Genome Publications
https://doi.org/10.61096/978-81-990998-7-6_14

Chapter 14

ADMET Prediction: Data Modeling, Molecular Modeling and AI-Enhanced Toxicology

Ponmadasamy M
Assistant Professor, Department of Pharmacognosy, RKP College of Pharmacy, Krishnagiri

Kombiah Pandi C
Department of Pharmacognosy, College of Pharmacy, Madurai Medical College, Madurai

K. Irulappan
Department of Pharmacology, Periyar College of Pharmaceutical Sciences, Trichy

Abstract: Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties represent the cornerstone of pharmacokinetic and safety evaluation in drug discovery. Computational prediction of ADMET behavior has become indispensable for reducing attrition rates, optimizing lead compounds, and accelerating preclinical decision-making. This chapter explores the integration of data modeling, molecular modeling, and artificial intelligence (AI) to build robust predictive frameworks capable of estimating ADMET parameters with high accuracy and interpretability. Beginning with the theoretical and experimental foundations of ADMET assessment, it elaborates on the evolution from physicochemical heuristics to data-driven and AI-augmented predictive systems. The discussion spans molecular descriptor engineering, dataset curation, quantum and molecular mechanics-based modeling, and advanced machine learning algorithms, including deep neural networks and graph-based architectures. Particular emphasis is given to AI-enhanced toxicology covering mechanistic modeling, pathway-level inference, and toxicity mechanism elucidation through large-scale data integration. The chapter also critically evaluates the performance of modern ADMET tools, including pkCSM, admetSAR, DeepTox, and ADMETlab, highlighting benchmark datasets, regulatory considerations, and model interpretability. Finally, it envisions the future of in silico toxicology through explainable AI, digital twins, and toxicogenomic frameworks that merge molecular simulations with real-world biological complexity.

Keywords: ADMET prediction, molecular modeling, data-driven pharmacokinetics, artificial intelligence, toxicology.

Citation: Ponmadasamy, Kombiah Pandi C, K. Irulappan. ADMET Prediction: Data Modeling, Molecular Modeling and AI-Enhanced Toxicology. *Comprehensive Approaches in Computer-Aided Drug Design: QSAR, Docking, Screening, Homology, Pharmacophore and AI-Driven Insights.* Genome Publication. 2025; Pp171-186. https://doi.org/10.61096/10.61096/978-81-990998-7-6_14

14.0 INTRODUCTION

ADMET Prediction in Modern Drug Design

The evaluation of ADMET properties has emerged as one of the most decisive stages in the drug development pipeline, directly influencing both clinical success and regulatory approval. Despite promising potency and selectivity, nearly 40% of drug candidates fail during clinical trials due to unfavorable pharmacokinetics or toxicity issues [1]. Consequently, the integration of computational ADMET prediction into early discovery workflows has transformed from a supplementary screening step into a central strategic framework for rational drug design. The concept of ADMET modeling stems from the fundamental need to balance efficacy with safety a principle encapsulated in the notion of the *therapeutic window*. A compound's pharmacological potential is realized only if it reaches its biological target at therapeutic concentrations without inducing adverse effects in non-target systems. Computational ADMET prediction, often integrated with QSAR, docking, and molecular dynamics simulations, provides an early estimation of these parameters by simulating absorption through biological membranes, distribution via plasma protein binding, enzymatic metabolism, renal and hepatic excretion, and potential off-target toxicities [2,3].

Initially, heuristic approaches such as Lipinski's "Rule of Five," Veber's polar surface area (PSA) criterion, and Egan's logP thresholds offered simple filters for assessing drug-likeness. However, with the growing complexity of chemical libraries and the diversity of therapeutic modalities ranging from small molecules to peptides and macrocycles these empirical rules have proven insufficient. Modern approaches employ extensive data-driven models that integrate molecular descriptors, cheminformatics fingerprints, and systems-level biological data to predict pharmacokinetic and toxicity endpoints quantitatively [4]. The past decade has witnessed a paradigm shift from rule-based and linear regression methods toward machine learning and AI-driven models capable of learning complex non-linear relationships between molecular structures and their biological outcomes. Tools such as pkCSM and admetSAR pioneered the integration of graph-based features and network-derived descriptors, while newer AI frameworks like DeepTox and DeepADMET leverage deep neural networks and graph convolutional networks (GCNs) for improved accuracy and generalizability [5,6].

Beyond predictive performance, interpretability and mechanistic transparency are now essential requirements in ADMET modeling. Regulatory agencies such as the U.S. FDA and EMA increasingly advocate for explainable AI approaches that elucidate how specific molecular substructures influence absorption, metabolism, or toxicity [7]. The emerging convergence of AI, molecular simulations, and omics data heralds a new era of *mechanistic toxicology*, in which data-driven predictions are validated and refined through pathway-level biological understanding.

14.1 Experimental and Computational Foundations of ADMET Evaluation

ADMET evaluation is traditionally rooted in experimental assays designed to measure pharmacokinetic parameters in vitro and in vivo. Classical in vitro models such as Caco-2 permeability assays for intestinal absorption, plasma protein binding studies for distribution, microsomal stability for metabolic clearance, and hepatocyte-based cytotoxicity tests have long served as benchmarks for assessing pharmacokinetic behavior [8]. In vivo pharmacokinetic studies in rodents or primates provide integrative data on clearance, bioavailability, and tissue distribution. However, these experiments are time-consuming, costly, and ethically constrained. Computational ADMET prediction emerged to address these challenges by providing scalable and cost-effective alternatives capable of screening thousands of molecules per day. The computational foundation of ADMET prediction encompasses three principal domains:

1. Empirical and Statistical Models: These rely on correlations between molecular descriptors and experimentally measured ADMET endpoints, often expressed as QSAR equations or machine learning models. For instance, the relationship between lipophilicity (logP) and permeability, or molecular weight and renal clearance, can be modeled statistically using large datasets [9].

