioactivated and not-bioactivated furan-, phenol-, nitroaromatic-, and thiophene-containing drugs with AUC performances of $100\%$, $73\%$, $93\%$, and $88\%$, respectively. In addition, we used our models to uncover bioactivation mechanisms that were previously under-appreciated. For example, N-dealkylation is the oxidation of an alkylated amine at the nitrogen-carbon bond, cleaving the parent compound into an amine and an aldehyde. Even though aldehydes can be toxic, metabolic studies usually neglect to report or investigate them because they are assumed to be efficiently detoxified into carboxylic acids and alcohols. Applying the N-dealkylation model to approved and withdrawn medicines, we found that aldehyde metabolites produced from N-dealkylation may explain the hepatotoxicity of several drugs: indinavir, piperacillin, verapamil, and ziprasidone. These results demonstrated the utility of comprehensive bioactivation models that systematically consider constituent metabolic steps in gauging toxicity risks.

# 1

# Introduction

## 1.1 Idiosyncratic adverse drug reactions.

Millions of Americans take medications on a daily basis to prevent and treat a variety of health conditions. Breakthroughs in drug discovery and development have improved and saved millions of lives [2]. However, medications can also cause harm. Adverse drug reactions (ADRs) have a major impact on public health [2]. They are responsible for more than 50% of acute liver failure cases [135] and 280, 000 hospital admissions [244]. ADRs are the fourth most frequent cause of death in the U.S., costing 100,000 lives annually [132, 134].

Idiosyncratic adverse drug reactions (IADRs), also known as type B ADRs [71], are immune-mediated hypersensitivity reactions that can affect virtually any organ. The skin, liver,

1

and blood cells are the most common sites. Major clinical manifestations of IADRs include skin rashes (e.g. urticaria, maculopapular eruption, Stevens-Johnson syndrome, and toxic epidermal necrolysis), hematologic adverse reactions (e.g. agranulocytosis, thrombocytopenia, and hemolytic or aplastic anemia), drug-induced autoimmunity (e.g. drug-induced Lupus-like syndrome), and liver injury (e.g. hepatocellular and cholestatic liver injury) [237, 239].

IADRs are a leading cause of acute liver failure, affecting $2,000$ patients per year in the U.S. and accounting for $25\%$ of cases in intensive care units [5, 188]. Also known as fulminant hepatic failure, acute liver failure is characterized by severe and sudden liver cell dysfunction leading to coagulopathy and hepatic encephalopathy in previously healthy persons with no known underlying liver disease. Without liver transplant, this illness can rapidly progress to cerebral edema and multi-organ failure and subsequently to coma and death. IADRs account for more than $17\%$ of liver transplant cases [188].

IADRs are the most difficult type of ADRs to predict and prevent. Due to the relatively low incidence rate, the lack of valid animal models, and limited knowledge of the underlying mechanism, IADRs are usually not detected during clinical trials and serious IADRs commonly emerge only after Food and Drug Administration (FDA) approval. One analysis found that approximately one out of five drugs were either withdrawn or acquired new black box warnings for IADRs after having been on the market for two to seven years [132].

IADRs are the most serious and expensive problem in drug development. They are the main reason for drug clinical trial termination and drug withdrawal [245]. Nevertheless, the majority of drugs are not associated with IADRs, offering hope that new medicines could avoid these adverse reactions if reliable predictors existed [226]. Accurate prediction and identification of molecules prone to IADRs would revolutionize drug development by screening out IADR-prone candidates early, before significant resource investment and exposure to patients.
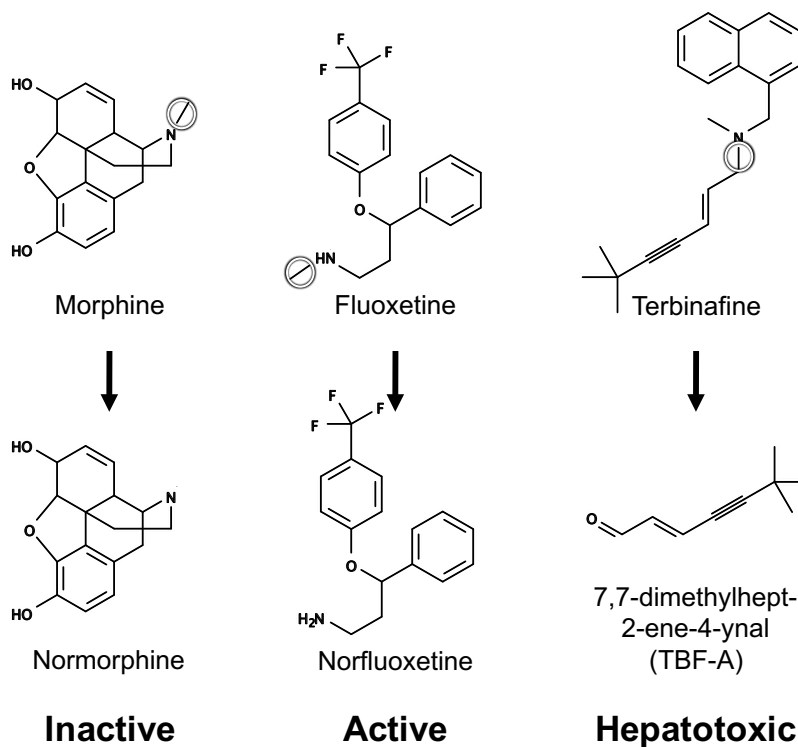
**Figure 1.2.1:** Three drugs exemplify three different outcomes of metabolism. The site of metabolism is circled in the parent molecules. (**A**) In the majority of cases, metabolism can lead to detoxification. For example, the metabolism of the analgesic morphine via N-dealkylation produces normorphine, a much less potent and more easily excreted metabolite than the parent compound. (**B**) Metabolism can also lead to bioactivation. For example, N-dealkylation of fluoxetine, an anti-depressive drug, generates norfluoxetine, the active metabolite responsible for the majority of fluoxetine's therapeutic effects. (**C**) Metabolism can sometimes lead to toxification. For instance, N-dealkylation of terbinafine gives rise to 7,7-dimethylhept-2-ene-4-ynal (TBF-A), a reactive metabolite that can conjugate to proteins and lead to toxicity.

## 1.2   Drug Metabolism.

One of the liver's major physiological roles is to biotransform lipophilic xenobiotics into hydrophilic metabolites that can be more readily excreted (Figure 1.2.1A). Liver enzymes like the cytochromes P450 and uridine diphosphate glucunosyltransferases (UGTs) catalyze phase I and phase II biotransformation (Table 1.2.1) of a remarkably diverse range of xenobiotics, including therapeutic agents [85]. Such compounds are concentrated in the liver due to hepatic circulation and active transport systems [245]. The liver is the most important organ in drug metabolism and is also the first to suffer damage from drug toxicity [135, 190].

Although the major role of drug metabolism is detoxification, it can also act as a toxification process (Figure 1.2.1C). A number of enzymes, especially the cytochromes P450, can generate, and in many instances, release reactive metabolites — electrophilic species that can bind covalently to macromolecules such as proteins and DNA. The diversity of cytochromes P450, together with the reactivity of their oxygen intermediates, enables them to bioactivate even relatively inert substrates, leading to the direct formation of chemically diverse reactive metabolites [85].

There is substantial evidence that most IADRs are caused by drugs that are not intrinsically toxic, but become harmful after being metabolically transformed into reactive metabolites [29, 78, 166, 167, 191]. Approximately 62% of drugs withdrawn from the market for IADRs, or with black box warning for IADRs, have been shown to produce reactive metabolites [243]. Reactive metabolites are a cause for concern because they can involve in covalent interactions with critical macromolecules such as DNA and proteins, and/or nonconvalent interactions causing oxidative and other intracellular stress [86, 113, 177, 231].

Reactive metabolites comprise hard and soft electrophiles. Hard electrophiles include alkyl/benzylic carbonium ions, iminium ions, and glyoxals; while soft electrophiles are typically $\alpha, \beta$-unsaturated ketones, quinones, and acrylamides [86]. Hard and soft electrophiles form covalent adducts with hard and soft nucleophiles in DNA and proteins. DNA conjugation can

4

lead to cancer [25, 166, 167, 199]. Protein conjugation can lead to cellular dysfunction, cell death, and sometimes, immune response cascades that can extend the damage to organ and systemic levels [5, 8, 164, 165].

Two main factors can influence whether the reactive metabolite formation from a molecule would lead to IADRs. The first factor is the relative contribution of all competitive bioactivation and detoxification pathways involved in the metabolism of that molecule. Although the liver is rich in bioactivating enzymes, it is also equipped with a plethora of detoxification pathways to remove the toxic reactive metabolites. Examples of such pathways are glutathione conjugation, glucuronidation, sulfation, arene oxide hydration, and formadehyde oxidation catalyzed by glutathione S-transferases, UGTs, sulfotransferases, expoxide hydrolases, and formadehyde dehydrogenases, respectively (Table 1.2.1). Only when a reactive metabolite escapes such detoxification pathways due to poor affinity and enzyme depletion, can it damage proteins and nucleic acids. The best example is the lack of liver injury with therapeutic doses of acetaminophen and the fact that acetaminophen overdose accounts for more than 50% of drug-induced primary liver failure cases in the U.S. [86, 123]. At therapeutic doses, acetaminophen's reactive metabolite, N-acetyl-p-benzo-quinoneimine (NAPQI), is neutralized and removed by reduced glutathione (GSH) through conjugation (Figure 1.2.2). After an overdose, depleted GSH reserves can no longer compensate for the excessive production of NAPQI [26, 121]. NAPQI starts covalently binding to proteins and DNA and cause damage. The second factor is patient specific attributes such as co-morbidities, co-administered medications and genetic variants. For example, individuals with certain alleles of human leukocyte antigen (HLA) gene, which encodes for major histocompatibility complexes (MHC) — immunologic receptors that display drug-derived antigens on the cell surface, have higher risk to develop IADRs than the generation population [30, 55, 237]. These idiosyncratic, individual-specific attributes explain why drug trials on selected populations for limited time periods often fail to detect inter-personal variation and adverse events that only emerged after long exposure to the drug.

**Table 1.2.1:** Reaction Types and Enzymes that Participate in Drug Metabolism

| Type | Reactions | Catalytic Enzymes |
|---|---|---|
| **Phase I** | Oxidation | Cytochromes P450 |
| | | Flavin-containing Monooxygenases |
| | | Xanthine Oxidases |
| | | Amine Oxidases |
| | | Monoamine Oxidases |
| | Reduction | Cytochromes P450 |
| | | NADP-quinone oxidoreductases |
| | | Carbonyl reductases |
| | | Glutathione peroxidases |
| | Ester hydrolysis | Carboxylesterases |
| | | Amidases |
| | Hydration | Epoxide hydrolases |
| **Phase II** | Glucuronidation | Uridine glucurosyl transferases |
| | Sulfonation | Sulfotransferases |
| | Glutathione conjugation | Glutathione S-transferases |
| | Acetylation | Acetylases |
| | Methylation | Methyltransferases |
| | Glucosylation | Glucosyltransferase |

Data from [64]
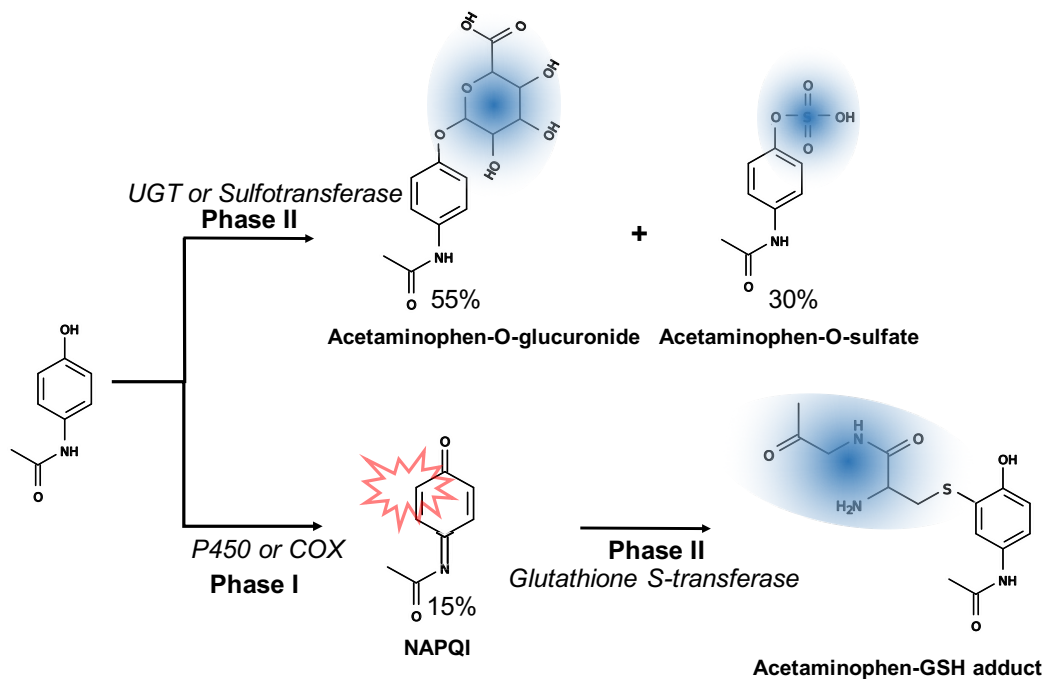
**Figure 1.2.2:** Metabolism of acetaminophen: competition between bioactivation and detoxification pathways. Each added conjugate is indicated by the blue shading. At therapeutic doses, $85\%$ of the administered acetaminophen undergoes phase II conjugation reactions and is primarily excreted in the urine as the corresponding O-glucuronide ($55\%$) and O-sulfate ($30\%$) [26]. Only $15\%$ of the administered dose undergoes phase I oxidation reaction to form the reactive N-acetyl-p-benzoquinoneimine (NAPQI) species. This small amount of NAPQI is neutralized and removed by reduced glutathione (GSH) through conjugation [121]. At higher doses, massive production of NAPQI overwhelms the GSH reserves and the reactive NAPQI starts covalently binding to proteins and DNA and eventually disrupts cellular homeostasis.

## 1.3 Current experimental approaches to directly screen for IADRs

Screening for potential toxicity is extremely important in drug development, especially in the early stages. *In vivo* animal studies and *in silico* quantitative structure-activity relationship (QSAR) models are examples of current approaches to assess IADR risk. The ultimate goal is to define predictive markers to reliably screen out lead compounds with potential toxicity early and thereby decrease late attrition rate.

There are very few good animal models to study IADRs *in vivo*. Similar to autoimmune diseases, IADR is believed to arise from a failure of immune tolerance [218]. This failure is idiosyncratic not only in humans but also in animals [218]. Strategies to develop IADR animal models include (1) testing each drug that causes IADRs on several species, (2) co-administering agents that activate the immune system or deplete GSH reserves, and (3) using genetically modified animals with impaired immune tolerance [178, 218]. Unfortunately, these attempts have been unsuccessful [178, 218]. Furthermore, due to differences in the immune system among species, results from IADR studies in animal models are often not applicable to human [237]. Manipulation of animal models to mimic the human immune system is underway but little success has been made. In addition, inter-species metabolic variation also compromises the value of animal models [10, 33, 152, 183]. Many drugs have been shown to form reactive metabolites in animal models but not in humans and vice versa. For example, efavirenz, a potent and specific HIV-1 reverse transcriptase inhibitor commonly used in HIV drug cocktails, was shown to form GSH and protein adducts in rats but not in humans, even though the drug is used at relatively high dose [174].

*In silico* quantitative structure-activity relationship (QSAR) analysis is a common screening tool for IADRs. Based on experimental data, QSAR analysis can discover the correlation between a certain biological or pharmacological activity and the physiochemical properties of compounds [131]. It can provide an understanding of how structure affects these activities, especially the interactions between functional groups of small molecules and those of their biological targets,

which may not be immediately obvious [7]. Because QSAR models can interpolate and make predictions on novel analogues within the training domain, they can be used to inform structural modification. However, because QSAR models rely heavily on biological data, which is prone to considerable experimental error, false correlations may arise. In addition, data used in QSAR models are often from multiple sets of assays whose experimental settings are not identical; consequently, QSAR model predictions have broad confidence intervals. Lastly, QSAR models often fail to extrapolate outside the range of the training domain. These disadvantages limit the potential of QSAR models in predicting IADR risk and guiding structural modification in drug design and development.

## 1.4 Current technologies in screening for reactive metabolite

Due to our inability to predict whether a reactive metabolite will be toxic, chemical intervention to minimize bioactivation is always at high priority in the drug development pipeline. From inception to completion of a drug development project, bioactivation mechanism investigation is tightly coupled with iterative structure modification using medicinal chemistry [72]. This proactive approach to addressing bioactivation is warranted when considering the issue of patient safety and the economic risks in pursuing drug candidates that covalently modify DNA and proteins.

Reactive metabolites are short-lived, with half-lives of typically less than one minute, and are usually undetectable in plasma [118]. However, the intracellular formation of reactive metabolites can be inferred from endogenous trapping reactions or physiochemistry techniques. Their formation may be modulated by enzyme induction, enzyme inhibition, and gene deletion in animals [86]. However, none of these experimental procedures are directly applicable to humans. Hence, human exposure to reactive metabolites is almost impossible to quantify. Many *in vitro* and *in vivo* experimental methods have been developed to screen for reactive metabolite formation [126]. The following explores the advantages and limitations of these approaches.

High-throughput *in vitro* covalent binding assays use GSH and cyanide ($CN^-$) as small

molecule trapping agents for soft and hard electrophiles, respectively [72]. Lead compounds are incubated in microsomal or S9 fraction [1] preparations before exposure to GSH and $CN^-$. The mixture containing stable adducts is then characterized by liquid chromatography-tandem mass spectrometry (LC-MS/MS) and/or NMR spectroscopy. This methodology has been frequently utilized to demonstrate bioactivation of structurally diverse functionalities or motifs. However, it is an indirect approach to estimating the potential for covalent binding to biomacromolecules like DNA and proteins. Not all reactive metabolites will react with GSH's nucleophilic sulfydryl group or $CN^-$; instead, some (e.g. aldehydes and $\alpha, \beta$-unsaturated ketones) may preferentially react with other nucleophilic amine-based groups such as lysine and histidine [98, 110].

Direct screening of covalent binding to biological macromolecules are conducted either *in vitro* using proteins that contain a representative array of amino acids (e.g. bovine ribonuclease A) or *in vivo* in preclinical species (e.g. rat) [32, 160]. These methods require incorporation of a radiolabel at a chemically and metabolically stable position of the molecule. The extent of protein modification is based on the amount of radiolabeled molecules covalently bound to proteins following *in vitro* incubation with human liver microsomes (sometimes hepatocytes) or *in vivo* administration in a preclinical species. However, it is impossible to differentiate between bound radiolabeled molecules and incorporation of radiolabel into macromolecules via decomposition of metabolites into precursors of endogenous compounds. In addition, direct screening for covalent binding is low-throughput and is not applicable in the early stages of drug development because radiolabeled molecule synthesis is expensive and technically challenging.

In some cases, a reactive metabolite can irreversibly inactivate the enzyme that generates it by covalently binding to an active site amino acid and/or the heme prosthetic group instead of exiting the active site [125]. Consequently, enzyme inactivation studies are another approach to screen for reactive metabolites [182]. Because bioactivation is required for enzyme inactivation, precursors of such reactive metabolites are classified as mechanism-based inactivators or suicide substrates. This screening approach is limited in scope because few reactive metabolites are

---

[1]S9 fraction preparation (post-mitochondrial supernatant fraction) contains both cytosolic and microsomal enzymes together with appropriate cofactors.

suicide substrates.

The outcome of metabolite-protein conjugation depends on the role of the protein target. While adduction of critical proteins required for important cellular functions and viability would lead to cell death, adduction of noncritical protein would only cause minor dysfunction. Furthermore, adduction of certain protein targets might be necessary for the desired pharmacological effect of a drug [71]. For example, omeprazole and exemestane work through the generation of a reactive metabolite directed at their therapeutic protein targets, $H^+,K^+$-ATPase [185] and aromatase [144], respectively. Immunochemical assays are a powerful technique for identification and characterization of cellular components that have undergone covalent modification by reactive metabolites. In this approach, a synthesized surrogate protein-hapten adduct is used as an immunogen to be administered to animals. The immunized animal would produce antibodies against the immunogen after several administrations. These antibodies are selected from the animal's antiserum using enzyme-linked immunosorbent assay (ELISA). They can then be used to identify haptenized proteins via Western blot analysis [175, 192, 254]. The major disadvantage of immunoassays is that they may not quantify all adducts that are formed because of epitope selectivity. Furthermore, screening for reactive metabolites through this approach is low-throughput and time consuming.

Lastly, direct metabolite identification through various chemical techniques is an important approach to understanding metabolic pathways and biochemical mechanisms by which metabolites are generated [137, 142]. Unfortunately, like previous methods, this approach is also low-throughput and labor-intensive.

Overall, most current *in vitro* and *in vivo* reactive metabolite screening approaches are too expensive and low-throughput to perform for thousands of candidates in the early stage of drug development. They also do not yield site-level information, which can sometimes be used to modify drugs to avoid bioactivation and make them safer, as we will see in the next section (Figure 1.5.1).

**Table 1.5.1:** Common structural alerts and their corresponding reactive metabolites after bioactivation

| Functional Group | Reactive Metabolite(s) | Bioactivation Pathway Reaction(s) |
|---|---|---|
| Aromatic amines | Quinone-imine or nitroso metabolite | Quinone formation, N-oxidation |
| Hydrazine, hydrazides | Diazene or diazonium ion | Oxidation, radical formation |
| Nitroarenes | Nitroso | Reduction |
| Fused azaheterocycles | Nitrenium ion | Dehalogenation, dehydrogenation |
| Benzylamines | Nitroso, oxime | Reduction |
| Foramides | Isocyanate | Hofmann rearrangement |
| Sulfonylureas | Isocynate | Hydrolysis |
| Thioureas | S-Oxides, isocyanate | S-oxidation |
| Hydroquinones | p-benzoquinone | Quinone formation |
| *o*- or *p*-Alkyphenols | *o*- or *p*-Quinone methide | Quinone formation |
| Methylenedioxyphenyl | *o*-quinone | Quinone formation |
| 3-Methyleneindoles | Imine-methide | Imine formation |
| Furans | $\alpha,\beta$-Unsaturated dicarbonyl | Ring scission, epoxidation, hydrolysis |
| Thiophenes | $\alpha,\beta$-Unsaturated dicarbonyl | S-oxidation, epoxidation |
| Thiazoles | Thioamide, glyoxal | Ring opening, C-oxidation, S-oxidation |
| Thiazolidinediones | Isocyanate, S-oxide | Quinone formation, S-oxidation, ring-opening |
| Arenes, bromoarenes | Arene oxide | Epoxidation |
| Alkynes | Ketenes, oxirene | Oxidation, re-arrangement |
| $\alpha,\beta$-Unsaturatedcarbonyl | Intrinsic electrophilicity | |
| Aliphatic amines | Iminium ion | Dealkylation,N-oxidation,radical formation |
| Alkylhalides | Acylhalides | C-oxidation |

## 1.5 Structural alerts

Structural alerts are chemical moieties that have the potential to generate reactive metabolites through bioactivation (Table 1.5.1). They are commonly used within the pharmaceutical industry, the FDA, and discovery tools to flag lead compounds with toxicity risk [16, 48, 69, 96]. Indeed, about 78-83% of drugs with high incidence of IADRs contain structural alerts, and 62-69% of them have evidence of reactive metabolite formation [226].

The structural alert approach is commonly used because it is the only available tool that is cheap and straightforward. Candidate compounds with one or more structural alerts are often removed from drug development pipeline or chemically modified to minimize toxicity risk. For example, when a lead compound MRL-A (a 3-acyl-N1-methylpiperazine derivative) was found to
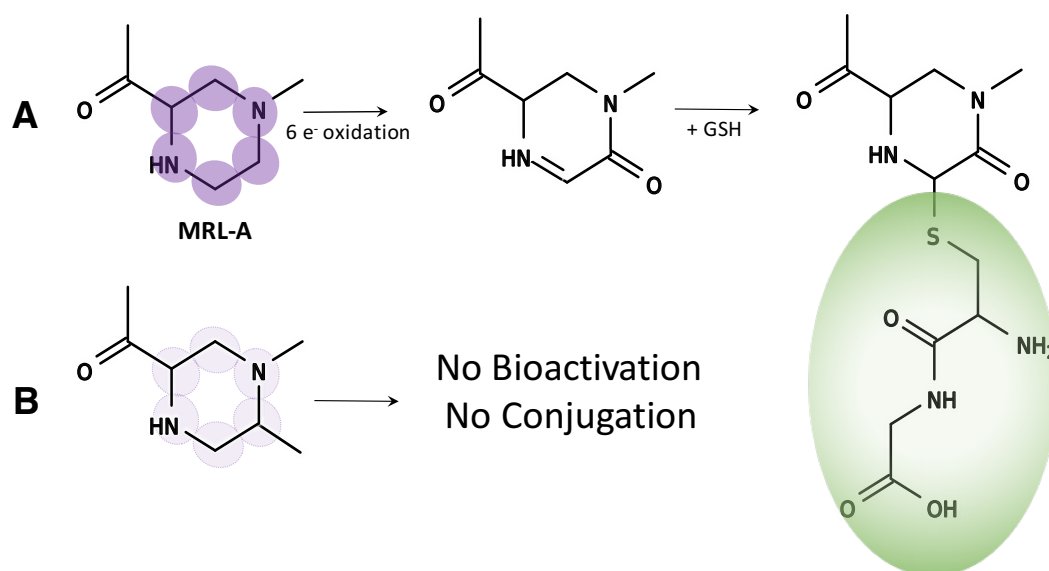
**Figure 1.5.1:** A drug development team at Merck Research Laboratories chemically modified a drug candidate, MRL-A (**A**), by adding an alkyl group at the $\alpha$ position to N1-methyl functionality in the piperazine structural alert (highlighted in purple) to create a new compound (**B**) [72]. The reactive MRL-A's metabolite was observed to conjugate to glutathione (indicated by the green shading). The new compound no longer contains piperazine structural alert and does not undergo bioactivation. This simple chemical intervention eliminates the formation of a potentially toxic reactive metabolite.

undergo bioactivation at the piperazine structural alert, researchers modified it by adding an alkyl group at the $\alpha$ position to N1-methyl functionality (Figure 1.5.1) [72]. This simple chemical intervention created a new drug candidate that was devoid of the piperazine structural alert and did not undergo metabolic bioactivation.

Iterative chemical modifications to avoid structural alerts are not always practical because these candidate compounds are simultaneously also being optimized for efficacy. More importantly, the safety hazard associated with structural alerts seems to be overestimated. Toxicity due to a specific alert structure is highly conditional, depending on metabolic transformations of the precursor and electronic interactions across the metabolite. Figure 1.5.2 shows a typical example: while the thiophene structural alert in suprofen is the site of bioactivation, the same alert structure in eprosartan is not. The differences in biotransformation lead to drastic differences in the safety profiles of the two compounds. While suprofen was withdrawn from the market due to hepatotoxicity, eprosartan is safe. In fact, of the 200 most frequently prescribed drugs in the US,

13

**A**

**Suprofen**

Cytochromes P450

**Reactive Metabolite**

IADR

**B**

**Eprosartan**

No Bioactivation

No IADR
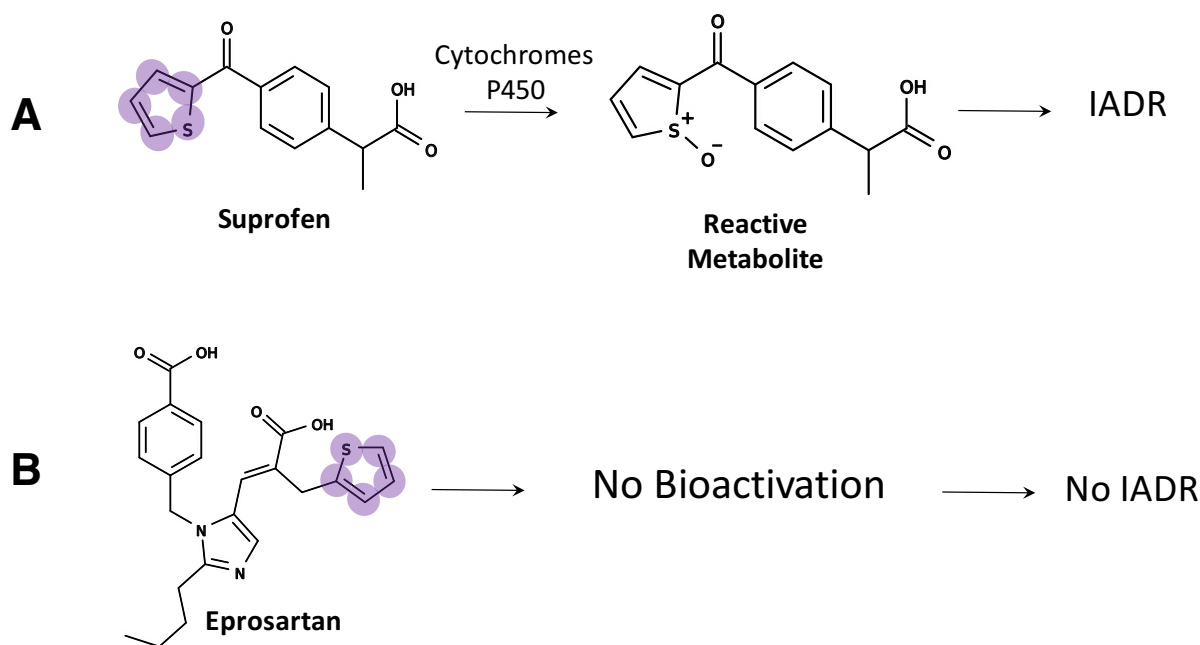
**Figure 1.5.2:** Suprofen and eprosartan both contain thiophene (highlighted in purple), a structural alert. Their toxicity profiles are very different. While suprofen (**A**), an NSAID, is withdrawn from the market due to hepatotoxicity caused by reactive metabolites, eprosartan (**B**) is a safe, commonly prescribed and used antihypertensive drug that does not form reactive metabolites [80].

$\sim 50\%$ of them contain more than one structural alert [226]. Yet, the vast majority of these drugs are not associated with adverse drug reactions despite prolonged use.

Current approaches using structural alerts fail to differentiate molecules with potential to form reactive metabolite from those without because they do not adequately model metabolism and its role in activating or evading reactive metabolite formation. Structural alerts are limited because they neither account for the likelihood (or unlikelihood) of metabolism to bioactivate the alert, nor do they account for the effect of substituents on the reactivity — the capability to form covalent adduct with macromolecules like DNA and proteins — of the bioactivated molecule [69, 119, 226]. Molecules are flagged even if (1) their structural alert is not bioactivated, (2) the resulting metabolite is not reactive, (3) the reactive metabolite is eventually metabolized into a non-reactive form, or (4) an alternative, non-activating pathway is responsible for metabolic clearance of the compound [119]. In addition, structural alerts can only identify toxic molecules with specific, well-known substructures (Table 1.5.1); they cannot identify substructures that have not yet been observed to generate toxic metabolites. Consequently, the recommendations of the

structural alert approach are very difficult to interpret; safe molecules are often flagged as toxic, and unsafe molecules slip through the process [119].

## 1.6   Metabolism and reactivity modeling to predict bioactivation

I am proposing a new method that would improve the current structural alert-based reactive metabolite screening approaches by explicitly modeling the impact of metabolism and reactivity. This method combines the existing knowledge on the biotransformation pathways of well-studied structural alerts with metabolism and reactivity models to differentiate molecules with potential to form reactive metabolite from those without. Specifically, known metabolic pathway(s) of each common structural alert can be broken down into individual reaction steps. Metabolism and metabolite structure prediction models for each step along these biotransformation pathways can then be developed using machine-learning software built from literature-derived data. Given the structure of a new compound, these models would be able to predict its biotransformation pathway(s), the structures of its metabolites, and the probabilities at which these metabolites are formed. The probabilistic output and modularity of these models allow easy interpretation and flexible application, i.e. the same model can be used in different biotransformation pathways. The reactivity of all compounds along the biotransformation pathway(s) can be calculated using the reactivity models recently developed in our laboratory [105, 106]. Figure 1.6.1 is a simple case study showing how this approach can be used. This research holds the potential to reduce the cost of drug development and decrease the chance of patient exposure to IADR-prone drugs.
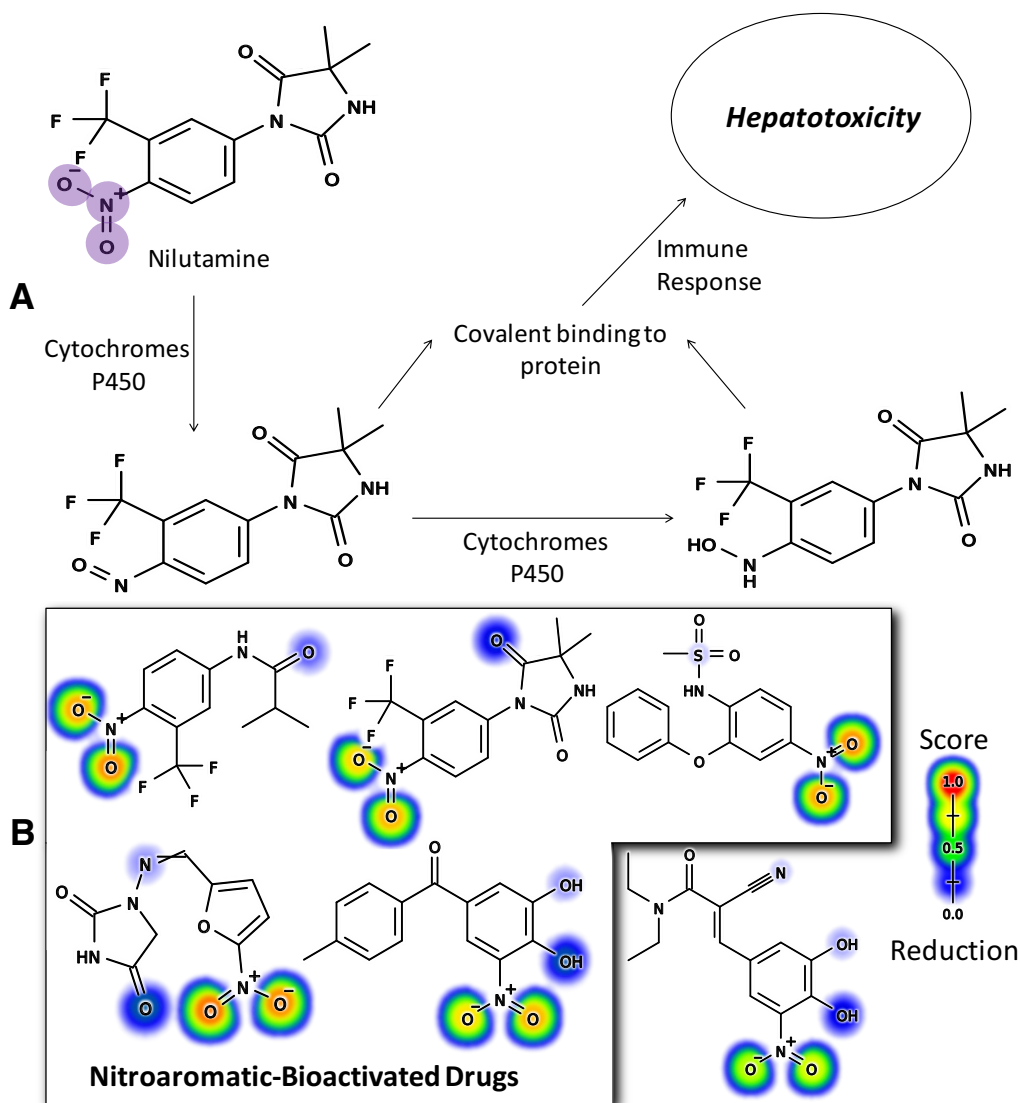
**Figure 1.6.1:** (**A**)The nitroaromatic structural alert (highlighted in purple) can undergo bioactivation by N-reduction or remain inactivated. Once the molecule is bioactivated, its metabolite could bind to proteins and trigger immune response cascade and eventually cause hepatotoxicity. (**B**) Here, I build a reduction model and use it to calculate the likelihood that an atom would undergo reduction based on the structure of the molecule containing that atom. The five drugs from left to right, top to bottom inside the box: flutatamide, nilutaminde, nimesulide, nitrofurantoin, and tolcapone form reactive metabolite by reduction at the nitrogen atom within the nitroaromatic structural alert [27, 133]. The remaining drug entacapone does not form reactive metabolite. In addition, it should be noted that tolcapone and entacapone are structural analogues. My reduction model is able to assign the bioactivated nitroaromatic structural alert nitrogen atoms with higher scores in comparison to the non-bioactived ones.

# 2

# Modeling UGT-Mediated Metabolism

This chapter is adapted from a manuscript published in *Bioinformatics*:

**Na Le Dang**, Tyler B. Hughes, Varun Krishnamurthy, and S. Joshua Swamidass, *A simple Model Predicts UGT-Mediated Metabolism* **2016**, *32*, 20, 3183-3189

## 2.1   Summary

Uridine diphosphate glucunosyltransferases (UGTs) metabolize 15% of FDA approved drugs. Lead optimization efforts benefit from knowing how candidate drugs are metabolized by UGTs. This chapter describes a computational method for predicting sites of UGT-mediated metabolism on drug-like molecules. The UGT metabolism predictor developed in this study is available at

`http://swami.wustl.edu/xenosite/p/ugt`. XenoSite correctly predicts test molecule's sites of glucoronidation in the Top-1 or Top-2 predictions at a rate of $86\%$ and $97\%$, respectively. In addition to predicting common sites of UGT conjugation, like hydroxyl groups, it can also accurately predict the glucoronidation of atypical sites, such as carbons. We also describe a simple heuristic model for predicting UGT-mediated sites of metabolism that performs nearly as well (with, respectively, $80\%$ and $91\%$ Top-1 and Top-2 accuracy), and can identify the most challenging molecules to predict on which to assess more complex models. Compared with prior studies, this model is more generally applicable, more accurate, and simpler (not requiring expensive quantum modeling).

## 2.2   Introduction

Uridine diphosphate glucuronosyltransferases (UGTs) are an important family of proteins that metabolize $15\%$ of FDA approved drugs [249]. UGTs conjugate glucuronic acid to a diverse range of molecules, rendering them more hydrophilic and more easily eliminated. Specifically, the glucuronic acid can conjugate to oxygens, nitrogens, sulfurs or carbons, in order of decreasing likelihood (Figure 2.2.1).

Understanding and modeling UGT metabolism is important because conjugation can affect the safety and efficacy of drugs. For example, a genetic polymorphism that inactivates a specific UGT significantly increases the bone marrow toxicity of irinotecan, by preventing its primary route of elimination [36]. Similarly, genetic polymorphisms that increase expression of a specific UGT reduce the efficacy of atorvastatin, a commonly used HMG CoA reductase inhibitor [197, 202]. More commonly, drug candidates are optimized to ensure that they are metabolically stable and, therefore, not too rapidly eliminated [129].

Identifying the specific atoms of a candidate drug that are glucoronidated during UGT-mediated metabolism—its sites of metabolism (SOMs)—is valuable to lead optimization efforts. Knowing the SOMs of a candidate drug allows medicinal chemists to design new

*O-Glucuronidation*    *N-Glucuronidation*    *S-Glucuronidation*    *C-Glucuronidation*
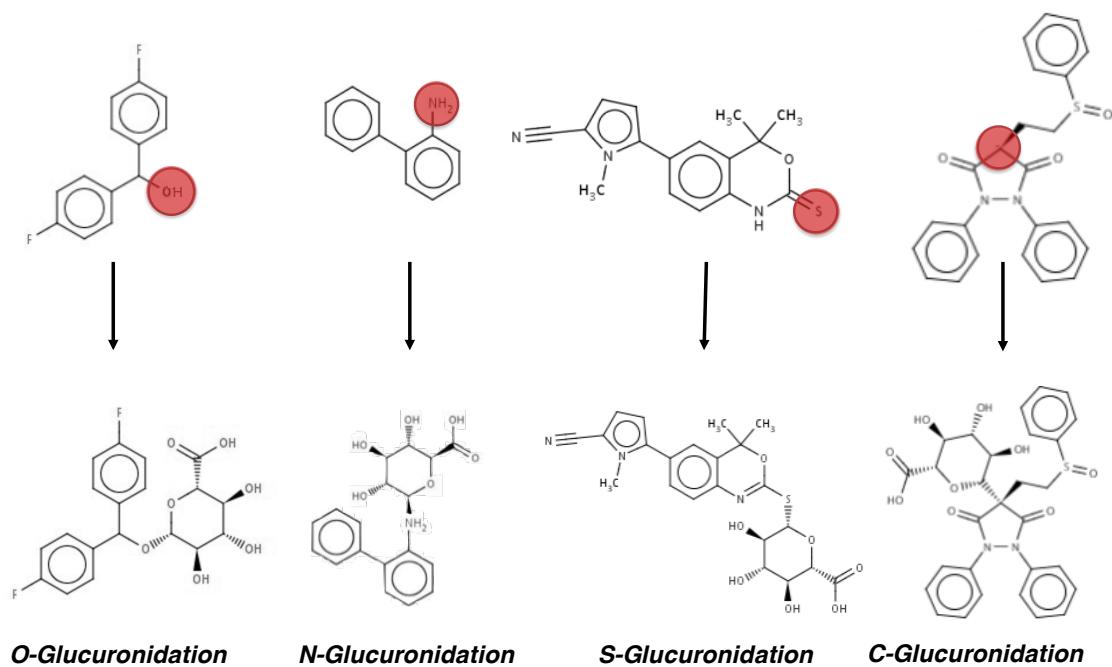
**Figure 2.2.1: Four types of UGT catalyzed reactions.**UGTs attach glucoronides to molecules to detoxify them and make them easier to excrete. Glucoronides can be attached to several atom types in a molecule, for example (in order of decreasing likelihood) oxygens, nitrogens, sulfurs and carbons. Specific examples, from the database, of each of these conjugation reactions is displayed. The site of conjugation is circled in the parent molecules.

compounds with improved bioavailability. Computational methods that predict UGT-mediated

SOMs for drug-like molecules are just now beginning to appear. So far, the literature reports two

approaches to predicting UGT sites of metabolism [194, 207]. Both methods use machine

learning to encode the chemical and biological rules represented in a set of reactions extracted

from the Accelrys Metabolite Database (AMD).

Here, we present two models to predict UGT metabolism. First, we propose a simple heuristic

model (based on global statistics) to predict UGT metabolism. This simple model correctly

predicts the site of UGT conjugation in more than $80\%$ of molecules. This heuristic model

provides both a baseline of performance against which other methods can be compared, and also a

method for identifying specific molecules in the dataset that are difficult to predict without

modeling. Second, we introduce XenoSite UGT, an adaptation of an algorithm previously

developed by our group to predict drug metabolism by cytochrome P450s [258]. The XenoSite

UGT model uses a similar approach as existing methods, by learning rules from a training dataset

derived from the AMD. We built the XenoSite model using $2898$ unique substrates containing $4557$

glucoronidation reactions, $3.2$ times more reactions than used in previous models.

XenoSite improves on existing approaches, including the simple heuristic model, in several

ways. First, it improves on the SOM-UGT model described by [194], who used support vector

machines to build four independent classification models to differentiate between observed and

non-observed SOMs for common substructures vulnerable to UGT-mediated

metabolism—aromatic and aliphatic hydroxyls, carboxylic acids and nitrogen containing groups.

Unlike XenoSite, SOM-UGT cannot predict glucuronidation of less common or atypical

substructures, such as ketones, thiols, or amides. Moreover, [194] do not describe or test any

strategies for combining the predictions of their four independent models into a single set of

predictions for a given molecule. This is a critical shortcoming, because in practice individual

molecules are being evaluated for metabolic soft spots, and there is no way to know the

performance of SOM-UGT at correctly identifying these soft spots on a per-molecule basis. In

contrast, XenoSite is a single model that produces predictions for all atoms in molecule, and can

20

even identify atypical sites of conjugation accurately. Second, XenoSite improves on the SOMP model described by [207], which uses fingerprint descriptors in a naïve Bayes classifier to rank all atoms in a molecule by their likelihood of being conjugated by UGTs. However, SOMP was only tested on small, uncharged molecules, and was not evaluated on difficult to predict molecules. On these challenging molecules, XenoSite was significantly more accurate than SOMP, SOM-UGT, and the heuristic model.

## 2.3    Materials and Methods

### 2.3.1    Training Data

The AMD was used to build a dataset of UGT substrates with their annotated sites of glucoronidation. Each reaction in the AMD contains reactant and product molecular structures, the catalyzing enzyme, and the species involved in the reaction. We extracted 4325 human reactions labeled as glucoronidation by the AMD. To confirm the AMD's classification, we used SMARTS string matching to check that each reaction product included an added glucoronide attached to an oxygen, nitrogen, carbon, or sulfur atom. UGT-mediated sites of metabolism were identified through analysis of reactant and product structures to identify which reactant atom(s) are glucoronidated. In total, 2839 unique substrates containing 3340 SOMs were curated from the AMD. All atoms in each molecule were classified into one of four substructure groups, matching those in Peng et al. [194]. These groups were aliphatic hydroxyls (AlOH), aromatic hydroxyls (ArOH), carboxylic acids (COOH), and nitrogen containing sites (Nitrogen). The Nitrogen group contained aromatic and aliphatic amino nitrogens, and aromatic and aliphatic heterocyclic nitrogens. All remaining oxygens, nitrogens, sulfurs and carbons were added to the Atypical group. For example, nitrogens in sulfoamides, urea and carboxamides were included in the Atypical group because they are only rarely conjugated. The exact definitions of these groups are included in the Supplementary Information. As expected, commonly glucoronidated substructures (Figure 2.3.1) were much more common than atypical substructures (Figure 2.3.2),

21
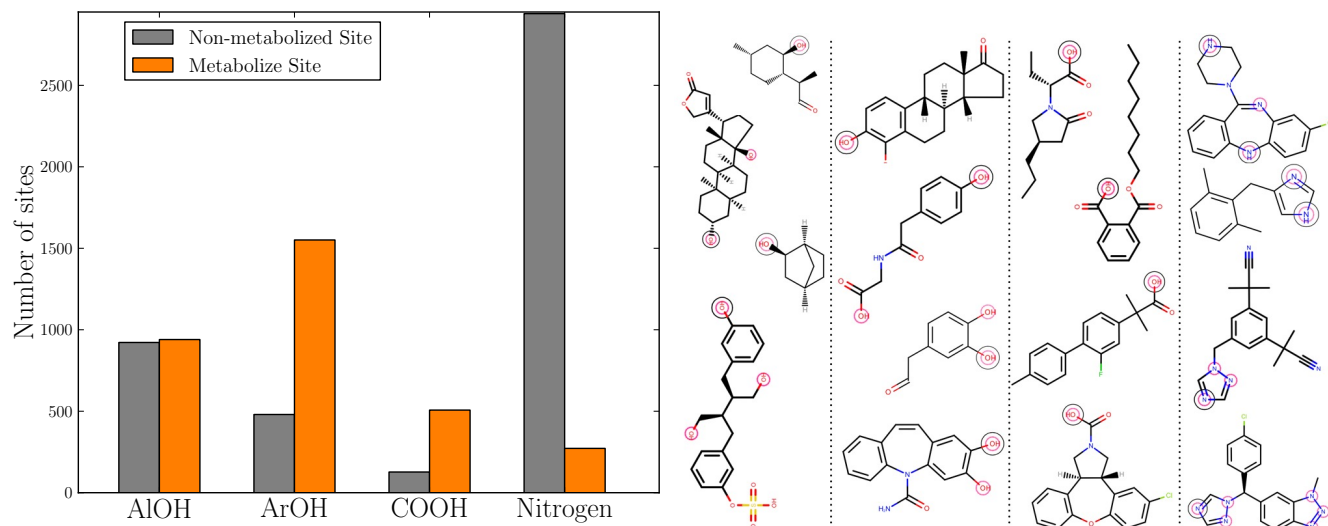
**Figure 2.3.1: The propensity of commonly glucoronidated chemical groups to undergo UGT-mediated metabolism in the data used in this study.** UGT metabolism prediction models are trained on $4557$ glucuronidation reactions from $2839$ unique substrates extracted from the Accelrys Metabolite Database. Example substrates shown above contain of commonly glucoronidated chemical groups—aliphatic hydroxyls (AlOH), aromatic hydroxyls (ArOH), carboxylic acids (COOH) and nitrogen containing groups (Nitrogen)—circled in red. Experimentally observed SOM are circled in black. In our data, the AlOH, ArOH, COOH, Nitrogen, and Atypical groups are conjugated, respectively, $49.6\%$, $76.3\%$, $80.0\%$, $8.5\%$, and $0.15\%$ of the time.

and there were enough examples to model even the atypical sites.

### 2.3.2 External Testing Data

Three external datasets were used to assess the generalizability of models built on the training data. The first testing set contains $141$ unique UGT substrates recently added to AMD (Jan-2015 version) that are not in our training dataset. The second and third test sets respectively composed of $54$ and $20$ substrates were used as validation sets by [207] and [194].

### 2.3.3 Heuristic Model

We constructed a simple heuristic model based on overall database statistics. This model was useful for two reasons. First, it provided a baseline of performance that more complex methods should outperform. Second, the molecules it cannot predict were good test cases for more complicated algorithms. In this heuristic model, all the potential sites of UGT metabolism in a
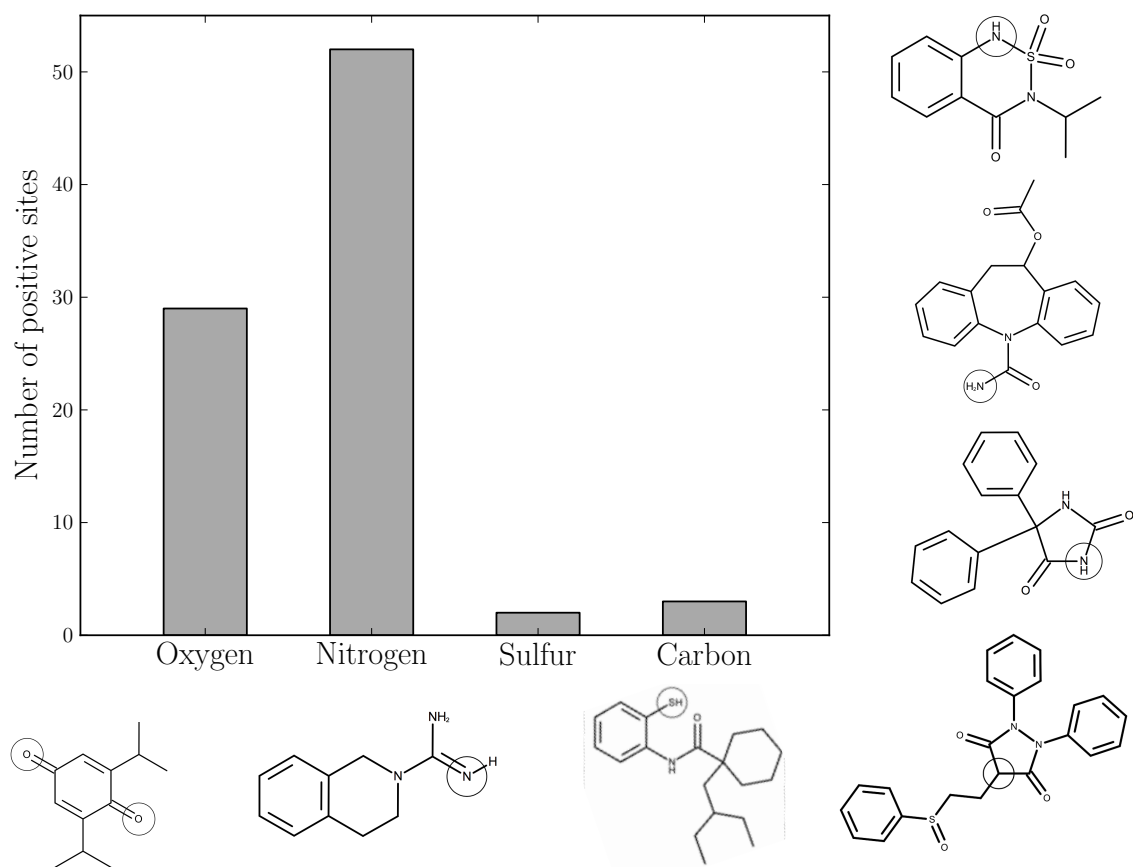
**Figure 2.3.2: Atypical chemical groups undergoing UGT-mediated glucorondiation.** In rare instances, ethers, ketones, ureas, carboxamides, sulfoamides, thiols and tertiary carbons are chemical groups that are glucoronidated . These chemical groups are collectively referred to as Atypical group in the text. Across 54965 atypical sites within thousands of molecules, only 85 positive were identified, where UGT was conjugated with oxygen, nitrogen, sulfur and carbon, respectively, 29, 51, 2, and 3 times. Examples of these atypicals are shown with sites of UGT metabolism circled.

23

test molecule were identified. Each potential site was labeled by its overall probability of being metabolized in the database. Matching these probabilities, the AlOH, ArOH, COOH, Nitrogen, and Atypical groups were assigned the initial scores of, respectively, $49.6\%$, $76.3\%$, $80.0\%$, $8.5\%$, and $0.14\%$. Next, these scores were summed across the whole molecule to compute a normalization term. Finally, the initial scores were divided by this normalization term to yield the final score. This score sums to one in molecules that have at least one potential site. The atoms of a molecule with no potential sites all receive a score of zero. A Python implementation of this model is included in the Supplementary Information to facilitate future studies.

### 2.3.4   Descriptors

A vector of numerical descriptors represented each atom in a test molecule; 98 descriptors in total, 62, 20, and 16 of which encode topological-, molecule-, and quantum chemical-derived chemical information, respectively. These descriptors were chosen because they are effective for modeling P450 metabolism in our prior work [258]. Descriptors were computed using in-house software applied to SDF files with 3D coordinates (generated using Open Babel) and explicit hydrogens [186]. We added an additional set of 7 statistical descriptors based on the heuristic model. Five of these descriptors encoded the number of certain substructural groups—AlOH, ArOH, COOH, Nitrogen (NR3, NHR2, NH2R) and thiols—that are contained in a given substrate. All atoms of the same molecule had the same values for these descriptors. Two atom-level descriptors, heuristic score (see 7.3.4) and the number of topological equivalent atoms were also added. A full description of all descriptors used in this work is available in the Supplemental Information.

### 2.3.5   Machine Learning Models

A matrix of descriptor encoded atoms was presented to a neural network with 10 hidden nodes. For comparison, we also trained a logistic regressor on the same data. During training, we learn a mapping between the descriptor values of each atom and the binary experimental response of that atom, metabolized or not metabolized. Also, atoms in the dataset were weighted so that the less

common elements were equally important to the model's error function as the most common atoms. The weights of the model were calibrated during training by performing gradient descent on the cross-entropy error of the difference between the predicted and actual response values of each atom. Predictions were obtained using a leave-one-molecule-out cross-validation procedure. Each molecule was predicted by an independent model trained using the remaining molecules of the training set.

