

EDGE ARTICLE

[View Article Online](#)
[View Journal](#) | [View Issue](#)



Cite this: *Chem. Sci.*, 2025, **16**, 18884

All publication charges for this article have been paid for by the Royal Society of Chemistry

DeepMetab: a comprehensive and mechanistically informed graph learning framework for end-to-end drug metabolism prediction

Yiling Zhou,^a Dejun Jiang,^a Xiao Wei,^a Jiacai Yi,^b Yikun Wang,^a Youchao Deng^a and Dongsheng Cao  ^{*a}

Predicting drug metabolism remains a long-standing challenge in pharmacokinetics due to the mechanistic complexity of enzymatic transformations and the fragmented nature of current computational tools. Existing models are typically limited to isolated tasks – substrate recognition, metabolic site identification, or metabolite generation – lacking mechanistic fidelity, holistic integration, and chemical interpretability. Here, we introduce DeepMetab, the first comprehensive and mechanistically informed deep graph learning framework for end-to-end prediction of CYP450-mediated drug metabolism. DeepMetab uniquely integrates three essential prediction tasks – substrate profiling, site-of-metabolism (SOM) localization, and metabolite generation – within a unified multi-task architecture. It employs a dual-labeling strategy that simultaneously captures atom- and bond-level reactivity, and infuses multi-scale features including quantum-informed and topological descriptors into a graph neural network (GNN) backbone. A curated knowledge base of expert-derived reaction rules further ensures mechanistic consistency during metabolite synthesis. DeepMetab consistently outperformed existing models across nine major CYP isoforms in all three prediction tasks. Its strong generalizability was further validated on 18 recently FDA-approved drugs, achieving 100% TOP-2 accuracy for SOM prediction and accurately recovering several experimentally confirmed metabolites absent from the training set. Visualization of learned representations reveals expert-level discernment of electronic characteristics, steric architecture, and regiochemical determinants, underscoring the model's interpretability. Together, DeepMetab represents a next-generation AI system that bridges symbolic reaction rules and deep graph reasoning to deliver accurate, interpretable, and end-to-end metabolism predictions, offering tangible value for both preclinical research and regulatory applications.

Received 24th June 2025
Accepted 4th September 2025

DOI: 10.1039/d5sc04631a
rsc.li/chemical-science

Introduction

Drug metabolism, also referred to as biotransformation, is the biochemical transformation of pharmaceutical substances by specialized enzymatic systems, predominantly occurring in the liver.¹ This sophisticated process, involving the structural transformation of xenobiotics through Phase I and/or Phase II reactions, is fundamental for maintaining homeostatic physiological functions and plays a pivotal role in drug development.² Moreover, the formation of toxic or reactive metabolites can lead to severe acute adverse effects, underscoring the importance of drug metabolism in safety evaluations. Over the past decades, hepatic metabolism, particularly metabolism mediated by cytochrome P450 (CYP450) enzymes, has been recognized as a primary contributor to liver toxicity, which is one of

the leading causes of drug withdrawals, accounting for approximately 15–30% of all cases.³ One well-known example is the global withdrawal of troglitazone in 1997, following reports of life-threatening liver failure caused by its CYP3A4-mediated metabolites.⁴

Among numerous metabolic enzymes in the human body, the cytochrome P450 enzyme system stands out as the most critical Phase I metabolic pathway, responsible for approximately 75% of drug metabolism processes in the human body.⁵ As a typical monooxygenase, CYP450 primarily introduces polar functional groups into substrate molecules *via* oxidation, reduction, hydrolysis, and other related transformations, thereby facilitating subsequent Phase II metabolism.⁶ Given the versatility of its catalytic functions, the CYP450 enzyme system is subdivided into 18 families and 43 subfamilies, with isoenzymes from the CYP1, CYP2, and CYP3 families accounting for the predominant metabolic activities.⁷ Notably, different isoenzymes exhibit substrate specificity, catalyzing the metabolism of specific drugs and their metabolites through distinct functional transformations.⁸ For instance, CYP2C19 catalyzes

^aXiangya School of Pharmaceutical Sciences, Central South University, Changsha 410013, Hunan, P.R. China. E-mail: oriental-cds@163.com

^bCollege of Computer, National University of Defense Technology, Changsha 410073, Hunan, China



the conversion of voriconazole to its inactive *N*-oxide metabolite;⁹ CYP2D6 mediates the transformation of codeine into morphine, enhancing its analgesic activity,¹⁰ while CYP3A4 catalyzes the conversion of nefazodone to its toxic metabolite 4-hydroxynefazodone.¹¹ These examples illustrate the functional diversity of CYP450 isoenzymes and their profound impact on drug efficacy and toxicity. Therefore, comprehensive understanding of drug metabolism processes, particularly the elucidation of metabolic pathways and metabolites facilitated by major CYP450 isoenzymes, holds significant theoretical value and practical implications for pharmaceutical development and safety assessment. However, traditional experimental approaches are often constrained by high resource demands, time-consuming procedures, and substantial economic costs.¹² In contrast, computational approaches offer an efficient alternative, significantly reducing resource consumption.

Current computational approaches on drug metabolism prediction primarily employs two approaches. The first is a data-driven method based on natural language processing (NLP) models, such as MetaPredictor.¹³ The second approach is mechanism-based, emphasizing the systematic prediction of metabolic enzymes, identification of sites of metabolism, and application of relevant metabolic transformation rules to generate metabolites, as exemplified by tools such as GLORYx¹⁴ and BioTransformer3.0.¹⁵ While the data-driven approach offers end-to-end metabolite prediction, it often overlooks critical intermediate metabolic representations, such as metabolic pathways, which are equally important. Additionally, its performance in metabolite prediction tasks remains suboptimal. For example, MetaPredictor,¹³ due to its reliance on promiscuous datasets and the inherent limitations of text generation algorithms, demonstrates limited accuracy in metabolite generation. Specifically, its site of metabolism (SOM) predictor achieves only 57.8% of TOP-1 accuracy on the test set. Moreover, MetaPredictor¹³ frequently generates meaningless text, requiring multiple filtering steps through automated scripts following the prompt-based metabolite predictor module.

Mechanism-based approaches, which offer more profound insights into the intricate stages of metabolism, have gained significant recognition. Nevertheless, current approaches face critical challenges. Firstly, previous studies are typically limited to isolated tasks and have not provided comprehensive predictions for metabolic characterization, particularly in terms of metabolite generation.¹⁶ For instance, CypReact¹⁷ and CYPstrate¹⁸ just concentrate on metabolizing enzyme prediction, while SMARTCyp3.0,¹⁹ FAME3,²⁰ and CyProduct²¹ (the CYP450 metabolism prediction module of BioTransformer3.0 (ref. 15)) only specialize in SOM prediction. Among these, FAME3 (ref. 20) does not differentiate between isoforms, while tools like SMARTCyp3.0 (ref. 22) only predict metabolism for just three CYP450 isoforms. Secondly, existing work constrained accuracy stemming from algorithmic limitations and insufficient exploitation of available features. While studies like BioTransformer 3.0,¹⁵ which integrates CypReact,¹⁷ CyProduct,²¹ and a metabolite generation module,²¹ claim to offer relatively comprehensive coverage of major isoforms and metabolic

phenotypes, key limitations remain. Specifically, the metabolite generation module lacks induction of several critical reaction rules (e.g., de-halogenation), its algorithmic framework relies solely on simple machine learning (ML) models such as random forest, and it utilizes only basic descriptors, restricting the richness of the information employed. Unlike traditional ML methods, deep learning (DL) also has been applied to SOM prediction for its superior ability in molecular representation learning. Vladimir *et al.*²³ (2023) demonstrated that Graph Convolutional Networks (GCNs), extracting features directly from molecular topology graphs, deliver more better performance in predicting enzyme metabolic sites than traditional ML methods. However, these methods rely on basic chemical features, lacking critical reactivity and physicochemical details necessary for accurate predictions, and depend on promiscuous enzyme data, which fail to effectively distinguish the metabolism of different enzymes. While ML and DL methods tend to offer increased convenience and faster processing times, quantum chemistry approaches are capable of delivering more comprehensive and detailed information regarding molecular interactions and mechanisms. In 2022, StarDrop²⁴ addressed this by integrating semi-empirical quantum chemical descriptors with ML techniques. Despite the limited dataset and the use of simpler algorithms, this approach enhanced efficiency and accuracy to a certain extent, suggesting that combining computational chemistry with DL may produce more robust and generalizable models for metabolic predictions. All in all, despite significant advances, contemporary mechanistic-driven metabolism prediction models remain largely confined to discrete tasks, with their predictive capacity further hampered by intrinsic algorithmic and representational limitations. These constraints substantially undermine both accuracy and translatability across the vast chemical and structural diversity of drug-like molecules. Moreover, the persistent absence of a comprehensive, systematized, and hierarchically structured formalization of metabolic rules continues to fundamentally compromise the robustness, reproducibility, and precision of metabolite generation.

To address these challenges, we present the first integrated and mechanistically grounded graph learning platform for end-to-end prediction of CYP450-mediated drug metabolism. Unlike prior integrated or end-to-end tools, BioTransformer 3.0 (ref. 15) relies on traditional ML heuristics with limited mechanistic fidelity and scalability; the commercial platforms StarDrop²⁴ and ADMET Predictor²⁵ employ semi-empirical/force-field methodologies that are computationally demanding and less amenable to end-to-end learning; and MetaPredictor¹³ targets mixed-enzyme metabolism without recovering pathway information. In contrast, DeepMetab's mechanism-informed GNN is the first deep-learning-based system that prioritizes pathway inference while localizing SOMs and generating metabolites, thereby broadening task coverage and mechanistic rigor, culminating in a comprehensive end-to-end platform for CYP450-mediated drug metabolism. Notably, its strong generalizability was further validated on 18 recently FDA-approved drugs, achieving 100% TOP-2 accuracy for SOM prediction and accurately recovering several experimentally confirmed



metabolites absent from the training set. Visual analysis of model outputs highlights advanced recognition of key electronic, steric, and regiochemical factors, attesting to its high interpretability. Altogether, DeepMetab offers a uniquely robust computational tool for advancing drug metabolism studies.

Methods and materials

Collection and compilation of substrate, SOM dataset, and metabolic rules

Since data scarcity is one of the biggest challenges in metabolism modeling, we have made extensive efforts to collect and expand the dataset and metabolic rules to address this limitation. Fig. 1 presents a systematic workflow of curation of substrate, SOM dataset and metabolic reaction rule compilation. For the substrate dataset, we constructed a molecular substrate dataset comprising over 3800 compounds across 9 CYP450 isoforms. The positive samples were primarily sourced from the research work of Zaretzki *et al.*²⁶ and Yang *et al.*,^{27,28} encompassing experimentally validated molecular substrates metabolized by at least one CYP450 isoform. The negative samples were systematically integrated from the datasets of existing models including CypReact¹⁷ and CYPstrate,¹⁸ along with the DrugBank²⁹ database. More information about the curation of the substrate dataset is available in the SI (S1.1 Section). During data processing, the metabolic enzyme information for each molecule was encoded into a binary vector (consisting of zeros and ones), where each of the nine elements corresponds to the metabolic status of a specific CYP450 isoform, as illustrated in Fig. 1.

The SOM data in our study were compiled from multiple sources, including the research work of Zaretzki *et al.*,²⁶ modeling data from CyProduct,²¹ established databases such as DrugBank²⁹ and BRENDA,³⁰ and relevant literature.³¹ Compared to the existing comprehensive EBoMD dataset,²¹ our expanded SOM dataset includes approximately one-third more compounds, reaching over 900 molecular substrates. This expansion substantially enhances both the richness and diversity of the available SOM data. Considering the current challenges of ambiguous matching and incomplete summarization of metabolic rules, we developed an innovative and efficient site-labeling methodology that integrates the advantages of both atom of metabolism (AOM) and bond of metabolism (BOM) to ensure accuracy and uniqueness in metabolite generation through reaction rules. This methodology encompasses two key aspects: first, we implemented a combined SOM labeling approach utilizing both AOM and BOM to prevent reaction type misclassification. In curation of the SOM dataset, reactions occurring at single atomic site (such as hydroxylation and heteroatom oxidation) are labeled using AOM, while reactions involving atomic pairs (chemical bonds), including dealkylation, hydrolysis, and epoxidation, are labeled using BOMs. This approach significantly reduces misclassification of multiple reactions at the same site. As illustrated in the right panel of Fig. 1 Rules, the parent molecule undergoes an *N*-dealkylation reaction, leading to the formation of an exposed amino group while the cleaved methyl group is converted to

formaldehyde as a byproduct. Due to the AOM labeling at the C adjacent to the N in amino nitrogen, conventional AOM annotation methodology could lead to ambiguity with C-hydroxylation reactions – a common error in GLORYx.¹⁴ To address this, our refined annotation strategy effectively differentiates between these reaction types by implementing *N*-dealkylation as BOM and C-hydroxylation as AOM.

