



MULTITASK DEEP LEARNING FOR CYP INHIBITION, INACTIVATION, AND METABOLIC SOFT-SPOT PREDICTION

Nikolai Ivanov^{1*}, Sergey Volkov¹, Elena Morozova²

1. *Department of Pharmaceutical AI Engineering, Faculty of Medicine, Novosibirsk State University, Novosibirsk, Russia.*
2. *Department of Computational Drug Sciences, Faculty of Pharmacy, Tomsk State University, Tomsk, Russia.*

ARTICLE INFO

Received:

12 December 2025

Received in revised form:

02 March 2026

Accepted:

06 March 2026

Available online:

28 April 2026

Keywords: Multitask deep learning, Cytochrome P450, CYP inhibition, Time-dependent inactivation, Site of metabolism, Graph neural networks

ABSTRACT

CYP enzymes dominate drug clearance, and early recognition of inhibition, inactivation, and metabolic soft spots is central to modern drug design. A unified computational view of these liabilities could help prioritize safer chemical series before costly downstream testing. Existing models often treat reversible inhibition, time-dependent inactivation, and site-of-metabolism prediction as separate tasks. This separation can obscure the shared chemical determinants that drive binding, bioactivation, and metabolic transformation. This article describes a multitask deep learning model that jointly predicts CYP reversible inhibition, time-dependent inactivation, and metabolic soft-spot location from molecular structure. The objective is to use a shared molecular representation to support more consistent and data-efficient metabolic profiling. The proposed model uses a graph neural network backbone shared across three prediction heads. These heads conceptually support isoform-specific inhibition prediction, TDI risk prediction, and atom-level soft-spot localization within the same molecular framework. Conceptually, the model would be expected to improve consistency across related metabolic endpoints compared with isolated single-task systems. It could also connect molecule-level liability predictions with atom-level explanations that guide medicinal chemistry interpretation. A unified metabolic profiling model could streamline CYP liability assessment in early discovery. By combining inhibition, inactivation, and soft-spot prediction, such a model could provide a comprehensive and interpretable metabolic hazard panel from a single molecular input.

This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/), which allows others to remix, and build upon the work non commercially.

To Cite This Article: Ivanov N, Volkov S, Morozova E. Multitask Deep Learning for CYP Inhibition, Inactivation, and Metabolic Soft-Spot Prediction. *Pharmacophore*. 2026;17(2):23-33. <https://doi.org/10.51847/7FS6HA9tSy>

Introduction

Cytochrome P450 enzymes are central determinants of drug clearance, and unanticipated CYP inhibition can create clinically relevant drug–drug interaction risk. Deep learning models for CYP inhibition have therefore become increasingly important, particularly when they address multiple isoforms rather than a single enzyme [1]. Time-dependent inhibition introduces an additional hazard because the liability may arise through metabolic activation and enzyme inactivation rather than reversible binding alone [2, 3]. At the same time, metabolic soft-spot prediction is needed because atom-specific transformations can influence clearance, metabolite exposure, and the opportunity for reactive metabolite formation [4, 5].

Current computational approaches often divide reversible CYP inhibition, TDI, and site-of-metabolism prediction into separate modeling silos. Models such as CYPlebrity focus on inhibitor classification across CYP enzymes [6], whereas recent QSAR approaches treat reversible and time-dependent inhibition as related but still separately curated endpoints [7]. Site-of-metabolism models based on graph learning or bond-level prediction address atom-specific reactivity without necessarily modeling the corresponding inhibition phenotype [8, 9]. This fragmentation duplicates modeling effort and may miss cross-task signals in which the same substructure contributes to CYP binding, metabolic activation, and local oxidation susceptibility. Multitask learning offers a principled way to share molecular representations across related endpoints. In CYP modeling, multitask systems have already been used for inhibitor prediction [10] and for explainable substrate prediction across enzymes [11]. Broader ADMET modeling has also shown that multi-task graph learning can use auxiliary endpoints to support more coherent molecular representations [12], while derivative-based ADMET learning suggests that shared chemical context can

Corresponding Author: Nikolai Ivanov; Department of Pharmaceutical AI Engineering, Faculty of Medicine, Novosibirsk State University, Novosibirsk, Russia. E-mail: nikolai.ivanov@gmail.com

guide optimization across several properties [13]. These precedents motivate a unified CYP metabolism model, even though the combined prediction of reversible inhibition, TDI, and atom-level soft spots remains less developed.

The thesis of this MDL article is that a single multitask deep learning model could learn CYP inhibition, time-dependent inactivation, and metabolic soft-spot prediction from molecular structure. A shared graph encoder with attention could capture local atom environments and global molecular features that are relevant to CYP liability [14]. Recent deep learning platforms for CYP inhibition and induction show how multi-endpoint CYP panels may be handled within modern architectures [15], while comprehensive graph learning for metabolism indicates that end-to-end metabolic prediction can be mechanistically structured [16]. The intended outcome is not a claimed experimental result but a model design that could deliver an integrated, interpretable metabolic profile for medicinal chemistry use [17].

Background

CYP-Mediated Drug Metabolism and Its Consequences

Major CYP isoforms differ in substrate scope, active-site preference, and vulnerability to inhibition, making isoform-specific prediction essential for drug metabolism assessment. Web servers and deep learning platforms for CYP activity prediction frame this problem as a multi-enzyme activity profile rather than a single binary endpoint [18, 19]. Reversible inhibition reflects competitive or noncompetitive interference with enzyme function, whereas TDI may involve enzyme-catalyzed conversion of a compound into an inactivating species [2]. Metabolic soft spots represent atoms or bonds most likely to undergo biotransformation, and prediction tools beyond classic CYP enzymes illustrate how metabolism is distributed across both enzyme families and local chemical environments [20].

