World Journal of Pharmaceutical Sciences
ISSN (Print): 2321-3310; ISSN (Online): 2321-3086
Available online at: https://wjpsonline.com/
Review Article

COMPUTATIONAL APPROACHES TO PREDICTING DRUG METABOLISM

Kavya Banala[1]*, Sridhar Siddiraju[2], Muvvala Sudhakar[2]

[1]Pharm.D, Malla Reddy College of Pharmacy (MRCP), Maisammaguda, Telangana, India – 500100
[2]Department of Pharmaceutical Chemistry, Malla Reddy College of Pharmacy (MRCP), Maisammaguda, Telangana, India – 500100

Received: 10-03-2026 / Revised Accepted: 13-03-2026 / Published: 15-04-2026

Abstract
One of the most important factors to be examined and improved in drug discovery initiatives is metabolism. It is crucial to identify metabolically unstable substances or possible inhibitors of crucial enzymes as soon as feasible. Strong computational techniques are required because there are more compounds being synthesized than can be studied experimentally. Goal: We provide a summary of the most recent in-silico techniques for forecasting experimental metabolic endpoints, emphasizing their suitability for the pharmaceutical sector. The metabolic fate and interaction with metabolizing enzymes are shown both macroscopically and microscopically. Methods: Approaches based on ligands, proteins, and rules are discussed. Conclusion: Despite significant advancements, the computations' outcomes still require close examination. It is necessary to consider both the models' domain of applicability and methodological constraints.

KEYWORDS: ADME, QSAR, drug metabolism, in-silico prediction, protein–ligand interaction, site of metabolism (SOM), cytochrome P450, metabolic stability, pharmacokinetics, quantum mechanics, descriptor-based modeling, metabolic pathways, enzyme-substrate interaction, predictive modeling, computational pharmacology.

INTRODUCTION

Drug discovery and development have changed significantly in recent years. Previously, it was frequently thought to be adequate to create a small molecule with high affinity to the primary target (and minimal affinity to targets that are not disease-related) (1). Poor physicochemical characteristics or pharmacokinetic behaviour were only addressed later in the development phase. This subset selection requires careful thought, and in-silico techniques are frequently used to guide a prioritization (2,3). Similarly, trustworthy computational techniques can impact the creation of novel compounds by forecasting ADME characteristics, either to prevent possible ADME-related problems or to identify which possible issues require experimental investigation (4,5). Among the different ADME endpoints, metabolism is arguably the most difficult to forecast since it is a complicated biological process involving several enzymatic systems and mechanisms that frequently compete (6,7). The overall metabolic process is influenced by the substrate's reactivity as well as the ligand-enzyme interaction; genetic and phenotypic variations make matters more complicated (8,9). The prediction of a compound's overall metabolic stability, the strength of its binding to metabolic enzymes, and the computation of microscopic reaction barriers are just a few of the many aspects of in-silico approaches to metabolic events (10,11). These potential endpoints are emphasized in Section 3 of this review, which is surrounded by other sections that give background information on human metabolism and specifics of the methodology.

All too frequently, theoretical and experimental approaches are still seen as distinct fields. The most crucial thing to understand is that there is a great deal of room for collaboration between the two fields, which will enable the analysis and forecasting of drug metabolism to advance significantly (3,6). Successful drug discovery and development depend on having a thorough understanding of metabolic processes at the molecular level (7,8). Understanding a molecule's metabolic characteristics can help to maximize the stability and consequently the drug's risk-benefit ratio and in vivo half-life (9,11). There are numerous experimental techniques available to examine drugs' metabolic fate in unprecedented detail (8,10). Nevertheless, the development of computational tools for drug metabolism prediction has been significantly influenced by the continued demands placed on these experimental approaches in terms of scientific equipment, human expertise, cost, and time (4,13,16). The

Address for Correspondence: Kavya Banala, Pharm.D, Malla Reddy College of Pharmacy (MRCP), Maisammaguda, Telangana, India – 500100, Email: kavyabanala05@gmail.com.