Second, we introduced the principles of minimal and optimal labeling to address multi-site reactions. This approach minimizes reaction site labeling when multiple sites are involved, thereby reducing misclassification risks. The left panel of Fig. 1 Rules demonstrates this principle with the reduction of an α,β -unsaturated ketone carbonyl. Simultaneous labeling of all changing chemical bonds could lead to misidentification of the carbon–carbon double bond as an epoxidation site—a limitation of BioTransformer rules.¹⁵ Our minimal labeling approach effectively resolves such multi-site reaction ambiguities. Finally, as shown in the SOM part of Fig. 1, we encoded both AOMs and BOMs labeling as binary vectors (consisting of zeros and ones), where element positions correspond to the indices of respective atoms or bonds within the molecule.

The metabolic reaction rules in this study were systematically compiled and summarized through extensive literature review.³² Building upon existing rules from BioTransformer¹⁵ and SyGMA,³³ combined with our innovative metabolic site labeling approach mentioned above, we developed a more comprehensive set of metabolic rules through extensive literature collection and analysis. To achieve systematic and comprehensive categorization, metabolic reactions were hierarchically classified into four primary categories: oxidation, hydrolysis, reduction, and dehalogenation. These were further refined into 15 secondary subcategories and several tertiary subcategories, exceeding the number of BioTransformer¹⁵ rules by about 25% and demonstrating an improvement of over 10% in comparative analyses with similar models, as illustrated by the pie chart in the Rules section of Fig. 1. The implementation of these rules was accomplished using the RDKit³⁴ package in Python.

The GNN algorithm used in DeepMetab

Our model architecture consists of two main components: the substrate prediction module, which determines the enzyme or enzymes for which a given molecule can serve as a substrate, and the SOM prediction module, which predicts the metabolic sites of the molecule for these enzymes. Both modules are based on a Graph Neural Network (GNN), a neural network tailored to learning representations from graph-structured data, which is particularly suitable for cheminformatics tasks due to its ability to capture molecular structures and relationships.

Specifically, the workflow of GNN architecture consists of three steps: initialization, message passing and updating, and final embedding generation. During initialization, edge features are derived from their individual properties (e.g., bond type) and the features of connected nodes (e.g., atom types), ensuring that the directed nature of the graph is respected as



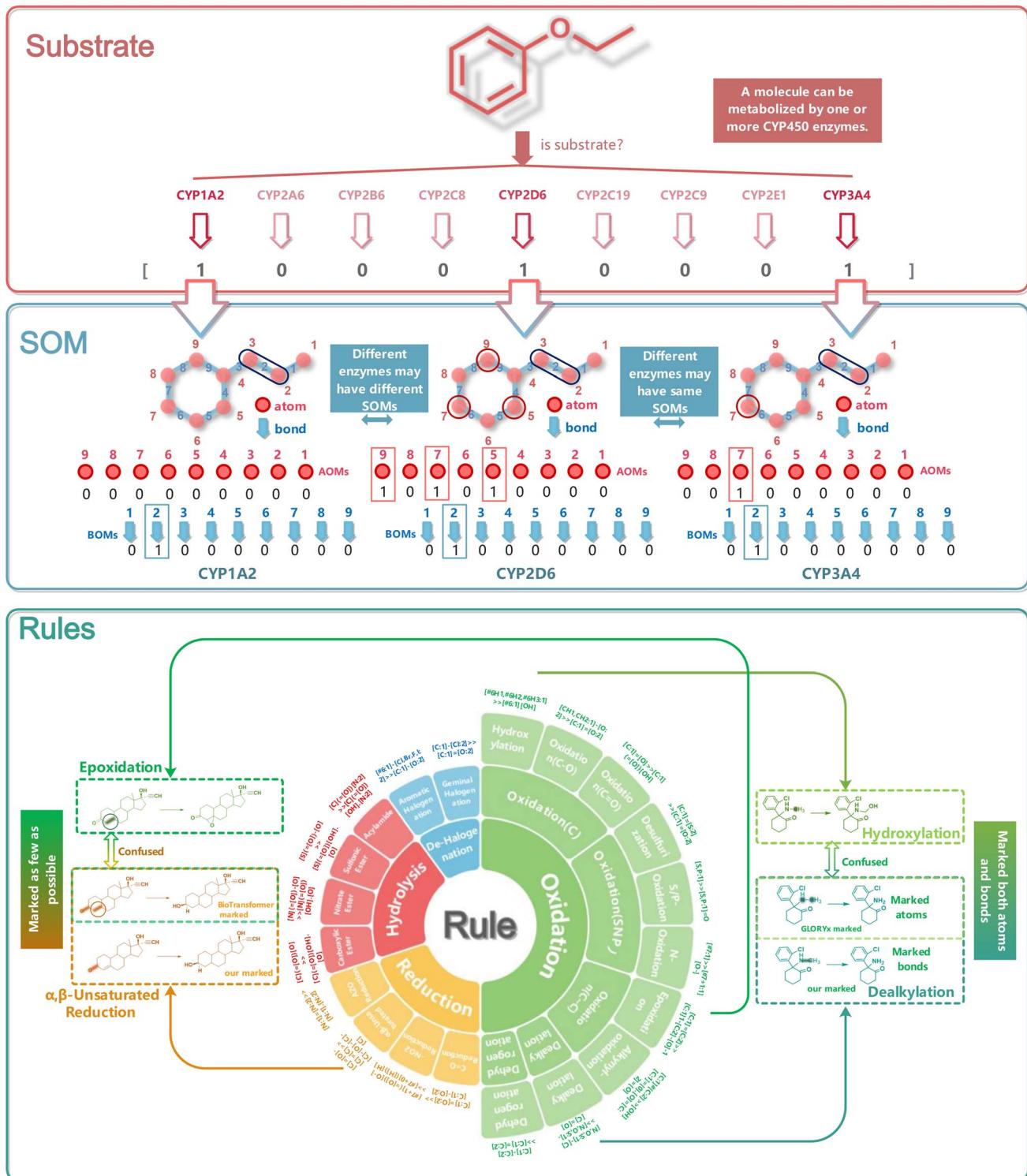


Fig. 1 The workflow of curation of substrate and SOM dataset, and compilation of metabolic reaction rules. The diagram illustrates how various substrates are processed by different CYP450 enzymes, leading to distinct SOM. It also showcases the hierarchical classification of metabolic reaction rules, including key categories such as oxidation, reduction, hydrolysis, and dehalogenation, represented in the central pie chart.

initial messages do not rely on reverse messages. That message m_{vw}^{t+1} does not depend on its reverse message m_{vw}^t from the previous iteration. Eqn (1) shows how to initialize the edge hidden states before the first step of message passing:

$$h_{\text{vw}}^0 = \tau(W_i \text{cat}(x_v, e_{\text{vw}})) \quad (1)$$

where $W_i \in R_i^{h \times h}$ is a learned matrix, $\text{cat}(x_v, e_{vw})$ is the concatenation of the atom features x_v for atom v and the bond features

e_{vw} for bond vw , and τ is the ReLU activation function. In the message-passing phase, edge features are updated by aggregating information from neighboring nodes and edges, enabling the model to capture higher-order interactions. Rather than performing message passing based on the node hidden states h_v^t and messages m_v^t , our study operates on the edge hidden states h_{vw}^t and messages m_{vw}^t during directed message passing:

$$h_{vw}^{t+1} = \tau(h_{vw}^0 + W_m m_{vw}^{t+1}) \quad (2)$$

$$m_{vw}^{t+1} = \sum_{k \in (N(v)/w)} h_{vw}^t \quad (3)$$

where $W_m \in R^{h \times h}$ is a learned matrix with hidden size h . For $t \in \{1, \dots, T\}$, the hidden states of atoms can be derived after the message passing phase. Similarly, the hidden states of bonds can be obtained using the same method.

$$m_v = \sum_{w \in N(v)} h_{vw}^T \quad (4)$$

$$h_v = \tau(W_\alpha \text{cat}(x_v, m_v)) \quad (5)$$

After several iterations, the final embeddings for nodes and edges are obtained. In the substrate prediction module, atom states are derived by aggregating the features of connected edges, while bond states are directly computed from the updated edge representations, resulting in comprehensive and informative molecular graph embeddings. In contrast, the SOM prediction module does not aggregate atom and bond embeddings into a full molecular graph representation. Instead, individual bond or atom embeddings are utilized directly for predictive tasks.

These embeddings (h_v) are then processed by task-specific feed-forward neural networks (FFNN), which produce outputs for each prediction.

$$y = f(h) \quad (6)$$

In this study, the model predicts the metabolism for nine CYP enzymes, with predictions represented as FFNN outputs. Specifically, the substrate prediction module generates predictions at the molecular level, indicating whether the entire molecule serves as a metabolic substrate. In contrast, the SOM prediction module performs predictions at the atomic (node) or bond (edge) level, identifying specific metabolic sites within the molecule.

Multi-scale features used in the SOM prediction module

It has been well established that computational chemistry methods offer unique advantages in metabolic site prediction, where they can extract richer details of reaction processes and demonstrate superior generalization performance for structurally novel molecules despite their computational intensity. Recent research by William *et al.* (2024)³⁵ further confirms that the incorporation of quantum chemistry (QM) descriptors significantly enhances model performance in molecular

property prediction, exhibiting exceptional accuracy and generalization capability, particularly in small-dataset scenarios. Based on those observations, this study leverages the complementary strengths of deep graph learning and computational chemistry methods by incorporating additional multi-scale quantum mechanical characterization into the GNN framework.

We implemented differentiated feature construction strategies for the two core modules: substrate prediction and SOM prediction. The substrate prediction module primarily utilizes basic topological information from molecular graphs given that the distinction between substrates and non-substrates for CYP enzymes depends largely on global molecular features such as overall shape, size, and structural connectivity, which can be effectively captured by simple graph representations, while the SOM module employs a multi-scale feature representation that systematically integrates molecular graph structural information, atom-level reactivity descriptors, and global molecular descriptors. Notably, the global molecular descriptors derived from the molecule are directly stacked onto each individual atomic feature within the molecular graph, enabling comprehensive incorporation of both local and global molecular properties. Further details are provided in the SI (Section S2.3).

The evaluation of DeepMetab

In the model evaluation section of DeepMetab, we employed distinct evaluation metrics for the Substrate, SOM, and Rules modules, considering the unique characteristics of each, to ensure a comprehensive and accurate assessment. For the Substrates module, we utilized multiple metrics to evaluate the model's discriminative capability between substrate and non-substrate molecules. Specifically, we employed Area Under the Receiver Operating Characteristic Curve (AUC) for overall discrimination ability, Accuracy (ACC) for prediction accuracy, along with additional metrics (Area Under the Precision-Recall Curve (PRC), Recall, Precision (PRE), Sensitivity (SEN), and Specificity (SPEC)) to assess performance on imbalanced data. In terms of the SOM module, considering the importance of holistic metabolic site prediction within molecules, we implemented two distinct calculation approaches: (1) computing SOM metrics across all molecules collectively (denoted as A), and (2) calculating average SOM metrics per molecule (denoted as R). While the former represents traditional evaluation methodology, the latter better reflects the practical significance of SOM prediction by accounting for individual molecular results. To comprehensively evaluate this module, we employed AUC, PRC, AUC, ACC, and Jaccard indices to assess the model's ability to distinguish between metabolic and non-metabolic sites. Additionally, TOP-N metrics were also introduced to evaluate the ranking accuracy of predicted metabolic sites. Finally, for the Rules module, we incorporated Jaccard and TOP-N metrics to thoroughly evaluate both the accuracy of metabolite generation and the reliability of top-ranked predictions.