In-Silico Models for CYP Inhibition and Inactivation

In-silico CYP inhibition models have progressed from conventional QSAR and fingerprint-based classifiers toward deep neural systems that can learn nonlinear molecular representations. A multitask autoencoder approach for CYP inhibition prediction demonstrated how shared hidden features can support multiple CYP isoform outputs [1], while substructure pattern recognition remains valuable for linking predictions to medicinal chemistry hypotheses [21]. Recent machine learning studies emphasize molecular properties and endpoint-specific modeling for CYP inhibition [22, 23], but small datasets for particular isoforms can still limit robust model development [12]. For TDI, QSAR models that jointly consider reversible and time-dependent inhibition highlight the need for curation strategies that distinguish ordinary binding from inactivation liability [7].

Site-of-Metabolism Prediction Algorithms

Site-of-metabolism prediction has evolved from rule-based and reactivity-based reasoning toward machine learning models that score atoms or bonds within a molecular graph. CypReact and CyProduct illustrate how CYP metabolism can be represented through reactant and product prediction rather than only as an enzyme activity label [24, 25]. FAME 3 broadened site-of-metabolism prediction across phase 1 and phase 2 enzyme systems [4], and graph neural network approaches have further reframed SOM prediction as atom-level learning on molecular structures [8]. Bond-centered oxidation prediction and newer CYP metabolic site tools show that atom-specific labeling remains challenging because metabolic outcomes depend on both intrinsic reactivity and enzyme accessibility [9, 26, 27].

Multitask and Multi-Output Learning for Drug Metabolism

Multitask and multi-output learning are attractive for drug metabolism because CYP endpoints are chemically related but incompletely observed. iCYP-MFE explicitly uses multitask learning for CYP inhibitor identification, showing how enzyme-specific outputs can share molecular encodings [10]. Explainable multitask deep learning for CYP substrates similarly suggests that related CYP tasks can be learned together while preserving interpretability [11]. Broader ADMET systems, including ADMETlab 2.0, HelixADMET, adaptive auxiliary task selection, DeepDelta, and hybrid fragment-SMILES tokenization, show that shared representations can support multi-endpoint pharmacokinetic modeling beyond CYP alone [12, 13, 28-30].

Interpretability for Metabolic Hazard Assessment

Interpretability is critical because CYP liability predictions must guide chemical redesign rather than merely label a compound as risky. Multimodal CYP inhibitor prediction with explainability demonstrates how model explanations can be tied to molecular features relevant to enzyme inhibition [31]. Coloring molecules with explainable artificial intelligence offers a useful paradigm for displaying atom- or substructure-level evidence in preclinical relevance assessment [17]. Quantitative evaluation of explainable graph neural networks and early graph convolutional molecular embedding work also support the idea that learned graph features should be inspected, validated, and connected to chemically meaningful patterns [32, 33].

Model Development Overview

High-Level Unified Metabolic Profiling Pipeline

The proposed pipeline would process a molecule through a shared molecular graph encoder that produces atom-level embeddings and a graph-level summary. From this shared representation, one head could output an isoform-resolved CYP inhibition vector, a second head could output TDI probability, and a third head could generate a soft-spot heat map across atoms. Graph attention models for CYP inhibitor prediction support this kind of shared molecular processing because they can

combine local substructure evidence with whole-molecule context [14]. Comprehensive graph learning for drug metabolism further suggests that metabolism prediction can be structured as an end-to-end framework rather than a set of isolated descriptors [16].

Core Input Representations and Tasks

The core input would be a molecular graph in which atoms and bonds encode chemical identity, connectivity, aromaticity, formal charge, hybridization, and other features relevant to CYP recognition. The first task would represent CYP inhibition as multi-label classification or activity regression across isoforms, consistent with multitask inhibitor modeling [1, 10]. The second task would classify TDI potential using structural and molecular evidence associated with inactivation risk, following the conceptual direction of CYP3A4 TDI modeling and broader reversible/TDI QSAR work [2, 7]. The third task would score non-hydrogen atoms as potential metabolic soft spots, drawing on SOM predictors that learn atom- or bond-level metabolic susceptibility [4, 8, 9].

Design Principles

The central design principle is that a shared encoder should exploit correlations among CYP binding, inactivation liability, and metabolic transformation while preserving task-specific outputs. Multitask CYP substrate prediction shows how shared representations can still support enzyme-specific interpretation [11], and multi-task graph learning for ADMET suggests that auxiliary tasks can be selected or weighted to benefit related endpoints [12]. The model should handle missing labels through masked losses, because many compounds will not have complete annotation across isoforms, TDI status, and SOM sites. It should also produce calibrated and interpretable predictions so that medicinal chemists can understand why a molecule is flagged rather than treating the model as a black box [17, 32].

Data Sources and Feature Engineering

CYP Inhibition and Inactivation Datasets

CYP inhibition data would be curated from public chemistry and pharmacology resources and from specialized studies that report isoform-specific inhibition labels or activity values. Multitask CYP inhibition models and CYP inhibitor classifiers provide templates for organizing multi-isoform activity labels under a shared compound representation [1, 6]. TDI data would require separate standardization because inactivation liability may depend on assay design, preincubation, and kinetic interpretation, as emphasized by CYP3A4 TDI modeling and experimental-variability comparisons [2, 3]. Recent QSAR work on reversible and time-dependent CYP inhibition further supports separating reversible inhibition labels from TDI labels while allowing them to inform a shared representation [7].