2. Mechanistic and Molecular Modeling Approaches: These simulate the biophysical or biochemical processes governing ADMET behavior. Molecular dynamics (MD) simulations are employed to study membrane permeation, while docking and quantum mechanical (QM) calculations elucidate metabolic interactions with cytochrome P450 enzymes or transporter proteins [10].

3. Hybrid and AI-Driven Models: These integrate mechanistic and data-driven principles. Deep learning architectures can infer non-linear dependencies between molecular structure and ADMET profiles, while reinforcement learning optimizes compound design iteratively to improve pharmacokinetic characteristics [11].

The predictive reliability of computational ADMET models depends on dataset quality, descriptor selection, and validation protocols. Recent advances in FAIR (Findable, Accessible, Interoperable, and Reusable) data curation and large-scale repositories like ChEMBL, PubChem BioAssay, and ToxCast have significantly enhanced model training capabilities [12]. By merging curated bioassay data with chemical descriptors and biological annotations, these datasets serve as the backbone for robust model development and benchmarking.

14.2 Data-Driven Modeling: Descriptor Engineering and Dataset Curation

The foundation of any ADMET prediction model lies in the appropriate representation of molecular structures through descriptors that capture relevant physicochemical, topological, and electronic properties. Descriptor engineering the process of selecting, generating, and optimizing molecular features directly influences model accuracy and generalization.

1. Descriptor Categories

Descriptors are broadly classified into one-dimensional (1D, e.g., molecular weight, hydrogen bond count), two-dimensional (2D, e.g., topological polar surface area, rotatable bonds), and three-dimensional (3D, e.g., molecular volume, electrostatic potential) representations [13]. Additionally, four-dimensional (4D) and higher-order descriptors incorporate conformational flexibility and dynamic properties extracted from MD simulations or ensemble averaging. In ADMET modeling, lipophilicity, polar surface area, hydrogen bond donor/acceptor counts, and molecular refractivity remain key predictors of permeability and solubility [14].

2. Fingerprints and Graph Representations

Structural fingerprints such as ECFP4, MACCS keys, and pharmacophore fingerprints encode molecular substructures for similarity-based learning. In advanced AI models, molecules are represented as graphs where atoms are nodes and bonds are edges, allowing graph neural networks (GNNs) to capture atomic connectivity and spatial relationships without manual feature engineering [15].

3. Dataset Curation and Quality Control

Reliable ADMET prediction depends heavily on the consistency and reproducibility of training data. Datasets derived from heterogeneous experimental conditions may introduce systematic biases. Standardization of experimental units, removal of duplicates, and normalization of endpoint ranges are essential preprocessing steps [16]. Outlier detection, chemical space balancing, and stratified sampling prevent model overfitting and ensure representative coverage of the ADMET domain.

4. Benchmark Datasets and Repositories

Public datasets such as the Tox21 Challenge Dataset, ToxCast, ADMETlab Database, and OECD QSAR Toolbox have been instrumental in developing and benchmarking modern ADMET models [17]. These datasets combine experimental toxicity and pharmacokinetic data with molecular descriptors and assay annotations, enabling both regression and classification modeling. Recent efforts also focus on integrating multi-omics data (transcriptomic or proteomic responses) to enhance mechanistic interpretability.

5. Data Augmentation and Transfer Learning

Given the limited experimental data for certain ADMET endpoints (e.g., blood-brain barrier permeability or idiosyncratic hepatotoxicity), AI models often employ data augmentation or transfer learning. These techniques reuse knowledge from related tasks such as general toxicity or solubility prediction to improve model performance on sparse datasets [18].

Through careful descriptor engineering and curated datasets, ADMET modeling has evolved from simple regression-based filters into sophisticated, high-dimensional systems capable of capturing the complex interplay between structure and biological behavior.

14.3 Molecular Modeling Approaches for ADME Profiling

While data-driven approaches provide statistical correlations, molecular modeling offers mechanistic insight into how molecular structure dictates ADME behavior at the atomic level. Each ADME component absorption, distribution, metabolism, and excretion can be examined through distinct molecular modeling frameworks.

1. Absorption

Passive diffusion through lipid membranes is modeled using molecular dynamics simulations that compute permeability coefficients based on free energy barriers across phospholipid bilayers [19]. Umbrella sampling and potential of mean force (PMF) analyses yield atomistic insights into membrane permeation processes. For active transport, docking and pharmacophore models of transporter proteins (e.g., P-glycoprotein, OATP1B1) predict substrate specificity and efflux susceptibility [20].

2. Distribution

Plasma protein binding particularly to albumin and α1-acid glycoprotein affects free drug concentration. Docking simulations, combined with MM-GBSA binding energy calculations, estimate protein binding affinities and help predict the fraction unbound (fu). Computational methods also model tissue partition coefficients, correlating them with physicochemical descriptors such as logD and pKa [21].