How to Cite this Article: Kavya Banala, COMPUTATIONAL APPROACHES TO PREDICTING DRUG METABOLISM, World J Pharm Sci 2026; 14(01): 105-111; https://doi.org/10.54037/WJPS.2022.100905

metabolic fate of virtual compounds can be predicted using in silico techniques, and the most promising approach to maximize the metabolic stability of a priori project compounds (12,15,18).

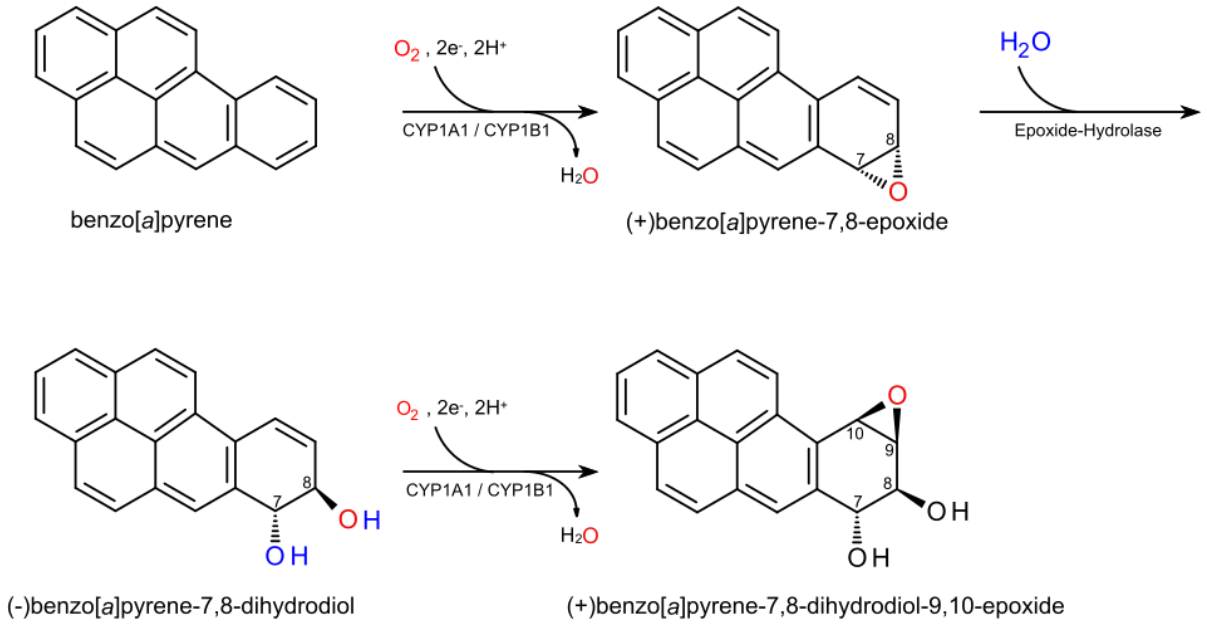

Fig. 1: Metabolic activation pathway of benzo[a]pyrene, showing Phase I transformation to a carcinogenic epoxide metabolite via cytochrome P450 enzymes. This illustrates key aspects of xenobiotic metabolism, including enzymatic transformation and metabolite formation.

(Attribution:

Image: Benzo[a]pyrene metabolism by Cancer Research UK, via Wikimedia Commons. Licensed under CC BY-SA 4.0.)

Drugs taken orally undergo metabolism by the gut microbiota, hepatocytes, enterocytes, and other body cells (15). Drug metabolism mediated by the gut microbiota and the host is impacted by a number of factors, including diet, disease, host genetics, and the use of both antibiotics and non-antibiotics (11,16). Apart from the enzymes that humans use to metabolize drugs, certain drugs are also metabolized by microbial species and enzymes (12,14). These can change the production of active and non-active metabolites and further regulate drug responses like pharmacokinetics and/or pharmacodynamics (4,5). Both untargeted metabolomics for metabolites derived from the host and microbiome and targeted metabolomics for drugs and their metabolites can aid in a better understanding of the mechanisms underlying host and microbiome-related metabolism (17,18).