Site-of-Metabolism Data

SOM data would be compiled from metabolite identification studies, curated metabolism benchmarks, and tools that represent CYP reactions at atom, bond, or product levels. FAME 3 provides a precedent for assigning phase 1 and phase 2 metabolism labels to candidate atoms [4], while CypReact and CyProduct show how reactants and CYP metabolic products can be linked computationally [24, 25]. Graph neural network SOM prediction and bond-level oxidation modeling support the use of atom-specific or bond-specific labels as supervision for the soft-spot head [8, 9]. Quantum-mechanical or reactivity descriptors could be used as auxiliary information when available, but they should support rather than replace experimentally grounded metabolic annotations.

Data Alignment and Handling of Sparse Labels

Data alignment would map each molecule to all available labels while preserving the distinction between molecule-level and atom-level endpoints. In CYP inhibition modeling, many compounds are profiled against only a subset of isoforms, so the inhibition loss should ignore missing isoform labels rather than treating them as inactive [10, 34]. In multi-endpoint ADMET learning, adaptive auxiliary task selection and self-supervised knowledge transfer provide useful precedents for learning from incomplete and heterogeneous datasets [12, 29]. For SOM, sparse atom labels could be augmented through metabolism-specific pretraining or transfer learning, but experimentally observed metabolic sites should remain the primary target for final supervision [5, 16].

Multitask Deep Learning Architecture

Shared Molecular Graph Encoder

The shared encoder would be implemented as a graph attention network or message-passing neural network that updates atom embeddings using neighboring atoms, bonds, and learned chemical context. Graph convolutional molecular embeddings provide the conceptual foundation for learning property-relevant molecular representations directly from attributed molecular graphs [33]. CYP inhibition models using graph convolution and attention show how such encoders can represent enzyme-relevant substructures while retaining global molecular context [14]. For metabolism, an end-to-end graph learning framework suggests that atom-level and molecule-level metabolic signals can be learned within the same structural representation [16].

Task-Specific Prediction Heads

The CYP inhibition head would generate isoform-specific outputs for major CYP enzymes, allowing each output to specialize while sharing the same upstream molecular representation. Existing multitask CYP inhibitor models support this design by assigning separate outputs to related CYP endpoints rather than building fully independent models [1, 10]. The TDI head would operate on the graph-level embedding and could be trained to flag potential inactivation liability without claiming mechanistic certainty, consistent with the conceptual distinction between reversible and time-dependent inhibition [2, 7]. The SOM head would operate on atom embeddings and assign a relative soft-spot likelihood to each candidate atom, following atom-level and bond-level metabolism prediction paradigms [4, 8, 9].

Table 1 defines the proposed multitask CYP liability architecture by linking each model component to its prediction target, representation level, supervision signal, and medicinal chemistry decision-use function.

Table 1. Multitask CYP liability architecture: endpoint structure, representation level, and decision-use logic

| Model component | Prediction target | Representation level | Primary supervision signal | Why the task belongs in the shared framework | Output format | Medicinal chemistry interpretation |
|---|--|--|--|--|--|---|
| Shared molecular graph encoder | Cross-task chemical representation | Atom-level embeddings plus graph-level summary | Gradients from all available CYP inhibition, TDI, and SOM labels | CYP binding, bioactivation, and metabolic transformation often arise from overlapping substructural and physicochemical determinants | Learned atom embeddings, bond-aware messages, and graph-level vector | Provides a common chemical context for interpreting multiple metabolic liabilities from one molecule |
| CYP inhibition head | Reversible inhibition across selected CYP isoforms | Molecule-level, isoform-resolved | Binary inhibition labels, activity thresholds, IC50/Ki categories, or continuous activity values | Isoform-specific inhibition and other CYP liabilities may share determinants such as lipophilicity, heteroatoms, aromatic systems, steric shape, and charge distribution | Multi-label probability vector or activity regression vector | Identifies which enzymes may require confirmatory inhibition assays and DDI-focused follow-up |
| Time-dependent inactivation head | Probability of TDI liability | Molecule-level | Curated TDI labels, inactivation assay outcomes, or kinetic inactivation annotations | TDI may depend on structural features that also influence CYP binding and metabolic activation, making it chemically related but not identical to reversible inhibition | TDI probability, risk category, or assay-prioritization flag | Flags candidates that may need preincubation-based CYP assays or mechanistic inactivation evaluation |
| Soft-spot localization head | Atom-level site-of-metabolism likelihood | Atom or bond level | Experimentally observed SOM labels, metabolite identification data, or atom/bond transformation labels | Metabolic soft spots provide local evidence that can help explain molecule-level clearance, bioactivation, or TDI concerns | Molecular heat map, atom-ranking list, or soft-spot score per atom | Directs analog redesign by identifying labile atoms that may be blocked, replaced, or sterically shielded |
| Sparse-label masking module | Correct handling of incomplete annotations | Dataset and loss-function level | Missingness indicators for each isoform, TDI endpoint, and SOM label | CYP datasets are rarely complete; untested endpoints must not be treated as | Masked losses applied only to observed labels | Reduces false-negative learning and allows heterogeneous datasets to |

