JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

Metabolic Forest: Predicting the Diverse Structures of Drug Metabolites

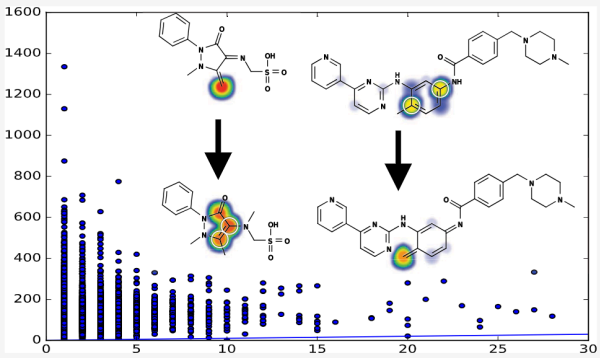
Tyler B. Hughes, Na Le Dang, Ayush Kumar, Noah R. Flynn, and S. Joshua Swamidass*

Cite This: J. Chem. Inf. Model. 2020, 60, 4702−4716

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Adverse drug metabolism often severely impacts patient morbidity and mortality. Unfortunately, drug metabolism experimental assays are costly, inefficient, and slow. Instead, computational modeling could rapidly flag potentially toxic molecules across thousands of candidates in the early stages of drug development. Most metabolism models focus on predicting sites of metabolism (SOMs): the specific substrate atoms targeted by metabolic enzymes. However, SOMs are merely a proxy for metabolic structures: knowledge of an SOM does not explicitly provide the actual metabolite structure. Without an explicit metabolite structure, computational systems cannot evaluate the new molecule's properties. For example, the metabolite's reactivity cannot be automatically predicted, a crucial limitation because reactive drug metabolites are a key driver of adverse drug reactions (ADRs). Additionally, further metabolic events cannot be forecast, even though the metabolic path of the majority of substrates includes two or more sequential steps. To overcome the myopia of the SOM paradigm, this study constructs a well-defined system—termed the metabolic forest—for generating exact metabolite structures. We validate the metabolic forest with the substrate and product structures from a large, chemically diverse, literature-derived dataset of 20 736 records. The metabolic forest finds a pathway linking each substrate and product for 79.42% of these records. By performing a breadth-first search of depth two or three, we improve performance to 88.43 and 88.77%, respectively. The metabolic forest includes a specialized algorithm for producing accurate quinone structures, the most common type of reactive metabolite. To our knowledge, this quinone structure algorithm is the first of its kind, as the diverse mechanisms of quinone formation are difficult to systematically reproduce. We validate the metabolic forest on a previously published dataset of 576 quinone reactions, predicting their structures with a depth three performance of 91.84%. The metabolic forest accurately enumerates metabolite structures, enabling promising new directions such as joint metabolism and reactivity modeling.



INTRODUCTION

Safety problems are one of the primary causes of drug candidate attrition.[1−3] Furthermore, idiosyncratic adverse drug reactions (IADRs) frequently only arise after approval, incurring significant resources.[4,5] Many IADRs present as drug-induced liver injury, the leading driver of drug withdrawal from the market.[6−10] Although IADRs are poorly understood, many are linked to bioactivation: enzymatic conversion of drugs to electrophilically reactive metabolites.[11−15] Reactive metabolites covalently bind to nucleophilic sites within biological macromolecules, including DNA[16−18] and off-target proteins.[19−21] Conjugation to DNA is frequently mutagenic,[22,23] and conjugation to proteins can disrupt their functions[24,25] or incite adverse immune responses leading to IADRs.[26−30]

Detecting reactive metabolites early in the development process is essential to minimize toxicity risk.[10,31−33] If bioactivation is observed for a particular lead compound, small structural differences may avoid reactive metabolite formation without significantly affecting its pharmacophore.[4,34] Mass spectrometry assays utilizing simple traps such as glutathione or cyanide is frequently used for initial screening of reactive metabolites-forming compounds.[35,36] However, while glutathione and cyanide both only have a single nucleophilic side, macromolecules such as protein and DNA are much larger and contain multiple nucleophilic sites.[37] Consequently, binding to glutathione and cyanide may not accurately predict binding to protein and DNA, hence a significant caveat for assays utilizing these simple traps.[38,39]

Received: April 13, 2020
Published: September 3, 2020

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

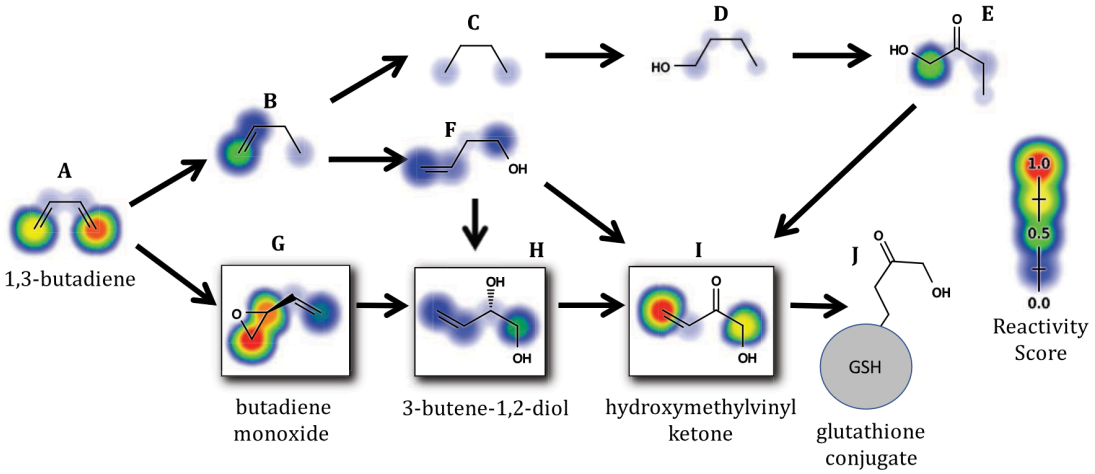


Figure 1. Metabolic forest explores metabolic pathways by sequentially applying metabolism rules. These pathways can reproduce experimentally observed structures or enumerate new structures. Oftentimes, there are several possible pathways between substrates and metabolites. For example, butadiene (A), a carcinogenic industrial chemical,[15,79−81] has several known metabolites.[81] The bottom pathway, (A) → (G) → (H) → (I) → (J), demonstrates an experimentally elucidated pathway. First, cytochromes P450 and myeloperoxidase epoxidize 1,3-butadiene[82−84] form butadiene monoxide (G). Second, epoxide hydrolase forms 3-butene-1,2-diol (H).[85] Third, cytochromes P450 form hydroxymethylvinyl ketone (I), a highly reactive Michael acceptor.[86] Fourth, glutathione conjugates to the reactive metabolite (J).[86] Three alternative pathways are shown. The (A) → (B) → (F) → (I) → (J) pathway has an equivalent number of metabolic steps, with two different intermediates (B, F) from the experimentally known pathway. The (A) → (B) → (C) → (D) → (E) → (I) → (J) pathway has several additional steps. The color shading on each site shows reactivity score assigned by our published reactivity model.[77] The score ranged from 0 to 1, displayed as rainbow shading on each site in accordance with the provided reactivity scale. The scale provides both color and size cue for prediction interpretation.

Binding to biological macromolecules can be more accurately quantified in experiments with radiolabeled compounds.[34,40] Still, radiolabeled compounds are generally not used until the late stage of preclinical development because they are expensive to synthesize.

Computational approaches could facilitate reactive metabolite detection by rapidly flagging drug candidates susceptible to bioactivation, thereby reducing both the total number of needed experiments and the chances of missing potentially toxic molecules. Currently, the dominant computational aid for avoiding bioactivation is the structural alert approach.[4,41−43] Using a database of structural alerts, which are simply motifs such as phenols or furans that are often bioactivated, molecules containing those substructures can easily be flagged.[44] However, structural alerts have several shortcomings: they are often not actually bioactivated due to specific molecular context, they do not match all drugs that produce reactive metabolites, and they are purely retrospective.[45−47]

Instead, several alternative, more flexible approaches have the potential to learn concepts from the data that generalize to never-before-seen structures. The majority of the work focus on predicting sites of metabolism (SOMs): the specific atoms that are metabolically modified by enzymes.[47−64] while others concentrate on metabolite structure inference from the parent compound.[65−72]

A molecule's SOMs can be used to infer the structures of its metabolites or suggest where a molecule might be rationally redesigned. However, SOM predictors also have limitations. First, they omit consideration of reactivity and therefore miss the second half of the bioactivation mechanism. As a result, benign metabolic steps like hydroxylation are indistinguishable from bioactivation events like epoxidation. For example, the epoxidation of carbamazepine to carbamazepine-10,11-epoxide, which is linked to adverse reactions,[73−75] would eventually form the same product as hydroxylation at both carbons of the bond where the epoxide forms. Second, SOM models are limited by datasets containing missing intermediates that elude experimental detection. For example, one systematic review of experimental metabolism literature reported that only 45.6% of all metabolites were formed by a single step, compared to 32.0 and 26.5% of metabolites that were generated by two or three or more steps, respectively.[76] Both shortcomings of the SOM paradigm are solvable by explicitly enumerating metabolite structures.

There are published computational metabolism tools that infer metabolite structures from parent compounds.[65−72] In this work, we built the metabolic forest: a collection of metabolism rules that rapidly enumerates metabolite structures for one or several sequential steps (Figure 1). To rigorously evaluate its ability to reproduce experimentally observed structures, we quantified performance across a large, literature-derived dataset of tens of thousands of reactions. By performing a systematic search between reported substrates and products, the metabolic forest automatically labels SOMs (and reveals manual annotation errors) and suggests missing intermediate structures. The metabolic forest predicts many different types of metabolism, including quinone formation. Although their formation is nontrivial to programmatically represent, quinone species are especially important to include in a useful method because they compose over 40% of known reactive metabolites.[76] Using the quinone and epoxide structures generated by the metabolic forest, we link previously developed metabolism[45,63] and reactivity[77] models to generate bioactivation hypotheses. A direct comparison of our approach to BioTransformer,[72] SyGMa,[70] and GLORY[71] is shown in our subsequent work.[78]

■ DATA AND METHODS

Phase I Metabolism Data. We measured the performance of the metabolite predictor on the phase I metabolism dataset,

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

detailed in our recent study.[64] In short, these data consisted of 20 736 *in vitro* and *in vivo* phase I human records from the literature-derived Accelrys Metabolite Database (AMD). We identified five categories of phase I metabolism: stable oxygenation, unstable oxygenation, dehydrogenation, hydrolysis, and reduction, and manually labeled 10 280, 5811, 2794, 3869, and 1590 sites of metabolism. Due to their complex, often multistep mechanisms, quinone formations do not fit well into the criteria used to extract and label the phase I dataset. Instead, to quantify performance at predicting quinone structures, we used a dataset from one of our previous studies consisting of 576 quinone formations extracted from the AMD.[45]

Metabolic Forest. We built a metabolite structure predictor that rapidly enumerates trees of metabolic pathways. This algorithm—codenamed the metabolic forest—was built in python using the 2017.09.01 release of RDKIT, an open-source cheminformatics package.[87] The metabolic forest included 24 reaction rules (Table 1). Each rule belonged to a ruleset, including (1) each of the five broad classes of phase I metabolism labeled in our simultaneous study,[64] (2) conjugation, (3) quinone formation, and (4) tautomerization. The conjugation and tautomerization rulesets encompass transformations that were sometimes implicitly included in our phase I and quinone formation datasets, perhaps due to experimental limitations.