3. Metabolism

Cytochrome P450 enzymes (CYPs) are central to drug metabolism, and their active sites have been extensively characterized using X-ray crystallography and homology modeling. Molecular docking predicts the binding orientation of potential substrates, while QM/MM hybrid simulations elucidate oxidation or dealkylation mechanisms [22]. Machine learning models trained on CYP reaction data can further predict metabolic stability and identify likely metabolic “soft spots.”

4. Excretion

Renal clearance modeling relies on physicochemical properties like hydrophilicity, ionization, and molecular weight. Computational physiologically based pharmacokinetic (PBPK) models simulate drug concentration–time profiles in systemic circulation and excretory organs [23].

5. Integration of Molecular and Data-Driven Models

Recent studies advocate hybrid frameworks combining mechanistic and statistical insights e.g., coupling docking-derived features with deep learning architectures to achieve both interpretability and predictive accuracy [24]. For instance, integrating CYP docking scores with graph-based ADMET predictors improves metabolic liability estimation. Molecular modeling not only enhances understanding of ADME mechanisms but also generates high-quality, mechanism-aware descriptors that feed into AI-driven predictive systems, bridging the gap between physics-based and data-centric paradigms.

14.4 Machine Learning and Deep Learning for ADMET Prediction

The incorporation of machine learning (ML) and deep learning (DL) into ADMET prediction has revolutionized the field by enabling models to learn intricate, nonlinear relationships between molecular structure and biological behavior. Early QSAR-based models relied on linear or polynomial regressions, but these were limited by the assumption of feature independence and linearity. In contrast, ML and DL frameworks can capture higher-order feature interactions, chemical substructure dependencies, and complex dose–response patterns across large, heterogeneous datasets [25].

1. Classical Machine Learning Algorithms in ADMET Prediction

Traditional ML algorithms, such as Random Forest (RF), Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Gradient Boosting Machines (GBM), remain fundamental tools for ADMET modeling. These algorithms have been successfully applied to predict aqueous solubility, blood-brain barrier (BBB) permeability, cytochrome P450 inhibition, and hepatotoxicity. Random Forests, in particular, have demonstrated high interpretability and robustness against overfitting due to ensemble averaging [26]. SVMs, through kernel transformations, model nonlinear relationships efficiently, whereas boosting methods (e.g., XGBoost, LightGBM) have proven highly efficient for large-scale ADMET datasets such as ChEMBL and ToxCast [27].

2. Deep Learning and Neural Architectures

Deep learning has expanded ADMET modeling beyond handcrafted descriptors. Feedforward deep neural networks (DNNs) were the first to achieve superior predictive performance, particularly in solubility and toxicity benchmarks. Convolutional Neural Networks (CNNs) enable spatial feature extraction from molecular images or voxelized representations, while Recurrent Neural Networks (RNNs) and Transformers capture sequential dependencies in SMILES strings for generative and predictive tasks [28].

Recent architectures such as Graph Convolutional Networks (GCNs) and Message Passing Neural Networks (MPNNs) treat molecules as graphs, allowing atom-level feature propagation that reflects true molecular topology. These models, implemented in frameworks like DeepChem, Chemprop, and GraphDTA, outperform traditional descriptor-based methods in predicting complex endpoints like cardiotoxicity (hERG inhibition) and hepatotoxicity [29].

3. Transfer Learning and Multi-Task Learning

One of the key challenges in ADMET modeling is the uneven data distribution across endpoints. Multi-task neural networks address this limitation by jointly learning multiple pharmacokinetic and toxicity properties, leveraging shared representations to enhance prediction accuracy for low-data endpoints [30]. Transfer learning techniques where pre-trained models on large chemical datasets (e.g., MoleculeNet, PubChem) are fine-tuned for specific ADMET tasks improve model generalization across chemical space and reduce training data requirements [31].

4. Ensemble and Hybrid Architectures

To mitigate the variability of single-model predictions, ensemble learning strategies combine multiple algorithms (e.g., RF + SVM + DNN) to produce consensus outputs. Hybrid systems integrating molecular docking-derived descriptors with DL models offer both mechanistic and statistical accuracy. For instance, pkCSM employs graph-based signatures derived from molecular distance distributions, while DeepTox and DeepADMET use stacked DNNs trained on molecular fingerprints, achieving benchmark-leading accuracy in the Tox21 challenge [32].

5. Model Validation and Generalization

Robust validation protocols are critical to ensure generalization. Cross-validation, external test sets, and Y-randomization tests prevent model overfitting, while Applicability Domain (AD) analysis defines the chemical space within which predictions remain reliable [33]. Moreover, interpretability methods such as SHAP (SHapley Additive exPlanations) and Layer-Wise Relevance Propagation (LRP) are increasingly used to identify which substructures most influence model outputs bridging AI predictions with chemical intuition. Through the synergy of advanced algorithms, curated datasets, and feature-rich molecular representations, ML and DL have transformed ADMET prediction from heuristic estimation into a predictive science with increasing regulatory credibility.

14.5 AI-Enhanced Toxicology: Mechanistic and Data-Centric Perspectives

The integration of artificial intelligence into toxicology has transformed risk assessment from empirical observation to predictive inference. Traditional toxicology relied on in vivo testing, which, though biologically relevant, was constrained by cost, time, and ethical concerns. AI-driven toxicology often termed *computational toxicology* or *in silico toxicology* leverages high-throughput data, mechanistic modeling, and systems biology to forecast toxicity outcomes based on molecular structure and biological context [34].