| | | | | | | |
|-----------------------------------|--|-----------------------------|--|--|--|---|
| | | | | negative observations | | contribute without artificial label inflation |
| Task-balancing strategy | Stable joint optimization across endpoints | Training-objective level | Weighted inhibition, TDI, and SOM losses | Data-rich tasks can otherwise dominate scarce but clinically important endpoints such as TDI or less common isoforms | Dynamic or pre-specified task weights | Helps preserve performance across all outputs rather than optimizing only the easiest or largest endpoint |
| Interpretability layer | Substructure and atom-level explanation | Molecular and atomic levels | Attribution methods, attention inspection, gradient-based maps, or SHAP-style explanations | A unified model is useful only if predicted liabilities can be connected to chemically meaningful features | Highlighted substructures, atom heat maps, explanatory feature reports | Converts model outputs into redesign hypotheses rather than opaque risk labels |
| Integrated liability panel | Actionable metabolic hazard summary | Compound-profile level | Combined outputs from all prediction heads | Drug discovery decisions require a coordinated profile, not isolated predictions from unrelated tools | CYP inhibition profile, TDI flag, soft-spot map, uncertainty, and follow-up recommendation | Supports compound triage, analog comparison, DDI-risk prioritization, and metabolite-study planning |

Figure 1 presents the proposed multitask deep learning architecture linking molecular graph representation, shared CYP-relevant encoding, task-specific inhibition, inactivation, and soft-spot outputs, and interpretable metabolic liability reporting.

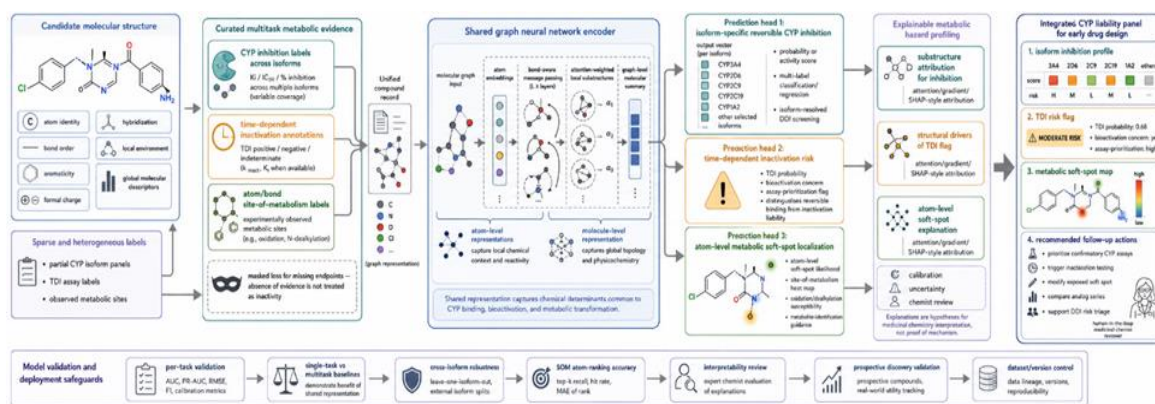


Figure 1. Multitask deep learning architecture for unified CYP inhibition, inactivation, and metabolic soft-spot prediction.

Loss Balancing and Training Strategy

The training objective would combine molecule-level inhibition loss, molecule-level TDI loss, and atom-level SOM loss while masking unavailable labels for each compound. Multitask ADMET systems indicate that task weighting and auxiliary endpoint selection are important when endpoints differ in sparsity, noise, and biological scope [12, 28]. Small-dataset CYP inhibition work also suggests that training should avoid allowing data-rich endpoints to dominate endpoints with limited observations [34]. A practical strategy would use mini-batches containing any available labels and update only the relevant heads, while the shared encoder receives gradients from all observed tasks and learns a representation that can support inhibition, inactivation, and soft-spot prediction together [11, 13].

Handling Sparse and Multi-Isoform Data

Handling Missing Isoform Labels

The inhibition head would output predictions for all selected CYP isoforms, but the loss function would be applied only where experimental labels are available. This masked-loss design is important because CYP inhibition datasets often contain uneven coverage across isoforms, as seen in multitask inhibitor prediction and small-dataset CYP modeling studies [10, 34]. Missing

CYP2C8 or CYP2B6 annotations, for example, should not be interpreted as inactivity simply because a compound was not tested. By separating absence of evidence from negative evidence, the model could learn from sparse multi-isoform matrices without introducing systematic false-negative labels.

Leveraging Cross-Isoform Correlations

The shared encoder would be expected to learn chemical features that are useful across several CYP isoforms while allowing each output head to specialize in isoform-specific preferences. CYP activity prediction platforms and multi-enzyme inhibitor models show that CYP endpoints can be represented as related outputs rather than isolated prediction problems [18, 19]. Cross-isoform learning may help the model distinguish broad hydrophobic CYP3A4 liabilities from more shape- or charge-sensitive patterns associated with other isoforms, although such distinctions should be validated rather than assumed. Multitask substrate prediction further supports the idea that shared representations can capture CYP-family relationships while preserving enzyme-specific interpretability [11].

Table 2 shows the conceptual structure of a shared-encoder, multi-head architecture for predicting CYP isoform activity and the corresponding functional roles of each component across isoform-specific outputs.