METHODS OF COMPUTATION FOR EXAMINING DRUG METABOLISM:

For the prediction of drug metabolism, a vast range of computational techniques and integrated approaches have been developed and are available as both free and paid software as well as web services (13,15). They might be categorized as either comprehensive ("global") or specific ("local") (14,16). While global models can theoretically be applied to a wide range of biological systems, including the majority of small organic compounds and all metabolic enzymes and bio-transformations, specific models are only applicable to specific biomolecules (primarily metabolic enzymes) and/or metabolic reactions (12,17). Many software packages for metabolism aim to combine different tools and techniques for the greatest number of drug metabolism targets rather than just one enzyme (15,18).

It is also possible to include additional functional proteins, like transporters (like P-glycoprotein), that might work in concert with metabolizing enzymes (11). Determining the basic structural, functional, and mechanistic characteristics of biomolecules linked to drug metabolism is one of the most popular uses of computational techniques, as it allows the identification of small organic molecules' metabolically labile positions and metabolite prediction (3,10,13). A solid foundation for forecasting the metabolites' reactivity, toxicity, bioactivity, and other pharmacokinetic and pharmacodynamic characteristics has been established once chemical structures have been determined (15,16,19).

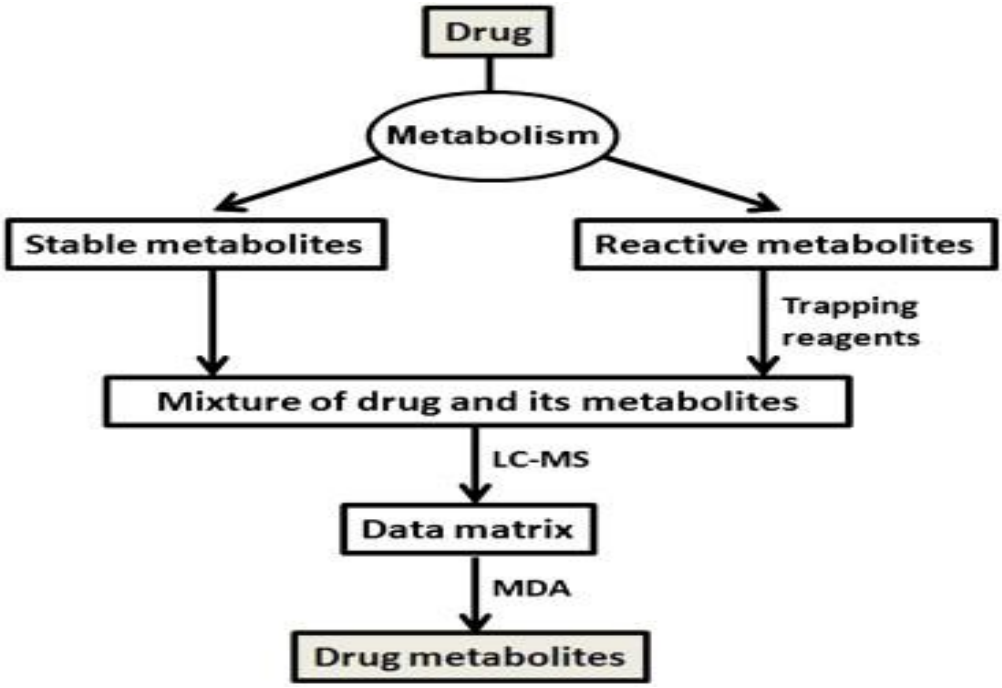

Figure 2: Drug Metabolism (Flow Chart)

ANALYTICAL TECHNIQUES AND BIOLOGICAL SYSTEMS FOR EXAMINING DRUG METABOLISM:

The most significant forms of CYPs, which are crucial for drug metabolism, are primarily expressed in the liver, though they are also expressed at lower levels in a wide range of other organs (7,8). Thus, liver or in vitro liver-derived systems are frequently the most practical, fascinating, and significant experimental model systems for studies of metabolism, such as when examining the first-pass metabolism of an oral medication (9,10). Since humans are typically the target species, it is crucial and beneficial to use in vitro systems with human-derived material early in drug discovery due to the significant metabolic variability among different species (11,13). There are numerous in vitro systems and novel in vivo techniques that can be applied specifically based on the problem or issue to be resolved (14,15). Recombinant enzymes, which are expressed in conjunction with coenzymes to achieve optimal catalytic activity, are the most basic systems. For example, they can be utilized as a single CYP isoform system for drug interaction research or to determine which isoforms are involved in a compound's metabolism (6,12). Centrifugation can be used to separate various enzyme families from native material, such as liver. While membrane-bound enzymes like CYPs and UDP-glucuronosyltransferases (UGTs) are enriched in the pellet and, following resuspension, the material known as "microsomes" (membrane vesicles of the endoplasmic reticulum) is obtained, soluble cytosolic enzymes, such as sulfotransferases (SULTs), remain in the supernatant (10,14).

APPLICATIONS:

1)METABOLIC ALTERATIONS:

Due to the complexity of underlying processes, metabolic stability predictions are typically based on ligand/descriptor-based classification models, distinguishing between stable and unstable compounds (1,16). One of the earliest models using Partial Least Squares Regression (PLS) was developed for calcitriol analogues, achieving accurate predictions for 17 out of 20 compounds (1).

Pfizer conducted machine learning research using proprietary data on ~15,000 compounds, with up to 80% accuracy in predicting stability on test sets. Prediction confidence was emphasized, showing a strong correlation between calculated probability and prediction accuracy (15). Another study compared various techniques, with Support Vector Machines (SVM) and Random Forest performing best (13,15).

B) IDENTIFICATION OF BIOTRANSFORMATIONS AND METABOLITES: Understanding drug metabolism requires identifying involved enzymes and reaction pathways (3,10). Tools like Bio-print, which include biological fingerprints from marketed drugs and ADME test data, use QSAR models and structural similarity to predict enzyme interactions (1,13). Similarly, the Meta-Drug database stores xenobiotic reactions and enzyme inhibition kinetics. Using this, kernel-partial least squares models predicted metabolic outcomes like N-dealkylation and aromatic hydroxylation with 61–79% success (14,15).

Despite promising results, larger datasets are needed to capture rare biotransformations. Erhardt reviewed key challenges, including lack of 3D structural data and incomplete drug-drug interaction (DDI) profiles (17,18). Expert systems like METEOR, MetaDrug, and Meta-Bol Expert—based on commercial databases such as Accelrys Metabolism and MDL—rank likely sites of metabolism (SOMs) using transformation rules (6,18). AstraZeneca applied molecular fingerprinting to predict SOMs, achieving 87% accuracy in blind testing (12).

Organon's SyGMa (Systematic Generation of potential Metabolites) leverages MDL data and expert rules to cover ~70% of known human metabolic reactions. Rule weighting based on reaction frequency improved experimental metabolite enrichment, supporting its use in in vitro detection workflows (6).

C)LIGAND REGIOSELECTIVITES, RATES, AND REACTIVITIES IN METABOLIC REACTIONS:

A metabolic enzyme's overall turnover rate depends on two key factors: the electronic reactivity at the site of metabolism (SOM) and the likelihood that the SOM is correctly oriented within the enzyme's active site (7,9). Reactivity estimation methods often address these aspects to varying degrees. Regioselectivity can be assessed by comparing metabolic rates within a compound.

Meta-Site software predicts SOMs by analyzing the complementarity of GRID molecular interaction fields between the ligand and the CYP active site, particularly the heme iron (15). The best-fitting orientation suggests the SOM based on proximity to the heme. Additionally, reactivity corrections using quantum chemical calculations for drug-like fragments further refine predictions (10,14). Based on CYP2D6 modeling, considering the top three predicted sites yielded ~80% accuracy in SOM identification (6).

Interestingly, studies showed that Meta-Site's performance in predicting the site of metabolism (SOM) for CYP3A4 was relatively unaffected by variations in protein structure, although predictions improved with crystal structures (80%) compared to homology models (50%) (10,15). In contrast, CYP2D6 predictions remained consistent regardless of structure (6).