Table 1. Twenty-Four Rules Used by the Metabolic Forest, Their Associated Ruleset, and Their Type

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

For phase I metabolism, the stable oxygenation ruleset included the epoxidation, hydroxylation, nitrogen oxidation, and sulfur oxidation rules, the unstable oxygenation ruleset included the dealkylation and oxidation dehalogenation rules, the dehydrogenation ruleset included a lone dehydrogenation rule, the hydrolysis ruleset included the dephosphorylation, epoxide opening, carbonyl cleavage, and azo splitting rules, and the reduction ruleset included the benzodioxole reduction, dehydration, hydrogenation, nitrogen reduction, sulfur reduction, oxygen reduction, and reductive dehalogenation rules.

The conjugation ruleset included four rules specifying the reactions acetylation, glucuronidation, glutathionation, and sulfation.

The quinone formation ruleset included a single quinone formation rule that both modeled the two-electron oxidation that directly forms quinones and several types of reactions that often set the stage for that oxidation, such as aromatic hydroxylation.

Similarly, the tautomerization ruleset had a single eponymous rule. Tautomerization, although not generally regarded as a type of metabolism, nevertheless plays a role in known metabolic pathways of drugs like clopidogrel[88] and ranitidine.[89]

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

SMARTS Rules. Many of the reaction rules used the reaction SMARTS syntax provided by the open-source cheminformatics library RDKit (Table 2).[87] Reaction SMARTS syntax is derived from SMARTS patterns, and has similarities to the alternative reaction languages "SMIRKS"[90,91] and "SMILES".[92] Reaction SMARTS are a compact method for encoding transformations that occur on a small number of localized atoms. For example, the string: "[#6h:1]≫[*:1]O" expresses a hydroxylation reaction where the oxygen ("O") is connected with a single bond to a carbon with at least one hydrogen ("[#6h:1]"). The oxygen will be protonated during a sanitization step before outputting the final structure. The exact Reaction SMARTS used for each rule—developed by manual inspection and adjustment—are listed in Table 2.

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

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

Fundamentally, this implementation considers various resonance structures in turn. However, rather than naively

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

Table 2. Reaction SMARTS Used by the Metabolite Prediction Algorithm

rule	reaction SMARTS
acetylation	[#7,#8,#16;h:1]≫[*:1][#6](=[#8])[#6]
azo splitting	[#7:1]=[#7:2]≫[*:1].[*:2]
benzodioxole reduction	[#6R:1]-[#8R:2]-[#6H2R:3]-[#8R:4]-[#6R:5]≫([*:1]-[*:2].[*:3].[*:4]-[*:5])
dealkylation	[#6H3:1][#7,#8H0,#16:2]≫([*:2].[*:1](=O)O)
	[#6H3:1][#7,#8H0,#16:2]≫([*:2].[*:1]=O)
	[#6H3:1][#7,#8H0,#16:2]≫([*:2].[*:1]-O)
	[#6H2:1][#7,#8H0,#16:2]≫([*:2].[*:1](=O)O)
	[#6H2:1][#7,#8H0,#16:2]≫([*:2].[*:1]=O)
	[#6H2:1][#7,#8H0,#16:2]≫([*:2].[*:1]-O)
	[#6H1:1][#7,#8H0,#16:2]≫([*:2].[*:1]=O)
	[#6H1:1][#7,#8H0,#16:2]≫([*:2].[*:1]-O)
	[#6H0:1][#7,#8H0,#16:2]≫([*:2].[*:1]-O)
	[#6:1][#6:2]≫(O-[*:1].[*:2])
	[#6h:1][#6:2]≫(O-[*:1].[*:2])
	[#6h:1][#6:2]≫(O=[*:1].[*:2])
	[#8H1:3]-[#6:1]-[#7,#8,#16:2]≫([*:3]=[*:1].[*:2])
dehydration	[#6,#7:1]-[#8H1:2]≫[*:1].[*:2]
	[#6:3]-[#6:1]-[#8H1:2]≫[*:3]=[*:1].[*:2]
	[#6,#7:1]=[#8:2]≫[*:1].[*:2]
dehydrogenation	[#16v4:1]-[Oh:2]≫[*:1]=[*:2]
	[#6h:1]-[#6D1H3,#6D2H2,#6D3H1,#7D2H1,#7D1H2,#7D3,#8H1:2]≫[*:1]=[*:2]
dephosphorylation	[#8:1]=[#15:2]([#8:3])([#8:4])[#8:5][#6:6]≫[*:1]=[*:2]([*:3])([*:4])[*:5].O[*:6]
epoxidation	[#6:1]=[#6,#7:2]≫[*:1]1-[*:2][O]1
epoxide opening	[#6:1]1[#8:2][#6:3]1≫([*:2][*:3][*:1])
	[#6:1]1[#8:2][#6:3]1≫([*:2][*:3][*:1]O)
glucuronidation	[#6:1][#6:2](=[O,N,P,S:3])[#8:4]≫O1C(C(=O)O)C(O)C(O)C(O)C([*:4][*:2](=[#8:3])[*:1])1
	[#8H1:1][#6:2]≫O1C(C(=O)O)C(O)C(O)C(O)C([*:1][*:2])1
glutathionation	[#6:1]1[#8:2][#6:3]1≫C(CC(=O)N[C@@H](CS([*:1][*:3][*:2]))C(=O)NCC(=O)O)[C@@H](C(=O)O)N
	[#6:1][Cl:2]≫C(CC(=O)N[C@@H](CS([*:1]))C(=O)NCC(=O)O)[C@@H](C(=O)O)N
	[#16h1:1]≫C(CC(=O)N[C@@H](CS([*:1]))C(=O)NCC(=O)O)[C@@H](C(=O)O)N
hydrogenation	[#6:1]#[#6:2]≫[*:1]=[*:2]
	[#6:1]=[#6:2]≫[*:1]-[*:2]
hydrolysis	[#8,#16:1]=[#6:2]-[#7,#8,#16:3]≫([*:1]=[*:2](O).[*:3])
	[#8,#16:1]=[#6:2]-[#7,#8,#16:3]≫([*:1]=[*:2].[*:3])
hydroxylation	[#6h:1]≫[*:1]O
	[#6h2:1]≫[*:1]=O
nitrogen oxidation	[#7v3:1]≫[*:1]O
	[#7v3:1]≫[*:1]=O
nitrogen reduction	[#7:1](=[#8:2])-[#8:3]≫[*:1].[*:2].[*:3]
	[#7:1]([#8:2]) [#8:3]≫[*:1].[*:2].[*:3]
	[#8:2]=[#7:1]-[#8:3]≫[*:1].[*:2].[*:3]
	[#8:2]=[#7:1]-[#8-1:3]≫[*:1]=[*:2].[*:3]
	[#7D2:1]=[#8:2]≫[*:1]
	[#7:1]-[#8:2]≫([*:1].[*:2])
oxidative dehalogenation	[#9,#17,#35,#53,#85:1]-[#6:2]≫[*:1].[*:2]O
	[#9,#17,#35,#53,#85:1]-[#6h1:2]≫[*:1].[*:2]=O
	[#9,#17,#35,#53,#85:1]-[#6H2:2]≫[*:1].[*:2](O)=O
	[#9,#17,#35,#53,#85:1]-[#6:2][#6H1:3]≫[*:2](O)[*:3]-[*:1]
	[#9,#17,#35,#53,#85:1]-[#6:2]-[#9,#17,#35,#53,#85:3]≫[*:1].[*:2](O)=O.[*:3]
	[#9,#17,#35,#53,#85:1]-[#6:2]-[#9,#17,#35,#53,#85:3]≫[*:1].[*:2](O)O.[*:3]
oxygen reduction	[#8:1]=[#6,#7:2]≫[*:1]-[*:2]
	[#8:1]-[#8:2]≫[*:1].[*:2]
reductive dehalogenation	[#9,#17,#35,#53,#85:1]-[#6:2]≫[*:1].[*:2]
	[#9,#17,#35,#53,#85:1]-[#6:2]-[#6:3]≫[*:1].[*:2]=[*:3]
sulfation	[#6:1][#8:2]≫[*:1][*:2]S(=O)(=O)O
	[#6:1]1=[#6:2][#6:3]2[#8:7][#6:4]2[#6:5]=[#6:6]1≫[*:1]1=[*:2][*:3]=[*:4](-S(C)(=O)(=O))[*:5]=[*:6]1
sulfur oxidation	[#16;v2,v4:1]≫[*:1]O
	[#16;v2,v4:1]≫[*:1]=O
sulfur reduction	[#16:1]=[#8:2]≫[*:1].[*:2]
	[#16:1]-[#16:2]≫[*:1].[*:2]

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

Table 3. Transforms Used by the Resonance Pair Rules

rule	endpoint (transforms)	system type
dehydrogenation	[#6h:1][#6D1H3,#6D2H2,#6D3H1,#7D2H1,#7D1H2,#8H:2] (“single2double”) [#6h:1][#7D3:2] (“single2double”, “addPlus1”)	conjugated
hydrogenation	[*:1] [#6R:1][#6D1H3,#6D2H2,#6D3H1,#7D2H1,#7D1H2,#8H:2] (“single2double”) [#6D2H1:1] (“addO”)	conjugated
quinone formation	[#6H0R:1]-[F,Cl,Br,I:2] (“replaceHalogenWithO”, “single2double”) [#6H0R:1][#7D3:2] (“single2double”, “addPlus1”) [#6R:1][#7,#8:2][#6:3] (“dealk”, “single2double”)	aromatic

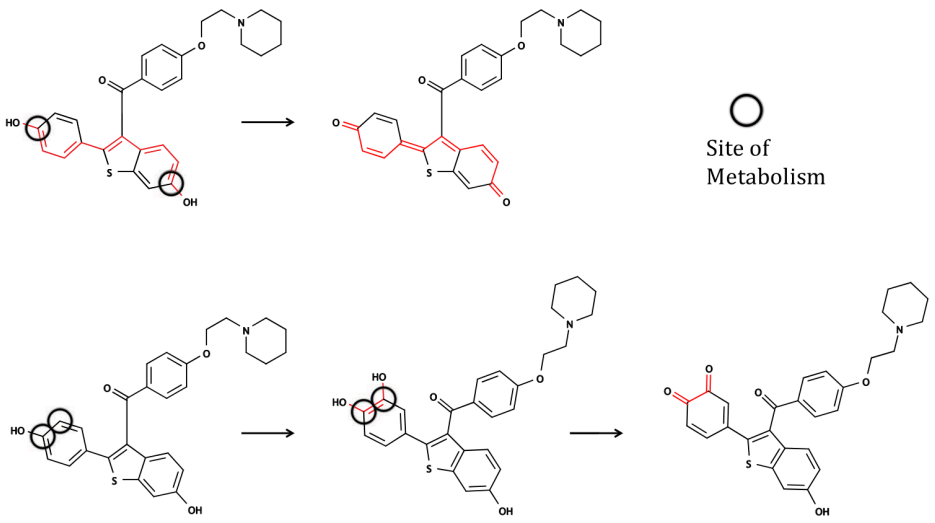


Figure 2. Raloxifene illustrates the quinone prediction method. Various resonance structures are considered in turn. For each structure, all paths between each pair of atoms are examined. Each atom pair is tested for compatibility with several predefined endpoint structures. A valid path is indicated by red bonds on the upper left. For this path, both endpoints are hydroxyls, in purple, which form a long-range quinone in a one-step, two-electron oxidation when the single and double bonds swap, as indicated by the arrows. Two-step quinone formations are also predicted by our method, such as that formed by hydroxylation followed by oxidation to form the lower right structure.