Table 2. Multitask learning framework for CYP isoform activity prediction using a shared encoder with isoform-specific output heads

| Component | Role in Model Architecture | Learning Function | Relevance to CYP Prediction |
|-------------------------------------|---|---|---|
| Shared encoder | Learns unified molecular representation from input structures | Extracts general chemical features (e.g., hydrophobicity, sterics, electronics) | Captures cross-isoform patterns shared across CYP family |
| CYP3A4 output head | Isoform-specific prediction layer | Learns CYP3A4-specific binding and metabolism preferences | Focuses on broad hydrophobic and large active-site substrates |
| CYP2D6 output head | Isoform-specific prediction layer | Learns charge-driven and polar interaction patterns | Captures sensitivity to ionizable groups and electrostatics |
| CYP2C9 output head | Isoform-specific prediction layer | Learns shape- and aromaticity-dependent metabolism rules | Reflects substrate selectivity and steric constraints |
| CYP1A2 output head | Isoform-specific prediction layer | Learns planar aromatic preference signals | Emphasizes planar, aromatic ligand recognition |
| Cross-task regularization | Aligns learning across outputs | Encourages shared structure-activity relationships | Improves generalization across CYP isoforms |
| Task-specific specialization | Divergence from shared features | Fine-tunes isoform-specific metabolic rules | Preserves interpretability and biological specificity |

Data Augmentation for SOM Labeling

Because experimentally annotated SOM data are often less abundant than molecule-level activity labels, the SOM branch could benefit from pretraining on auxiliary metabolism-related tasks. The metabolic rainbow framework illustrates how phase I metabolism can be learned as a structured prediction problem across reaction classes [5], while active learning for site-of-metabolism data generation suggests that new labels can be prioritized where model uncertainty is highest [35]. Reactivity descriptors, bond environments, and CYP product-prediction signals could provide surrogate objectives before fine-tuning on experimentally observed soft spots [9, 25]. Such augmentation should be treated as representation learning rather than a substitute for direct metabolite evidence.

Model Interpretability and Metabolic Profiling

Explaining Inhibition and TDI Predictions

For graph-level CYP inhibition and TDI outputs, explainability methods could identify substructures that drive predicted liability and help medicinal chemists judge whether the prediction is chemically plausible. Explainable multimodal CYP inhibitor modeling shows how molecular features can be connected to CYP450 inhibition predictions [31], and substructure-based deep learning for CYP inhibition supports the value of linking predictions to recognizable chemical patterns [21]. SHAP-style, attention-based, or gradient-based attributions could highlight motifs associated with reversible binding or inactivation risk, such as electrophilic precursors or oxidizable heteroaromatic systems. These explanations should be interpreted as hypotheses for follow-up chemistry rather than proof of a specific bioactivation mechanism.

Atom-Level Explanation for Soft-Spot Predictions

For soft-spot prediction, interpretability should operate directly at the atom level so that the output can be visualized as a molecular heat map. FAME 3 and graph neural network SOM models demonstrate that atom-level metabolism prediction can guide attention to likely sites of biotransformation [4, 8]. Explainable graph neural network evaluation is relevant because high-quality visual explanations should correspond to chemically meaningful atoms rather than arbitrary graph artifacts [32]. A combined model could therefore connect a molecule-level TDI flag with the atom-level site most likely to initiate metabolic activation, creating a more useful metabolic profile than either output alone.

*Integration Into Drug Discovery And Ddi Risk Assessment**Early-Stage Metabolic Hazard Screening*

In early discovery, the model could be applied to virtual libraries to identify compounds with predicted CYP inhibition, potential TDI liability, and metabolically labile atoms before synthesis. ADMETlab 2.0 and HelixADMET illustrate how integrated computational ADMET platforms can support early triage across pharmacokinetic endpoints [28, 29]. A unified CYP-specific model would extend this idea by returning a focused metabolic hazard panel rather than separate outputs from unrelated tools. Medicinal chemists could then consider structural modifications that reduce predicted inhibition or block an exposed soft spot while preserving desired activity.

Supporting DDI Risk Assessment in Development

During lead optimization and early development, isoform-specific inhibition outputs could be used to prioritize confirmatory in-vitro CYP inhibition assays. Models for CYP inhibition, CYP3A4 TDI, and reversible-versus-time-dependent inhibition suggest that computational screening can help organize follow-up testing when many candidates compete for experimental resources [2, 6, 7]. A TDI flag could trigger dedicated inactivation assays, while atom-level soft-spot predictions could guide metabolite identification experiments. The model should therefore be positioned as a decision-support system for DDI risk assessment rather than a replacement for experimental evaluation.

Table 3 shows how isoform-specific inhibition predictions, TDI flags, and atom-level soft-spot identification can be translated into a structured experimental prioritization strategy during lead optimization and early development.

Table 3. Model-derived CYP inhibition outputs and their role in prioritizing experimental follow-up during lead optimization

| Model output | Interpretation | Suggested experimental follow-up | Role in decision-making |
|---|---|---|--|
| Isoform-specific inhibition score (e.g., CYP3A4, CYP2D6, CYP2C9) | Predicted likelihood of inhibition for each CYP isoform | Confirmatory in-vitro CYP inhibition assays per isoform | Prioritizes which CYP isoforms require immediate experimental validation |
| CYP3A4 time-dependent inhibition (TDI) flag | Indicates potential mechanism-based enzyme inactivation | Dedicated TDI inactivation kinetic assays (e.g., pre-incubation studies) | Triggers specialized assays for mechanism-based inhibition risk |
| Reversible inhibition probability | Likelihood of competitive or non-covalent inhibition | Standard reversible inhibition IC50/Ki determination assays | Helps classify inhibition type for DDI risk assessment |
| Atom-level metabolic soft-spot prediction | Identified molecular regions prone to CYP-mediated metabolism | Metabolite identification studies (LC-MS/MS) and structural modification design | Guides structural optimization to reduce metabolic liabilities |
| Integrated DDI risk score | Combined prediction across isoforms and inhibition mechanisms | Tiered in-vitro assay strategy (screen → confirm → mechanistic studies) | Supports prioritization of compounds in multi-candidate selection |

*Evaluation Strategy**Per-Task Predictive Performance*

Evaluation should compare each task against appropriate single-task and multitask baselines without relying on one endpoint to represent overall success. CYP inhibition could be assessed separately by isoform, consistent with multi-isoform inhibitor prediction studies [1, 10], while TDI evaluation should reflect the specific challenge of distinguishing reversible inhibition from time-dependent inactivation [3, 7]. SOM evaluation should assess whether predicted atom rankings align with experimentally observed sites of metabolism, following atom-level metabolism modeling traditions [4, 26]. Metrics such as classification discrimination, calibration, and atom-ranking behavior could be reported in a future validation study, but this conceptual article does not claim numerical outcomes.