### 2.3.6   Performance Metrics

Two different metrics were used to evaluate the prediction accuracy. The first was the Top-$N$ metric, which considered a substrate correctly predicted if any of its experimentally observed SOMs were predicted in the top $N$ rank-positions out of all possible SOMs of the substrate. Tied scores are handled appropriately, by averaging Top-$N$ scores over all permutations of the tied sites within each molecule. The Top-2 metric is the standard metric for evaluation CYP site of metabolism models and is sensible in this context as well [210, 221]. We used a paired $t$-test to compute the statistic significance between Top-$N$ accuracies. The second metric was the area under the ROC curve (AUC) [228] of the predictions of metabolized and not metabolized atoms that all belong to the same chemical group: AlOH , ArOH, COOH, Nitrogen, or Atypical. This is a standard metric employed in machine learning, as well as by Peng et al. [194], to determine how well a method is able to distinguish between positives and negative test cases. This metric does not measure the within-molecule accuracy, which is much more important for this specific application. At the same time, it exposes which types of sites are best predicted. We used fisher's exact test to compute the significance of ROC differences, choosing the score cutoff at the point closest to the upper left corner of the ROC plot. All significance tests used a threshold of 0.05.

|  | Heuristic | Logistic | Neural Net |
|---|---|---|---|
| Top-1 | 79.9 | 85.5 | 86.1 |
| Top-2 | 91.6 | 96.3 | 97.1 |
| Top-3 | 94.7 | 98.4 | 98.9 |
| AlOH-AUC | 84.6 | 86.4 | 87.1 |
| ArOH-AUC | 80.0 | 81.3 | 83.1 |
| COOH-AUC | 85.4 | 86.8 | 87.5 |
| Nitrogen-AUC | 87.8 | 90.8 | 93.3 |
| Atypical-AUC | 86.7 | 98.6 | 98.4 |

**Table 2.4.1: Leave-one out cross-validated prediction accuracies on the training dataset.** For each metric, the highest performance is bold, along with any scores not statistically different from the best performance (using a p-value cutoff of 0.05). XenoSite is the best performing models across all metrics. The number of metabolized to nonmetabolized AlOH , ArOH, COOH, Nitrogen, and Atypical sites are, respectively, $925/937$, $1551/480$, $507/127$, $272/2940$, and $85/54880$.

## 2.4 Results and Discussion

### 2.4.1 Accuracy in Identifying SOMs

XenoSite model very accurately predicted UGT metabolism, and the heuristic model was nearly as accurate. XenoSite model, using a neural network, had cross-validated Top-1, Top-2, and Top-3 accuracies of $86.1\%$, $97.1\%$ and $98.9\%$ for the training set (Table 2.4.1 and Figure 2.4.1). The AUC accuracies of XenoSite for AlOH , ArOH, COOH, Nitrogen, and Atypical chemical groups were, respectively, $87.1\%$, $83.1\%$, $87.5\%$, $93.3\%$, and $98.4\%$. It appears that modeling UGT metabolism is easy for most molecules; according to the Top-2 metric, the XenoSite models were $10\%$ more accurate than CYP models, and the heuristic model alone was $91\%$ accurate. According to the Top-1, Top-2, and Top-3 metrics the performance difference between XenoSite and the baseline heuristic was respectively $6\%$, $5\%$, and $4\%$. Likewise, the accuracy of the neural network and logistic regressor models differed by less than one percent for all performance metrics, but this improvement was consistent so we decided to continue using the neural network.

At the same time, the neural network identified Atypical and Nitrogen sites much better than
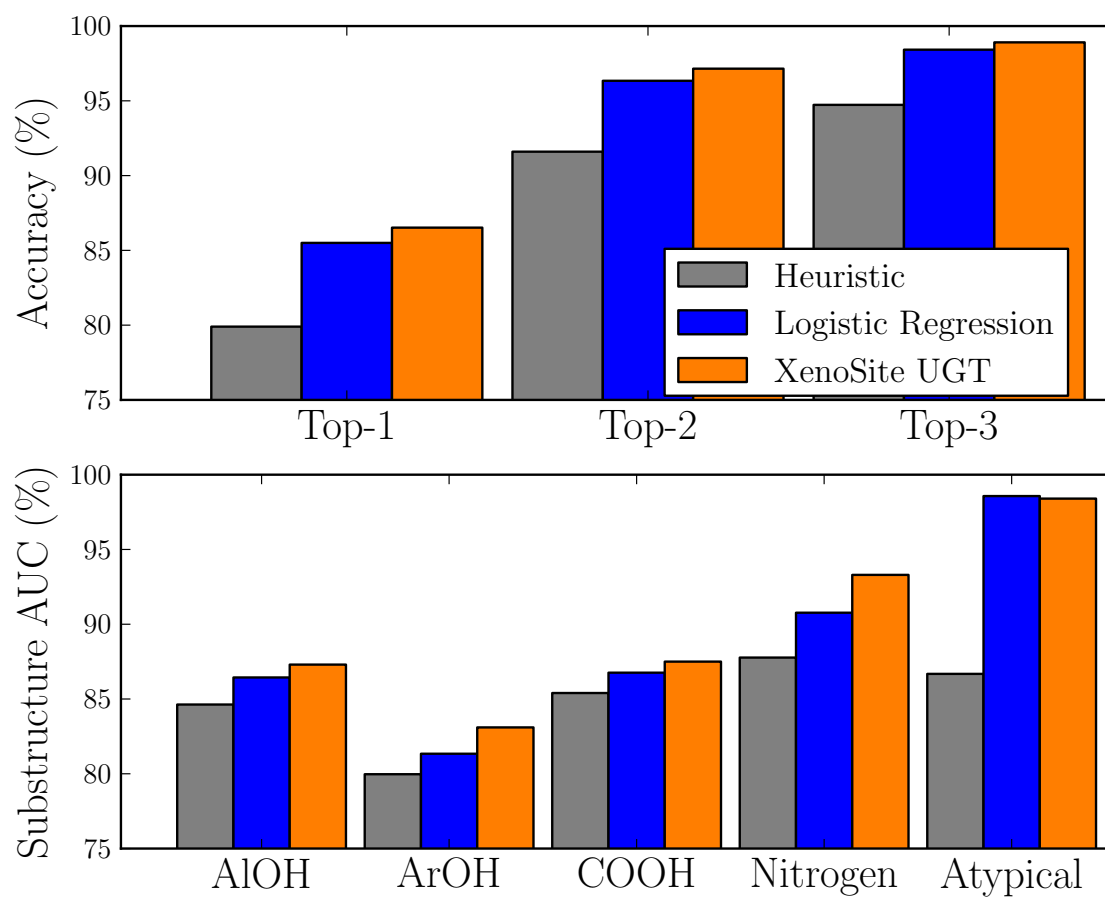
**Figure 2.4.1: Comparison between neural network, logistic regression and heuristic models.** Leave-one out cross-validated prediction accuracies on the training dataset are shown. Across all performance metrics, the neural network performs best, followed closely by the logistic regressor, and finally by the heuristic method.

the heuristic model. About 80% of the training set molecules were very easy, and predicted correctly by the heuristic model. The remaining molecules were difficult, but often predicted correctly by XenoSite (Figure 2.4.2).

## 2.4.2 Descriptors Driving Accuracy

A permutation sensitivity analysis identified the descriptors driving model accuracy [107]. We started with the trained model. Next, we selected the 508 molecules that were poorly predicted by the heuristic model (with no validated SOMs within the molecules' first ranked heuristic scores). Next, we randomly permuted each descriptor column (or group of closely related descriptors) in the input data for these molecules. The trained model was applied to the permuted data, and the performance drop across all the molecules was recorded. The higher the performance drop, the more important the descriptor to the model's performance.

We saw similar results using all performance metrics (Figure 2.4.3). Using the aggregate Top-2 performance as a guide, this analysis identified topological descriptors (identities of the atom and its neighbors (atoms one, two and three bond away), number of bound hydrogens and heavy atoms and size of ring containing the atom) and heuristic descriptors (heuristic score and number of substructures) as the most important descriptors for differentiating metabolized sites from non-metabolized sites. The result revealed that XenoSite heavily relies on local topology for calculating SOM score, and does not need descriptors from the quantum simulation. Similar results were seen in the substructure-specific sensitivities. Here, once again, topological and heuristic descriptors were the most important, and no quantum chemical descriptors were necessary. At the same time, a few molecule-level descriptors (like logP) were also important. Notably, the two heuristic descriptors were consistently among the most important.
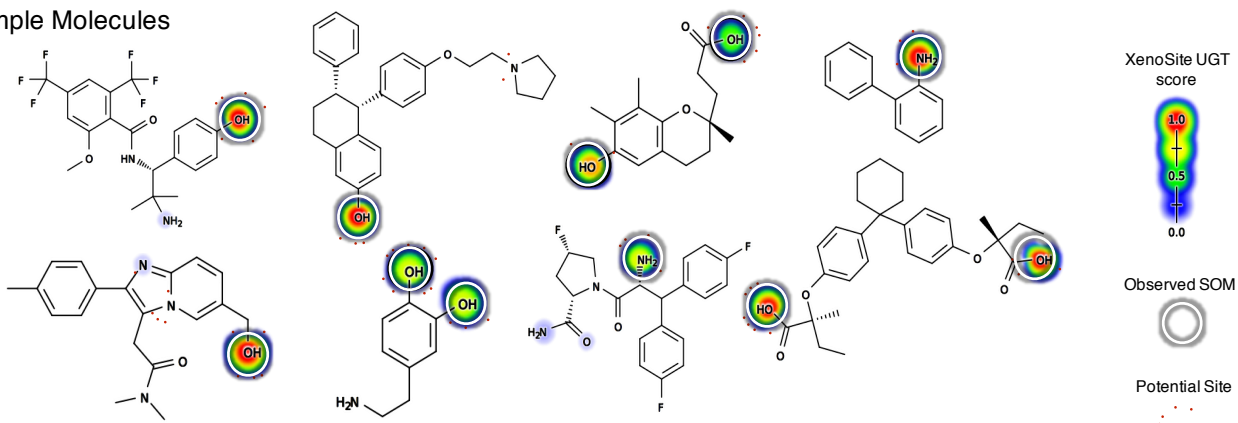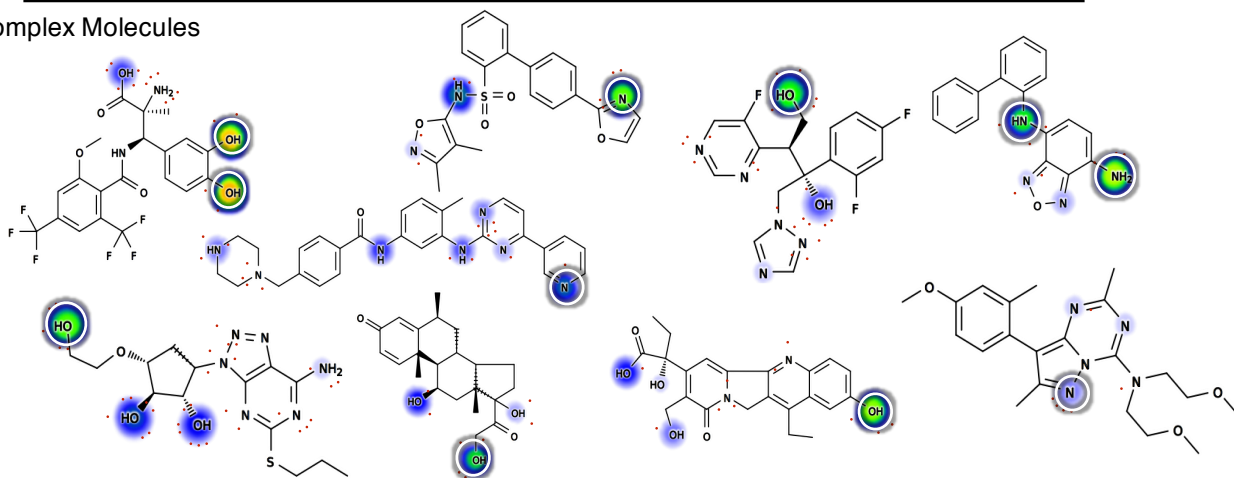
**Figure 2.4.2: XenoSite predictions.** Example molecules are shown with their experimentally observed UGT site of metabolism (SOM) circled in black. Four commonly glucoronidated chemical groups—AlOH , ArOH, COOH, Nitrogen—are in small dashed-circles. Possible atypical sites are not circled, even though our model can predict them as sites. The shading on each atom plots the cross-validated prediction score. The top panel shows simple molecules with SOMs that can be readily identified by the heuristic model. The bottom panel shows complex molecules with SOMs that are accurately predicted by the XenoSite but not the heuristic model. The complex molecules are real drug candidates and drugs: GSK101892 [193], Imatinib mesylate [84], Voriconazole [204], 10074-G5 [49], Oxycodone [19], Methylprednisone [242], Ticagrelor [232], SN-38 [230] and Dasatinib [45].
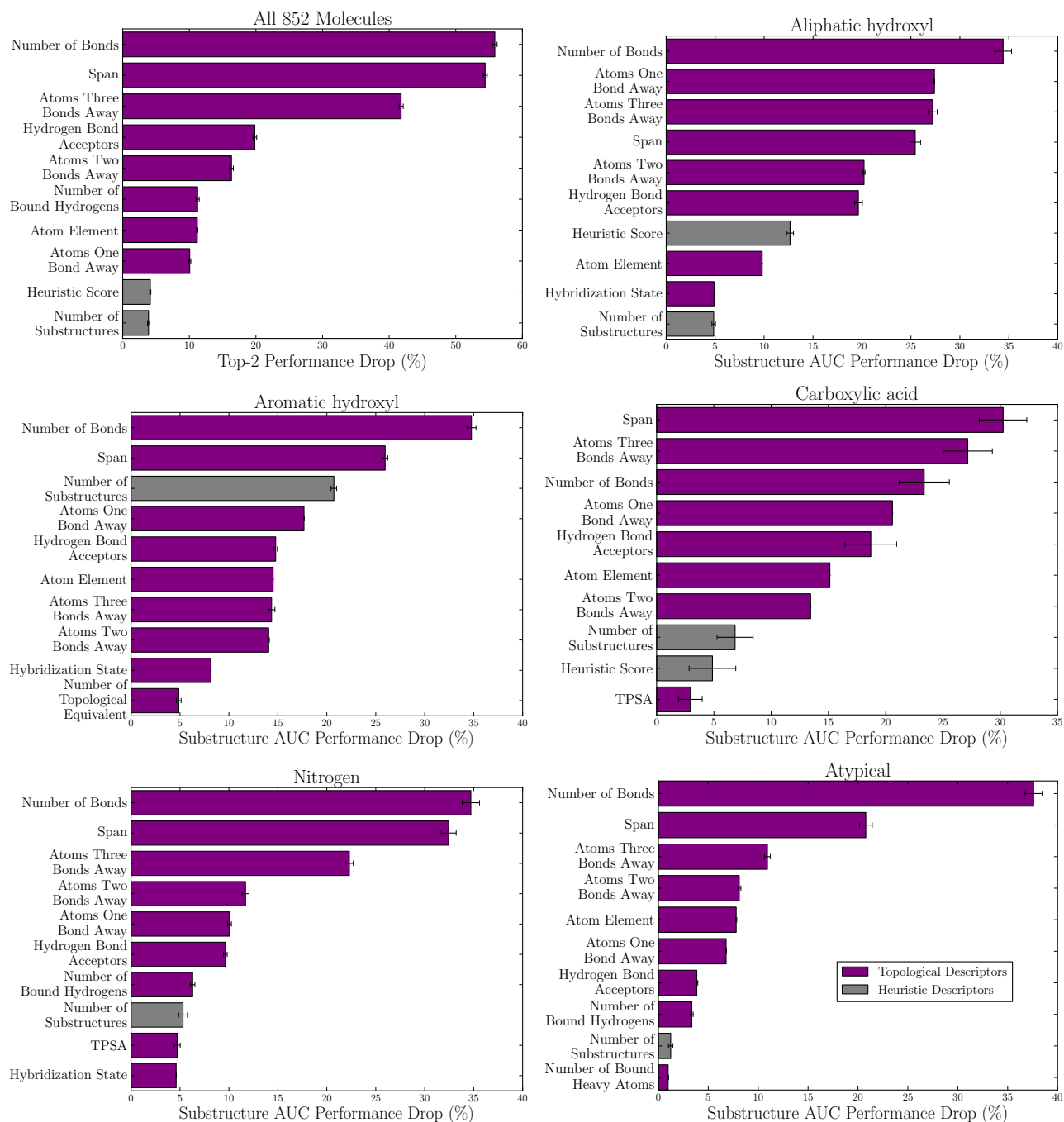
**Figure 2.4.3: The importance of specific descriptors to predicting 852 difficult molecules.** A permutation sensitivity analysis quantified the importance of descriptors for the human UGT metabolism model to differentiate metabolized from non metabolized sites across the whole dataset (Top-2 accuracy) and within all potential sites of the same substructure groups (AUC accuracy). The 10 most important descriptors are plotted. The heuristic and topological descriptors were most important. Surprisingly, none of the quantum level descriptors were important. Consequently, the final model we publish in the website does not run a quantum simulation to make predictions.

30
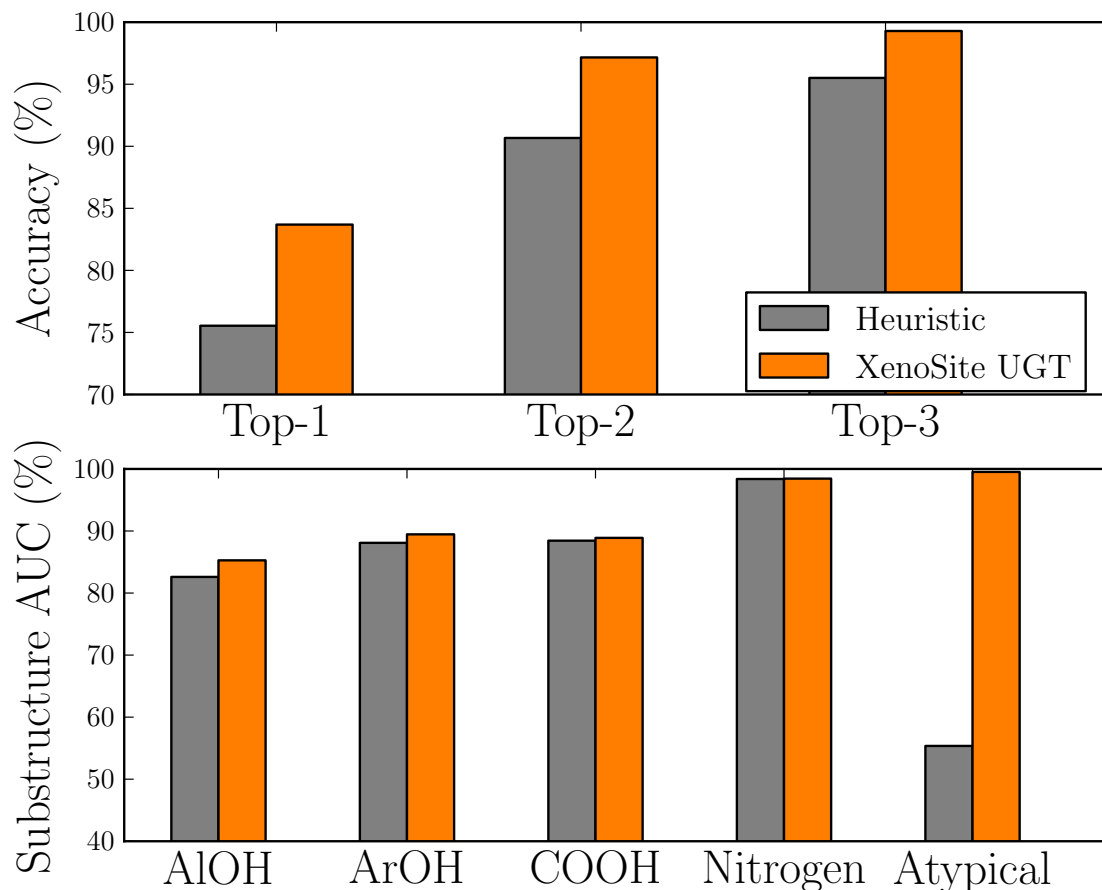
**Figure 2.4.4: Performance on external test sets:** A test set of 141 unique molecules that were not in our training set were collected by filtering the recently added glucuronidation reactions in the Accelrys Metabolite Database. The neural network performed better than the heuristic model across all metrics.

### 2.4.3  Performance on External Dataset

We settled on using a neural network trained on heuristic and topological descriptors for our final model, but also saw value in the heuristic model's simplicity. Both the neural network and the heuristic models performed well on the external testing dataset of 141 molecules (Figure 2.4.4). Just like in the the training set, the neural network performed better than the heuristic model (83.7% versus 75.5% Top-1 accuracy). Accuracies in substructure AUCs, however, were essentially identical except with atypical sites. Here, the neural network achieved an AUC of 99.5%, compared to 55.4% for the heuristic model. Overall, performance on the testing dataset of either model was comparable to the cross-validated performance on the training dataset.

### 2.4.4 Comparison to Prior Studies

Performance comparisons to prior methods using previously published external datasets were inconclusive. We first compared XenoSite to SOMP [207]. SOMP's test set included 20 molecules. Prediction accuracies of XenoSite and SOMP [207] on this test—Top-1, -2 and -3 scores of 90%, 95%, and 95%, respectively for SOMP versus 95%, 95%, and 100% for XenoSite—were not statically significant. Similarly, the SOM-UGT model in Peng et al. [194] was validated using an external dataset of 54 molecules, reporting only sensitivity and specificity for each chemical group. Difference in accuracies were not statistically significant ( Figure 2.4.5), probably because this dataset is small and is mainly composed of very easy to predict molecules.

A test set built using 49 molecules that the heuristic model does not predict accurately according to the Top-1 metric, showed XenoSite is significantly more accurate than the heuristic model, SOM-UGT and SOMP. A subset of 49 difficult molecules was identified from the 141 external test set molecules. Each of these 49 molecules was not correctly predicted by the heuristic model. The sites of UGT metabolism for these 49 molecules were predicted using SOMP's website and SOM-UGT software. Top-*N* and substructure AUC metrics were used to assess the performance of the two models. The true positive, true negative, false positive and false negative values were calculated using the point closet to the upper left conner on the ROC curve. SOMP was unable to make predictions on two molecules with positive charges in this testing set while all other models could. As shown in Table 2.4.2, XenoSite outperforms all other models across all accuracy metrics. Specifically, XenoSite has statistically-significant better performance than SOMP, SOM-UGT and the heuristic model according to the Top-2,-3 and Atypical AUC metrics.

XenoSite and the heuristic model run very quickly, taking less than a second per molecule. We expect SOMP, which is based on fingerprints, is similarly fast. However, SOM-UGT requires 59 minutes to predict the 49 molecules in the test set. In this regard, SOM-UGT is substantially slower than other approaches, and therefore less useful for screening large numbers of molecules.

These comparisons show that our modeling approach outperformed existing methods in tests,
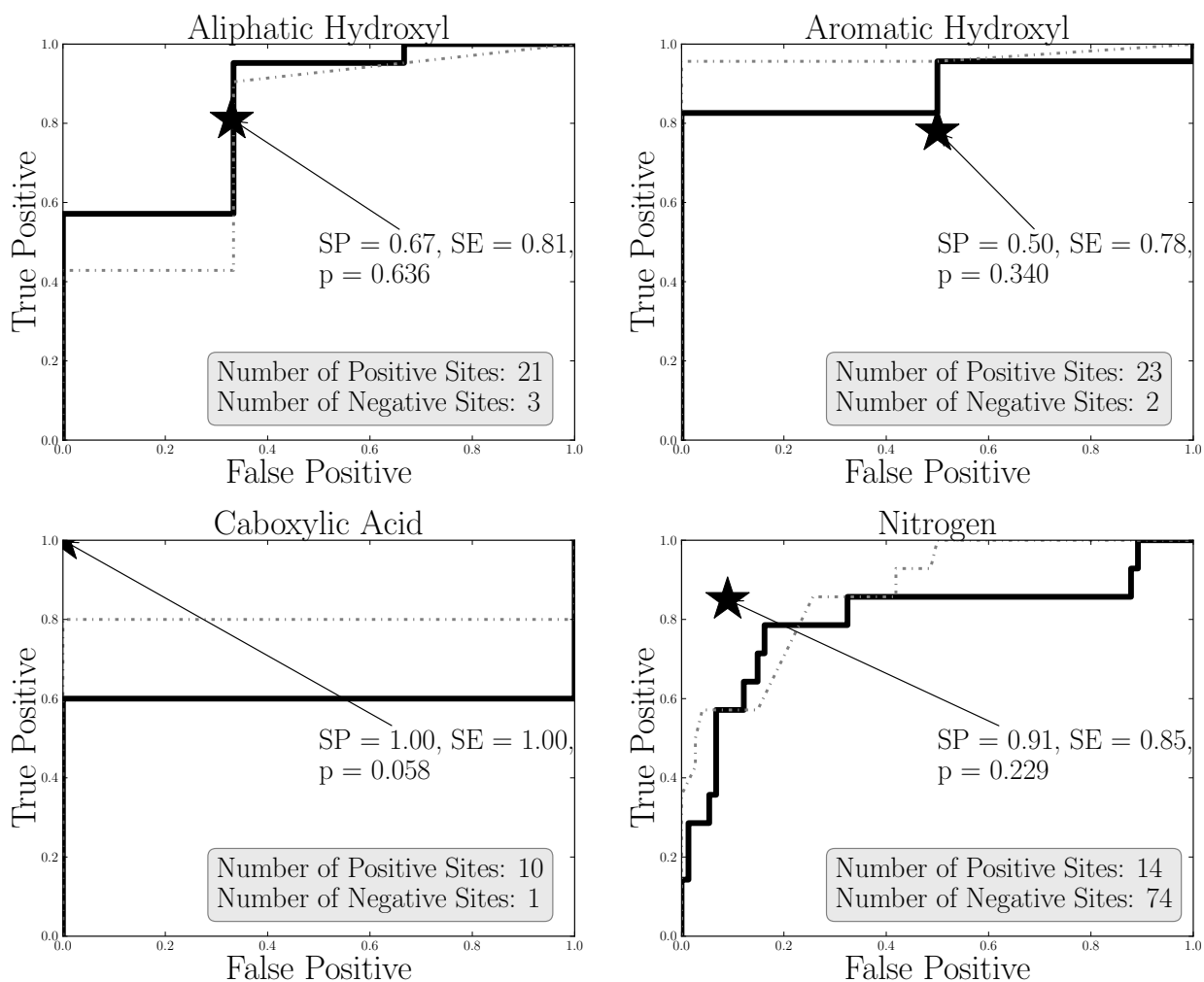
**Figure 2.4.5: The XenoSite predictions on an external dataset of 54 molecules** The external dataset in the [194] study is used to test our model. The published performance of SOM-UGT are depicted as stars. XenoSite ROC plots are depicted as solid black lines. Heuristic ROC plots are depicted as dash lines. The points closest to the upper left corner of the XenoSite ROC plots are chosen for comparison to SOM-UGT. The performance difference between our model and the SOM-UGT on this test set is not statistically significant (Fisher's exact test (2x4) two-tailed $p$-values: AlOH: 0.636, ArOH: 0.340, COOH:0.058, Nitrogen: 0.229).

**Table 2.4.2: XenoSite is more accurate than all other methods on 49 difficult to predict molecules.** To compute the Top-$N$ performances, a global SOM-UGT model was constructed predicting positive sites that were predicted positive by any of the four published substructure-specific models. For each metric, the highest performance is bold, along with any scores not statistically different from the best performance (using a p-value cutoff of 0.05). XenoSite is the best performing model. The number of metabolized to nonmetabolized AlOH , ArOH, COOH, Nitrogen, and Atypical sites are, respectively, $26/44$, $17/23$, $2/15$, $8/36$, and $2/1062$.

|  | Heuristic | XenoSite | SOM-UGT | SOMP |
|---|---|---|---|---|
| Top-1 | 0.00 | 53.06 | 33.70 | 49.88 |
| Top-2 | 82.86 | 91.84 | 56.13 | 71.43 |
| Top-3 | 94.29 | 97.96 | 63.70 | 79.59 |
| AlOH-AUC | 75.61 | 77.53 | 63.32 | 74.97 |
| ArOH-AUC | 60.33 | 68.03 | 48.59 | 64.05 |
| COOH-AUC | 60.71 | 56.67 | 43.33 | 48.21 |
| Nitrogen-AUC | 86.50 | 91.91 | 49.26 | 92.13 |
| Atypical-AUC | 89.82 | 99.43 | 50.00 | 93.67 |

but also highlights a key deficiency in the literature. Accuracy in this problem is strongly driven by the number of easy molecules that are trivially predicted by a heuristic method. We recommend future modeling effort should be directed towards predicting molecules performance according to the heuristic model on test sets in future studies.

## 2.5   Conclusion

This study introduces two approaches to predicting UGT Sites of Metabolism: (1) a statistics-based heuristic model and (2) XenoSite, a neural network trained on a large database of UGT metabolism. XenoSite accurately predicts observed SOMs of known substrates $86\%$ of the time, and outperforms existing methods, including the heuristic model, on difficult molecules. XenoSite might might be most useful in contexts where atypical sites of UGT metabolism are important and, therefore, the heuristic model is less accurate.

## 2.6   Abbreviations

ADR, Adverse Drug Reaction; AMD, Accelrys Metabolite Database; AUC, Area Under the

Receiver Operating Characteristic Curve; UGTs, Uridine diphosphate glucunosyltransferases

SOM, Site of Metabolism

# 3

# Modeling Phase II Metabolism

This chapter is adapted from a manuscript in preparation:

**Na Le Dang**, Matthew K. Matlock, and S. Joshua Swamidass, *Deep Learning Phase II Metabolism in Four Colors*

## 3.1 Summary

Phase II metabolism works in concert with Phase I metabolism to influence the absorption, disposition, metabolism excretion, and toxicity of drugs. However, there have been few modeling efforts to predict Phase II metabolism. In addition, because metabolic reactions have been studied in isolation, a common limitation of both *in vitro* and *in silico* studies is that they cannot integrate

these isolated systems to predict their net effect *in vivo*. To fill this knowledge gap, we built a Phase II metabolism model that predicts both the site and substrate specificity of four major Phase II reactions: glucuronidation, sulfonation, acetylation and methylation. To learn the net effect of competing pathways, we implement a model architecture with multi-score output layers and cross-talk layers. The multi-score output layers allow information-sharing between metabolism prediction tasks. The cross-talk layers allow integration of different reaction types and different sites, scaling the output of a specific reaction type to reflect the presence of competing reaction types. Our model can predict the site specificity with an average area under the ROC curve (AUC) of $98\%$ and the substrate specificity with an AUC of $83\%$.

## 3.2   Introduction

Given the importance of metabolism in the absorption, disposition, metabolism and excretion (ADME) profile of a drug, the field of drug metabolism has rapidly evolved both experimentally and computationally. While much effort has been focused on Phase I metabolism, specifically that of P450 enzymes, very few models for Phase II have been developed. [57–59, 97, 103, 155, 207, 208, 210, 257–259, 261] Our group and others had previously developed site of metabolism models for uridine diphosphate glucuronosyltransferases (UGT)[56, 194, 207] and glutathione transferases (GST).[105, 106] Despite accounting for only $\sim 20\%$ of all metabolism reactions,[249] Phase II metabolism significantly contributes to the ADME profile of a drug.[41, 51, 162, 219] As the results, having both accurate Phase I and Phase II models is imperative to computational metabolism assessment.

Uridine diphosphate glucuronosyltransferases (UGT), sulfotransferases (SULT), N-acetyltransferases (NAT), thiopurine S-methyl transferase (TPMT), and glutathione transferases (GST) are the five most important Phase II metabolizing enzyme families.[88, 115, 246] Together, these enzyme families account for $87\%$ of all Phase II metabolism.[115] In this study, we developed a model that simultaneously predicts UGT, SULT,
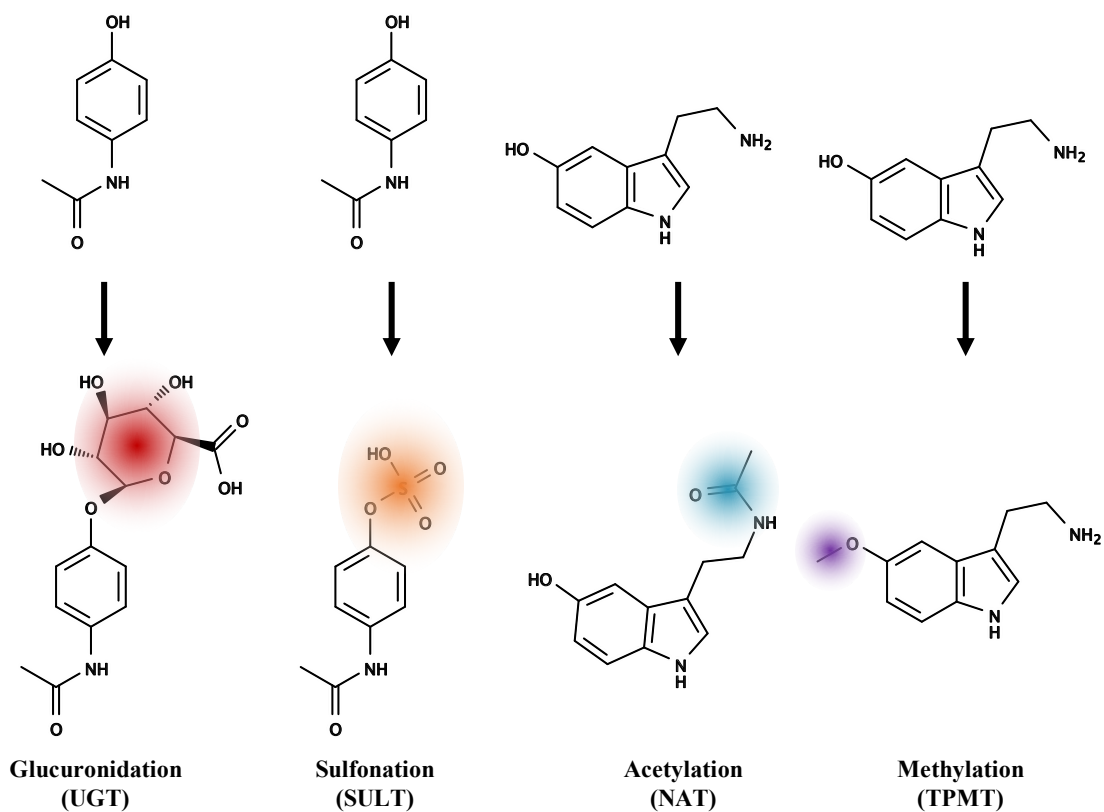
**Figure 3.2.1: Four Phase II reaction types covered in this study**. We developed a model that simultaneously predicts glucuronidation (UGT), sulfonation (SULT), acetylation(NAT), and methylation (TPMT) sites of metabolism. Example reactions above are glucuronidation and sulfonation of acetaminophen, and acetylation and methylation of serotonin.

NAT, and TPMT sites of metabolism (Firgure 3.3.1). GST metabolism has been assessed extensively by our group, and is not covered in this study. [105, 106]

TO DO: Add a paragraph about the model prediction integration.

## 3.3    Methods

### 3.3.1    Training Datasets

We collected a chemically diverse Phase II substrate data set from the literature-derived Accelrys Metabolite Database (AMD). A total of 17181 Phase II reaction records were collated. Multiple reactions involving the same substrate were mapped onto a single substrate instance using maximum common substructure matching. [39] The final Phase II data set contained 8750 unique substrates. There were 6417, 2211, 918 and 910 UGT, SULT, NAT, and TPMT substrates, respectively. 7468, 2535, 995, and 1125 sites of UGT, SULT, NAT, and TPMT metabolism, respectively, were registered (Figure 3.3.1). Under the AMD license agreement, we were not allowed to disclose the structures of molecules in the data set. However, to enable the reconstruction of our complete database and reproduction of our results, we provided all reaction and molecule AMD registry numbers in the Supplemental Materials.

### 3.3.2    Phase II Model

To predict the susceptibility to UGT, SULT, NAT, and TMPT metabolism, we trained a deep neural network with one input layer, three hidden layers, one cross-talk layer, and two output layers (Figure 7.3.2). For an input molecule, atom-level and molecule-level descriptors were calculated for all atoms. There were 15 molecule-level and 390 atom-level descriptors. The descriptors are detailed in the Supplemental Materials. The input layer fed to the first hidden layer of 20 nodes, which in turn fed to the second hidden layer of four nodes. For each atom, this second hidden layer would generate four scores, corresponding to the four types of Phase II

| | UGT | SULT | NAT | TPMT |
|---|---|---|---|---|
| UGT | | 1360 | 73 | 259 |
| SULT | 1263 | | 34 | 183 |
| NAT | 117 | 52 | | 19 |
| TPMT | 265 | 184 | 27 | |
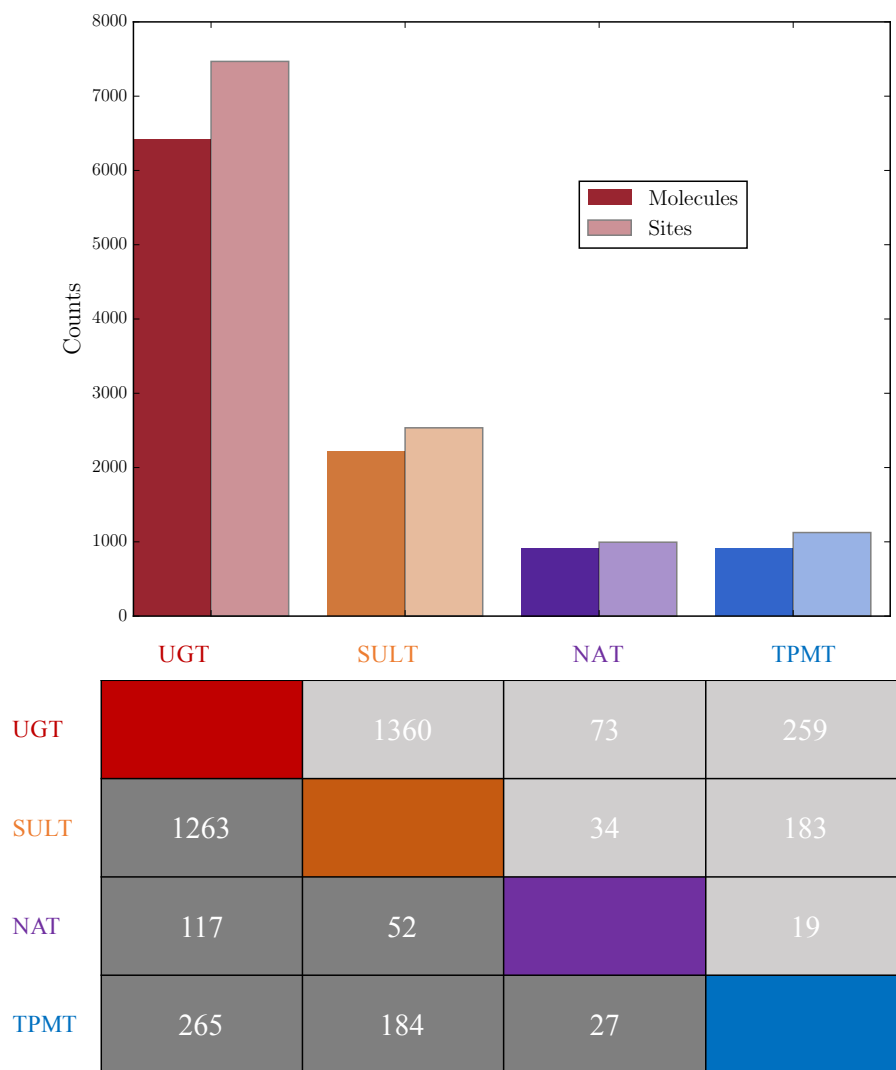
**Figure 3.3.1: Training data set**. (Upper) The numbers of molecules and sites metabolized by each enzyme family are shown. There were 17181 compounds in the XenoSite Phase II training data set. 6417, 2211, 918 and 910 were UGT, SULT, NAT, and TPMT substrates, respectively. There were 7468, 2535, 995, and 1125, respectively, sites of UGT, SULT, NAT, and TPMT metabolism. (Lower) The numbers of molecules (dark-gray cells) and sites (light-grey cells) that metabolized by pairs of enzymes are shown.
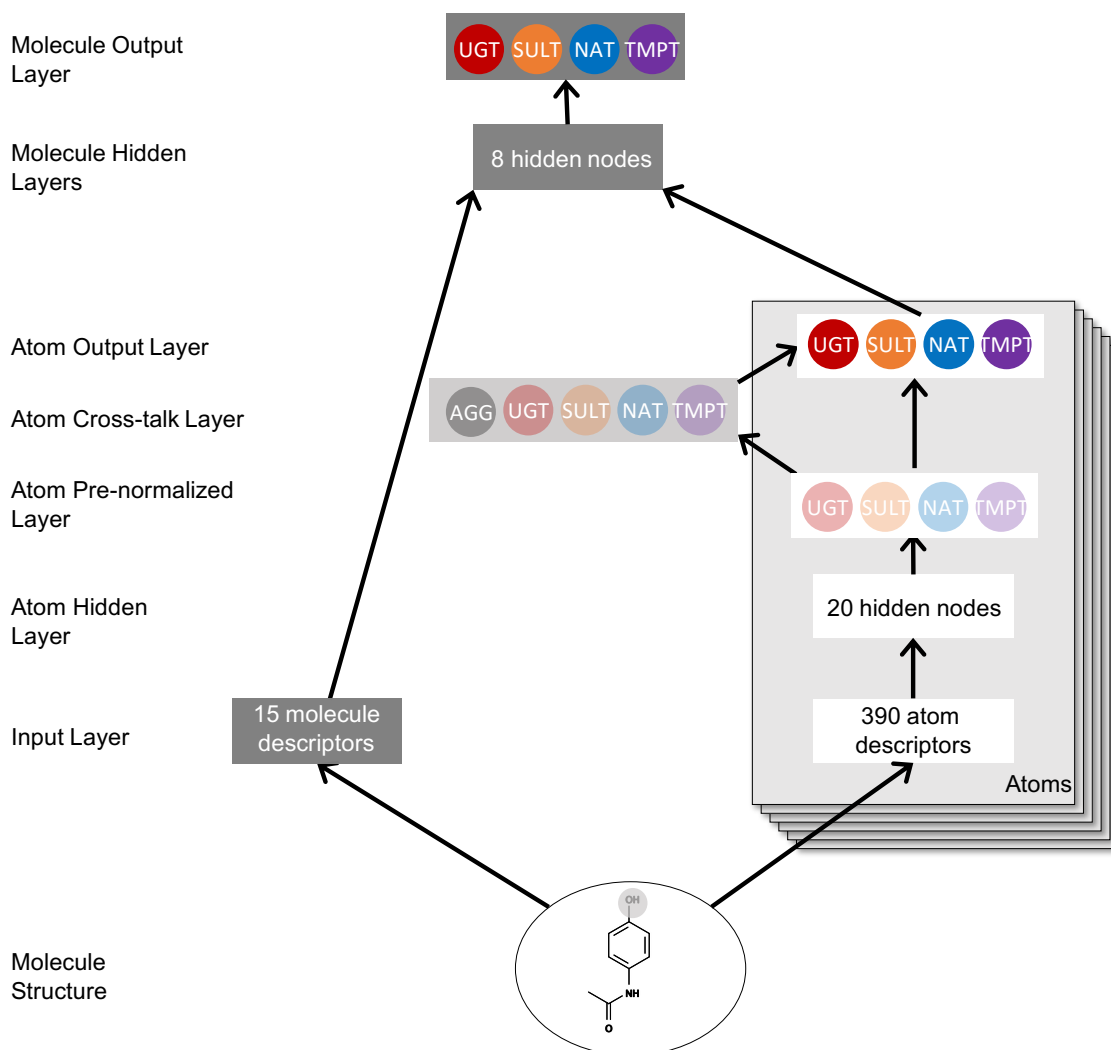
**Figure 3.3.2: The Structure of the Phase II Reaction Model.** Our neural network model contained one input layer, two hidden layers, and two output layers. From the structure of an input molecule, 15 molecule-level and 390 atom-level descriptors were calculated and fed to the input layer. Both molecule- and atom-level descriptors were submitted to the atom neural network, which computed four atom-level scores, corresponding to the four Phase I reaction types, for each atom in the input molecule. The atom network was trained using site-level Phase II metabolism data, where sites were involved in a certain type of reaction and labeled positive for that class. As the result, each bond is predicted with four atom-level scores corresponding to four Phase II reaction types. Each of these score reflects the probability that the atom undergoes the corresponding type of reaction. Next, the top five atom-level scores for each reaction type (20 in total) and all molecule-level descriptors were presented to the molecule network, which yielded four molecule-level scores. The molecule network was trained using molecule-level Phase II metabolism data, where molecules were involved in each type of reaction classes and labeled positive for that type. Each molecule is predicted with four scores. Each of these score reflects the probability that the molecule undergoes one of the four corresponding types of Phase II biotransformation.

reactions. Outputs of the second hidden layer were aggregated on a per-molecule basis at the cross-talk layer of five nodes, corresponding to the four reaction types and the composite of all types. The cross-talk layer and the second hidden layer both fed into the site-level output layer. This output layer would generate four site-level scores corresponding to the four types of Phase II reactions for each atom. Site-level scores were aggregated on a per-molecule basis. Top 5 site-level scores for each reaction types (20 in total), in addition to molecule-level descriptors, were fed to the third hidden layer of 10 nodes. The third hidden layer fed to the molecule-level output layer which would generate four scores for each molecule. Each of these four scores represented the probability that the molecule was a Phase II type-specific (UGT, SULT, NAT, and TMPT) substrate.

The network was trained to produce accurate site-level and molecule-level scores. The weights of the model were calibrated during training using gradient descent on the cross-entropy error of the difference between the predicted and actual response values of each site/molecule using TensorFlow.[3]

## 3.4   Results and Discussion

The following sections examined the inner working of the Phase II model. First, we reported the model performance at the site level. Second, we assessed the accuracy at the molecule level. Third, we examined the accuracy at predicting enzyme specificity. Fourth, we examined the factors that are important to each target at the site level. Fifth, we examined the factors that are important to each target at the molecule level. The accuracy of the model was evaluated using a ten-fold cross-validation procedure. The data set was divided into ten groups with an equal number of substrates in each group. Predictions were generated for each tenth of the data by a model trained using the remaining nine tenths of the set. The resulting predictions covered the complete data set and were used to compute the accuracy metrics in this study.

### 3.4.1  Site Prediction Accuracy

Considering each reaction type separately, the Phase II model accurately predicts SOMs within metabolized molecules. Reaction-type specific performance is measured using "top-two" and "average site AUC". First, we measure the"top-two" performance on a type-by-type base, where a molecule is considered to be correctly predicted if any of its reported SOMs is ranked in the top two sites sorted by that type-specific prediction score. The top-two metric is commonly used to evaluate CYP site of metabolism models. The top-two metric is helpful for assessing individual drug candidates. Glucuronidation, sulfuration, acetylation, and methylation top-two accuracies are $96.5\%$, $95.3\%$, $94.4\%$, and $95.3\%$, respectively (Figure 3.4.1a, top). Second, we measure "average site AUC", a metric based on the area under the ROC curve (AUC). AUC is a standard metric employed in machine learning to quantify how well a model distinguishes between positive and negative test cases. For each reaction type, we compute the individual AUC of each molecule in the data set, and then averaging across the whole data set to produce a single summary statistic. Average-site AUCs for glucuronidation, sulfation, acetylation, and methylation are $98.8\%$, $98.6\%$, $97.6\%$, and $98.6\%$, respectively (Figure 3.4.1a, bottom).

The model also accurately differentiates metabolized and non-metabolized sites considering the entire data set. We use the "global-siteAUC" metric to evaluate this ability. For each reaction type, a ROC curve across all atoms in the data set is constructed (Figure 3.4.1b). The global-site AUCs for glucuronidation, sulfuration, acetylation, and methylation are $98.5\%$, $98.3\%$, $97.1\%$, and $98.0$, respectively.

Considering all reaction types together, the Phase II model accurately predicted type-specific SOMs within metabolized molecules. In contrast with the previous metrics, the model's ability to simultaneously predict the type and site is evaluated here. In this evaluation, each reaction-site pair is considered as an individual entity. All pairs are ranked by the scores, and compared with the known metabolism to verify if observed SOMs are separated from the rest. Considered all reaction types together, the model has an average top-two, average-site AUC, and global-site
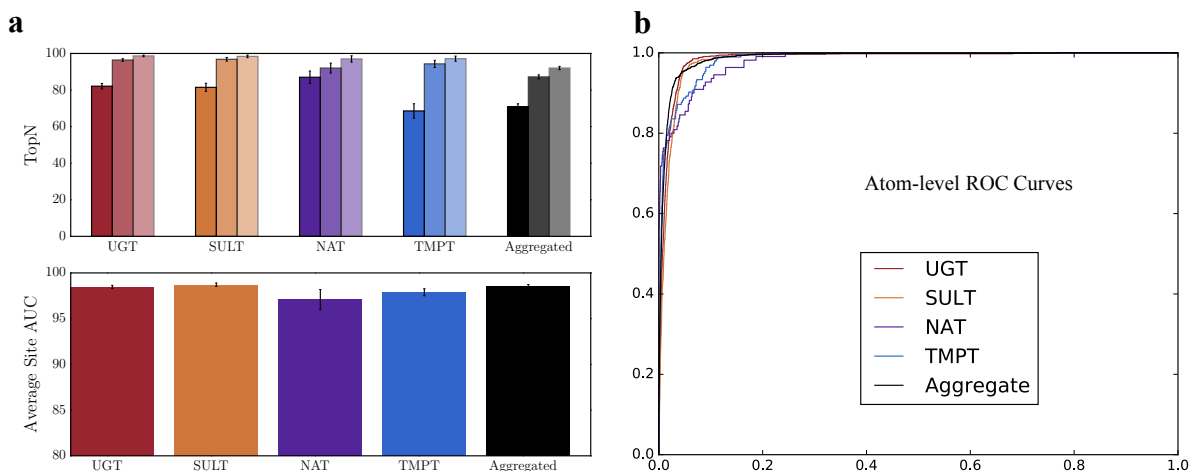
**Figure 3.4.1: Phase II Model Accurately Predicts Site-Level Metabolism**. Considered each reaction type separately, the Phase II model accurately predicted SOMs within metabolized molecules. Glucuronidation, sulfuration, acetylation, and methylation top-two accuracies were, respectively, $96.5\%$, $95.3\%$, $94.4\%$, and $95.3\%$. The global AUC for glucuronidation, sulfuration, acetylation, and methylation were $98.5\%$, $98.3\%$, $97.1\%$, and $98.0\%$, respectively. Similar results were observed considering all reaction types together. Considered all four reaction types together (Aggregated–black), the model had an average cross-validated top-two accuracy and global AUC of $95.3\%$ and $98.6\%$, respecitively.

AUC of $95.3\%$, $98.2\%$, and $98.6\%$, respectively.

The site-level output of the model is a well-scaled probability. When we binned atom-reaction pairs by prediction score, the proportion of sites of metabolism in each bin strongly correlates with the bin's score (Figure 7.4.2, $R^2 = 0.947$). Thus, the model's site-level output is interpretable as the probability and can be combined with other probabilistic outputs.

### 3.4.2 Molecule Prediction Accuracy

Considering each reaction type separately, the Phase II model accurately predicts which molecules are metabolized. The global AUCs at the molecule level are $80.9\%$, $77.2\%$, $89.4\%$ and $85.2\%$ for UGT, SULT, NAT and TPMT, respectively. In addition, the Phase II model correctly predicts the type of reaction for each site. The aggregated global AUC score is $90.9\%$ (Figure 3.4.3).

The molecule-level output of the model is a well-scaled probability. When we binned atom-reaction pairs by prediction score, the proportion of sites of metabolism in each bin closely
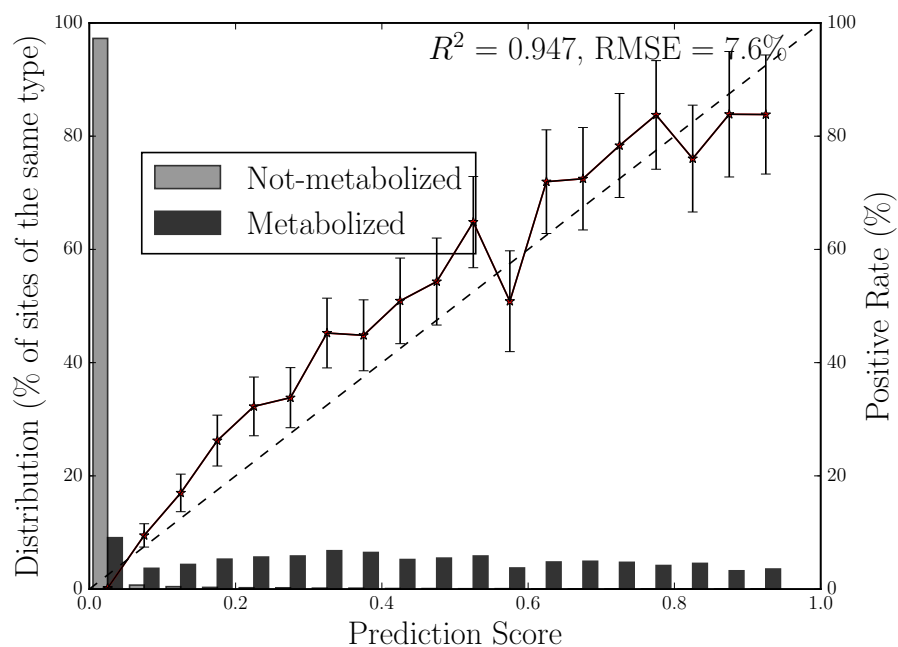
**Figure 3.4.2: The model makes well-scaled site-level predictions, corresponding to the probabilities.**
Phase II site-level score has a strong correlation to a perfectly scaled prediction ($R^2$ value of 0.947 and RMSE of 7.6%). This means that the score is interpretable as the probability that an atom undergoes one of the four Phase II reaction types. The bar graph plots the distribution of scores across reaction-type-specific metabolized and non-metabolized heavy atoms. The solid line plots the percentage of metabolized atoms in each bin. The diagonal dashed line indicates a hypothetical perfectly scaled prediction. Reliability diagrams for each reaction type are in the Supplemental Materials.
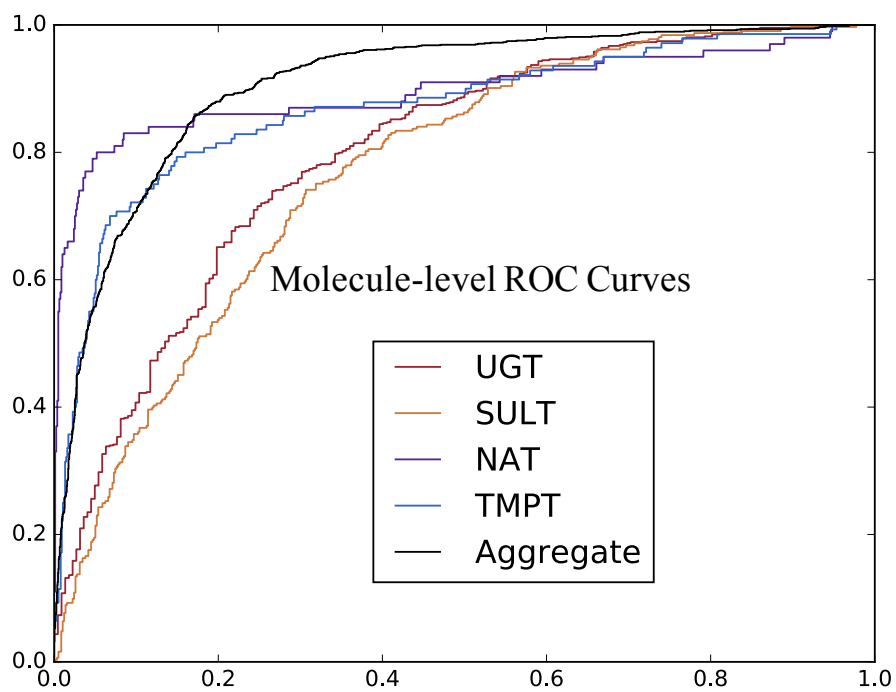
**Figure 3.4.3: The model accurately predict molecule-level metabolism.** Considering each reaction type separately, the Phase II model accurately predictes which molecule is metabolized. The global AUCs at the molecule level are $80.9\%$, $77.2\%$, $89.4\%$ and $85.2\%$. Considering all targets together, the global AUC score is $90.9\%$.
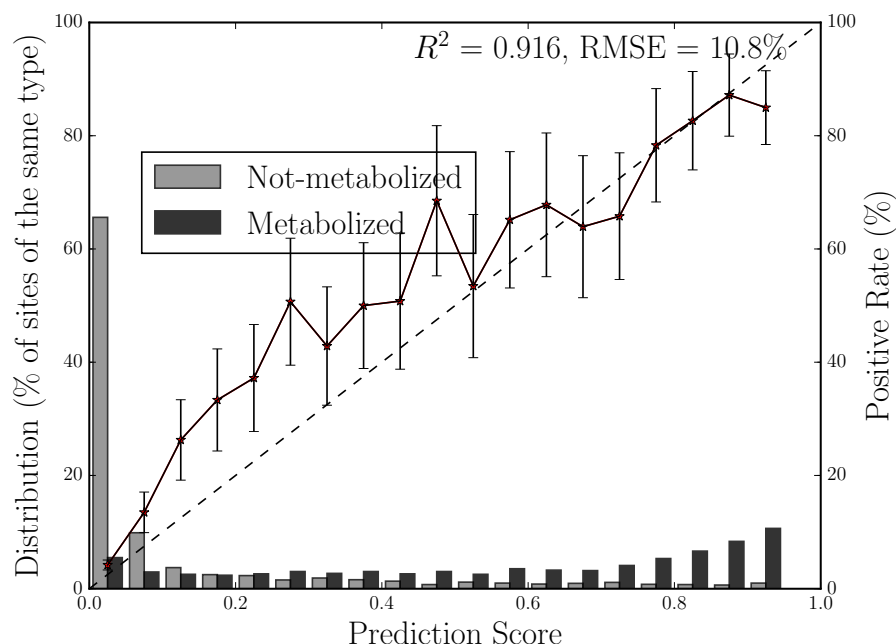
**Figure 3.4.4: The model makes well-scaled molecule-level predictions, corresponding to probabilities.**
Phase II molecule-level score has a strong correlation to a perfectly scaled prediction ($R^2$ value of 0.916 and
RMSE of 10.8%). This means that the score is interpretable as the probability that a molecule is metabolized
by a modeled Phase II reaction. The bar graphs plot the distributions of scores across 35,000 reaction type-
specific metabolized and non-metabolized molecules (8750 for each reaction type). The solid lines plot the
percentage of molecules that undergoes one of the four modeled Phase II reactions in each bin. The diagonal
dashed lines indicate a hypothetical perfectly scaled prediction. Reliability diagrams for each of the reaction
type are in the Supplemental Materials.

correlates with the bin's score (Figure 7.4.2, $R^2 = 0.916$). Thus the model's molecule-level output

is interpretable as a probability and can be combined with other probabilistic outputs.