Table 1 Curated training and test datasets for the substrate module of nine cytochrome P450 enzymes

	Enzyme	CYP1A2	CYP2A6	CYP2B6	CYP2C8	CYP2C9	CYP2C19	CYP2D6	CYP2E1	CYP3A4
Training	Substrate	416	102	154	157	369	314	396	148	1081
	Non-substrate	1719	1736	1684	1681	1779	1776	1694	1690	1910
Test	Substrate	40	14	12	12	33	29	27	8	63
	Non-substrate	115	120	123	120	114	120	117	122	113

Results and discussion

Overview of dataset and metabolic rules

The substrate training dataset comprises over 3500 molecules (Table 1), with approximately 1500 being metabolized by one or

more metabolic enzymes. Fig. 2B shows a comparison of substrate and non-substrate molecule counts for each of the nine CYP isoforms, revealing a significant class imbalance. This imbalance is particularly pronounced in 'niche' isoforms such as CYP2A6 and CYP2E1, where the ratio of positive to negative

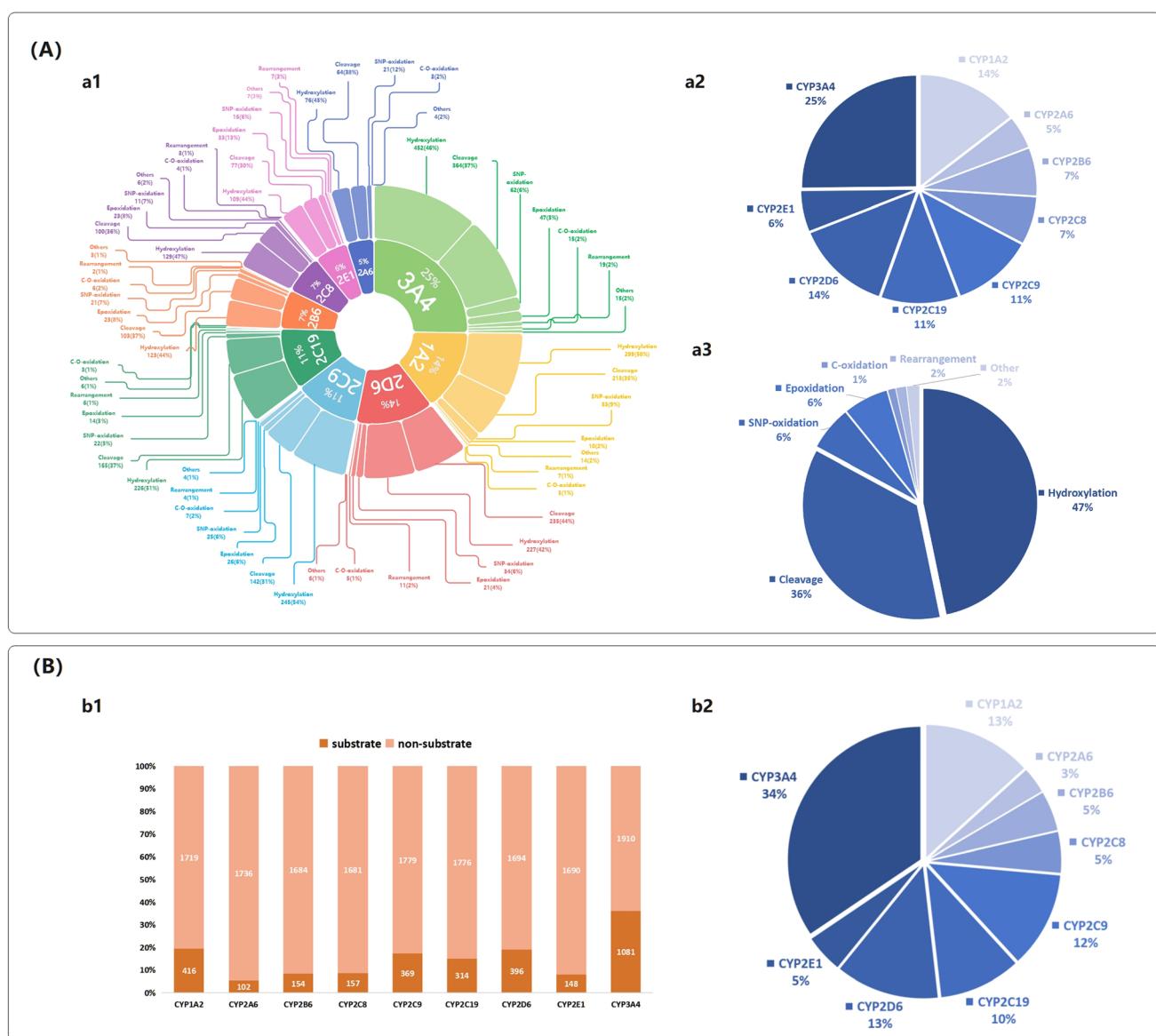


Fig. 2 Comprehensive analysis of substrate distribution and SOM dataset characteristics. (A) SOM training dataset analysis. (a1) The overall composition of reaction types highlights the proportion of metabolic reactions mediated by CYP450 enzymes. (a2) Differential proportions of CYP450 isoforms to metabolic processes, illustrating their relative importance. (a3) Detailed classification of reaction types and their respective proportions to the overall metabolism. (B) Substrate training dataset analysis. (b1) The analysis of the substrate dataset presents the class distribution of substrates and non-substrates among CYP450 isoforms. (b2) Distribution of substrate compounds categorized by specific CYP450 isoforms.



Table 2 Curated training and test datasets for the SOM module of nine cytochrome P450 enzymes

	Enzyme	CYP1A2	CYP2A6	CYP2B6	CYP2C8	CYP2C9	CYP2C19	CYP2D6	CYP2E1	CYP3A4
Training	Substrates	313	104	152	150	251	238	297	129	550
	AOM	276	80	125	115	224	210	209	98	420
	Non-AOM	2515	718	1135	1374	2502	2201	2694	724	6105
	BOM	214	72	108	105	152	142	217	90	361
	Non-BOM	3272	890	1609	1850	2371	2441	3854	923	7561
Test	Substrates	17	9	13	14	17	16	27	10	40
	AOM	293	86	123	112	220	197	213	107	390
	Non-AOM	2649	793	1136	1310	2532	2145	2671	798	5713
	BOM	230	83	113	108	157	153	225	115	348
	Non-BOM	3438	1083	1657	1884	2627	2578	3949	1169	7182

cases is less than 1 : 10. In contrast, more common isoforms display a relatively better balance. For example, CYP3A4 has a positive-to-negative ratio of approximately 1 : 2. This ratio reflects not only its metabolic activity but also the greater availability of data for these more prevalent enzymes. The pie chart in Fig. 2B highlights the distribution of substrate counts among the different CYP isoforms, with CYP3A4 accounting for 34% of the total substrates, underscoring its prevalence and significance in metabolism compared to the other isoforms. The further UpSet plot (Fig. S1.1A) reveals that over half of the molecules are metabolized by multiple enzymes, with 16 molecules being substrates for all enzymes. These findings demonstrate strong correlations among the metabolic tasks of different isoforms, suggesting that a multi-task modeling approach would be beneficial.

SOM training dataset contains nearly 1500 reactions from 874 substrate molecules (Table 2). We performed a comprehensive statistical analysis of the nine CYP enzymes, focusing on the number of substrates classified as AOM and non-AOM, as well as BOM and non-BOM. The results reveal that the largest number of substrates are metabolized by CYP3A4, CYP1A2, and CYP2D6, with CYP3A4 accounting for more than half of the total, reflecting its association in the SOM training dataset with 420 AOMs and 361 BOMs, while CYP1A2 follows with 276 AOMs and 214 BOMs, and CYP2D6 shows 724 AOMs and 90 BOMs. This highlights the metabolic significance of these enzymes, which facilitates the development of models based on this dataset to achieve more effective applications.

Further analysis is performed on the datasets. It is evident that hydroxylation reactions (Fig. 2a3) are the most prevalent, constituting nearly half of the reactions. The next most common type is cleavage reactions, which include *O*-dealkylation, *N*-dealkylation and ester hydrolysis and account for about one-third of the reactions. Fig. 2a1 shows the distribution of the proportions of each reaction type in each enzyme, with certain reaction types showing significantly elevated frequencies in specific enzymes. For instance, SNP-oxidation reactions occur at approximately twice the average frequency in CYP2A6 compared to other isoforms, while epoxidation reactions demonstrate notably higher prevalence in CYP2E1 than in other variants. In contrast, hydroxylation and cleavage reactions maintain consistently high proportions across all isoforms, whereas rearrangement reactions exhibit uniformly low

frequencies across all variants. These findings highlight that modeling each enzymatic isoform respectively is crucial for accurately predicting metabolic outcomes and understanding drug interactions, efficacy, and toxicity. To display the shared and unique substrate molecules of the SOM dataset more intuitively among different metabolic enzymes, an UpSet plot was constructed as shown in Fig. S1.1B. It can be observed that the metabolic enzyme with the highest number of unique substrate molecules is CYP3A4, with 193 unique substrates. However, these unique molecules account for only about 1/3 of all CYP3A4-metabolized molecules. Additionally, in the entire dataset, the sum of shared substrates metabolized by more than one enzyme molecule amounts to 473, which exceeds the number of unique substrate molecules, totaling 401. These observations suggest a high degree of interconnectivity among the metabolism of these enzymes.

To elucidate the reaction rules we formulated, and the SOM annotation methodologies employed for various metabolic processes, Table 3 delineates the principal categories along with illustrative examples of metabolic reactions mediated by CYP450 enzymes. The dataset comprises a total of 6 AOMs and 14 BOMs as examples, meticulously classified according to their labeling types, which reflect the specific enzymatic activities involved. In our classification of reaction types, we adopted a streamlined approach to enhance clarity and comprehensiveness. For instance, within the category of dealkylation reactions, we have effectively consolidated *N*-dealkylation, *O*-dealkylation, and *S*-dealkylation into a unified category termed *X*-dealkylation (shown in No. 12 of Table 3). This category captures the broader concept of heteroatom dealkylation, facilitating a clearer understanding of the underlying mechanisms. Conversely, when examining ester hydrolysis reactions, we have distinguished among assorted groups, such as thioesters and nitrate esters, differentiating them from general carbon esters due to their significant biochemical variances (shown in No. 13 and No. 14 of Table 3). This distinction underscores the need for precise labeling in the study of metabolic pathways. To further exemplify these classifications, the “SOM annotation method” column provides representative instances of specific reaction types, illuminating the signaling pathways involved. Notably, the red markings indicate the labeled sites on the molecular structures, drawing attention to



Table 3 A systematic compilation of metabolic reaction rules governing cytochrome P450 enzymes

No.	Type	Reaction type	SOM annotation method	No.	Type	Reaction type	SOM annotation method
1	AOM	NO ₂ -reduction		11	BOM	α,β-Unsaturated carbonyl reduction	
2	AOM	N-Oxidation		12	BOM	X-Dealkylation	
3	AOM	Hydroxylation (N)		13	BOM	Hydrolysis 1	
4	AOM	S/P-Oxidation		14	BOM	Hydrolysis 2	
5	AOM	C=O oxidation		15	BOM	Azo-reduction	
6	AOM	Hydroxylation		16	BOM	Desulfurization	
7	BOM	Rearrangement		17	BOM	Geminal halide oxidation	
8	BOM	Epoxidation		18	BOM	Dehalogenation	
9	BOM	C-O oxidation		19	BOM	Dehydrogenation	
10	BOM	C=O reduction		20	BOM	Alkynyl-oxidation	

critical functional groups affected during the metabolic transformations.

Overview of DeepMetab workflow

As illustrated in Fig. 3, DeepMetab consists of five distinct modules, including the input module, responsible for processing initial molecular features; the substrate module, which predicts the metabolic enzymes for a given molecule; the SOM module, dedicated to identifying the metabolic sites based on the predicted metabolic enzymes for a given molecule; the rules module, which defines metabolic transformation rules; and the output module, which seamlessly integrates these components to generate comprehensive metabolic pathways and products.

The input module was used to convert molecular SMILES into molecular graphs with multi-scaled features. For both substrate and SOM components, we implemented multi-task GNN³⁶ architecture to construct models for metabolic substrate prediction and metabolic site prediction. This multi-task approach shares underlying feature representations and learning parameters across different CYP isoform prediction tasks, enabling the model to capture common characteristics among isoforms while preserving their unique features. In the training of multi-task models, loss weighting strategies³⁷ have been introduced to address the imbalance between positive and negative samples in the dataset, enabling the model to pay more attention to underrepresented classes during training. In our

pipeline, the substrate module initially analyzes compounds to predict which metabolic enzymes will process them. Once these enzyme predictions are established, the SOM module leverages this information to determine which enzyme-specific model should be applied, enabling accurate prediction of sites of metabolism. What's more, to achieve precise metabolic site prediction, we developed another two complementary multi-task models within the SOM module: one dedicated to identifying AOM and another focused on predicting BOM. This dual-model architecture enables comprehensive characterization of metabolic sites from both atomic and chemical bond perspectives.