Multitask Benefit and Transfer Learning

The multitask benefit should be evaluated by comparing jointly trained models against models trained independently for inhibition, TDI, and SOM. Multi-task graph learning under adaptive auxiliary task selection provides a useful precedent for testing whether auxiliary endpoints help or harm a target task [12]. Transfer learning could also be examined by pretraining on broader ADMET or metabolism datasets and then fine-tuning on CYP-specific endpoints, following the general logic of self-supervised knowledge transfer and derivative-aware ADMET modeling [13, 29]. The key question is whether shared representations improve consistency and robustness, especially for endpoints with limited labels.

Interpretability Validation

Interpretability validation should test whether model-highlighted atoms and substructures correspond to chemically credible CYP liabilities. Molecular coloring with explainable artificial intelligence provides a precedent for using visual attributions in preclinical relevance assessment [17], and quantitative explainability studies emphasize that explanations themselves require evaluation rather than automatic acceptance [32]. Medicinal chemists could review whether highlighted structural alerts align

with known inhibition or inactivation hypotheses, while metabolism specialists could assess whether SOM heat maps match plausible oxidation or dealkylation chemistry. Such validation would be qualitative and mechanistic, complementing but not replacing predictive evaluation.

Table 4 provides an evaluation framework for determining whether the unified model improves endpoint-specific prediction, cross-task consistency, interpretability, and prospective decision support compared with isolated CYP liability models.

Table 4. Evaluation and interpretation framework for a unified CYP inhibition, TDI, and soft-spot model

| Evaluation domain | Core question | CYP inhibition assessment | TDI assessment | Soft-spot assessment | Multitask-specific test | Interpretation standard | Deployment implication |
|---------------------------------|---|---|--|--|--|--|--|
| Per-task discrimination | Can each endpoint be predicted accurately on its own terms? | Evaluate isoform-specific classification or regression performance separately for each CYP enzyme | Evaluate ability to distinguish TDI-positive from TDI-negative compounds, especially among reversible inhibitors | Evaluate whether known metabolic atoms or bonds receive high predicted ranks | Compare each head against matched single-task baselines | High aggregate performance should not hide weak performance for rare isoforms or sparse endpoints | Determines whether the model is reliable enough for endpoint-specific screening decisions |
| Calibration and uncertainty | Are predicted probabilities meaningful for decision triage? | Assess whether predicted inhibition probabilities correspond to observed inhibition frequency | Assess whether TDI risk categories align with observed inactivation outcomes | Assess confidence in atom-level soft-spot rankings, especially when multiple plausible sites exist | Test whether joint training improves or worsens calibration relative to isolated models | Uncertain predictions should be visibly separated from low-risk predictions | Supports rational prioritization of confirmatory assays rather than overconfident automation |
| Cross-task consistency | Do related outputs form a chemically coherent metabolic profile? | Examine whether strong inhibition predictions align with plausible CYP-recognition features | Examine whether TDI flags are connected to metabolic activation or reactive structural hypotheses | Examine whether soft spots occur near chemically plausible sites for transformation | Test whether shared representations reduce contradictory predictions across heads | A molecule flagged for TDI should ideally have interpretable structural or soft-spot evidence | Helps medicinal chemists understand whether the combined profile is plausible or internally inconsistent |
| Sparse-label robustness | Does the model handle incomplete CYP, TDI, and SOM annotation without bias? | Test performance under uneven isoform label coverage | Test whether scarce TDI labels are overwhelmed by larger inhibition datasets | Test robustness when SOM labels are partial or limited to observed metabolites | Compare masked-loss training against naive missing-as-negative training | Missing experimental data should not be interpreted as true inactivity or absence of metabolism | Protects against systematic false reassurance in under-tested chemical regions |
| Multitask benefit | Does joint learning improve over separate models? | Compare inhibition performance under single-task and multitask training | Test whether inhibition and SOM signals improve TDI prediction | Test whether molecule-level CYP information improves atom-level localization | Use ablation models: inhibition-only, TDI-only, SOM-only, pairwise multitask, and full multitask | Multitask learning should be retained only when it improves accuracy, calibration, or interpretability | Justifies the added complexity of a unified architecture |
| Atom-level explanation validity | Are highlighted soft spots chemically meaningful? | Link inhibition attributions to recognizable CYP-binding substructures | Link TDI attribution to plausible bioactivation motifs, without claiming proof | Compare highlighted atoms with experimentally observed metabolic sites | Test whether shared encoder attributions remain stable across related tasks | Explanations should be reviewed by metabolism and medicinal chemistry experts | Enables redesign suggestions such as blocking, replacing, or shielding labile atoms |

| | | | | | | | |
|---------------------------------------|--|---|--|--|---|---|---|
| Prospective utility | Does the model improve real discovery decisions? | Track whether predicted CYP inhibition flags anticipate confirmatory assay outcomes | Track whether TDI flags help prioritize inactivation testing | Track whether predicted soft spots guide successful analog modification or metabolite identification | Compare prospective triage using the unified panel versus independent tools | The model should assist decisions, not replace experimental metabolism studies | Establishes whether the system has practical value in lead optimization and DDI-risk planning |
| Governance and reproducibility | Can the model be audited and updated safely? | Maintain isoform-specific dataset provenance and assay definitions | Document TDI assay conditions, thresholds, and label harmonization rules | Record SOM annotation sources and metabolite-evidence quality | Version datasets, model weights, task weights, and evaluation splits | Transparent reporting is required because CYP endpoints are assay-sensitive and heterogeneous | Supports reproducible benchmarking, regulatory-facing documentation, and responsible deployment |