### 3.4.3   Enzyme Specificity

Our model can discriminate pairwise enzyme specificity at both site- and molecule-levels with

high accuracies. We measure this ability through the "signed" and "unsigned enzyme specificity"

metrics. The "signed enzyme specificity" measures how often the model assigns higher score to

the metabolizing enzyme than the non-metabolizing enzyme among pairs with different

metabolism status. The "unsigned enzyme specificity" measures how often the model assign pairs

of scores with larger differences for differentially metabolized pairs than those of identically

metabolized pairs. The average pairwise site- and molecule-level signed enzyme specificity are 90.4% and 90.6%, respectively. The average site- and molecule-level pairwise unsigned enzyme specificities are 98.0% and 77.7%, respectively (Figure 3.4.6). Overall, the model is best at differentiating NAT metabolism from that of other enzyme families.

Our model could predict which enzyme(s) of the four responsible for metabolism at both site- and molecule-level with high accuracies. For each metabolized site/molecule, we constructed a ROC curve from its four metabolism scores and corresponding metabolism statuses. The "average enzyme-specificity AUC" reflects the model ability to accurately predict enzymes specificity for sites and molecules. The average enzyme-specifity AUC at site- and molecule-levels were 90.6% and 90.1%, respectively.

### 3.4.4 Descriptors Driving Accuracy

A permutation sensitivity analysis identifies the descriptors driving model accuracy [107]. We started with the trained model. Next, we randomly permuted each descriptor column (or group of closely related descriptors) in the input data for these molecules. The trained model was applied to the permuted data, and the performance drop across all the molecules was recorded. The higher the performance drop, the more important the descriptor to the model's performance.

We saw similar results using all performance metrics (Figure 3.4.8). Using the aggregate Global AUC performance as a guide, this analysis identified topological descriptors (identities of the atom and its neighbors (atoms one, two and three bond away), number of bound hydrogens and heavy atoms and size of ring containing the atom) and heuristic descriptors (heuristic score and number of substructures) as the most important descriptors for differentiating metabolized sites from non-metabolized sites. The result revealed that XenoSite heavily relies on local topology for calculating SOM score, and does not need descriptors from the quantum simulation. Similar results were seen in the substructure-specific sensitivities. Here, once again, topological and heuristic descriptors were the most important, and no quantum chemical descriptors were

**Figure 3.4.5: Our Phase II Model Have High Enzyme Specificity**. The "signed enzyme specificity" measures how often the model assign higher score to metabolizing enzyme than non-metabolizing enzyme among pairs with different metabolism status. The "unsigned enzyme specificity" measures how often the model assign pairs of scores with larger differences for differentially metabolized pairs than those of identically metabolized pairs. The average pairwise site- and molecule-level signed enzyme specificity are $90.4\%$ and $90.6\%$, respectively. The average site- and molecule-level pairwise unsigned enzyme specificity are $98.0\%$ and $77.7\%$, respectively.
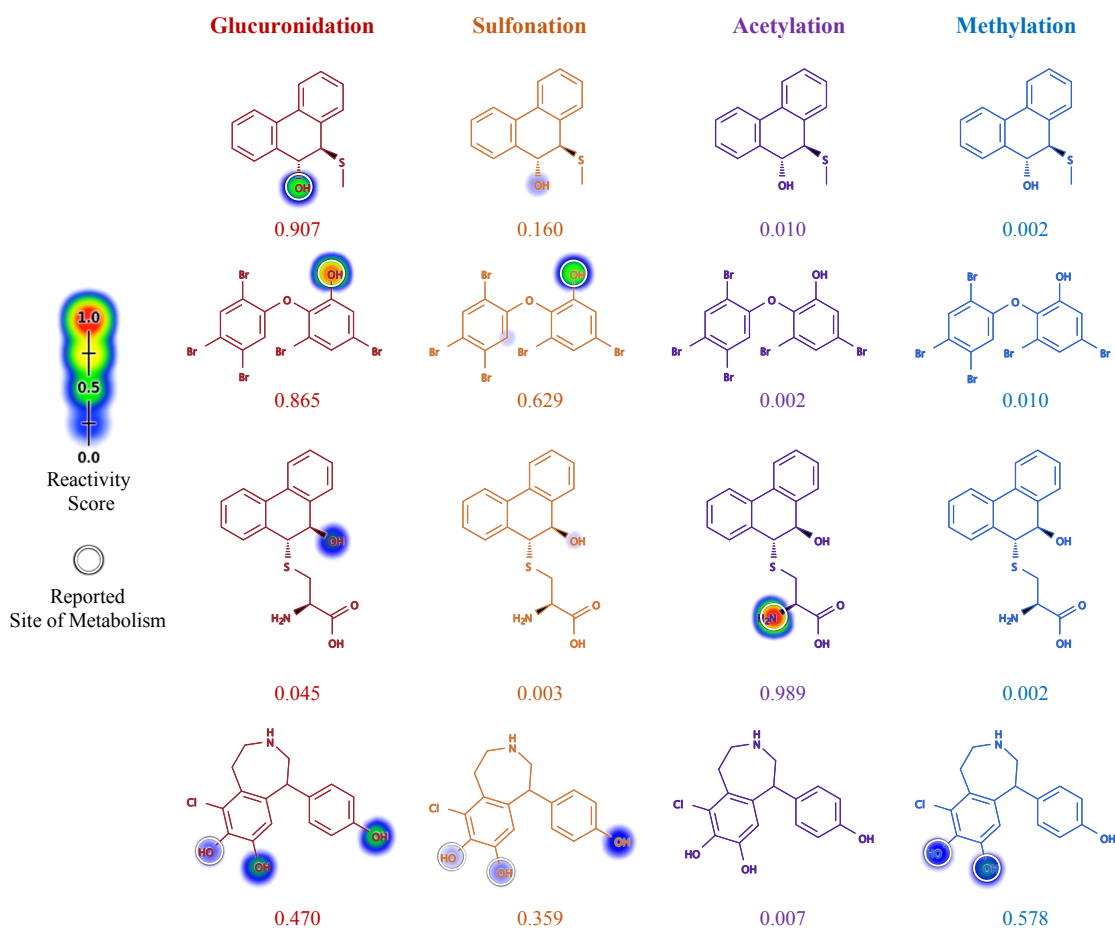
**Figure 3.4.6: Sample Predictions**. Four example compounds (row-wise) with ten-fold cross-validated reaction-specific (column-wise) predictions are shown. Experimentally reported sites of metabolism are circled. Site-level reaction-specific predictions are represented by the glow at each atom. Molecule-level reaction-specific predictions are the number under each molecule structure. XenoSite Phase II model assigns SOMs with the highest site-level scores in each molecule and occurring reactions with the highest molecule-level scores.

necessary. At the same time, a few molecule-level descriptors (like logP) were also important. Notably, the two heuristic descriptors were consistently among the most important.

At molecule-level, a sensitivity analysis reveal that most important factors to predicting substrate specificity, with the exception of sulfonation (SULT) target, are top scores from the site-level. Top1 UGT score was the most important descriptor for glucuronidation (UGT), acetylation (NAT), and methylation (TPMT) targets. Numbers of hydrogen bonds within the molecules (HBD, HBA1, HBA2) and logP, correlated to the compound's water solubility, are also among the most significant descriptors.

## 3.5 Conclusion

This study introduces an accurate model to predict four major types of Phase II metabolism: glucuronidation, sulfation, acetylation, and methylation. Phase II XenoSite can predict SOM within metabolized molecule with an average top-two $95.3\%$ and substrate specifcity with an average molecule AUC of $85.2\%$. We will combine this model with other models to predict the complete metabolism profile of a novel compound.
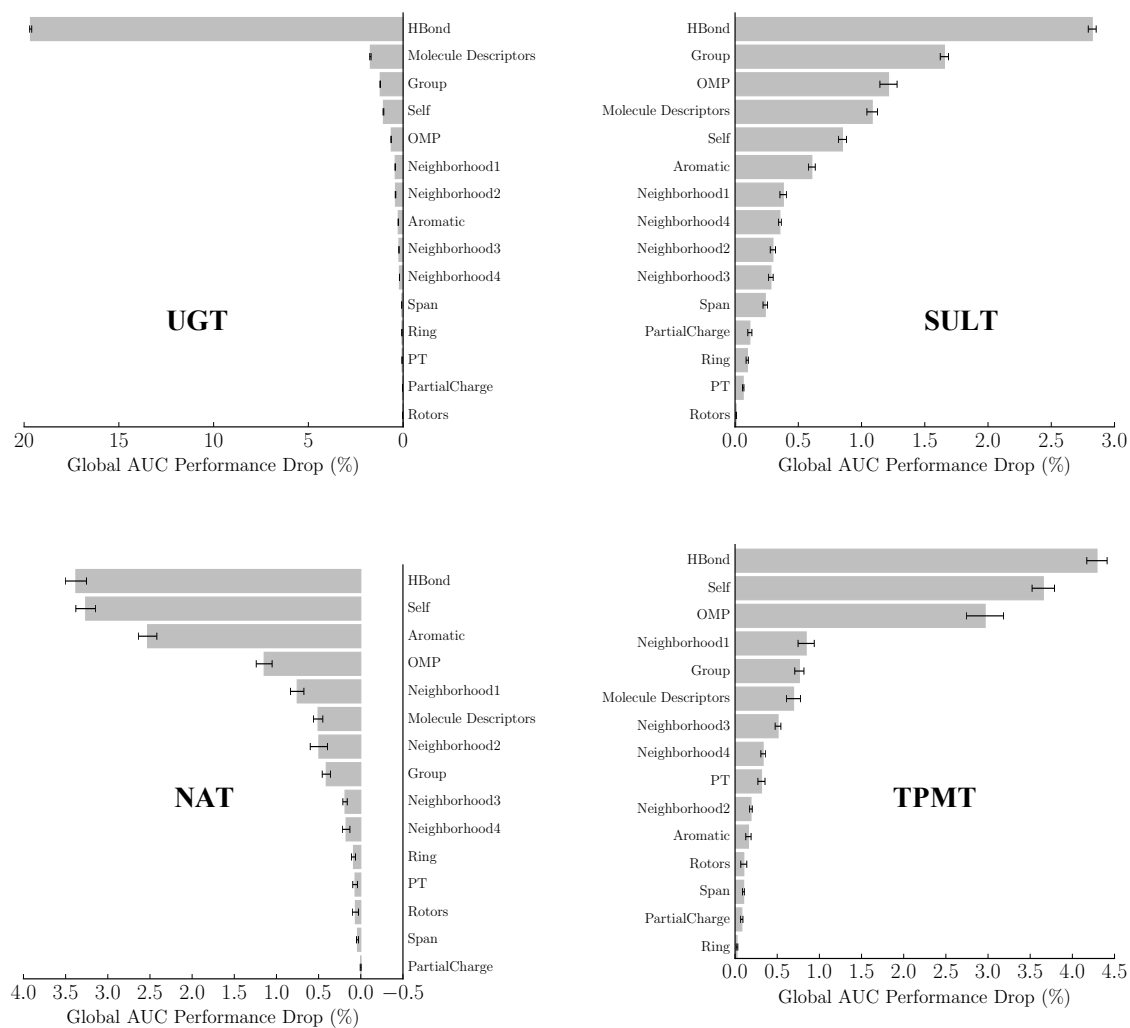
**Figure 3.4.7: The importance of specific descriptors to predicting site specificity.** A permutation sensitivity analysis quantified the importance of descriptors for the Phase II model to differentiate metabolized from non metabolized molecules of each reaction type. The 10 most important descriptors are plotted for each target. The number of hydrogen bonds of the site is always the most important descriptors for predicting Phase II SOM. Glucuronidation (UGT) target are most sensitive to numbers of hydrogen bonds.
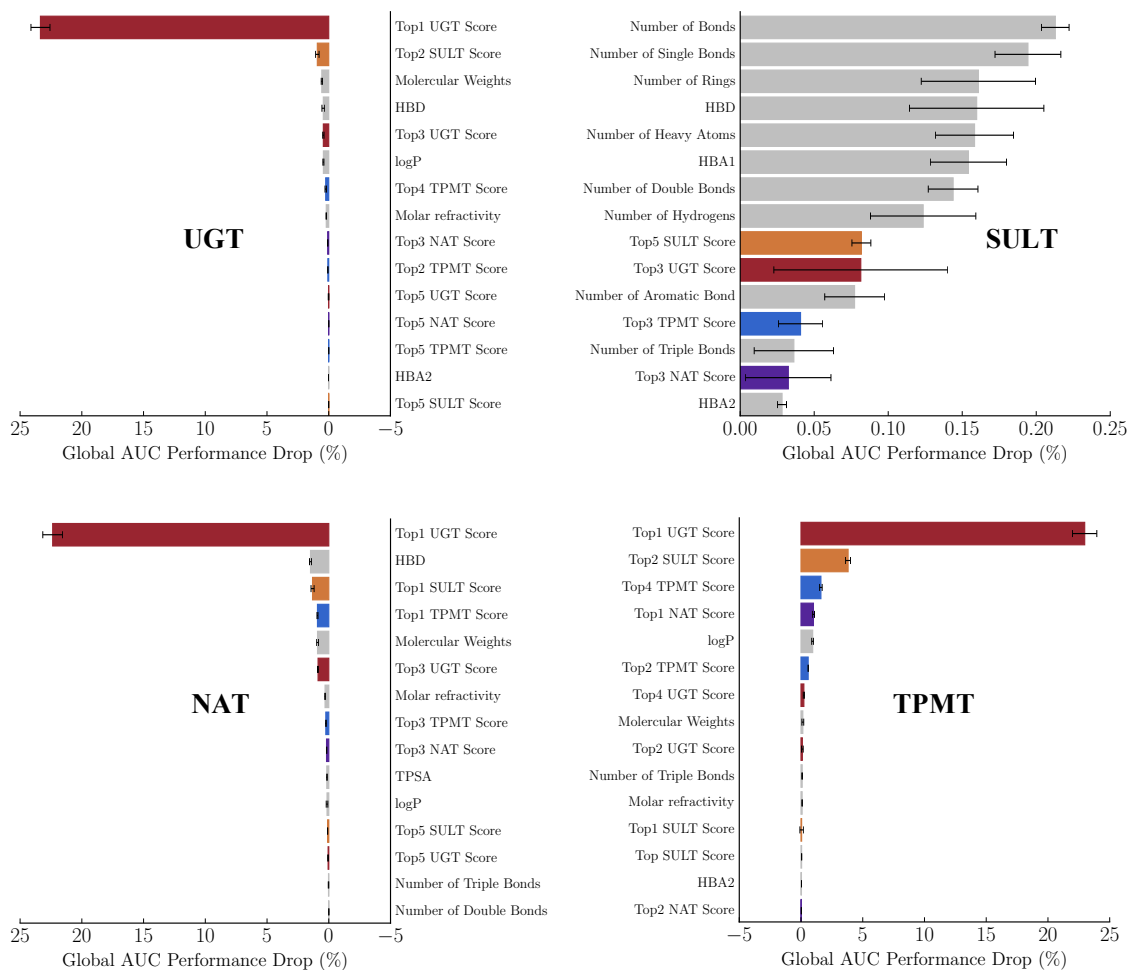
**Figure 3.4.8: The importance of specific descriptors to predicting molecule specificity.** A permutation sensitivity analysis quantified the importance of descriptors for the Phase II model to differentiate metabolized from non metabolized molecules of each reaction type. The 10 most important descriptors are plotted for each target. Top1 UGT was the most important descriptors for glucuronidation. Suprisingly, Top 1 UGT score was also the most important descriptor for acetylation and methylation. The magnitude of performance drop is 10 fold smaller for sulfonation than that of other target.

# 4

# Modeling Phase-I Metabolism.

This chapter is adopted from a manuscript submitted to ACS Central Science:

**Na Le Dang**, Tyler B. Hughes, Matthew K. Matlock, and S. Joshua Swamidass, *The Metabolic Rainbow: Deep Learning Phase 1 Metabolism in Five Colors*, **2018**

## 4.1    Summary

Metabolism of drugs their affects absorption, efficacy, excretion, and toxicity profiles. *In vitro* experiments with human liver microsomes and *in vivo* animal models are commonly used to predict the metabolism of a drug candidate in humans. Unfortunately, it is expensive to comprehensively test the metabolism of vast numbers of candidates during early stages drug

development. Moreover, *in vitro* experiments do not correspond with *in vivo* metabolism, so pathways giving rise to toxicity in patients are not always understood. Computational metabolism models can rapidly and cost-effectively predict which atoms or bonds of a molecule are modified during metabolism—site of metabolism—on thousands of drug candidates, thereby significantly reducing the number of experiments needed. However, current computational metabolism models are unable to predict what specific metabolic transformations would happen at certain sites. Identification of metabolism type is important for metabolite prediction, especially in the case of Phase I metabolism which includes many reaction types. Without knowledge of potential metabolite structures, medicinal chemists cannot differentiate harmful metabolic transformations from beneficial ones. To address this shortcoming, the current study modeled not only the site but also the type of human Phase I metabolism. We trained a neural network model on a literature-derived dataset of 20736 human Phase I reaction records. Among the covered reaction types were dealkylation, dehydrogenation, epoxidation, hydrolysis, hydroxylation, and reduction. Our model differentiated between these metabolic reactions with cross-validated accuracy of 97.1% area under the curve. The model developed in this study is available at http://swami.wustl.edu/xenosite.

## 4.2   Introduction

Metabolism has a significant implication on the absorption, distribution, efficacy, excretion, and toxicity profiles of pharmaceutical agents. In this study, we focused on human Phase I metabolism by cytochrome P450s (P450), other oxidoreductases, and hydrolases. These enzymes are responsible for the majority of xenobiotic metabolism. For example, P450s alone participate in the metabolism of more than 75% of FDA approved drugs.[249]

Phase I enzymes like P450 catalyze a wide variety of substrates, forming a diverse range of metabolites. [85] Phase I metabolism by these enzyme can introduce or reveal a functional group such as a hydroxyl, carboxyl, or amino group through a wide range of chemical transformations. Examples of Phase I reactions are hydroxylation, epoxidation, S- and N- oxidation, dealkylation,

deamination, dehalogenation, and dehydrogenation, among others.[60]

Although metabolism mainly detoxifies and deactivates xenobiotics, it can also bioactivate them into reactive metabolites.[135, 190] Reactive metabolites are electrophilic species that disturb cellular homeostasis through covalent/non-covalent interactions with macromolecules[100, 167, 177, 191] and can trigger cascades of toxic responses.[86, 113, 177, 231] Whether molecules are activated into toxic metabolites, or detoxified, often depends on the precise metabolic transformations they undergo in Phase I metabolism. For example, only some Phase I metabolites of chloramphenicol are reactive, and expected to be toxic (Figure 4.2.1).

There is substantial evidence that most immune-mediated hypersensitivity reactions—for example, idiosyncratic adverse drug reactions (IADRs) [71]—are caused by reactive metabolites of drugs. Approximately 62% of drugs either withdrawn from the market or labeled with a black box warning due to IADRs have been demonstrate to be metabolized into reactive metabolites [243]. For this reason, substantial effort is put towards determining which molecules are metabolized into reactive metabolites.

There are three existing approaches to model the complexities of Phase I sites of metabolism: (1) generic SOM prediction, (2) reaction-type specific prediction, and (3) multiple reaction-type prediction. Each of these three approaches has different strengths and weaknesses. However, for predicting bioactivation, all three are fundamentally limited. To address these limitations, a key innovation put forward is a "five color" classes for metabolic sites of metabolims, a metabolic rainbow. This is as an approach for building a rainbow SOM model that, based on context, can be unambiguously decoded into a much larger number of reaction-types.

### 4.2.1   Generic Metabolism Model

In the first approach, generic SOM prediction, models are trained to predict which atoms are modified by metabolism. Typically, atoms are labeled as metabolized or not metabolized, but in some cases bonds may be labeled.[207, 208, 210, 257, 258] In this approach, metabolism models
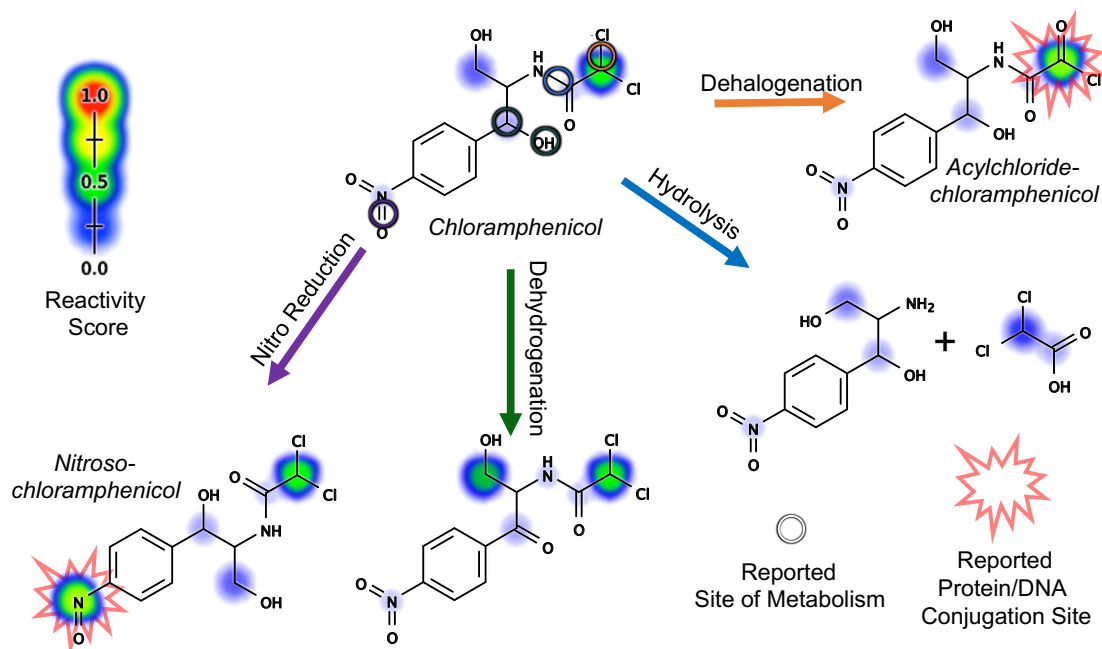
**Figure 4.2.1: Both site and type identifications are important for determining whether a metabolic transformation is beneficial or harmful.** Antibiotic chloramphenicol is known for its severe toxicities such as "gray baby syndrome" and bone marrow suppression.[74] These toxicities were attributed to the reactivity of both chloramphenicol and several of its Phase I metabolites, nitroso-chloramphenicol and acylchloride-chloramphenicol.[52, 83, 256] However, not all chloramphenicol's Phase I metabolites are reactive. Both metabolites generated through hydrolysis and dehydrogenation were reported and predicted by our reactivity model[106] to be inactive. This example illustrates the importance of knowing not only the site but the type of reaction in discriminating between beneficial and harmful metabolic transformations.

predict which atoms and bonds are metabolically modified within a given molecule (its sites of metabolism), but not the specific type of metabolic transformations take place at these sites. These models are very effective at identifying metabolically labile sites, have become widely used in industry as a guide in designing molecules with increased half life.

Many sites, however, have potential to be metabolized in multiple ways. In generic SOM modeling, the specific metabolic transformation, the specific reaction type, is not specified. Consequently, these models do not provide enough information to predict metabolites, and are unhelpful in modeling bioactivation pathways.

### 4.2.2   Single Reaction-Type Models

In the second approach, specific reaction types are modeled individually in separate models. For example, our group has published quinone formation, epoxidation, N-dealkylation, reduction, and sulfur-oxidation models.[57, 59, 103, 104] This approach has significant advantages. The nuances of each reaction type can be carefully considered in isolation. Expert knowledge can guide design of features, that can substantially improve accuracy. Curation of datasets of a uniform reaction-type encourages consistent annotation, and identify when data must be supplemented for reasonable performance. Most importantly, predictions are not ambiguous regarding the metabolic transformation, and give enough information to predict metabolite structures. This approach has been invaluable in studying bioactivation pathways, including modeling how metabolism makes some molecules with structural alerts toxic, but leaves others inert.[57, 59]

However, this approach is limited in several key respects. First, each model is trained on entirely different datasets, so their outputs are not scaled to predict the relative likelihood of different reaction-types. Second, information common to different reaction-types is not exploited to improve accuracy. Third, only a minority of reaction-types have been modeled this way.

### 4.2.3 Multiple Reaction-Type Models

In the third approach, a single model is used to simultaneously a large number of Phase I reactions.[97, 261] Only a few attempts have been made in this, all were very far from comprehensive. As we will see, these models only cover 42-48% Phase I metabolism and leave out several reaction types that frequently generate reactive metabolites like epoxidation[130] and quinone formation.[214] Moreover, it is difficult for users to interpret these models' predictions because they are not probabilistic. Detailed review of these models is in the Result and Discussion section. Those this approach might have potential, it has not useable to study bioactivation for these reasons.

A looming problem for multiple reaction-type models is the difficulty of visualizing and displaying results. As the number of reaction types proliferates, our ability to visualize and organize results into coherent views becomes more and more difficult. With a large number of reaction types in Phase I metabolism, the scalability of these approaches is in doubt.

### 4.2.4 Five Colors of Metabolism

The central innovation put forward here is a "five color" classification of reaction types into a metabolic rainbow (Table 4.2.1). This is as an approach for building SOM models that predict just five different classes of reactions taking place on bonds, which can, based on a bond's context, be unambiguously decoded into a much larger number of reaction-types. In this framework, bonds are labeled, not atoms, and they are labeled by color. Depending on the precise substructure the label appears and the color of its reaction, SOM sites can be decoded into one of 21 reaction-types.

Two design goals guided classification of reaction types into into five colors. First, knowledge of the bond and color of a reaction, should almost always be sufficient to determine the expected metabolite. Second, semantically related reaction types should be grouped in to the same color. Notably, in this case, bonds are labeled, not atoms. Moreover, bonds between heavy atoms and hydrogen are included too, as are pseudo-bonds between heavy-atoms and lone pairs. Redundancy is reduced by training on only one bond from each topologically equivalent group.
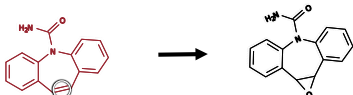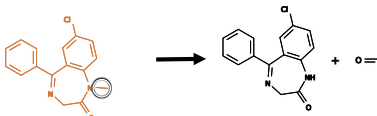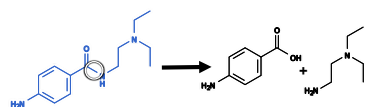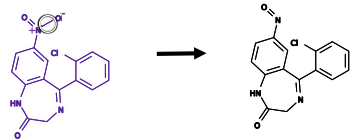
This conceptual innovation might overcome several limitations of prior approaches. First, in using five colors, predictions corresponding to 21 reaction types can be displayed on just five copies of a molecule, each one corresponding to a different color. Second, because training is done jointly across all reaction-types and colors at the same time, information can be shared across reaction types, which reduces the total number of trainable weights and may enable modeling of the reaction types for which less data is available. Third, by using the same molecules to train all reaction-types, the model outputs are expected to predict the relative probabilities between different reaction-types with greater accuracy.

## 4.3    Materials and Methods

### 4.3.1    Five Colors of Metabolism

Using the five color class approach, we identified 21 reaction-types, and grouped them into five colors: red (stable oxygenation), orange (unstable oxygenation), green (dehydrogenation), blue (hydrolysis), and purple (reduction) (Table 4.2.1).

The first color is red, for stable oxygenation. In a red reaction, an oxygen is added to the substrate and resulted a single observable metabolite. Red reactions include aromatic/aliphatic hydroxylation, epoxidation, nitrogen oxidation, and sulfur oxidation (Table 4.3.1). The second color is orange, for unstable oxygenation. In an orange reaction, addition of an oxygen to the substrate destabilizes one of the molecule's heavy bond (bond connecting between two non-hydrogen atoms). Eventually this bond breaks and gives rise to two metabolites, one contains the added oxygen and the other does not. Orange reactions include nitrogen-, oxygen-, sulfur-, and carbon-dealkylation, and oxidative dehalogenation (Table 4.3.2). The third color is green, for dehydrogenation. In a green reaction, hydrogen(s) is extracted to form double bond, triple bond, iminium, or quinone. Green reactions include double/triple bond formation and quinone/imine/methide formation (Table 4.3.3).The forth color is blue, for hydrolysis. In a blue reaction, addition of water molecule causes amide-, ester-, or ether-bond breakage and gives rise

| Color | Definition | Reaction Types | Number | |
|---|---|---|---|---|
| | | | Molecule | Site |
| Red | Stable Oxygenation (SO) | aromatic hydroxylation<br>aliphatic hydroxylation<br>aromatic epoxidation<br>aromatic epoxidation<br>N-oxidation<br>S-oxidation | 3098 | 10280 |
| Orange | Unstable Oxygenation (UO) | N-dealkylation<br>O-dealkylation<br>S-dealkylation<br>C-dealkylation<br>oxidative deamination<br>oxidative dehalogenation | 3749 | 5811 |
| Green | Dehydrogenation (DH) | dehydrogenation<br>alcohol to aldehyde or ketone<br>single- to double-bond<br>double- to triple-bond<br>quinone formation<br>iminium formation | 762 | 2794 |
| Blue | Hydrolysis (HD) | ester hydrolysis<br>amide hydrolysis<br>ether hydrolysis<br>cyanide hydrolysis | 3188 | 3969 |
| Purple | Reduction (RD) | carbonyl reduction<br>nitro reduction<br>sulfo reduction<br>reductive dehalogenation<br>hydrogenation | 1213 | 1590 |

**Table 4.2.1: Five Colors of Phase I Metabolism**

to two metabolites. Blue reactions include amide, ester, and ether hydrolysis (Table 4.3.4). The fifth color is purple, for reduction. In a purple reaction, an oxygen is removed or hydrogen(s) is added to the substrate. Purple reaction include nitro-, carbonyl-, and sulfo-reduction, reductive dehalogenation, and hydrogenation (Table 4.3.5).

We built the first model to produce probabilistic predictions on both the site and the type of human Phase I metabolism for small, drug-like molecules. Our model was trained on a diverse dataset of literature-derived human Phase I reactions. The rainbow model was a multi-target, deep neural network model that computed site- and molecule-level vectors of prediction scores, each corresponding to the probability that the site or molecule being metabolized by a certain class of Phase I reactions.

The training dataset of our model contained 21 types of Phase I reactions, ranging from the most common types like hydroxylation and dealkylation to less common yet important reactive-metabolite generating reaction types like epoxidation and quinone formation. We grouped these diverse Phase I reaction types into five classes based the structural changes that they introduce to the substrates: stable oxygenation, unstable oxygenation, dehydrogenation, hydrolysis, and reduction. Detailed definitions of each class is in the Material and Method section.

While abundant data is available for Phase I reactions like hydroxylation and dealkylations, reactions such as quinone formation and cyanide hydrolysis have only a few examples in our dataset. Were individual models constructed for each reaction type, some of these models would be susceptible to ascertainment bias due to data limitations.[153]

4.3.2   Metabolism Training Data

We collected a chemically diverse dataset from the literature-derived Accelrys Metabolite Database (AMD). A total of 20736 human *in vitro* and *in vivo* records were gathered. Each record was composed of one or more chemical transformations. A diverse set of Phase I reaction types, including hydroxylation, dealkylation, deamination, dehalogenation, and hydrolysis, among others were present in our training dataset (Table 4.2.1). Our study accounted for $92.3\%$ of all

human Phase I metabolism in the AMD dataset. The remaining $7.7\%$ Phase I reaction types excluded from the current study, in order from most to least common, were tautomerization, isomerization, rearrangement, radical formation, hydration, deacylation, denitrogenation, and decarbonylation. Reactions starting from the same substrate were mapped onto a single training molecule using maximum common substructure matching. [39] Our final dataset contained $9674$ unique molecules. Under the AMD license agreement, we were not allowed to disclose the structures of molecules in the dataset. However, to enable rebuilding of our complete database and reproduction of our results, we provided all reaction and molecule AMD registry numbers in the Supporting Information.

### 4.3.3 Rainbow Sites of Metabolism

For each class, we implemented an automated labeling algorithm that uses the structures of substrate and product to identify the specific bond or atom, called site(s) of metabolism (SOMs), at which a metabolic reaction takes place. The specific labeling rules for each reaction type are shown in **??**. The final dataset had $10280$, $5811$, $2794$, $3869$, and $1590$ sites of stable oxygenation, unstable oxygenation, dehydrogenation, hydrolysis, and reduction, respectively (Table 4.2.1).

These labeling algorithms could identify SOMs for the majority of AMD reaction records, but they were not perfect. Each pair of reactant and product was inspected to identify mislabeled SOMs, and $15\%$ of all SOMs were identified as incorrect and manually corrected. As with other manual annotation processes, there were human errors in the labeling process. We estimated such errors by randomly selecting a subset of $200$ reaction records for further evaluation. For these sampled reaction records, multiple independent annotators repeated the manual inspection. Next, all annotators examined and discussed each reaction together to reach consensus on the correct labeling. Of the $200$ sampled reaction records, ten records had different annotations from the established correct labeling. Approximately $5\%$ of annotated reaction records contained errors.
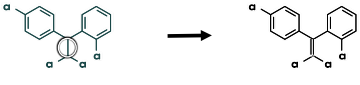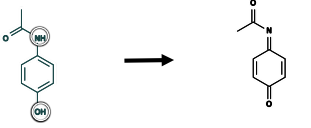
| Reaction Type | Example | SOM Labeling | Site |
|---|---|---|---|
| Epoxidation |  | Double/aromatic bond between two carbon atoms | 636 |
| Hydroxylation |  | Bond between a heavy atom and a hydrogen | 8475 |
| S-oxidation |  | Lone pair on surfur atom | 424 |
| N-oxidation |  | Lone pair on nitrogen atom | 665 |
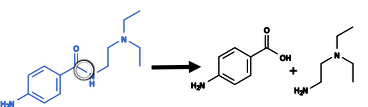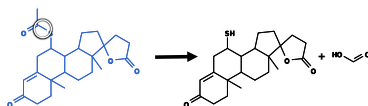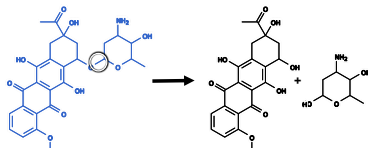
**Table 4.3.1: Red Reactions are Stable Oxygenations (SO).** The table shows the "red" reaction types, how each is marked on a molecule, and the number of times each appears in the data.

| Reaction Type | Example | SOM Labeling | Site |
|---|---|---|---|
| N-dealkylation |  | Bond between nitrogen and carbon | 2970 |
| O-dealkylation |  | Bond between oxygen and carbon | 1712 |
| S-dealkylation |  | Bond between sulfur and carbon | 93 |
| C-dealkylation |  | Bond between two carbon | 577 |
| Oxidative Dehalogenation |  | Bond between halogen and carbon | 309 |

**Table 4.3.2: Orange Reactions are Unstable Oxygenations (UO).** The table shows the "orange" reaction types, how each is marked on a molecule, and the number of times each appears in the data.

| Reaction Type | Example | SOM Labeling | Site |
|---|---|---|---|
| Double/triple bond formation | | Bond between heavy atom and hydrogen | 2741 |
| Quinone/ Imine/Methide formation | | Bond between heavy atom and hydrogen | 90 |
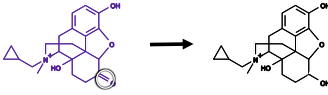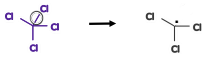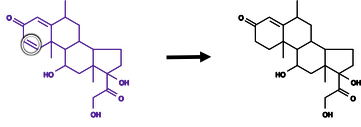
**Table 4.3.3: Green Reactions are Dehydrogenations (DH).** The table shows the "green' reaction types, how each is marked on a molecule, and the number of times each appears in the data.

| Reaction Type | Example | SOM Labeling | Site |
|---|---|---|---|
| Amide hydrolysis | | Amide bond | 1118 |
| Ester hydrolysis | | Ester bond | 2112 |
| Ether hydrolysis | | Ether bond | 47 |

**Table 4.3.4: Blue Reactions are Hydrolyses (HD).** The table shows the "blue" reaction types, how each is marked on a molecule, and the number of times each appears in the data.

### 4.3.4   Identification of Possible Sites of Metabolism

Each type of metabolic reaction occurs at some but not all sites. For example, epoxidation only occurs at double and aromatic bonds while hydrolysis takes place at amide, ester, and ether bonds. These sets of potential sites could be used to heuristically estimate where in the molecule a specific reaction happens. To assess our model's ability to predict site of metabolism for individual reaction types within their specific domain, we identified all reaction type-specific potential sites (Table **??**) using SMARTS patterns. [63] Of 691349 bonds and lone pairs in our dataset, 461876 (66.81%) sites were identified as potential sites for at least one of the covered

65

| Reaction Type | Example | SOM Labeling | Site |
|---|---|---|---|
| Nitro reduction |  | Bond between nitrogen and attached oxygen | 313 |
| Carbonyl reduction |  | Carbonyl bond | 869 |
| Sulfo reduction |  | Bond between sulfur and attached oxygen | 46 |
| Reductive dehalogenation |  | Bond between halogen and attached carbon | 2 |
| Hydrogenation |  | Double, triple bond between pair of hydrogenated atoms | 1435 |

**Table 4.3.5: Purple Reactions are Reductions (RD).** The table shows the "purple" reaction types, how each is marked on a molecule, and the number of times each appears in the data.

Phase I reaction types. On average, $99.97\%$ of class-specific labeled SOM were in the corresponding set of potential sites. Class-specific labeled SOM that did not fall within the corresponding domain for their class were in fact mislabeled sites.

## 4.3.5   Descriptors

To predict the susceptibility to metabolism of each bond/lone pair in the dataset, our model used structural, physical and chemical information encoded in numerical descriptors. Each bond was represented by a vector of 404 descriptors that describe its properties at atom, bond and molecule levels. All of these descriptors are calculated by in-house software from 2D Open Babel SDF files.[186] The majority of the descriptors used in this project are atom-level descriptors (e.g. atom identity, charge, hybridization) previously developed for XenoSite metabolism models, as described in previous studies from our group.[104, 105, 258] For each bond, there were 189 descriptors for each atom on either side of the bond (Table **??**), 10 bond descriptors (e.g. bond

type, bond length) (Table **??**) and 16 molecule descriptors (e.g. molecular weight, topological polar surface area, molar refractivity) (Table **??**). The complete list of descriptors used in this study is provided in the Supporting Information Tables **??**, **??**, and **??**.

### 4.3.6    Multitarget Deep Neural Network Model

Our Phase I model was a multitarget, deep convolutional neural network with one input layer, four hidden layers, and two output layers (Figure 7.3.2). Multitask neural networks can reuse features learned for one class of reactions to predict another, data-limited class, and can significantly reduce the number of model parameters required. [53, 106, 189]

   The top output layer computed molecule prediction vectors, each contained five scores corresponding to the five classes of Phase I reactions. Each of these molecule prediction scores correlated to the probability that a given molecule was observed to be metabolized via the corresponding reaction class. Similarly, the bottom output layer computed bond prediction vectors, each contained five scores. These scores correlated to the probabilities that a given bond/lone pair was an actual site of metabolism for each of the five reaction classes.

   The network was trained in two stages. First, the bond-level network was trained to produce accurate bond-level scores. For an input molecule, atom-level, bond-level, and molecule-level descriptors were calculated for all bonds and lone pairs. These descriptors were inputs to a neural network with two hidden layers (each contains 25 nodes, ReLU activation) that generated vectors of five scores corresponding to the five classes of Phase I reactions. Each element in the vector represented the probability that a bond/lone pair was a Phase I reaction-class SOM. The weights of the model were calibrated during training by performing gradient descent on the cross-entropy error (with L2 penalty of 0.3) of the difference between the predicted and actual response values of each bond using TensorFlow.[3] Second, the molecule-level network of two hidden layers (each contains 5 nodes, ReLu activation) was trained. In this training step, each row of the data matrix represented a molecule, and each column was descriptor. Descriptors included the top five bond-level scores for each reaction class (25 in total), and all molecule-level descriptors. Similar

to the first stage, we trained the weights of the networks using gradient descent on the cross-entropy error (with L2 penalty of 0.3) so that molecules observed being metabolized by certain class of reaction received higher score for that class than other molecules.

The accuracy of the model was evaluated using a ten-fold cross-validation procedure. The dataset was divided into ten groups with an equal number of substrates in each group. Predictions were generated for each tenth of the data by a model trained using the remaining nine tenths of the set. The resulting predictions covered the complete dataset and were used to compute the accuracy metrics in this study.

### 4.3.7   Visualizing Metabolism in Colors

Specifying the five colors more precisely, each was mapped to colorblind safe hex codes. [128, 252] Rainbow coloring is often though inappropriate for visualization. [28, 138] In this case, however, it is appropriate because there are no natural ordering to the reaction classes to which the color is keyed. Moreover, the rainbow shading for the predictions also includes a size cue, which immediately clarifies scale and ordering of colors, though a perceptually uniform shading might be important to consider in future work. [31]

In displaying metabolism data, the 2D depiction of the molecule is always used, but it is annotated with both (1) circles to denote each atom and bond that are metabolized and (2) background shading that denotes with color and size the strength of predictions at each atom and bond. Usually, the 2D depiction of the molecule is black and the annotation circles are white with a black shadow. Here, we use color in the 2D depiction and the circles to indicate reaction color.

For the most part, SOMs are visualized exactly as used for modeling, and described in the prior section, but in some cases adjustments are made. For a bond-between-hydrogen-and-heavy-atom SOM, the circle is displayed at the corresponding heavy atom. Similarly, a lone-pair SOM is visualized at the bearing atom. Likewise, prediction for a bond-between-hydrogen-and-heavy-atom or a lone-pair is visualized at the corresponding heavy atom. These modifications allow for displaying all SOMs and predictions on the molecule 2D
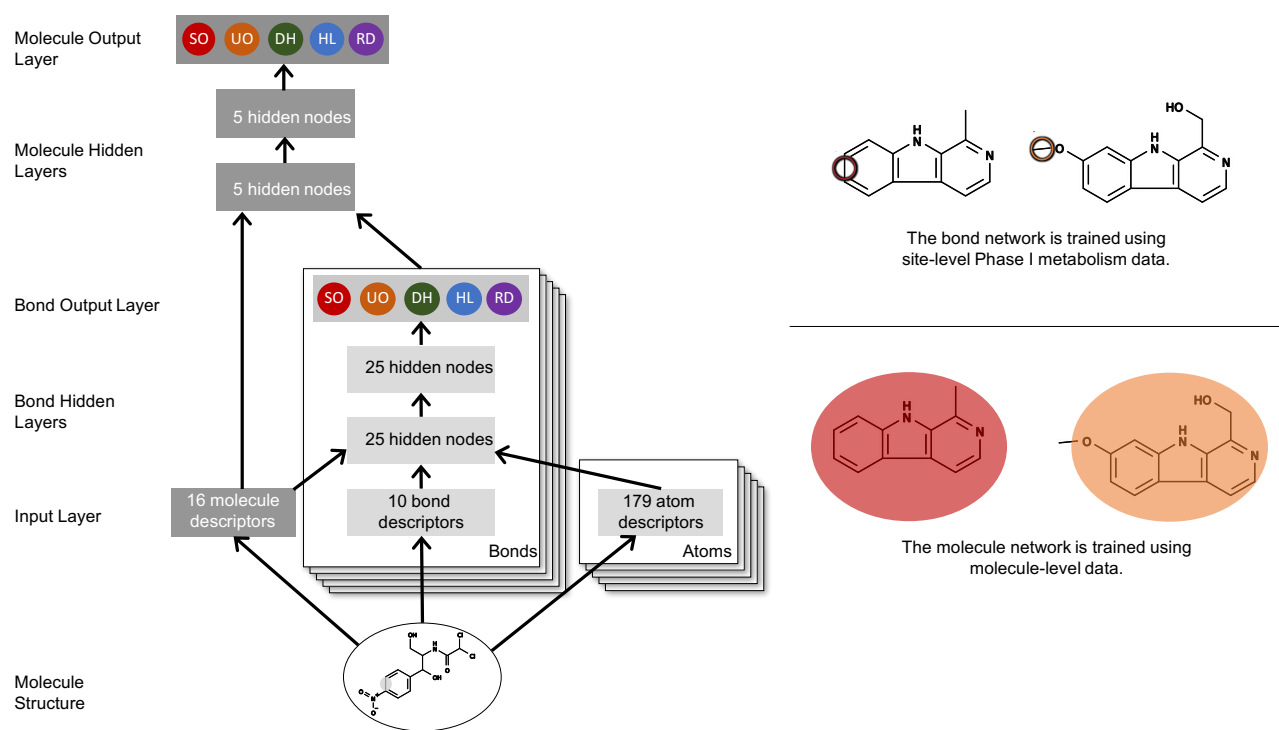
**Figure 4.3.1: The Structure of the Phase I Reaction Model**. (**Left**) The diagram on the left shows how information propagated through the neural network model, which contained one input layer, two hidden layers, and two output layers. From the structure of an input molecule, 16 molecule-level, 10 bond-level and two sets of 179 atom-level descriptors were calculated and fed to the input layer. Both molecule and bond level descriptors were submitted to the bond neural network, which computed five bond-level scores, corresponding to the five Phase I reaction classes, for each bond in the input molecule. The bond network was trained using site-level Phase I metabolism data, where sites involved in a certain class of biotransformation were labeled positive for that class. As the result, each bond is predicted with five bond-level scores correspond to five Phase I reaction classes. Each of these score reflects the probability that the bond undergoes the corresponding class of biotransformation. Next, the top five bond-level scores for each reaction class (25 in total) and all molecule-level descriptors were presented to the molecule network, which yielded five molecule-level scores. The molecule network was trained using molecule-level Phase I metabolism data, where molecules involved in each type of reaction classes were labeled positive for that class. Each molecule is predicted with five scores. Each of these score reflects the probability that the molecule undergoes one of the five corresponding classes of Phase I biotransformation. In the diagram, color circles are predictions ranging from 0 to 1; bars are vectors of real numbers. The five classes of Phase I metabolism covered in this studies are stable oxygenation (SO), unstable oxygenation (UO), dehydrogenation (DH), hydrolysis (HD), and reduction (RD). (**Right**) Examples from the data used in the two sequential training stages, on bond and molecule neural network, are shown. Positively labeled stable oxygenation bond and molecule are circled.

depiction that only contains heavy atoms.

### 4.3.8   Prediction Aggregation and Scaling

To improve visualization, the predictions were, at times, aggregated and scaled. The highest (summarized) visualization described in the prior section contains aggregated and scaled predictions. First, each atom and bond is assigned with the probabilistic-OR of its five color-prediction scores. Next, predictions at both the summarized and reaction-class levels are scaled by other predictions on the molecule by molecule and level by level basis: $\rho_x = p_x/(1 - \prod_{i=1}^{n}(1 - p_i))$, where $n$ is the number of prediction scores at the same level in the molecule to which prediction is scaled against. These predictions are visualized with a rainbow and size based color map that shades the background behind each molecule (Figure 4.4.6).

The lowest (reaction-type level) visualizations described in the prior section contains scaled predictions. The predictions for a reaction type are displayed only at its possible sites, and are the scaled prediction of the corresponding class reaction. Because some of reaction types are rare, their predictions are low and the corresponding color on the visualization scale are difficult for human eyes to differentiable. To improve the visibility of probabilistic differences between sites of the same reaction type, second, we scaled these prediction using the following equation: $\rho_x = \sigma(logit(p_x) - \mu_{logit(p)})$ where $\mu_{logit(p)}$ is the mean *logit* of prediction across all possible sites. Predictions at the reaction-class level scaled by the applying the same equation are also included. These predictions are visualized using the red-blue color scheme, where red indicates probabilities higher than average and blue indicates probabilities lower than average (Figure 4.4.6).

## 4.4   Results and Discussion

The following sections study a rainbow SOM deep neural network model of Phase I metabolism. This model was trained on a large training set of molecules, covering observed metabolic transformations in human liver microsome, recombinant isozyme, and in human *in vivo* studies. All Phase I metabolism was included, as is detailed in the Materials and Methods. First, we

compared the reaction coverage and properties of our Phase I model to published SOM prediction models. Second, we assessed the accuracy of our metabolism model in predicting both sites of metabolism with metabolized molecules, and which molecules are metabolized. Third, we evaluated the accuracy of the model in discriminating which molecules with specific structural alerts are bioactivated.[121, 226] Finally, we discussed the limitations of our approach and suggest future studies to improve toxicity risk assessment.

### 4.4.1 Reaction Coverage Comparison

We compared our rainbow model (Rainbow Xenosite) to previously published Phase I SOM prediction models: SMARTCyp,[210] RSPredictor,[257] SOMP,[207, 208] Site of Metabolism Estimator (SOME),[261] He et.al.,[97] and single reaction type models.[57, 59, 103, 104]. Our metabolic models, the rainbow and single reaction type models, are the only ones that produce well-scaled, probabilistic outputs (Figure 7.4.2). More importantly, Rainbow XenoSite cover sthe largest proportion of known human Phase I metabolic reactions of all Phase I models (Table 2): While previously published models cover up to $48\%$ of Phase I reactions, our rainbow model covers $92.3\%$ of AMD Phase I reactions. The large difference in reaction coverage makes direct comparison between Rainbow XenoSite and previously published models difficult.

SMARTCyp,[210] Xenosite 1.0/RSPredictor,[257, 258] and SOMP,[207, 208] are three published generic SOM Phase I prediction models. SMARTCyp, XenoSite 1.0/RSPredictor, and SOMP account for, respectively, $45.5\%$, $48\%$, and $46.1\%$ of all human Phase I reactions from the AMD database (Table **??**).

We published several models that focused on specific reactions like epoxidation,[104] quinone species formation,[103] N-dealkylation [59], S-oxidation, and N-reduction.[57] These reactions play important roles the bioactivation of multiple structural alerts leading to the formation of deleterious reactive metabolites.[57**?** ] Given a molecular structure, each of these models output probabilistic prediction scores corresponding to the likelihood that each atom or bond in the molecule is a site of metabolism for that specific reaction. Together, all these single reactions

71

account for $21.0\%$ of all Phase I reaction (Table 2).

Site of Metabolism Estimator (SOME)[261] and He et.al.[97] are two published multi-reaction Phase I prediction models. SOME and He et. al. account for $42.6\%$ and $37.6\%$, respectively, of all human Phase I reactions from the AMD database (Table 2).

Table 2. Rainbow Model had improved reaction coverage in comparison to previously published models.

| | Rainbow Model | SMARTCyp[20] | RSPredictor[21] | SOMP[23,24] | Combined SRMs[25-28] | SOME[29] | He et. al.[30] |
|---|---|---|---|---|---|---|---|
| Phase I Coverage | 92.3% | 45.5% | 48.0% | 46.1% | 21.0%[a] | 42.6% | 37.6% |
| Red | Ar-hydroxylation[b]<br>Al-hydroxylation[c]<br>N-oxidation<br>S-oxidation<br>P-oxidation<br>epoxidation | Ar-hydroxylation<br>Al-hydroxylation<br>N-oxidation<br>S-oxidation<br>epoxidation | Ar-hydroxylation<br>Al-hydroxylation<br>N-oxidation<br>S-oxidation | Ar-hydroxylation<br>Al-hydroxylation<br>N-oxidation<br>S-oxidation<br>epoxidation | S-oxidation<br>epoxidation | Ar-hydroxylation<br>Al-hydroxylation<br>N-oxidation<br>S-oxidation | Ar-hydroxylation<br>Al-hydroxylation |
| Orange | N-dealkylation<br>O-dealkylation<br>S-dealkylation<br>C-dealkylation<br>P-dealkylation<br>oxidative-deamination<br>oxidative-dehalogenation | N-dealkylation<br>O-dealkylation<br>S-dealkylation | N-dealkylation<br>O-dealkylation<br>oxidative-dehalogenation | N-dealkylation<br>O-dealkylation | N-dealkylation | N-dealkylation<br>O-dealkylation | N-dealkylation<br>oxidative-deamination |
| Green | alcohol to aldehyde or ketone<br>double/triple-bond formation<br>quinone formation<br>iminium formation | | alcohol to aldehyde or ketone | alcohol to aldehyde or ketone<br>double/triple bond formation | quinone formation<br>iminium formation | | alcohol to aldehyde or ketone |
| Blue | ester hydrolysis<br>amide hydrolysis | | | | | | |

[a]Combined SRMs: combined single reaction models
[b]aromatic hydroxylation
[c]aliphatic hydroxylation

**Table 2. Rainbow Model had improved reaction coverage in comparison to previously published models.**

| | Rainbow Model | SMARTCyp[20] | RSPredictor[21] | SOMP[23,24] | Combined SRMs[25–28] | SOME[29] | He et. al.[30] |
|---|---|---|---|---|---|---|---|
| Purple | ether hydrolysis cyanide hydrolysis sulfo-reduction nitro-reduction carbonyl-reduction phosphate-reduction | | nitro-reduction | | nitro-reduction | | |
| Descriptors | Topological | SMARTCyp—estimated energy required for CYP to access certain functional groups (defined by SMARTS string), accessibility | Topological, quantum mechanic, SMARTCyp | Chemical fingerprints | Topological | Quantum mechanic | Topological, quantum mechanic |
| Machine Learning Algorithm | Multitarget, deep Neural Network | Linear Model | Ensemble of SVM models | Bayesian | Single-target Neural Networks | Six Support Vector Machine (SVM) models, each makes SOM prediction for a Phase I reaction type | Ensemble of machine learning algorithms |
| SOM Labeling Strategy | five color SOM | generic atom SOM | generic atom SOM | generic atom SOM | atom/bond SOM | reaction-type specific atom SOM | reaction-type specific bond SOM |
| Prediction Format | Probabilistic | Ranking | Ranking | Binary Output | Probabilistic | Probabilistic?[d] | Unspecified |

[d]The authors[29] claimed that the output from SVM were probabilistic after Platt scaling but failed to provide evidence that the scaling indeed worked. In addition, Platt scaling of SVM output has been shown to fail in certain instances.[22] Platt scaling is based on the assumption that the distance from a point to the decision plane in SVM is correlated to the confidence of the classification prediction. In reality, the distance of a point

**Table 2. Rainbow Model had improved reaction coverage in comparison to previously published models.**

| | Rainbow Model | SMARTCyp[20] | RSPredictor[21] | SOMP[23,24] | Combined SRMs[25–28] | SOME[29] | He et. al.[30] |
|---|---|---|---|---|---|---|---|
| Data Source | AMD | Literature curation | Literature curation | AMD | AMD | AMD | BKM-react |
| Reproducibility | Training data published | Training data published | Training data published | Training data not published | Training data published | Training data not published | Training data published |
| Availability to test new molecules | Web server | Web server | No | Web server | Web server | Software | No |

form the decision plane is dependent on the sparsity of the data and does not always reflect the confidence of the prediction.