Next, in the rules module, we established a systematic metabolite generation framework that precisely matches predicted metabolic sites from the SOM module with a pre-constructed metabolic reaction rule database to generate corresponding metabolite structures. To enhance prediction efficiency and accuracy, we implemented a differentiated matching mechanism based on site types, employing distinct reaction matching strategies for AOM and BOM respectively. Finally, in the last output module, we developed a systematic integration to provide comprehensive and structured prediction results. The system provides structured data output systematically documenting all predicted metabolizing enzymes, metabolic sites, metabolites, and scoring metrics. This comprehensive output format ensures thorough



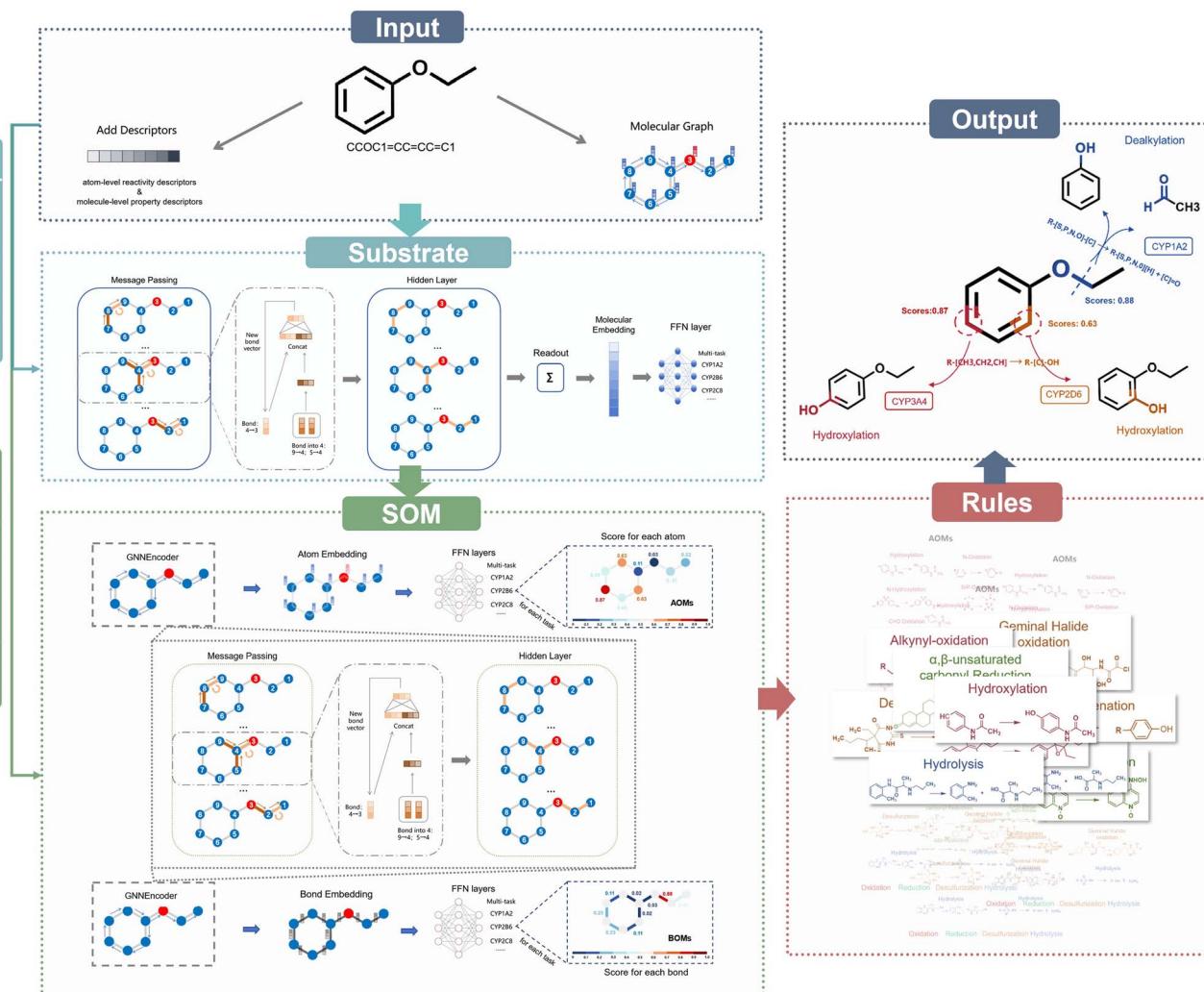


Fig. 3 The workflow of DeepMetab. The input module begins by converting molecules into molecular graphs or graph structures seamlessly integrating multi-scale descriptors into the atomic feature representations. From there, the substrate module analyzes graph representations to predict which CYP450 isoforms will participate in metabolism. The SOM module then builds upon this analysis to precisely identify specific sites of metabolism within the molecule. Once these sites are determined, the rules module applies established reaction mechanisms to generate potential metabolites. The process culminates in the output module, which presents a comprehensive integration of the results, clearly displaying the predicted substrates, metabolic pathways, and resulting metabolites in a cohesive manner.

documentation and facilitates subsequent analysis and interpretation of the prediction results.

The performance of DeepMetab for substrates and SOM prediction

In the substrate prediction module, we conducted a comprehensive comparison between single-task and multi-task GNN models through five-fold cross-validation of the training data-set, with results presented in Fig. 4. The optimal parameters and models were determined through five-fold cross-validation and grid search, with detailed hyperparameter specifications provided in the SI (Section S2.1). Fig. 4D illustrates the comparative analysis of performance metrics between multi-task and single-task approaches, with more detailed enzyme-specific results provided in Table S3.1. In Fig. 4D, bar plots represent the weighted average values across tasks for each

metric, calculated according to the number of substrates for each task, and line plots indicate the variance in five-fold cross-validation. The analysis reveals that the multi-task model outperforms its single-task counterpart in AUC, ACC, PRE, Jaccard and SPEC metrics. A particularly noteworthy observation emerges from the variance analysis: across all evaluation metrics, the multi-task model exhibits remarkably lower variance in five-fold cross-validation, approximately an order of magnitude smaller than the single-task model. This substantial reduction in variance strongly indicates that multi-task learning significantly enhances model stability and robustness. Fig. 4E presents the detailed ROC and PRC curves for individual tasks, demonstrating excellent performance characteristics for multi-task strategy.

In the SOM prediction module, both single-task and multi-task models employ the GNN algorithm framework across nine



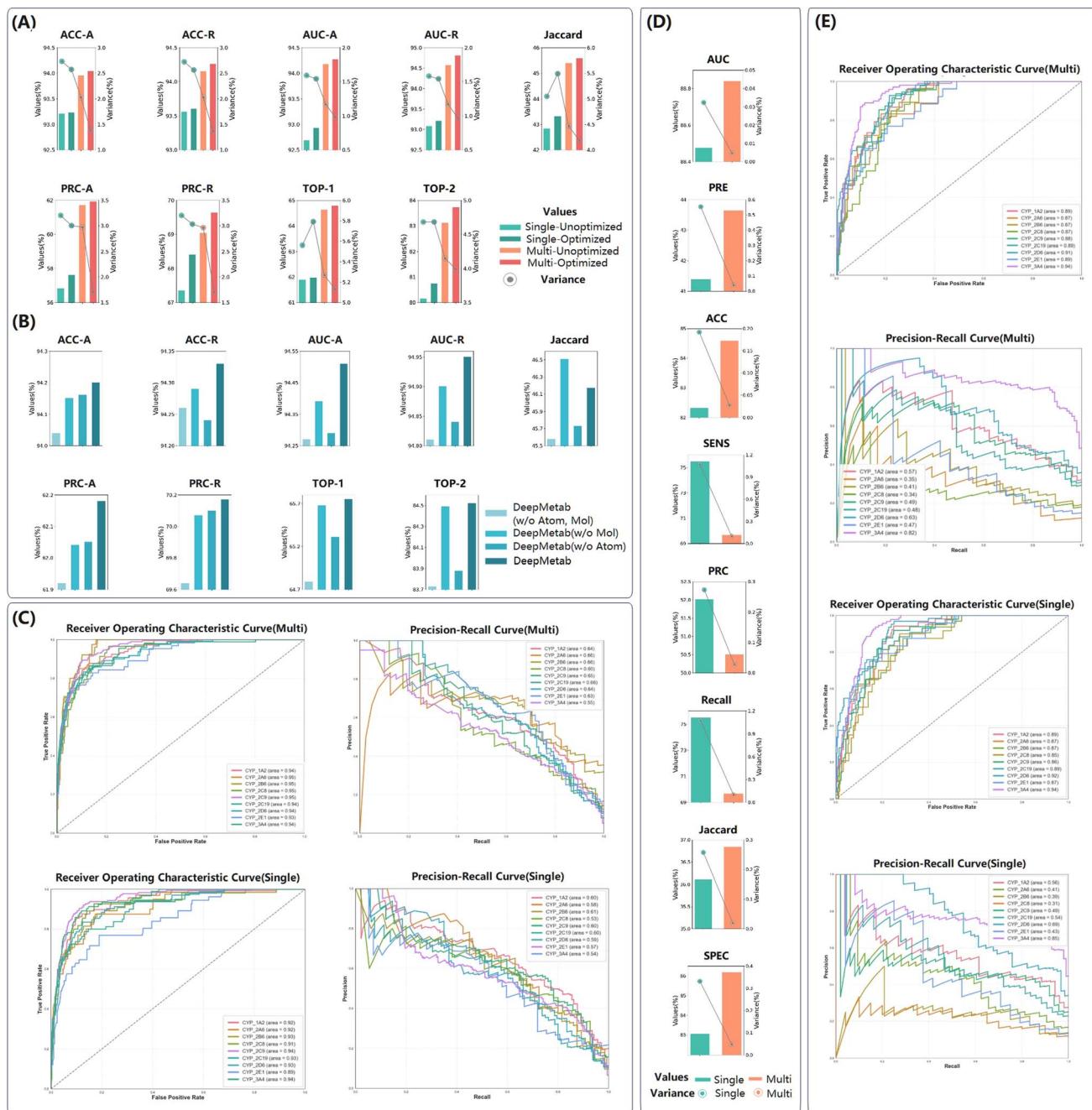


Fig. 4 Performance evaluation of DeepMetab. (A) Five-fold cross-validation results of SOM prediction showing average performance metrics as bars and variance as lines; orange and green represent multi-task and single-task results respectively, with the darker color indicating optimized dataset split. (B) Five-fold cross-validation results with progressive descriptor integration from light to dark blue (all are based on an optimized multi-task method). (C) ROC and PRC curves comparing single task *versus* multi-task performance for SOM models. (D) Five-fold cross-validation results of substrate model showing performance metrics as bars and variance as lines; orange and green represent multi-task and single-task results respectively. (E) ROC and PRC curves comparing single task *versus* multi-task approaches for the substrate model.

CYP enzyme isoforms. For predicting the SOM of each enzyme isoform, our approach inherently encompasses predictions for both AOM and BOM. Fig. 4A reveals that the multi-task model consistently outperforms its single-task counterpart across all evaluation metrics, demonstrating an average improvement of 2–3%. Notably, the most substantial improvement is observed in the PRC-A metric, reaching approximately 5%. Furthermore,

the multi-task approach exhibits superior performance in terms of variance, achieving lower values across all evaluation metrics. As illustrated in Fig. 4C, the performance analysis across different CYP isoforms reveals distinct patterns for SOM prediction. The major CYP enzymes including CYP1A2, CYP2D6, and CYP3A4 demonstrate remarkable performance in terms of AUC and PRC metrics, achieving AUC scores of 0.92,

0.93, and 0.94 respectively in the single-task framework. However, enzymes with limited substrate data, such as CYP2E1, failed to achieve an AUC score exceeding 0.9, presumably due to data scarcity constraints affecting model performance. Notably, the implementation of multi-task learning substantially improved the performance on these data-limited tasks, with CYP2E1 showing a significant enhancement of 4% in AUC score. Similar performance improvements were consistently observed in the PRC curves across these isoforms. This consistent reduction in variance, coupled with improved performance metrics, strongly indicates that the multi-task learning strategy not only significantly enhances various aspects of model performance in SOM prediction but also substantially improves the stability of models.

To further enhance the reliability of models, we implemented an optimized data split strategy for the SOM dataset. While the initial approach solely considered substrate proportions, our refined methodology ensures proportional representation of various reaction types across all folds (detailed optimization procedures are documented in the SI (Section S2.2)). This optimization strategy resulted in a more balanced dataset distribution. As demonstrated in Fig. 4A, this refined methodology yielded improvements across all weighted average evaluation metrics. Notably, the PRC-R metric showed the most substantial enhancement, with an improvement of approximately 1%. The multi-task model incorporating optimized data partitioning demonstrated superior performance, achieving significant improvements across all evaluation metrics. This comprehensive enhancement suggests that the combination of multi-task learning and optimized data partitioning creates a synergistic effect, leading to more robust and reliable model performance for SOM prediction.