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

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

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

```
procedure ResonancePairAlgorithm(Molecule,Transforms,Systems)
    T ← all matches in Molecule to Transforms
    for S in Systems do
        for KekuleForm (K) in S do
            for PairedMatches (A1, A2) in both S and T do
                P ← path of alternating single/double bonds between A1 and A2, with start and end double bond
                M ← transform S in Molecule to match K
                M ← apply transform from A1 to M
                M ← apply transform from A2 to M
                M ← swap single and double bonds along P in M
                yield M
```

The exact specifications of the three resonance pair rules, dehydrogenation, hydrogenation, and quinone formation rules, are detailed in Table 3. Dehydrogenation and hydrogenation considered conjugated systems, as previously described. However, due to its mechanism, the quinone formation rule only needed to consider aromatic systems (Figure 2).

Resonance Structure Rules. The tautomerization and epoxidation rules also used resonance structures, but are distinct from the resonance pair rules because they do not perform combinations of modifications on pairs of atoms. The tautomerization rule works by iterating through resonance structures and enumerating all atoms one-bond away from each conjugated system (Algorithm 2). For each of these neighboring atoms, all paths from the neighbor through each atom within the conjugated system, where the number of

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

double and single bonds was the same, were considered. For each path, all double and single bonds were swapped to emit a tautomer. The tautomerization rule does not change the overall number of double bonds, single bonds, or hydrogens. The epoxidation rule used resonance structures and a pattern listed in Table 2 to replace double bonds between carbons with epoxides.

Algorithm 2 The tautomerization algorithm.

```
procedure TautomerizationRule(Molecule)
    for S in ConjugatedSystems do
        for KekuleForm (K) in S do
            for Neighbor (N) in Molecule not in S do
                for Atom (A) in S do
                    P ← path of alternating single/double bonds between A and N, with equal number of single and double bonds
                    M ← swap single and double bonds along P in M
                    yield M
```

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

RESULTS AND DISCUSSION

We explored several applications of the metabolic forest. First, we quantified performance at accurately reproducing metabolite structures and the corresponding pathways across a large, literature-derived dataset of phase I reactions. Second, using these pathways, we automatically labeled SOMs, thereby also fixing any mistakes in the original, manual labels. Third, we measured the intermediate metabolites generated when finding paths between substrates and metabolites. Fourth, we quantified how well the metabolite predictor produced quinone structures, a reaction type that is both especially challenging to represent and especially important for anticipating toxicity.

Accurate Predictions of Metabolite Structures. This study aimed to develop a system for accurately predicting metabolite structures. With this in mind, a critical metric was the percentage of AMD records for which the metabolite predictor generated the exact experimentally observed metabolite. Each AMD record includes the structures of both substrate and product. Using these structures, a useful method should find a series of transformations linking substrate and product, for any record that adheres to known metabolism patterns.

We performed several experiments to measure the ability of the metabolic forest to accurately predict the product structures of 20 736 AMD records (Table 4). First, we ran an annotated search, whereby we generated all metabolite structures for all of a molecule's manually labeled SOMs; which resulted in an overall performance of 78.36%. Second, we ran a depth one search (Algorithm 3), which improved over the known site search with an overall performance of 79.42%. Deepening the search to depths two or three depth significantly increased the overall performance to 88.43 and 88.77%, respectively. Previous experience with the AMD suggested that many reactions reported in a single record may in fact be two or more metabolic steps, and this is confirmed by the superior performance of the two and three depth search compared to the one depth search.

Table 4. Performance of Several Methods at Reproducing the Structure(s) Linking the Substrates and Products of 20 736 AMD Records[a]

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

[a]For each reaction type, the highest performance is in bold. Any score not statistically different from the best performance is in italics (using a *P*-value cutoff of 0.05).

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

```
procedure BFSMetabolite(Substrate,Metabolite,RuleSet,Depth)
    Intermediates(I) = []                                   ▷ Empty list
    for Rule(R) in RuleSet do
        for Transforms in R do
            T ← all matches in Substrate to Transforms
            for Site(S) in T do
                M ← apply T at S on Substrate
                if M == Metabolite then return M
                else
                    I ← append M to I
    if Depth >= 1 then
        for M in I do
            BFSMetabolite(M,Metabolite,RuleSet,Depth − 1)
```

For a given input molecule, the total number of structures generated in multiple steps combinatorially explodes. Con-

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

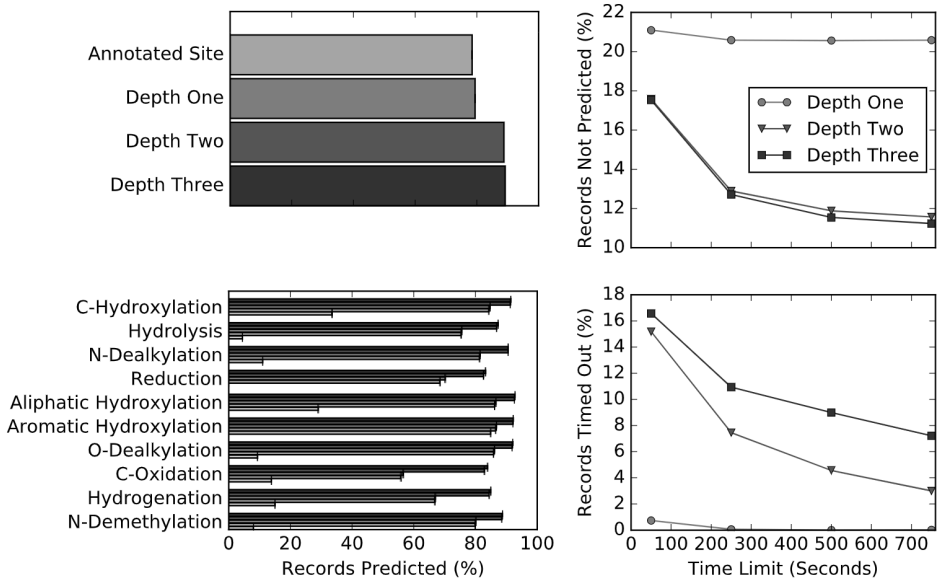


Figure 3. Metabolic forest took as input the reactant and product structures of AMD records. For each of the 20 736 records, the algorithm performed a breadth-first search to link the reactant and product using a collection of metabolic rules. These searches continued until a pathway to the exact product structure is found, the depth-limited was reached, or a maximum time limit was exceeded. Top left: the overall percentage of records for which a valid pathway was found, using a time limit of 1000 s. In the annotated site comparison experiment, a depth one search was restricted to using sites of metabolism matching those that were manually annotated during construction of the dataset. Bottom left: performance is broken down for the 10 most frequent reaction types of the dataset, as recorded in the AMD. Detailed results are reported in Table 4. Top right: percent error as a function of the maximum time limit allowed per record. The depth one search requires little running time to exhaust all possibilities. In contrast, performances of the depth two and depth three search are improved by allowing for a longer search. Bottom right: the corresponding number of records that timed out for each depth and time limit.

sequently, for each breadth-first search, we measured performance at five time cutoffs of 50, 125, 250, 500, and 750 s/reaction record. For searches of depth one, the time cutoff had minimal effect on performance. This is as expected because all depth one possibilities were explored within the allotted time window. In contrast, searches of greater depth often required a longer run time to find a valid solution (Figure 3).

Examples of records for which a path was successfully elucidated are shown in Figure 4. For example, for zotepine, it was straightforward to find a one-step sulfur oxidation linking substrate and its metabolite. Also highlighted in Figure 4 are examples of records for which a valid path was not found. For example, for a capsaicin metabolite, glutathione attaches to an unsubstituted carbon on a phenyl ring and a hydroxyl is simultaneously added at a different carbon meta to the glutathionation site. Some steps seem to be missing, as the simultaneous glutathionation and meta hydroxylation is not a known metabolic pathway. Although we might hope to capture this metabolite with a deeper or faster search by combining metabolism rules, encoding a rule for such an idiosyncratic case would suggest mere memorization of the data, rather than generation of metabolically logical structures.

Inferring Intermediate Metabolite Structures. Many AMD records have missing intermediates (Figure 5). Such records are often immediately obvious during manual inspection, for they often entail multiple transformations at different parts of the molecule, such as two hydroxylations. Another frequent variety is sequential transformations at a single site being conflated into a single reaction. For example, some halogenated aromatic rings are reported to immediately form quinones, without reporting the intermediate dehalogenation event that needs to take place before the two-electron oxidation to a quinone. As previously discussed, we implemented a depth-first search algorithm to infer intermediate metabolite structures. To quantify the number of missing metabolites discovered, we subtract the number of metabolites found by the depth one search from the number of metabolites found by the depth two search. This resulted in 1868 metabolites, which can be interpreted as missing metabolites in AMD records.

Automatic Detection of Mislabeled Sites. Comparing the annotated site search to the depth one search revealed cases where SOMs were mislabeled (Figure 6). Across 20 736 phase I reaction records, the depth one search found a reaction connecting reactant and product in 79.42% of cases. In contrast, an annotated site search that limited rules to generating structures by modifying at least one of the manually annotated SOMs only had a performance of 78.36%. The comparable performance of 79.42 and 78.36% implies that the metabolic forest produces SOMs that are of similar validity as those produced by manual labor. Furthermore, mistakes in the human annotations are revealed by examining cases where the depth one search found a solution, but the annotated-site-limited search failed.