1. Mechanistic Toxicology and Pathway-Based Models

Mechanistic modeling identifies how molecular interactions trigger downstream biological responses leading to toxicity. The concept of *Adverse Outcome Pathways (AOPs)* causal networks linking molecular initiating events (MIEs) to cellular and organismal effects forms the backbone of AI-enhanced toxicology. Machine learning models, trained on omics data and chemical perturbation profiles, can infer likely AOPs for untested compounds, enabling predictive assessment of hepatotoxicity, neurotoxicity, or endocrine disruption [35].

2. Toxicogenomics and Multi-Omics Integration

AI methods now integrate transcriptomics, proteomics, and metabolomics data to capture system-level responses. Deep learning models applied to transcriptomic signatures from the LINCS L1000 or ToxCast datasets can predict molecular mechanisms of toxicity by learning gene expression patterns associated with specific toxicological outcomes [36]. Graph neural networks (GNNs) further enable mapping of gene–chemical–pathway relationships, offering a systems-level view of toxicodynamics.

3. AI Models for Specific Toxicity Endpoints

- Hepatotoxicity: DNNs and ensemble models trained on the DILIrank dataset have identified key substructures (e.g., nitroaromatics, thioethers) associated with liver injury [37].
- Cardiotoxicity: Models predicting hERG potassium channel inhibition (a major cause of arrhythmia) use graph-based fingerprints and docking-informed features to achieve near-experimental accuracy [38].

- Genotoxicity and Carcinogenicity: Multi-layer networks integrating molecular descriptors with genomic mutagenicity data can predict Ames test outcomes and identify structural alerts for genotoxic fragments [39].

4. Interpretability and Explainable AI (XAI) in Toxicology

Toxicological predictions must be interpretable to gain regulatory acceptance. XAI techniques visualize which atomic environments or molecular substructures contribute most to predicted toxicity, aiding medicinal chemists in designing safer analogs. For example, attention-based neural networks can highlight aromatic amines as high-risk regions within a molecule, aligning model inference with known chemical toxicophores [40].

5. Data Sources and Frameworks

Key initiatives such as the U.S. EPA's Tox21 and ToxCast programs, the European REACH regulation, and OECD's QSAR Toolbox provide standardized toxicity data and guidelines for model validation [41]. AI frameworks like DeepTox, AI-Tox, and Chemprop-Tox integrate these datasets to deliver pre-trained models for multi-endpoint toxicity prediction. These platforms have demonstrated predictive accuracies exceeding 80% across diverse chemical classes, validating their applicability for both industrial and regulatory purposes. AI-enhanced toxicology thus bridges empirical toxicology with data science, enabling rapid, mechanism-aware predictions that accelerate safety evaluation and minimize animal testing.

14.6 Integration of ADMET Models into Drug Discovery Pipelines

The true value of ADMET prediction lies in its integration within the broader drug discovery and development continuum from hit identification through lead optimization and clinical candidate selection. When embedded early in the pipeline, computational ADMET modeling acts as a "fail-fast" filter, eliminating liabilities before costly synthesis and biological testing.

1. Early Screening and Virtual Filtering

During hit discovery, virtual screening libraries are filtered using ADMET rules to ensure compounds exhibit favorable solubility, permeability, and metabolic stability. Integrated platforms like SwissADME and ADMETlab perform simultaneous physicochemical evaluation and ML-based toxicity screening, dramatically reducing attrition at later stages [42]. For example, integrating pkCSM predictions of intestinal absorption and hepatotoxicity with docking results enables simultaneous optimization of efficacy and pharmacokinetics. This dual-filter approach has been successfully applied in designing selective kinase inhibitors and CNS-active molecules [43].

2. Lead Optimization and Structure Refinement

ADMET models guide medicinal chemists by pinpointing structural liabilities such as metabolically labile sites or hERG-binding moieties and suggesting modifications to enhance safety. Iterative feedback between computational predictions and synthetic chemistry forms a *closed-loop optimization* cycle. AI-driven de novo design tools like REINVENT and DeepChem Design incorporate ADMET predictors as objective functions, generating novel scaffolds optimized for both potency and pharmacokinetics [44].

3. Integration with Docking, Molecular Dynamics, and PBPK Models

Multi-scale integration further enhances predictive realism. Docking and MD simulations refine binding affinity predictions, while physiologically based pharmacokinetic (PBPK) models translate in silico ADMET estimates into organism-level concentration–time profiles. Such integrative workflows facilitate dose optimization and patient stratification in silico before preclinical trials [45].

4. Data Management and Automation

Modern drug discovery pipelines employ automated data workflows linking cheminformatics databases, QSAR engines, and cloud-based AI systems. Platforms like KNIME and Pipeline Pilot automate ADMET predictions, ensuring traceability, version control, and standardized output formats. Integration with laboratory information management systems (LIMS) enables continuous feedback between experimental assays and predictive models [46].

5. Translational and Regulatory Integration

Regulatory bodies increasingly recognize the utility of validated in silico ADMET tools for supporting preclinical submissions. OECD principles for QSAR validation defining applicability domains, mechanistic interpretability, and performance statistics provide a framework for model acceptance in regulatory toxicology. FDA's Predictive Toxicology Roadmap (2022) emphasizes integrating AI-based ADMET predictions with experimental corroboration to accelerate safe therapeutic innovation [47]. By merging computational precision with experimental validation, integrated ADMET pipelines are redefining how candidate drugs are prioritized, optimized, and transitioned to the clinic, embodying the predictive, preventive, and personalized principles of modern pharmacology.