The CypScore approach utilized ~850 compounds and 2,400 CYP-mediated transformations, using quantum mechanics-based descriptors in MLR models for various reaction types, including hydroxylation and N-/O-dealkylation. It accurately predicted >75% of SOMs in public datasets and up to 90% using Bayer Schering's internal data (6,14). Blocking predicted SOMs led to improved metabolic stability.

Warfarin's orientation in the CYP2C9 active site was examined via molecular dynamics, aligning with experimental regioselectivity (8,20). Ligand- and protein-based SOM prediction methods were compared using ROCS for molecular alignment. ROCS achieved 82% accuracy versus 71% for docking methods when matched with the reference compound flurbiprofen (13,18). A dataset of 30 compounds validated the method, identifying 28 as non-CYP2C9 substrates.

Quantum mechanical calculations have been used to estimate reaction rates, particularly for CYP-mediated hydroxylation, with the hydrogen abstraction step representing the highest energy barrier in the rebound mechanism (7,9).

The stability of radicals formed after hydrogen abstraction can help estimate intrinsic group reactivity, often linked to Fukui function maxima with moderate success in predicting metabolism sites (9,17). Advanced methods directly estimate hydrogen abstraction barriers using simplified cytochrome models or full iron-porphyrin systems. Recent developments like QM/MM methods consider the protein environment, though they remain computationally expensive (7,10).

These studies highlight the importance of both substrate accessibility and electronic factors, especially the reactive Fe=O group in the enzyme's active site (15,20). For CYP3A4, where steric constraints are lower, QM-based predictions are more reliable, but accuracy drops with other isoforms (14,18). Combined QM calculations (e.g., energy of abstraction and surface exposure) can identify metabolic "hot spots," though predictions succeeded in only ~50% of CYP3A4 cases (6,10).

A low-cost trend vector model was developed for CYP3A4 regioselectivity, which performed comparably to Meta-Site (15). For CYP2D6 and CYP3A4, a QSMR model using 83 N-dealkylation reactions and neural networks identified steric hindrance and electronic descriptors as most relevant (13,19). Additional models using random forest and substructure fragments achieved good accuracy, correctly predicting the SOM in 67–84% of test compounds (14,20).

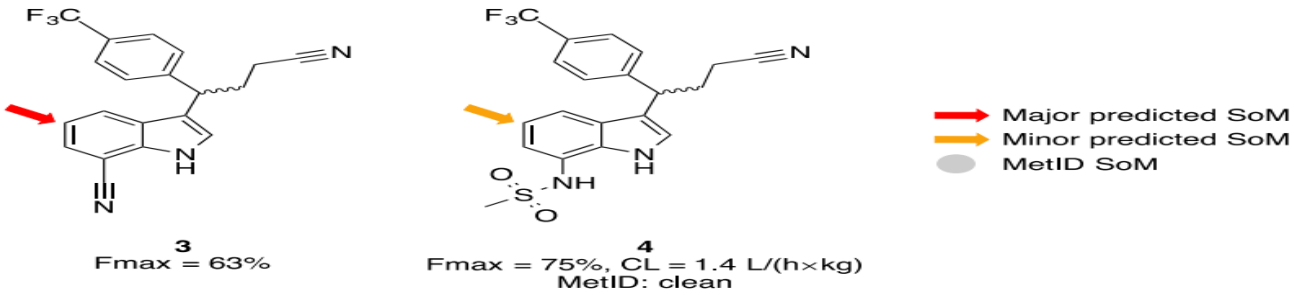

Fig. 3 Illustration of predicted major (red arrow) and minor (orange arrow) sites of metabolism (SoM) for two drug candidates. Compound 3 shows a major predicted SoM with Fmax = 63%, while compound 4 demonstrates improved metabolic stability (Fmax = 75%) and reduced clearance (CL = 1.4 L/h×kg), with a clean MetID profile.