### 4.4.2 Ambiguity of Generic SOMs

We first compared ambiguity of the rainbow labeling scheme to generic SOMs in reported reactions. As an exemplar of the generic SOM labels, we used scheme adopted by XenoSite[258] (XenoSite 1.0), which was among the first published computational models that predicted the likely SOMs for P450 isozymes based on the structure of an input molecule. XenoSite 1.0 used a generic SOM labeling scheme where atoms of a compound were labeled as SOM based on the structural change from substrate to product. There are several methods reported in the literature that generically predict SOM, and all use a similar labeling scheme. Compared to this generic SOM approach, the rainbow labels specify both sites and reaction types by identifying modified bonds and the reaction classes of reactions observed at those sites.

Among reported sites of metabolism, the rainbow labeling scheme significantly reduced the ambiguity in the generic labeling algorithm. A single site can correspond to multiple types of reaction, and generic labeling leaves ambiguity as to which metabolite is generated by metabolism. A SOM was classified as able to undergo a certain reaction type if the SOM is part of a potential site defined by SMARTS strings of that particular reaction. For sites that can undergo more than one type of reactions, the ambiguous sites, we quantified how often knowledge of the site of metabolism alone is sufficient or insufficient to infer the structure of the metabolite.

Most generic sites of metabolism, using atom labels, are ambiguous. Using generic atom labeling approach (Table **??**), $90.58\%$ ($18544$ / $20472$) of the SOMs can undergo more than one reaction type. We identified 20 major groupings of ambiguous SOMs under this labeling scheme (Figure 4.4.1). Most sites of metabolism, using generic bond labeling scheme, are ambiguous too. Using bonds reduced the percentage of ambiguous SOM, but still left most sites of ambiguous $71.47\%$ ($17061/23872$). In contrast, only $0.3\%$ of rainbow labeled bonds are ambiguous. Despite these improvements, the rainbow labels, however, are ambiguous for a small subset of C-dealkylation reactions wherein the oxygen can attach to either of the two carbons in the bond. Rainbow labels nearly eliminate all ambiguity in sites of metabolism, which might enable

76

| Reported | Clashed | Count |
|---|---|---|
| hydroxylation | C-dealkylation | 4707 |
| N-dealkylation | hydroxylation | 2754 |
| hydroxylation | double/triple bond formation | 2446 |
| hydroxylation | hydrogenation | 2124 |
| ester hydrolysis | carbonyl reduction | 2109 |
| hydroxylation | epoxidation | 2077 |
| ester hydrolysis | O-dealkylation | 2036 |
| ester hydrolysis | C-dealkylation | 2025 |
| N-dealkylation | C-dealkylation | 1671 |
| O-dealkylation | hydroxylation | 1519 |
| O-dealkylation | ether hydrolysis | 1445 |
| N-dealkylation | double bond formation | 1377 |
| amide hydrolysis | N-dealkylation | 1116 |
| amide hydrolysis | carbonyl reduction | 1116 |
| amide hydrolysis | C-dealkylation | 1014 |
| epoxidation | adjacent reduction | 1007 |
| epoxidation | C-dealkylation | 912 |
| C-dealkylation | hydroxylation | 856 |
| carbonyl reduction | C-dealkylation | 853 |
| epoxidation | triple bond formation | 848 |

N-dealkylation & hydroxylation

hydroxylation & epoxidation

hydroxylation & dehydrogenation

amide hydrolysis & carbonyl reduction

**Figure 4.4.1: Generic SOM labels are ambiguous.** Many sites of metabolism (SOM) if labeled using the traditional method can undergo more than one type of reactions. Top 20 groups of such ambiguous sites are listed in the table. These sites were reported to undergoes the reaction annotated in the left but could also undergoes the reaction on the right. The number of known sites that can undergo at least two types of reaction are listed. Example molecules are known drugs with known SOM circled.

metabolite prediction from effective rainbow SOM models of metabolism.

### 4.4.3   Reaction-Class Specific Site-Level Accuracy

Considering each color separately, the rainbow model accurately predicted colored SOMs within metabolized molecules. First, we measured "top-two" performance on a color by color bases, where a molecule was considered to be correctly predicted if any of its reported SOMs were ranked in the top two sites sorted by that class-specific prediction score. The top-two metric is commonly used to evaluate CYP site of metabolism models and is useful for assessing individual drug candidates. Stable oxygenation, unstable oxygenation, dehydrogenation, reduction, and hydrolysis top-two accuracies were, respectively, $72.2\%$, $80.4\%$, $74.2\%$, $94.9\%$, and $94.7\%$ (Figure 4.4.2). Across all five classes, the model had an average cross-validated top-two accuracy of $88.2\%$.

Similar results results were obtained using the "average site AUC" for each reaction class, a
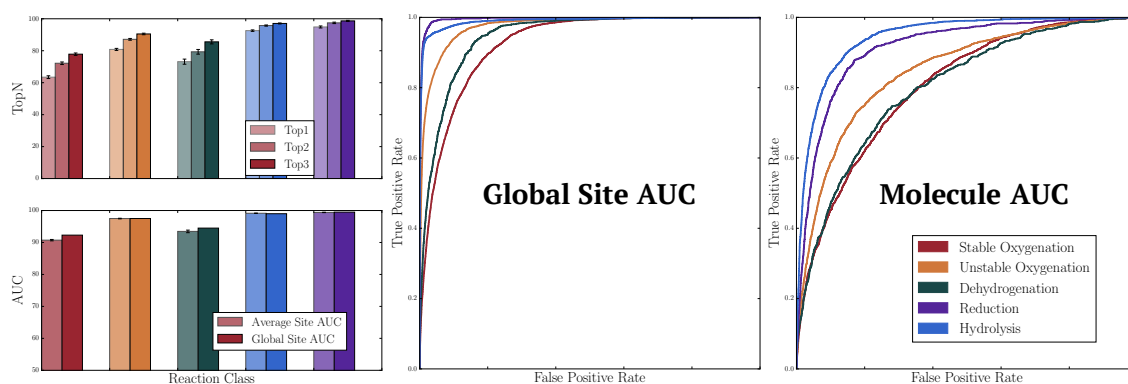
**Figure 4.4.2: XenoSite accurately predicts site of Phase I metabolism.** Ten-fold cross-validated accuracies are shown. Across all five classes, the model had an average cross-validated top-two and average AUC accuracies of $88.2\%$ and $92.3\%$, respectively.
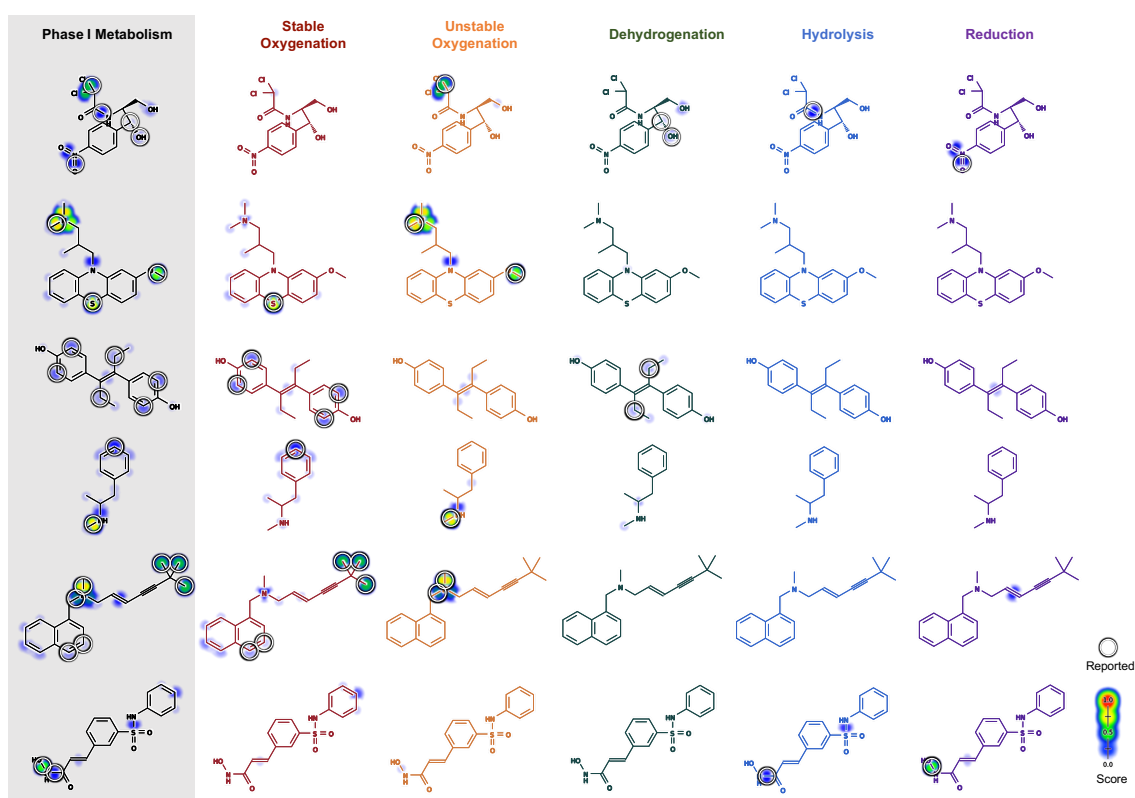


**Figure 4.4.3: XenoSite accurately predicts site of Phase I metabolism on drugs.** Example drugs with ten-fold cross-validated predictions are shown. Experimentally reported sites of metabolism are circled. XenoSite assigns the molecule's SOMs with the highest Phase I metabolism scores in every case. Five example drugs are visualized, including from top to bottom, are chloramphenicol, levomepromazine, diethylstibestrol, terbinafine, and belinostat.

metric based on the area under the ROC curve (AUC). AUC is a standard metric employed in machine learning to quantify how well a model is able to distinguish between positives and negative test cases of a specific chemical class. It is employed here by computing the individual AUCs are computed for each molecule in the dataset, color by color, and then averaging across the whole dataset to produce a single summary statistic for each reaction class. Stable oxygenation, unstable oxygenation, dehydrogenation, reduction, and hydrolysis average site AUC accuracies were 91.1%, 97.6%, 94.7%, 99.5%, and 99.7%, respectively. Across all five classes, the average AUC was 97.5% (Figure 4.4.2). Third, we measured the model's ability to distinguish metabolized sites from all other potential sites across the entire dataset using "possible site AUC. These results were consistently obtained, even when sites unable to undergo metabolism, because they did not correspond to any reaction type, were removed from consideration and the performance measure was recomputed. The "possible site AUC" [228] of the predictions of metabolized and non-metabolized sites across all possible sites in the entire dataset. Stable oxygenation, unstable oxygenation, dehydrogenation, reduction, and hydrolysis possible site AUC accuracies were 85.4%, 94.3%, 86.8%, 98.2%, and 93.9%, respectively. Across all five targets, the model had an average possible site AUC of 92.3%.

Considering all colors together, the rainbow model accurately predicted class-specific SOMs within metabolized molecules. In contrast with the previous section, each color is considered together, and the model's ability to correctly predict the color and site at the same time is measured. In this evaluation, each color-site pair is considered an individual entity, and is assigned a score by the model. All pairs are ranked by the scores, and this is compared with the known metabolism to test if observed sites of metabolism are separated from the rest.

We evaluated the model using three metrics. Across all metrics, we see high performance. First, we measured "average all-color AUC". This metric is computed by averaging the AUC over five scores assigned to each metabolized site in the dataset. The model had cross-validated average all-color AUC accuracy of 97.4%. Second, we calculated "all-color top-two" performance, where a molecule was considered to be correctly predicted if any of its observed SOMs were ranked in

**Table 4.4.1: Reaction-Type Specific Accuracy**.

| | | Rainbow | | | |
| --- | --- | --- | --- | --- | --- |
| | Reaction Type | PS | All | SOME[261] | He et. al.[97] |
| red | S-oxidation | 89.4% | 99.9% | 92.9%* | |
| | C-hydroxylation | 79.8% | 93.8% | | 89.4%* |
| | N-oxidation | 91.1% | 99.6% | 96.9%* | |
| | aromatic epoxidation | 95.8% | 99.5% | | |
| | aliphatic epoxidation | 89.5% | 97.5% | | |
| orange | N-dealkylation | 94.2% | 99.5% | 95.4%* | 95.5%* |
| | oxidative dehalogenation | 82.8% | 99.9% | | |
| | O-dealkylation | 92.3% | 99.6% | 94.3%* | |
| | S-dealkylation | 76.1% | 99.9% | | |
| | C-dealkylation | 86.0% | 96.5% | | |
| green | aliphatic dehydegenation | 87.8% | 94.6% | | |
| | quinone formation | 92.1% | 99.8% | | |
| | aromatic dehydrogenation | 90.4% | 97.2% | | |
| blue | amide hydrolysis | 77.3% | 99.8% | | |
| | ester hydrolysis | 81.1% | 99.9% | | |
| | ether hydrolysis | 79.6% | 99.6% | | |
| purple | nitro reduction | 91.8% | 99.9% | | |
| | aliphatic hydrogenation | 98.7% | 99.9% | | |
| | carbonyl/sulfo reduction | 95.2% | 99.9% | | |
| | reductive dehalogenation | 99.2% | 99.9% | | |
| | aromatic hydrogenation | 97.9% | 99.9% | | |

*Each model in this table uses a different dataset. The reported accuracies are not comparable.

the top two sites sorted by all site prediction scores. rainbow has an "all-color top-two" of 78.2%. Third, we calculated "average all-color site AUC". This is the average AUC of all molecule's ROC curves, each containing all points of reaction-color prediction and target pairs across five reaction classes. The model had cross-validated average all-reaction AUC accuracy of 97.3%.

### 4.4.4   Reaction-Type Specific Site-Level Accuracy

We evaluated the model's ability to separate metabolized sites from all potential sites for specific reaction types across the entire dataset. Reported accuracies from [261] He et.al. [97] models on the same reaction types are also included in Table **??**. Across 21 reaction types, the model had an average cross-validated global AUC accuracy of 99.0%.

### 4.4.5   Molecule-Level Model Accuracy

The rainbow model accurately predicted whether a molecule undergoes a specific class of Phase I reaction. In our training dataset, each substrate was metabolized through one or more reaction classes but rarely through all five types. For each reaction class, we used the "molecule AUC" metric to evaluate how well the model separated molecules metabolized through that specific reaction class from those that undergo reactions of other classes. Following site-level training, we investigated several methods of discriminating between type-specific metabolized and non-metabolized molecules (Figure 7.4.1). The best performing model was a neural network that takes as input the top five prediction scores for each reaction class and molecule-level descriptors (Figure 7.3.2). The best performing model had molecule AUCs of $78.3\%$, $83.9\%$, $77.3\%$, $90.4\%$ and $92.7\%$ for stable oxygenation, unstable oxygenation, dehydrogenation, reduction, and hydrolysis, respectively (Figure 4.4.2).

### 4.4.6   Probability Scaling of Model Output

The model's output can be interpreted as probabilities. When we binned class-specific sites by the Phase I prediction score, the proportion of class-specific SOMs in each bin closely correlated with the bin's score (Figure 7.4.2). Likewise, when we binned molecules by the Phase I molecule score, the proportion of class-specific metabolized molecules in each bin also correlates with the bin score. Quantitatively, Pearson regression coefficients of site and molecule levels are $0.996$ and $0.947$, respectively.

### 4.4.7   Reactive Metabolite Forming Reactions

Detecting or predicting the formation electrophilically reactive metabolites is an important step in avoiding idiosyncratic adverse drug reactions.[5, 196, 218, 225] Experimental methods for detecting reactive metabolites, most frequently *in vitro* incubation with glutathione and cyanide, are well established.[116, 206, 263] However, these methods require significant time and resources, especially when screening thousands of compounds during drug development. Our

previous works and others have shown the instrumental role of computational methods in directing these experimental efforts.[57, 103, 104, 106] We published several models that focused on specific reactions like epoxidation,[104] quinone species formation,[103] S-oxidation,[57] nitroaromatic reduction,[57] and N-dealkylation.[59] These reactions play important roles the bioactivation of multiple structural alerts leading to the formation of deleterious reactive metabolites.[57] Given a molecular structure, each of these models output probabilistic prediction scores corresponding to the likelihood that each atom or bond in the molecule is a site of metabolism for that specific reaction. The following sections demonstrate that Phase I can predict reactive metabolite formation reaction with high accuracy, comparable to previous models of individual reaction types.

### 4.4.8   Epoxidation

Epoxides, the second most common type of reactive metabolites, are often formed via P450-mediated oxidation of double bonds or aromatic rings.[130, 211] Epoxides are detoxified by epoxide hydrolases and the glutathione system.[75, 172, 215, 224] When these pathways are overwhelmed, epoxides can covalently bind to proteins and cause toxic side effects and immune reactions.

The rainbow model can predict the formation of this key reactive metabolite, and is comparable in accuracy to our previous epoxidation model.[104] The epoxidation model was trained on a dataset containing 524 molecules, 389 of which contained sites of epoxidation. Our model was able to predict this dataset with an accuracy of 92.7% possible site AUC (Figure 4.4.4A), compared to the possible site AUC of 94.3% of the previous epoxidation model (one-sided $p = 0.097$). Reliability plot of unscaled stable oxygenation on possible epoxidation sites demonstrate that the scores can be interpreted as probabilities (Figure 4.4.5A).
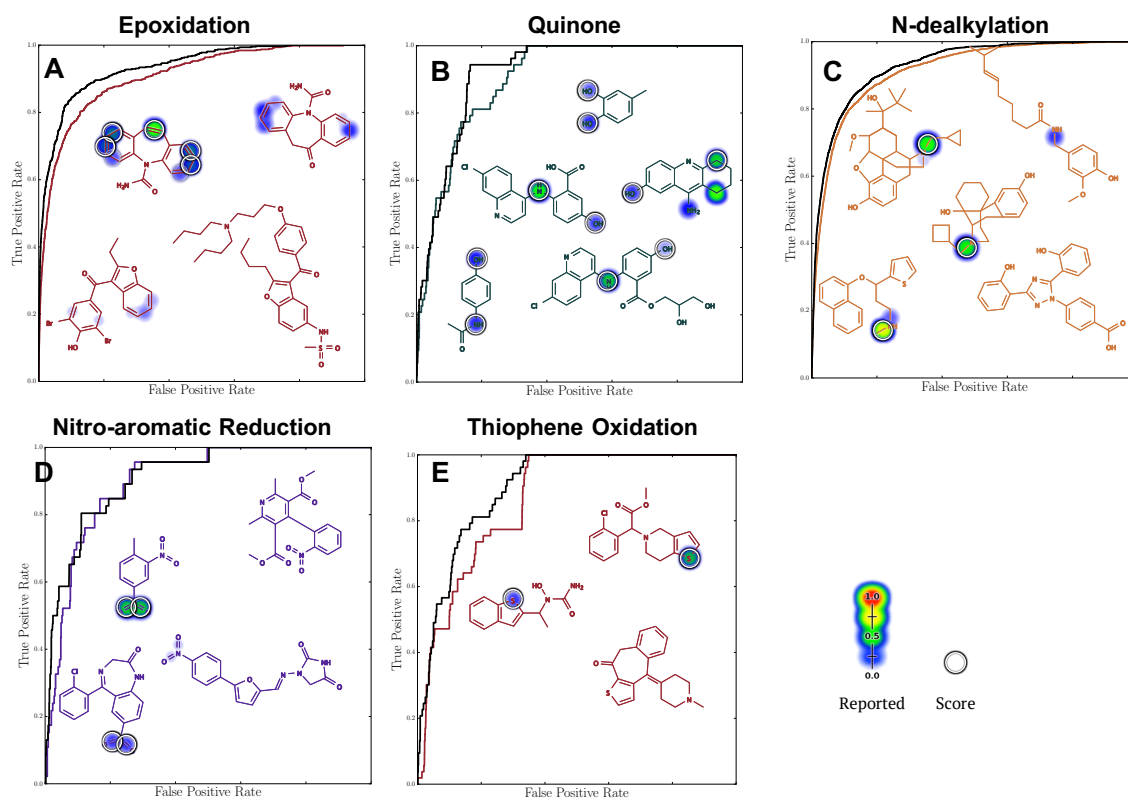
**Figure 4.4.4: Rainbow Model Accurately Predicts Key Reactive Metabolite Formation Reactions.**
Datasets of 524, 39, 883, 98, and 50 molecules were used to test rainbow performance on epoxidation, one-step quinone formation, N-dealkylation, nitroaromatic reduction, and thiophene S-oxidation. The model's possible site AUC accuracies are 92.7%, 90.1%, 92.2%, 90.1% and 86% on epoxidation, one-step quinone formation, N-dealkylation, nitroaromatic reduction, and thiophene S-oxidation datasets, respectively. Previously, we trained models for each specific reactive metabolite formation reactions. These models' possible site AUC accuracies on epoxidation, one-step quinone formation, N-dealkylation, nitroaromatic reduction, and thiophene S-oxidation datasets are 94.3%, 92.5%, 94.7%, 93.0% and 88.0%. The differences between each pair of performance are statistically insignificant. Example molecules are shown with cross-validated scaled prediction (refer to Prediction Visualization).

### 4.4.9   One-Step Quinone Formation

Quinone is the most common class of reactive metabolites, accounting for $40\%$ of known reactive metabolites.[233] Quinone species are frequently highly electrophilically reactive.[214] Dehydrogenation is the direct metabolic reaction that form quinone.

The rainbow model predicted the formation of this important reactive metabolite. Our model predicted a dataset of $39$ molecules that form quinone through one-step (dehydrogenation) with an accuracy of $90.1\%$ AUC across all quinone formation possible sites (Figure 4.4.4B). This performance is comparable the possible site AUC of $92.5\%$ of previous quinone formation model (one-sided $p = 0.36$).[103] Reliability plot of unscaled dehydrogenation scores on possible quinone sites demonstrate that the scores are well-scaled probabilities (Figure 4.4.5B).

The rainbow model does not predict multiple step quinone formation reactions. Only a small subset of quinones are formed after a single-step of metabolism. This is an intrinsic limitation of this approach, but it might be addressed by an algorithm that chain predictions of the rainbow model into multiple steps. This possibility is out of scope of this study, and will be considered future work.

### 4.4.10   N-Dealkylation

Alkylated amines are a class of compounds with important biological and pharmaceutical functions. [21, 220] N-dealkylation, the replacement of an alkyl group by hydrogen, is the main metabolism pathway for many alkylated amines.[87, 99, 253] This removed alkyl group is often transformed into a reactive aldehyde. Unfortunately, metabolic studies often focus on the amine metabolite and do not report the aldehyde metabolite because aldehydes are presumed to be quickly detoxified to carboxylic acids and alcohols by multiple enzyme systems. However, some reactive aldehydes escape these detoxification pathways, and form adducts that can give rise to adverse drug events, especially in patients where the detoxification pathways are inhibited by other drugs.[20, 212]

The rainbow model accurately predicts sites of N-dealkylation on the dataset of $883$ molecules

used to train our previous N-dealkylation model.[59] Our model predicted N-dealkylation with an accuracy of 92.2% global AUC (Figure 4.4.4C). This accuracy is comparable than the global AUC accuracy of 94.7% of our previous N-dealkylation model ($p = 0.08$). Reliability plot of unscaled unstable oxygenation on possible N-dealkylation sites demonstrate that the scores can be interpreted as probabilities (Figure 4.4.5C).

### 4.4.11   Nitroaromatics Reduction

Nitroaromatics are common in the environment and are important pharmacophores in many drugs.[176, 198, 203, 247] Unfortunately, nitroaromatic groups can metabolically transform into electrophilic, highly unstable intermediates such as nitroso and nitro anion radicals through reduction.[27, 57, 124, 236]

The rainbow model accurately predicts nitroaromatic reduction. Previously, we published the prediction from an early version of rainbow on nitroaromatic-containing drugs. The early version model was trained on a dataset containing 3061 molecules, 98 of which contain nitroaromatics.[57] In comparison, the training dataset of the rainbow model contains 314 nitroaromatic molecules. The rainbow model of Phase I model predicted nitro-reduction of the 98 molecules with an AUC accuracy of 90.1% (Figure 4.4.4D), compared to 93.0% obtained by the early version of the model (one-sided $p = 0.16$). Reliability plot of unscaled reduction on nitroaromatics demonstrate that the scores can be interpreted as probabilities (Figure 4.4.5D).

### 4.4.12   Thiophene S-Oxidation

Thiophenes are five-membered sulfur-containing aromatic rings that are important pharmacophores in many drugs.[17, 38, 54, 101, 112, 148, 157] Thiophenes can undergo P450-mediated S-oxidation to form a reactive sulfur oxide, which can subsequently bind to proteins causing toxic side effects. For example, suprofen is a thiophene-containing analgesic medication that failed in late-stage clinical trials because of the formation of a short-lived reactive metabolite that caused irreversible inhibition of phase I enzymes.[182] However, many nontoxic
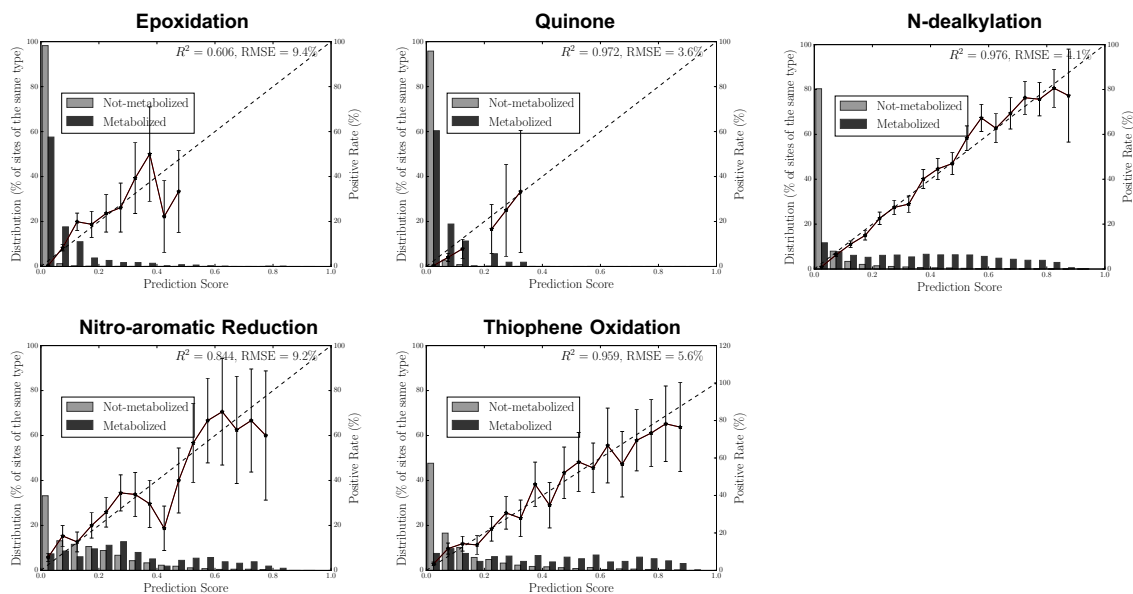
**Figure 4.4.5: The model makes well scaled predictions, corresponding to probabilities.** The bar graphs plot the distributions of scores across metabolized and non-metabolized potential sites for each reaction class. The solid lines plot the percentage of bonds and lone pairs that are metabolized via a specific Phase I reaction class (using non-normalized frequencies) in each bin. The diagonal dashed lines indicate a hypothetical perfectly scaled prediction. Phase I XenoSite score has a strong correlation to a perfectly scaled prediction. This means that the class-specific score is interpretable as the probability that potential sites is metabolized via the corresponding bioactivation reaction.

drugs also contain thiophenes that are not bioactivated, such as eprosartan, a closely related analogue to suprofen.[80]

The rainbow can predict thiophene S-oxidation. Previously, we published the predictions from an early version of rainbow on thiophene-containing compounds.[57] The training dataset of the earlier model contained 50 thiophene containing compounds [57], compared to 105 in the training data for the current study. Our new model predicted thiophene S-oxidation with a global AUC accuracy of 86.0% (Figure 4.4.4D), compared to the previous model's performance of 88.0% ($p = 0.398$). Reliability plot of unscaled stable oxygenation on thiophenes demonstrate that the scores can be interpreted as probabilities (Figure 4.4.5).

### 4.4.13    Visualizing Metabolism in Colors

There is no established way of displaying site of metabolism data from a large number of reaction types in a coherent or efficient way. One motive in defining reaction classes as "colors" is to introduce a mnemonic metaphor as a foundation for efficient visualizations of metabolism. Reaction color becomes a visual cue for reaction classes in visualizations with multiple reaction types.

We define three levels of visualizations at which to display SOM predictions and observations. Three three levels of visualization work together (Figure 4.4.6). Large lists of molecules can be displayed with just the highest level predictions, and lower level visualizations can clarify details when they are requested. Also, multiple levels can be displayed at the same time to display a large amount of metabolism data. At the highest level, we show predictions as a single molecule, so this may be most useful for screening large numbers of molecules in the first pass. This visualization uses a black molecule depiction with white circles. In the future, we might add color to the circles to indicate reaction class, perhaps including multiple colors when multiple reaction colors are observed at a single site (Figure 4.2.1).

At the middle level, five depictions are shown, one for each color. This visualization makes immediately apparent the reaction type of most predictions and sites of metabolism. Here, each of the five molecules are depicted each of the reaction colors, and observed sites of metabolism are annotated with white circles. The model predictions for each color are displayed as shading behind the molecule of the corresponding color. The color of the depicted molecule is a critical visual cue. With practice, one can distinguish reaction-types based on structural context and the reaction color.

At the lowest level, all reaction types are displayed in separate depictions. The visualization still colors the molecule depiction with the class color, but there are multiple molecules with the same color. Reported sites of metabolism, again, are marked with white circles. The reaction type name is included too, to clarify which depiction corresponds to which reaction type. This is the most expansive visualization, which is most suitable for drilling down into the details of

molecules being considered most closely.

Though the predictions in the model can be visualized, it does take effort to learn now to navigate the figure. In future we plan to integrate this model as a building block into composite algorithms designed to answer specific questions in drug development. For example, we hope to models that can predict bioactivation pathways with this model. Ultimately, we expect this visualization may be more useful as a diagnostic for tools that make use of predictions from the model.

## 4.5    Model Limitations

The rainbow model has several limitations. First, our model does not cover all Phase I reactions and *in vivo* transformation. Although rare, excluded reactions include tautomerization, isomerization, and rearrangement, which can have important implications on the efficacy and safety profile of drugs. In the future, improved models may be able to label and group these reaction types with a similar approach. Second, while the current SOM labeling scheme significantly reduced ambiguity, it does not provide sufficient information to predict the exact metabolites of a small number of reactions, about $0.3\%$ of the total. In future studies, we could add an ordering to the labeling, such that each bond is corresponds to two oppositely oriented edges. SOMs such as these might remove the ambiguity in these rare cases.. Third, the model slightly underperformed previously developed models of specific reactive metabolite generating reactions, though the drop in performance was not statistically significant. It is possible that improved descriptors or training protocols could change this general trend. Fourth, rainbow model only consider one aspect of IADRs. IADRs are complex phenomenon with several additional factors, including dosage, route of administration, competing bioactive and detoxification pathways, co-morbidities, co-administered medications, and genetic variants.[30, 55, 237]
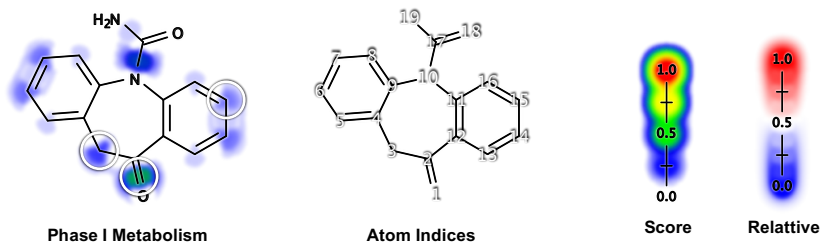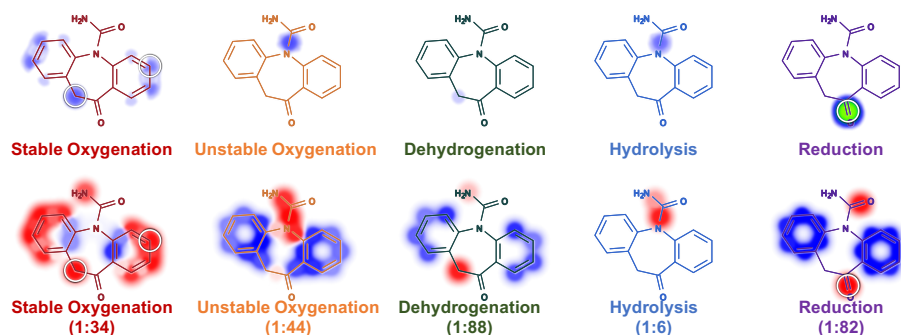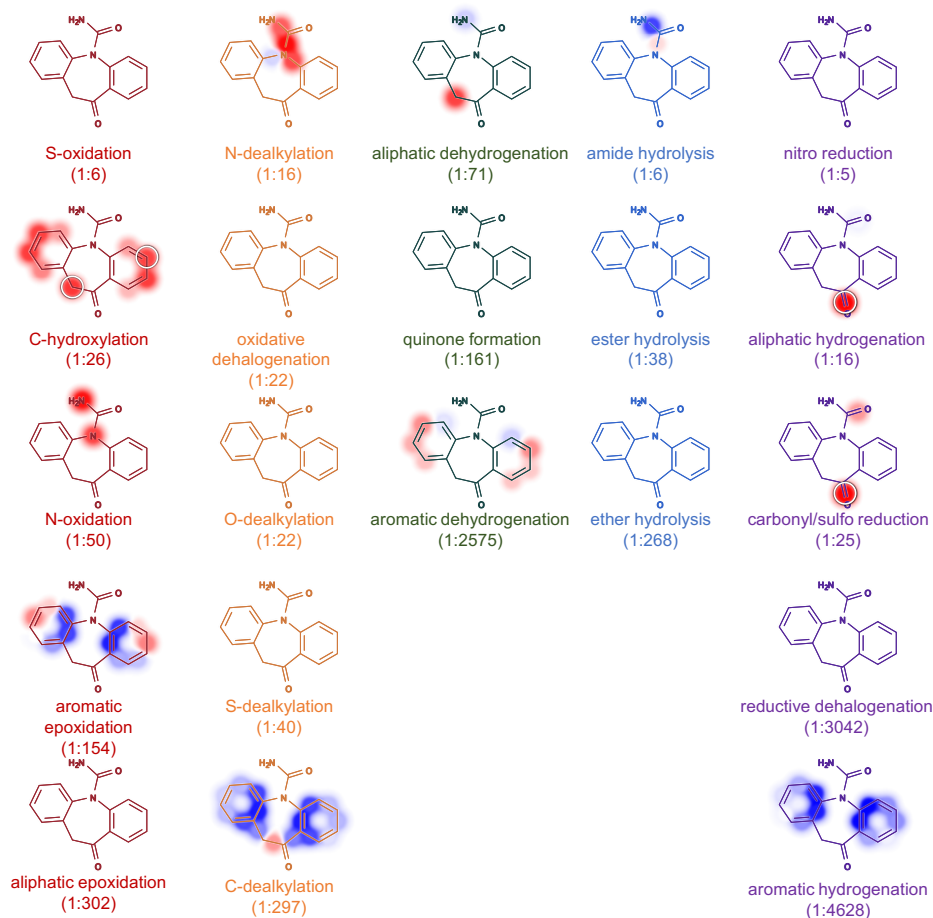
**Figure 4.4.6: Phase I metabolism in colors.** Phase I model predictions on Oxcarbazepine are displayed at three levels. The top level displays summarized predictions to highlight metabolic "hot spots". The middle level contains predictions for each of the five reaction colors. The bottom level displays predictions for each of the 21 reaction types.

## 4.6   Conclusions

Published site of metabolism models can provide both accurate and rapid screening tool for large numbers of small molecules. Unfortunately, they are unable to predict the type of reaction a site would undergo and thereby ambiguous in discriminating whether a harmful metabolic transformation would happen at a certain site. This study demonstrated that modeling both site and type of reaction would enable unambiguous metabolite structure prediction and thereby discern which site of metabolism should be eliminate to avoid harmful metabolic transformation. We developed a labeling scheme that reduced the number of ambiguous SOMs from 90.6% to 0.3%. We also showed that it is possible to develop a single model for a wide variety of Phase I reaction types by grouping them into classes and using a multi-target model. In comparison to published models, we doubled the Phase I reaction coverage to 92.3%. Our model predicted the five classes of Phase I metabolism with top-two, average site AUC, and global AUC accuracies of 78.2%, 97.3%, and 97.1%, respectively. In comparison to previous models exclusively designed for specific reaction types, our model performed equally well on several important structural alert bioactivation pathways. The results underscored the feasibility of reducing ascertainment bias without sacrificing performance. While we have not comprehensively handle all the relevant complexities of predicting a compound's likelihood of metabolism-induced toxicity, we are moving closer to that goal. Ultimately, we envision that combined models of metabolism, metabolite structure prediction, and reactivity prediction will provide a powerful tool in assessing ADR risk of drug candidates.

## 4.7   Abbreviations

ADR, Adverse Drug Reaction; AMD, Accelrys Metabolite Database; AUC, Area Under the Receiver Operating Characteristic Curve; CYP, Cytochromes P450; SOM, Site of Metabolism

# 5

# Modeling Phase-I and Phase-II *in vivo* Competition

This chapter is a manuscript in preparation:

**Na Le Dang**, Matthew K. Matlock, and S. Joshua Swamidass, *In vivo* competition between Phase I and Phase II.

## 5.1   Summary

As the field of metabolism modeling has matured, many important advances have been made. We now can effectively and accurately predict many aspects of metabolism based on the structure of

the candidate molecule. However, most work has focused on cytochrome P450 metabolism, which is the major part of Phase I metabolism. We recently published Rainbow model[58] that is capable of predicting not only the site but also the type of diverse Phase I reactions with high accuracy. Much less work has been done on Phase II metabolism, even though our group and others have published on glucuronidation[56, 194, 207] and glutathionation.[105, 106] We for the less common nevertheless clinically significant[41, 51, 162, 219] Phase II reactions like sulfuration, acetylation, and methylation.

A looming critical challenge is combining models of Phase I and 2 to predict metabolism *in vivo*. This challenge is difficult not only for computational but also for experimental approaches. Two major factors hinder using *in vitro* systems to measure the *in vivo* balance of Phase I versus Phase II metabolism. First, *in vitro* studies entail using enzymes, substrates, as well as cofactors at concentrations that are substantially higher than the typical *in vivo* conditions. As the result, *in vitro* systems often overestimate *in vivo* substrate affinity. For example, uridine glucuronide transferases (UGT) are capable of metabolizing even low affinity substrates given high enough concentration.[88, 246] Second, it is very difficult to emulate *in vivo* compartmentalization *in vitro*. Consequently, *in vitro* metabolism are not as diverse as *in vivo* metabolism.

The balance between Phase I and 2 metabolism *in vivo* has important clinical consequences. While both Phase I and 2 are capable of generating reactive metabolite with deleterious effects, [13] they often act as competitive and/or sequential pathways to limit reactive metabolite formation and/or deactivate the reactive compound.[249] For example, while Phase I activate acetaminophen into reactive quinone, Phase II reaction such as glucuronidation and sulfonation are competitive pathways that limit the quinone formation and glutathionation is the deactivating step (Figure 7.2.1). Furthermore, altered balance between Phase I and 2 metabolism can lead to serious consequence in understudied subpopulations. For example, young children rely more heavily on Phase II metabolism than healthy adults do: defect in UGT system increase toxicity risk to drugs like chloramphenicol.[12, 173] Similarlly, SULT1A2 polymorphism may affect the individual susceptibility towards procarcinogens, in particular certain aromatic amines and
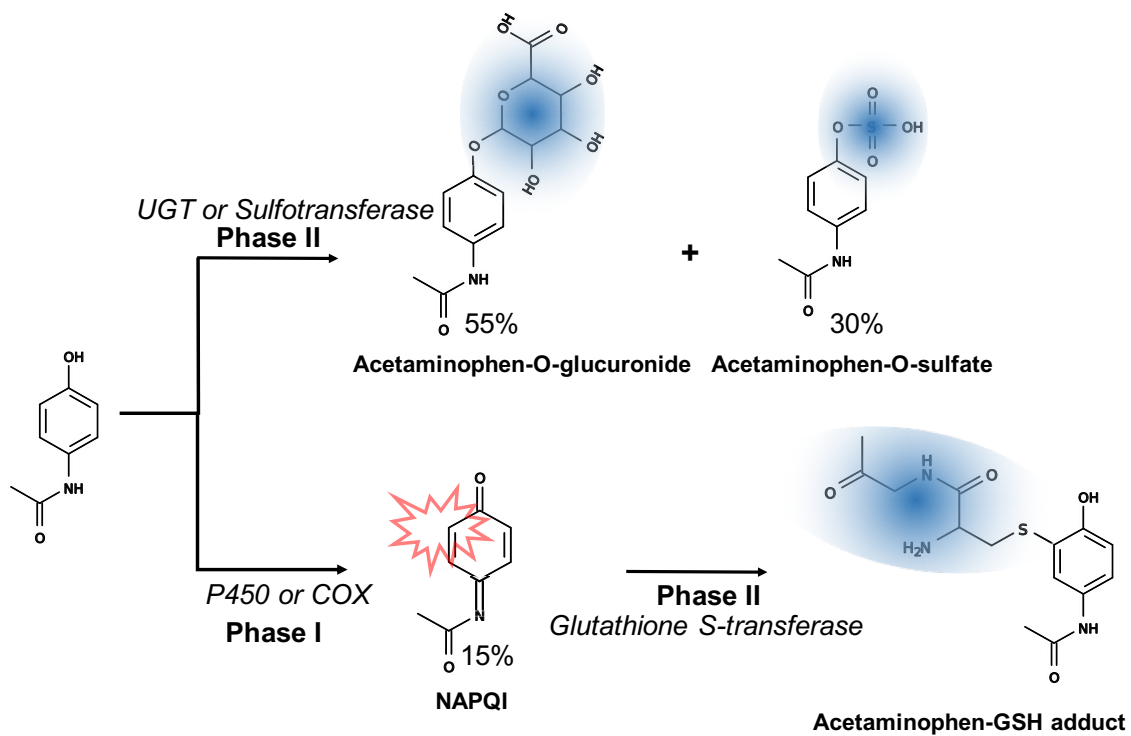
**Figure 5.1.1: The balance between Phase I and 2 metabolism *in vivo* has important clinical consequences..** For example, while phase 1 activate acetaminophen into reactive quinone, phase 2 reaction such as glucuronide and sulfo conjugation are competitive pathways that limit the formation and glutathione conjugation is the deactivating step

amides, including diuretic agent triamterene and hair follicle stimulant minoxidil.[115, 162] Defective GSTA1*B allele was associated with increased survival in breast cancer patients treated with cyclophosphamide [229] Defective NAT2 alleles increases risk of drug-induced lupus erythematosus of procainamide.[200]

To address the aforementioned problem, we develop a machine learning model to predict *in vivo* Phase I and 2 substrate specificity. We collated large databases of $200,000$ human *in vivo* and *XXX* Phase II reactions. We built a Phase II metabolism model that predict four of the five major Phase II conjugation reaction: UGT, SUTL, NAT, and TMPT. [115] This comprehensive site of Phase II metabolism is the first of its kind and is highly accurate. Next, we combine outputs from the Phase II metabolism model, our published Phase I metabolism[58] and reactivity[103] model and other descriptors to build an accurate *in vivo* Phase I and 2 metabolism model. This *in vivo* metabolism model is the an accurately predict when a molecule and its metabolite go down Phase I and Phase II pathways.

## 5.2   Methods

### 5.2.1   Training Datasets

We collected a chemically diverse *in vivo* substrate data sets from the literature-derived Accelrys Metabolite Database (AMD). A total of $11888$ *in vivo* reaction records were collated. Our *in vivo* data set contained $380016$ -reference specific compounds. Under the AMD license agreement, we were not allowed to disclose the structures of molecules in the data set. However, to enable rebuilding of our complete database and reproduction of our results, we provided all reaction and molecule AMD registry numbers in the Supplemental Materials.

### 5.2.2   Outputs from Phase I, Phase II, and Reactivity Models

A subset of descriptors for the *in vivo* models are cross-validated predictions from the Phase II model and our published Phase I and Reactivity models. Our Phase I model produces 5 scores for
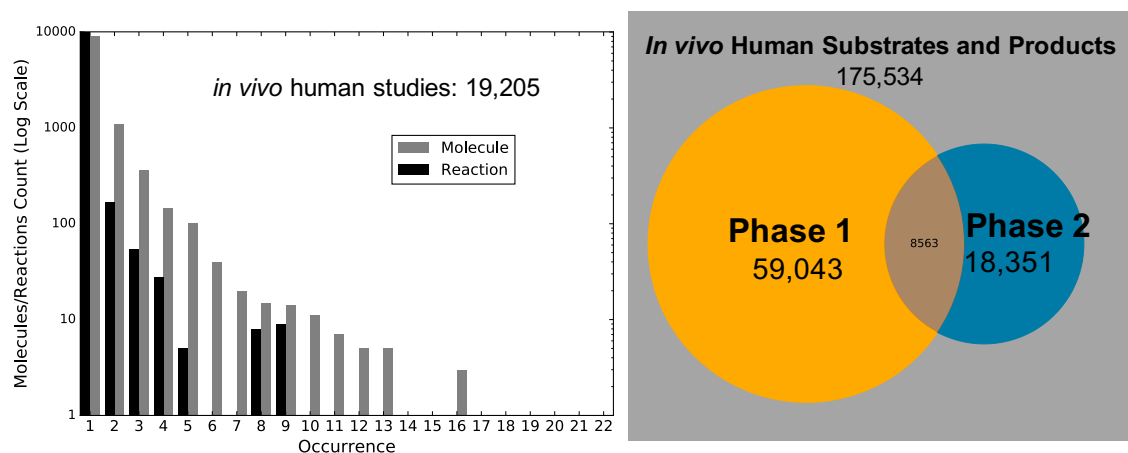
**Figure 5.2.1:** *In vivo* **training dataset**. TODO: include numbers from other Phase3 reactions: SULT, NAT, and TMPT

each site and molecule, each correspond to a class of Phase I reactions: stable oxygenation, unstable oxygenation, dehydrogenation, hydrolysis, and reduction. For each molecule in the *in vivo* data set, five molecular scores and five highest site scores from each class (30 in total) are used as input descriptors to the *in vivo* model.

Our reactivity model calculates 4 prediction scores for each site and molecule, indicating the probabilities that these sites and molecules are conjugated to protein, glutathione, DNA, and cyanide. Conjugation to glutathione is a type of Phase II reaction and conjugation to protein and DNA can lead to hypersensitivity reactions and cancers. For each molecule in the *in vivo* data set, four molecular scores and five highest scores from each conjugation type (24 in total) are used as input descriptors to the *in vivo* model.

Similarly, for each molecule in the *in vivo* data set, four molecular scores and five highest scores from each conjugation type (24 in total) from the Phase II model are used as input descriptors to the *in vivo* model.

## 5.2.3   In Vivo Model Descriptors

To predict the susceptibility to Phase I and Phase II metabolism of molecule in the data set, our model use numerical descriptors that include outputs from our previous models as well as context and chemical/physcial descriptors. Each molecules was represented by a vector of 118 descriptors

| Context Descriptor | A | B | C |
|---|---|---|---|
| Min number of steps | 1 | 1 | 2 |
| Max number of steps | 2 | 1 | 2 |
| Number of paths from the parent compound | 2 | 1 | 1 |
| Number of compounds with equal or small metabolic steps | 1 | 1 | 10 |

4 context descriptors
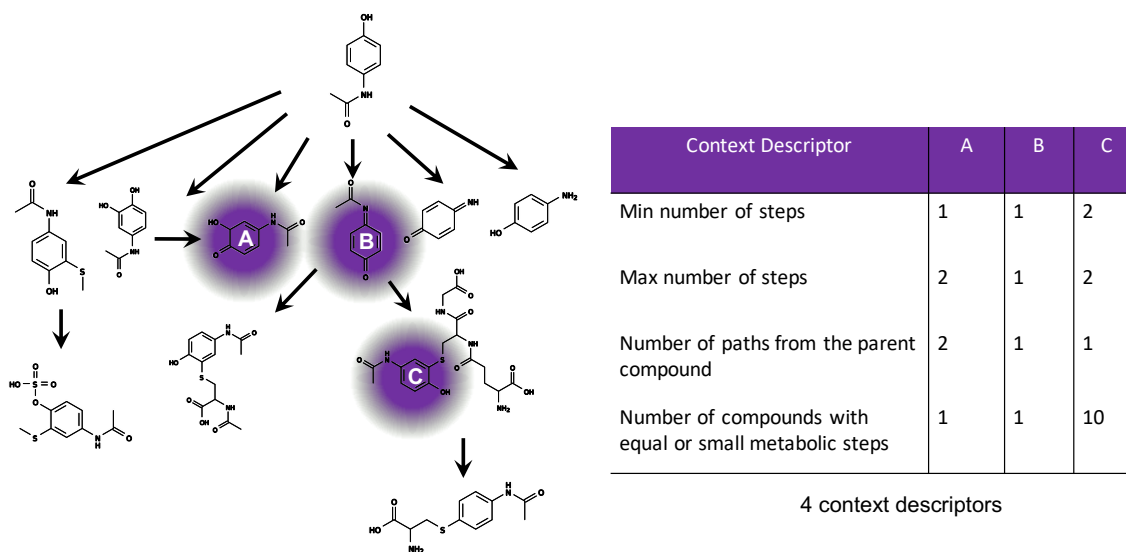
**Figure 5.2.2: Context descriptors from metabolic network.**Four context descriptors are calculated from metabolic networks where compounds are nodes and metabolic transformations are edges. A metabolic network is defined by the nodes and edges annotated from a certain study. Each study is associated with a single metabolic network. Each metabolic network can be associated with multiple studies. The administered compounds are denoted as the parents. The minimum and maximum transformations from the parent to a compound of interest are the min_steps and max_steps descriptors of the compound. The number of pathways from the parent to a certain compound and the total number of compounds with fewer transformation steps are the num_paths and higher_rank descriptors of the compound.

that describe its properties. 78 descriptors are molecule-level predictions from our Phase I, PhaseII, and reactivity models that were trained on *in vitro* metabolism studies (See above section).[56**? ?** ] 16 molecule descriptors (e.g. molecular weight, topological polar surface area, molar refractivity) (Table ) were developed in previous works. Four context descriptors (Figure 5.2.2) were calculated from metabolic networks where compounds are nodes and metabolic transformations are edges. A metabolic network is defined by the nodes and edges annotated from a certain study. Each study is associated with a single metabolic network. Each metabolic network can be associated with multiple studies. The administered compounds are denoted as the parents. The minimum and maximum transformations from the parent to a compound of interest are the min_steps and max_steps descriptors of the compound. The number of pathways from the parent to a certain compound and the total number of compounds with fewer transformation steps are the num_paths and higher_rank descriptors of the compound.

96

**Figure 5.2.3: The Structure of the Phase I Reaction Model.**) The diagram on the shows how information propagated through the neural network model, which contained one input layer, five hidden layers, and one output layer.

### 5.2.4 In Vivo Model

Our *in vivo* model have one input layer, serverl hidden layers and one output layers (Figure 7.3.2).

## 5.3 Results and Discussion

The following sections examined the inner working of the Phase II and *in vivo* models. First, we reported on the accuracy of the Phase II model. Second, we assessed the prediction of the *in vivo* model. Third, we examined the factors that are important to each target of the *in vivo* model. Fourth, we presented several case studies to demonstrated the utility of the *in vivo* model.

### 5.3.1 Phase II Model Accuracy

The accuracy of the model was evaluated using a ten-fold cross-validation procedure. The data set was divided into ten groups with an equal number of substrates in each group. Predictions were generated for each tenth of the data by a model trained using the remaining nine tenths of the set. The resulting predictions covered the complete data set and were used to compute the accuracy metrics in this study.

### 5.3.2 Phase II Descriptor Importance

SULT metabolites can be reactive.[] SULT is high affinity, low capacity while UGT is low affinity, high capacity.[] Both sensitivity analysis and perturbation studies result is in agreement with the observation.