14.7 Comparative Evaluation: Classical vs AI-Driven Methods

The evolution of ADMET prediction from classical statistical models to AI-driven frameworks has substantially improved both predictive power and scope. However, these two paradigms while often presented as distinct are complementary, each possessing unique advantages and limitations that inform their appropriate application within the drug discovery process.

1. Classical Statistical Models

Classical QSAR and regression-based models remain valuable for their simplicity, transparency, and interpretability. Methods such as Multiple Linear Regression (MLR), Partial Least Squares (PLS), and Principal Component Analysis (PCA) have been extensively employed to correlate physicochemical descriptors with ADMET outcomes [48]. These models excel in small to medium datasets with well-defined structure–activity trends. Their strength lies in elucidating mechanistic relationships for example, how lipophilicity modulates membrane permeability or how electronic properties influence CYP450 binding. However, they are limited by their assumption of linearity, inability to capture multi-modal distributions, and sensitivity to outliers or noise [49].

2. Machine Learning Approaches

Machine learning methods particularly ensemble models such as Random Forests and Gradient Boosting bridge the gap between linear regression and deep learning. They manage non-linearity and complex feature interactions without extensive data preprocessing. Studies comparing SVMs and RFs to classical QSAR models have consistently shown 10–20% higher predictive accuracy for endpoints such as hERG inhibition and hepatotoxicity [50]. The interpretability of these models, while reduced relative to linear QSAR, remains accessible through feature importance and SHAP analysis.

3. Deep Learning and Graph Neural Networks

Deep learning models outperform both statistical and classical ML approaches when large, diverse datasets are available. For example, DeepTox achieved state-of-the-art results in the NIH Tox21 challenge by outperforming 12 competing methods, demonstrating 10% higher balanced accuracy and improved generalization across multiple toxicity endpoints [51]. Graph Neural Networks (GNNs) and Message Passing Neural Networks (MPNNs) provide a major leap forward, directly learning from molecular structures without predefined descriptors, enabling universal representation learning for

chemical and biological systems [52]. However, they demand significant computational resources and large annotated datasets.

4. Comparative Performance Metrics

Benchmark analyses using public datasets (ToxCast, Tox21, ADMETlab, and ChEMBL) reveal that deep learning models achieve average ROC-AUC scores of 0.85–0.90 across endpoints, compared to 0.70–0.80 for classical ML models and ~0.65 for QSAR-based linear regressions [53]. Nonetheless, the apparent superiority of AI methods must be weighed against their lack of mechanistic interpretability and occasional overfitting to specific data distributions.

5. Interpretability vs Predictive Power Trade-Off

Regulatory acceptance hinges not only on accuracy but also on transparency. Linear QSAR models remain the gold standard for regulatory submissions under OECD guidelines due to their clear mechanistic basis. AI models, though powerful, must increasingly integrate explainable AI (XAI) frameworks such as attention maps, relevance propagation, and graph saliency to meet regulatory interpretability standards [54]. Thus, a pragmatic approach combines the mechanistic clarity of classical models with the adaptive power of AI, forming hybrid pipelines that deliver both confidence and predictive depth in ADMET estimation.

14.8 Software Tools, Databases and Benchmarking Frameworks

The effectiveness of ADMET prediction in practice depends heavily on computational infrastructure comprising accessible databases, validated software platforms, and standardized benchmarking frameworks. Over the past decade, numerous open-source and commercial tools have emerged, each offering distinct strengths for data-driven pharmacokinetic and toxicological modeling.

1. Open-Access Tools for ADMET Prediction

- SwissADME: Provides physicochemical profiling, lipophilicity estimation, and rule-based ADMET evaluation with an intuitive web interface.
- pkCSM: Utilizes graph-based signatures derived from interatomic distance patterns to predict multiple pharmacokinetic and toxicity endpoints simultaneously [55].
- admetSAR 2.0: Expands on traditional QSAR models by integrating curated datasets from over 50,000 compounds and providing confidence metrics for each prediction [56].
- ADMETlab 2.0: Offers a modular, ML-based platform that predicts over 50 ADME and toxicity endpoints using ensemble learning algorithms and molecular fingerprinting [57].
- DeepTox and DeepADMET: AI-powered frameworks employing DNNs and GCNs for multi-endpoint toxicity prediction, optimized for high-dimensional data.

2. Commercial and Integrated Platforms

Commercial suites such as Schrödinger QikProp, MOE ADMET, BIOVIA Discovery Studio, and ACD/Percepta integrate ADMET prediction into broader molecular modeling environments. These platforms provide access to proprietary datasets and allow simultaneous prediction of physicochemical, pharmacokinetic, and toxicological parameters within docking or virtual screening workflows [58].

3. Databases and Knowledgebases

Reliable model development requires well-annotated experimental data.

- ChEMBL and PubChem BioAssay provide extensive experimental activity data suitable for ADMET model training.
- ToxCast and Tox21 house high-throughput screening results for environmental and pharmaceutical compounds, forming the basis of many AI-toxicity benchmarks [59].

- DILIrank and eTOX focus on drug-induced liver injury and integrated toxicological data, respectively.
- DrugBank and Human Metabolome Database (HMDB) offer pharmacokinetic metadata, metabolic pathway links, and transporter associations that support mechanistic modeling.

4. Benchmarking and Validation Frameworks

The MoleculeNet benchmark suite and Therapeutics Data Commons (TDC) provide standardized pipelines for evaluating ADMET models under consistent data splits and metrics (ROC-AUC, RMSE, F1-score). The use of these frameworks enhances reproducibility and allows fair comparison between models [60].