Accurate Generation of Quinone Structures. Accurate quinone structure generation is an especially important task for a metabolite structure predictor. Quinone species, including quinone-imines, quinone-methides, and imine-methides, represent over 40% of reactive metabolites.[76] Many drugs are vulnerable to quinone formation due to the ubiquity of the phenyl ring in drug design. Due to their abundance, quinone formation in drug metabolism has been extensively studied experimentally.[105−109] Recently, we published the first study that explicitly predicted quinone formation.[45] However, that

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

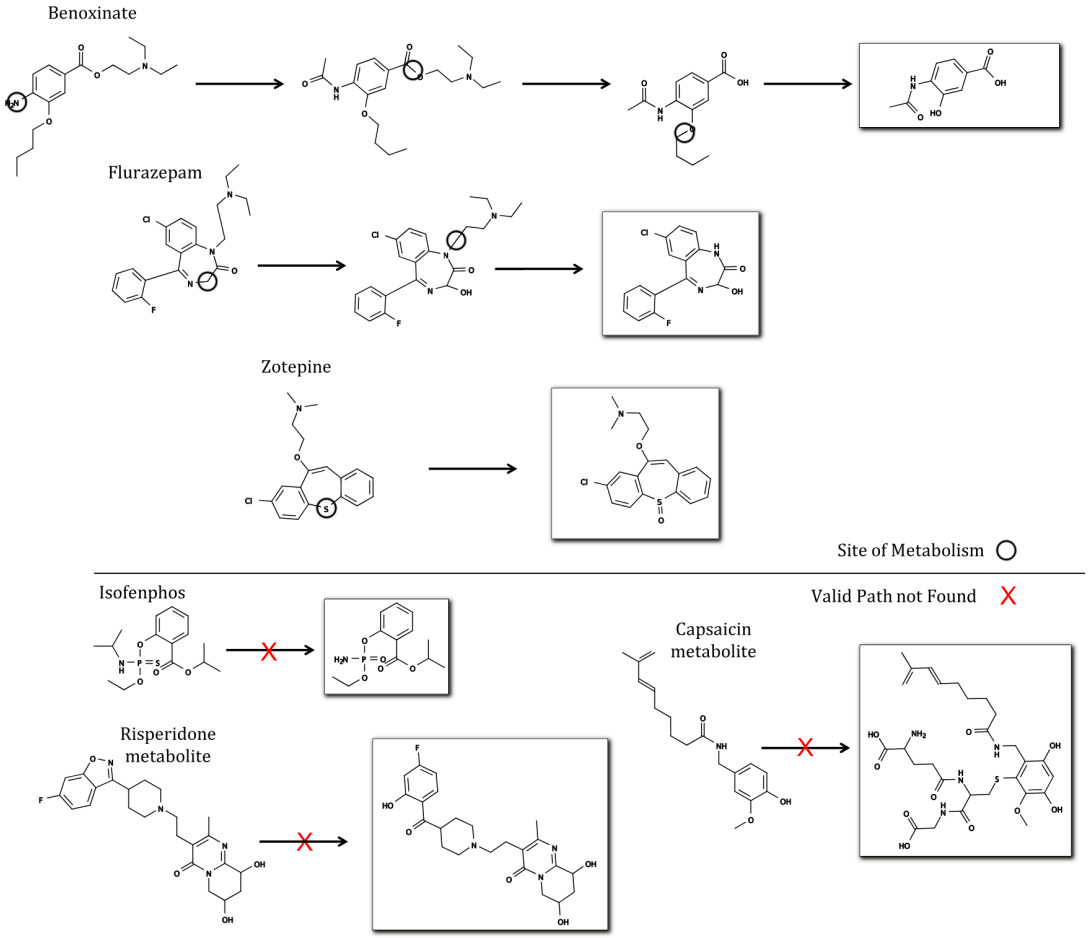


Figure 4. Examples of success and failure reaction records from metabolism literature. Top: transformation pathways successfully found by the metabolic forest, including a three-step reaction pathway of benoxinate,[93] a two-step reaction pathway of flurazepam,[94] and a one-step pathway of zotepine.[95] Bottom: records where the metabolic forest failed to find a pathway between substrate and product, including a simultaneous dealkylation and desulfuration of isofenphos,[96] a ring opening of a risperidone metabolite,[97] and a simultaneous glutathionation and hydroxylation of a capsaicin metabolite.[98] Experimentally observed metabolites are enclosed within boxes. Inferred sites of metabolism are circled. Records for which no pathway was found linking substrate and metabolite are marked with red crosses.

model had the limitation—common to most previous metabolism models—of only making predictions on the input molecule, and not producing actual metabolite structures.

For the present study, we designed a specific rule for quinone structure prediction. To evaluate performance at predicting quinone structures, we used the exact dataset from our previous quinone study.[45] Quinone formation reactions are challenging to encode because they can involve a variety of substituents and by definition entail a loss of aromaticity. Nevertheless, we accurately modeled quinone formation using a combination of SMARTS reactions rules and more fine-coded chemical programming (Algorithm 1). To our knowledge, this is the first published algorithm for predicting the structure of quinones.

The algorithm finds quinone not only in simple cases but also across aromatic systems spanning multiple rings (Figure 2). Across the 576 quinone formations reactions, the metabolic forest finds a formation pathway 91.84, 91.84, and 76.22% of the time, for depth three, two, and one searches, respectively (Figure 7).

Limitations. The metabolic forest accurately produced the metabolite structures across a large, literature-derived dataset of 20 736 phase I reaction records. Although diverse, this data may have biases that limit our results. For example, many short-lived intermediates are difficult or impossible to experimentally detect. As a result, such transitory molecules are likely underreported in metabolic studies. Due to our focus on accurately reproducing literature-derived data, our current tool may not be as well tuned to these underreported intermediates.

Secondly, the search algorithm naively considers all possible combinations of rules, without regard to biological patterns that make some combinations more likely than others. For example, the current method blindly mixes phase I and phase II reactions, ignoring the well-known paradigm of redox reactions often introducing functional groups for subsequent conjugation reactions.[76] Additionally, the search does not take into account the connectivity distance between the sites modified by subsequent metabolic steps. It would be logical to start the search at each depth near the site most recently modified, to reduce the number of false paths considered for reactions that are highly correlated, such as hydroxylation followed by quinone formation.[15]

A third possible caveat is the risk of incorrect metabolite structures in the data misinforming the design of the metabolic

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

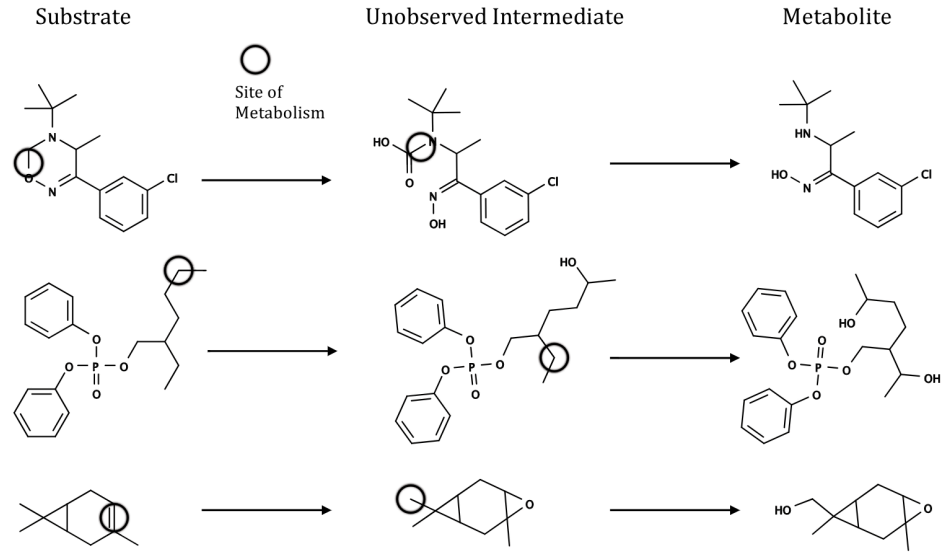


Figure 5. Missing intermediate structures can be predicted from those records for which a depth two or depth three search found a metabolic pathway not found at depth one. Three examples are visualized. Each substrate−metabolite pair could only be linked by searching a depth greater than 1. Such hypothesized intermediates can be short-lived and consequently difficult to observe experimentally, yet may have important biological consequences. Top row: a candidate prodrug of the antidepressant bupropion.[99] Middle row: the plasticizer additive and flame retardant 2-ethylhexyl diphenyl phosphate.[100] Bottom row: Δ^3-carene, a natural monoterpene.[101]

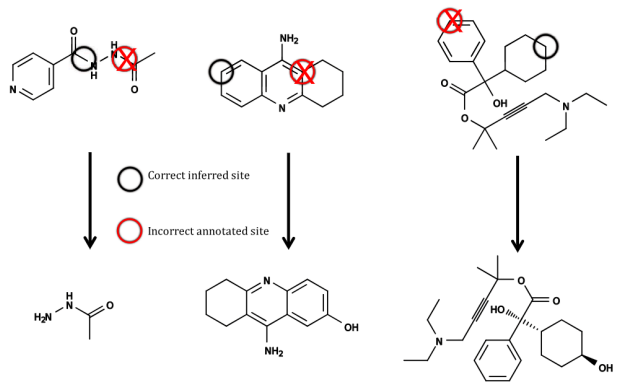


Figure 6. Across 20 736 phase I reactions, the depth one search found a reaction connecting reactant and product in 79.42% of cases, whereas the manually annotated sites only had a performance of 78.36%. By examining reactions solved by the depth one search but not the annotated sites, labeling mistakes were revealed. Three examples are shown, where the incorrect, manually annotated sites are circled in red and crossed, and the correct sites uncovered by the depth one search are circled in black. From left to right: acetylisoniazid,[102] tacrine,[103] and (NS-21), a former drug candidate for bladder disorders.[104]

rules. The potential impact of this possibility on our reported accuracies seems low because the large dataset size of 20 736 records made the effect of any single record negligible. Furthermore, when constructing the metabolic forest, only a small fraction of the data was inspected. It seems more likely that incorrect or unorthodox metabolite structures resulted in a performance underestimation. For example, one poorly performing reaction type was glutathionation reaction because glutathione was reported as several slightly different forms in the data and therefore did match any product structure. Explicitly encoding all of these glutathione forms would be undesirable because it would unnecessarily expand the total number of possible structures at each depth and place undue emphasis on a single reaction type.

Case Study. Building a bioactivation model by linking models of metabolism[47,63,77] and reactivity[45,116] was a primary motivation for constructing an accurate method of metabolite structure generation. The full construction and analysis of this unified bioactivation model is beyond the scope of this study. Nevertheless, to demonstrate the value of the metabolic forest, we built a prototype implementation. In this system, we used the explicit metabolite structures generated by the metabolic forest to unite three, heretofore incompatible, models: the previously discussed quinone formation model,[45] a model of epoxidation,[63] and a model of reactivity.[77]

The quinone formation and epoxidation models were originally conceived as specialized metabolism studies, with their respective focuses chosen because quinones and epoxides together represent around 50% of all known reactive metabolites.[76,117] However, classifying quinones and epoxides as either reactive or nonreactive is a generalization because reactivity is actually a continuum. For example, there are naturally occurring epoxides[118] and other small-molecule-containing epoxides that are known to be nonreactive[119] In these cases, the presence of electron-donating groups on the carbons in an epoxide often stabilized the motif, by reducing the electron deficiency. Similarly, nearly identical quinones can vary widely in their reactivity and toxicity.[120] Consequently, evaluating the reactivity of possible quinones and epoxides enables a ranking of possible structures that may drive toxicity.