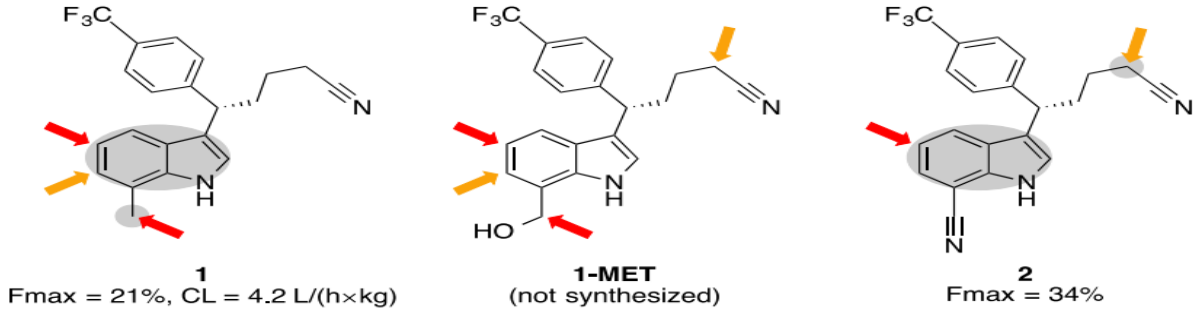

Fig. 4 Prediction and Identification of Major and Minor Sites of Metabolism in Drug Candidates

2) BINDING TO SPECIFI C METABOLIC ENZYMES:

Most in-silico studies on enzyme–drug interactions focus on the cytochrome P450 (CYP) superfamily. Computational methods have been applied to model compounds as substrates or inhibitors of specific CYP isoforms (3,6). Early studies often used small, congeneric compound series to develop correlations with parameters like lipophilicity or through 3D-QSAR models (1,13).

Multivariate data analysis techniques—such as principal component analysis (PCA) and partial least squares (PLS)—have been employed using 2D and 3D molecular descriptors to build multiparameter linear models. These approaches have been extended to predict inhibition across multiple CYP isoforms simultaneously (12,14,15).

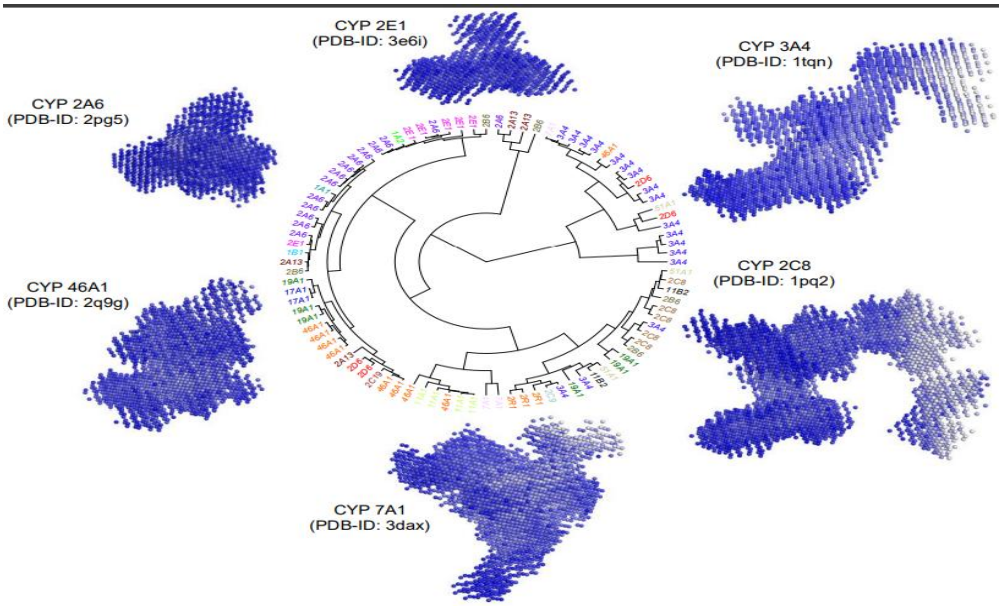

Fig. 5: Structural and Phylogenetic Comparison of Human Cytochrome P450 Isoforms Involved in Drug Metabolism

Cytochrome P450 (CYP) binding pockets vary in shape and buriedness depending on the isoform, as visualized using the PocketPicker tool (6). Central clustering diagrams group enzymes based on pocket geometry, with substrate-free pockets shown for major isoforms. While CYP2 and CYP3 primarily metabolize xenobiotics, CYP11 and CYP46 play key physiological roles (11,15).