Furthermore, we conducted comprehensive ablation studies to investigate the impact of multi-scale information integration on the performance of SOM prediction. Based on the multi-task framework, we designed three experimental groups categorized by information source: (1) DeepMetab without atom-level reactivity descriptors and global molecular descriptors (DeepMetab (w/o atom, mol)), (2) DeepMetab without global molecular descriptors (DeepMetab (w/o mol)), (3) DeepMetab without atom-level reactivity descriptors (DeepMetab (w/o atom)), (4) full integration of baseline, atomic, and molecular descriptors, DeepMetab. Fig. 4B demonstrates that both the incorporation of atom-level reactivity descriptors and molecular descriptors outperforms the DeepMetab (w/o atom, mol) across all evaluation metrics. Furthermore, DeepMetab achieves the best performance on all metrics except for Jaccard, where it trails behind DeepMetab (w/o mol) by only 0.4%, which highlights that the integration of the multi-scale information into the GNN framework is important for reliable prediction of SOM.

DeepMetab outperforms other prediction tools for comprehensive drug metabolism prediction

We first compared the substrate prediction performance of DeepMetab with existing models (CYPstrate and CypReact)

using the testing dataset. The evaluation employed two metrics: ACC and Jaccard index, with results presented in Fig. 5A. Fig. 5a1 shows the average performance across all CYP450 isoforms, where DeepMetab demonstrated modest improvements over both CYPstrate and CypReact. Radar plots in Fig. 5a2–3 detail the isoform-specific comparisons, revealing consistent superiority of DeepMetab in ACC metrics across almost all isoforms and overall better performance in Jaccard indices.

Then, we conducted comprehensive external validation of the SOM module by comparing DeepMetab's predictive capabilities with established models in the field, including FAME3, SMARTCyp3.0, and the CYP450 component (CyProduct) of BiTransformer3.0. Since FAME3 does not differentiate between CYP450 isoforms, the comparison focuses on overall CYP450 metabolic site prediction capability (Fig. 5b1). DeepMetab consistently outperformed FAME3 across all metrics, with particularly notable improvements of approximately 10% in TOP-1 and TOP-2 metrics (Fig. 5b1). Fig. 5b2 presents a weighted average comparison with CyProduct, where weights were assigned based on substrate quantities for each enzyme isoform. DeepMetab demonstrated superior performance across all metrics, with a particularly significant advantage exceeding 10% in the TOP-1 and TOP-2 metrics. The comparison with SMARTCyp3.0, shown in Fig. 5b3, focuses on the three CYP450 isoforms (CYP2C9, CYP2D6, and CYP3A4) that SMARTCyp3.0 can predict. The weighted average results across these three enzymes demonstrate DeepMetab's substantial advantages, with performance improvements exceeding 10% in multiple metrics (Jaccard, TOP-1 and TOP-2). To provide detailed insights into isoform-specific performance, Fig. 5b4–8 presents radar plots comparing performances across different CYP450 isoforms. DeepMetab exhibited superior performance across most isoforms, with particularly notable improvements in low-data isoforms such as CYP2A6, CYP2B6, and CYP2C8. Substantial improvements were also observed across other isoform-specific tasks.

To further evaluate DeepMetab's rule coverage and metabolite prediction accuracy, we conducted a comparative analysis with two other rule-based metabolite prediction platforms, with detailed comparison methodology and specific results presented in the SI (Section S3.3). The results are presented in Fig. 5c, where performance was assessed using Jaccard, TOP-1, and TOP-2 metrics for the final metabolite predictions. Fig. 5c1 illustrates the overall comparison of CYP450 metabolite predictions without isoform differentiation, while Fig. 5c2–4 present isoform-specific comparisons across different CYP450 isoforms. The analysis demonstrates DeepMetab's overall superior overall performance, with only a relatively notable exception in the Jaccard metric for CYP2D6. In all other cases, DeepMetab either significantly outperformed or matched the competing platforms, with particularly impressive advantages in TOP-1 and TOP-2 metrics, where improvements exceeded 10%. These results not only validate DeepMetab's excellence in metabolic site prediction but also demonstrate the robustness of our metabolic rule system. The consistent superior performance across multiple metrics and isoforms underscores the



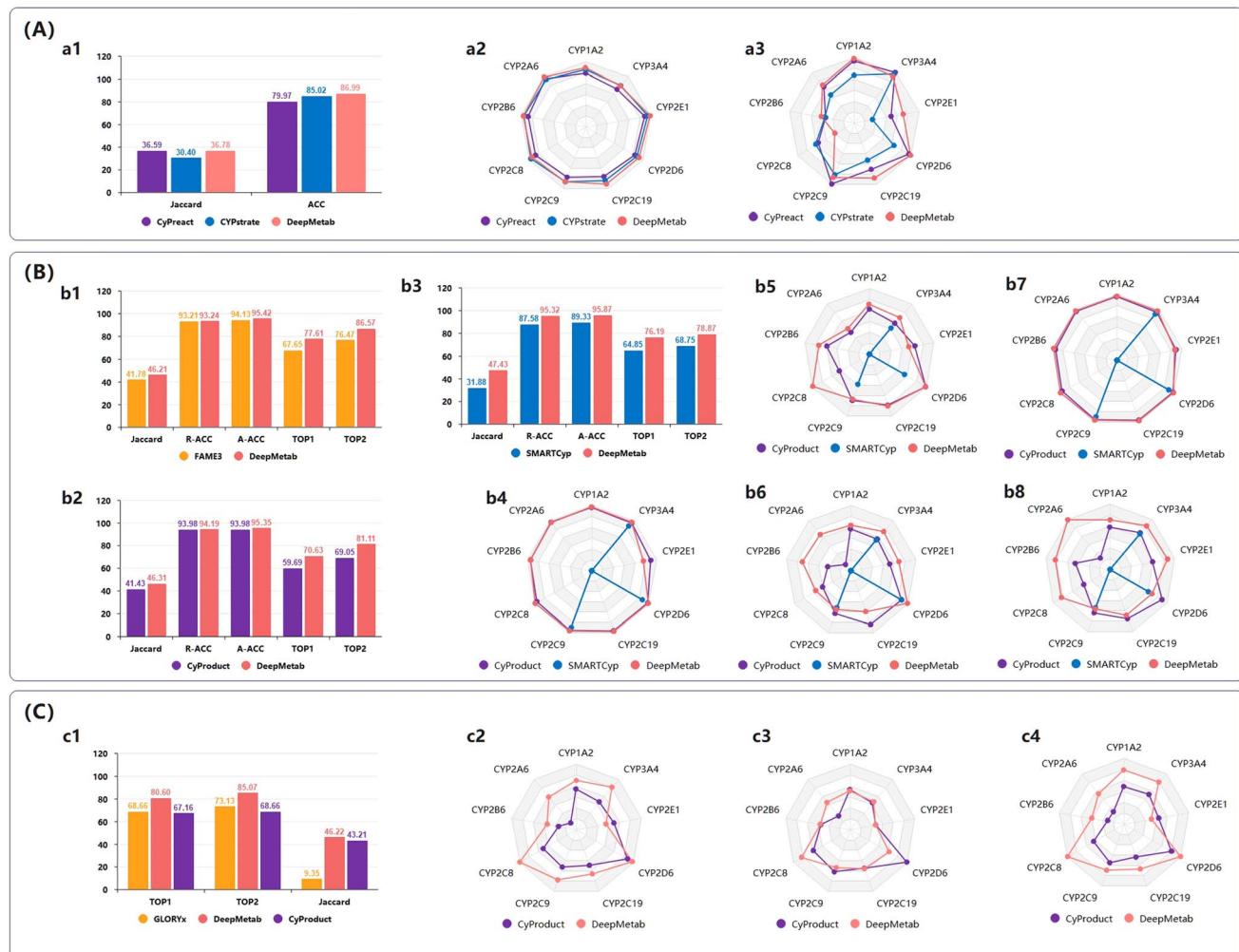


Fig. 5 Comprehensive performance evaluation of DeepMetab against established prediction tools. (A) Substrate prediction model comparative analysis. (a1) Aggregated performance metrics (mean values) across all evaluated CYP450 isoforms. (a2) and (a3) Isoform-specific ACC (a2) and Jaccard (a3) comparison with CyProduct and CYPstrate. (B) SOM model comparative assessment. (b1)–(b3) Substrate-weighted average performance metrics comparing DeepMetab with FAME3 (b1), CyProduct (b2), and SMARTCyp3.0 (b3). (b4)–(b8) Detailed isoform-specific performance comparisons utilizing multiple evaluation metrics: ACC-R (b4), Jaccard (b5), TOP-1 (b6), ACC-A (b7), and TOP-2 (b8). (C) Metabolite generation rules effectiveness comparison. (c1) Comprehensive evaluation of metabolite prediction performance against GLORYx and CyProduct. (c2)–(c4) Comparative analysis of prediction performance across CYP450 isoforms against GLORYx and CyProduct using TOP-1 (c2), Jaccard (c3), and TOP-2 (c4) evaluation metrics.

comprehensive nature of our approach to metabolite prediction.

DeepMetab reveals complex chemical environment learning of metabolic sites

Chemical environments, determined by electronic, steric, and stereo electronic characteristics, are crucial for predicting metabolic sites. Our graph-based message-passing neural network effectively learns these features for accurate predictions. To further analyze DeepMetab's learning and uncover patterns of metabolic sites, we visualized the hidden representations of the SOM module, as shown in Fig. 6. Fig. 6A illustrates the t-SNE clustering results of hidden layer representations for atoms in various metabolic site environments from a random selection of over 300 molecules in the SOM training set. The

color scheme represents the cosine similarity between atoms in specific environments and actual metabolic atoms (shown in the SI (Section S3.4)), with deeper blue indicating a higher likelihood of being a metabolic atom. The visualization analysis reveals several significant patterns: aromatic carbon atoms, particularly in *ortho*- and *para*-positions, predominantly display blue coloration, aligning with empirical knowledge and expert rules regarding phenolic hydroxylation reactions. α -Carbon positions exhibit a distinctive bimodal distribution, with detailed case analysis (Fig. 6C) revealing that red-colored sites adjacent to the $-\text{SO}_2-$ group may be metabolically inhibited due to high polarity (Case 1 and 2). Sulfur and phosphorus atoms show approximately equal distribution between metabolic and non-metabolic tendencies, with Case 5 and Case 6 indicating that red-colored phosphorus sites typically represent atoms at