The quinone model predicts how likely quinones are to form at both the site and molecule levels. Similarly, the epoxidation model yields probabilistic scores for epoxide formation, also at the site- and molecule level. The reactivity model predicts reactivity to biological macromolecules, including DNA and protein, as well as glutathione and cyanide, which are frequently used experimentally to detect reactive molecules.[38,39] To build our bioactivation model, we constructed a pipeline where, for a given input molecule, all of the

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

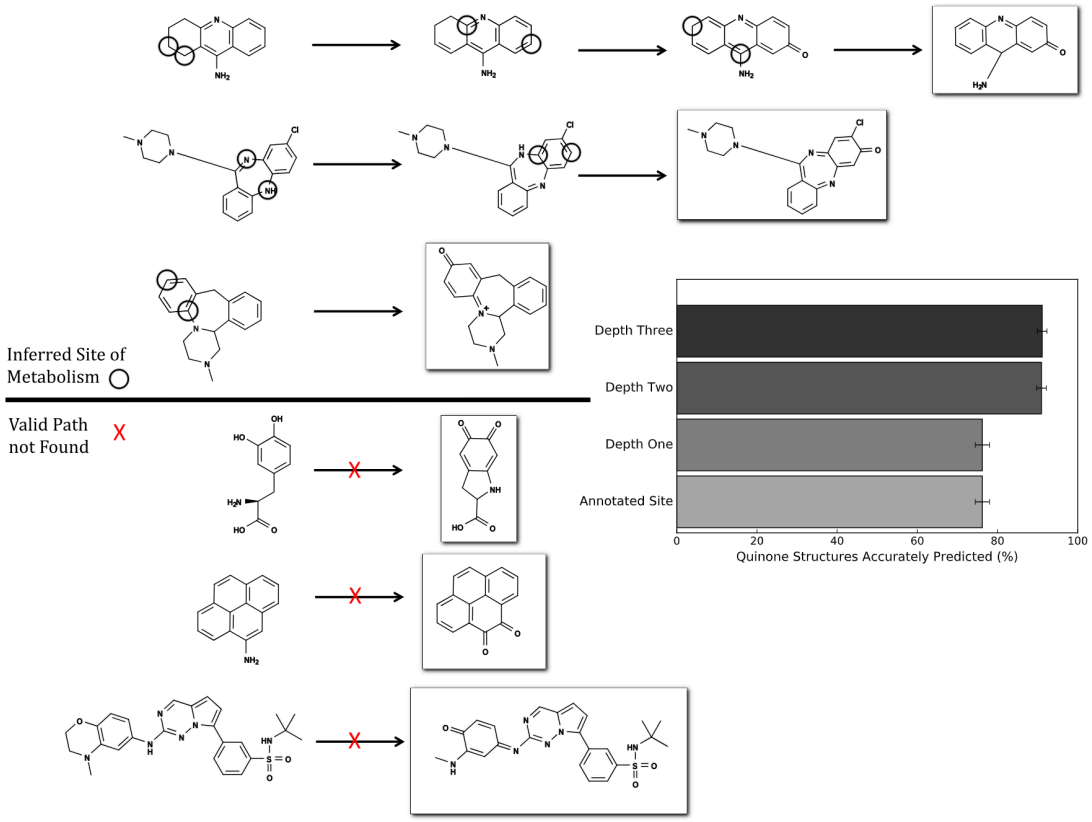


Figure 7. Metabolic forest accurately produced quinone structures. Across a dataset of 576 quinone formations,[45] the metabolic forest found a formation pathway 91.84, 91.84, and 76.22% of the time, for depth three, two, and one searches, respectively. Bottom half: examples of experimentally observed quinones successfully reproduced by depth one, two, and three searches. From top to bottom: the starting drugs are tacrine, a cholinesterase inhibitor;[110,111] clozapine, an atypical antipsychotic;[39] and mianserin, an antidepressant.[112] Third to bottom: dihydroxyphenylalanine.[113] Second to bottom row: a quinone is formed from 4-nitropyrene, an urban air pollutant from diesel engines.[16,114] Bottom: a kinase inhibitor drug candidate.[115]

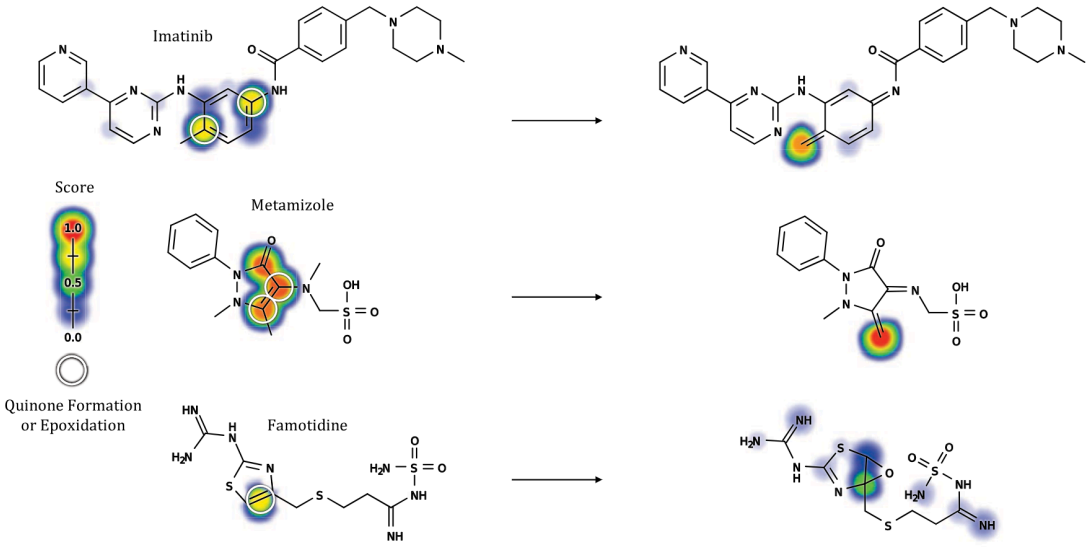


Figure 8. Previously developed models of metabolism and reactivity are connected by the metabolic forest to generated bioactivation hypotheses for several IADR-associated drugs, whose IADR drivers were never elucidated. Top: the chemotherapy drug imatinib has been implicated in several cases of idiosyncratic, severe hepatotoxicity.[4,121] Middle: the analgesic metamizole was withdrawn from the market in several countries due to cases of agranulocytosis.[123−127] Bottom: famotidine, a histamine H_2 receptor antagonist, has been associated with idiosyncratic reactions, including toxic epidermal necrolysis.[130] The colored shading represents the quinone formation score[45] for imatinib and metamizole, the epoxidation score[63] for famotidine, and the reactivity scores[77] of all three predicted metabolites.

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

structures of all possible epoxides and quinones were generated, and these structures submitted to the reactivity model. By multiplying the probabilistic metabolism scores from the epoxidation and quinone models by the reactivity scores, metabolites can be identified that are both likely to form and likely to be reactive. In the future, we plan to quantitatively use these bioactivation scores to predict reactive metabolite formation, and to systematically evaluate its performance at this task. Here, we highlight a few examples where the metabolite predictor has generated hypotheses about reactive metabolites that may be responsible for IADRs (Figure 8).

The chemotherapy drug imatinib has been implicated in several cases of idiosyncratic, severe hepatotoxicity.[4,121] However, no reactive metabolite has been detected.[122] Our models predict formation of a highly reactive quinone-methide. Our quinone formation model[45] predicted formation of a quinone-imide that is also predicted to be highly reactive by our reactivity model.[77] The potent nonopioid analgesic metamizole was withdrawn from the market in several countries, including the United States, due to cases of agranulocytosis.[123−127] The mechanism of this toxicity has not been elucidated, although evidence has been found for an immunoallergic origin, suggestive of a bioactivation mechanism.[128] A cation radical has been observed in incubations with myeloperoxidase,[129] but without evidence of its formation *in vivo* the culprit of metamizole remains an open question. Our bioactivation model presents an alternative hypothesis: the formation of a reactive quinone on metamizole's pyrazolone motif. Famotidine, a histamine H_2 receptor antagonist, has been associated with unexplained idiosyncratic reactions, including toxic epidermal necrolysis.[130] A reactive epoxide forming on the thiazole within famotidine is a possible explanation suggested by epoxidation and reactivity scores.

CONCLUSIONS

This study established a validated, accurate tool for predicting metabolic structures across one or several metabolic steps. Our method combined simple rules encoded by reaction SMARTS with novel algorithms for complex, resonance-structure-based transformations, including quinone formation, hydrogenation, dehydrogenation, and tautomerization. We validated the metabolite structure predictor on a diverse collection of 20 736 records from a literature-derived database. Beginning with the substrate of each record, a breadth-first search successfully found a transformation resulting in the exact experimentally observed product 79.42, 88.43, and 88.77% of the time when generating a metabolite tree limited to one, two, or three successive rules, respectively. We also validated our method on a previously published dataset of 576 quinone formations,[45] producing the correct quinone structure with accuracies of 76.22, 91.84, and 91.84% with searches of depths one, two, and three, respectively. Our tool can also be used to infer missing intermediate structures, and to automatically label sites of metabolism. Most importantly, the metabolite predictor enables integration of metabolism and reactivity models to construct a bioactivation model. Until now, metabolite structures were the missing link in that endeavor. We constructed a prototype model that predicted novel putative reactive metabolites for the drugs imatinib, metamizole, and famotidine, each of which is associated with IADRs, the causes of which have never been elucidated. These reactive metabolites are specific, testable hypotheses about the mechanism of their parent drug's idiosyncratic toxicity. We plan to comprehensively model bioactivation using the metabolite predictor and expect that accurate enumeration of possible metabolite structures will become a cornerstone of many other future investigations.

AUTHOR INFORMATION

Corresponding Author

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

Authors

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

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

Ayush Kumar − *Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, United States*

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

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.0c00360

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the developers of the open-source chemoinformatics tools Open Babel[131] and RDKit. Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under award numbers R01LM012222 and R01LM012482. Computations were performed using the facilities of the Washington University Center for High Performance Computing, which were partially funded by National Institutes of Health grant numbers 1S10RR022984-01A1 and 1S10OD018091-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors also thank both the Department of Immunology and Pathology at the Washington University School of Medicine and the Washington University Center for Biological Systems Engineering for their generous support of this work.

ABBREVIATIONS

ADR, adverse drug reaction; AMD, Accelrys Metabolite Database; IADR, idiosyncratic adverse drug reaction; SOM, site of metabolism

REFERENCES

(1) Roberts, R. A.; Kavanagh, S. L.; Mellor, H. R.; Pollard, C. E.; Robinson, S.; Platz, S. J. Reducing Attrition in Drug Development: Smart Loading Preclinical Safety Assessment. *Drug Discovery Today* **2014**, *19*, 341−347.

(2) Kola, I.; Landis, J. Can the Pharmaceutical Industry Reduce Attrition Rates. *Nat. Rev. Drug Discovery* **2004**, *3*, 711−716.