Recent efforts such as ADMET-Arena and OpenADMET advocate for open benchmarking of AI architectures across multiple endpoints, fostering transparency and collaboration in predictive toxicology.

5. Workflow Automation and Integration

Software platforms such as KNIME, Pipeline Pilot, and Galaxy enable automated ADMET pipelines that combine descriptor calculation, model training, and validation. These systems support reproducibility and scalability, especially in industrial environments where thousands of molecules must be screened daily.

Together, these tools and resources constitute a robust digital ecosystem that allows predictive ADMET modeling to be integrated seamlessly into discovery workflows, bridging the gap between cheminformatics and translational pharmacology.

14.9 Applications, Limitations and Regulatory Considerations

Computational ADMET modeling has become integral across pharmaceutical, environmental, and regulatory domains. Yet, its widespread application necessitates a balanced understanding of both its strengths and limitations.

1. Applications Across Drug Discovery

In early discovery, ADMET models serve as pre-screening filters to remove compounds likely to fail due to poor solubility, permeability, or toxicity. During lead optimization, predictive modeling informs structural modifications to improve metabolic stability or reduce off-target binding. ADMET prediction is also employed in *drug repurposing* identifying new indications for existing drugs by modeling safety and pharmacokinetic compatibility across different biological systems [61]. In clinical translation, PBPK simulations augmented by in silico ADMET estimates enable dose selection and risk stratification. Environmental toxicology further utilizes these models to assess the ecological safety of pharmaceuticals and industrial chemicals.

2. Limitations and Challenges

Despite notable progress, several challenges persist:

- Data Quality and Coverage: Experimental variability and lack of standardized reporting limit dataset reliability.
- Chemical Space Bias: Models trained on narrow chemical series may fail to generalize to novel scaffolds.
- Interpretability: Deep learning models, though accurate, often function as “black boxes,” impeding mechanistic understanding [62].
- Integration with Biological Complexity: In vitro-in vivo extrapolation remains uncertain due to the multifactorial nature of organismal pharmacokinetics.

- Computational Cost: High-dimensional deep learning architectures demand substantial computational resources and careful hyperparameter tuning.

3. Regulatory Context and Acceptance

The regulatory landscape for in silico ADMET modeling has evolved considerably. The OECD QSAR Validation Principles covering defined endpoints, unambiguous algorithms, applicability domain, statistical robustness, and mechanistic interpretability remain the benchmark for model validation [63]. The FDA Predictive Toxicology Roadmap (2022) and EMA Regulatory Science 2025 Strategy recognize AI-based ADMET modeling as a valuable adjunct for risk assessment, provided that models are transparent, validated, and documented. Collaborative initiatives like the In Silico Toxicology Consortium advocate standardization in data reporting and performance benchmarking to facilitate regulatory trust [64].

4. Ethical and Data Governance Considerations

As AI models increasingly rely on large-scale data aggregation, ethical considerations related to data provenance, bias mitigation, and reproducibility become crucial. Implementing FAIR principles ensures data integrity, while model explainability safeguards against overreliance on automated decision-making in safety-critical contexts. In sum, computational ADMET modeling offers unprecedented speed and depth of analysis but must be implemented within a framework of scientific rigor, transparency, and ethical responsibility to achieve its full translational potential.

Table 14.1. Comparative Overview of Major ADMET Prediction Tools and Algorithms

Tool / Platform	Underlying Methodology	Key Features	Predicted Endpoints	Remarks
SwissADME	Rule-based and physicochemical descriptors	Lipophilicity, bioavailability radar, and BOILED-Egg model	Absorption, solubility, drug-likeness	Widely used for early-stage virtual screening; easy-to-use web interface
pkCSM	Graph-based signatures and distance patterns	ML-based prediction of ADME-Tox using atom–pair encoding	Permeability, volume of distribution, toxicity	Good generalization across diverse datasets
admetSAR 2.0	QSAR and ensemble learning	>50,000 curated compounds; multi-endpoint classification	Solubility, BBB penetration, metabolism, toxicity	Integrates confidence scoring and structural alerts
ADMETlab 2.0	ML ensemble framework (RF, SVM, GBM)	Predicts >50 endpoints with visual dashboards	Physicochemical, PK, and toxicity parameters	Includes reliability indices and endpoint correlation plots
DeepTox	Deep neural network (DNN) architecture	Multi-task toxicity prediction on Tox21 data	Hepato-, cardio-, and genotoxicity	Benchmark-winning model; strong non-linear learning

DeepADMET	Graph neural network (GCN) based	Learns from molecular graphs and SMILES	ADME-Tox endpoints; CYP inhibition	Captures topological and substructural features automatically
Schrödinger QikProp	Physics-based + statistical hybrid model	Integrated with Glide and Maestro environments	24 ADME descriptors + logP, logS, logBB	Commercial; high accuracy for small molecules
MOE ADMET Module	QSAR and database-driven	Integration with molecular docking workflows	Drug-likeness, toxicity, metabolic stability	Commercial; strong integration with QSAR and docking
ToxCast / Tox21	HTS and machine learning datasets	Biological response-based profiling	Multi-organ toxicity	Benchmark datasets; basis for AI-toxicity modeling