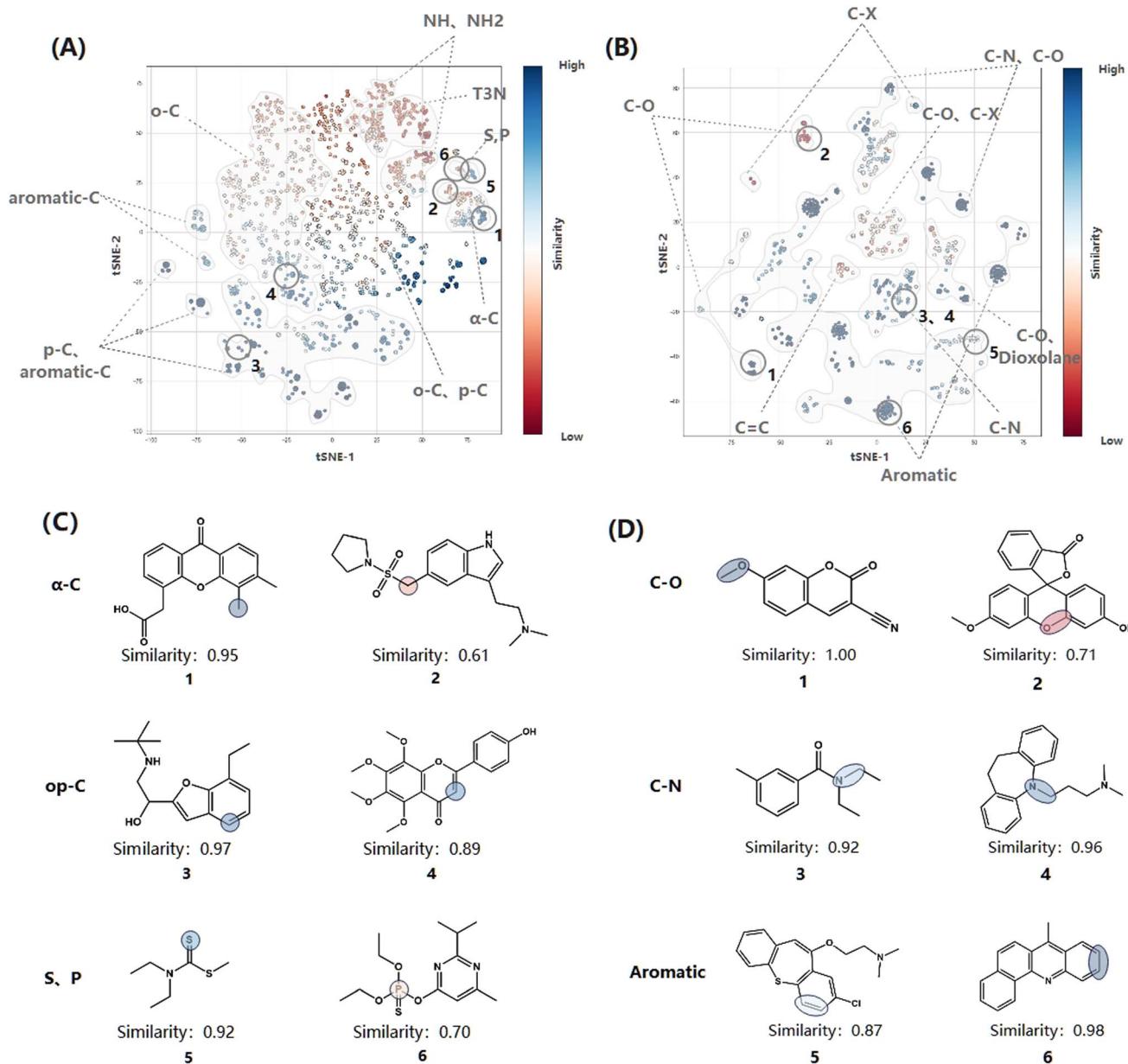


Fig. 6 Model interpretation. (A) t-SNE visualization of atom hidden layer representations, showing clusters of NH_2 (primary N), NH (secondary N), T3N (tertiary N), S (sulfur), P (phosphorus), $\alpha\text{-C}$ (alpha carbon), o-C (ortho carbon), and p-C (para carbon). Colors indicate COS similarity scores (avg. top 10) with actual metabolic sites. (B) t-SNE visualization of bond hidden layer representations, displaying clusters of C-O (carbon–oxygen single bond), C-N (carbon–nitrogen single bond), C=C (carbon–carbon double bond), C-X (carbon–halogen single bond), aromatic (aromatic bond), and dioxolane (C-O bonds in the dioxolane structure). Colors indicate COS similarity scores (avg. top 20) with actual metabolic sites. (C) Representative case studies of atoms in their corresponding chemical environments. (D) Representative case studies of bonds in their corresponding chemical environments.

their highest oxidation state, making further metabolic reactions thermodynamically unfavorable. Additionally, large regions of quaternary nitrogen (T3N) sites display red coloration, correlating with the empirical observation that quaternary nitrogen atoms, lacking hydrogen atoms, are generally less susceptible to oxidation.

Fig. 6B presents the visualization analysis of chemical bonds across various environmental contexts, employing the same methodology. The analysis reveals that C-N bonds and aromatic bonds predominantly exhibit blue coloration, indicating higher

predicted reactivity. Similarly, certain C-O , C-X , and aromatic bonds demonstrate elevated reactivity potential. Specific examples illustrated in Fig. 6D provide deeper insights into these patterns. For instance, the red coloration of certain C-O bonds can be attributed to the absence of available hydrogen atoms on the carbon center (Case 2), which precludes the empirically common dealkylation reactions at these sites. Conversely, other examples highlighted in the figure represent sites with historically documented high metabolic reactivity (Case 1, 3, 4). In addition, Case 5 and Case 6 are examples of

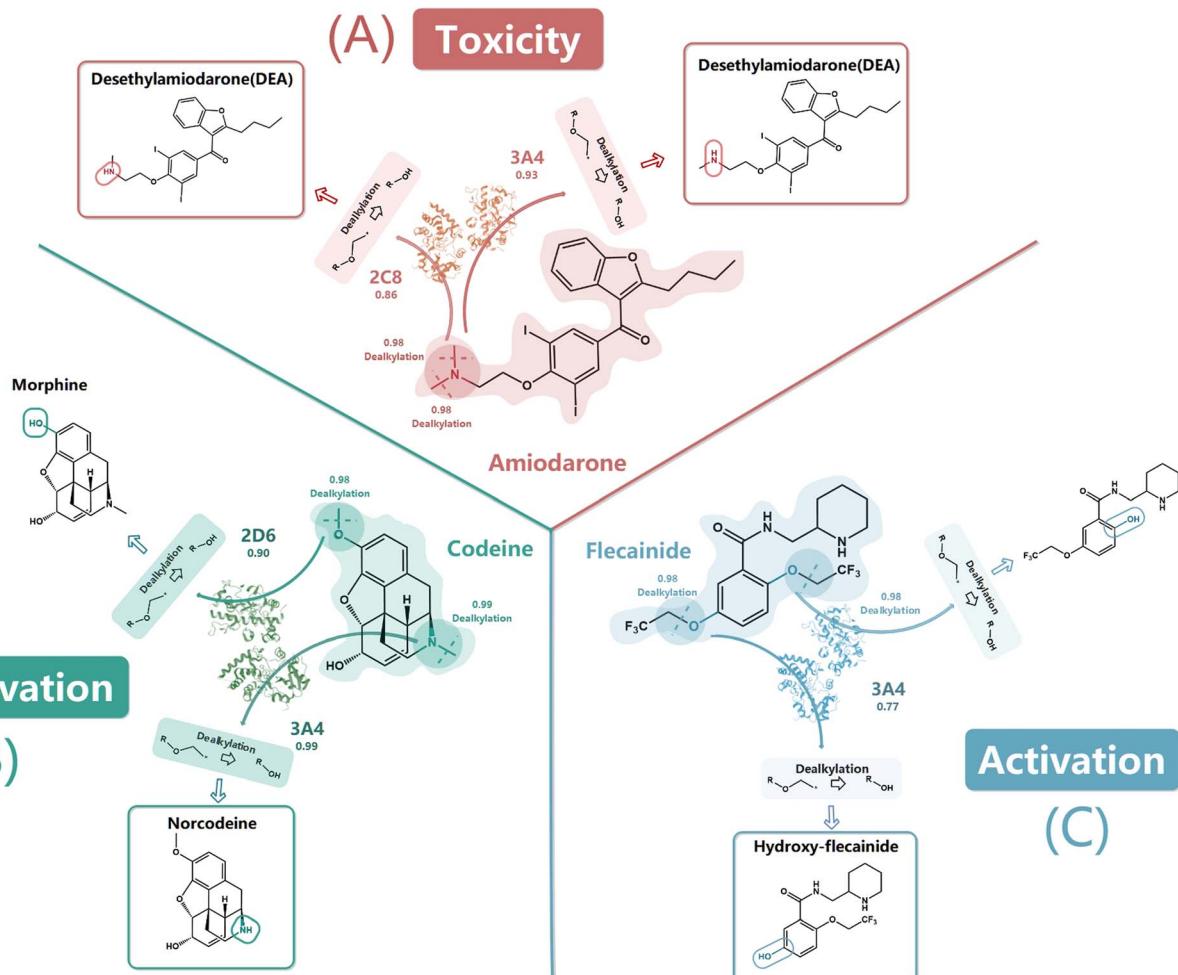


Fig. 7 Comprehensive visualization of DeepMetab's metabolic pathway predictions for three drug cases. The diagram illustrates predicted drug metabolism patterns categorized by functional outcomes. (A) Toxicity pathway of amiodarone leading to desethylamiodarone (DEA), (B) inactivation pathway of codeine metabolizing to norcodeine, and (C) activation pathway of flecainide transforming into hydroxy-flecainide. Key SOMs are highlighted within each molecular structure, with prediction scores displayed adjacent to each metabolic transformation. The involvement of specific CYP450 isoforms is mapped to each reaction pathway, demonstrating the model's ability to predict both the metabolites and the responsible enzymes in drug metabolism.

aromatic carbon–carbon bonds with high similarity scores. Based on empirical knowledge, epoxidation typically occurs at this site. The slightly lower similarity in Case 5 compared to Case 6 may be due to the influence of the electron-withdrawing chlorine substituent on the benzene ring.

Through this comprehensive hidden layer visualization analysis, we discovered that the model can deeply learn and understand the complex chemical environments of metabolic sites, thereby developing expert-like “insight” for accurately predicting metabolic sites. This visualization not only uncovers multiple potential patterns in CYP450-catalyzed metabolism but also significantly enhances the interpretability of the model.

Case studies of clinically approved drugs

To demonstrate the practical utility of DeepMetab, we investigated its application in several clinically relevant scenarios. To further highlight the significance of predicting functional drug metabolites, we carefully selected a range of classic cases

encompassing key scenarios such as metabolic toxicity, metabolic inactivation, and metabolic activation. Notably, these compounds possess novel structures that were not included in the training dataset, underscoring the model's robust generalization ability and its capacity to address critical challenges in drug metabolism prediction.

First, we examined amiodarone (AMD), an FDA-approved antiarrhythmic drug (1985) known for its potential hepatotoxicity risk,³⁸ where CYP450-mediated metabolites are suspected to be significant contributors to liver injury. DeepMetab identified CYP2C8 and CYP3A4 as the responsible enzymes (with scores of 0.86 and 0.93), predicted the metabolic site with a probability of 0.98, and – by applying the corresponding *N*-dealkylation biotransformation rule – accurately anticipated the formation of desethylamiodarone (DEA). This metabolite has been experimentally validated by Shohei *et al.*³⁹ to induce hepatotoxicity through oxidative stress mechanisms. To further evaluate DeepMetab's capability in predicting functionally significant metabolites, we analyzed codeine metabolism (Fig. 7B). Codeine, one of

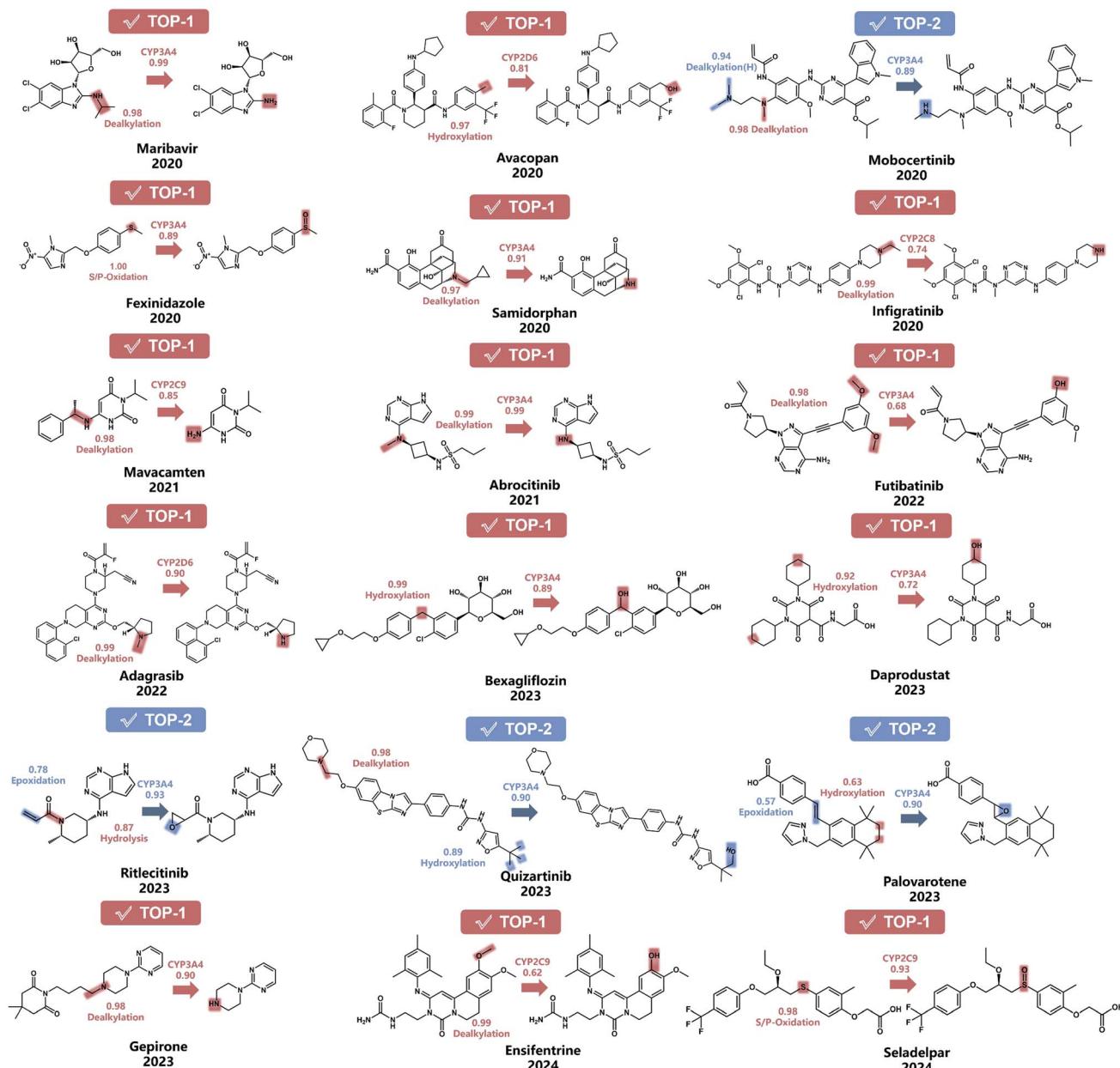


Fig. 8 The prediction for orally administered FDA-approved drugs (2020–2024). The figure presents the entire DeepMetab prediction workflow, where predicted metabolic enzymes and their associated scores are annotated in text, while predicted metabolic sites are distinctly highlighted using a light color. The figure also specifies the prediction probabilities and the categories of applied metabolic transformation rules. Specifically, TOP-1 indicates that the true metabolite ranks first among the predictions, while TOP-2 denotes that it appears within the top two predictions.