(3) Arrowsmith, J.; Miller, P. Phase II and Phase III Attrition Rates 2011-2012. *Nat. Rev. Drug Discovery* **2013**, *12*, 569.
(4) Stepan, A. F.; Walker, D. P.; Bauman, J.; Price, D. A.; Baillie, T. A.; Kalgutkar, A. S.; Aleo, M. D. Structural Alert/Reactive Metabolite Concept as Applied in Medicinal Chemistry to Mitigate the Risk of Idiosyncratic Drug Toxicity: A Perspective Based on the Critical Examination of Trends in the Top 200 Drugs Marketed in the United States. *Chem. Res. Toxicol.* **2011**, *24*, 1345−1410.
(5) Uetrecht, J.; Naisbitt, D. J. Idiosyncratic Adverse Drug Reactions: Current Concepts. *Pharmacol. Rev.* **2013**, *65*, 779−808.
(6) Almario, E. E.; Borlak, J.; Suzuki, A.; Chen, M. Drug-Induced Liver Injury. *BioMed Res. Int.* **2017**, *2017*, No. 2461694.
(7) Guengerich, F. P. Mechanisms of Drug Toxicity and Relevance to Pharmaceutical Development. *Drug Metab. Pharmacokinet.* **2011**, *26*, 3−14.
(8) Kaplowitz, N. Avoiding Idiosyncratic DILI: Two Is Better Than One. *Hepatology* **2013**, *58*, 15−17.
(9) Deng, X.; Luyendyk, J. P.; Ganey, P. E.; Roth, R. A. Inflammatory Stress and Idiosyncratic Hepatotoxicity: Hints From Animal Models. *Pharmacol. Rev.* **2009**, *61*, 262−282.
(10) Smith, G. F. *Progress in Medicinal Chemistry*; Elsevier, 2011; Vol. *50*, pp 1−47.
(11) Adams, D. H.; Ju, C.; Ramaiah, S. K.; Uetrecht, J.; Jaeschke, H. Mechanisms of Immune-Mediated Liver Injury. *Toxicol. Sci.* **2010**, *115*, 307−321.
(12) Williams, D.; Kitteringham, N.; Naisbitt, D.; Pirmohamed, M.; Smith, D.; Park, B. Are Chemically Reactive Metabolites Responsible for Adverse Reactions to Drugs. *Curr. Drug Metab.* **2002**, *3*, 351−366.
(13) Pichler, W. J.; Naisbitt, D. J.; Park, B. K. Immune Pathomechanism of Drug Hypersensitivity Reactions. *J. Allergy Clin. Immunol.* **2011**, *127*, S74−S81.
(14) Shenton, J. M.; Chen, J.; Uetrecht, J. P. Animal Models of Idiosyncratic Drug Reactions. *Chem.-Biol. Interact.* **2004**, *150*, 53−70.
(15) Kalgutkar, A. S.; Gardner, I.; Obach, R. S.; Shaffer, C. L.; Callegari, E.; Henne, K. R.; Mutlib, A. E.; Dalvie, D. K.; Lee, J. S.; Nakai, Y.; O'Donnell, J. P.; Boer, J.; Harriman, S. P. A Comprehensive Listing of Bioactivation Pathways of Organic Functional Groups. *Curr. Drug Metab.* **2005**, *6*, 161−225.
(16) Gallagher, J.; Heinrich, U.; George, M.; Hendee, L.; Phillips, D.; Lewtas, J. Formation of DNA Adducts in Rat Lung Following Chronic Inhalation of Diesel Emissions, Carbon Black and Titanium Dioxide Particles. *Carcinogenesis* **1994**, *15*, 1291−1299.
(17) Ishii, Y.; Inoue, K.; Takasu, S.; Jin, M.; Matsushita, K.; Kuroda, K.; Fukuhara, K.; Nishikawa, A.; Umemura, T. Determination of Lucidin-Specific DNA Adducts by Liquid Chromatography With Tandem Mass Spectrometry in the Livers and Kidneys of Rats Given Lucidin-3-O-Primeveroside. *Chem. Res. Toxicol.* **2012**, *25*, 1112−1118.
(18) van der Woude, H.; Alink, G. M.; van Rossum, B. E.; Walle, K.; van Steeg, H.; Walle, T.; Rietjens, I. M. Formation of Transient Covalent Protein and DNA Adducts by Quercetin in Cells With and Without Oxidative Enzyme Activity. *Chem. Res. Toxicol.* **2005**, *18*, 1907−1916.
(19) Lillibridge, J.; Amore, B.; Slattery, J.; Kalhorn, T.; Nelson, S.; Finnell, R.; Bennett, G. Protein-Reactive Metabolites of Carbamazepine in Mouse Liver Microsomes. *Drug Metab. Dispos.* **1996**, *24*, 509−514.
(20) Takakusa, H.; Masumoto, H.; Yukinaga, H.; Makino, C.; Nakayama, S.; Okazaki, O.; Sudo, K. Covalent Binding and Tissue Distribution/Retention Assessment of Drugs Associated With Idiosyncratic Drug Toxicity. *Drug Metab. Dispos.* **2008**, *36*, 1770−1779.
(21) Evans, D. C.; Watt, A. P.; Nicoll-Griffith, D. A.; Baillie, T. A. Drug-Protein Adducts: An Industry Perspective on Minimizing the Potential for Drug Bioactivation in Drug Discovery and Development. *Chem. Res. Toxicol.* **2004**, *17*, 3−16.
(22) DeMarini, D.; Brooks, L.; Bhatnagar, V.; Hayes, R.; Eischen, B.; Shelton, M.; Zenser, T.; Talaska, G.; Kashyap, S.; Dosemeci, M.; Kashyap, R.; Parikh, D.; Lakshmi, V.; Hsu, F.; Davis, B.; Jaeger, M.; Rothman, N. Urinary Mutagenicity as a Biomarker in Workers Exposed to Benzidine: Correlation With Urinary Metabolites and Urothelial DNA Adducts. *Carcinogenesis* **1997**, *18*, 981−988.
(23) Purohit, V.; Basu, A. K. Mutagenicity of Nitroaromatic Compounds. *Chem. Res. Toxicol.* **2000**, *13*, 673−692.
(24) Ikehata, K.; Duzhak, T. G.; Galeva, N. A.; Ji, T.; Koen, Y. M.; Hanzlik, R. P. Protein Targets of Reactive Metabolites of Thiobenzamide in Rat Liver in Vivo. *Chem. Res. Toxicol.* **2008**, *21*, 1432−1442.
(25) Dennehy, M. K.; Richards, K. A.; Wernke, G. R.; Shyr, Y.; Liebler, D. C. Cytosolic and Nuclear Protein Targets of Thiol-Reactive Electrophiles. *Chem. Res. Toxicol.* **2006**, *19*, 20−29.
(26) Stachulski, A. V.; Baillie, T. A.; Kevin Park, B.; Scott Obach, R.; Dalvie, D. K.; Williams, D. P.; Srivastava, A.; Regan, S. L.; Antoine, D. J.; Goldring, C. E. P.; Chia, A. J. L.; Kitteringham, N. R.; Randle, L. E.; Callan, H.; Castrejon, J. L.; Farrell, J.; Naisbitt, D. J.; Lennard, M. S. The Generation, Detection, and Effects of Reactive Drug Metabolites. *Med. Res. Rev.* **2013**, *33*, 985−1080.
(27) Srivastava, A.; Maggs, J.; Antoine, D.; Williams, D.; Smith, D.; Park, B. *Adverse Drug Reactions*; Springer: Berlin, 2010; pp 165−194.
(28) Liu, Z.-X.; Kaplowitz, N. Immune-Mediated Drug-Induced Liver Disease. *Clin. Liver Dis.* **2002**, *6*, 755−774.
(29) Naisbitt, D. J.; Williams, D. P.; Pirmohamed, M.; Kitteringham, N. R.; Park, B. K. Reactive Metabolites and Their Role in Drug Reactions. *Curr. Opin. Allergy Clin. Immunol.* **2001**, *1*, 317−325.
(30) Alempijevic, T.; Zec, S.; Milosavljevic, T. Drug-Induced Liver Injury: do We Know Everything. *World J. Hepatol.* **2017**, *9*, No. 491.
(31) Brink, A.; Pähler, A.; Funk, C.; Schuler, F.; Schadt, S. Minimizing the Risk of Chemically Reactive Metabolite Formation of New Drug Candidates: Implications for Preclinical Drug Design. *Drug Discovery Today* **2017**, *22*, 751−756.
(32) Mazerbourg, S.; Kuntz, S.; Grillier-Vuissoz, I.; Berthe, A.; Geoffroy, M.; Flament, S.; Bordessa, A.; Boisbrun, M. Reprofiling of Troglitazone Towards More Active and Less Toxic Derivatives: A New Hope for Cancer Treatment. *Curr. Top. Med. Chem.* **2016**, *16*, 2115−2124.
(33) Uetrecht, J. Screening for the Potential of a Drug Candidate to Cause Idiosyncratic Drug Reactions. *Drug Discovery Today* **2003**, *8*, 832−837.
(34) Park, B. K.; Boobis, A.; Clarke, S.; Goldring, C. E. P.; Jones, D.; Kenna, J. G.; Lambert, C.; Laverty, H. G.; Naisbitt, D. J.; Nelson, S.; Nicoll-Griffith, D. A.; Obach, R. S.; Routledge, P.; Smith, D. A.; Tweedie, D. J.; Vermeulen, N.; Williams, D. P.; Wilson, I. D.; Baillie, T. A. Managing the Challenge of Chemically Reactive Metabolites in Drug Development. *Nat. Rev. Drug Discovery* **2011**, *10*, 292−306.
(35) Zhu, M.; Ma, L.; Zhang, H.; Humphreys, W. G. Detection and Structural Characterization of Glutathione-Trapped Reactive Metabolites Using Liquid Chromatography-High-Resolution Mass Spectrometry and Mass Defect Filtering. *Anal. Chem.* **2007**, *79*, 8333−8341.
(36) Argoti, D.; Liang, L.; Conteh, A.; Chen, L.; Bershas, D.; Yu, C.-P.; Vouros, P.; Yang, E. Cyanide Trapping of Iminium Ion Reactive Intermediates Followed by Detection and Structure Identification Using Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS). *Chem. Res. Toxicol.* **2005**, *18*, 1537−1544.
(37) Attia, S. M. Deleterious Effects of Reactive Metabolites. *Oxid. Med. Cell. Longevity* **2010**, *3*, 238−253.
(38) Ma, S.; Subramanian, R. Detecting and Characterizing Reactive Metabolites by Liquid Chromatography/Tandem Mass Spectrometry. *J. Mass Spectrom.* **2006**, *41*, 1121−1139.
(39) Meneses-Lorente, G.; Sakatis, M. Z.; Schulz-Utermoehl, T.; De Nardi, C.; Watt, A. P. A Quantitative High-Throughput Trapping Assay as a Measurement of Potential for Bioactivation. *Anal. Biochem.* **2006**, *351*, 266−272.
(40) Day, S.; Mao, A.; White, R.; Schulz-Utermoehl, T.; Miller, R.; Beconi, M. A Semi-Automated Method for Measuring the Potential for Protein Covalent Binding in Drug Discovery. *J. Pharmacol. Toxicol. Methods* **2005**, *52*, 278−285.