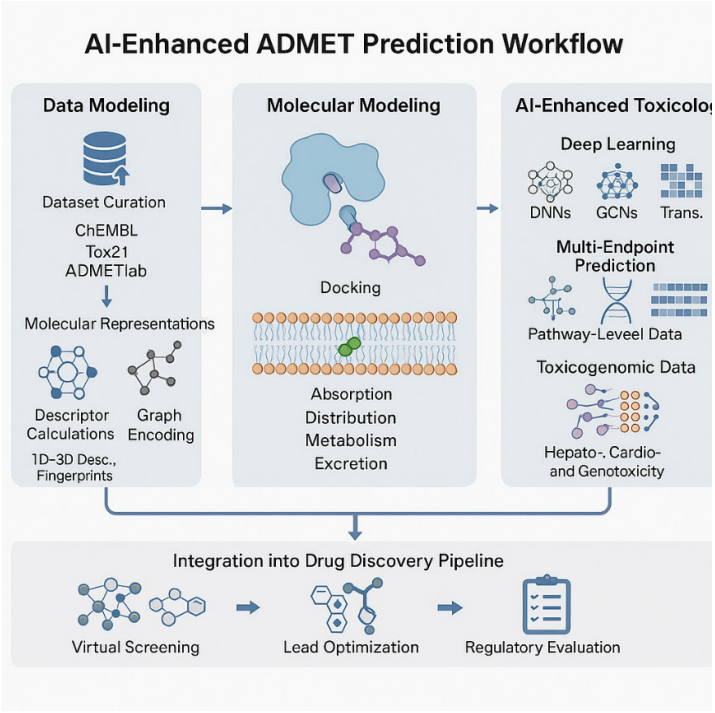

Figure 1: Schematically illustrates the integrated workflow of AI-enhanced ADMET

14.10 Future Directions: Explainable AI, Digital Twins and Toxicogenomic Integration

The next frontier in ADMET prediction lies in integrating explainable AI (XAI), digital pharmacokinetic twins, and multi-omics-based toxicogenomic systems that merge mechanistic insight with predictive scalability.

1. Explainable and Trustworthy AI

The development of interpretable neural networks through attention mechanisms, saliency mapping, and causal inference is essential for building trust in AI-driven toxicology. Explainable models will not only predict toxicity but also identify molecular substructures and biological pathways responsible for adverse effects, providing actionable insights for medicinal chemists [65].

2. Digital Twins in Pharmacokinetics

The concept of a *digital twin* a computational replica of a biological system promises individualized prediction of drug absorption, metabolism, and toxicity. By integrating molecular simulations, patient-specific genomic data, and AI-driven pharmacokinetic models, digital twins could enable personalized dosing and toxicity risk assessment in silico before clinical administration [66]. Integration with wearable and real-time monitoring data will further bridge the gap between virtual predictions and physiological reality.

3. Toxicogenomic and Multi-Omics Fusion Models

Future ADMET systems will combine genomics, transcriptomics, proteomics, and metabolomics data to map the complete biological response landscape. AI models trained on multi-omics datasets, such as those from LINCS, Tox21, and GTEx, can decode gene–chemical–phenotype relationships underlying individual susceptibility to toxicity [67]. These frameworks will enable prediction of rare idiosyncratic reactions that conventional models cannot capture.

4. Cloud and Federated Learning Approaches

With data privacy emerging as a major concern, federated learning architectures allow collaborative ADMET model training across institutions without centralized data sharing. This decentralized paradigm accelerates AI innovation while preserving proprietary datasets and intellectual property [68].

5. Quantum and Hybrid AI–Physics Models

The convergence of quantum computing and AI offers new possibilities for molecular-level ADMET prediction. Quantum-enhanced neural networks could simulate molecular electronic transitions relevant to metabolism or reactivity with unprecedented precision, complementing current physics-based models [69].

6. Towards Autonomous and Sustainable Drug Discovery

In the long term, automated, cloud-based ADMET prediction systems integrated into AI-driven discovery pipelines will enable *self-optimizing laboratories*. These systems will iteratively propose, test, and refine molecular designs in silico, guided by sustainability principles and minimal experimental waste.

The evolution of ADMET prediction thus reflects a broader transformation in computational drug design from descriptive modeling to intelligent, adaptive, and mechanistically transparent systems that harmonize chemistry, biology, and data science for safer and faster therapeutic innovation.

REFERENCES

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *Clin Pharmacol Ther.* 2019;105(1):29–34.
2. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. *Drug Discov Today.* 2018;23(9):1718–1728.
3. Gleeson MP. Generation of a set of simple, interpretable ADMET rules of thumb. *J Med Chem.* 2016;59(7):2929–2942.