Limitations

Data Scarcity for Certain Isoforms and TDI

A major limitation is that some CYP isoforms and TDI endpoints may have sparse, heterogeneous, or assay-dependent data. Small-dataset CYP inhibition work shows that limited endpoint coverage can restrict the reliability of deep models for less frequently studied isoforms [34]. TDI modeling is further complicated by experimental variability and dependence on assay conditions, which can make labels harder to harmonize across sources [3]. Pretraining on broader ADMET or metabolism tasks may help, but it cannot fully remove the need for high-quality endpoint-specific data.

The Challenge of Bioactivation and Reactive Metabolites

The proposed model would predict inhibition, TDI risk, and metabolic soft spots, but it would not directly establish the toxicity of downstream metabolites. CYP product prediction tools can suggest likely metabolic products [25], and comprehensive graph learning frameworks for drug metabolism indicate how reaction-aware prediction could be integrated into broader systems [16]. However, reactive metabolite toxicity depends on additional factors such as covalent binding, detoxification capacity, exposure, and biological target susceptibility. A future extension would need to connect soft-spot and product prediction with mechanistic models of bioactivation and cellular consequence.

Conclusion

A multitask deep learning model for CYP inhibition, inactivation, and metabolic soft-spot prediction could provide a unified computational view of metabolic liability. By processing a molecular graph through a shared encoder and task-specific heads, the model could jointly estimate isoform-specific inhibition, TDI risk, and atom-level metabolic susceptibility. This design would reflect the chemical reality that CYP binding, metabolic activation, and soft-spot formation are related rather than fully independent phenomena.

The main strength of the proposed framework is its use of a shared molecular representation to improve data efficiency across related endpoints. A single model could reduce duplicated modeling effort and return a more coherent metabolic profile for each candidate molecule. Atom-level soft-spot visualization would also make the system more actionable for medicinal chemists than a model that provides only molecule-level risk labels.

Important challenges remain for practical deployment. Sparse data for certain isoforms, inconsistent TDI annotations, and incomplete SOM labels could limit generalizability unless datasets are carefully curated and prospectively validated. The complexity of reactive metabolite toxicity also means that soft-spot prediction should be interpreted as a guide to metabolic transformation rather than as a complete safety assessment.