Pharmacophore models—some incorporating structural protein details—have been developed, but their predictive power often varies across compound classes due to diverse binding modes. Although limited in generalizability, such models are useful within specific chemical classes and may provide early warnings when used together (12,14).

Modern machine learning techniques, including supervised and unsupervised approaches using a variety of molecular descriptors, have significantly enhanced CYP interaction prediction. Classification models have reported accuracies above 70%, and binary models distinguishing strong/weak inhibitors show even better performance (16,17). Regression models on external test sets yield R^2 values of 0.55–0.68 (17,18).

The ability of predictive models to rank compounds within a series is essential for pharmaceutical applications. Such models are used to eliminate potent CYP inhibitors or retain benign compounds. When misclassifications across classes are minimal, regression or multiclass models that include medium-affinity inhibitors can enhance screening efficiency (16,17).

To optimize model performance, feature selection is often performed prior to training, but care must be taken to prevent overfitting (19). An essential part of QSAR modeling is defining the applicability domain—ensuring the compound being predicted lies within the chemical space of the training set.

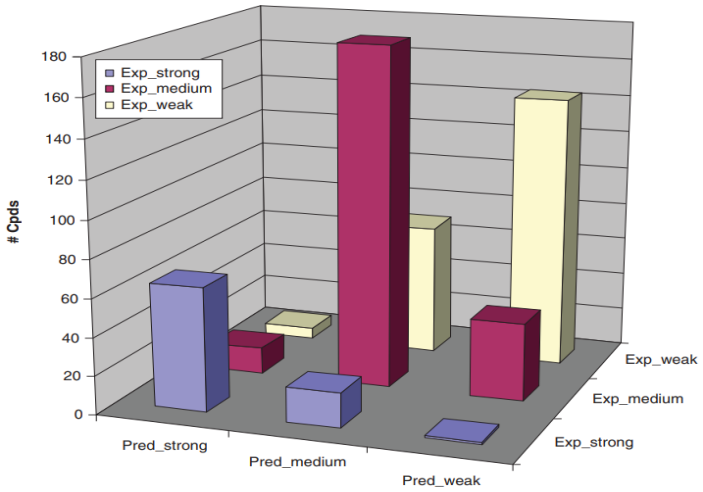

Fig. 6: Comparison of predicted and experimental CYP inhibition strengths (strong, medium, weak) for various compounds using a classification model

3) INDUCTION: CYP3A4 expression is regulated by the nuclear receptor PXR (pregane X receptor), the most well-known receptor linked to the induction of metabolic enzymes.

A recent study examined three previously published pharmacophore models using reliable, high-quality data. Two of them were deemed to be fairly predictive for PXR agonists by the authors, who recommended using a consensus pharmacophore model. Remarkably, the most varied training set served as the foundation for the worst model .Classification was done using a variety of machine learning techniques substances that either activate or do not activate PXR. The prediction accuracy of these models ranged from 73 to 87% for a small test of only 15 compounds.

SCOPE AND LIMITATIONS OF COMPUTATIONAL METHODS:

Table 1: Overview of Computational Approaches in Drug Metabolism Prediction

Investigation Area	Computational Methods	Scope & Limitations
Enzyme Structure & Mechanism	Homology modeling, QM, MD simulations	Detailed analysis of enzyme function; useful for short-lived intermediates.
Sites of Metabolism (SoMs)	ML, QSAR, docking, shape-based models	Accurately predicts top 3 likely SoMs in 70–90% of cases.
Metabolite Structure	Knowledge-based systems, data mining	Can generate many metabolites; ranking them remains challenging.
Metabolic Rates	QM, MD, limited QSAR	Poor general prediction; works only in narrow chemical spaces.

Investigation Area	Computational Methods	Scope & Limitations
Drug-Enzyme Interactions	QSAR, free energy methods	Predicts affinity/inhibition with enough data; costly and complex.
Bioactivity & Toxicity	Ligand/structure-based, rule-based	Useful but limited by data; high false positives and challenges with metabolites.
Metabolite ID (Met ID)	Metabolite generation & spectral tools	Rapidly improving with open-source tools and better data access.