### 5.3.3 In Vivo Model Accuracy

Ten fold cross-validation resulst: CYP $82.15$ UGT $80.69$ SULT $86.92$ NAT $89.28$ TMPT $73.76$ Aggregated $91.44$

### 5.3.4 Context-Dependence of In Vivo Model

### 5.3.5 Phase I and Phase II Selectivity of In Vivo Model

### 5.3.6 Case Studies

**Acetaminophen** Acetaminophen is the most common drug that cause liver failure.

**Irinotecan**

**Raloxifene** Scheme 36, Stepan2011structuralalert

Raloxifene form quinone by P450, but the alternative detoxification pathway is glucuronide conjugation.
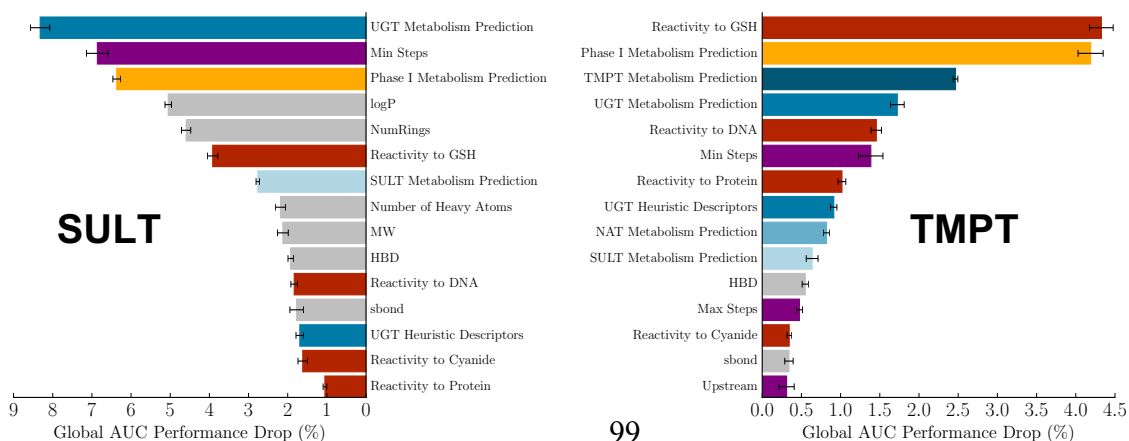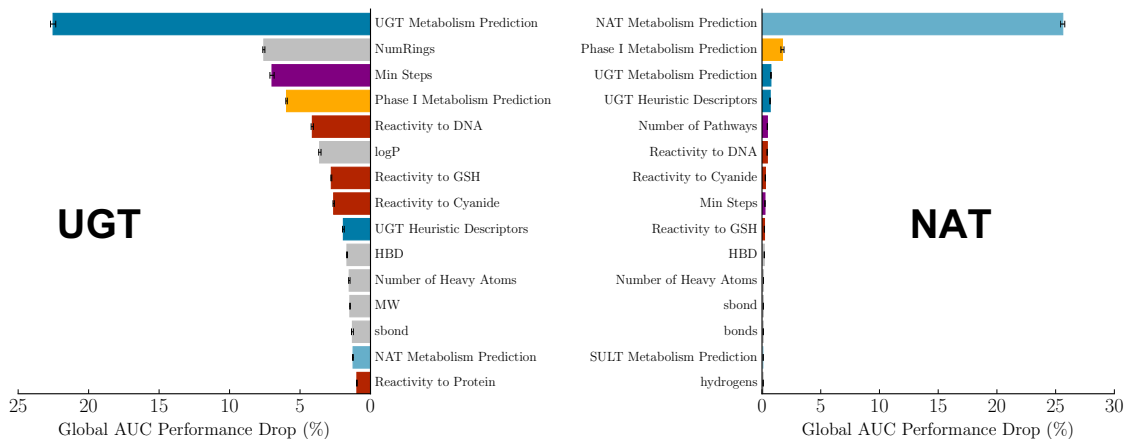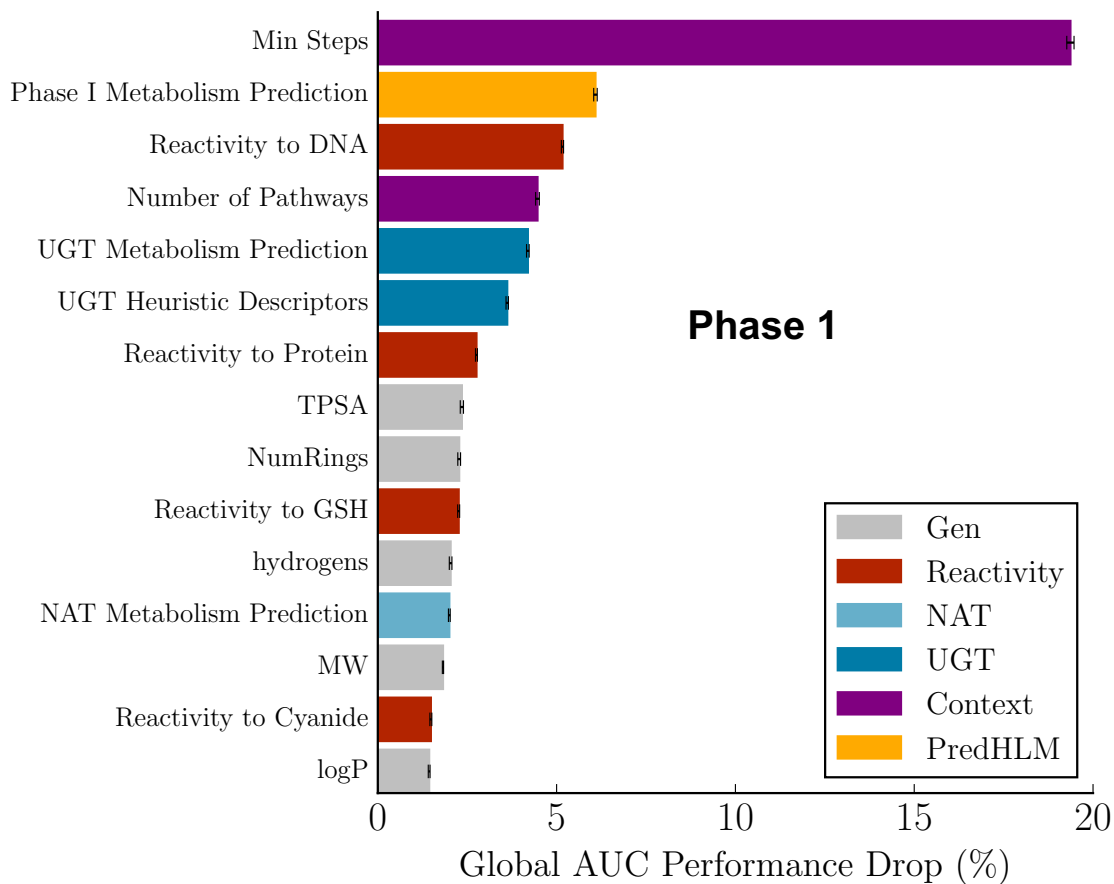
**Figure 5.3.1: Context descriptors and prediction from each in vitro model output are important to each model target.**

Estradiol
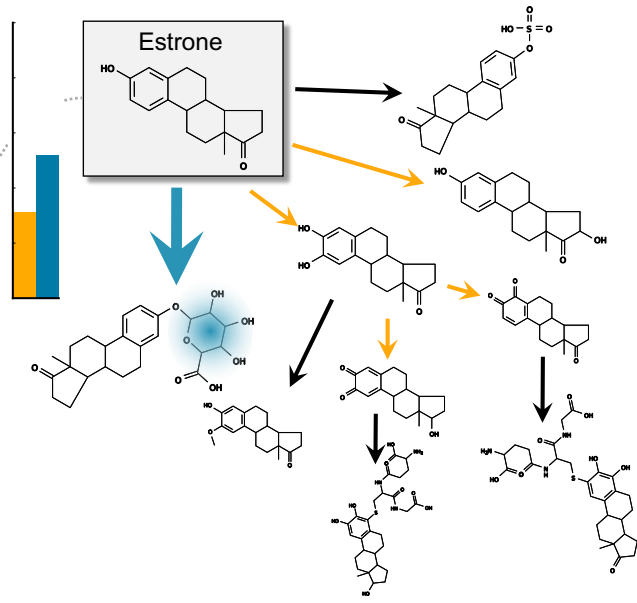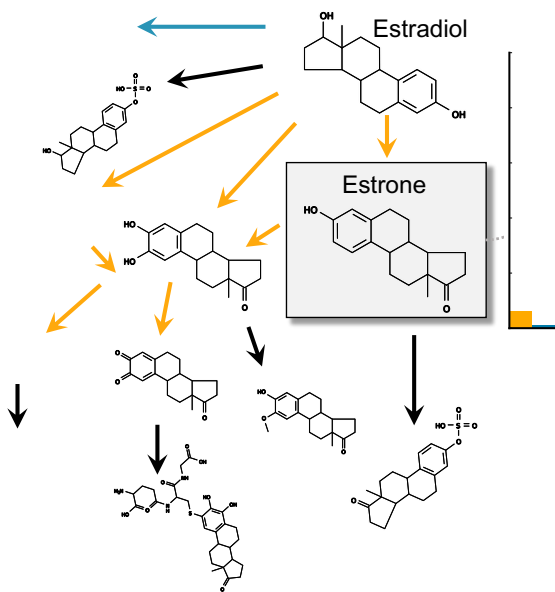
Estrone

Estrone

**Figure 5.3.2: The model accurately predict importance of UGT metabolism in irinotecan toxicity**.

Irinotecan

SN-38

Normal    w/o UGT
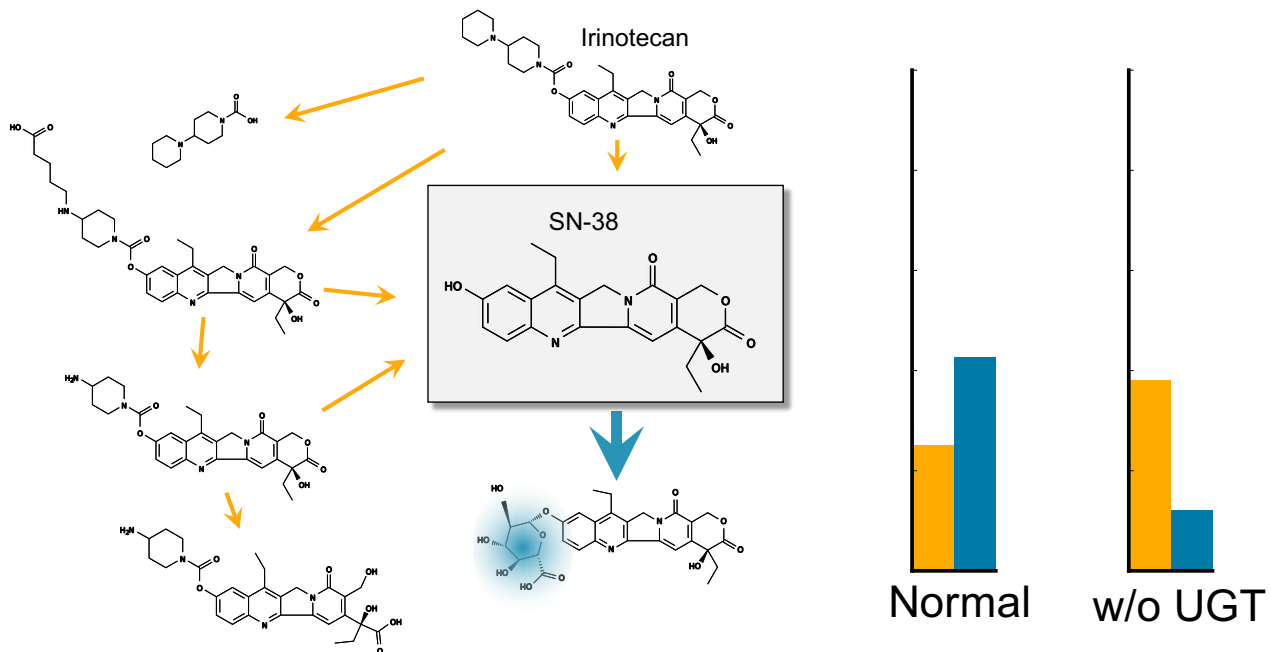
**Figure 5.3.3: The model accurately predict importance of UGT metabolism in irinotecan toxicity**.

**Tolcapone**

# 6

# Predicting Biotransformation of Structural-Alert Containing Compounds

## 6.1 Summary

Structural alerts are commonly used in drug discovery to identify molecules likely to form reactive metabolites, and thereby become toxic. Unfortunately, as useful as structural alerts are, they do not effectively model if, when, and why metabolism renders safe molecules toxic. Toxicity due to a specific structural alert is highly conditional, depending on the metabolism of the alert, the reactivity of its metabolites, dosage, and competing detoxification pathways. A systems approach, which explicitly models these pathways, could more effectively assess the toxicity risk of drug candidates. In this study, we demonstrated that mathematical models of P450 metabolism can predict the context-specific probability that a structural alert will be bioactivated in a given molecule. This study focuses on the furan, phenol, nitroaromatic, and thiophene alerts. Each of these structural alerts can produce reactive metabolites through certain metabolic pathways, but not always. We tested whether our metabolism modeling approach, XenoSite, can predict when a given molecule's alerts will be bioactivated.

Specifically, we used models of epoxidation, quinone formation, reduction, and sulfur-oxidation to predict the bioactivation of furan-, phenol-, nitroaromatic-, and thiophene-containing drugs. Our models separated bioactivated and not-bioactivated furan-, phenol-, nitroaromatic-, and thiophene-containing drugs with AUC performances of $100\%$, $73\%$, $93\%$, and $88\%$, respectively. Metabolism models accurately predict whether alerts are bioactivated and thus serve as a practical approach to improve the interpretability and usefulness of structural alerts. We expect that this same computational approach can be extended to most other structural alerts and later integrated into toxicity risk models. This advance is one necessary step towards our long term goal of building comprehensive metabolic models of bioactivation and detoxification to guide assessment and design of new therapeutic molecules.

## 6.2 Introduction

Idiosyncratic adverse drug reactions (IADRs) are a challenging problem in drug development. They are the leading reason for both termination of clinical investigation and withdrawal from the market.[132, 245] Most IADRs are hypersensitivity-driven adverse drug reactions, and arise when drugs are bioactivated into reactive metabolites.[29, 78, 166, 167, 191] Reactive metabolites frequently form covalent and noncovalent interactions with cellular macromolecules such as DNA, proteins, and lipids.[167, 177, 191] Covalent interactions can lead to cancer or trigger hypersensitivity reactions. Nonconvalent interactions can cause oxidative and other intracellular stress.[86, 113, 177, 231]

Structural alerts or toxicophores are chemical structures that can be bioactivated to generate reactive metabolites.[121] About 78-83% of drugs with a high incidence of IADRs contain structural alerts, and around 62-69% of these drugs form reactive metabolites.[226] Because structural alerts are understandable and inexpensive to apply, they are commonly used within the pharmaceutical industry, the FDA, and drug discovery tools to flag lead compounds with toxicity risk.[16, 48, 69, 96] Candidate compounds with one or more toxicophores are often chemically modified to remove the structural alerts and minimize toxicity risk.[72, 122] However, avoiding structural alerts is not always practical because alerts may be required for efficacy. Structural alerts like furan, thiophene, nitroaromatic, phenol, and aniline are building blocks with important pharmacological properties.[54, 176, 248] They can give rise to pharmacological activity, or provide pharmacokinetic benefits.

More importantly, toxicity due to a specific structural alert is conditional. A structural-alert containing compound would become harmful or remain safe depending on its metabolic pathways[209] and the reactivity of its metabolites. For example, the thiophene structural alert is ambiguous; in the case of methapyrilene, the structural alert undergoes bioactivation, while no activation of the thiophene occurs for eprosartan (Figure 6.2.1). The differences in biotransformation lead to drastic differences in the safety profiles of the two compounds. While

104

methapyrilene was withdrawn from the market due to hepatotoxicity, eprosartan is a safe and commonly prescribed anti-hypertensive. In fact, of the 200 most frequently prescribed drugs in the US, about 50% of them contain more than one structural alert.[226] However, the vast majority of these drugs are not associated with IADRs.

Because structural alerts do not adequately model metabolism, they often fail to predict whether molecules will form reactive metabolites (Figure 6.2.2).[69, 119, 226] Molecules are flagged even if (1) the structural alert is not bioactivated, (2) the reactive metabolite is quickly metabolized into a non-reactive form, or (3) an alternative, non-activating metabolic pathway is responsible for the clearance of the parent compound.[119] In addition, structural alerts can only identify toxic molecules with specific, well-known substructures; they do not identify substructures that have not yet been observed to generate reactive metabolites. For example, the formation of terbinafine's reactive metabolite—an aldehyde generated through N-dealkylation—was missed by the structural alert approach.[111] Consequently, the recommendations of structural alerts are very difficult to interpret: safe molecules are often flagged as toxic, and unsafe molecules may slip through.[9, 119]

In our prior work, we built several models of metabolism[103, 104, 156]. We hypothesize these models might more specifically identify when alerts are bioactivated, and the current study tests this hypothesis. Our method combines the knowledge of the biotransformation pathways of well-studied structural alerts with metabolism models to predict the formation of reactive metabolites. As a practical assessment of our approach, we apply individual metabolism models to predict the bioactivation of multiple structural alerts.
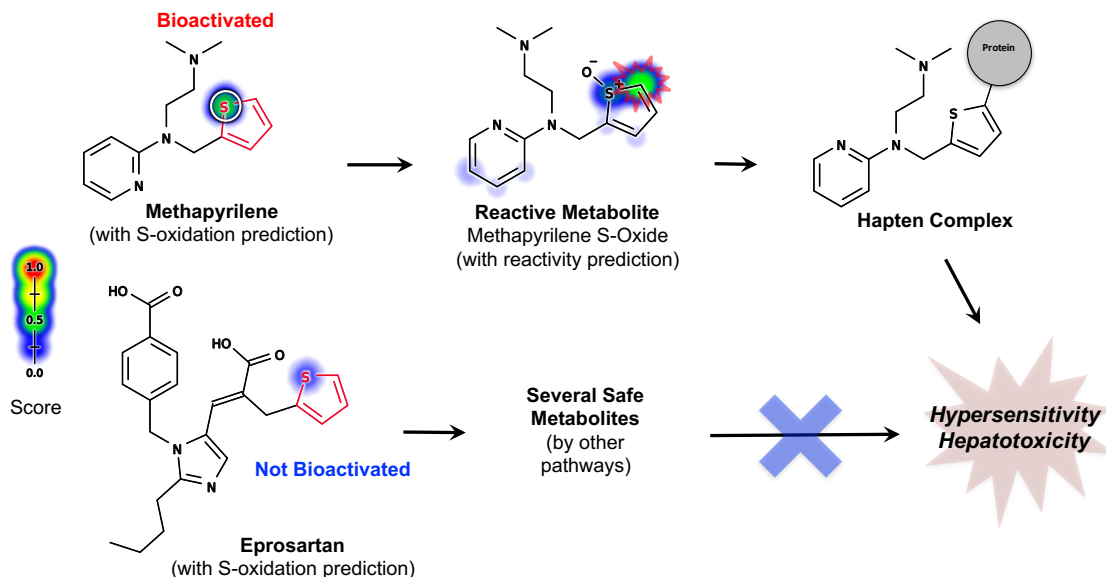
**Figure 6.2.1:** Structural alerts (highlighted in red) incorrectly flag safe drugs, because they do not adequately model metabolism. Known sites of metabolism are marked with white circles. The metabolism and reactivity[106] predictions are plotted against each atom in the molecule, with color shading ranging from red (1.0, likely) to white (0.0, unlikely). Structural alerts indiscriminately flag both bioactivated and not-bioactivated compounds as problematic. For example, both methapyrilene and eprosartan contain the thiophene structural alert, yet their toxicity profiles are very different, as predicted by metabolism models (atom shading). While methapyrilene, an antacid, was withdrawn from the market due to hepatotoxicity caused by reactive metabolites,[79] eprosartan is a safe anti-hypertensive that does not form reactive metabolites[80].

## 6.3  Materials and Methods

### 6.3.1  Epoxidation Model

Furans and thiophenes can be bioactivated by epoxidation. We use a hierarchical deep neural network to predict the probability that each alert is epoxidized. This model is based on 524 molecules, including 14 furans (Table 6.3.1), and was previously published by our group.[104] All molecules epoxidized by Human Liver Microsomes (HLMs) in the literature-derived Accelrys Metabolite Database (AMD) were used in the training set. Non-epoxidized molecules were selected for structural similarity to the epoxidized molecules. This model is available on the XenoSite web server.[156]
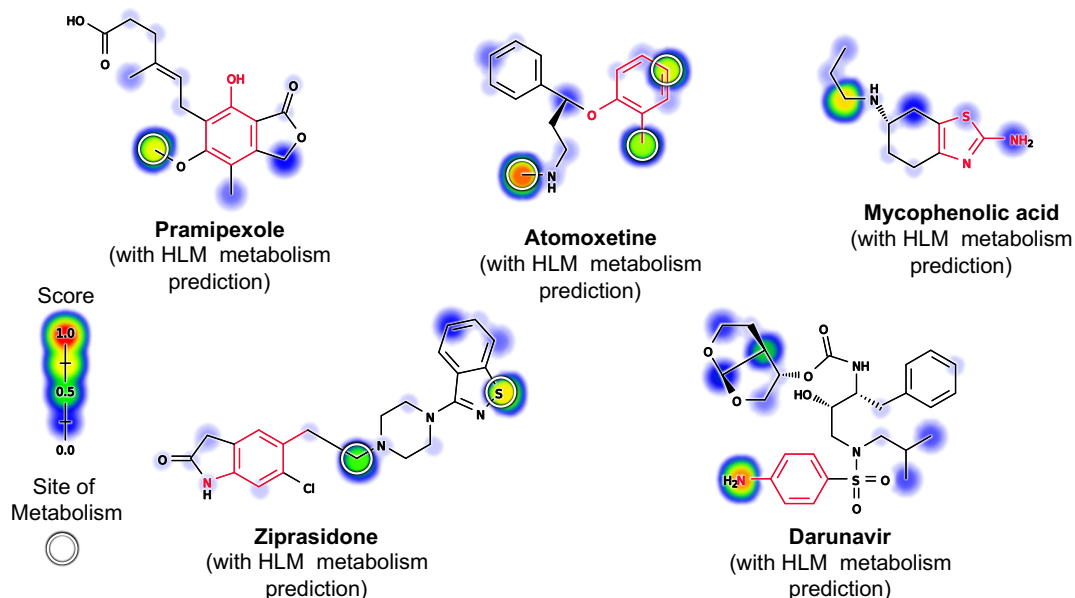
**Figure 6.2.2:** Metabolism models can identify safe molecules containing structural alerts. This figure overlays model predictions on a reproduction of a figure from Stepan et al. [226] In contrast, structural alerts (highlighted in red) incorrectly flag many safe drugs. Known sites of metabolism are marked with white circles. Predictions by one of our metabolism model[258] are plotted against each atom in the molecule, with color shading ranging from red (1.0, likely) to white (0.0, unlikely). The model predicted that mycophenolic acid, pramipexole, and ziprasidone would be metabolized outside their structural alerts to form safe metabolites. Atomoxetine is predicted to be metabolized at the *o*-alkyl aromatic ether structural alert, but the predicted metabolite is not the reactive quinone (Figure 6.4.3). Darunavir is the only incorrect prediction, because the drug is not oxidized at its aniline to form a reactive nitroso. These results were promising but preliminary, based on a previously published model that does not make metabolite specific predictions.[156, 258, 260] Building on this initial and encouraging result, this study aims to systematically test more advanced metabolism models in predicting the bioactivation of structural alerts.

### 6.3.2 Quinone Model

Phenols are bioactivated by quinone formation. We use a deep neural network that predicts whether phenols are metabolized to form quinones. This model is trained on 718 molecules, including 277 phenols (Table 6.3.1), from the AMD. This model is published in a separate study,[103] and is available on the XenoSite web server.[156]

### 6.3.3 Reduction and S-Oxidation Model

Nitroaromatics can be bioactivated by reduction and thiophenes can be bioactivated by sulfur-oxidation (S-oxidation). For these two structural alerts, we use a HLM metabolism model
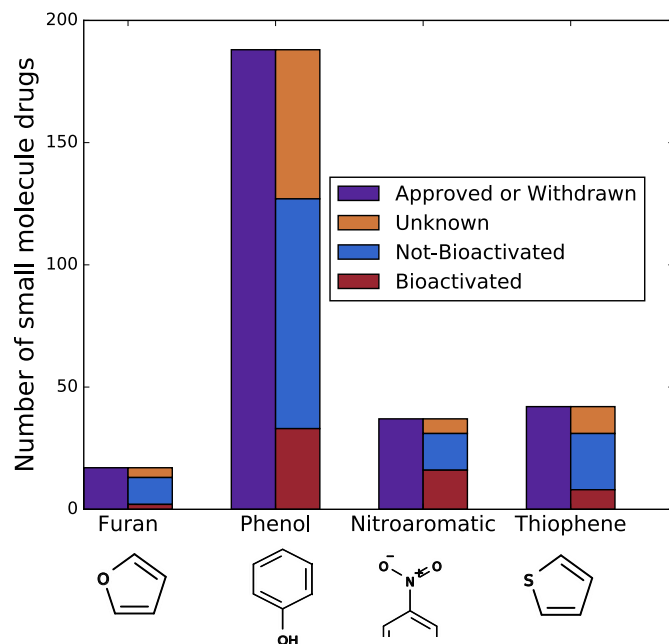
**Figure 6.3.1:** Commonly observed structural alerts that bioactivation can render toxic. Our furan, phenol, nitroaromatic, and thiophene structural alert evaluation sets contain 17, 188, 37, and 42 FDA approved or withdrawn small molecule drugs (in purple), respectively. Among those, 13 furan-, 127 phenol-, 31 nitroaromatic-, and 31 thiophene-containing drugs have been metabolically studied (blue and red). Of those metabolically studied drugs, the furan, phenol, nitroaromatic, and thiophene structural alerts are bioactivated, respectively, 23%, 26%, 51%, and 26% of the time. Nitroaromatics can include aromatic rings of any size, so only a fragment of the ring is visualized.

that predicts which part of the molecule is metabolized and what types of biotransformation (e.g. S-oxidation, hydroxylation, and/or reduction) the molecules undergoes. This model is a deep neural network that has an output for reduction and an output for sulfur oxidation. It was trained on 3061 molecules, including 98 nitroaromatics and 50 thiophenes from the AMD (Table 6.3.1). This model will be published in a separate study, but is currently available on the XenoSite web server.[156]

### 6.3.4 Scaled Predictions

The three models produce probabilistic predictions that range from zero to one. Still, it is possible the predictions are not ideally scaled to each other. We investigated rescaling the models' output by fitting the formula $z = 1/\left(1 + \exp\left(k \log\left[y/(1-y)\right] + w\right)\right)$, where $z$ is the rescaled score, $y$ is the unscaled model output, and $k$ and $w$ are tunable weights for each of the four models. This

**Table 6.3.1: Composition of the training and evaluation sets by structural alerts and metabolism models**. In the table, the number of each structural alert that is bioactivated (BA) and not-bioactivated, (NBA) is listed alongside the total size of the data set (Size). The bioactivated alerts are included in the training sets, but those not bioactivated are not always included. This is because, for the epoxidation and quinone formation models, negative examples are chosen to structurally match the positive examples, rather than to include all examples of the alert.

| Structural Alert | Model | Training Set BA | Training Set NBA | Training Set Size | Evaluation Set BA | Evaluation Set NBA | Evaluation Set Size |
|---|---|---|---|---|---|---|---|
| Furan | Epoxidation | 5 | 9 | 524 | 3 | 10 | 17 |
| Phenol | Quinone Formation | 169 | 108 | 718 | 33 | 94 | 188 |
| Nitroaromatic | Reduction | 24 | 74 | 3061 | 16 | 15 | 37 |
| Thiophene | Epoxidation | 7 | 2 | 524 | 4 | 27 | 42 |
| Thiophene | S-oxidation | 15 | 35 | 3061 | 8 | 23 | 42 |

rescaling preserves the order in which sites of metabolism are predicted by each model within its assigned structural alert. Only requiring two weights for each structural alert, the score can be trained with very small amounts of data (See Materials Section) For assessment, final rescaled scores were obtained using leave-one-out cross-validation. As we will see, rescaling does not lead to a statistically significant improvement in performance, so it may not be necessary.

## 6.3.5   Performance Comparison

We used the area under the receiver operating characteristic curve (AUC)[228] to evaluate the prediction accuracy of the metabolism model and structural alert approaches. To assess the statistical significance of different AUCs, two-tailed $p$-values were computed using the Hanley and McNeil formula.[95] To establish a baseline, we used logP: the octanol/water partition coefficient. Lipinski's well-known "rule of five" advises that highly hydrophobic drugs should be avoided, because they are more likely to be extensively metabolised.[139] Indeed, logP has been reported to directly correlate with *in vivo* toxicity, partly due to increased metabolism.[102] Because logP is an easily-computable, biologically-relevant parameter, it served as a good comparison point for our much more complex models.

### 6.3.6 Evaluation Set

We collected the structures of all FDA-approved and -withdrawn small molecule drugs from the DrugBank Database (May 2016).[251] We used SMARTS patterns specifying the structural alert of interest to filter each of the four evaluation sets: furan, phenol, nitroaromatic, and thiophene (Table 6.3.1, Figure 6.3.1). Complete lists of molecules in these evaluation sets are provided in the Supporting Information. Next, we used the AMD database to identify which of these molecules were bioactivated at their structural alert. We counted all furans, phenols, nitroaromatics, and thiophenes bioactivated for which there was a corresponding experimentally-observed reactive metabolites (or their downstream products) resulting from epoxidation, quinone formation, reduction, and S-oxidation/epoxidation, respectively.

Sufficient molecules for each structural alert were identified to test our approach. Furans were identified in $17$ approved or withdrawn drugs, $13$ of which were in AMD, and $3$ were bioactivated (Table **??**). All of the bioactivated furan examples are activated by epoxidation. Phenols were identified in $188$ approved or withdrawn drugs (Table **??**), $127$ of which were in AMD, and $33$ were bioactivated into quinones. Nitroaromatics were identified in $37$ approved or withdrawn drugs (Table **??**), $31$ of which were in AMD, and $16$ were bioactivated. All of the bioactivated nitroaromatic examples are activated by reduction. Thiophenes were identified in $42$ approved or withdrawn drugs, $31$ of which were in AMD, and $8$ were bioactivated (Tables **??**). Of the bioactivated thiophene drugs, $8$ are reported to be bioactivated by S-oxidation, $4$ are reported to be bioactivated by epoxidation, and $8$ are reported to be bioactivated by both S-oxidation and epoxidation.

### 6.3.7 Evaluation Set Predictions

In some cases, molecules in the evaluation set were also in the training sets. To ensure unbiased predictions on these molecules, we used a hold-out prediction. In turn for each validation molecule, the model was retrained with all the molecules except the validation molecule, and the predictions from the newly trained model was used. This approach ensured unbiased predictions,

even when molecules were in both sets.

## 6.4   Results and Discussion

The following sections use metabolism models to predict the bioactivation pathways for four commonly observed structural alerts: furans, phenols, nitroaromatics, and thiophenes. In turn, we evaluate this approach's performance with each alert. Next, we combine all four models to predict bioactivation of molecules containing any of the four structural alerts. This combined model is a more practical tool for identifying possible problematic metabolic pathways for functionally and structurally diverse molecules. Finally, we discuss the limitations of this approach and solutions for future studies to improve estimations of toxicity risk.

### 6.4.1   Furans

Furans are oxygen-containing five member aromatic rings that are commonly found in drugs, food, nutraceuticals, the environment, and industrial pollutants.[195, 213, 227, 248, 262] Furans can be bioactivated via epoxidation (Figure 6.4.1). For example, furosemide is a frequently prescribed diuretic that sometimes causes idiosyncratic hepatitis, which may be due to epoxidation of its furan ring.[154, 169–171, 226] This epoxide metabolite is electrophilically reactive and conjugates to nucleophilic sites within proteins. The resulting adducted protein serves as a hapten to induce toxic immune responses.[154, 226, 250] However, many furan-containing drugs are not toxic. For example, the H2 antagonist ranitidine does not provoke toxicity despite a high therapeutic dose because its furan is not bioactivated.[23, 226]. Significantly, many drugs contain furans, but only 23% undergo bioactivation (Figure 6.3.1).

We applied the epoxidation model to predict which furan-containing drugs undergo bioactivation at the alert.[104] Bioactivated furans were identified with an AUC of 100%, statistically outperforming using the structural alert alone, with a two-sided $p$-value of 0.01 (Figure 6.4.2). This performance also is significantly better than the baseline AUC of 60% using
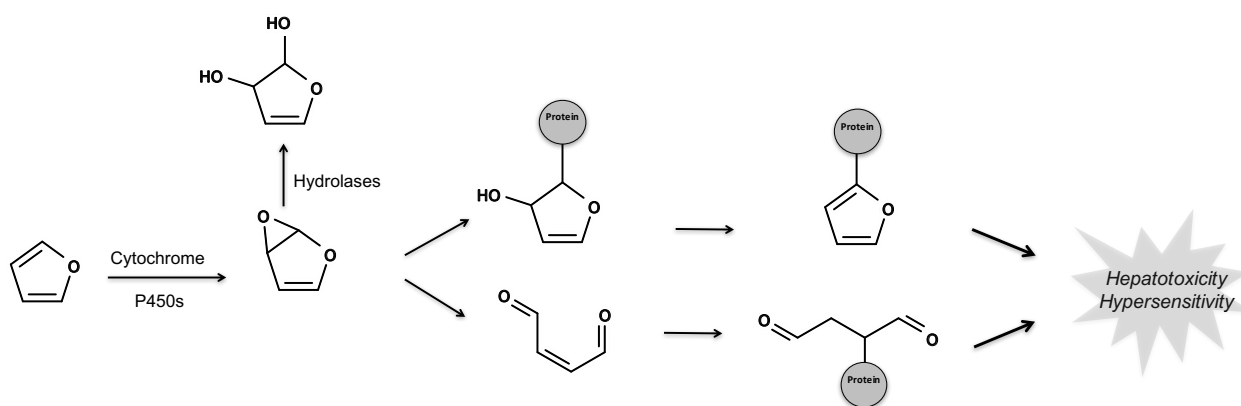
**Figure 6.4.1:** Furan is bioactivated by epoxidation. Cytochrome P450s bioactivate furans via epoxidation. The resulting epoxide is highly reactive to ring strain and polarized carbon-oxygen bonds. Consequently, the epoxide can react directly with proteins, or first undergo ring scission to form a reactive cis-enedione that then conjugates to proteins.[195]

logP (two-sided *p*-value of 0.022). The bioactivated drugs included furosemide[154, 226, 250], methoxsalen,[127] and prazosin[70], and the not-bioactivated drugs were vilazodone, fluticasone, lapatinib, ranitidine, amiodarone, dantrolene, nitrofurantoin, furazolidone, nitrofural, and dronedarone. Encouragingly, the model perfectly identified bioactivated furans with 100% accuracy. Of note, we even correctly predicted that methoxsalen is epoxidized, despite the drug being mislabeled as non-epoxidized in our training data drawn from the AMD. Upon further investigation, we found a source omitted by the AMD that reported methoxsalen's epoxidation at its furan structural alert.[127] We found it reassuring that our model revealed an error in our curated source data, because it is an evidence that the model was not overtrained.

## 6.4.2 Phenols

Phenols contribute significantly to biological and pharmacological properties and thus are found in many drugs. In fact, phenols are one of the most frequently observed structural alerts, present in about 10% of all drugs on the market. Unfortunately, this important structure can be readily converted into quinones (Figure 6.4.3). Quinone species, such as quinone-imines and quinone-methides, are electrophilic Michael acceptors that are often highly reactive, and comprise over 40% of all known reactive metabolites.[233] At the same time, many safe drugs contain

112
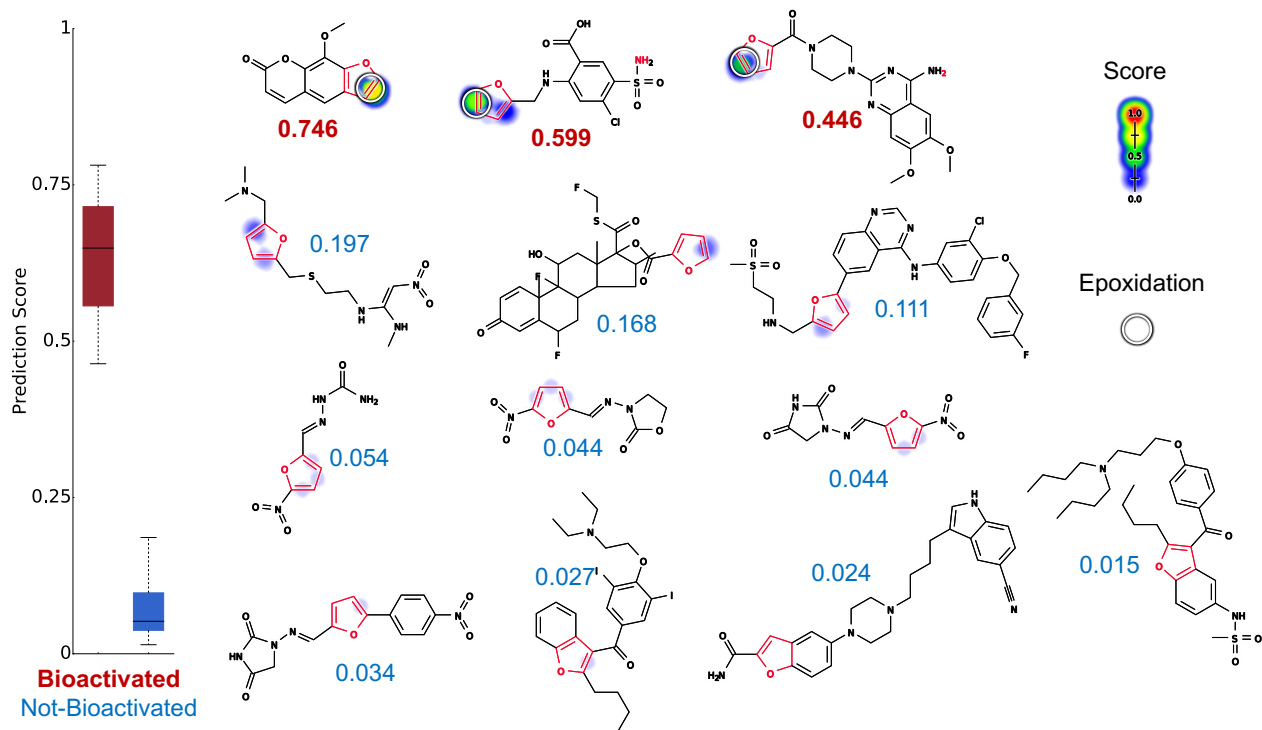
**Figure 6.4.2:** Metabolism model identifies which furans (highlighted in red) are bioactivated. From left to right, top to bottom, the molecules are methoxsalen,[127] furosemide,[154, 226, 250] prazosin,[70] ranitidine, fluticasone, lapatinib, nitrofural, furazolidone, nitrofurantoin, dantrolene, amiodarone, vilazodone, and dronedarone. Experimentally-observed sites of epoxidation are indicated by white circles. For each molecule, the colored shading represents bond epoxidation scores, which range from 0 to 0.746. The model's AUC accuracy on the furan evaluation set is 100% (statistically outperforms structural alert approach, two-sided $p$-value = 0.01). Notably, methoxsalen (the highest ranked molecule) was not epoxidized in the AMD. However, this is an omission in the AMD data set; methoxsalen is actually epoxidized, and is counted here as a positive drug.[127] Markedly, the model correctly notes that terminal furans (with just one substituent) are most likely to be bioactivated, but is still correctly recognizes the one exception (fluticasone) to the rule.

phenol and do not form quinones. Consequently, phenol's presence alone is not necessarily indicative of quinone formation. Furthermore, phenols may be "hidden" until phenyl rings undergo hydroxylation *in vivo* to form the phenol. Such cases would be missed by the structural alert approach. Avoiding phenyl rings in drug development is impractical. Therefore, an accurate method for identifying which phenols are actually at risk for quinone formation would be of great value.

We recently developed the quinone formation model, which accurately predicts quinone formation across diverse chemicals.[103] On the phenol data set, the quinone formation model
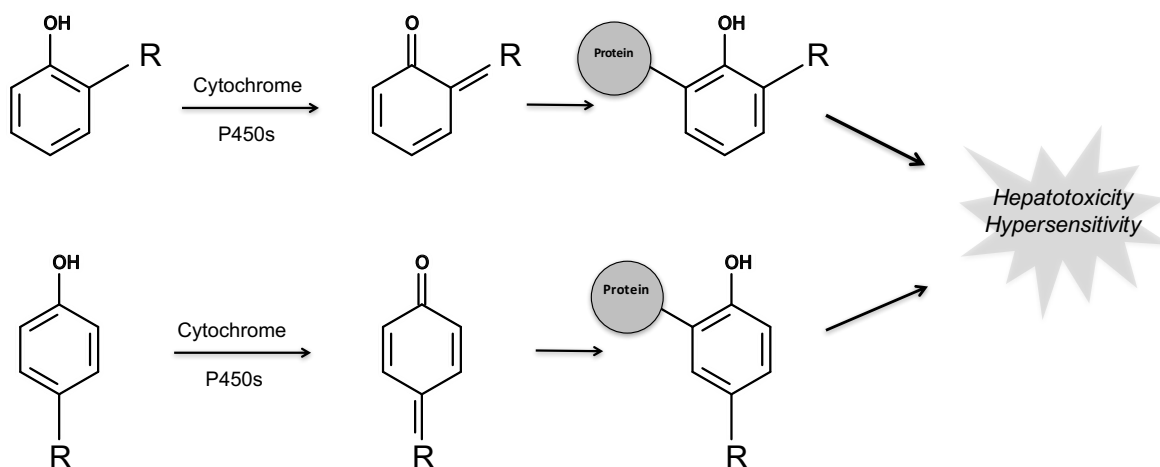
**Figure 6.4.3:** The phenol structural alert is bioactivated to form quinones by cytochrome P450. Quinones, such as quinone-imines and quinone-methides, are electrophilic Michael acceptors that are often highly reactive, and comprise over $40\%$ of all known reactive metabolites.[233] Consequently phenols are one of the most important types of structural alerts. Furthermore, any phenyl ring—an unavoidable building block of many drugs—can be subject to aromatic hydroxylation, thereby forming a phenol. "R" represents a carbon, oxygen, or nitrogen, which in conjunction with the phenol oxygen can form a quinone-methide, quinone, or quinone-imine, respectively. The R group can be either *ortho* or *para* to the phenol oxygen.

separated the quinone-forming molecules from the other molecules with an AUC of $73\%$ (Figure 6.4.4).

Quinone formation is more complicated than the other pathways in this study, because the process can involve multiple steps. The complexity of this biotransformation likely explains the lower accuracy of $73\%$. Nevertheless, this performance is still better than the structural alert AUC of $50\%$ (two-sided *p*-value = 0.0026) and the baseline logP AUC of $50.5\%$ (two-sided *p*-value of 0.003).

## 6.4.3   Nitroaromatics

Nitroaromatics are abundant in nature and the urban environment.[198, 203] We are frequently exposed to nitroaromatic compounds through daily activities like smoking, inhaling combustion gases, and consumption of grilled food.[198, 203] The nitroaromatic alert is also an important building block in many pharmaceutical agents.[176, 247] Unfortunately, some of these drugs cause adverse effects due to reactive metabolite formation.[176, 247]
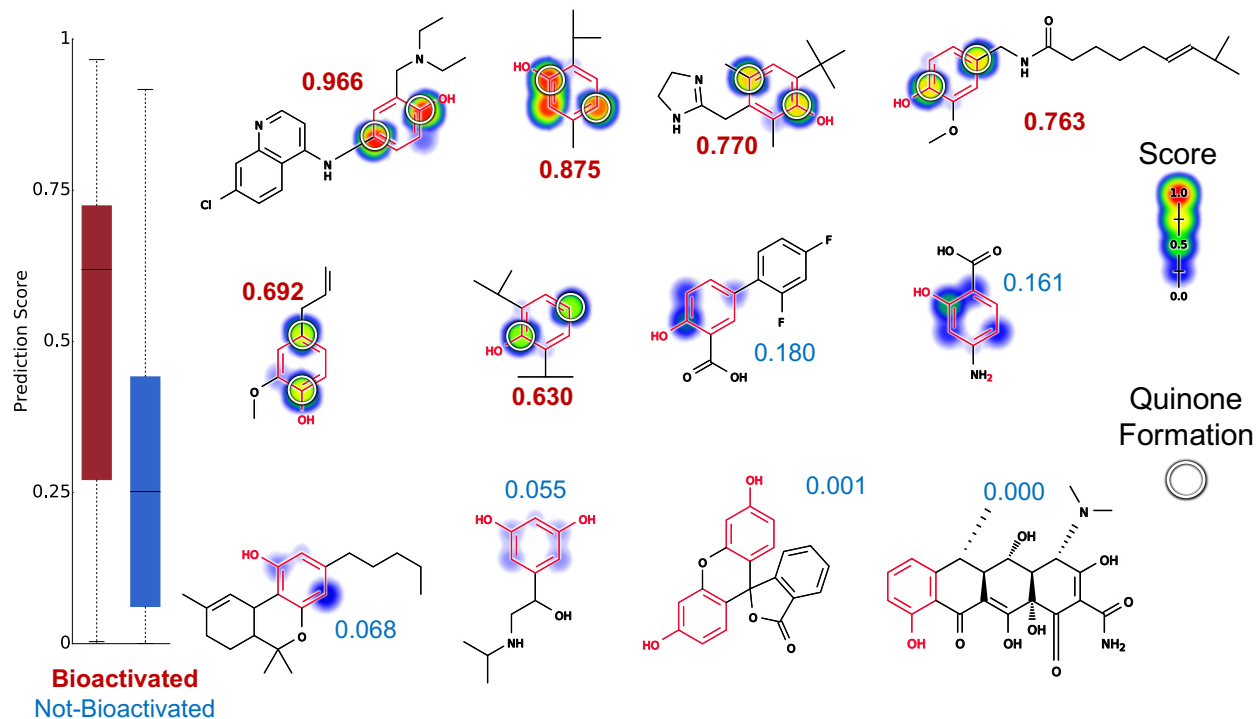
**Figure 6.4.4:** Metabolism model identifies which phenols (highlighted in red) are bioactivated into quinones. Twelve examples from the phenol evaluation set, from left to right, top to bottom: amodiaquine,[44, 46, 146] thymol,[234] oxymetazoline,[147] capsaicin,[201] eugenol,[168] propofol,[34] diflunisal, aminosalicylic acid, dronabinol, orciprenaline, fluorescein, and doxycycline. The remaining molecules and their prediction is reported in the Supporting Information (Table **??**). Attached numbers are the molecule quinone formation score, with red for the bioactivated phenols and blue for the rest. Experimentally-observed sites of quinone formation are indicated by white circles. For each molecule, the colored shading represents quinone site scores, which range from 0 to 0.97. The model's AUC accuracy on the phenol evaluation set is 73%, and better than the structural alert alone (two-sided $p$-value = 0.0026).

Nitroaromatic-containing compounds can undergo reductive metabolism through different pathways to form electrophilic, unstable intermediates such as nitroso and nitro anion radical (Figure 6.4.5). These reactive metabolites are responsible for the toxicity of some nitroaromatic containing drugs like tolcapone[184] and nimesulide.[136] On the other hand, many drugs like aranidipine[235] and nifedipine[205] are safe and lack reduction at their nitroaromatic group.

We used a reduction model to predict which nitroaromatic-containing drugs are bioactivated through reduction (Figure 6.4.6). The reduction model identified bioactivated nitroaromatics with an AUC of 93%, which significantly outperforms the structural alert alone (two-sided $p$-value = 0.0002). This performance also is significantly better than the baseline AUC of 77% using logP
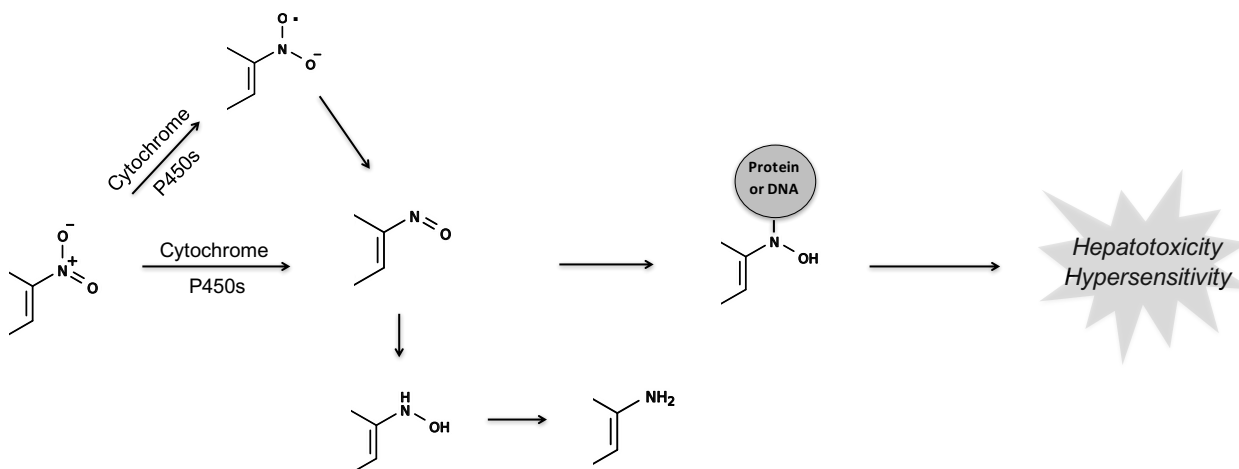
**Figure 6.4.5:** Nitroaromatics are bioactivated through reduction. Nitroaromatic compounds undergo sequential two-electron reductive steps to the nitroso, N-hydroxy, and amine. Alternatively, they can form nitro anion radical through a one-electron reduction in the absence of oxygen. The reaction chain can also be reversed when an aromatic amine is oxidized to the N-hydroxy/nitroso compound. However, because the intracellular environment is reducing at physiological conditions, the equilibrium usually shifts toward the right.[27, 124]

(two-sided $p$-value of 0.048). Compared to the other structural alerts examined in this study, the nitroaromatic moiety is the most frequently bioactivated (51%) (Table 6.3.1). Nevertheless, nitroaromatic groups can be part of the pharmacophore and nitro reduction is required for drugs like antibiotic nitazoxanide and metronidazole to exert their pharmacological effects.[109, 222] So, this strong performance is encouraging, and might effectively guide safe use of nitroaromatics.

### 6.4.4  Thiophenes

Thiophene derivatives are ubiquitous in the environment, as well as many drugs on the market.[54] Thiophene-containing compounds have a wide range of pharmacological properties, such as nematocidal, insecticidal, antifungal, antiviral, and antioxidant effects[17, 38, 101, 112, 148, 157]. For example, thiophene is a critical pharmacophore of anti-thrombotic drugs like clopidogrel.[73]

Thiophenes can undergo oxidative metabolism through different pathways to form electrophilic, unstable intermediates such as thiophene S-oxides, thiophene epoxides, and sulfenic acids (Figure 6.4.7).[54, 61, 62, 150] Formation of reactive, electrophilic intermediates by

**Table 6.4.1: AUC Accuracies of metabolism models on the thiophene training and testing data sets.** Rows correspond to prediction scores and columns correspond to classification tasks (e.g. S-oxidized versus not-S-oxidized). Highest accuracies are bolded, accuracies that are not statistically different from the highest prediction are italicized, and AUC accuracies that are statistically better than the accuracy of the random model are asterisked.

| | Training Set | | | Evaluation Set | | |
|---|---|---|---|---|---|---|
| | S-oxidized | Epoxidized | Bioactivated | S-oxidized | Epoxidized | Bioactivated |
| S-Oxidation score | 92.2%* | 69.6% | 93.7%* | 87.9%* | 85.3%* | 87.9%* |
| Epoxidation score | 57.8% | 65.8% | 67.6% | 67.9% | 60.0% | 67.9% |
| Combined score | 82.5%* | 66.4% | 86.7%* | 75.4%* | 70.7% | 75.4%* |

oxidative metabolism of thiophenes can induce toxicity, as reported for suprofen and tienilic acid. [22, 61, 67, 150, 159, 182]. On the other hand, thiophene-containing drugs like eprosartan and rivaroxaban are neither bioactivated nor toxic.

There is disagreement in the literature about the dominant pathway of bioactivation in thiophenes. Some have argued that, at least in specific cases like suprofen, epoxidation is more important than S-oxidation.[114, 182] However, others have argued that S-oxidation is generally more important.[50, 79, 82, 91, 117, 149, 161, 163] In this analysis, therefore, we consider both S-oxidation and epoxidation, and aim to study empirically which model best discriminates bioactivated molecules.

We evaluated three models: the epoxidation model, the S-oxidation model, and a model that mathematically combines predictions from both models. The final model combines epoxidation and S-oxidation predictions using the probabilistic OR function. This combined score reflects the probability that a thiophene will be bioactivated by either pathway if both models are relevant. These three models were assessed by their performances at identifying bioactivated thiophenes (Table 6.4.1).

We find that the S-oxidation model identifies bioactivated thiophenes with an AUC of 88%, better than both the epoxidation and combined models. Remarkably, the S-oxidation model separates epoxidized from not-epoxidized thiophenes better than the epoxidation model (AUC of

85% versus 60%). This surprising result reflects the controversy in the literature about the dominant pathway of thiophene bioactivation.

Certainly, some thiophenes are epoxidized,[182] but in most research studies the experiments necessary to discriminate between epoxidation and S-oxidation are not performed. Both the S-oxide and epoxide metabolites result from transfer of oxygen to the nucleophilic S-atom or double bond center of the thiophene. Consequently, S-oxide and epoxide metabolites look identical in mass spectrometric analysis used in most drug metabolism studies, as both are 16 mass units higher than the parent compound without a change in charge.[68, 179, 182]. Indeed, uncertainties in the site or type of reactions are quite common in metabolic data, but for this case are consequential because most studies cannot reliably determine which pathway leads to the reactive metabolite. Moreover, both pathways can produce 5-hydroxyl thiophene and other downstream metabolites (Figure 6.4.7). When a downstream metabolite is observed, the presence of both short-lived S-oxide and epoxide thiophene are inferred. In this context, it is possible that there is a tendency to misreport S-oxidations as epoxidations, especially if thiophene epoxidations are less common than S-oxidation.

The model's output in suprofen has direct relevance to the debate in the literature about S-oxidation and epoxidation of thiophene rings. O'Donnell et al. conclude that suprofen 5-hydroxythiophene was formed via epoxidation based on $^{18}$O incorporation analysis.[182] This evidence, however, does not exclude suprofen bioactivation via S-oxidation. Our model predicts that epoxidation is slightly more likely than S-oxidation in this case (probability 0.63 vs 0.49), but this is far from definitive. Encouragingly, our model's assessment is echoed by the same authors two years later in a review of bioactivation[121], where they note that suprofen might be epoxidized instead of being S-oxidized, as they had originally reported. It would be interesting to see if the epoxidation pathway, which is predicted slightly more likely, could be confirmed in a more discriminative experiment. This, however, is beyond the scope of the current study, and we are encouraged that our model produces results consistent with the literature.[80]

Figure 6.4.8 depicts 12 example molecules: suprofen,[80, 182] zileuton,[117] ticrynafen

(tienilic acid),[143] methapyrilene,[93] duloxetine,[37] tiaprofenic acid,[] eprosartan,[158] brotizolam,[216] rotigotine, rivaroxaban,[] dorzolamide, and olanzapine.[35, 47]. The first four drugs have been reported to form reactive metabolite through either S-oxidation or epoxidation catalyzed by cytochrome P450, while the other seven drugs do not undergo bioactivation within thiophene rings. All of these bioactivated drugs have been withdrawn from the market for reactive-metabolite-related toxicity (Table **??**). The model AUC of $88\%$ accuracy is better than the structural alert AUC of $50\%$ (two-sided $p$-value $= 0.009$) and the baseline logP AUC of $50.8\%$ (two-sided $p$-value of $0.01$).

## 6.5   Combining the Alerts

In practice, molecules can have more than one structural alert at a time. An integrated model that can predict the bioactivation of all alerts in a molecule is important in this case. Theoretically speaking, it is possible for each model to correctly predict the bioactivation of each alert (Table 6.5.1), but then fail when combined if they are poorly scaled and require different cutoff scores for each alert. Encouragingly, the global AUC across all alerts in the combined evaluation data set of $188$ molecules is $74.0\%$, indicating that this issue is not a limiting problem with this approach (Figure 6.5.1). Moreover, the AUC can be improved to $81.1\%$ by rescaling the scores of each model using the relevant structural alerts evaluation set (Figure 6.5.1). This improvement is not statistically significant. Nonetheless, the result is encouraging and suggests that a similar scaling may be useful as we expand to additional structural alerts in the future. Furthermore, across all molecules in the combined evaluation set that are bioactivated, the combined model assigns the bioactivated structural alert with higher score than the rest of molecule $100\%$ of the time (both unscaled and scaled).

119

**Table 6.5.1: Accuracy on training and evaluation sets of four structural alerts**. Sensitivity and specificity were calculated with the optimal cutoff point (closest to the upper left corner) on the ROC curve. AUC accuracies that are statistically better than the accuracy of the random model are asterisked.

| Structural Alert | Model | Training Set | | | Evaluation Set | | |
|---|---|---|---|---|---|---|---|
| | | AUC | Sensitivity | Specificity | AUC | Sensitivity | Specificity |
| Furan | Epoxidation | 53.3% | 60.0% | 66.7% | 100.0%* | 100.0% | 100.0% |
| Phenol | Quinone Formation | 85.7%* | 81.1% | 78.7% | 72.4%* | 61.3% | 76.1% |
| Nitroaromatic | Reduction | 71.4%* | 70.8% | 74.3% | 92.9%* | 87.5% | 86.7% |
| Thiophene | S-Oxidation | 87.4%* | 93.3% | 74.3% | 87.9%* | 87.5% | 78.2% |

**Table 6.5.2: Sensitivity and specificity at the optimal cutoff points on the ROC curves of unscaled and scaled scores on the combined evaluation set.** The scaling does not reorder predictions, so the AUC remains unchanged from Table 6.5.1.

| Structural Alert | Model | Unscaled Score | | Scaled Score | |
|---|---|---|---|---|---|
| | | Sensitivity | Specificity | Sensitivity | Specificity |
| Furan | Epoxidation | 100.0% | 100.0% | 100.0% | 100.0% |
| Phenol | Quinone Formation | 80.6% | 45.6% | 54.8% | 89.1% |
| Nitroaromatic | Reduction | 62.5% | 100.0% | 93.7% | 73.3% |
| Thiophene | S-Oxidation | 62.5% | 73.9% | 50.0% | 91.3% |

## 6.6 Limitations and Future Work

The most obvious limitation of this approach is that it only includes four structural alerts. Clearly, bioactivation of other structural alerts will not be handled by the current method. Nevertheless, this quality is not an intrinsic limitation, and our future work will expand this approach to all commonly observed alerts. While not definitive, these initial results are encouraging, and suggest that metabolism modeling is broadly applicable. Our models performed far better than structural alerts alone, and made quantitative predictions about how drugs will be bioactivated. As is expected in any modeling strategy, the current models do not make perfect predictions all cases. However, this strategy benefits from improved modeling approaches and data as they are developed, and we expect performance improvements as we refine our models. Ultimately, this approach might be most useful in identifying and prioritizing the most likely bioactivation pathways for a given molecule for follow-up. Mechanistic and systematic predictions of

bioactivation enable focused experimental studies to confirm or rule out reactive metabolites in specific cases. More significantly, some metabolites that are formed from structural alerts are not reactive. For example, only a subset of quinones formed from phenols are actually reactive. Side chains can tune quinone reactivity up or down, and we expect better utility by combining predictions of quinone formation with predictions of the reactivity of quinone metabolites. We have already published accurate models of metabolite reactivity,[105, 106] and plan to combine these with predictions of metabolites in our future work.