the earliest opioid medications, undergoes metabolic transformations that are critical for its therapeutic effects.⁴⁰ Specifically, DeepMetab accurately predicted the formation of morphine, the CYP2D6-mediated active metabolite included in the training dataset. Moreover, DeepMetab successfully predicted norcodeine, the inactive metabolite primarily formed *via* CYP3A4 – with enzyme and metabolic site prediction scores both at 0.99 – by applying the corresponding dealkylation biotransformation rule, despite this metabolite being absent from the training data. Additionally, we investigated flecainide, a classical sodium channel blocker first marketed in Europe in 1982. Its metabolism involves CYP450-mediated pathways, with CYP3A4 being the

primary enzyme responsible for the formation of hydroxy-flecainide. This metabolite retains similar electrophysiological properties to those of the parent drug, as demonstrated in previous studies.⁴¹ DeepMetab accurately predicted the formation of hydroxy-flecainide, identifying CYP3A4 as the responsible enzyme with a score of 0.77 and the metabolic site with a score of 0.98, before successfully applying the dealkylation biotransformation rule-highlighting CYP3A4's role in the dealkylation process (Fig. 7C). These applications demonstrate DeepMetab's reliability and practical value in predicting bioactive metabolites beyond the training dataset and show excellent generalization ability across diverse chemical structures and metabolic pathways.

Retrospective validation of DeepMetab based on the latest approved drugs by FDA

To further assess the predictive capability of DeepMetab for the metabolites of previously uncharacterized drug molecules, we conducted a validation study using oral drugs approved by the FDA within the past four years. Importantly, these drugs were not part of the training data used to develop the DeepMetab model, ensuring an independent evaluation of its performance. For this step, we utilized DeepMetab to predict the CYP-mediated metabolism of these novel drugs. A significant challenge in this task is that the metabolic product information for most newly approved drugs remains undisclosed, and even when available, it typically includes only the most critical 1–2 metabolites rather than offering a comprehensive profile. To address this constraint, we identified 18 drugs from the literature that have clearly documented metabolic pathways and associated metabolites.^{42–59} Using these as benchmarks, we systematically evaluated the predictions made by DeepMetab, with the results presented in Fig. 8.

DeepMetab accurately predicted the primary CYP-mediated metabolite (TOP-1) for 14 out of 18 drugs, achieving approximately 78% accuracy, while for the remaining four drugs (Mobocertinib, Ritlecitinib, Quizartinib, and Palovarotene), the correct metabolite appeared as the second-best prediction (TOP-2), yielding an overall TOP-2 accuracy of 100%. This performance reflects the model's effective end-to-end prediction capability: it successfully identified several major metabolizing CYP enzymes, such as CYP3A4 and CYP2D6, with scores ranging from 0.81 to 1.00. It also predicted fewer common enzymes, for example, the metabolism of Infigratinib by CYP2C8 with a score of 0.74. Concurrently, DeepMetab precisely localized the metabolic sites, often scoring above 0.9, exemplified by the dealkylation site on avacopan (0.97) and hydroxylation on daprodustat (0.92). Finally, the model applied specific metabolic transformation rules – such as dealkylation, hydroxylation, epoxidation, S/P-oxidation, and hydrolysis – to generate detailed metabolite structures consistent with enzyme and site predictions, resulting in accurate and comprehensive metabolic predictions. This high performance highlights not only the reliability and accuracy of DeepMetab but also its ability to prioritize the most relevant metabolites of a drug. Specifically, DeepMetab excelled at ranking the most important metabolites at the top of its predictions, demonstrating their practical value in drug metabolism research. Furthermore, DeepMetab demonstrated the ability to identify and prioritize the most significant metabolites with high accuracy, even in cases where detailed metabolic pathway information was limited. These findings underscore the potential of DeepMetab to significantly advance the study of drug metabolism, particularly for new and unexplored molecules, by providing accurate and actionable predictions that can guide further experimental investigations.

Conclusions

In this study, we introduced DeepMetab, a novel end-to-end deep learning framework specifically designed to improve the prediction of CYP-mediated drug metabolism. DeepMetab tackles critical limitations of prevailing approaches—including

confinement to isolated tasks, insufficient mechanistic fidelity, lack of holistic integration, and limited chemical interpretability—as well as challenges arising from inadequate datasets and the absence of comprehensive predictive frameworks. DeepMetab significantly advances the state-of-the-art in metabolic predictions and integrates substrate profiling, SOM localization, and metabolite generation through advanced data representation techniques, achieving substantial performance improvements over existing methods while providing valuable mechanistic insights.

Despite these notable advancements, several methodological and practical limitations remain to be addressed in subsequent research. The expansion of training datasets to encompass a more diverse chemical space and rare metabolic transformation patterns would significantly enhance the model's generalizability across the pharmaceutical chemical space. Furthermore, the incorporation of more sophisticated enzymatic interaction mechanisms and phase II conjugation pathways would yield a more comprehensive metabolic prediction system that better reflects the complexity of *in vivo* drug biotransformation.

Future methodological enhancements will concentrate on integrating additional cytochrome P450 isoforms and non-CYP metabolic enzyme systems to facilitate comprehensive metabolic prediction and construction of more holistic drug metabolism networks. We intend to extend DeepMetab's predictive domain to encompass a broader spectrum of metabolic enzyme systems, thereby providing more exhaustive metabolic prediction capabilities across multiple biotransformation pathways. Concurrently, we will develop an intuitive web-based interface to facilitate accessibility for researchers and pharmaceutical scientists, enabling seamless integration of these computational predictions into established drug discovery and development workflows. Collectively, these advances position DeepMetab to compress preclinical timelines by rapidly surfacing metabolic hotspots and high-risk metabolites, thereby guiding targeted *in vitro* follow-ups (e.g., isoform panels, TDI/reactivity screens) and rational prioritization of metabolically favorable scaffolds; in turn, earlier identification of CYP-mediated bioactivation and reactive species is expected to curtail late-stage attrition, while mechanism-aware, end-to-end predictions—informing SOM-blocking substitutions, soft-spot hopping, and isoform-selective optimization—enable safer compound design and tighter integration into DMTA workflows, ultimately improving decision quality across discovery and development.

Author contributions

Yiling Zhou and Dongsheng Cao designed the research study. Yiling Zhou developed the method and wrote the code. Yiling Zhou and Xiao Wei performed the analysis. Yiling Zhou, Dejun Jiang, Yi Ding, Jiabei Yi, and Yikun Wang wrote and checked the paper. All authors read and approved the manuscript.

Conflicts of interest

There are no conflicts to declare.



Data availability

The source data, running code and trained models are available at our GitHub repository: <https://github.com/YilingZhou/DeepMetab>.

Supplementary information is available. See DOI: <https://doi.org/10.1039/d5sc04631a>.

Acknowledgements

This work was financially supported by the National Science and Technology Major Project of the Ministry of Science and Technology of China [2023ZD0507104], National Natural Science Foundation of China (22220102001, 22307112, 82304316), The Center for Computational Biology and Bioinformatics, Furong Laboratory and Bioinformatics Center, Xiangya Hospital, Central South University. We acknowledge Haikun Xu, and the High-Performance Computing Center of Central South University for support.

References

- 1 J. Lee, J. L. Beers, R. M. Geffert and K. D. Jackson, A Review of CYP-Mediated Drug Interactions: Mechanisms and In Vitro Drug-Drug Interaction Assessment, *Biomolecules*, 2024, **14**, 99.
- 2 J. Shanu-Wilson, L. Evans, S. Wrigley, J. Steele, J. Atherton and J. Boer, Biotransformation: Impact and Application of Metabolism in Drug Discovery, *ACS Med. Chem. Lett.*, 2020, **11**, 2087–2107.
- 3 I. J. Onakpoya, C. J. Heneghan and J. K. Aronson, Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature, *BMC Med.*, 2016, **14**, 10.
- 4 N. Méndez-Sánchez, M. Arrese, D. Zamora-Valdés and M. Uribe, Current concepts in the pathogenesis of nonalcoholic fatty liver disease, *Liver Int.*, 2007, **27**, 423–433.
- 5 U. M. Zanger and M. Schwab, Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation, *Pharmacol. Ther.*, 2013, **138**, 103–141.
- 6 B. Hossam Abdelmonem, N. M. Abdelaal, E. K. E. Anwer, A. A. Rashwan, M. A. Hussein, Y. F. Ahmed, R. Khashana, M. M. Hanna and A. Abdelnaser, Decoding the Role of CYP450 Enzymes in Metabolism and Disease: A Comprehensive Review, *Biomedicines*, 2024, **12**(7), 1467.
- 7 Y. Wei, L. Palazzolo, O. Ben Mariem, D. Bianchi, T. Laurenzi, U. Guerrini and I. Eberini, Investigation of in silico studies for cytochrome P450 isoforms specificity, *Comput. Struct. Biotechnol. J.*, 2024, **23**, 3090–3103.
- 8 Y. Feng, M. D. Cameron, B. Frackowiak, E. Griffin, L. Lin, C. Ruiz, T. Schröter and P. LoGrasso, Structure-activity relationships, and drug metabolism and pharmacokinetic properties for indazole piperazine and indazole piperidine inhibitors of ROCK-II, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 2355–2360.
- 9 R. Hyland, B. C. Jones and D. A. Smith, Identification of the Cytochrome P450 Enzymes Involved in the N-Oxidation of Voriconazole, *Drug Metab. Dispos.*, 2003, **31**, 540.
- 10 D. W. Lee, R. Gardner, D. L. Porter, C. U. Louis, N. Ahmed, M. Jensen, S. A. Grupp and C. L. Mackall, Current concepts in the diagnosis and management of cytokine release syndrome, *Blood*, 2014, **124**, 188–195.
- 11 R. K. Handa, J. W. Harding and S. M. Simasko, Characterization and Function of the Bovine Kidney Epithelial Angiotensin Receptor Subtype 4 Using Angiotensin IV and Divalinal Angiotensin IV as Receptor Ligands, *J. Pharmacol. Exp. Ther.*, 1999, **291**, 1242.
- 12 Y. Li, Q. Meng, M. Yang, D. Liu, X. Hou, L. Tang, X. Wang, Y. Lyu, X. Chen, K. Liu, A.-M. Yu, Z. Zuo and H. Bi, Current trends in drug metabolism and pharmacokinetics, *Acta Pharm. Sin. B*, 2019, **9**, 1113–1144.
- 13 K. Zhu, M. Huang, Y. Wang, Y. Gu, W. Li, G. Liu and Y. Tang, MetaPredictor: in silico prediction of drug metabolites based on deep language models with prompt engineering, *Briefings Bioinf.*, 2024, **25**(5), bbae374.
- 14 C. de Bruyn Kops, M. Šicho, A. Mazzolari and J. Kirchmair, GLORYx: Prediction of the Metabolites Resulting from Phase 1 and Phase 2 Biotransformations of Xenobiotics, *Chem. Res. Toxicol.*, 2021, **34**, 286–299.
- 15 D. S. Wishart, S. Tian, D. Allen, E. Oler, H. Peters, V. W. Lui, V. Gautam, Y. Djoumbou-Feunang, R. Greiner and T. O. Metz, BioTransformer 3.0-a web server for accurately predicting metabolic transformation products, *Nucleic Acids Res.*, 2022, **50**, W115–W123.
- 16 J. D. Tyzack and J. Kirchmair, Computational methods and tools to predict cytochrome P450 metabolism for drug discovery, *Chem. Biol. Drug Des.*, 2019, **93**, 377–386.
- 17 S. Tian, Y. Djoumbou-Feunang, R. Greiner and D. S. Wishart, CypReact: A Software Tool for in Silico Reactant Prediction for Human Cytochrome P450 Enzymes, *J. Chem. Inf. Model.*, 2018, **58**, 1282–1291.
- 18 M. Holmer, C. de Bruyn Kops, C. Stork and J. Kirchmair, CYPstrate: A Set of Machine Learning Models for the Accurate Classification of Cytochrome P450 Enzyme Substrates and Non-Substrates, *Molecules*, 2021, **26**(15), 4678.
- 19 L. Olsen, M. Montefiori, K. P. Tran and F. S. Jørgensen, SMARTCyp 3.0: enhanced cytochrome P450 site-of-metabolism prediction server, *Bioinformatics*, 2019, **35**, 3174–3175.
- 20 M. Šicho, C. Stork, A. Mazzolari, C. de Bruyn Kops, A. Pedretti, B. Testa, G. Vistoli, D. Svozil and J. Kirchmair, 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.
- 21 S. Tian, X. Cao, R. Greiner, C. Li, A. Guo and D. S. Wishart, CyProduct: A Software Tool for Accurately Predicting the Byproducts of Human Cytochrome P450 Metabolism, *J. Chem. Inf. Model.*, 2021, **61**, 3128–3140.
- 22 L. Olsen, M. Montefiori, K. P. Tran and F. S. Jørgensen, SMARTCyp 3.0: enhanced cytochrome P450 site-of-



metabolism prediction server, *Bioinformatics*, 2019, **35**, 3174–3175.