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

(41) Kalgutkar, A. Should the Incorporation of Structural Alerts Be Restricted in Drug Design? An Analysis of Structure-Toxicity Trends With Aniline-Based Drugs. *Curr. Med. Chem.* **2014**, *22*, 438−464.
(42) Kalgutkar, A. S.; Didiuk, M. T. Structural Alerts, Reactive Metabolites, and Protein Covalent Binding: How Reliable Are These Attributes as Predictors of Drug Toxicity. *Chem. Biodiversity* **2009**, *6*, 2115−2137.
(43) Sushko, I.; Salmina, E.; Potemkin, V. A.; Poda, G.; Tetko, I. V. ToxAlerts: AWeb Server of Structural Alerts for Toxic Chemicals and Compounds With Potential Adverse Reactions. *J. Chem. Inf. Model.* **2012**, *52*, 2310−2316.
(44) Edwards, P. J.; Sturino, C. Managing the Liabilities Arising From Structural Alerts: A Safe Philosophy for Medicinal Chemists. *Curr. Med. Chem.* **2011**, *18*, 3116−3135.
(45) Hughes, T. B.; Swamidass, S. J. Deep Learning to Predict the Formation of Quinone Species in Drug Metabolism. *Chem. Res. Toxicol.* **2017**, *30*, 642−656.
(46) Dang, N. L.; Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Computational Approach to Structural Alerts: Furans, Phenols, Nitroaromatics, and Thiophenes. *Chem. Res. Toxicol.* **2017**, *30*, 1046−1059.
(47) Dang, N. L.; Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Computationally Assessing the Bioactivation of Drugs by N-Dealkylation. *Chem. Res. Toxicol.* **2018**, *31*, 68−80.
(48) Zaretzki, J.; Browning, M. R.; Hughes, T. B.; Swamidass, S. J. Extending P450 Site-Of-Metabolism Models With Region-Resolution Data. *Bioinformatics* **2015**, *31*, 1966−1973.
(49) Zaretzki, J.; Bergeron, C.; Rydberg, P.; Huang, T.-w.; Bennett, K. P.; Breneman, C. M. RS-Predictor: A New Tool for Predicting Sites of Cytochrome P450-Mediated Metabolism Applied to CYP 3A4. *J. Chem. Inf. Model.* **2011**, *51*, 1667−1689.
(50) Zaretzki, J.; Rydberg, P.; Bergeron, C.; Bennett, K. P.; Olsen, L.; Breneman, C. M. RS-Predictor Models Augmented With SMARTCyp Reactivities: Robust Metabolic Regioselectivity Predictions for Nine CYP Isozymes. *J. Chem. Inf. Model.* **2012**, *52*, 1637−1659.
(51) Zaretzki, J.; Bergeron, C.; Huang, T.-w.; Rydberg, P.; Swamidass, S. J.; Breneman, C. M. RS-WebPredictor: A Server for Predicting CYP-mediated Sites of Metabolism on Drug-Like Molecules. *Bioinformatics* **2013**, *29*, 497−498.
(52) Rydberg, P.; Gloriam, D. E.; Zaretzki, J.; Breneman, C.; Olsen, L. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. *ACS Med. Chem. Lett.* **2010**, *1*, 96−100.
(53) Rydberg, P.; Gloriam, D. E.; Olsen, L. The SMARTCyp Cytochrome P450 Metabolism Prediction Server. *Bioinformatics* **2010**, *26*, 2988−2989.
(54) Rydberg, P.; Olsen, L. Predicting Drug Metabolism by Cytochrome P450 2C9: Comparison With the 2D6 and 3A4 Isoforms. *ChemMedChem* **2012**, *7*, 1202−1209.
(55) Zaretzki, J.; Matlock, M.; Swamidass, S. J. XenoSite: Accurately Predicting CYP-mediated Sites of Metabolism With Neural Networks. *J. Chem. Inf. Model.* **2013**, *53*, 3373−3383.
(56) Matlock, M. K.; Hughes, T. B.; Swamidass, S. J. XenoSite Server: A Web-Available Site of Metabolism Prediction Tool. *Bioinformatics* **2015**, *31*, 1136−1137.
(57) Kim, D. N.; Cho, K.-H.; Oh, W. S.; Lee, C. J.; Lee, S. K.; Jung, J.; No, K. T. EaMEAD: Activation Energy Prediction of Cytochrome P450 Mediated Metabolism With Effective Atomic Descriptors. *J. Chem. Inf. Model.* **2009**, *49*, 1643−1654.
(58) Zaretzki, J.; Boehm, K. M.; Swamidass, S. J. Improved Prediction of CYP-Mediated Metabolism With Chemical Fingerprints. *J. Chem. Inf. Model.* **2015**, *55*, 972−982.
(59) Dang, N. L.; Hughes, T. B.; Krishnamurthy, V.; Swamidass, S. J. A Simple Model Predicts UGT-mediated Metabolism. *Bioinformatics* **2016**, *32*, 3183−3189.
(60) Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. P. A Model for Predicting Likely Sites of CYP3A4-mediated Metabolism on Drug-Like Molecules. *J. Med. Chem.* **2003**, *46*, 1330−1336.
(61) Peng, J.; Lu, J.; Shen, Q.; Zheng, M.; Luo, X.; Zhu, W.; Jiang, H.; Chen, K. In Silico Site of Metabolism Prediction for Human UGT-catalyzed Reactions. *Bioinformatics* **2014**, *30*, 398−405.
(62) Rudik, A.; Dmitriev, A.; Lagunin, A.; Filimonov, D.; Poroikov, V. SOMP: Web-Server for in Silico Prediction of Sites of Metabolism for Drug-Like Compounds. *Bioinformatics* **2015**, *31*, 2046−2048.
(63) Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Modeling Epoxidation of Drug-Like Molecules With a Deep Machine Learning Network. *ACS Cent. Sci.* **2015**, *1*, 168−180.
(64) Dang, N. L.; Matlock, M. K.; Hughes, T. B.; Swamidass, S. J. The Metabolic Rainbow: Deep Learning Phase 1 Metabolism in Five Colors. *J. Chem. Inf. Model.* **2020**, *60*, 1146−1164.
(65) Yousofshahi, M.; Manteiga, S.; Wu, C.; Lee, K.; Hassoun, S. PROXIMAL: a method for Prediction of Xenobiotic Metabolism. *BMC Syst. Biol.* **2015**, *9*, No. 94.
(66) Adams, S. E. Molecular Similarity and Xenobiotic Metabolism. Ph.D. Thesis, University of Cambridge, 2010.
(67) Šicho, M.; Stork, C.; Mazzolari, A.; de Bruyn Kops, C.; Pedretti, A.; Testa, B.; Vistoli, G.; Svozil, D.; Kirchmair, J. FAME 3: predicting the sites of metabolism in synthetic compounds and natural products for phase 1 and phase 2 metabolic enzymes. *J. Chem. Inf. Model.* **2019**, *59*, 3400−3412.
(68) Rudik, A. V.; Bezhentsev, V. M.; Dmitriev, A. V.; Druzhilovskiy, D. S.; Lagunin, A. A.; Filimonov, D. A.; Poroikov, V. V. MetaTox: Web Application for Predicting Structure and Toxicity of Xenobiotics' Metabolites. *J. Chem. Inf. Model.* **2017**, 638.
(69) Meng, J.; Li, S.; Liu, X.; Zheng, M.; Li, H. RD-Metabolizer: An Integrated and Reaction Types Extensive Approach to Predict Metabolic Sites and Metabolites of Drug-Like Molecules. *Chem. Cent. J.* **2017**, *11*, No. 65.
(70) Ridder, L.; Wagener, M. SyGMa: Combining Expert Knowledge and Empirical Scoring in the Prediction of Metabolites. *ChemMedChem* **2008**, *3*, 821−832.
(71) de Bruyn Kops, C.; Stork, C.; Šícho, M.; Kochev, N.; Svozil, D.; Jeliazkova, N.; Kirchmair, J. GLORY: Generator of the Structures of Likely Cytochrome P450 Metabolites Based on Predicted Sites of Metabolism. *Front. Chem.* **2019**, *7*, No. 402.
(72) Djoumbou-Feunang, Y.; Fiamoncini, J.; Gil-de-la Fuente, A.; Greiner, R.; Manach, C.; Wishart, D. S. BioTransformer: a comprehensive computational tool for small molecule metabolism prediction and metabolite identification. *J. Cheminf.* **2019**, *11*, No. 2.
(73) Kalis, M. M.; Huff, N. A. Oxcarbazepine, an Antiepileptic Agent. *Clin. Ther.* **2001**, *23*, 680−700.
(74) Harder, J. L.; Heung, M.; Vilay, A. M.; Mueller, B. A.; Segal, J. H. Carbamazepine and the Active Epoxide Metabolite Are Effectively Cleared by Hemodialysis Followed by Continuous Venovenous Hemodialysis in an Acute Overdose. *Hemodial. Int.* **2011**, *15*, 412−415.
(75) Jayasutha, J.; Bhargavdilip, S.; Kishore, K.; Ramasamy, C. Comparision of Efficacy and Safety of Carbamazepine and Eslicarbazepine in Adult Partial and Generalized Seizures. *Asian J. Pharm. Clin. Res.* **2014**, *7*, 144.
(76) Testa, B.; Pedretti, A.; Vistoli, G. Reactions and Enzymes in the Metabolism of Drugs and Other Xenobiotics. *Drug Discovery Today* **2012**, *17*, 549−560.
(77) Hughes, T. B.; Dang, N. L.; Miller, G. P.; Swamidass, S. J. Modeling Reactivity to Biological Macromolecules With a Deep Multitask Network. *ACS Cent. Sci.* **2016**, *2*, 529−537.
(78) Flynn, N. R.; Dang, N. L.; Ward, M. D.; Swamidass, S. J. XenoNet: Inference and Likelihood of Intermediated Metabolite Formation. *J. Chem. Inf. Model.* **2020**, 3431.
(79) Melnick, R. L.; Kohn, M. C. Mechanistic Data Indicate That 1, 3-Butadiene Is a Human Carcinogen. *Carcinogenesis* **1995**, *16*, 157−163.
(80) Bond, J.; Csanady, G.; Leavens, T.; Medinsky, M. Research Strategy for Assessing Target Tissue Dosimetry of 1, 3-Butadiene in Laboratory Animals and Humans. *IARC Sci. Publ.* **1993**, 45−55.
(81) U.S. Environmental Protection Agency. *Health Assessment of 1,3-Butadiene*; EPA, 2002.