Perhaps most importantly, the toxicity profile of a drug reflects an intricate interplay of multiple factors, including dosage, competing metabolic pathways, the reactivity of its metabolites, and coadministration of other medicines. For example, reducing daily dose to under 20 mg/day can substantially reduce toxicity risk.[119] These factors can be complex and hard to model. For instance, while acetaminophen overdose accounts for more than 50% of drug-induced primary liver failure cases in the U.S., the drug is generally safe when consumed at its therapeutic dose.[86, 123] At therapeutic doses, 85% of the administered acetaminophen undergoes phase II conjugation reactions and is primarily excreted in the urine as the corresponding O-glucuronide or O-sulfate[26]. Only 15% of the administered dose undergoes phase I oxidation reaction to form the reactive $N$-acetyl-$p$-benzoquinoneimine (NAPQI) species. At safe doses of acetaminophen, the small amount of NAPQI is neutralized and removed by reduced glutathione (GSH) through conjugation.[121] However, at higher doses, elevated amounts of NAPQI overwhelm GSH reserves and the reactive NAPQI starts covalently binding to macromolecules and eventually disrupts cellular homeostasis. In future studies, we will move closer to predicting toxicity by integrating individual metabolism models[56, 258–260] with further tools to incorporate competing and sequential processes that occur *in vivo*, and then couple them to models of reactivity,[104, 105] a key driver of drug toxicity.

## 6.7 Conclusion

Structural alerts are commonly used to identify molecules likely to produce reactive metabolites. Unfortunately, alerts are not precise, and incorrectly flag many safe molecules. This study demonstrated that metabolism models can improve the specificity of structural alerts by computationally modeling the relevant metabolism pathways. Our models predicted epoxidation of furans, quinone formation of phenols, nitrogen-reduction of nitroaromatics, and sulfur-oxidation of thiophenes with AUC performances of 100%, 73%, 93%, and 88%, respectively. While we have not comprehensively covered all structural alerts, or handled all the relevant complexities, our success suggests that computationally modeling metabolism could improve the interpretability of many structural alerts. Ultimately, we envision that models of metabolism coupled to models of toxicity will form a powerful new approach for assessing the IADR risk of drug candidates.

## 6.8 Abbreviations

AMD, Accelrys Metabolite Database; AUC, Area Under the Receiver Operating Characteristic Curve; BA, Bioactivated; NBA, Not Bioactivated; FDA, U.S. Food and Drug Administration GSH, Glutathione; HLM, Human Liver Microsomes; IADR, Idiosyncratic Adverse Drug Reactions; NAPQI, *N*-acetyl-*p*-benzoquinoneimine; P450, Cytochrome P450s
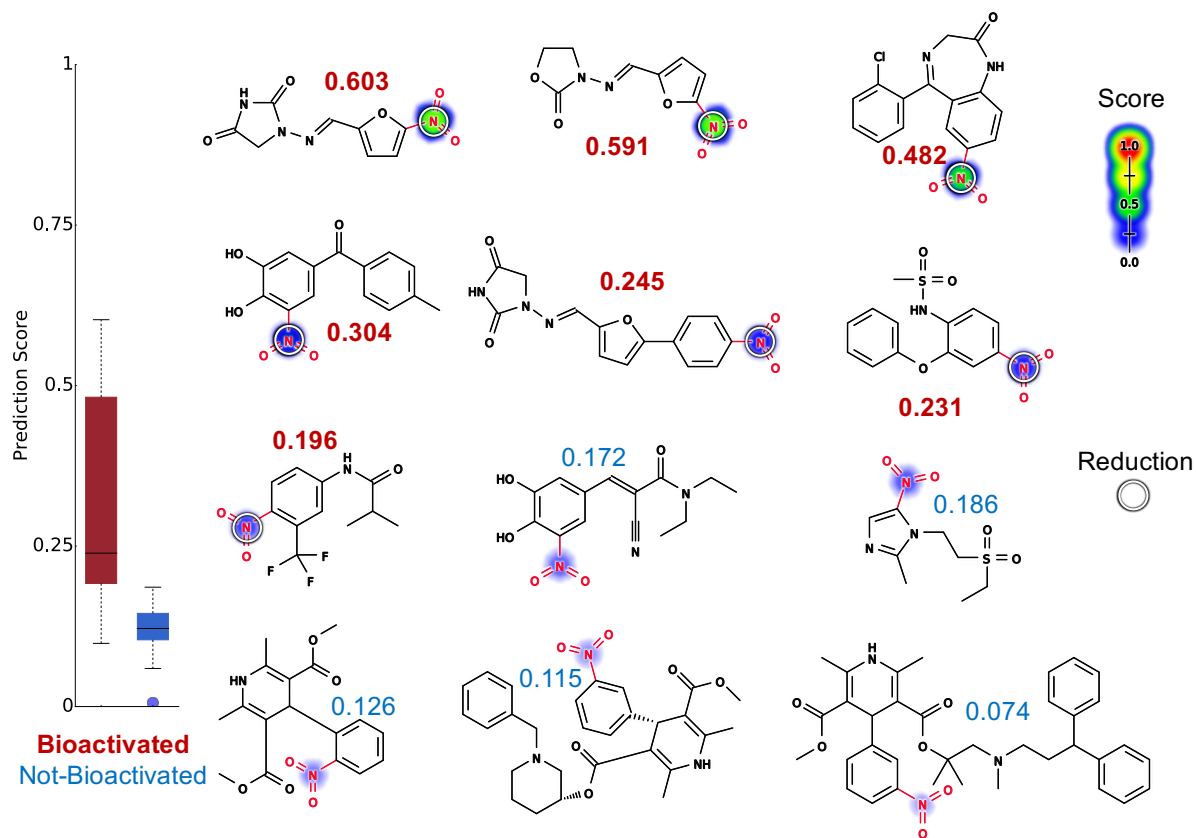
**Figure 6.4.6:** Metabolism model identifies which nitroaromatics (highlighted in red) are bioactivated. Twelve examples from the nitroaromatic evaluation set, from left to right, top to bottom: nitrofurantoin,[180] furazolidone,[6] clonazepam,[205] tolcapone,[184] dantrolene,[136, 238] nimesulide,[136] flutamide,[136] entacapone,[27, 133] tinidazole,[76] nifedipine,[205] benidipine,[] and lercanidipine.[] The remaining molecules and their prediction is reported in the Supporting Information (Table **??**). Site reduction score are indicated by the color shading, ranging from red (1.0 highly probable) to white (0.0, low probability). Experimentally-observed sites of reduction are indicated by white circles. Attached numbers are molecule reduction scores. The reduction model was able to assign drugs that are known to be bioactivated at nitroaromatic structural alerts (scores in red) with higher score than those that do not (scores in blue). The model AUC accuracy on the nitroaromatic evaluation set is $93\%$ (significantly outperforms structural alert approach, two-sided *p*-value = 0.0002).
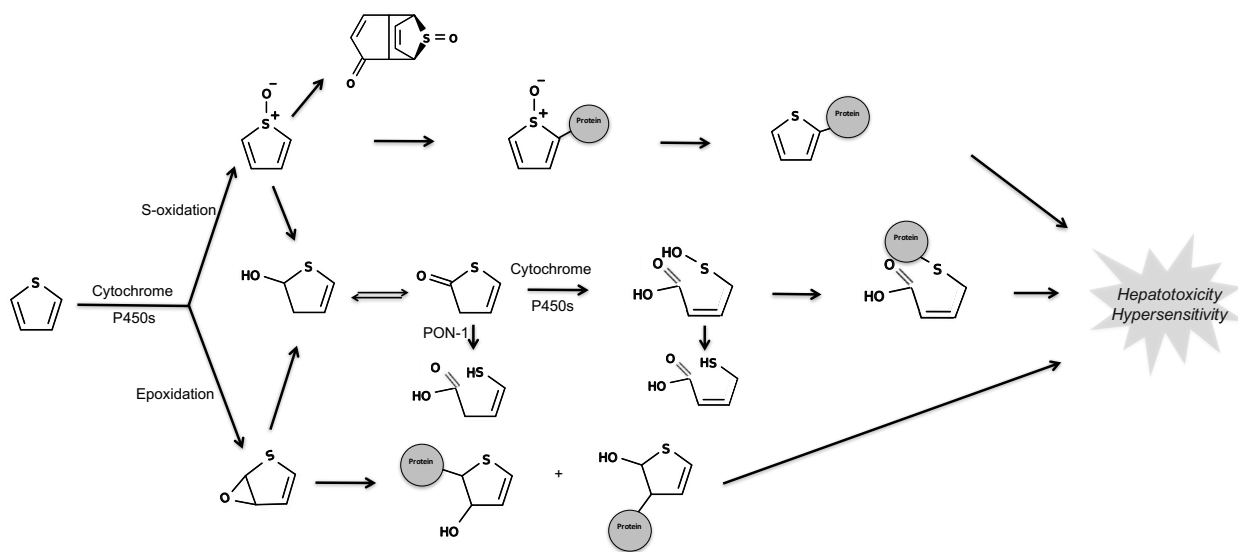
**Figure 6.4.7:** Thiophene is bioactivated through S-oxidation and epoxidation. The two bioactivation pathways of thiophene structural alerts are epoxidation and S-oxidation.[54, 61, 80] Cytochrome P450s are the major catalyzing enzymes of both pathways. These pathways yield electrophilic, unstable thiophene S-oxides,[62] thiophene epoxides,[182] and sulfenic acids.[150] Notably, mass spectrometry experiments cannot reliably determine which bioactivation pathway leads to reactive metabolites, because the thiophene S-oxide and the thiophene epoxide have the same mass and charge. This creates some ambiguity in the literature about the exact pathway by which some thiophenes are bioactivated.
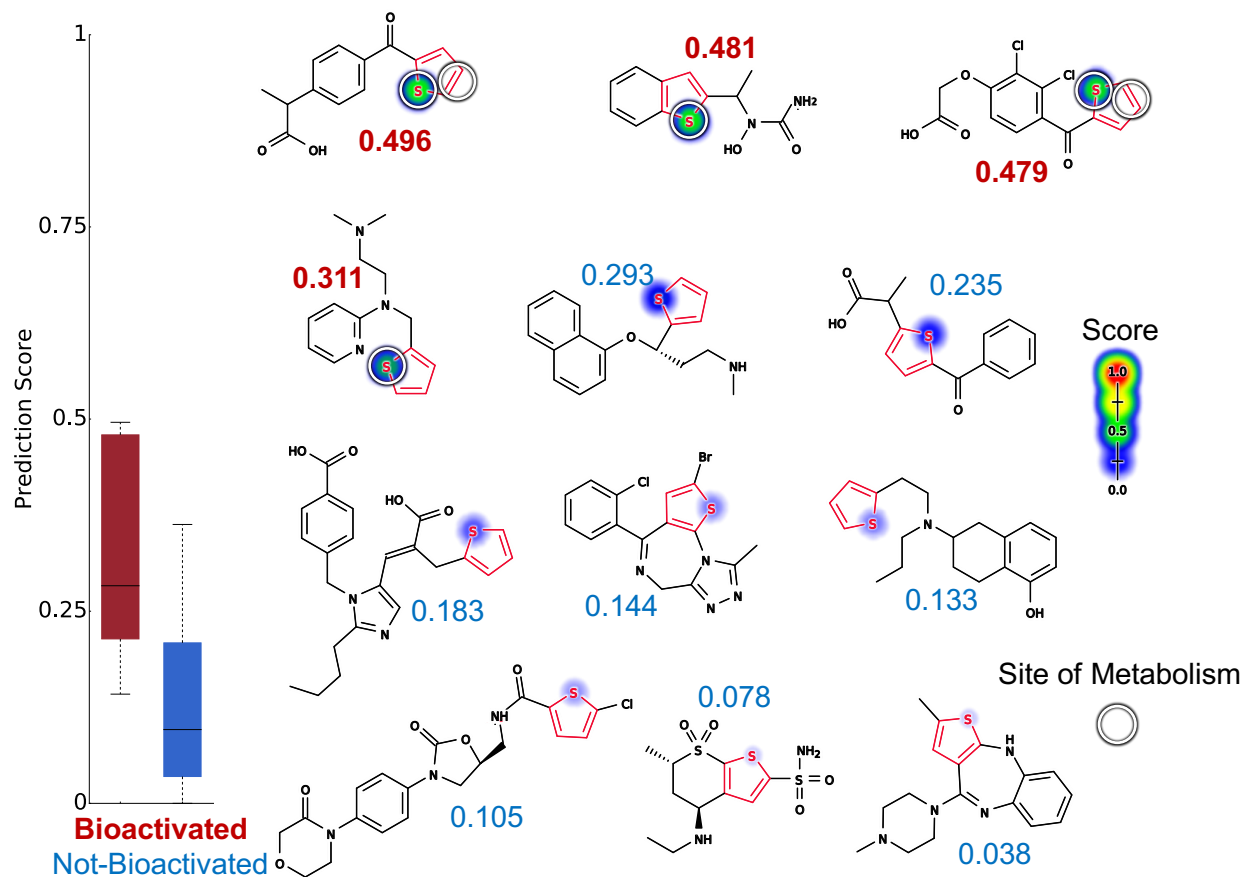
**Figure 6.4.8:** Metabolism model identifies whether thiophenes (highlighted in red) are bioactivated. Twelve examples from the thiophene evaluation set, from left to right, top to bottom: suprofen,[80, 182] zileuton,[117] ticrynafen (tienilic acid),[143] methapyrilene,[93] duloxetine,[37] tiaprofenic acid,[] eprosartan,[158] brotizolam,[216] rotigotine, rivaroxaban,[] dorzolamide, and olanzapine.[35, 47] The remaining molecules and their predictions are reported in the Supporting Information (Table **??**). S-oxidation model predictions are used to shade atoms, ranging from red (1.0, likely) to white (0.0, unlikely). Experimentally-observed sites of S-oxidation (atoms) and epoxidation (bonds) are highlighted by white circles. The numbers are the numerical scores of the S-oxidation model, with bioactivated molecules scored in red, and not-bioactivated molecules scored in blue. The model's AUC accuracy on the thiophene evaluation set is $88\%$ (statistically outperforming the structural alert approach, two-sided $p$-value = 0.009).
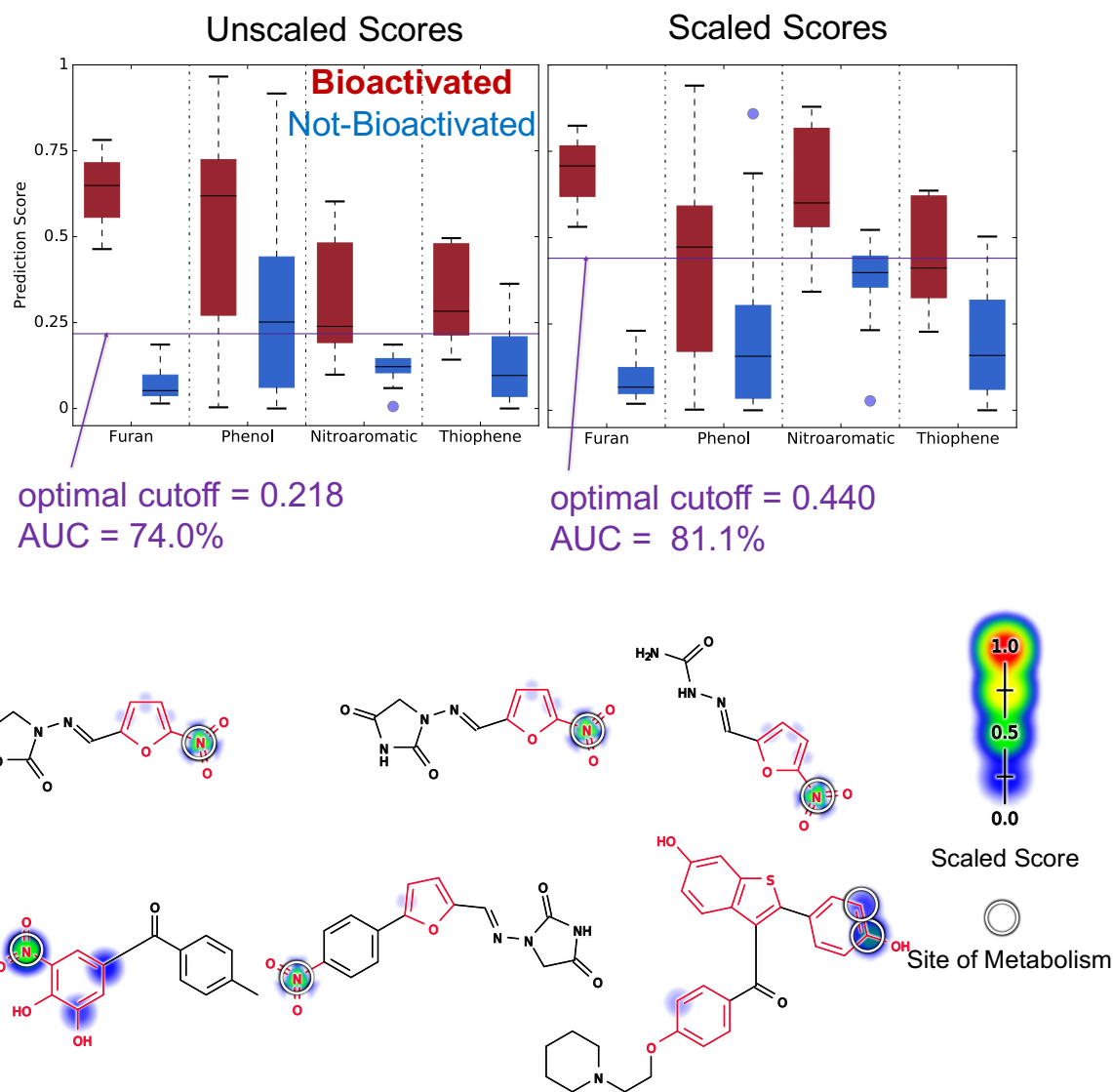
**Figure 6.5.1:** The integrated model identifies which structural alerts in the combined evaluation set are bioactivated. Cross-validated scaled predictions within structural alerts are indicated by the colored shading on atoms and bonds. Experimentally-observed sites of metabolism are indicated by white circles. Unscaled score global AUC 74.0%, scaled score global AUC 81.1%. The two accuracies are not statistically different (two-sided $p$-value = 0.199). Six molecules in the combined evaluation set that have more than one type of structural alerts (highlighted in red) and are bioactivated: from left to right, top to bottom are nitrofurantoin, furazolidone, nitrofural, tolcapone, dantrolene, and raloxifene. The combined model, both scaled and unscaled, correctly identifies the bioactivated structural alert of these compounds 100% of the time.

<div style="text-align: right; font-size: 4em; color: #8B0000;">**7**</div>

# Modeling a Rare Bioactivation Pathway

This chapter is adapted from a manuscript published in *Chemical Research Toxicology*:

**Na Le Dang**, Tyler B. Hughes, Varun Krishnamurthy, and S. Joshua Swamidass,

*Computationally Assessing the Bioactivation of Drugs by N-Dealkylation*, **2018**, *31*, 2, 68-80

## 7.1 Summary

Cytochromes P450 (CYPs) oxidize alkylated amines commonly found in drugs and other biologically active molecules, cleaving them into an amine and an aldehyde. Metabolic studies usually neglect to report or investigate aldehydes, even though they can be toxic. It is assumed that they are efficiently detoxified into carboxylic acids and alcohols. Nevertheless, some

aldehydes are reactive and escape detoxification pathways to cause adverse events by forming DNA and protein adducts. Herein, we modeled N-dealkylations that produce both amine and aldehyde metabolites and then predicted the reactivity of the aldehyde. This model used a deep learning approach previously developed by our group to predict other types of drug metabolism. In this study, we trained the model to predict N-dealkylation by human liver microsomes (HLM), finding that including isozyme-specific metabolism data alongside HLM data significantly improved results. The final HLM model accurately predicted the site of N-dealkylation within metabolized substrates (97% top-two and 94% area under the ROC curve). Next, we combined the metabolism, metabolite structure prediction, and previously published reactivity models into a bioactivation model. This combined model predicted the structure of the most likely reactive metabolite of a small validation set of drug-like molecules known to be bioactivated by N-dealkylation. Applying this model to approved and withdrawn medicines, we found that aldehyde metabolites produced from N-dealkylation may explain the hepatotoxicity of several drugs: indinavir, piperacillin, verapamil, and ziprasidone.

Our results suggest that N-dealkylation may be an under-appreciated bioactivation pathway, especially in clinical contexts where aldehyde detoxification pathways are inhibited. Moreover, this is the first report of a bioactivation model constructed by combining a metabolism and reactivity model. These results raise hope that more comprehensive models of bioactivation are possible. The model developed in this study is available at http://swami.wustl.edu/xenosite/.

## 7.2   Introduction

Alkylated amines are often important determinants of the bioactivity of organic molecules. The alkylated amine groups in biological molecules, like acetylcholine and epinephrine, often are required for their intended pharmacological functionality. As the result, alkylated amines are frequently used in medicinal chemistry.

When metabolized by cytochromes P450 (CYPs), alkylated amines usually undergo N-C bond cleavage (N-dealkylation) and give rise to an amine and an aldehyde.

N-dealkylation impacts clearance as well as pharmacodynamic properties (Figure 7.2.1).[87] Typically, N-dealkylation inactivates drugs and facilitates their elimination. For example, morphine undergoes N-demethylation to form the more readily excretable normorphine.[94] Alternatively, N-dealkylations may activate pro-drugs, like the N-demethylation of fluoxetine (Prozac) to the more potent metabolite norfluoxetine.[99, 253]

Frequently, metabolic studies focus on the amine metabolite, without reporting the aldehyde metabolite. Aldehydes are presumed to be detoxified to carboxylic acids or alcohols, which are suitable for conjugation and excretion. These detoxification pathways are very efficient for common aldehyde metabolites like formaldehyde, so researchers frequently ignore aldehydes and their effects on biological systems.

However, a subset of aldehydes can cause adverse events when they escape detoxification pathways. While most aldehydes form unstable adducts with protein, glutathione or DNA, some can form stable adducts. We use the term "reactive" to refer to aldehydes that form stable adducts with biological molecules.[5, 106] Reactive metabolites are especially problematic in patients where the detoxification pathways are inhibited by other drugs.[20, 212] For example, N-dealkylation of terbinafine yields a conjugated unsaturated aldehyde (TBF-A), a reactive metabolite that forms a transiently stable glutathione adduct that appears to cause liver toxicity in patients.[90, 111] Biochemical studies also suggest that several drugs—including clozapine and chlorpromazine—inhibit detoxification pathways enough to increase the toxicity of aldehyde metabolites.[20, 181] In addition, while animal models are frequently used to study competing bioactivation and detoxification pathways, species-specific differences in these detoxification pathways make these *in vivo* models less reliable in predicting metabolic outcome in human.[240] In summary, aldehyde formation is understudied and their toxicity may be contingent on several additional factors that are only now being elucidated.[212] As the result, aldehydes' importance may be under-appreciated and, at least on some cases, they might cause idiosyncratic toxicity in patients.

Data-driven models are already used in scientific and regulatory context. In lead optimization,

very similar machine learning models are used to identify metabolic hotspots on molecules.[210, 257] Likewise, alert structures are commonly used to identify liabilities in molecule structure,[120, 121, 226] and we previously demonstrated that metabolism models can improve the specificity of alerts.[57] The hope is that modeling of metabolism and reactivity will improve substantially on purely statistical approaches to understanding toxicity, as are commonly relied upon within industry and regulatory agencies.[255] Metabolism modeling with machine learning does not produce detailed enzyme mechanics, but they do provide finer grain information that is currently used, and for this reason are a significant step forward.

Our group has published extensively on deep learning models of metabolism and reactivity, with the ultimate goal of comprehensively predicting bioactivation pathways of drug-like molecules. These studies have included models for reaction types yielding specific metabolites such as epoxides[104] and quinone species. [**?** ]

We have also modeled the reactivity of metabolites toward adduction with glutathione, DNA, proteins, and cyanide, which often are biomarkers for toxicity as well as drivers of toxicological mechanisms.[**? ?** ] DNA conjugation can lead to cancer[166, 167] and protein conjugation can lead to cellular dysfunction, cell death, and sometimes immune response cascades that extend the damage to organ and systemic levels.[5] As insight into those possibilities, our models yield probabilities for individual sites on the molecule to undergo the specific reaction, enabling prediction of metabolite structures too. Taken together, metabolism modeling can predict if a bioactivation reaction is likely to take place, and this information can guide downstream analyses.

In this study, we aim to study a specific bioactivation pathway, the formation of reactive aldehydes by N-dealkylation. Though detailed enzyme mechanisms are not considered here, the question is mechanistic in that a specific metabolic pathway is being assessed for its mechanistic role in clinically observable toxicity. Our goal is to assess the impact of the N-dealkylation pathway, specifically. In contrast, other studies have focused on fitting statistical models to toxicity data without regard to bioactivation pathways.[255] We approach this question by developing a new model of N-dealkylation, combining it with a previously reported reactivity

model, and assessing drug-like molecules in the literature for bioactivation by this pathway. To develop the metabolism model, we used a deep learning algorithm, previously developed by our group to predict other types of drug metabolism[56, 57, 103–105, 258]. In this study, we trained the model to predict N-dealkylations by human liver microsomes (HLM). We also assessed the value of including isozyme-specific metabolism data alongside the HLM data, in building metabolism models. Next, we combined the metabolism and reactivity models into a bioactivation model. This combined model predicted the structure of the most likely reactive metabolites formed by N-dealkylation of an input molecule. Molecules known to form reactive metabolites by N-dealkylation were used to assess the bioactivation model. Finally, we used this model to study the importance of N-dealkylation as a bioactivation pathway.

## 7.3    Materials and Methods

### 7.3.1    Training Data

We collected a data set from the literature-derived Accelrys Metabolite Database (AMD) as a foundation for modeling efforts. A total of 1290 N-dealkylation reactions were extracted, each of which catalyzed by human liver microsome (HLM) or one of the nine most common human P450 isozymes: CYP1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, and 3A4 (Figure 7.3.1). We implemented an automated labeling algorithm that used the structure of each N-dealkylation product to identify the bond within the reactant that breaks, as the site of N-dealkylation. Duplicate molecules were merged into a single representation, with all metabolized bonds labeled. After merging, our final data set contained 883 unique substrates with each nitrogen-carbon bond assigned ten binary values (including missing if not-known), corresponding to metabolism by HLM and the individual P450 isozymes. There were 25, 506 bond-level training labels in the dataset.

Under the AMD license agreement, we were not allowed to disclose the structures of molecules in the data set. Nevertheless, to enable reproduction of our results, we provided the AMD reaction
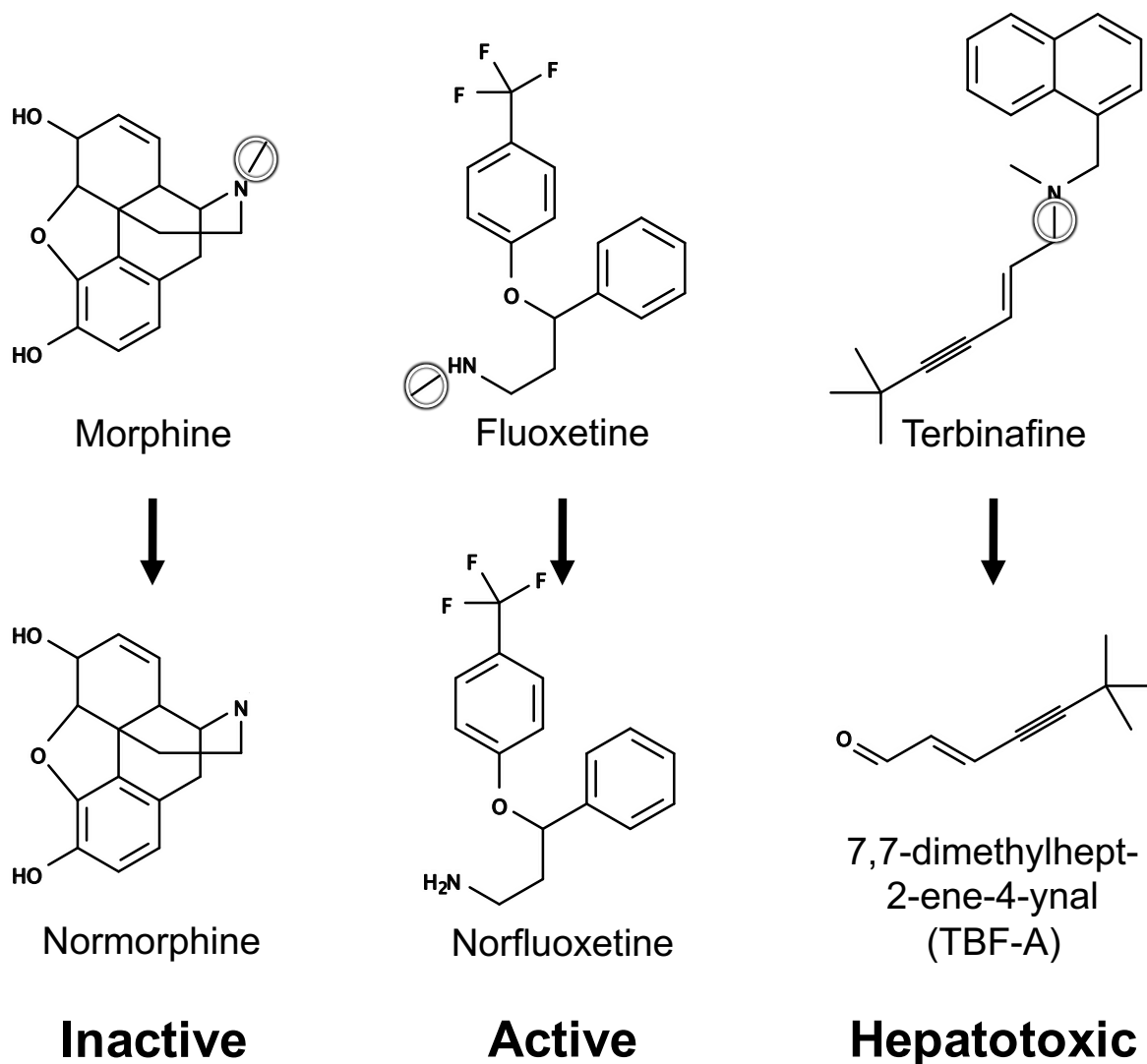
**Figure 7.2.1: N-dealkylation can alter the efficacy and safety profile of drugs.** Three drugs (with site of metabolism circled) demonstrate different outcomes of N-dealkylation. Although N-dealkylation produces two metabolites, only one observed metabolite for each reaction is shown.
(**Left**) N-dealkylation of a morphine, an analgesic drug, produces normorphine, pharmacologically inactive and more soluble metabolite that can be easily excreted in comparison to the parent compound.[223] (**Middle**) N-dealkylation of fluoxetine, an anti-depressive drug, generates norfluoxetine, the active metabolite responsible for the majority of fluoxetine's therapeutic effects.[99, 253] (**Right**) N-dealkylation of terbinafine gives rise to 7,7-dimethylhept-2-ene-4-ynal (TBF-A), a reactive metabolite that can conjugate to proteins and lead to toxicity.[111]
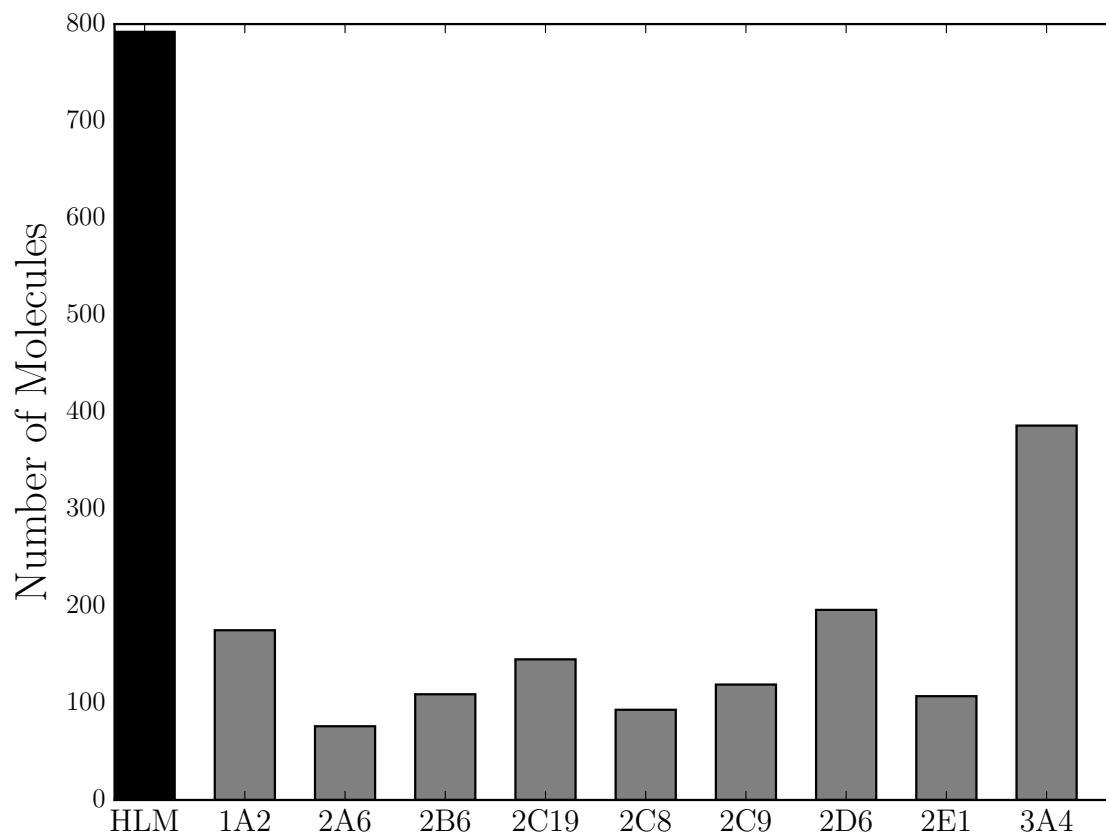
**Figure 7.3.1: Size of each data set.** Each data set contains unique N-dealkylation substrates metabolized by the corresponding Cytochromes P450 (CYP) isozyme or in human liver microsome (HLM). The size of each data set reflects the contribution of each enzymatic entity to N-dealkylation metabolism. Many of the substrates are metabolized by multiple CYP isozymes and HLM. The combined data set contains 883 molecules.

and molecule registry numbers and the metabolism status of every compound in the final data set in the Supporting Information.

## 7.3.2 External Validation Data

An external data set was used to assess the generalizability of models built on the training data. This testing set contained 108 unique HLM-mediated N-dealkylation substrates recently added to AMD (Jan-2017 version) that are not in our training dataset. We provided the AMD reaction and molecule registry numbers and the metabolism status of every compound in the test data set in the Supporting Information.

### 7.3.3 Descriptors
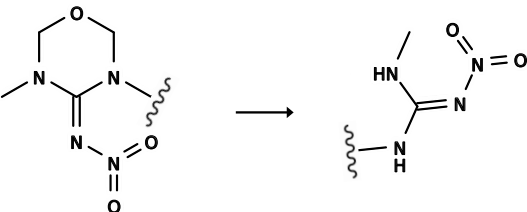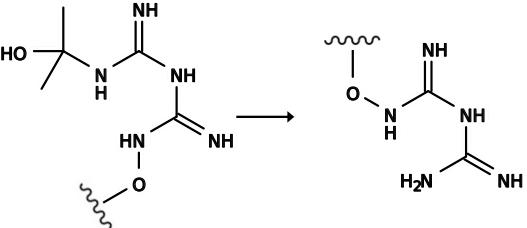
To model N-dealkylation, we encoded chemical information in numerical descriptors. Each bond was represented by a vector of 381 descriptors that described its properties at the atom, bond, and molecule levels. All of these descriptors were calculated by in-house software from Open Babel 2D SDF files.[186] Specifically, the 381 descriptors included two sets of 179 atom descriptors, 7 bond descriptors (e.g. bond type, bond length), and 16 molecule descriptors (e.g. molecular weight, topological polar surface area, or molar refractivity). Atom-level descriptors (e.g. atom identity, charge, or hybridization), representing atoms on either side of each bond, were developed for our previous models of metabolism and reactivity.[104, 105, 258] The complete list of descriptors used in this study is provided in the Supporting Information.

### 7.3.4 Heuristic Model

We constructed a simple heuristic model using overall dataset statistics to provide a baseline of performance against which to compare to more complex methods.[56]. In this heuristic model, we identified all nitrogen-carbon bonds as potential sites of N-dealkylation in a test molecule.

Based on the carbon chain attached to the nitrogen, potential site of N-dealkylation could be classified into four groups: methyl, short alkyl chains, in non-aromatic ring, and the remaining nitrogen-carbon bonds. The probability of being labeled as metabolized in the combined data set across all members of each group was calculated. This strategy yielded probabilities of 65.32%, 64.51%, 5.41%, and 16.98% for methyl, short alkyl chains, in non-aromatic ring, and the remaining nitrogen-carbon bonds, respectively (Table7.3.1). Each potential site was initially assigned with the probability of its group. All other bonds were assigned initial scores of zero. Next, these scores were linearly scaled so that they summed to one in molecules that have at least one potential site. The bonds of a molecule with no potential sites all received a score of zero. A python implementation of this model and an excel file containing the detailed statistics are included in the Supporting Information to facilitate future studies.

**Table 7.3.1:** Classes of N-C bonds in the heuristic model and the probability of being N-dealkylated by HLM.

| Group | Percent Metabolized | Example Reaction |
|---|---|---|
| Methyl[24] | 64.51% |  |
| Short alkyl[42] | 64.51% |  |
| Non-aromatic ring[81] | 5.41% |  |
| Other[217] | 16.98% |  |

### 7.3.5 Multi-Target Modular Model

For each input molecule, atom-level, bond-level, and molecule-level descriptors were calculated for all bonds between pair of heavy atoms. These descriptors were inputs to a neural network with two hidden layers that generates vectors of ten scores corresponding to one for HLM and one for each P450 isozyme. Each element in the vector represented the probability that a bond is metabolized, mediated by the corresponding isozyme or HLM. The weights of the model were calibrated during training by performing gradient descent on the cross-entropy error of the difference between the predicted and actual response values of each bond.

For assessment, predictions were computed using ten-fold cross-validation. Cross-validation is a jackknifing technique widely used in machine learning for providing an estimate of generalization accuracy. In our case, a model generated by nine tenths of the data set was used to predict the remaining tenth of the training data. The process was then repeated until all data were used in a test set exactly one time. The aggregated accuracy was tracked and compiled to assess the cross-validated accuracy as a reasonable estimate of model accuracy toward new data.

The model predicted N-dealkylation at the bond level using a multi-target, modular network [11, 14, 15, 18, 145] with one input layer, three hidden layers and one output layer (Figure 7.3.2). This architecture had several advantages over a standard neural network. First, it mirrors the local structure of some of the input descriptors. The contribution of each descriptor to the final model depended on the distance from the atom or bond it described.[104, 105, 258] Accordingly, the modular network combined descriptors by their distances to the atom of interest into neighborhood groups. Second, the multitask architecture solves related problems simultaneously. In this case, P450 share the same catalytic mechanism while differing in substrate specificity. This information from the isozyme specific datasets could then be used to improve the accuracy of HLM predictions in our model design. Finally, the multitask model architecture has nearly the same number of parameters as the single-task HLM model, but makes use of much more data.
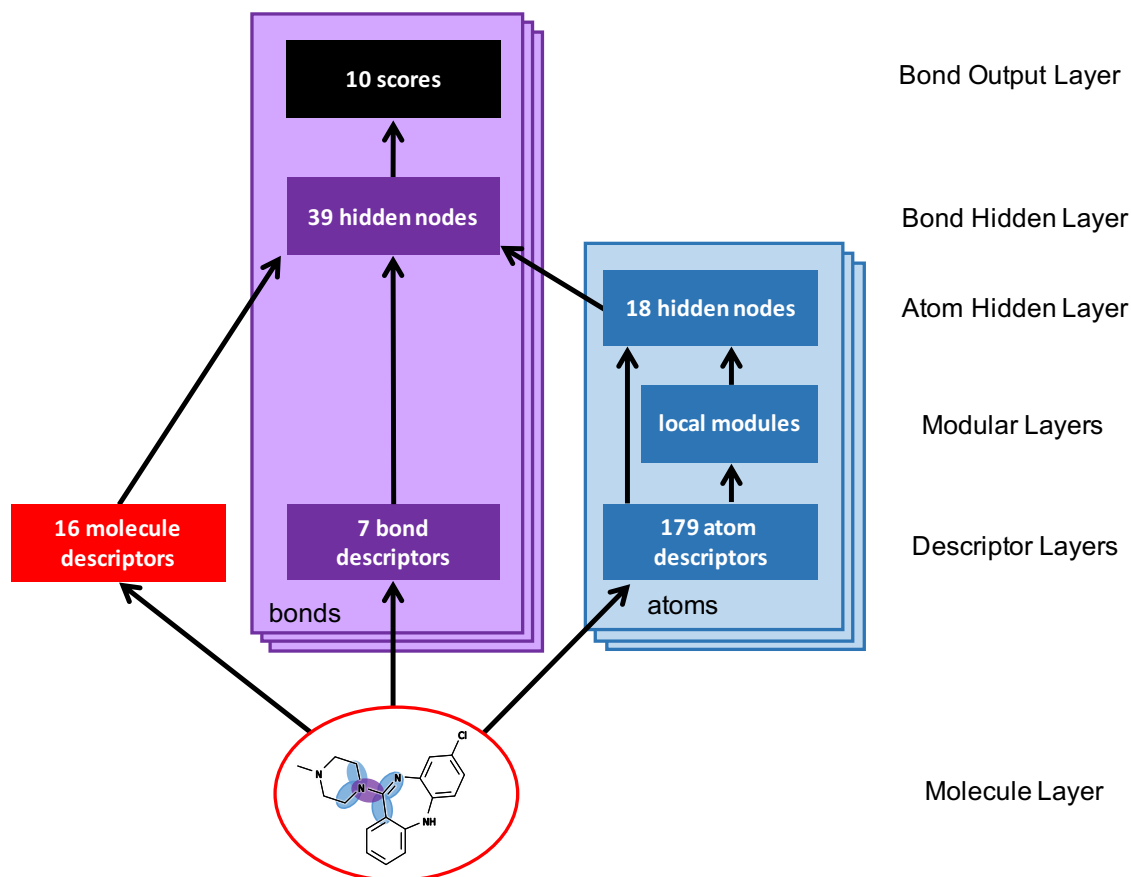
**Figure 7.3.2: The structure of the N-dealkylation model.** This diagram shows how information flows through the model, which is composed of one input layer, three hidden layers and one output layer. This model computes predictions for each nitrogen-carbon bond in a test molecule. From the 2D structure of an input molecule, 16 molecule-level, 7 bond-level, and two sets of 179 atom-level descriptors are calculated. All nodes in the second hidden layer feed to the output layer of ten nodes (each corresponds to a cytochrome P450 isozyme or human liver microsome). In the diagram, bars represent vectors of numbers. Layers are colored red, purple and blue corresponding to, respectively, molecules, bonds, and atoms. The model has $6,375$ weights.

### 7.3.6 Single Target Models

Multi-targeted models were more effective than single target models for predicting molecule reactivity.[106] We tested if this was true also for N-dealkylation by training ten single-target neural network models, each on a P450 isozyme or HML data set to provide a baseline of performance against which to compare to the multi-target modular deep neural network. A matrix of descriptor encoded bonds between heavy atoms in a data set was presented to a neural network with 20 hidden nodes. During training, the model learned a mapping between the descriptor values of each bond and the binary experimental response of that bond, metabolized or not metabolized by HLM or each P450 isozyme. The weight calibration and cross-validation procedures were identical to those of the multi-target modular deep neural network.

### 7.3.7 Reactivity Model

Previously, we developed a model to predict reactivity to biological molecules.[106] This model was a deep neural network that predicted, given a molecule's structure, its probabilities of conjugating to protein, DNA, glutathione (GSH), and cyanide. The model was trained on 1364 electrophilic and 1439 nonreactive molecules. On the original training data, the reactivity model predicted reactive atoms within reactive molecules with average site AUC of 96.6%, 89.8%, 92.8%, and 94.4% for cyanide, DNA, GSH, and protein, respectively. It is unreported how accurately the reactivity model predicts the reactivity of aldehydes in particular, and the cross-validated predictions were used to assess the performance on subset of molecules with aldehydes.

### 7.3.8 Metabolite Structures

The metabolite algorithm constructed a pair of amine and aldehyde/ketone products for every possible nitrogen-carbon bond cleavage. However, N-demethylations were excluded because the formaldehydes produced by these reactions are rapidly detoxified.[65] A python implementation of this N-dealkylation metabolite structure generator is included in the Supporting Information to
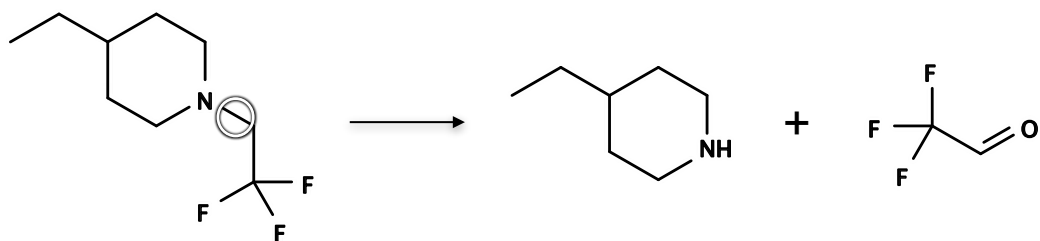
**Figure 7.3.3: Metabolites formed by N-dealkylation.** Carbon-nitrogen bond breakage during N-dealkylation creates two metabolites, one from each side of the bond. The nitrogen-side metabolite ($R_2NH$) has a hydrogen in place of the leaving alkyl group. The carbon-side metabolite is an aldehyde/ketone (RCOH/RCOR). Trifluoroacetaldehyde (TFALD) precursor's site of N-dealkylation is circled.[66]

facilitate future studies.

When a molecule is N-dealkylated at a nitrogen-carbon bond, this reaction generates two metabolites, i.e. a lower order amine from the nitrogen-side and aldehyde/ketone from the carbon-side.[87] Usually only the nitrogen side of the product is reported (Figure 7.3.3). Even when reported, carbon-side products are only described as aldehydes/ketones $48.5\%$ of the time in the AMD. In some cases, the reported reactions instead included downstream metabolites of the aldehydes, i.e. alcohols (26.1%) and acids (25.4%), generated by subsequent detoxification reactions by aldehyde reductases and aldehyde dehydrogenases.[65, 151, 187] For this reason, the metabolite algorithm always inferred the formation of aldehydes (or ketones) and does not model subsequent transformations.

### 7.3.9 N-Dealkylation Bioactivation Model

We built a bioactivation model that linked metabolism, metabolite structure prediction, and reactivity models.

This combined model predicted the most likely reactive metabolites generated by N-dealkylation from precursor molecules. For each potential metabolite, the combined model defined its output as the product of two predictions, the probability of formation times the probability of being reactive. The first prediction was the probability of formation as computed

by HLM N-dealkylation model. The second prediction was the maximum of the probabilities of reactivity toward protein and DNA as computed by the reactivity model. The metabolite with the highest score (the product of both predictions) was predicted to be the most likely reactive metabolite generated by N-dealkylation.

## 7.4 Results and Discussion

The following sections study the performance of the models, and then use them to assess N-dealkylation as a bioactivation pathway. First, we evaluated the N-dealkylation model in its accuracy of identifying metabolized sites within a metabolized molecule. Second, we analyzed the model predictions at the global level by calculating classification performance on nitrogen-carbon bonds in the entire data set. Third, we evaluated the accuracy of the N-dealklyation model on an external testing data set. Fourth, we assessed the reactivity model predictions on aldehyde containing molecules. Fifth, we tested the bioactivation model's prediction of reactive metabolites from molecules known to produce reactive aldehydes. Finally, we present three case studies of drugs known to produce reactive aldehydes by N-dealkylation.

### 7.4.1 Metabolism Model Performance

A key objective was to identify the specific bond within a substrate that undergoes N-dealkylation: its site of N-dealkylation. For every bond in a test molecule, the N-dealkylation model generated ten scores, each ranging from zero to one and corresponding to the probability that the bond was metabolized by HLM or a certain P450 isozyme. The HLM model predicts most sites correctly (Figure 7.4.1).

An informative model should assign metabolized sites with higher scores than non-metabolized sites within each molecule. We evaluated the model's success at this objective using two metrics. First, we calculated "top-two" performance, where a substrate was considered to be correctly predicted if any of its experimentally-observed sites of metabolism were predicted in the top two rank-positions out of all potential sites in the substrate. The top-two metric is commonly used to

**Table 7.4.1: Scores and dataset sizes for reactive and non-reactive molecules.** For the HLM model, a reasonable cutoff might be a score of 0.1 to 0.15. However, our model uses the score itself as a probability in downstream analysis.

|         | Mean Score | | Number of Compounds | |
|---------|----------|--------------|----------|--------------|
|         | Reactive | Non-Reactive | Reactive | Non-Reactive |
| Protein | 0.23     | 0.06         | 20       | 222          |
| GSH     | 0.32     | 0.03         | 48       | 1082         |
| DNA     | 0.31     | 0.03         | 10       | 195          |
| Cyanide | 0.16     | 0.02         | 1        | 128          |

evaluate CYP site of metabolism models.[210, 221] Second, we calculated "average nitrogen-carbon (N-C) AUC. This metric is computed by measuring the area under the ROC curves(AUC),[228] for all nitrogen-carbon bonds in the dataset.[104, 105]

The model accurately predicted metabolized sites within metabolized molecules. HLM model had cross-validated top-two and average N-C AUC accuracies of 96.6% and 93.7%, respectively. Across all ten targets, the model had cross-validated top-two and average N-C AUC accuracies of 97.2% and 95.3%, respectively (Tables 7.4.2 and 7.4.1). As an additional test for overfitting, we constructed a negative control where the metabolism targets were permuted randomly. The model was trained and tested, in cross-validation, on the permuted data. In this negative control, performance was poor, with an average AUC across isozymes of 49.4%, worse than a random classifier. The poor performance on this negative control further confirms the model is not overfitting and the high performances observed are not artifacts.

As another control designed to test the value of multitask training, we trained ten single-target models, each on a CYP isozyme or HML data set, for comparison. The multitask model performed slightly better than the single target models, but this improvement was not statistically significant ($P = 0.351$ and $0.425$ for top-two and average N-C AUC respectively). Both multitask and single target models significantly outperformed the heuristic model, which had top-two and average N-C AUC accuracies of 80.8% and 81.4%, respectively.

An effective model should distinguish metabolized sites globally in the dataset across all

**Table 7.4.2: The model accurately identifies sites of N-dealkylation.** The table contains ten-fold cross-validated top-two, average N-C AUC, and global N-C AUC performance of the multitask (MT), single target (ST) and heuristic (HR) models. Accuracies of the MT model on the training data set (Self) are also included for reference. For each metric, the highest cross-validated performance is bolded. Any scores not statistically different from the best performance are italicized. In all cases, the neural networks are significantly better than the heuristic model. The performance difference between the HLM single target and the multitask models by top-two and average N-C AUC is not significant ($P = 0.352$ and $0.425$, respectively, by Mann-Wittney U test). The performance difference by global N-C AUC, is statistically significant ($P = 0.021$ by paired permutation test).

| | Top-two | | | | Average N-C AUC | | | | Global N-C AUC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Isozyme | Self | MT | ST | HR | Self | MT | ST | HR | Self | MT | ST | HR |
| HLM | 98.9 | 96.6 | 95.7 | 80.8 | 96.1 | 93.7 | 92.9 | 81.4 | 97.5 | 95.6 | 95.4 | 87.1 |
| CYP1A2 | 98.9 | 96.6 | 94.9 | 86.5 | 97.0 | 95.0 | 91.5 | 87.2 | 95.1 | 90.1 | 90.5 | 83.4 |
| CYP2A6 | 100 | 98.7 | 98.7 | 88.4 | 96.2 | 94.8 | 86.4 | 80.9 | 96.3 | 88.2 | 87.7 | 84.4 |
| CYP2B6 | 99.1 | 99.1 | 96.3 | 90.2 | 99.7 | 98.9 | 94.4 | 89.9 | 97.3 | 92.4 | 90.8 | 87.1 |
| CYP2C19 | 100 | 98.6 | 95.9 | 86.3 | 98.7 | 97.3 | 96.4 | 91.4 | 97.0 | 92.8 | 92.3 | 84.8 |
| CYP2C8 | 100 | 97.8 | 96.8 | 85.4 | 98.3 | 96.4 | 93.9 | 88.2 | 95.4 | 89.3 | 88.6 | 82.9 |
| CYP2C9 | 99.2 | 97.5 | 95.8 | 88.7 | 98.4 | 97.4 | 96.5 | 93.4 | 96.3 | 90.6 | 89.9 | 85.8 |
| CYP2D6 | 100 | 99.0 | 98.5 | 89.6 | 98.2 | 97.0 | 97.0 | 91.9 | 95.7 | 91.4 | 90.7 | 84.4 |
| CYP2E1 | 98.1 | 97.2 | 96.3 | 84.8 | 95.8 | 93.9 | 89.6 | 81.2 | 96.7 | 91.6 | 89.4 | 83.9 |
| CYP3A4 | 98.2 | 96.1 | 96.6 | 80.1 | 98.2 | 95.8 | 95.5 | 86.8 | 95.9 | 90.9 | 90.2 | 82.8 |

nitrogen-carbon bonds. To quantify this performance, we computed the area under the ROC curve globally across all metabolized and non-metabolized nitrogen-carbon bonds (Table 7.4.2).[56, 194]

The model accurately predicted sites of N-dealkylation by this metric. Across all ten targets, the model had an average cross-validated global N-C AUC of 92.6% (Table 7.4.2). Notably, the multitask model predicted HLM-mediated sites with global N-C AUC of 95.6%, significantly outperforming the single target model ($P = 0.021$ by paired permutation test[228]). Both multitask and single target models significantly outperformed the heuristic model, which had an average global nitrogen-carbon AUC accuracy of 85.2%.