23 V. Porokhin, L. P. Liu and S. Hassoun, Using graph neural networks for site-of-metabolism prediction and its applications to ranking promiscuous enzymatic products, *Bioinformatics*, 2023, **39**(3), btad089.

24 M. Öeren, P. J. Walton, J. Suri, D. J. Ponting, P. A. Hunt and M. D. Segall, Predicting Regioselectivity of AO, CYP, FMO, and UGT Metabolism Using Quantum Mechanical Simulations and Machine Learning, *J. Med. Chem.*, 2022, **65**, 14066–14081.

25 S. Plus, *ADMET Predictor*, <https://www.simulations-plus.com/software/admetpredictor/metabolism/>.

26 J. Zaretzki, M. Matlock and S. J. Swamidass, XenoSite: Accurately Predicting CYP-Mediated Sites of Metabolism with Neural Networks, *J. Chem. Inf. Model.*, 2013, **53**, 3373–3383.

27 N.-N. Wang, X.-G. Wang, G.-L. Xiong, Z.-Y. Yang, A.-P. Lu, X. Chen, S. Liu, T.-J. Hou and D.-S. Cao, Machine learning to predict metabolic drug interactions related to cytochrome P450 isozymes, *J. Cheminf.*, 2022, **14**, 23.

28 L. Fu, S. Shi, J. Yi, N. Wang, Y. He, Z. Wu, J. Peng, Y. Deng, W. Wang, C. Wu, A. Lyu, X. Zeng, W. Zhao, T. Hou and D. Cao, ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support, *Nucleic Acids Res.*, 2024, **52**, W422–W431.

29 C. Knox, M. Wilson, C. M. Klinger, M. Franklin, E. Oler, A. Wilson, A. Pon, J. Cox, N. E. L. Chin, S. A. Strawbridge, M. Garcia-Patino, R. Kruger, A. Sivakumaran, S. Sanford, R. Doshi, N. Khetarpal, O. Fatokun, D. Doucet, A. Zubkowski, D. Y. Rayat, H. Jackson, K. Harford, A. Anjum, M. Zakir, F. Wang, S. Tian, B. Lee, J. Liigand, H. Peters, R. Q. R. Wang, T. Nguyen, D. So, M. Sharp, R. da Silva, C. Gabriel, J. Scantlebury, M. Jasinski, D. Ackerman, T. Jewison, T. Sajed, V. Gautam and D. S. Wishart, DrugBank 6.0: the DrugBank Knowledgebase for 2024, *Nucleic Acids Res.*, 2024, **52**, D1265–D1275.

30 A. Chang, L. Jeske, S. Ulbrich, J. Hofmann, J. Koblitz, I. Schomburg, M. Neumann-Schaal, D. Jahn and D. Schomburg, BRENDA, the ELIXIR core data resource in 2021: new developments and updates, *Nucleic Acids Res.*, 2021, **49**, D498–D508.

31 P. Lee, *Handbook of Metabolic Pathways of Xenobiotics*, Wiley, 2015.

32 *Cytochromes P450: Role in the Metabolism and Toxicity of Drugs and other Xenobiotics*, The Royal Society of Chemistry, 2008.

33 L. Ridder and M. Wagener, SyGMA: combining expert knowledge and empirical scoring in the prediction of metabolites, *ChemMedChem*, 2008, **3**, 821–832.

34 RDKit: Open-Source Cheminformatics Software, <https://www.rdkit.org/>.

35 S.-C. Li, H. Wu, A. Menon, K. A. Spiekermann, Y.-P. Li and W. H. Green, When Do Quantum Mechanical Descriptors Help Graph Neural Networks to Predict Chemical Properties?, *J. Am. Chem. Soc.*, 2024, **146**, 23103–23120.

36 E. Heid, K. P. Greenman, Y. Chung, S.-C. Li, D. E. Graff, F. H. Vermeire, H. Wu, W. H. Green and C. J. McGill, Chemprop: A Machine Learning Package for Chemical Property Prediction, *J. Chem. Inf. Model.*, 2024, **64**, 9–17.

37 M. Buda, A. Maki and M. A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, *Neural Netw.*, 2018, **106**, 249–259.

38 M. Chen, V. Vijay, Q. Shi, Z. Liu, H. Fang and W. Tong, FDA-approved drug labeling for the study of drug-induced liver injury, *Drug Discovery Today*, 2011, **16**, 697–703.

39 S. Takai, S. Oda, K. Tsuneyama, T. Fukami, M. Nakajima and T. Yokoi, Establishment of a mouse model for amiodarone-induced liver injury and analyses of its hepatotoxic mechanism, *J. Appl. Toxicol.*, 2016, **36**, 35–47.

40 Q. Y. Yue and J. Säwe, Different effects of inhibitors on the O- and N-demethylation of codeine in human liver microsomes, *Eur. J. Clin. Pharmacol.*, 1997, **52**, 41–47.

41 J. Guehler, C. C. Gornick, H. G. Tobler, A. Almquist, J. R. Schmid, D. W. Benson Jr and D. G. Benditt, Electrophysiologic effects of flecainide acetate and its major metabolites in the canine heart, *Am. J. Cardiol.*, 1985, **55**, 807–812.

42 K. Sun, K. Ilic, P. Xu, R. Ye, J. Wu and I. H. Song, Effect of Food, Crushing of Tablets, and Antacid Coadministration on Maribavir Pharmacokinetics in Healthy Adult Participants: Results From 2 Phase 1, Open-Label, Randomized, Crossover Studies, *Clin. Pharmacol. Drug Dev.*, 2024, **13**, 644–654.

43 S. Miao, P. Bekker, D. Armas, M. Lor, R. Hanada, S. Okamura, Y. Umezawa and A. Trivedi, Food Effect and Pharmacokinetic Bridging of Avacopan in Caucasian and Japanese Healthy Participants, *Clin. Pharmacol. Drug Dev.*, 2024, **13**, 1011–1023.

44 F. Gonzalvez, S. Vincent, T. E. Baker, A. E. Gould, S. Li, S. D. Wardwell, S. Nadworny, Y. Ning, S. Zhang, W. S. Huang, Y. Hu, F. Li, M. T. Greenfield, S. G. Zech, B. Das, N. I. Narasimhan, T. Clackson, D. Dalgarno, W. C. Shakespeare, M. Fitzgerald, J. Chouitar, R. J. Griffin, S. Liu, K. K. Wong, X. Zhu and V. M. Rivera, Mobocertinib (TAK-788): A Targeted Inhibitor of EGFR Exon 20 Insertion Mutants in Non-Small Cell Lung Cancer, *Cancer Discovery*, 2021, **11**, 1672–1687.

45 E. Torreele, B. Bourdin Trunz, D. Tweats, M. Kaiser, R. Brun, G. Mazué, M. A. Bray and B. Pécul, Fexinidazole – a new oral nitroimidazole drug candidate entering clinical development for the treatment of sleeping sickness, *PLOS Neglected Trop. Dis.*, 2010, **4**, e923.

46 FDA Approved Drug Products: LYBALVI (olanzapine and samidorphan) tablets, https://www.accessdata.fda.gov/drugsatfda_docs/label/2021/213378s000lbl.pdf.

47 S. Liu, S. Zhao, X. Zhang, E. Chun Yong Chan, Z. Wang, H. Li and X. Tian, Identification of the human Cytochrome P450 enzymes (P450s) responsible for metabolizing infigratinib to its pharmacologically active Metabolites, BHS697, and CQM157, and assessment of their in vitro inhibition of P450s and UDP-glucuronosyltransferases (UGTs), *Biochem. Pharmacol.*, 2024, **226**, 116390.



48 M. P. Grillo, J. C. L. Erve, R. Dick, J. P. Driscoll, N. Haste, S. Markova, P. Brun, T. J. Carlson and M. Evanchik, In vitro and in vivo pharmacokinetic characterization of mavacamten, a first-in-class small molecule allosteric modulator of beta cardiac myosin, *Xenobiotica*, 2019, **49**, 718–733.

49 *FDA Approved Drug Products: CIBINQO (abrocitinib) tablets, for oral use*, [https://cdn\(pfizer.com/pfizercom/USPI_Med_Guide_CIBINQO_Abrocitinib_tablet.pdf](https://cdn(pfizer.com/pfizercom/USPI_Med_Guide_CIBINQO_Abrocitinib_tablet.pdf).

50 I. Yamamiya, A. Hunt, F. Yamashita, D. Sonnichsen, T. Muto, Y. He and K. A. Benhadji, Evaluation of the Mass Balance and Metabolic Profile of Futibatinib in Healthy Participants, *Clin. Pharmacol. Drug Dev.*, 2023, **12**, 927–939.

51 L. Rahbaek, C. Cilliers, C. J. Wegerski, N. Nguyen, J. Otten, L. Hargis, M. A. Marx, J. G. Christensen and J. Q. Tran, Absorption, single-dose and steady-state metabolism, excretion, and pharmacokinetics of adagrasib, a KRAS(G12C) inhibitor, *Cancer Chemother. Pharmacol.*, 2024, **95**, 7.

52 W. Zhang, X. Li, H. Ding, Y. Lu, G. E. Stilwell, Y. D. Halvorsen and A. Welihinda, Metabolism and disposition of the SGLT2 inhibitor bexagliflozin in rats, monkeys and humans, *Xenobiotica*, 2020, **50**, 559–569.

53 G. Tai, F. Xia, C. Chen, A. Pereira, J. Pirhalla, X. Miao, G. Young, C. Beaumont and L. Chen, Investigation of the human metabolism and disposition of the prolyl hydrolase inhibitor daprodustat using IV microtracer with Enterotest bile string, *Pharmacol. Res. Perspect.*, 2023, **11**, e1145.

54 J. N. Bauman, A. C. Doran, G. M. Gualtieri, B. Hee, T. Strelevitz, M. A. Cerny, C. Banfield, A. Plotka, X. Wang, V. S. Purohit and M. E. Dowty, The Pharmacokinetics, Metabolism, and Clearance Mechanisms of Ritlecitinib, a Janus Kinase 3 and Tyrosine-Protein Kinase Family Inhibitor, in Humans, *Drug Metab. Dispos.*, 2024, **52**, 1124–1136.

55 M. Sanga, J. James, J. Marini, G. Gammon, C. Hale and J. Li, An open-label, single-dose, phase 1 study of the absorption, metabolism and excretion of quizartinib, a highly selective and potent FLT3 tyrosine kinase inhibitor, in healthy male subjects, for the treatment of acute myeloid leukemia, *Xenobiotica*, 2017, **47**, 856–869.

56 *Health Canada Product Monograph: Sohonos (palovarotene) capsules for oral use* https://pdf.hres.ca/dpd_pm/00064435.PDF.

57 A. Kaur Gill, Y. Bansal, R. Bhandari, S. Kaur, J. Kaur, R. Singh, A. Kuhad and A. Kuhad, Gepirone hydrochloride: a novel antidepressant with 5-HT1A agonistic properties, *Drugs Today*, 2019, **55**, 423–437.

58 *FDA Approved Drug Products: OHTUVAYRE (ensifentrine) inhalation suspension, for oral inhalation use*, https://www.accessdata.fda.gov/drugsatfda_docs/label/2024/217389s000lbl.pdf.

59 *FDA Approved Drug Products: LIVDELZI (seladelpar) capsules, for oral use*, https://www.accessdata.fda.gov/drugsatfda_docs/label/2024/217899s000lbl.pdf.