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

(82) Csanády, G. A.; Guengerich, F. P.; Bond, J. A. Comparison of the Biotransformation of 1, 3-Butadiene and Its Metabolite, Butadiene Monoepoxide, by Hepatic and Pulmonary Tissues From Humans, Rats and Mice. *Carcinogenesis* **1992**, *13*, 1143−1153.
(83) Duescher, R. J.; Elfarra, A. A. Human Liver Microsomes Are Efficient Catalysts of 1, 3-Butadiene Oxidation: Evidence for Major Roles by Cytochromes P450 2A6 and 2E1. *Arch. Biochem. Biophys.* **1994**, *311*, 342−349.
(84) Duescher, R.; Elfarra, A. 1,3-Butadiene Oxidation by Human Myeloperoxidase. Role of Chloride Ion in Catalysis of Divergent Pathways. *J. Biol. Chem.* **1992**, *267*, 19859−19865.
(85) Krause, R. J.; Sharer, J. E.; Elfarra, A. A. Epoxide Hydrolase-Dependent Metabolism of Butadiene Monoxide to 3-Butene-1, 2-Diol in Mouse, Rat, and Human Liver. *Drug Metab. Dispos.* **1997**, *25*, 1013−1015.
(86) Krause, R. J.; Kemper, R. A.; Elfarra, A. A. Hydroxymethylvinyl Ketone: A Reactive Michael Acceptor Formed by the Oxidation of 3-Butene-1, 2-Diol by cDNA-expressed Human Cytochrome P450s and Mouse, Rat, and Human Liver Microsomes. *Chem. Res. Toxicol.* **2001**, *14*, 1590−1595.
(87) Landrum, G. *RDKit: Open-Source Cheminformatics*; BibSonomy, 2006.
(88) Ortiz de Montellano, P. R. Cytochrome P450-Activated Prodrugs. *Future Med. Chem.* **2013**, *5*, 213−228.
(89) Basit, A. W.; Lacey, L. F. Colonic Metabolism of Ranitidine: Implications for Its Delivery and Absorption. *Int. J. Pharm.* **2001**, *227*, 157−165.
(90) Guasch, L.; Sitzmann, M.; Nicklaus, M. C. Enumeration of Ring-Chain Tautomers Based on SMIRKS Rules. *J. Chem. Inf. Model.* **2014**, *54*, 2423−2432.
(91) Deng, W.; Schneider, G.; So, W. V. Mapping Chemical Structures to Markush Structures Using SMIRKS. *Mol. Inf.* **2011**, *30*, 665−671.
(92) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31−36.
(93) Kasuya, F.; Igarashi, K.; Fukui, M. Metabolism of Benoxinate in Humans. *J. Pharm. Sci.* **1987**, *76*, 303−305.
(94) Meatherall, R. Benzodiazepine Screening Using EMIT II extregistered and TDx extregistered: Urine Hydrolysis Pretreatment Required. *J. Anal. Toxicol.* **1994**, *18*, 385−390.
(95) Shiraga, T.; Kaneko, H.; Iwasaki, K.; Tozuka, Z.; Suzuki, A.; Hata, T. Identification of Cytochrome P450 Enzymes Involved in the Metabolism of Zotepine, an Antipsychotic Drug, in Human Liver Microsomes. *Xenobiotica* **1999**, *29*, 217−229.
(96) Knaak, J.; Albayati, M.; Raabe, O.; Blancato, J. Development of in VitroVmax and Km Values for the Metabolism of Isofenphos by P-450 Liver Enzymes in Animals and Human. *Toxicol. Appl. Pharmacol.* **1993**, *120*, 106−113.
(97) Meuldermans, W.; Hendrickx, J.; Mannens, G.; Lavrijsen, K.; Janssen, C.; Bracke, J.; Le Jeune, L.; Lauwers, W.; Heykants, J. The Metabolism and Excretion of Risperidone After Oral Administration in Rats and Dogs. *Drug Metab. Dispos.* **1994**, *22*, 129−138.
(98) Reilly, C. A.; Henion, F.; Bugni, T. S.; Ethirajan, M.; Stockmann, C.; Pramanik, K. C.; Srivastava, S. K.; Yost, G. S. Reactive Intermediates Produced From the Metabolism of the Vanilloid Ring of Capsaicinoids by P450 Enzymes. *Chem. Res. Toxicol.* **2013**, *26*, 55−66.
(99) O'Byrne, P. M.; Williams, R.; Walsh, J. J.; Gilmer, J. F. Synthesis, Screening and Pharmacokinetic Evaluation of Potential Prodrugs of Bupropion. Part One: In Vitro Development. *Pharmaceuticals* **2014**, *7*, 595−620.
(100) Ballesteros-Gómez, A.; Erratico, C. A.; Van den Eede, N.; Ionas, A. C.; Leonards, P. E.; Covaci, A. In Vitro Metabolism of 2-Ethylhexyldiphenyl Phosphate (EHDPHP) by Human Liver Microsomes. *Toxicol. Lett.* **2015**, *232*, 203−212.
(101) Schmidt, L.; Belov, V. N.; Göen, T. Human Metabolism of Δ 3-Carene and Renal Elimination of Δ 3-Caren-10-Carboxylic Acid (Chaminic Acid) After Oral Administration. *Arch. Toxicol.* **2015**, *89*, 381−392.
(102) Li, F.; Miao, Y.; Zhang, L.; Neuenswander, S. A.; Douglas, J. T.; Ma, X. Metabolomic Analysis Reveals Novel Isoniazid Metabolites and Hydrazones in Human Urine. *Drug Metab. Pharmacokinet.* **2011**, *26*, 569−576.
(103) Marques, L. A.; Kool, J.; de Kanter, F.; Lingeman, H.; Niessen, W.; Irth, H. Production and On-Line Acetylcholinesterase Bioactivity Profiling of Chemical and Biological Degradation Products of Tacrine. *J. Pharm. Biomed. Anal.* **2010**, *53*, 609−616.
(104) Nakamura, A. Cytochrome P450 Isoforms Responsible for the N-Deethylation and Cyclohexane-Hydroxylation of NS-21. *Xenobiotica* **1999**, *29*, 243−252.
(105) Monks, T. J.; Hanzlik, R. P.; Cohen, G. M.; Ross, D.; Graham, D. G. Quinone Chemistry and Toxicity. *Toxicol. Appl. Pharmacol.* **1992**, *112*, 2−16.
(106) Xiong, R.; Siegel, D.; Ross, D. Quinone-Induced Protein Handling Changes: Implications for Major Protein Handling Systems in Quinone-Mediated Toxicity. *Toxicol. Appl. Pharmacol.* **2014**, *280*, 285−295.
(107) Christie, G.; Breckenridge, A.; Park, B. Drug-Protein Conjugates−Xviii: Detection of Antibodies Towards the Antimalarial Amodiaquine and Its Quinone Imine Metabolite in Man and the Rat. *Biochem. Pharmacol.* **1989**, *38*, 1451−1458.
(108) Yamamoto, Y.; Yamazaki, H.; Ikeda, T.; Watanabe, T.; Iwabuchi, H.; Nakajima, M.; Yokoi, T. Formation of a Novel Quinone Epoxide Metabolite of Troglitazone With Cytotoxic to HepG2 Cells. *Drug Metab. Dispos.* **2002**, *30*, 155−160.
(109) Awad, H. M.; Boersma, M. G.; Vervoort, J.; Rietjens, I. M. Peroxidase-Catalyzed Formation of Quercetin Quinone Methide-Glutathione Adducts. *Arch. Biochem. Biophys.* **2000**, *378*, 224−233.
(110) Spaldin, V.; Madden, S.; Pool, W.; Woolf, T.; Park, B. The Effect of Enzyme Inhibition on the Metabolism and Activation of Tacrine by Human Liver Microsomes. *Br. J. Clin. Pharmacol.* **1994**, *38*, 15−22.
(111) Madden, S.; Spaldin, V.; Hayes, R.; Woolf, T.; Pool, W.; Park, B. Species Variation in the Bioactivation of Tacrine by Hepatic Microsomes. *Xenobiotica* **1995**, *25*, 103−116.
(112) Wen, B.; Fitch, W. L. Screening and Characterization of Reactive Metabolites Using Glutathione Ethyl Ester in Combination With Q-Trap Mass Spectrometry. *J. Mass Spectrom.* **2009**, *44*, 90−100.
(113) Choi, S. Y.; Kim, S.; Hwang, J. S.; Lee, B. G.; Kim, H.; Kim, S. Y. Benzylamide Derivative Compound Attenuates the Ultraviolet B-Induced Hyperpigmentation in the Brownish Guinea Pig Skin. *Biochem. Pharmacol.* **2004**, *67*, 707−715.
(114) Sun, Y.-W.; Guengerich, F. P.; Sharma, A. K.; Boyiri, T.; Amin, S.; El-Bayoumy, K. Human Cytochromes P450 1A1 and 1B1 Catalyze Ring Oxidation but Not Nitroreduction of Environmental Pollutant Mononitropyrene Isomers in Primary Cultures of Human Breast Cells and Cultured McF-10a and McF-7 Cell Lines. *Chem. Res. Toxicol.* **2004**, *17*, 1077−1085.
(115) Wells-Knecht, K. J.; Ott, G. R.; Cheng, M.; Wells, G. J.; Breslin, H. J.; Gingrich, D. E.; Weinberg, L.; Mesaros, E. F.; Huang, Z.; Yazdanian, M.; Ator, M. A.; Aimone, L.; Zeigler, K.; Dorsey, B. D. 2, 7-Disubstituted-Pyrrolotriazine Kinase Inhibitors With an Unusually High Degree of Reactive Metabolite Formation. *Chem. Res. Toxicol.* **2011**, *24*, 1994−2003.
(116) Hughes, T. B.; Miller, G. P.; Swamidass, S. J. Site of Reactivity Models Predict Molecular Reactivity of Diverse Chemicals With Glutathione. *Chem. Res. Toxicol.* **2015**, *28*, 797−809.
(117) Rendic, S.; Guengerich, F. P. Contributions of Human Enzymes in Carcinogen Metabolism. *Chem. Res. Toxicol.* **2012**, *25*, 1316−1383.
(118) Marco-Contelles, J.; Molina, M. T.; Anjum, S. Naturally Occurring Cyclohexane Epoxides: Sources, Biological Activities, and Synthesis. *Chem. Rev.* **2004**, *104*, 2857−2900.
(119) Delaine, T.; Ponting, D. J.; Niklasson, I. B.; Emter, R.; Hagvall, L.; Norrby, P.-O.; Natsch, A.; Luthman, K.; Karlberg, A.-T.

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716

Epoxyalcohols: Bioactivation and Conjugation Required for Skin Sensitization. *Chem. Res. Toxicol.* **2014**, *27*, 1860−1870.
(120) Chan, K.; Jensen, N.; O'Brien, P. J. Structure-Activity Relationships for Thiol Reactivity and Rat or Human Hepatocyte Toxicity Induced by Substituted P-Benzoquinone Compounds. *J. Appl. Toxicol.* **2008**, *28*, 608−620.
(121) Ramanarayanan, J.; Krishnan, G. Hepatotoxicity and EGFR Inhibition. *Clin. Adv. Hematol. Oncol.* **2008**, *6*, 200−201.
(122) Gschwind, H.-P.; Pfaar, U.; Waldmeier, F.; Zollinger, M.; Sayer, C.; Zbinden, P.; Hayes, M.; Pokorny, R.; Seiberling, M.; Ben-Am, M.; Peng, B.; Gross, G. Metabolism and Disposition of Imatinib Mesylate in Healthy Volunteers. *Drug Metab. Dispos.* **2005**, *33*, 1503−1512.
(123) Baumgartner, C. M.; Koenighaus, H.; Ebner, J. K.; Henke, J.; Schuster, T.; Erhardt, W. D. Cardiovascular Effects of Dipyrone and Propofol on Hemodynamic Function in Rabbits. *Am. J. Vet. Res.* **2009**, *70*, 1407−1415.
(124) Schug, S. A.; Manopas, A. Update on the Role of Non-Opioids for Postoperative Pain Treatment. *Best Pract. Res., Clin. Anaesthesiol.* **2007**, *21*, 15−30.
(125) Wessel, J. C.; Matyja, M.; Neugebauer, M.; Kiefer, H.; Daldrup, T.; Tarbah, F. A.; Weber, H. Characterization of Oxalic Acid Derivatives as New Metabolites of Metamizol (Dipyrone) in Incubated Hen's Egg and Human. *Eur. J. Pharm. Sci.* **2006**, *28*, 15−25.
(126) Basak, G.; Drozd-Sokołowska, J.; Wiktor-Jedrzejczak, W. Update on the Incidence of Metamizole Sodium-Induced Blood Dyscrasias in Poland. *J. Int. Med. Res.* **2010**, *38*, 1374−1380.
(127) Jasiecka, A.; Maślanka, T.; Jaroszewski, J. Pharmacological Characteristics of Metamizole. *Pol. J. Vet. Sci.* **2014**, *17*, 207−214.
(128) García-Martínez, J.; Fresno, J. V.; Lastres, P.; Bernabéu, C.; Betés, P.; Martín-Pérez, J. Effect of Metamizol on Promyelocytic and Terminally Differentiated Granulocytic Cells. Comparative Analysis With Acetylsalicylic Acid and Diclofenac. *Biochem. Pharmacol.* **2003**, *65*, 209−217.
(129) Uetrecht, J. P.; Ma, H. M.; MacKnight, E.; McClelland, R. Oxidation of Aminopyrine by Hypochlorite to a Reactive Dication: Possible Implications for Aminopyrine-Induced Agranulocytosis. *Chem. Res. Toxicol.* **1995**, *8*, 226−233.
(130) Noah, S.; Karen, W.; Patricia, P.; Jeffery, W. Acute Generalized Exanthematous Pustulosis Resembling Toxic Epidermal Necrolysis Caused by Famotidine. *Acta Derm.-Venereol.* **2003**, *83*, 76.
(131) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011**, *3*, No. 33.

https://dx.doi.org/10.1021/acs.jcim.0c00360
J. Chem. Inf. Model. 2020, 60, 4702−4716