The output of the model is a well-scaled probability. When we binned N-C bonds by HLM prediction score, the proportion of sites of HLM-mediated N-dealkylation in each bin closely
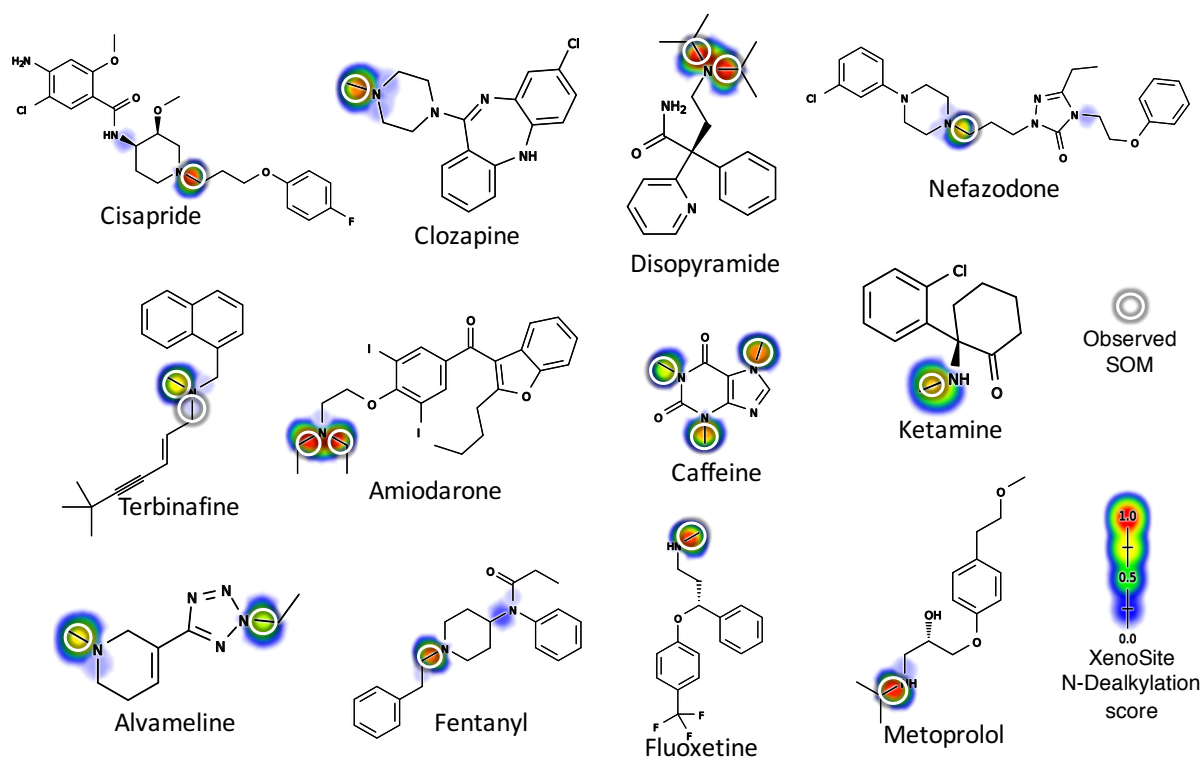
**Figure 7.4.1: The HLM accurately predicts metabolism of most molecules.** Cross-validated human liver microsome (HLM) N-dealkylation scores on example drugs from our training data are shown. Experimentally observed HLM sites of N-dealkylation are circled. The model gives the highest score to the correct sites in all these cases.
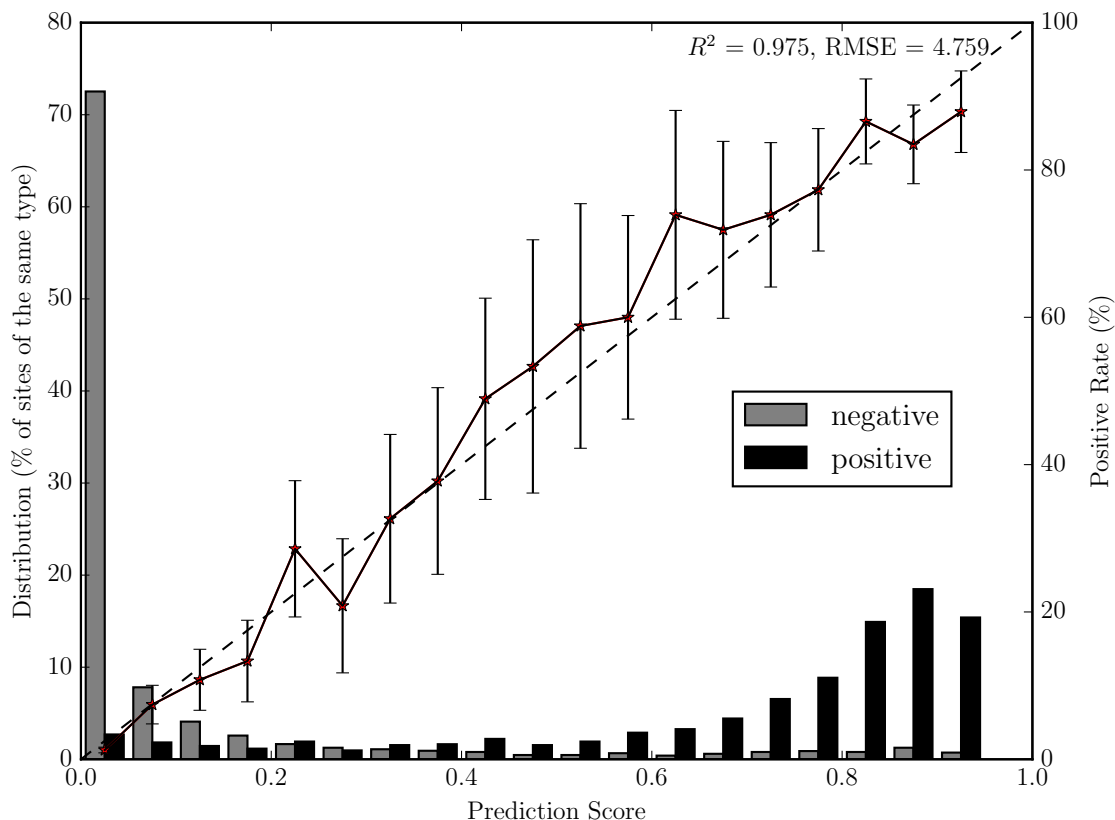
**Figure 7.4.2: The model makes well-scaled predictions, corresponding to probabilities.** The bar graphs plot the distributions of scores across 4071 dealkylated and non-dealkylated nitrogen-carbon bonds. The solid lines plot the percentage of nitrogen-carbon bonds that are dealkylated by human liver microsome (HLM) (using non-normalized frequencies) in each bin. The diagonal dashed lines indicate a hypothetical perfectly scaled prediction. HLM N-dealkylation score has a strong correlation to a perfectly scaled prediction ($R^2$ value of 0.975 and RMSE of 4.759%). This means that the score is interpretable as the probability that a nitrogen-carbon bond is metabolized in HLM. Reliability diagrams for all cytochrome P450 isozymes in the multi-target model are in the Supporting Information.

correlates with the bin's score (Figure 7.4.2, $R^2 = 0.975$). Thus the model's output is interpretable as a probability and can be combined with other probabilistic outputs.

## 7.4.2 External Validation of N-Dealkylation

The multitask N-dealkylation model performed well on the external testing data set of 108 HLM-metabolized molecules. Prediction accuracies of our model on this external testing data set were 95.5%, 94.1% and 94.8%, Top-two, Average N-C AUC, and Global N-C AUC, respectively. Overall, performance on the testing data set was comparable to the cross-validated performance

on the training data set.

### 7.4.3 Reactivity Predictions on Aldehydes

The accuracy of the reactivity model on aldehydes, specifically, was not assessed. A effective reactivity model for aldehydes is important for this study, and assessed here.

We find that the reactivity model accurately predicted aldehydes reactivity. We identified 746 aldehyde containing molecules and 1128 aldehyde sites from the reactivity model's training data set and used their cross-validated predictions to access the model performance on this subset of data. First, we assessed the ability of model to identify the correct atom in the molecule as reactive. The average site AUC is computed by averaging the AUC of sites computed within each molecule separately. The reactivity model predicted reactive atoms of 746 aldehyde containing molecules with average site AUC accuracies of 97.0%, 91.7%, 95.7%, and 94.9% for cyanide, DNA, GSH, and protein, respectively. Next, we assessed the ability of the model to separate reactive and non-reactive aldehyde molecules. Across the full database 1128 aldehydes, the model can accurately separate reactive and non-reactive substructures with AUCs of 93.0%, 80.3%, 89.1%, and 71.0% for cyanide, DNA, GSH, and protein, respectively.' The model?s performance was similarly strong on the 313 alpha-beta unsaturated aldehyde in this same dataset (Supplementary Materials). These assessments demonstrate that the reactivity model can accurately model the reactivity of aldehyde containing compounds.

### 7.4.4 A Census of Reactive Aldehyde Metabolites

The aldehyde products from the carbon-side of N-dealkylation reactions are frequently omitted in the literature. The training data set contained 1290 N-dealkylation producing aldehyde metabolites, yet only 26 of these reactions were reported with aldehyde metabolites. Of these 26 aldehydes, the literature reports 16 and 21 conjugate to GSH and DNA, respectively. Moreover, several of the aldehydes unreported from these reactions appear to be reactive (Figure 7.4.3). This omission underscores the large gaps in the literature, and suggests the potential importance of

aldehydes in toxicity pathways too.

Next, we studied the aldehyde metabolites of the reported N-dealkylation reactions of approved and withdrawn drugs. The complete list of these drugs, their aldehyde metabolites, and reactivity scores are provided in the Supporting Information. Several hepatotoxic drugs with unknown mechanisms of hepatotoxicity[1]—like indinavir, piperacillin, verapamil, and ziprasidone—are predicted by the reactivity model to have aldehyde metabolites more reactive than the parent drug and all the rest of their reported metabolites in the AMD (Figure 7.4.4). Although piperacillin toxicity has been recently associated with beta-lactam ring opening and subsequent protein conjugation, metabolic studies on this antibiotic did not assessed the reactivity of its acetyl aldehyde metabolite produced via N-demethylation, a major in vitro and in vivo metabolic pathway of piperacillin.[77] This omission adds to the evidence that N-dealkylation may be an under-appreciated bioactivation pathway. Reactive aldehydes could contribute to hepatotoxicity of these drugs. In future work, we plan to study these pathways, and confirm them in appropriate *in vitro* studies.

When a N-dealkylation reaction is reported in the literature without reporting the aldehyde metabolite, it is unclear if the aldehyde is unreported because (1) it was observed but not deemed important enough to report, (2) it was not observed because experiments were not tuned to detect them, or (3) it was not observed because it is a reactive metabolite that conjugates to protein. This final case is most concerning because these aldehydes could contribute to the toxicity of a large number of drugs. N-dealkylation appears to be a neglected bioactivation pathway that merits systematic study, especially in cases with evidence of idiosyncratic adverse events typically associated with reactive metabolites.

7.4.5  Modeling Bioactivation Into Reactive Aldehydes

We tested the bioactivation model on molecules that are known to produce reactive metabolites by N-dealkylation. The metabolism model's training data set contains 112, 52 and 3 molecules with
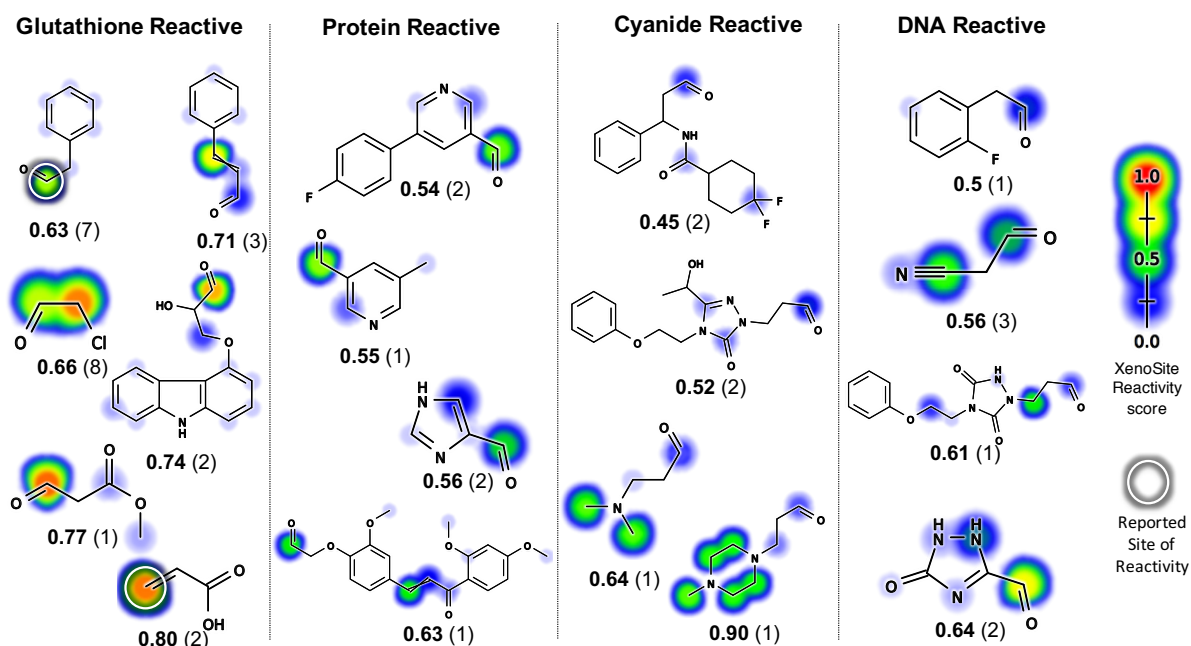
**Figure 7.4.3: Reactive aldehyde metabolites.** We computed the structures of all aldehyde metabolites generated from observed N-dealkylation reactions. Next, we applied the reactivity model to these predicted metabolites to assess their reactivity.[?] Examples of the aldehyde metabolites and their reactivity predictions are shown. The molecular reactivity scores (MRS) are bolded. The number of molecules known to produce this metabolite are shown in parentheses. Experimentally determined sites of reactivity are circled.
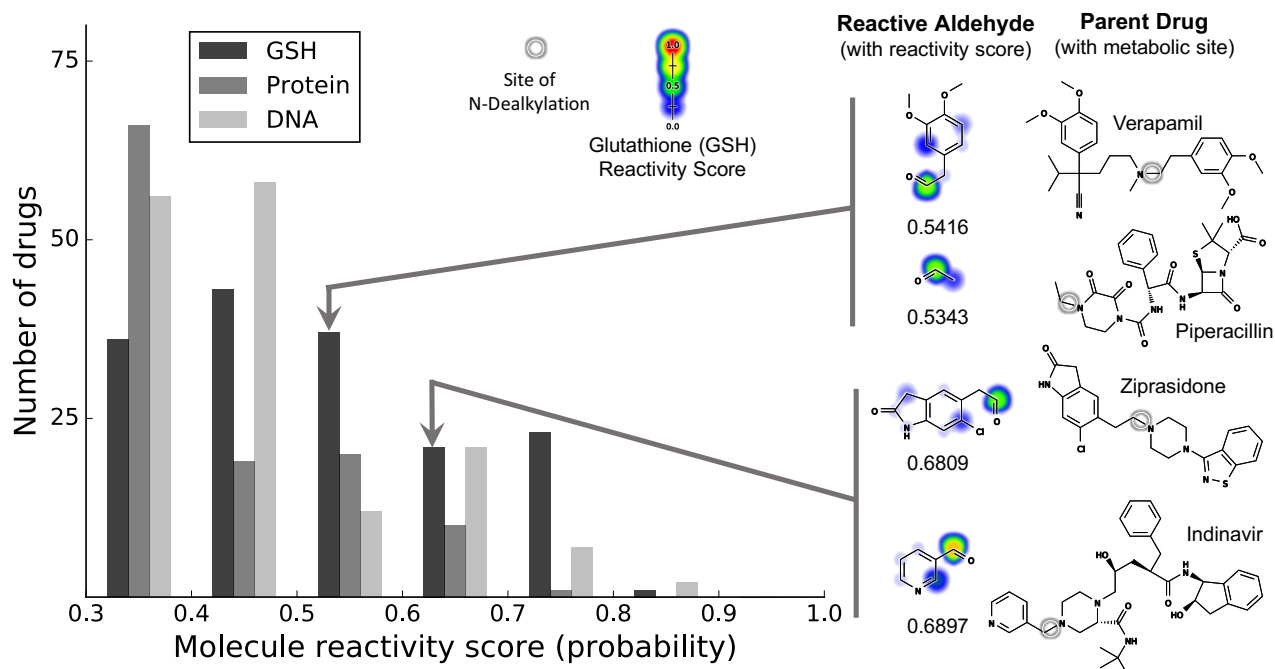
**Figure 7.4.4: Reactive aldehyde metabolites formed by N-dealkylation reactions.** Of 1925 approved and withdrawn drugs, the literature reports N-dealkylation reactions for 380. The aldehydes formed by these N-dealkylation reactions were inferred, and their reactivity was assessed with a previously published reactivity model.[?] (**Left**) The distribution of reactivity scores above 0.3 for glutathione (GSH), protein and DNA molecule reactivity. For reference, the reactivity score of N-acetyl-p-benzoquinone imine, a well known electrophile responsible for acetaminophen's toxicity, are 0.75, 0.54, and 0.38 for GSH, protein, and DNA, respectively. (**Right**) Several drugs are (1) known to be hepatotoxic (2) by unknown mechanisms, also (3) appear to form reactive aldehydes. These aldehyde metabolites were predicted to be more reactive than the parent drug as well as their other observed metabolites, so these aldehydes might be the mechanism of toxicity.

148
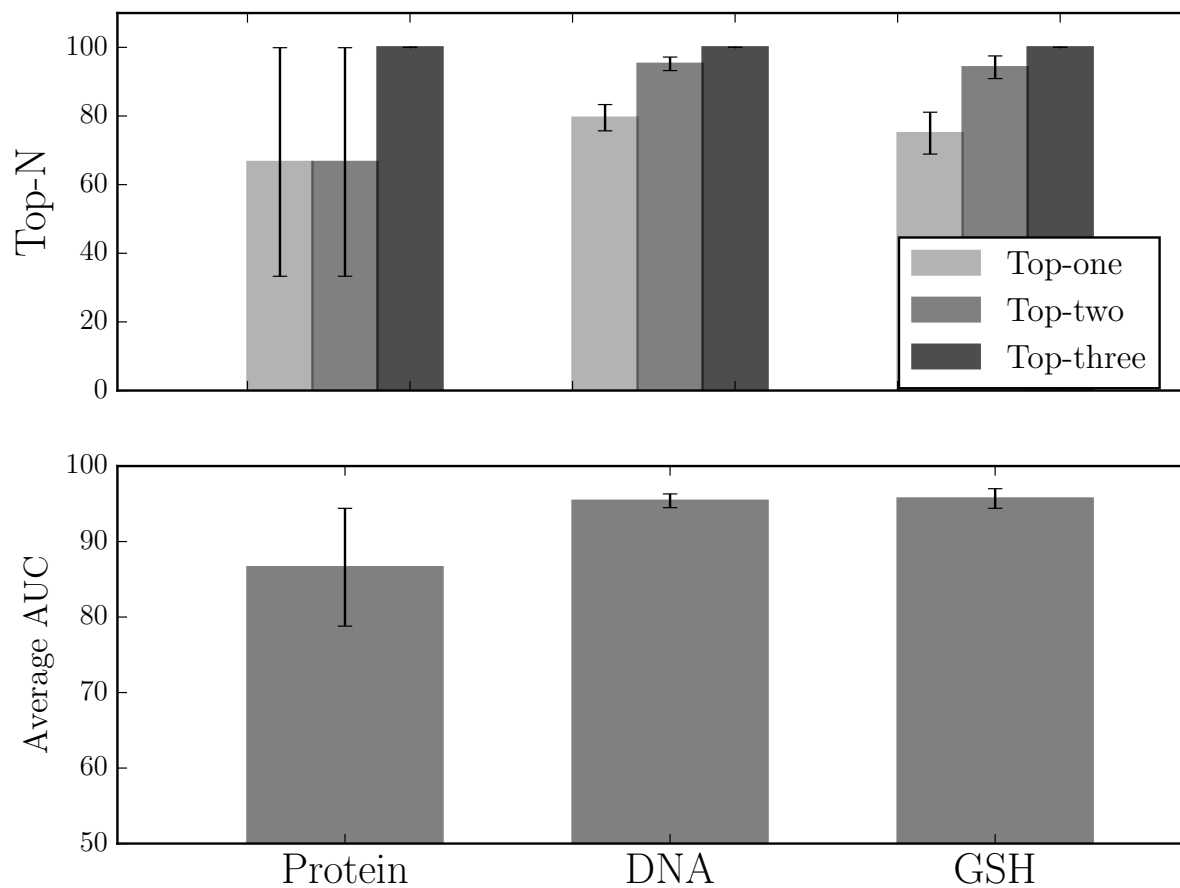
**Figure 7.4.5: The bioactivation model accurately identifies reactive metabolites produced by N-dealkylation.** The upper panel shows top-N performance metrics for 112, 52, 3 molecules that produce through N-dealkylation metabolites reported to conjugate to DNA, GSH, and protein, respectively. A molecule was considered correctly predicted if any of its metabolites observed to be reactive were predicted in the first-, second-, or third-rank position. Bottom panel, the AUC for predictions of all metabolites produced via N-dealkylation from each molecule is computed and then averaged across the whole data set, measuring the per-molecule performance.
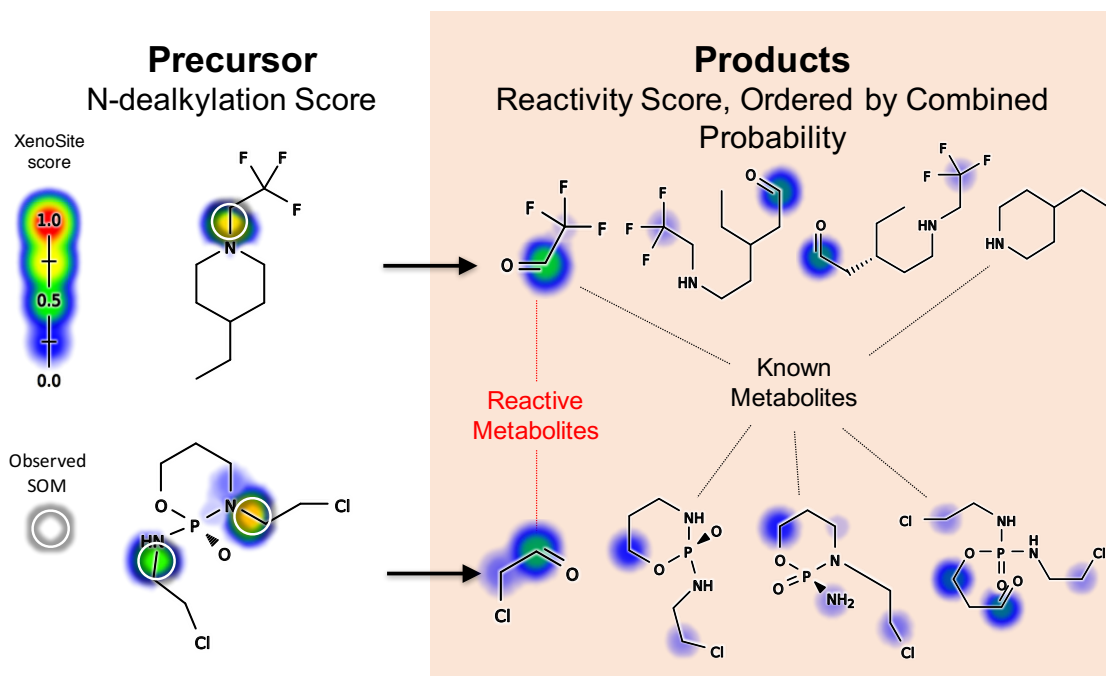
**Figure 7.4.6: Combination of metabolism, metabolite structure prediction and reactivity models to predict reactive metabolites.** The combined model predicts the most likely reactive metabolites produced by N-dealkylation from precursor molecules. Example molecules are trifluoroacetaldehyde precursor[66] and ifosfamide.[40, 43] Terbinafine is discussed in the next section and figure.

N-dealkylation metabolites reported to be reactive with DNA, GSH, and protein, respectively. No training is done in this experiment, so the molecules here are just used for validation. In this task, the aldehyde metabolites that are not formed or are not reactive are intrinsic negative controls. For each molecule, the goal is to predict the known reactive first, above all possible N-dealkylation metabolites.

The bioactivation model can predict reactive metabolite generated through N-dealkylation with high accuracy (Figure 7.4.5). The bioactivation model can rank the metabolites observed to be reactive with DNA, GSH, and protein at the first position $79.4\%$, $75.0\%$, and $66.6\%$ of the time. Averaging the AUC of the metabolites associated with each molecule, the bioactivation model ranking predict observed reactive metabolites with accuracies of $95.4\%$, $95.7\%$, and $86.6\%$ for DNA, GSH, and protein, respectively. The Supporting Information includes the metabolite structures and their bioactivation model scores, with the reactive aldehyde labeled.

The three molecules known to have toxicity associated with reactive metabolites generated by

N-dealkylation: a trifluoroacetaldehyde (TFALD) precursor,[66] isfosamide,[40, 43] and terbinafine (TBF)[111] illustrate the model's predictions (Figures 7.4.6 and 7.4.7). The bioactivation model correctly identifies the toxic metabolite in all three cases (Figures 7.4.6 and 7.4.7). In the first case, the TFALD precursor, TFALD is identified with high probability, and is known to cause cytotoxicity[140, 141] The second case, ifosamide is an anticancer prodrug that is extensively metabolized by P450.[40] Ifosamide's hepato- and nephrotoxicity-profile has been attributed to chloroacetaldehyde—a metabolite formed through N-dealkylation of ifosamide.[43] For both cases of TFALD precursor[66] and ifosfamide, [40, 43] the combined model correctly ranks known reactive metabolites (TFALD and chloroacetaldehyde) as the most likely reactive metabolites produced by N-dealkylation from the parent molecules (Figure 7.4.6).The anti-fungal TBF, the third case, is more complicated. The correct metabolite is identified (Figure 7.4.7), however it is assigned a low probability, just 6%. As we will see, this is because the reactive metabolite may be formed by double rather than single dealkylation(s) from TBF.

These results are encouraging. No training is need to sequentially apply the metabolism and reactivity models. So without fitting an parameters, it appears the bioactivation model can correctly identify reactive aldehydes in drug-like molecules. A better analysis might include more molecules, but aldehyde toxicity is not studied enough to be sure any aldehydes safe.

### 7.4.6    Bioactivation of Terbinafine

TBF is a widely used anti-fungal agent. The most common adverse side effects associated with oral terbinafine treatment are mild to severe gastrointestinal and cutaneous reactions.[92] However, TBF causes transient liver injury in at least 1% of patients, which may progress to death or fulminant liver failure.[4, 89] Consequently, oral TBF prescriptions require regular monitoring of liver enzymes.[89, 90] Hepatotoxicity is thought to be caused by 7,7-dimethylhept-2-ene-4-ynal (TBF-A, Figure 7.4.7). This reactive metabolite was only demonstrated 15 years after marketing TBF worldwide, and was not identified by standard reactive metabolite screens.[111]
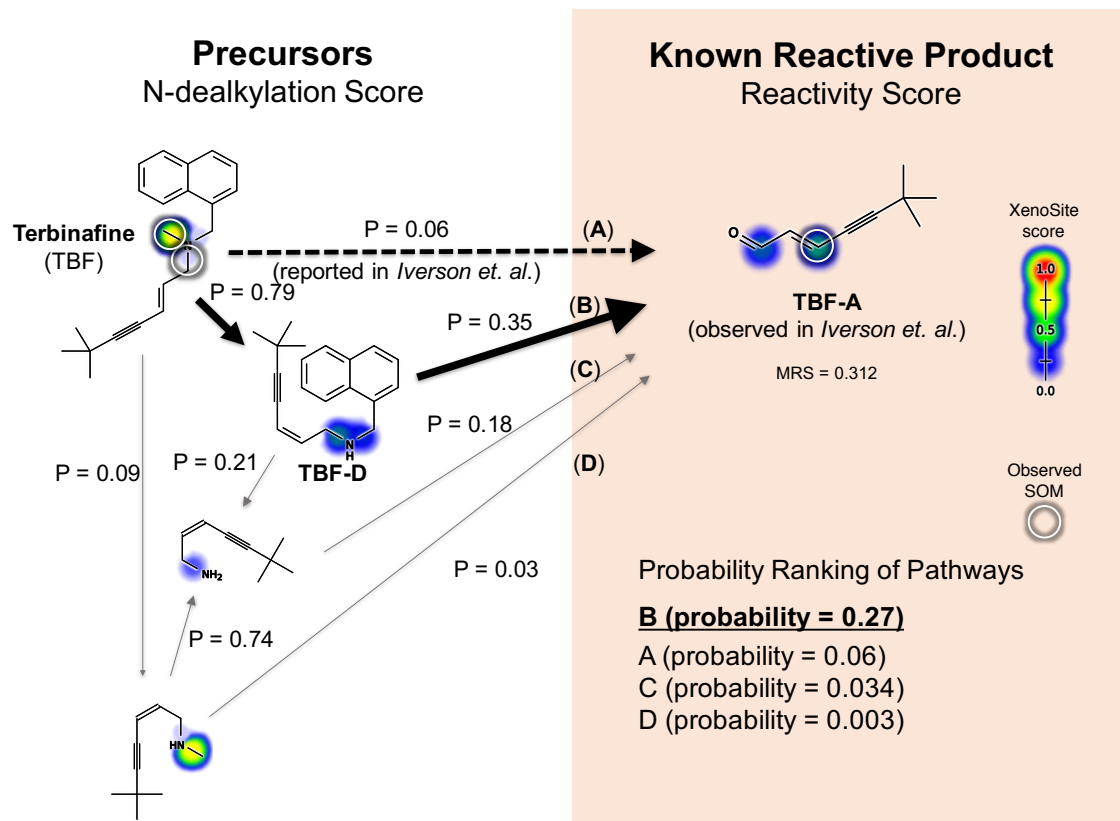
**Figure 7.4.7: Terbinafine is predicted to form its reactive metabolite by two sequential dealkylations.** Multiple metabolites generated by N-dealkylation of terbinafine have been observed.[111, 241] The reactive metabolite 7,7-dimethylhept-2-ene-4-ynal (TBF-A, depicted) was identified as a key mediator of hepatotoxicity.[111] It is proposed that TBF-A is formed directly from terbinafine single dealkylation (dashed arrows, $P = 0.06$), but several alternate pathways are possible. The model computes the probability of each pathway, finding that TBF-A is most likely formed by two sequential dealkylations (path B, $P = 0.27$) through the intermediate desmethyl terbinafine (TBF-D). Supporting this prediction, TBF-D is reported as a metabolite of TBF by some studies.[241] This suggests a revision of the single dealkylation pathway reported in the literature (path A, $P = 0.06$).

Although TBF-A was eventually identified as a reactive metabolite, the exact mechanism of its formation is unclear.[18, 111, 241] TBF-A has been proposed as a metabolite of TBF,[111], but there are several pathways possible, and the precise pathway was not identified. TBF possesses three N-alkyl groups that are susceptible to oxidative cleavage. These metabolic reactions may generate reactive TBF-A in a single dealkylation (path A, Figure 7.4.7) or multiple dealkylations (paths B, C, and D, Figure 7.4.7). The model identifies the most probable pathway (path B, Figure 7.4.7). Though beyond the scope of this computational study, preliminary experimental data suggest the double- dealkylation pathway is more important.

A sequential dealkylation pathway to forming TBF-A might explain why it was not detected in the glutathione trapping experiments commonly used to screen for reactive metabolites. Likewise, TBF-A may be more reactive with proteins than with GSH, as many aldehydes are, and therefore not be reliably trapped by GSH. These are all reasons TBF-A's role as a reactive metabolite might have been missed for so long and highlight the promise of using computational modeling alongside experimental screens. We envision a joint approach. Computational models might identify reactive metabolites missed in the experimental screens. Further study would then verify or rule out these metabolites, so the risk factors of a molecule can be more reliably understood.

At the same time, the TBF case study demonstrated the limitations of the current bioactivation model, which only considers single-transformation bioactivation even though multiple-transformation bioactivation is important too. Our immediate goals are to expand the range of metabolism that can be modeled, and then use these models to predicting multiple generations of metabolites. Combining reactivity models with this comprehensive metabolism model might enable bioactivation modeling for most drugs.

## 7.5   Limitations and Future Directions

The most prominent limitation of the current approach is that it does not consider competing metabolic pathways and multiple metabolic transformations. For example, a molecule can undergo either hydroxylation and then dehydrogenation at one site or N- dealkylation and then

reduction at another site. We are currently developing comprehensive models to predict competing Phase I and II metabolic pathways. Our next goal is being able to predict the complete metabolic network for a test molecule. Ultimately, we aim to build comprehensive metabolic and reactivity models. Perhaps a first step might extend this approach to O-dealkylation , another reaction that produces aldehydes. Alternatively, we could include epoxidation or quinone formation, both of which often produce reactive molecules and are predicted by models previously published by our group.

Just as importantly, this model does not account for all important factors. Dosage, co-administered medicines, and patient factors (age, co-morbidities, genetic variations) also play important roles in adverse drug reactions (ADR). For example, limiting daily dose to under 20 mg/day can significantly decrease ADR.[119] Similarly, understanding the factors governing which specific proteins and amino acid are bound by reactive metabolites may be important to understanding toxicity.[108] While taking all of these factors in to account will be a daunting task, we believe that models solving smaller problem like ours are steady steps toward this goal.

Finally, as with any machine learning approaches, there is no guarantee that the models' applicability domain will extend beyond their training domain of metabolically studied small molecules. Proprietary chemical domains such as natural products and peptides-based drugs that are actively explored by pharmaceutical companies may not be well suited to the current models trained on literature-derived data. However, this approach could be easily applied to new data to expand its domain of applicability.

## 7.6   Conclusion

This study demonstrated a deep network model for predicting N-dealkylation leading to the formation of reactive aldehydes. The model predicted sites of N-dealkylation for HLM substrates with top-two, average N-C AUC, and global N-C AUC accuracies of 96.6%, 93.7%, and 95.6%, respectively. We also predicted reactive and potentially toxic metabolites by combining metabolism and reactivity models into a bioactivation model. One limitation to the current

approach is that it does not automatically consider multiple metabolic steps. More comprehensive modeling of bioactivation pathways are on the horizon, and actively being developed by our group. Even so, we identified multistep pathways for terbinafine. Likewise, we identified several drugs (indinavir, piperacillin, verapamil, and ziprasidone) that are hepatotoxic by unknown mechanisms of hepatotoxicity, and produce reactive aldehyde metabolites by N-dealkylation. This suggests that N-dealkylation may be an under-appreciated bioactivation pathway, and suggest they should be more carefully reported and assessed in metabolic studies. These results also encourage continued work towards building comprehensive models of bioactivation, to consider the full range of metabolic transformations.

## 7.7   Abbreviations

ADR, Adverse Drug Reaction AMD, Accelrys Metabolite Database; AUC, Area Under the Receiver Operating Characteristic Curve; CYP, Cytochromes P450; HLM, Human Liver Microsomes; N-C, nitrogen-carbon; N-dealkylation, nitrogen dealkylation; TBF, terbinafine; TBF-A, 7,7-dimethylhept-2-ene-4-ynal; TBF-D, desmethyl terbinafine TFALD, Trifluoroacetaldehyde

# 8

# Conclusion and Future Directions

Predicting idiosyncratic adverse drug reactions (IADRs), toxicity due to bioactivated reactive metabolites, is a very challenging task. Multiple, competing metabolism pathways contribute to the complexity of drug metabolism, disposition, and toxicity. In this work, we developed models which were able to learn reactions comprising a substantial portion of these complex, interconnected metabolism networks. While each model presented in this thesis can accurately predict the corresponding metabolic components, we are still exploring different approaches for building a comprehensive metabolism model. Furthermore, the models developed in this thesis work do not account for all the important factors contributing to IADR risk. Dosage, co-administered medicines, and patient factors (age, co-morbidities, genetic variations) also play

important roles in IADR. For example, limiting daily dose to under 20 mg/day can significantly decrease IADR [119]. In addition, the relationship between IADRs and specific hapten complexes of reactive metabolite and proteins is important to predicting toxicity [108].

In order to address the aspects of metabolism necessary to predict IADRs, several addition models must be developed. First, several key metabolic reactions including tautomerization, isomerization, and rearrangement are not covered by our current modeling efforts. Second, we must develop methods for integrating models of individual reactions to predict the complete metabolic network for a test molecule. Third, quantifying exposure to reactive metabolites requires accurate modeling of kinetics, which are not addressed in our current work. Fourth, to predict IADR risk for individual patients, we must model patient-specific differences in metabolism cause by environmental and genetic factors.

While taking all of complex factors in to account to predict IADR is a daunting task, we believe that models solving smaller problems like those in this thesis are steady steps toward this goal. The thesis work is a part of our laboratory's ongoing effort to develop XenoSite, an online collection of metabolism and reactivity mathematical models that systematically summarize the data from thousands of papers into a condensed and useable computational tool that predicts adduction and potential toxicity of molecules. If successful, these modeling approaches will help make new medicines and current medical practice safer and more effective.

# References

[1] Clinical and research information on drug-induced liver injury. *https://livertox.nlm.nih.gov (accessed Dec 2017)*.

[2] Faers reporting by patient outcomes by year, 11 2015. URL `http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/ Surveillance/AdverseDrugEffects/ucm070461.htm`.

[3] Martın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org*, 2015.

[4] SM Abdel-Rahman and MC Nahata. Oral terbinafine: a new antifungal agent. *Ann. Pharmacother.*, 31(4):445–456, 1997.

[5] David H Adams, Cynthia Ju, Shashi K Ramaiah, Jack Uetrecht, and Hartmut Jaeschke. Mechanisms of immune-mediated liver injury. *Toxicol. Sci.*, 115(2):307–321, 2010.

[6] BH Ali. Pharmacology and toxicity of furazolidone in man and animals: some recent research. *Gen. Pharmacol.*, 20(5):557–563, 1989.

[7] Ignacio Aliagas, Alberto Gobbi, Timothy Heffron, Man-Ling Lee, Daniel F Ortwine, Mark Zak, and S Cyrus Khojasteh. A probabilistic method to report predictions from a human liver microsomes stability qsar model: a practical tool for drug discovery. *J. Comput.-Aided Mol. Des.*, 29(4):327–338, 2015.

[8] Rubén Alvarez-Sánchez, David Basketter, Camilla Pease, and Jean-Pierre Lepoittevin. Studies of chemical selectivity of hapten, reactivity, and skin sensitization potency. 3. synthesis and studies on the reactivity toward model nucleophiles of the 13c-labeled skin sensitizers, 5-chloro-2-methylisothiazol-3-one (mci) and 2-methylisothiazol-3-one (mi). *Chem. Res. Toxicol.*, 16(5):627–636, 2003.

[9] Vinicius M Alves, Eugene N Muratov, Stephen J Capuzzi, Regina Politi, Yen Low, Rodolpho C Braga, Alexey V Zakharov, Alexander Sedykh, Elena Mokshyna, Sherif

Farag, Carolina H. Andrade, Victor E. Kuz'min, Denis Fourchesh, and Alexander Tropsha. Alarms about structural alerts. *Green Chem.*, 18(16):4348–4360, 2016.

[10] Manuel Ammerschlaeger, Jürgen Beigel, Kai-Uwe Klein, and Stefan O Mueller. Characterization of the species-specificity of peroxisome proliferators in rat and human hepatocytes. *Toxicol. Sci.*, 78(2):229–240, 2004.

[11] Rangachari Anand, Kishan Mehrotra, Chilukuri K Mohan, and Sanjay Ranka. Efficient classification for multiclass problems using modular neural networks. *IEEE Trans. Neural Networks*, 6(1):117–124, 1995.

[12] Gail D Anderson. Children versus adults: Pharmacokinetic and adverse-effect differences. *Epilepsia*, 43(s3):53–59, 2002.

[13] Sabry M Attia. Deleterious effects of reactive metabolites. *Oxid. Med. Cell. Longevity*, 3 (4):238–253, 2010.

[14] Gasser Auda and Mohamed Kamel. Modular neural network classifiers: A comparative study. *Journal of Intelligent and Robotic Systems*, 21(2):117–129, 1998.

[15] Farooq Azam. *Biologically inspired modular neural networks*. PhD thesis, Citeseer, 2000.

[16] Jonathan B Baell and Georgina A Holloway. New substructure filters for removal of pan assay interference compounds (pains) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.*, 53(7):2719–2740, 2010.

[17] J Bakker, FJ Gommers, I Nieuwenhuis, and H Wynberg. Photoactivation of the nematicidal compound alpha-terthienyl from roots of marigolds (tagetes species). a possible singlet oxygen role. *J. Biol. Chem.*, 254(6):1841–1844, 1979.

[18] Konstantin V Balakin, Sean Ekins, Andrey Bugrim, Yan A Ivanenkov, Dmitry Korolev, Yuri V Nikolsky, Andrey A Ivashchenko, Nikolay P Savchuk, and Tatiana Nikolskaya. Quantitative structure-metabolism relationship modeling of metabolic n-dealkylation reaction rates. *Drug Metab. Dispos.*, 32(10):1111–1120, 2004.

[19] Andrea Baldacci and Wolfgang Thormann. Analysis of oxycodol and noroxycodol stereoisomers in biological samples by capillary electrophoresis. *Electrophoresis*, 26(10): 1969–1977, 2005.

[20] John T Barr and Jeffrey P Jones. Evidence for substrate-dependent inhibition profiles for human liver aldehyde oxidase. *Drug Metabolism and Disposition*, 41(1):24–29, 2013.

[21] Marcus Baumann and Ian R Baxendale. An overview of the synthetic routes to the best selling drugs containing 6-membered heterocycles. *Beilstein journal of organic chemistry*, 9:2265, 2013.

[22] PPMDLMCJPJC Beaune, PM Dansette, D Mansuy, L Kiffel, M Finck, C Amar, JP Leroux, and JC Homberg. Human anti-endoplasmic reticulum autoantibodies appearing in a drug-induced hepatitis are directed against a human liver cytochrome p-450 that hydroxylates the drug. *Proc. Natl. Acad. Sci. U. S. A.*, 84(2):551–555, 1987.

[23] JA Bell, FA Dallas, WN Jenner, and LE Martin. The metabolism of ranitidine in animals and man. *Biochem. Soc. Trans.*, 8(1):93–93, 1980.

[24] Gwénaëlle Bellec, Yvonne Dréano, Patrick Lozach, Jean François Ménez, and François Berthou. Cytochrome p450 metabolic dealkylation of nine n-nitrosodialkylamines by human liver microsomes. *Carcinogenesis*, 17(9):2029–2034, 1996.

[25] John S Bertram. The molecular biology of cancer. *Mol. Aspects Med.*, 21(6):167–223, 2000.

[26] Jos GM Bessems and Nico PE Vermeulen. Paracetamol (acetaminophen)-induced toxicity: molecular and biochemical mechanisms, analogues and protective approaches. *Crit. Rev. Toxicol.*, 31(1):55–138, 2001.

[27] Urs A Boelsterli, Han K Ho, Shufeng Zhou, and Koon Yeow Leow. Bioactivation and hepatotoxicity of nitroaromatic drugs. *Curr. Drug Metab.*, 7(7):715–727, 2006.

[28] David Borland and Russell M Taylor Ii. Rainbow color map (still) considered harmful. *IEEE computer graphics and applications*, 27(2), 2007.

[29] Bernard B Brodie, Watson D Reid, Arthur K Cho, Glenn Sipes, Gopal Krishna, and James R Gillette. Possible mechanism of liver necrosis caused by aromatic organic compounds. *Proc. Natl. Acad. Sci. U. S. A.*, 68(1):160–164, 1971.

[30] Peter J Bugelski. Genetic aspects of immune-mediated adverse drug effects. *Nat. Rev. Drug Discovery*, 4(1):59–69, 2005.

[31] Roxana Bujack, Terece L Turton, Francesca Samsel, Colin Ware, David H Rogers, and James Ahrens. The good, the bad, and the ugly: A theoretical framework for the assessment of continuous colormaps. *IEEE Transactions on Visualization and Computer Graphics*, 2017.

[32] Philip C Burcham, Frank R Fontaine, Lisa M Kaminskas, Dennis R Petersen, and Simon M Pyke. Protein adduct-trapping by hydrazinophthalazine drugs: mechanisms of cytoprotection against acrolein-mediated toxicity. *Mol. Pharmacol.*, 65(3):655–664, 2004.

[33] John Caldwell. Problems and opportunities in toxicity testing arising from species differences in xenobiotic metabolism. *Toxicol. Lett.*, 64:651–659, 1992.

[34] Rita Calé, Irene Aragão, Helena Martins, Germano Cardoso, Luisa M Ferreira, Paula S Branco, M Lourdes Bastos, and Paula Guedes de Pinho. Propofol and metabolites monitoring in serum of patients with induced sedation. *Toxicol. Lett.*, 189:S113–S114, 2009.

[35] John T Callaghan, Richard F Bergstrom, Louis R Ptak, and Charles M Beasley. Olanzapine. *Clin. Pharmacokinet.*, 37(3):177–193, 1999.

[36] Leslie E Carlini, Neal J Meropol, John Bever, Michael L Andria, Todd Hill, Philip Gold, Andre Rogatko, Hao Wang, and Rebecca L Blanchard. Ugt1a7 and ugt1a9 polymorphisms predict response and toxicity in colorectal cancer patients treated with capecitabine/irinotecan. *Clin. Cancer Res.*, 11(3):1226–1236, 2005.

[37] Chun Yip Chan, Lee Sun New, Han Kiat Ho, and Eric Chun Yong Chan. Reversible time-dependent inhibition of cytochrome p450 enzymes by duloxetine and inertness of its thiophene ring towards bioactivation. *Toxicol. Lett.*, 206(3):314–324, 2011.

[38] GFQ Chan, GH Neil Towers, and JC Mitchell. Ultraviolet-mediated antibiotic activity of thiophene compounds of tagetes. *Phytochemistry*, 14(10):2295–2296, 1975.

[39] Open-Source Cheminformatics and Machine Learning. http://www.rdkit.org/.

[40] Chong-Sheng Chen, Youssef Jounaidi, and David J Waxman. Enantioselective metabolism and cytotoxicity of r-ifosfamide and s-ifosfamide by tumor cell-expressed cytochromes p450. *Drug Metab. Dispos.*, 33(9):1261–1267, 2005.

[41] Min Chen, Bing Xia, Bixiao Chen, Qiusha Guo, Jin Li, Mei Ye, and Zhengguo Hu. N-acetyltransferase 2 slow acetylator genotype associated with adverse effects of sulfasalazine in the treatment of inflammatory bowel disease. *Canadian Journal of Gastroenterology and Hepatology*, 21(3):155–158, 2007.

[42] Taehyeon M Cho, Randy L Rose, and Ernest Hodgson. The effect of chlorpyrifos-oxon and other xenobiotics on the human cytochrome p450-dependent metabolism of naphthalene and deet. *Drug Metab. Dispos.*, 22(4):235, 2007.

[43] Leila Choucha-Snouber, Caroline Aninat, Laurent Grsicom, Geoffrey Madalinski, Céline Brochot, Paul Emile Poleni, Florence Razan, Christiane Guguen Guillouzo, Cécile Legallais, Anne Corlu, and Leclerc Eric. Investigation of ifosfamide nephrotoxicity induced in a liver–kidney co-culture biochip. *Biotechnol. Bioeng.*, 110(2):597–608, 2013.

[44] 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. *Biochem. Pharmacol.*, 38(9):1451–1458, 1989.

[45] Lisa J Christopher, Donghui Cui, Wenying Li, Anthony Barros, Vinod K Arora, Haiying Zhang, Lifei Wang, Donglu Zhang, James A Manning, Kan He, et al. Biotransformation of [14c] dasatinib: in vitro studies in rat, monkey, and human and disposition after administration to rats and monkeys. *Drug Metab. Dispos.*, 36(7):1341–1356, 2008.

[46] Frederick C Churchill, Leslie C Patchen, Carlos C Campbell, Ira K Schwartz, Phuc Nguyen-Dinh, and Charlotte M Dickinson. Amodiaquine as a prodrug: importance of metabolite (s) in the antimalarial effect of amodiaquine in humans. *Life Sci.*, 36(1):53–62, 1985.

[47] Leslie Citrome, Virginia L Stauffer, Lei Chen, Bruce J Kinon, Darcie L Kurtz, Jennie G Jacobson, and Richard F Bergstrom. Olanzapine plasma concentrations after treatment

with 10, 20, and 40 mg/d in patients with schizophrenia: an analysis of correlations with efficacy, weight gain, and prolactin concentration. *J. Clin. Psychopharmacol.*, 29(3): 278–283, 2009.

[48] Alf Claesson and Ola Spjuth. On mechanisms of reactive metabolite formation from drugs. *Mini-Rev. Med. Chem.*, 13(5):720–729, 2013.

[49] Dana M Clausen, Jianxia Guo, Robert A Parise, Jan H Beumer, Merrill J Egorin, John S Lazo, Edward V Prochownik, and Julie L Eiseman. In vitro cytotoxicity and in vivo efficacy, pharmacokinetics, and metabolism of 10074-g5, a novel small-molecule inhibitor of c-myc/max dimerization. *J. Pharmacol. Exp. Ther.*, 335(3):715–727, 2010.

[50] Samuel M Cohen, Shoji Fukushima, Nigel J Gooderham, F Peter Guengerich, Stephen S Hecht, Ivonne MCM Rietjens, Robert L Smith, Maria Bastaki, Christie L Harman, Margaret M McGowen, Luis G Valerio, and Sean V. Taylor. Safety evaluation of substituted thiophenes used as flavoring ingredients. *Food Chem. Toxicol.*, 99:40–59, 2017.

[51] Roger Collier. Drug development cost estimates hard to swallow, 2009.

[52] Jean Pierre Cravedi, Elisabeth Perdu-Durand, Maryse Baradat, Jacques Alary, Laurent Debrauwer, and Georges Bories. Chloramphenicol oxamylethanolamine as an end product of chloramphenicol metabolism in rat and humans: evidence for the formation of a phospholipid adduct. *Chemical research in toxicology*, 8(5):642–648, 1995.

[53] George E Dahl, Navdeep Jaitly, and Ruslan Salakhutdinov. Multi-task neural networks for qsar predictions. *arXiv preprint arXiv:1406.1231*, 2014.

[54] Deepak K Dalvie, Amit S Kalgutkar, S Cyrus Khojasteh-Bakht, R Scott Obach, and John P O'Donnell. Biotransformation reactions of five-membered aromatic heterocyclic rings. *Chem. Res. Toxicol.*, 15(3):269–299, 2002.

[55] Ann K Daly. Using genome-wide association studies to identify genes important in serious adverse drug reactions. *Annu. Rev. Pharmacol. Toxicol.*, 52:21–35, 2012.

[56] Na Le Dang, Tyler B Hughes, Varun Krishnamurthy, and S Joshua Swamidass. A simple model predicts ugt-mediated metabolism. *Bioinformatics*, 32(20):3183–3189, 2016.

[57] Na Le Dang, Tyler B. Hughes, Grover P. Miller, and S. Joshua Swamidass. Computational approach to structural alerts: Furans, phenols, nitroaromatics, and thiophenes. *Chem. Res. Toxicol.*, 2017. doi: 10.1021/acs.chemrestox.6b00336. URL http://dx.doi.org/10.1021/acs.chemrestox.6b00336.

[58] Na Le Dang, Tyler B. Hughes, Matthew K. Matlock, and S. Joshua Swamidass. The metabolic rainbow: Deep learning phase 1 metabolism in five colors. *ACS Cent. Sci.*, 2018.

[59] Na Le Dang, Tyler B Hughes, Grover Paul Miller, and S Joshua Swamidass. Computationally assessing the bioactivation of drugs by n-dealkylation. *Chem. Res. Toxicol.*, 31(2):68–80, Feb 19 2018.

[60] PB Danielson. The cytochrome p450 superfamily: biochemistry, evolution and drug metabolism in humans. *Current drug metabolism*, 3(6):561–597, 2002.

[61] Patrick M Dansette, Gildas Bertho, and Daniel Mansuy. First evidence that cytochrome p450 may catalyze both s-oxidation and epoxidation of thiophene derivatives. *Biochem. Biophys. Res. Commun.*, 338(1):450–455, 2005.

[62] PM Dansette, Cao Thang Do, H. El Amri, and D Mansuy. Evidence for thiophene-s-oxide as a primary reactive metabolite of thiophene in vivo: formation of a dihydrothiophene sulfoxide mercapturic acid. *Biochem. Biophys. Res. Commun.*, 186(3):1624–1630, 1992.

[63] Inc. DAYLIGHT Chemical Information Systems. Smarts - a language for describing molecular patterns. http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.

[64] JD DeBethizy and Johnnie R Hayes. Metabolism: a determinant of toxicity. *Princ. Methods Toxicol. (6th Ed.)*, pages 59–100, 1994.

[65] Sundeep S Dhareshwar and Valentino J Stella. Your prodrug releases formaldehyde: should you be concerned? no! *J. Pharm. Sci.*, 97(10):4184–4193, 2008.

[66] Martin E Dowty, George Hu, Fengmei Hua, F Barclay Shilliday, and Heather V Dowty. Drug design structural alert formation of trifluoroacetaldehyde through n-dealkylation is linked to testicular lesions in rat. *Int. J. Toxicol.*, 30(5):546–550, 2011.

[67] Anne Dreiem and Frode Fonnum. Thiophene is toxic to cerebellar granule cells in culture after bioactivation by rat liver enzymes. *Neurotoxicology*, 25(6):959–966, 2004.

[68] Fuying Du, Qian Ruan, Mingshe Zhu, and Jie Xing. Detection and characterization of ticlopidine conjugates in rat bile using high-resolution mass spectrometry: applications of various data acquisition and processing tools. *J. Mass Spectrom.*, 48(3):413–422, 2013.

[69] Claire M Ellison, Richard Sherhod, Mark TD Cronin, Steven J Enoch, Judith C Madden, and Philip N Judson. Assessment of methods to define the applicability domain of structural alert models. *J. Chem. Inf. Model.*, 51(5):975–985, 2011.

[70] JCL Erve, SC Vashishtha, O Ojewoye, A Adedoyin, R Espina, W DeMaio, and RE Talaat. Metabolism of prazosin in rat and characterization of metabolites in plasma, urine, faeces, brain and bile using liquid chromatography/mass spectrometry (lc/ms). *Xenobiotica*, 38(5): 540–558, 2008.

[71] John CL Erve. Chemical toxicology: reactive intermediates and their role in pharmacology and toxicology. 2006.

[72] David C Evans, Alan P Watt, Deborah A Nicoll-Griffith, and Thomas A Baillie. Drug-protein adducts: an industry perspective on minimizing the potential for drug bioactivation in drug discovery and development. *Chem. Res. Toxicol.*, 17(1):3–16, 2004.

[73] Nagy A Farid, Atsushi Kurihara, and Steven A Wrighton. Metabolism and disposition of the thienopyridine antiplatelet drugs ticlopidine, clopidogrel, and prasugrel in humans. *J. Clin. Pharmacol.*, 50(2):126–142, 2010.

[74] Henry M Jr Feder, Carl Osier, and Eufronio G Maderazo. Chloramphenicol: a review of its use in clinical practice. *Reviews of infectious diseases*, 3(3):479–491, 1981.

[75] Adrian J Fretland and Curtis J Omiecinski. Epoxide hydrolases: biochemistry and molecular biology. *Chem. Biol. Interact.*, 129(1):41–59, 2000.

[76] Horatio B Fung and Thien-Ly Doan. Tinidazole: a nitroimidazole antiprotozoal agent. *Clin. Ther.*, 27(12):1859–1884, 2005.

[77] Giulia Ghibellini, Arlene S Bridges, Claudia N Generaux, and Kim LR Brouwer. In vitro and in vivo determination of piperacillin metabolism in humans. *Drug Metab. Dispos.*, 35 (3):345–349, 2007.

[78] James R Gillette, Jerry R Mitchell, and Bernard B Brodie. Biochemical mechanisms of drug toxicity. *Annu. Rev. Pharmacol.*, 14(1):271–288, 1974.

[79] Emma E Graham, Rachel J Walsh, Charlotte M Hirst, James L Maggs, Scott Martin, Martin J Wild, Ian D Wilson, John R Harding, J Gerald Kenna, Raimund M Peter, Dominic P. Williams, and B. Kevin Park. Identification of the thiophene ring of methapyrilene as a novel bioactivation-dependent hepatic toxicophore. *J. Pharmacol. Exp. Ther.*, 326(2):657–671, 2008.

[80] Darja Gramec, Lucija Peterlin Mašič, and Marija Sollner Dolenc. Bioactivation potential of thiophene-containing drugs. *Chem. Res. Toxicol.*, 27(8):1344–1358, 2014.