Conclusion: Computational methods have become essential in predicting drug metabolism, offering fast and cost-effective insights into metabolic stability, enzyme binding, and metabolite formation. While these tools show strong performance in specific areas, challenges like predicting metabolic rates and metabolite bioactivity remain. Ongoing advancements in AI, data integration, and experimental validation will further enhance their accuracy and utility in drug discovery.

REFERENCES:

1. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 2003 ; 22 : 69-77.
2. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affi nity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl 2002 ; 41 : 2644 -76.
3. De Groot MJ, Kirton SB, Sutcliffe MJ. In silico methods for predicting ligand binding determinants of cytochromes P450. Curr Top Med Chem 2004 ; 4 : 1803 -24.
4. Ekins S, Nikolsky Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. Trends Pharmacol Sci 2005 ; 26 : 202 -9.
5. Ekins S, Nikolsky Y, Nikolskaya T. Techniques: application of systems biology to absorption, distribution, metabolism, excretion and toxicity. Trends Pharmacol Sci 2005 ; 26 : 202 -9.
6. Göller A. H, Hennemann M, Hillisch, A. et al. CypScore – a quantum chemistry based approach for the prediction of likely sites of P450-mediated metabolism. 4th Joint Sheffi eld Conference on Chemoinformatics.
7. Ogliaro F, Harris N, Cohen S, et al. A model ‘rebound’ mechanism of hydroxylation by cytochrome P450: stepwise and effectively concerted pathways, and their reactivity patterns. J Am Chem S.
8. Collins JR, Loew GH. Theoretical study of the product specifi city in the hydroxylation of camphor, norcamphor, 5,5-difl uorocamphor, and pericyclocamphanone by cytochrome P-450cam. J Biol Chem 1988 ; 263 : 3164 -70.
9. Korzekwa KR, Jones JP, Gillette JR. Theoretical studies on cytochrome P-450 mediated hydroxylation: A predictive model for hydrogen atom abstractions. J Am Chem Soc 199.
10. Pudzianowski AT, Loew GH. Quantum-chemical studies of model cytochrome P450 hydrocarbon oxidation mechanisms. 1. A MINDO/3 study of hydroxylation and epoxidation pathways for methane and ethylene. J Am Chem Soc 1980 ; 102 : 5443 -9.
11. van Waterschoot, R. A. B. & Schinkel, A. H. A critical analysis of the interplay between cytochrome P450 3A and P-glycoprotein: Recent insights from knockout and transgenic mice. Pharmacol. Rev. 63, 390-410 .
12. Koutsoukas, A. et al. From in silico target prediction to multi-target drug design: Current databases, methods and applications. J. Proteomics 74, 2554- 2574.
13. Bender, A. in Drug Metabolism Prediction (ed J. Kirchmair) 397-411 (WileyVCH, Weinheim).
14. Leach, A. G. & Kidley, N. J. in Drug Metabolism Prediction (ed J. Kirchmair) 103-132.
15. Kirchmair, J. et al. Computational prediction of metabolism: Sites, products, SAR, P450 enzyme dynamics, and mechanisms. J. Chem. Inf. Model. 52, 617- 648.
16. Gleeson, M. P. et al. The challenges involved in modeling toxicity data in silico: A review. Curr. Pharm. Des. 18, 1266-1291.
17. Phuangsawai, O., Hannongbua, S. & Gleeson, M. P. in Drug Metabolism Prediction (ed J. Kirchmair) 321-349.
18. Zamora, I., Fontaine, F., Serra, B. & Plasencia, G. High-throughput, computer assisted, specific MetID. A revolution for drug discovery. Drug Discov. Today Technol. 10, e199-e205.
19. Eriksson L, Jaworska J, Worth AP, et al. Methods for reliability and uncertainty assessment and for applicability evaluations of classifi cation- and regression-based QSARs. Environ Health Perspect 2003 ; 111 : 1361 -75.
20. Heady L, Fernandez-Serra M, Mancera RL, et al. Novel structural features of CDK inhibition revealed by an ab initio computational method combined with dynamic simulations. J Med Chem 2006 ; 49 : 5141 -53.