Future work should emphasize open-source models, transparent benchmarks, and collaborative datasets that integrate CYP inhibition, inactivation, and metabolism annotations. Prospective validation would be essential to determine whether multitask learning improves decision-making in real discovery projects. With careful evaluation and interpretable outputs, unified metabolic hazard prediction could become a valuable part of early drug design workflows.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Li X, Xu Y, Lai L, Pei J. Prediction of human cytochrome P450 inhibition using a multitask deep autoencoder neural network. *Mol Pharm.* 2018;15(10):4336-45.
2. Xu M, Lu Z, Wu Z, Gui M, Liu G, Tang Y, et al. Development of in silico models for predicting potential time-dependent inhibitors of cytochrome P450 3A4. *Mol Pharm.* 2022;20(1):194-205.
3. Fluetsch A, Trunzer M, Gerebtzoff G, Rodriguez-Perez R. Deep learning models compared to experimental variability for the prediction of CYP3A4 time-dependent inhibition. *Chem Res Toxicol.* 2024;37(4):549-60.
4. Šicho M, Stork C, Mazzolari A, de Bruyn Kops C, Pedretti A, Vistoli G, et al. 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(8):3400-12.
5. Dang NL, Matlock MK, Hughes TB, Swamidass SJ. The metabolic rainbow: deep learning phase I metabolism in five colors. *J Chem Inf Model.* 2020;60(3):1146-64.
6. Plonka W, Stork C, Šicho M, Kirchmair J. CYPlebrity: machine learning models for the prediction of inhibitors of cytochrome P450 enzymes. *Bioorg Med Chem.* 2021;46:116388.
7. Faramarzi S, Bassan A, Cross KP, Yang X, Myatt GJ, Volpe DA, et al. Novel (Q)SAR models for prediction of reversible and time-dependent inhibition of cytochrome P450 enzymes. *Front Pharmacol.* 2025;15:1451164.
8. Porokhin V, Liu LP, Hassoun S. Using graph neural networks for site-of-metabolism prediction and its applications to ranking promiscuous enzymatic products. *Bioinformatics.* 2023;39(3):btad089.
9. He S, Li M, Ye X, Wang H, Yu W, He W, et al. Site of metabolism prediction for oxidation reactions mediated by oxidoreductases based on chemical bond. *Bioinformatics.* 2017;33(3):363-72.
10. Nguyen-Vo TH, Trinh QH, Nguyen L, Nguyen-Hoang PU, Nguyen TN, Nguyen DT, et al. iCYP-MFE: identifying human cytochrome P450 inhibitors using multitask learning and molecular fingerprint-embedded encoding. *J Chem Inf Model.* 2021;62(21):5059-68.
11. Fang J, Tang Y, Gong C, Huang Z, Feng Y, Liu G, et al. Prediction of cytochrome P450 substrates using the explainable multitask deep learning models. *Chem Res Toxicol.* 2024;37(9):1535-48.
12. Du BX, Xu Y, Yiu SM, Yu H, Shi JY. ADMET property prediction via multi-task graph learning under adaptive auxiliary task selection. *iScience.* 2023;26(11):108242.
13. Fralish Z, Chen A, Skaluba P, Reker D. DeepDelta: predicting ADMET improvements of molecular derivatives with deep learning. *J Cheminform.* 2023;15(1):101.
14. Qiu M, Liang X, Deng S, Li Y, Ke Y, Wang P, et al. A unified GCNN model for predicting CYP450 inhibitors by using graph convolutional neural networks with attention mechanism. *Comput Biol Med.* 2022;150:106177.
15. Xiao Z, Hirao H. Deep learning models for predicting human cytochrome P450 inhibition and induction. *J Chem Inf Model.* 2025;65(19):9947-61.
16. Zhou Y, Jiang D, Wei X, Yi J, Wang Y, Deng Y, et al. DeepMetab: a comprehensive and mechanistically informed graph learning framework for end-to-end drug metabolism prediction. *Chem Sci.* 2025;16(40):18884-902.
17. Jiménez-Luna J, Skalic M, Weskamp N, Schneider G. Coloring molecules with explainable artificial intelligence for preclinical relevance assessment. *J Chem Inf Model.* 2021;61(3):1083-94.
18. Banerjee P, Dunkel M, Kemmler E, Preissner R. SuperCYPsPred—a web server for the prediction of cytochrome activity. *Nucleic Acids Res.* 2020;48(W1):W580-5.
19. Ai D, Cai H, Wei J, Zhao D, Chen Y, Wang L. DEEPCYPs: a deep learning platform for enhanced cytochrome P450 activity prediction. *Front Pharmacol.* 2023;14:1099093.
20. Finkelmann AR, Goldmann D, Schneider G, Göller AH. MetScore: site of metabolism prediction beyond cytochrome P450 enzymes. *ChemMedChem.* 2018;13(21):2281-9.
21. Chen Z, Zhang L, Zhang P, Guo H, Zhang R, Li L, et al. Prediction of cytochrome P450 inhibition using a deep learning approach and substructure pattern recognition. *J Chem Inf Model.* 2023;64(7):2528-38.
22. Zahid H, Tayara H, Chong KT. Harnessing machine learning to predict cytochrome P450 inhibition through molecular properties. *Arch Toxicol.* 2024;98(8):2647-58.
23. Gong C, Feng Y, Zhu J, Liu G, Tang Y, Li W. Evaluation of machine learning models for cytochrome P450 3A4, 2D6, and 2C9 inhibition. *J Appl Toxicol.* 2024;44(7):1050-66.
24. Tian S, Djoumbou-Feunang Y, Greiner R, Wishart DS. CypReact: a software tool for in silico reactant prediction for human cytochrome P450 enzymes. *J Chem Inf Model.* 2018;58(6):1282-91.
25. Tian S, Cao X, Greiner R, Li C, Guo A, Wishart DS. CyProduct: a software tool for accurately predicting the byproducts of human cytochrome P450 metabolism. *J Chem Inf Model.* 2021;61(6):3128-40.
26. Yang H, Liu J, Chen K, Cong S, Cai S, Li Y, et al. D-CyPre: a machine learning-based tool for accurate prediction of human CYP450 enzyme metabolic sites. *PeerJ Comput Sci.* 2024;10:e2040.
27. Jacob RA, Gaskin L, Seidel T, Chen Y, Mazzolari A, Kirchmair J. FAME3R: an efficient, practical and reliable open-source tool for predicting phase 1 and phase 2 sites of metabolism. *J Cheminform.* 2026;18:14.

28. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 2021;49(W1):W5-14.
29. Zhang S, Yan Z, Huang Y, Liu L, He D, Wang W, et al. HelixADMET: a robust and endpoint extensible ADMET system incorporating self-supervised knowledge transfer. *Bioinformatics.* 2022;38(13):3444-53.
30. Aksamit N, Tchagang A, Li Y, Ombuki-Berman B. Hybrid fragment-SMILES tokenization for ADMET prediction in drug discovery. *BMC Bioinformatics.* 2024;25(1):255.
31. Atwereboannah AA, Wu WP, Al-Antari MA, Yussif SB, Ejiyi CJ, Gu YH, et al. MEN: leveraging explainable multimodal encoding network for precision prediction of CYP450 inhibitors. *Sci Rep.* 2025;15(1):21820.
32. Rao J, Zheng S, Lu Y, Yang Y. Quantitative evaluation of explainable graph neural networks for molecular property prediction. *Patterns.* 2022;3(12):100628.
33. Coley CW, Barzilay R, Green WH, Jaakkola TS, Jensen KF. Convolutional embedding of attributed molecular graphs for physical property prediction. *J Chem Inf Model.* 2017;57(8):1757-72.
34. Permadi EE, Watanabe R, Mizuguchi K. Improving the accuracy of prediction models for small datasets of cytochrome P450 inhibition with deep learning. *J Cheminform.* 2025;17(1):66.
35. Chen Y, Seidel T, Jacob RA, Hirte S, Mazzolari A, Pedretti A, et al. Active learning approach for guiding site-of-metabolism measurement and annotation. *J Chem Inf Model.* 2024;64(2):348-58.