[81] Trevor Green, Alison Toghill, Robert Lee, Felix Waechter, Edgar Weber, and James Noakes. Thiamethoxam induced mouse liver tumors and their relevance to humans part 1: mode of action studies in the mouse. *Toxicol. Sci.*, 86(1):36–47, 2005.

[82] Mark P Grillo. Detecting reactive drug metabolites for reducing the potential for drug toxicity. *Expert Opin. Drug Metab. Toxicol.*, 11(8):1281–1302, 2015.

[83] Brion J Gross, Richard V Branchflower, Terrence R Burke, David E Lees, and Lance R Pohl. Bone marrow toxicity in vitro of chloramphenicol and its metabolites. *Toxicology and applied pharmacology*, 64(3):557–565, 1982.

[84] Hans-Peter Gschwind, Ulrike Pfaar, Felix Waldmeier, Markus Zollinger, Claudia Sayer, Peter Zbinden, Michael Hayes, Rolf Pokorny, Michael Seiberling, Monique Ben-Am, et al. Metabolism and disposition of imatinib mesylate in healthy volunteers. *Drug Metab. Dispos.*, 33(10):1503–1512, 2005.

[85] F Peter Guengerich. Cytochromes p450, drugs, and diseases. *Mol. Interventions*, 3(4):194, 2003.

[86] F Peter Guengerich and James S MacDonald. Applying mechanisms of chemical toxicity to predict drug safety. *Chem. Res. Toxicol.*, 20(3):344–369, 2007.

[87] FP Guengerich, O Okazaki, Y Seto, and TL Macdonald. Radical cation intermediates in n-dealkylation reactions. *Xenobiotica*, 25(7):689–709, 1995.

[88] C Guillemette. Pharmacogenomics of human udp-glucuronosyltransferase enzymes. *Pharmacogenomics J.*, 3(3):136–158, 2003.

[89] Aditya K Gupta and Neil H Shear. Terbinafine: an update. *J. Am. Acad. Dermatol.*, 37(6): 979–988, 1997.

[90] AK Gupta, JQ del Rosso, CW Lynde, GH Brown, and NH Shear. Hepatitis associated with terbinafine therapy: three case reports and a review of the literature. *Clin. Exp. Dermatol.*, 23(2):64–67, 1998.

[91] Nguyêt-Thanh Ha-Duong, Sylvie Dijols, Anne-Christine Macherey, Joyce A Goldstein, Patrick M Dansette, and Daniel Mansuy. Ticlopidine as a selective mechanism-based inhibitor of human cytochrome p450 2c19. *Biochemistry*, 40(40):12112–12122, 2001.

[92] Michael Hall, Carmen Monka, Pierre Krupp, and Desmond O'Sullivan. Safety of oral terbinafine: results of a postmarketing surveillance study in 25 884 patients. *Arch. Dermatol.*, 133(10):1213–1219, 1997.

[93] Hisham K Hamadeh, Brian L Knight, Astrid C Haugen, Stella Sieber, Rupesh P Amin, Pierre R Bushel, Raymond Stoll, Kerry Blanchard, Supriya Jayadev, Raymond W Tennant, Michael L. Cunningham, Cynthia A. Afshari, and Richard S. Paules. Methapyrilene toxicity: anchorage of pathologic observations to gene expression alterations. *Toxicol. Pathol.*, 30(4):470–482, 2002.

[94] CW Hand, RA Moore, HJ McQuay, MC Allen, and JW Sear. Analysis of morphine and its major metabolites by differential radioimmunoassay. *Ann. Clin. Biochem.*, 24(2):153–160, 1987.

[95] James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1):29–36, 1982.

[96] Mike Hann, Brian Hudson, Xiao Lewell, Rob Lifely, Luke Miller, and Nigel Ramsden. Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci.*, 39(5):897–902, 1999.

[97] Shuaibing He, Manman Li, Xiaotong Ye, Hongyu Wang, Wenkang Yu, Wenjing He, Yun Wang, and Yanjiang Qiao. Site of metabolism prediction for oxidation reactions mediated by oxidoreductases based on chemical bond. *Bioinformatics*, page btw617, 2016.

[98] Alistair P Henderson, Christine Bleasdale, William Clegg, and Bernard T Golding. 2, 6-diarylaminotetrahydropyrans from reactions of glutaraldehyde with anilines: models for biomolecule cross-linking. *Chem. Res. Toxicol.*, 17(3):378–382, 2004.

[99] Christoph Hiemke and Sebastian Härtter. Pharmacokinetics of selective serotonin reuptake inhibitors. *Pharmacol. Ther.*, 85(1):11–28, 2000.

[100] Jordan L Holtzman. Role of reactive oxygen and metabolite binding in drug toxicity. *Life sciences*, 30(1):1–9, 1982.

[101] JB Hudson, EA Graham, N Miki, GHN Towers, LL Hudson, R Rossi, A Carpita, and D Neri. Photoactive antiviral and cytotoxic activities of synthetic thiophenes and their acetylenic derivatives. *Chemosphere*, 19(8):1329–1343, 1989.

[102] Jason D Hughes, Julian Blagg, David A Price, Simon Bailey, Gary A DeCrescenzo, Rajesh V Devraj, Edmund Ellsworth, Yvette M Fobian, Michael E Gibbs, Richard W Gilles, Nigel Greene, Enoch Huang, Teresa Krieger-Burke, Jens Loesel, Travis Wager, Larry Whiteley, and Yao Zhang. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorg. Med. Chem. Lett.*, 18(17):4872–4875, 2008.

[103] T. B. Hughes and S. J. Swamidass. Deep learning to predict the formation of quinone species in drug metabolism. *Chem. Res. Toxicol.*, 30(2):642–656, 2017.

[104] Tyler B Hughes, Grover P Miller, and S Joshua Swamidass. Modeling epoxidation of drug-like molecules with a deep machine learning network. *ACS Cent. Sci.*, 1(4):168–180, 2015.

[105] Tyler B Hughes, Grover P Miller, and S Joshua Swamidass. Site of reactivity models predict molecular reactivity of diverse chemicals with glutathione. *Chem. Res. Toxicol.*, 28 (4):797–809, 2015.

[106] Tyler B Hughes, Na Le Dang, Grover P Miller, and S Joshua Swamidass. Modeling reactivity to biological macromolecules with a deep multitask network. *ACS Cent. Sci.*, 2 (8):529–537, 2016.

[107] Andrew Hunter, Lee Kennedy, Jenny Henry, and Ian Ferguson. Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer methods and programs in biomedicine*, 62(1):11–19, 2000.

[108] Keisuke Ikehata, Tatyana G Duzhak, Nadezhda A Galeva, Tao Ji, Yakov M Koen, and Robert P Hanzlik. Protein targets of reactive metabolites of thiobenzamide in rat liver in vivo. *Chemical research in toxicology*, 21(7):1432–1442, 2008.

[109] Robert MJ Ings, James A McFadzean, and Walter E Ormerod. The mode of action of metronidazole in trichomonas vaginalis and other micro-organisms. *Biochem. Pharmacol.*, 23(9):1421–1429, 1974.

[110] Koichi Itakura, Atsunori Furuhata, Noriyuki Shibata, Makio Kobayashi, and Koji Uchida. Maillard reaction-like lysine modification by a lipid peroxidation product: immunochemical detection of protein-bound 2-hydroxyheptanal in vivo. *Biochem. Biophys. Res. Commun.*, 308(3):452–457, 2003.

[111] Suzanne L Iverson and Jack P Uetrecht. Identification of a reactive metabolite of terbinafine: insights into terbinafine-induced hepatotoxicity. *Chem. Res. Toxicol.*, 14(2): 175–181, 2001.

[112] S Iyengar, JT Arnason, BJR Philogene, P Morand, NH Werstiuk, and G Timmins. Toxicokinetics of the phototoxic allelochemical *α*-terthienyl in three herbivorous lepidoptera. *Pestic. Biochem. Physiol.*, 29(1):1–9, 1987.

[113] Hartmut Jaeschke, Gregory J Gores, Arthur I Cederbaum, Jack A Hinson, Dominique Pessayre, and John J Lemasters. Mechanisms of hepatotoxicity. *Toxicol. Sci.*, 65(2): 166–176, 2002.

[114] Chaitanya K Jaladanki, Nikhil Taxak, Rohith A Varikoti, and Prasad V Bharatam. Toxicity originating from thiophene containing drugs: Exploring the mechanism using quantum chemical methods. *Chem. Res. Toxicol.*, 28(12):2364–2376, 2015.

[115] Petra Jancova, Pavel Anzenbacher, and Eva Anzenbacherova. Phase ii drug metabolizing enzymes. *Biomed. Pap.*, 154(2):103–116, 2010.

[116] Wenying Jian, Hua-Fen Liu, Weiping Zhao, Elliott Jones, and Mingshe Zhu. Simultaneous screening of glutathione and cyanide adducts using precursor ion and neutral loss scans-dependent product ion spectral acquisition and data mining tools. *J. Am. Soc. Mass. Spectrom.*, 23(5):964–976, 2012.

[117] Elizabeth M Joshi, Brian H Heasley, Mahendra D Chordia, and Timothy L Macdonald. In vitro metabolism of 2-acetylbenzothiophene: relevance to zileuton hepatotoxicity. *Chem. Res. Toxicol.*, 17(2):137–143, 2004.

[118] Amit S Kalgutkar. Handling reactive metabolite positives in drug discovery: What has retrospective structure–toxicity analyses taught us? *Chem.-Biol. Interact.*, 192(1):46–55, 2011.

[119] 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? *Chem. Biodiversity*, 6(11):2115–2137, 2009.

[120] Amit S Kalgutkar, Deepak Dalvie, R Scott Obach, and Dennis A Smith. Bioactivation and natural products. *Reactive Drug Metabolites*, pages 203–224.

[121] Amit S Kalgutkar, Iain Gardner, R Scott Obach, Christopher L Shaffer, Ernesto Callegari, Kirk R Henne, Abdul E Mutlib, Deepak K Dalvie, Jae S Lee, Yasuhiro Nakai, John P O'Donnell, Jason Boer, and Shawn P Harriman. A comprehensive listing of bioactivation pathways of organic functional groups. *Curr. Drug Metab.*, 6(3):161–225, 2005.

[122] AS Kalgutkar. Should the incorporation of structural alerts be restricted in drug design? an analysis of structure-toxicity trends with aniline-based drugs. *Curr. Drug Metab.*, 22(4): 438–464, 2015.

[123] Neil Kaplowitz. Idiosyncratic drug hepatotoxicity. *Nat. Rev. Drug Discovery*, 4(6): 489–499, 2005.

[124] Gregory L Kedderis and Gerald T Miwa. The metabolic activation of nitroheterocyclic therapeutic agents. *Drug Metab. Rev.*, 19(1):33–62, 1988.

[125] Ute M Kent, MI Jushchhyshyn, and Paul F Hollenberg. Mechanism-based inactivators as probes of cytochrome p450 structure and function. *Curr. Drug Metab.*, 2(3):215–243, 2001.

[126] Johannes Kirchmair, Andreas H Göller, Dieter Lang, Jens Kunze, Bernard Testa, Ian D Wilson, Robert C Glen, and Gisbert Schneider. Predicting drug metabolism: experiment and/or computation? *Nat. Rev. Drug Discovery*, 2015.

[127] Luke L Koenigs and William F Trager. Mechanism-based inactivation of cytochrome p450 2b1 by 8-methoxypsoralen and several other furanocoumarins. *Biochemistry*, 37(38): 13184–13193, 1998.

[128] Martin Krzywinski, Inanc Birol, Steven Jones, and Marco Marra. Getting into visualization of large biological data sets. *BioVis2012, http://mkweb. bcgsc. ca/biovis2012*, 2012.

[129] Gondi N Kumar and Sekhar Surapaneni. Role of drug metabolism in drug discovery and development. *Med. Res. Rev.*, 21(5):397–411, 2001.

[130] David C Lamb, Michael R Waterman, and Bin Zhao. Streptomyces cytochromes p450: applications in drug metabolism. *Expert Opin. Drug Metab. Toxicol.*, 9(10):1279–1294, 2013.

[131] Silvia Lapenna, Mojca Fuart-Gatnik, and Andrew Worth. *Review of QSAR models and software tools for predicting acute and chronic systemic toxicity*. Publications Office, 2010.

[132] Karen E Lasser, Paul D Allen, Steffie J Woolhandler, David U Himmelstein, Sidney M Wolfe, and David H Bor. Timing of new black box warnings and withdrawals for prescription medications. *J. Am. Med. Assoc.*, 287(17):2215–2220, 2002.

[133] Pia Lautala, Brian T Ethell, Jyrki Taskinen, and Brian Burchell. The specificity of glucuronidation of entacapone and tolcapone by recombinant human udp-glucuronosyltransferases. *Drug Metab. Dispos.*, 28(11):1385–1389, 2000.

[134] Jason Lazarou, Bruce H Pomeranz, and Paul N Corey. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *J. Am. Med. Assoc.*, 279 (15):1200–1205, 1998.

[135] William M Lee. Drug-induced hepatotoxicity. *N. Engl. J. Med.*, 349(5):474–485, 2003.

[136] Angelique Leone, Alex Nie, J Brandon Parker, Sharmilee Sawant, Leigh-Anne Piechta, Michael F Kelley, L Mark Kao, S Jim Proctor, Geert Verheyen, Mark D Johnson, Peter G. Lord, and Michael K. McMillian. Oxidative stress/reactive metabolite gene expression signature in rat liver detects idiosyncratic hepatotoxicants. *Toxicol. Appl. Pharmacol.*, 275 (3):189–197, 2014.

[137] Feng Li, Frank J Gonzalez, and Xiaochao Ma. Lc–ms-based metabolomics in profiling of drug metabolism and bioactivation. *Acta Pharm. Sin. B*, 2(2):118–125, 2012.

[138] Adam Light and Patrick J Bartlein. The end of the rainbow? color schemes for improved data graphics. *Trans., Am. Geophys. Union*, 85(40):385–391, 2004.

[139] Christopher A Lipinski, Franco Lombardo, Beryl W Dominy, and Paul J Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.*, 23(1-3):3–25, 1997.

[140] SC Lloyd, DM Blackburn, and PMD Foster. Trifluoroethanol and its oxidative metabolites: comparison of in vivo and in vitro effects in the rat testis. *Fd. Chem. Toxic.*, 24(6-7): 653–654, 1986.

[141] SC Lloyd, DM Blackburn, and PMD Foster. Trifluoroethanol and its oxidative metabolites: Comparison of in vivo and in vitro effects of rat testis. *Toxicol Appl Pharmacol.*, 92(3): 390–401, 1988.

[142] Wiebke Lohmann and Uwe Karst. Generation and identification of reactive metabolites by electrochemistry and immobilized enzymes coupled on-line to liquid chromatography/mass spectrometry. *Anal. Chem.*, 79(17):6831–6839, 2007.

[143] M Pilar López-García, Patrick M Dansette, and Julio Coloma. Kinetics of tienilic acid bioactivation and functional generation of drug–protein adducts in intact rat hepatocytes. *Biochem. Pharmacol.*, 70(12):1870–1882, 2005.

[144] Per E Loünning. Aromatase inhibition for breast cancer treatment. *Acta Oncol.*, 35(S5): 38–43, 1996.

[145] 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*, pages 330–339. Springer, 1997.

[146] JL Maggs, MD Tingle, NR Kitteringham, and BK Park. Drug-protein conjugates—xiv: mechanisms of formation of protein-arylating intermediates from amodiaquine, a myelotoxin and hepatotoxin in man. *Biochem. Pharmacol.*, 37(2):303–311, 1988.

[147] Mukesh K Mahajan, Vinita Uttamsingh, J Scott Daniels, Liang-Shang Gan, Barbara W LeDuc, and David A Williams. In vitro metabolism of oxymetazoline: evidence for bioactivation to a reactive metabolite. *Drug Metab. Dispos.*, 39(4):693–702, 2011.

[148] Jonas Malmström, Mats Jonsson, Ian A Cotgreave, Leif Hammarström, Martin Sjödin, and Lars Engman. The antioxidant profile of 2, 3-dihydrobenzo [b] furan-5-ol and its 1-thio, 1-seleno, and 1-telluro analogues. *J. Am. Chem. Soc.*, 123(15):3434–3440, 2001.

[149] Annah Mancy, Pierre Broto, Sylvie Dijols, Patrick M Dansette, and Daniel Mansuy. The substrate binding site of human liver cytochrome p450 2c9: an approach using designed tienilic acid derivatives and molecular modeling. *Biochemistry*, 34(33):10365–10375, 1995.

[150] Daniel Mansuy and Patrick M Dansette. Sulfenic acids as reactive intermediates in xenobiotic metabolism. *Arch. Biochem. Biophys.*, 507(1):174–185, 2011.

[151] Satori A Marchitti, Chad Brocker, Dimitrios Stagos, and Vasilis Vasiliou. Non-p450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily. *Expert Opin. Drug Metab. Toxicol.*, 4(6):697–720, 2008.

169

[152] Marcella Martignoni, Geny MM Groothuis, and Ruben de Kanter. Species differences between mouse, rat, dog, monkey and human cyp-mediated drug metabolism, inhibition and induction. 2006.

[153] Eric Martin, Prasenjit Mukherjee, David Sullivan, and Johanna Jansen. Profile-qsar: a novel meta-qsar method that combines activities across the kinase family to accurately predict affinity, selectivity, and cellular activity. *Journal of chemical information and modeling*, 51(8):1942–1956, 2011.

[154] 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. *Chem. Res. Toxicol.*, 20(3):455–464, 2007.

[155] Matt K Matlock, Tyler B Hughes, and S Joshua Swamidass. Xenosite-server: A web-available site of metabolism prediction tool. *Bioinformatics*, page btu761, 2014.

[156] Matthew K Matlock, Tyler B Hughes, and S Joshua Swamidass. Xenosite server: a web-available site of metabolism prediction tool. *Bioinformatics*, 31(7):1136–1137, 2015.

[157] H Matsuura, G Saxena, SW Farmer, REW Hancock, and GHN Towers. Antibacterial and antifungal polyine compounds from glehnia littoralis ssp. leiocarpa. *Planta Med.*, 62(3): 256–259, 1996.

[158] Karen J McClellan and Julia A Balfour. Eprosartan. *Drugs*, 55(5):713–8, 1998.

[159] Randolph J McMurtry and Jerry R Mitchell. Renal and hepatic necrosis after metabolic activation of 2-substituted furans and thiophenes, including furosemide and cephaloridine. *Toxicol. Appl. Pharmacol.*, 42(2):285–300, 1977.

[160] Susie J Meade, Antonia G Miller, and Juliet A Gerrard. The role of dicarbonyl compounds in non-enzymatic crosslinking: a structure–activity study. *Bioorg. Med. Chem.*, 11(6): 853–862, 2003.

[161] Christine Medower, Lian Wen, and William W Johnson. Cytochrome p450 oxidation of the thiophene-containing anticancer drug 3-[(quinolin-4-ylmethyl)-amino]-thiophene-2-carboxylic acid (4-trifluoromethoxy-phenyl)-amide to an electrophilic intermediate. *Chem. Res. Toxicol.*, 21(8):1570–1577, 2008.

[162] Walter Meinl, John HN Meerman, and Hansruedi Glatt. Differential activation of promutagens by alloenzymes of human sulfotransferase 1a2 expressed in salmonella typhimurium. *Pharmacogenetics and Genomics*, 12(9):677–689, 2002.

[163] Amy E Mercer, Sophie L Regan, Charlotte M Hirst, Emma E Graham, Daniel J Antoine, Craig A Benson, Dominic P Williams, John Foster, J Gerry Kenna, and B Kevin Park. Functional and toxicological consequences of metabolic bioactivation of methapyrilene via thiophene s-oxidation: Induction of cell defence, apoptosis and hepatic necrosis. *Toxicol. Appl. Pharmacol.*, 239(3):297–305, 2009.

[164] Emmanuel Meschkat, Martin D Barratt, and Jean-Pierre Lepoittevin. Studies of the chemical selectivity of hapten, reactivity, and skin sensitization potency. 1. synthesis and studies on the reactivity toward model nucleophiles of the 13c-labeled skin sensitizers hex-1-ene-and hexane-1, 3-sultones. *Chem. Res. Toxicol.*, 14(1):110–117, 2001.

[165] Emmanuel Meschkat, Martin D Barratt, and Jean-Pierre Lepoittevin. Studies of the chemical selectivity of hapten, reactivity, and skin sensitization potency. 2. nmr studies of the covalent binding of the 13c-labeled skin sensitizers 2-[13c]-and 3-[13c] hex-1-ene-and 3-[13c] hexane-1, 3-sultones to human serum albumin. *Chem. Res. Toxicol.*, 14(1): 118–126, 2001.

[166] Elizabeth C Miller and James A Miller. The presence and significance of bound aminoazo dyes in the livers of rats fed p-dimethylaminoazobenzene. *Cancer Res.*, 7(7):468–480, 1947.

[167] Elizabeth C Miller and James A Miller. In vivo combinations between carcinogens and tissue constituents and their possible role in carcinogenesis. *Cancer Res.*, 12(8):547–556, 1952.

[168] Emmanuel F Minet, Gentile Daniela, Clive Meredith, and Eian D Massey. A comparative in vitro kinetic study of [14c]-eugenol and [14c]-methyleugenol activation and detoxification in human, mouse, and rat liver and lung fractions. *Xenobiotica*, 42(5): 429–441, 2012.

[169] Jerry R Mitchell, Wayne R Snodgrass, and James R Gillette. The role of biotransformation in chemical-induced liver injury. *Environ. Health Perspect.*, 15:27, 1976.

[170] JR Mitchell, WZ Potter, and DJ Jollow. Furosemide-induced hepatic and renal tubular necrosis. In *Proceedings of the Federation of American Societies for Experimental Biology*, volume 32, page 305. FEDERATION AMER SOC EXP BIOL 9650 ROCKVILLE PIKE, BETHESDA, MD 20814-3998, 1973.

[171] JR Mitchell, WZ Potter, JA Hinson, and DJ Jollow. Hepatic necrosis caused by furosemide. *Nature*, 251(5475):508, 1974.

[172] Christophe Morisseau. Role of epoxide hydrolases in lipid metabolism. *Biochimie*, 95(1): 91–95, 2013.

[173] Anne Mulhall, John de Louvois, and Rosalinde Hurley. Chloramphenicol toxicity in neonates: its incidence and prevention. *Br Med J (Clin Res Ed)*, 287(6403):1424–1427, 1983.

[174] AE Mutlib, H Chen, GA Nemeth, JA Markwalder, SP Seitz, LS Gan, and DD Christ. Identification and characterization of efavirenz metabolites by liquid chromatography/mass spectrometry and high field nmr: species differences in the metabolism of efavirenz. *Drug Metab. Dispos.*, 27(11):1319–1333, 1999.

171

[175] Dean J Naisbitt, Dominic P Williams, Munir Pirmohamed, Neil R Kitteringham, and B Kevin Park. Reactive metabolites and their role in drug reactions. *Curr. Opin. Allergy Clin. Immunol.*, 1(4):317–325, 2001.

[176] Sidney D Nelson. Structure toxicity relationships-how useful are they in predicting toxicities of new drugs? In *Biological reactive intermediates VI*, pages 33–43. Springer: Berlin, Germany, 2001.

[177] Sidney D Nelson and Paul G Pearson. Covalent and noncovalent interactions in acute lethal cell injury caused by chemicals. *Annu. Rev. Pharmacol. Toxicol.*, 30(1):169–195, 1990.

[178] Winnie Ng, Alexandra RM Lobach, Xu Zhu, Xin Chen, Feng Liu, Imir G Metushi, Amy Sharma, Jinze Li, Ping Cai, Julia Ip, et al. Animal models of idiosyncratic drug reactions. *Adv. Pharmacol.*, 63:81, 2012.

[179] Yumi Nishiya, Katsunobu Hagihara, Takashi Ito, Masami Tajima, Shin-ichi Miura, Atsushi Kurihara, Nagy A Farid, and Toshihiko Ikeda. Mechanism-based inhibition of human cytochrome p450 2b6 by ticlopidine, clopidogrel, and the thiolactone metabolite of prasugrel. *Drug Metab. Dispos.*, 37(3):589–593, 2009.

[180] Luis J Núñez-Vergara, JC Sturm, Claudio Olea-Azar, P Navarrete-Encina, S Bollo, and JA Squella. Electrochemical, uv-visible and epr studies on nitrofurantoin: Nitro radical anion generation and its interaction with glutathione. *Free Radical Res.*, 32(5):399–409, 2000.

[181] R Scott Obach, Phuong Huynh, Mary C Allen, and Christine Beedham. Human liver aldehyde oxidase: inhibition by 239 drugs. *The Journal of Clinical Pharmacology*, 44(1): 7–19, 2004.

[182] John P O'Donnell, Deepak K Dalvie, Amit S Kalgutkar, and R Scott Obach. Mechanism-based inactivation of human recombinant p450 2c9 by the nonsteroidal anti-inflammatory drug suprofen. *Drug Metab. Dispos.*, 31(11):1369–1377, 2003.

[183] John Okyere, Ekow Oppon, Daniel Dzidzienyo, Lav Sharma, and Graham Ball. Cross-species gene expression analysis of species specific differences in the preclinical assessment of pharmaceutical compounds. 2014.

[184] C Warren Olanow. Tolcapone and hepatotoxic effects. *Arch. Neurol.*, 57(2):263–267, 2000.

[185] Lars Olbe, Enar Carlsson, and Per Lindberg. A proton-pump inhibitor expedition: the case histories of omeprazole and esomeprazole. *Nature Reviews Drug Discovery*, 2(2):132–139, 2003.

[186] Noel M OLBoyle, Michael Banck, Craig A James, Chris Morley, Tim Vandermeersch, and Geoffrey R Hutchison. Open babel: An open chemical toolbox. *J Cheminf*, 3:33, 2011.

[187] Udo Oppermann. Carbonyl reductases: the complex relationships of mammalian carbonyl-and quinone-reducing enzymes and their role in physiology. *Annu. Rev. Pharmacol. Toxicol.*, 47:293–322, 2007.

[188] George Ostapowicz, Robert J Fontana, Frank V Schiødt, Anne Larson, Timothy J Davern, Steven HB Han, Timothy M McCashland, A Obaid Shakil, J Eileen Hay, Linda Hynan, et al. Results of a prospective study of acute liver failure at 17 tertiary care centers in the united states. *Ann. Clin. Med.*, 137(12):947–954, 2002.

[189] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[190] Aashish Pandit, Tarun Sachdeva, and Pallavi Bafna. Drug-induced hepatotoxicity: A review. 2012.

[191] B Kevin Park, Neil R Kitteringham, James L Maggs, Munir Pirmohamed, and Dominic P Williams. The role of metabolic activation in drug-induced hepatotoxicity. *Annu. Rev. Pharmacol. Toxicol.*, 45:177–202, 2005.

[192] BK Park, DJ Naisbitt, SF Gordon, NR Kitteringham, and M Pirmohamed. Metabolic activation in drug allergies. *Toxicology*, 158(1):11–23, 2001.

[193] Mario Pellegatti. Metabolites in safety testing: issues and approaches to the safety evaluation of human metabolites in a drug that is extensively metabolized. *J. Drug Metab. Toxicol.*, 1, 2010.

[194] Jianlong Peng, Jing Lu, Qiancheng Shen, Mingyue Zheng, Xiaomin Luo, Weiliang Zhu, Hualiang Jiang, and Kaixian Chen. In silico site of metabolism prediction for human ugt-catalyzed reactions. *Bioinformatics*, 30(3):398–405, 2014.

[195] Lisa A Peterson. Reactive metabolites in the biotransformation of molecules containing a furan ring. *Chem. Res. Toxicol.*, 26(1):6–25, 2012.

[196] Werner J Pichler, Dean J Naisbitt, and B Kevin Park. Immune pathomechanism of drug hypersensitivity reactions. *J. Allergy Clin. Immunol.*, 127(3):S74–S81, 2011.

[197] Thomayant Prueksaritanont, Raju Subramanian, Xiaojun Fang, Bennett Ma, Yue Qiu, Jiunn H Lin, Paul G Pearson, and Thomas A Baillie. Glucuronidation of statins in animals and humans: a novel mechanism of statin lactonization. *Drug Metab. Dispos.*, 30(5): 505–512, 2002.

[198] Vandana Purohit and Ashis K Basu. Mutagenicity of nitroaromatic compounds. *Chem. Res. Toxicol.*, 13(8):673–692, 2000.

[199] Olga Rechkoblit, Alexander Kolbanovskiy, Lucy Malinina, Nicholas E Geacintov, Suse Broyde, and Dinshaw J Patel. Mechanism of error-free and semitargeted mutagenic bypass of an aromatic amine lesion by y-family polymerase dpo4. *Nat. Struct. Mol. Biol.*, 17(3): 379–388, 2010.

[200] Marcus M Reidenberg, Dennis E Drayer, Beverly Lorenzo, Brian L Strom, Suzanne L West, Ellen S Snyder, Bruce Freundlich, and Paul D Stolley. Acetylation phenotypes and environmental chemical exposure of people with idiopathic systemic lupus erythematosus. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 36 (7):971–973, 1993.

[201] Christopher A Reilly, Fred Henion, Tim S Bugni, Manivannan Ethirajan, Chris Stockmann, Kartick C Pramanik, Sanjay K Srivastava, and Garold S Yost. Reactive intermediates produced from the metabolism of the vanilloid ring of capsaicinoids by p450 enzymes. *Chem. Res. Toxicol.*, 26(1):55–66, 2012.

[202] S Riedmaier, K Klein, U Hofmann, JE Keskitalo, PJ Neuvonen, M Schwab, M Niemi, and UM Zanger. Udp-glucuronosyltransferase (ugt) polymorphisms affect atorvastatin lactonization in vitro and in vivo. *Clin. Pharmacol. Ther.*, 87(1):65–73, 2010.

[203] Clare L Ritter and Danuta Malejka-Giganti. Nitroreduction of nitrated and c-9 oxidized fluorenes in vitro. *Chem. Res. Toxicol.*, 11(11):1361–1367, 1998.

[204] SJ Roffey, S Cole, P Comby, D Gibson, SG Jezequel, ANR Nedderman, DA Smith, DK Walker, and N Wood. The disposition of voriconazole in mouse, rat, rabbit, guinea pig, dog, and human. *Drug Metab. Dispos.*, 31(6):731–741, 2003.

[205] Gerald M Rosen, Helen A Demos, and Elmer J Rauckman. Not all aromatic nitro compounds form free radicals. *Toxicol. Lett.*, 22(2):145–152, 1984.

[206] Timo Rousu, Olavi Pelkonen, and Ari Tolonen. Rapid detection and characterization of reactive drug metabolites in vitro using several isotope-labeled trapping agents and ultra-performance liquid chromatography/time-of-flight mass spectrometry. *Rapid Rapid Commun. Mass. Spectrom.*, 23(6):843–855, 2009.

[207] Anastasia Rudik, Alexander Dmitriev, Alexey Lagunin, Dmitry Filimonov, and Vladimir Poroikov. Somp: web-server for in silico prediction of sites of metabolism for drug-like compounds. *Bioinformatics*, 10:btv087, 2015.

[208] Anastasia V Rudik, Alexander V Dmitriev, Alexey A Lagunin, Dmitry A Filimonov, and Vladimir V Poroikov. Metabolism site prediction based on xenobiotic structural formulas and pass prediction algorithm. *J. Chem. Inf. Model.*, 54(2):498–507, 2014.

[209] Aleksandra Rybacka, Christina Rudén, Igor V Tetko, and Patrik L Andersson. Identifying potential endocrine disruptors among industrial chemicals and their metabolites–development and evaluation of in silico tools. *Chemosphere*, 139:372–378, 2015.

[210] Patrik Rydberg, David E Gloriam, Jed Zaretzki, Curt Breneman, and Lars Olsen. Smartcyp: a 2d method for prediction of cytochrome p450-mediated drug metabolism. *ACS Med. Chem. Lett.*, 1(3):96–100, 2010.

[211] Patrik Rydberg, Richard Lonsdale, Jeremy N Harvey, Adrian J Mulholland, and Lars Olsen. Trends in predicted chemoselectivity of cytochrome p450 oxidation: B3lyp barrier heights for epoxidation and hydroxylation reactions. *Mol. Graphics Model.*, 52:30–35, 2014.

[212] Seigo Sanoh, Yoshitaka Tayama, Kazumi Sugihara, Shigeyuki Kitamura, and Shigeru Ohta. Significance of aldehyde oxidase during drug development: Effects on drug metabolism, pharmacokinetics, toxicity, and efficacy. *Drug Metab. Pharmacokinet.*, 30(1):52–63, 2015.

[213] RA Saunders, JR Griffith, and FE Saalfeld. Identification of some organic smog components based on rain water analysis. *Biol. Mass Spectrom.*, 1(3):192–194, 1974.

[214] Johannes AH Schwöbel, Dominik Wondrousch, Yana K Koleva, Judith C Madden, Mark TD Cronin, and Gerrit Schüürmann. Prediction of michael-type acceptor reactivity toward glutathione. *Chem. Res. Toxicol.*, 23(10):1576–1585, 2010.

[215] Janeric Seidegård and Gunilla Ekström. The role of human glutathione transferases and epoxide hydrolases in the metabolism of xenobiotics. *Environ. Health Perspect.*, 105 (Suppl 4):791, 1997.

[216] C Senda, W Kishimoto, K Sakai, A Nagakura, and T IGARASHI*. Identification of human cytochrome p450 isoforms involved in the metabolism of brotizolam. *Xenobiotica*, 27(9):913–922, 1997.

[217] Todd W Shearer, Michael P Kozar, Michael T O'Neil, Philip L Smith, Guy A Schiehser, David P Jacobus, Damaris S Diaz, Young-Sun Yang, Wilbur K Milhous, and Donald R Skillman. In vitro metabolism of phenoxypropoxybiguanide analogues in human liver microsomes to potent antimalarial dihydrotriazines. *J. Med. Chem.*, 48(8):2805–2813, 2005.

[218] Jacintha M Shenton, Jie Chen, and Jack P Uetrecht. Animal models of idiosyncratic drug reactions. *Chem.-Biol. Interact.*, 150(1):53–70, 2004.

[219] Edith Sim, Lesley Stanley, Edward W Gill, and Alison Jones. Metabolites of procainamide and practolol inhibit complement components c3 and c4. *Biochemical Journal*, 251(2): 323–326, 1988.

[220] Ana L Simplício, John M Clancy, and John F Gilmer. Prodrugs for amines. *Molecules*, 13 (3):519–547, 2008.

[221] Suresh B Singh, Lucy Q Shen, Matthew J Walker, and Robert P Sheridan. A model for predicting likely sites of cyp3a4-mediated metabolism on drug-like molecules. *J. Med. Chem.*, 46(8):1330–1336, 2003.

[222] Gary Sisson, Avery Goodwin, Ausra Raudonikiene, Nicky J Hughes, Asish K Mukhopadhyay, Douglas E Berg, and Paul S Hoffman. Enzymes associated with reductive activation and action of nitazoxanide, nitrofurans, and metronidazole in helicobacter pylori. *Antimicrob. Agents Chemother.*, 46(7):2116–2123, 2002.

[223] Howard S Smith. Opioid metabolism. In *Mayo Clinic Proceedings*, volume 84, pages 613–624. Elsevier, 2009.

[224] Sanjay K Srivastava, Simon C Watkins, Erin Schuetz, and Shivendra V Singh. Role of glutathione conjugate efflux in cellular protection against benzo [a] pyrene-7, 8-diol-9, 10-epoxide–induced dna damage. *Mol. Carcinog.*, 33(3):156–162, 2002.

[225] Andrew V Stachulski, Thomas A Baillie, B Kevin Park, R Scott Obach, Deepak K Dalvie, Dominic P Williams, Abhishek Srivastava, Sophie L Regan, Daniel J Antoine, Christopher EP Goldring, Alvin J.L. Chia, Neil R Kitteringham, Laura E. Randle, Hayley Callan, Castrejon J. Luis, John Farrell, Dean J Naisbitt, and Martin S. Lennard. The generation, detection, and effects of reactive drug metabolites. *Med. Res. Rev.*, 33(5): 985–1080, 2013.

[226] Antonia F Stepan, Daniel P Walker, Jonathan Bauman, David A Price, Thomas A Baillie, Amit S Kalgutkar, and Michael D Aleo. Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the united states. *Chem. Res. Toxicol.*, 24(9):1345–1410, 2011.

[227] A Sunesson, W Vaes, C Nilsson, Goran Blomquist, Barbro Andersson, and Rolf Carlson. Identification of volatile metabolites from five fungal species cultivated on two media. *Appl. Environ. Microbiol.*, 61(8):2911–2918, 1995.

[228] S Joshua Swamidass, Chloé-Agathe Azencott, Kenny Daily, and Pierre Baldi. A croc stronger than roc: measuring, visualizing and optimizing early retrieval. *Bioinformatics*, 26 (10):1348–1356, 2010.

[229] Carol Sweeney, Christine B Ambrosone, Lija Joseph, Angie Stone, Laura F Hutchins, Fred F Kadlubar, and Brian F Coles. Association between a glutathione s-transferase a1 promoter polymorphism and survival after breast cancer treatment. *International journal of cancer*, 103(6):810–814, 2003.

[230] Melanie N Tallman, Joseph K Ritter, and Philip C Smith. Differential rates of glucuronidation for 7-ethyl-10-hydroxy-camptothecin (sn-38) lactone and carboxylate in human and rat microsomes and recombinant udp-glucuronosyltransferase isoforms. *Drug Metab. Dispos.*, 33(7):977–983, 2005.

[231] Wei Tang and Anthony YH Lu. Metabolic bioactivation and drug-related adverse effects: current status and future directions from a pharmaceutical research perspective. *Drug Metab. Rev.*, 42(2):225–249, 2010.

[232] Renli Teng, Stuart Oliver, Martin A Hayes, and Kathleen Butler. Absorption, distribution, metabolism, and excretion of ticagrelor in healthy subjects. *Drug Metab. Dispos.*, 38(9): 1514–1521, 2010.

[233] Bernard Testa, Alessandro Pedretti, and Giulio Vistoli. Reactions and enzymes in the metabolism of drugs and other xenobiotics. *Drug Discovery Today*, 17(11):549–560, 2012.

[234] Bernhard Thalhamer, Wolfgang Buchberger, and Mario Waser. Identification of thymol phase i metabolites in human urine by headspace sorptive extraction combined with thermal desorption and gas chromatography mass spectrometry. *J. Pharm. Biomed. Anal.*, 56(1):64–69, 2011.

[235] Lei Tian, Juanjuan Jiang, Yiling Huang, Li Xu, Hong Liu, and Yishi Li. Determination of aranidipine and its active metabolite in human plasma by liquid chromatography/negative electrospray ionization tandem mass spectrometry. *Rapid Commun. Mass Spectrom.*, 20 (19):2871–2877, 2006.

[236] Jack Uetrecht. N-oxidation of drugs associated with idiosyncratic drug reactions. *Drug Metab. Rev.*, 34(3):651–665, 2002.

[237] Jack Uetrecht and Dean J Naisbitt. Idiosyncratic adverse drug reactions: current concepts. *Pharmacol. Rev.*, 65(2):779–808, 2013.

[238] Riccardo Utili, John K Boitnott, and HYMAN J Zimmerman. Dantrolene-associated hepatic injury. incidence and character. *Gastroenterology*, 72(4 Pt 1):610–616, 1977.

[239] S Vasoo. Drug-induced lupus: an update. *Lupus*, 15(11):757–761, 2006.

[240] Jeffrey D Vassallo, Sarah M Hicks, George P Daston, and Lois D Lehman-McKeeman. Metabolic detoxification determines species differences in coumarin-induced hepatotoxicity. *Toxicological Sciences*, 80(2):249–257, 2004.

[241] Alison EM Vickers, John R Sinclair, Markus Zollinger, Francis Heitz, Ulrike Glänzel, Laurie Johanson, and Volker Fischer. Multiple cytochrome p-450s involved in the metabolism of terbinafine suggest a limited potential for drug-drug interactions. *Drug Metab. Dispos.*, 27(9):1029–1038, 1999.

[242] Tom B Vree, Louis Maljers, Noud Borg, Nico MM Nibbering, Corrien PWGM Verwey-van Wissen, Aart Lagerwerf, Robert AA Maes, and P Jozef H Jongen. High-performance liquid-chromatographic-atmospheric-pressure chemical-ionization ion-trap mass-spectrometric identification of isomeric c6-hydroxy and c20-hydroxy metabolites of methylprednisolone in the urine of patients receiving high-dose pulse therapy. *J. Pharm. Pharmacol.*, 51(10):1155–1166, 1999.

[243] Jennie L Walgren, Michael D Mitchell, and David C Thompson. Role of metabolism in drug-induced idiosyncratic hepatotoxicity. *Crit. Rev. Toxicol.*, 35(4):325–361, 2005.

[244] DC Washington. National action plan for adverse drug event prevention, 2014.

[245] Paul B Watkins and Leonard B Seeff. Drug-induced liver injury: Summary of a single topic clinical research conference. *Hepatology*, 43(3):618–631, 2006.

[246] Peter G Wells, Peter I Mackenzie, Jayanta Roy Chowdhury, Chantal Guillemette, Philip A Gregory, Yuji Ishii, Antony J Hansen, Fay K Kessler, Perry M Kim, Namita Roy Chowdhury, and Joseph K Ritter. Glucuronidation and the udp-glucuronosyltransferases in health and disease. *Drug Metab. Dispos.*, 32(3):281–290, 2004.

[247] Dominic P Williams and Dean J Naisbitt. Toxicophores: groups and metabolic routes associated with increased safety risk. *Curr. Opin. Drug Discovery Dev.*, 5(1):104–115, 2002.

177

[248] GM Williams, A Mattia, and A Renwick. *Safety evaluation of certain food additives: Furan-substituted aliphatic hydrocarbons, alcohols, aldehydes, ketones, carboxylic acids and related esters, sulfides, disulfides and ethers.(addendum)*. World Health Organization: Geneva, Switzerland, 2009.

[249] J Andrew Williams, Ruth Hyland, Barry C Jones, Dennis A Smith, Susan Hurst, Theunis C Goosen, Vincent Peterkin, Jeffrey R Koup, and Simon E Ball. Drug-drug interactions for udp-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (auci/auc) ratios. *Drug Metab. Dispos.*, 32(11):1201–1208, 2004.

[250] 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. *Mol. Pharmacol.*, 12(5):759–768, 1976.

[251] David S Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, 36(suppl 1):D901–D906, 2008.

[252] Bang Wong. Points of view: Color coding. *Nat. Methods*, 7(8):573–573, 2010.

[253] David T Wong, Frank P Bymaster, Leroy R Reid, Douglas A Mayle, Joseph H Krushinski, and David W Robertson. Norfluoxetine enantiomers as inhibitors of serotonin uptake in rat brain. *Neuropsychopharmacology*, 8(4):337–344, 1993.

[254] Marty Wulferink, Sabine Dierkes, and Ernst Gleichmann. Cross-sensitization to haptens: formation of common haptenic metabolites, t cell recognition of cryptic peptides, and true t cell cross-reactivity. *Eur. J. Immunol.*, 32(5):1338–1348, 2002.

[255] Ke Yu, Xingchao Geng, Minjun Chen, Jie Zhang, Bingshun Wang, Katarina Ilic, and Weida Tong. High daily dose and being a substrate of cytochrome p450 enzymes are two important predictors of drug-induced liver injury. *Drug Metab. Dispos.*, 42(4):744–750, 2014.

[256] Adel A Yunis. Differential in-vitro toxicity of chloramphenicoi, nitroso-chloramphenicol, and thiamphenicol. *Sexually transmitted diseases*, 11:340–342, 1984.

[257] Jed Zaretzki, Patrik Rydberg, Charles Bergeron, Kristin P Bennett, Lars Olsen, and Curt M Breneman. Rs-predictor models augmented with smartcyp reactivities: robust metabolic regioselectivity predictions for nine cyp isozymes. *J. Chem. Inf. Model.*, 52(6):1637–1659, 2012.

[258] Jed Zaretzki, Matthew Matlock, and S Joshua Swamidass. Xenosite: Accurately predicting cyp-mediated sites of metabolism with neural networks. *J. Chem. Inf. Model.*, 53(12): 3373–3383, 2013.

[259] Jed Zaretzki, Kevin Boehm, and S Joshua Swamidass. Improved prediction of cyp-mediated metabolism with chemical fingerprints. *J. Chem. Inf. Model.*, 55(5): 972–982, 2015.

178

[260] JM Zaretzki, MR Browning, TB Hughes, and SJ Swamidass. Extending p450 site-of-metabolism models with region-resolution data. *Bioinformatics*, 31(12):1966, 2015.

[261] Mingyue Zheng, Xiaomin Luo, Qiancheng Shen, Yong Wang, Yun Du, Weiliang Zhu, and Hualiang Jiang. Site of metabolism prediction for six biotransformations mediated by cytochromes p450. *Bioinformatics*, 25(10):1251–1258, 2009.

[262] Shufeng Zhou, Hwee-Ling Koh, Yihuai Gao, Zhi-yuan Gong, and Edmund Jon Deoon Lee. Herbal bioactivation: the good, the bad and the ugly. *Life Sci.*, 74(8):935–968, 2004.

[263] Mingshe Zhu, Li Ma, Haiying Zhang, and W Griffith Humphreys. Detection and structural characterization of glutathione-trapped reactive metabolites using liquid chromatography-high-resolution mass spectrometry and mass defect filtering. *Anal. Chem.*, 79(21): 8333–8341, 2007.

# Na Le Dang

| CONTACT INFORMATION | 5544 Waterman Blvd Apt. 2W<br>St. Louis, MO 63112 | 314-899-6049<br>dangnl@wustl.edu |
|---|---|---|

RESEARCH INTERESTS

Machine learning, neural network, chemical informatics, metabolism modeling

EDUCATION

**Washington University School of Medicine in St. Louis**, St. Louis, MO

Ph.D., Computational and Systems Biology, *Expected:* May 2021
- Thesis Topic: *Modeling Small Molecule Metabolism in Human Liver Microsome to Better Predict Toxicity Risk*
- Advisor: S Joshua Swamidass, M.D. Ph.D.

M.D., Expected: May 2021

**Wesleyan University**, Middletown, CT

M.A., Chemistry , May 2012
- Thesis Topic: *Molecular Dynamics Studies of Free Energy Landscapes for Protein Folding and Binding: The U1A-RNA Complex and Related Systems*
- Advisor: David Beveridge, Ph.D.

B.A., Molecular Biology and Biochemistry Double Major, May 2011

PEER REVIEWED ARTICLES

11. Arghya Datta, Matthew K. Matlock, **Na Le Dang**,Thiago Moulin, Keith F. Woeltje, Elizabeth L. Yanik, and S. Joshua Swamidass. 'Black Box' to 'Conversational' Machine Learning: Ondansetron Reduces Risk of Hospital-Acquired Venous Thromboembolism. IEEE Journal of Biomedical and Health Informatics, Oct. 2020. PMID: 33095721

10. Tyler B. Hughes, **Na Le Dang**, Ayush Kumar, Noah Flynn, and S. Joshua Swamidass. The Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites *J. Chem. Inf. Model.*, Sep. 2020. PMID: 32881497

9. Noah Flynn, **Na Le Dang**, Michael D. Ward, and S. Joshua Swamidass. XenoNet: Inference and Likelihood of Intermediate Metabolite Formation, *J. Chem. Inf. Model.*, June 2020. PMID: 32525671

8. **Na Le Dang**, Tyler B. Hughes, Mathew K. Matlock, and S. Joshua Swamidass. The Metabolic Rainbow: Deep Learning of Phase I Metabolism in Five Color, *J. Chem. Inf. Model.*, Feb. 2020. PMID: 32040319

7. Dusty Barnette, Mary A. Davis, **Na Le Dang**, Anirudh S. Pidugu, Tyler Hughes, S. Joshua Swamidass, Gunnar Boysen, and Grover Miller. Lamisil (terbinafine) Toxicity: Determining pathways to bioactivation through computational and experimental approaches. *Biochem. Pharm.*, 2018. PMID: 30076845

6. **Na Le Dang**, Tyler B. Hughes, Grover P. Miller and S. Joshua Swamidass. Computationally Assessing the Bioactivation of Drugs by N-Dealkylation. *Chem. Res. Toxicol.*, 2018. PMID: 29355304

5. Mathew K. Matlock, **Na Le Dang**, and S. Joshua Swamidass. Learning a Local-Variable Model of Aromatic and Conjugated Systems. *ACS Cent. Sci.*, 2018. PMID: 29392176

4. **Na Le Dang**, Tyler B. Hughes, Grover P. Miller and S. Joshua Swamidass. Computational Approach to Structural Alerts: Furans, Phenols, Nitroaromatics, and Thiophenes. *Chem. Res. Toxicol.*, 2017. PMID: 28256829

3. Tyler B. Hughes, **Na Le Dang**, Grover P. Miller, and S. Joshua Swamidass. Modeling Reactivity to Biological Macromolecules with a Deep Multitask Network. *ACS Cent. Sci.*, 2016. PMID: 27610414

2. West C. Van Voorhis, W.C., ..., **Na Le Dang**, ..., Paul Willis. Open-source discovery with the Malaria Box compound collection for neglected diseases and beyond. *PLoS Medicine*, 2016. PMID: 27467575

1. **Na Le Dang**, Tyler B. Hughes, Varun Krishnamurthy, and S. Joshua Swamidass. A Simple Formula Predicts UGT-Mediated Metabolism. *Bioinformatics*, 2016. PMID: 27324196

PRE-PRINT ARTICLES

2. Matthew K Matlock, Arghya Datta, **Na Le Dang**, Kevin Jiang, and S Joshua Swamidass. Deep learning long-range information in undirected graphs with wave networks. *arXiv*, doi: preprint arXiv:1810.12153, 2018

1. Matthew Matlock, **Na Le Dang**, David Brown, an S. Joshua Swamidass. A Sign of Disparity: Racial/Ethnic Composition of Treatment Centers is an Independent Risk Factor in SPRINT Trial. *bioRxiv*, doi: https://doi.org/10.1101/102111, 2017

PAPERS UNDER REVIEW

1. Tyler B. Hughes, Noah R. Flynn, **Na Le Dang**, and S. Joshua Swamidass. Modeling the Bioactivation and Subsequent Reactivity of Drugs. (CRT, Submitted December, 2020)

PAPERS IN PREPARATION

3. **Na Le Dang**, Noah R. Flynn, Lev Morgan, and S. Joshua Swamidass. Predicting Bioactivation Potential of Therapeutic Agents for COVID-19 and Related Human Coronavirus Diseases.

2. **Na Le Dang**, Mathew K. Matlock, and S. Joshua Swamidass. Modeling *in vivo* competition between Phase I and Phase II Metabolism.

1. **Na Le Dang**, Arghya Datta, Jon Silva and S. Joshua Swamidass. Modeling Cytochrome P450 Isozyme Specific Metabolism.

PRESENTATIONS

Talks
- Washington University in St. Louis MSTP Reatreat      April 2018
  *Deep Learning Mechanism of Drug Toxicity*
- Washington University in St. Louis MSTP WIP      November 2017
  *Deep Learning Mechanism of Drug Toxicity*
- Washington University in St. Louis MGG/CSB retreat      September 2016
  *Modeling Phase I Metabolism of Small Molecules*
- Biomolecular Structure and Dynamics Meeting, Guilderland, NY      June 2011
  *Insight into Transition States in Protein RNA Recognition through MD Simulation*
- Science Computing and Informatics Center Seminar Series      December 2011
  *Graph Theory and Its Applications*

Posters
- ACS National Meeting      August 2018
  *Deep Learning of Isozyme Specific Phase I Metabolism*
- Gordon Research Conference in Drug Metabolism      July 2018
  *Deep Learning of Isozyme Specific Phase I Metabolism*
- Experimental Biology Annual Meeting      April 2018
  *Deep Learning of Phase I Metabolism*
- 13th AAP/ASCI/APSA Annual Meeting      April 2017
  *Modeling Phase 1 Reactions: Towards Metabolite Prediction*
- 20th North American ISSX meeting      October 2015
  *Predicting N-Dealkylation Reactions: Sites of Metabolism, Metabolites and Reactivity*
- Washington University in St. Louis MGG/CSB retreat      September 2015
  *Predicting Sites of UGT-Mediated Metabolism*
- Wesleyan HHMI Summer Research Poster Session      August 2010
  *Molecular Dynamics Studies of Thermal Unfolding of the RNA Recognition Motif Protein, U1A*

- Wesleyan HHMI Summer Research Poster Session                              August 2009
  *Investigating the Role of Promoter Motifs in the Transcriptional Regulation of RRB Genes*
- Wesleyan HHMI Summer Research Poster Session                              August 2008
  *Impacts of Glucose Concentration on Ribosomal Biogenesis (RRB) Genes Expression Profiles*

Teaching Experience

**Teaching Assistant**                                                     Fall 2014-Spring 2015
Cell and Organ Systems Biology M75 503: Histology
Instructor: Paul C. Bridgman, Ph.D
Anatomy and Neurobiology Department
Washington University School of Medicine in St Louis

**Teaching Assistant**                                                     Fall 2010-Spring 2011
Organic Chemistry
Instructor: Michael A. Calter, Ph.D
Chemistry Department
Wesleyan University

Research Experience

**Post-doc researcher**                                                    April 2020 to present
PIs: Jon Silva, Ph.D. and S Joshua Swamidass, M.D. Ph.D.
McKelvey School of Engineering,
Washington University in St. Louis

**Research Assistant**                                                     July 2014 to August 2018
PIs: S Joshua Swamidass, M.D. Ph.D.
Pathology and Immunology Department,
Washington University School of Medicine in St. Louis

**Research Assistant**                                                     June 2012 to Aug 2012
PIs: Gary Stormo, Ph.D and S Joshua Swamidass, M.D. Ph.D.
Computational and Systems Biology Department,
Washington University School of Medicine in St Louis

**Research Assistant**                                                     Sep 2009 to May 2012
PIs: David Beveridge, Ph.D.
Chemistry Department,
Wesleyan University

**Research Assistant**                                                     December 2007 to May 2009
PIs: Michael McAlear, Ph.D.
Molecular Biology and Biochemistry,
Wesleyan University

Extracuricular Activities

**Epidemiology Research**                                                  July 2020 to present
PIs: Victoria J. Fraser, M.D.
John T. Milliken Department of Medicine,
Washington University School of Medicine in St Louis

This clinical trial investigates epidemiology, transmission, and outcomes of COVID-19 among patients and healthcare providers in three hemodialysis units associated with Washington University.

**Clinical Trial Research**                                                April 2020 to present
PIs: Brian Kim, M.D. and Dina Katabi, Ph.D.
Washington University School of Medicine in St Louis and Massachusetts Institute of

Technology

The project is a student-led clinical trial to monitor and predict clinical deterioration among COVID-19 patients at their homes using an artificial intelligence-based technology. We use a wireless device designed by MIT researchers to monitor physiology parameters such as respiratory rate, heart rate, and other motions.

**Volunteering at Saturday Neighborhood Health Clinic**     August 2012 to September 2018
Work as preclinical and clinical student volunteer at a medical student-run free clinic that serves the adult uninsured population of St. Louis.

Awards        Student Awards — Wesleyan University
- Freeman Asian Scholarship                                         August 2007
  Merit-based four-year scholarship for bachelor studies at Wesleyan University awarded to eleven international students from Asia to promote cross-cultural understanding.
- Hawk Prize                                                        May 2011
  A prize awarded annually to Wesleyan seniors who have done the most effective research in biochemistry
- Dean's List                                          December 2010, May 2011
  High academic achievement Award

References    S. Joshua Swamidass, M.D. Ph.D.
Associate Professor                                      Phone: 949-394-9906
Department of Pathology and Immunology           E-mail: swamidass@wustl.edu
Washington University School of Medicine in St. Louis

David Beveridge, Ph.D.
Professor                                                Phone: 860-865-2575
Chemistry Department                           E-mail: dbeveridge@wesleyan.edu
Wesleyan University

Technical Skills
- Methods: Statistical machine learning methods including neural networks, random forests, SVMs, and others
- Compter Languages: Python, Perl, R, LaTeX
- Software: Tensorflow, Matlab, Mathematica