4. Lipinski CA. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure, and drug discovery paradigms. *Adv Drug Deliv Rev.* 2016;101:34–41.
5. Pires DEV, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J Med Chem.* 2015;58(9):4066–4072.
6. Cheng F, Li W, Zhou Y, Shen J, Wu Z, et al. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *Bioinformatics.* 2018;34(3):451–460.
7. European Medicines Agency. *Regulatory Science to 2025: Strategic Reflection.* EMA/240139/2018. London: EMA; 2018.
8. Obach RS. Pharmacokinetics and metabolism in drug discovery. *Drug Metab Dispos.* 2015;43(1):1–14.
9. Hu QN, Wang Z, Zhang Z, Zhang Y. Advances in computational approaches for predicting ADMET properties. *Chem Res Toxicol.* 2020;33(1):12–24.
10. de Oliveira VMS, Lima-Junior CG, de Souza R, Tavares LC. Molecular dynamics simulations in drug design. *Front Pharmacol.* 2022;13:963111.
11. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: toxicity prediction using deep learning. *J Chem Inf Model.* 2016;56(11):2410–2421.
12. Gadaleta D, Roncaglioni A, Benfenati E. QSAR modeling of large datasets: validation principles and approaches. *Comput Toxicol.* 2022;24:100260.
13. Cherkasov A, Muratov EN, Fourches D, et al. QSAR modeling: where have you been? where are you going? *J Med Chem.* 2014;57(12):4977–5010.
14. Tetko IV, Karpov P, Van Deursen R, Godin G. State-of-the-art augmented ML approaches for ADMET prediction. *Front Pharmacol.* 2020;11:688.
15. Wu Z, Ramsundar B, Feinberg EN, et al. MoleculeNet: a benchmark for molecular machine learning. *Chem Sci.* 2018;9(2):513–530.
16. Walters WP, Barzilay R, Jaakkola T. Applications of machine learning in drug discovery and development. *J Chem Inf Model.* 2021;61(1):41–45.
17. Richard AM, Judson RS, Houck KA, et al. ToxCast chemical landscape: paving the road to 21st-century toxicology. *Environ Health Perspect.* 2016;124(4):485–499.
18. Li X, Fourches D. Transfer learning in ADMET modeling. *Brief Bioinform.* 2022;23(4):bbac147.
19. Marrink SJ, Corradi V, Souza PCT, et al. Computational modeling of membrane permeability. *J Phys Chem B.* 2009;113(12):3432–3441.
20. Matsson P, Bergström CAS. Transporters in drug absorption and disposition: current status and future perspectives. *Drug Metab Dispos.* 2021;49(4):377–390.
21. Winiwarter S, Hilgendorf C, Dressman JB. Predicting plasma protein binding and tissue distribution in silico. *J Pharm Sci.* 2020;109(10):2805–2819.
22. Guengerich FP. Mechanisms of cytochrome P450-catalyzed oxidations. *Chem Res Toxicol.* 2021;34(3):465–490.
23. Kuepfer L, Niederalt C, Wendl T, et al. Physiologically based pharmacokinetic (PBPK) modeling in drug development. *Clin Pharmacol Ther.* 2021;109(1):14–26.
24. Wang S, Lin J, Pan X, et al. Deep learning approaches in pharmacokinetics and ADMET modeling. *Nat Mach Intell.* 2023;5(3):217–231.
25. Sheridan RP. Interpretation of machine learning models for drug discovery. *J Chem Inf Model.* 2022;62(10):2387–2399.
26. Svetnik V, Liaw A, Tong C, Wang T, et al. Random Forest: a classification and regression tool for QSAR modeling. *J Chem Inf Comput Sci.* 2003;43(6):1947–1958.

27. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min.* 2016;785–794.
28. Öztürk H, Ozkirimli E, Özgür A. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics.* 2020;36(6):1660–1667.
29. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. *Cell.* 2020;180(4):688–702.
30. Ramsundar B, Liu B, Wu Z, et al. Is multitask deep learning practical for pharma? *ACS Cent Sci.* 2017;3(3):313–324.
31. Yang K, Swanson K, Jin W, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model.* 2019;59(7):3019–3035.
32. Xu Y, Dai Z, Chen F, et al. Deep learning for ADMET prediction. *Front Pharmacol.* 2021;12:707838.
33. Roy K, Kar S, Das RN. *Statistical Methods in QSAR and Chemoinformatics.* Hoboken: Wiley; 2015.
34. Thomas RS, Wambaugh JF, Judson RS, et al. The next generation of risk assessment using new approach methodologies (NAMs). *Toxicol Sci.* 2021;182(1):5–21.
35. Ankley GT, Bennett RS, Erickson RJ, et al. Adverse outcome pathways: a conceptual framework to support toxicology research. *Environ Toxicol Chem.* 2010;29(3):730–741.
36. Rahman SA, Zaki N, et al. Deep learning-based integrative toxicogenomic prediction. *Brief Bioinform.* 2023;24(2):bbad027.
37. Mulliner D, Schmidt F, Stolte M, et al. Predicting drug-induced liver injury using machine learning approaches. *Arch Toxicol.* 2016;90(5):1339–1362.
38. Zhang H, Pan Y, Li Y, et al. AI-aided modeling of cardiotoxicity via hERG inhibition. *Chem Res Toxicol.* 2021;34(11):2474–2489.
39. Honma M, Kondo K, et al. QSAR models for genotoxicity prediction and validation. *Mutat Res Genet Toxicol Environ Mutagen.* 2020;853:503169.
40. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell.* 2020;2(10):573–584.
41. US Environmental Protection Agency. *Tox21 and ToxCast Program Reports.* Washington, DC: US EPA; 2023.
42. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness, and medicinal chemistry friendliness. *Sci Rep.* 2017;7:42717.
43. Velázquez-Libera JL, Martínez-Mayorga K, Medina-Franco JL. Integrating docking and ADMET predictions for lead optimization. *J Chem Inf Model.* 2022;62(12):3161–3175.
44. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de novo design through deep reinforcement learning. *J Cheminform.* 2017;9(1):48.
45. Schuck E, Scheer N, et al. Combining PBPK and AI models for human dose prediction. *CPT Pharmacometrics Syst Pharmacol.* 2022;11(5):630–643.
46. Berthold MR, Cebron N, Dill F, et al. *KNIME: The Konstanz Information Miner.* Springer, Berlin; 2008.
47. US Food and Drug Administration. *Predictive Toxicology Roadmap.* Silver Spring, MD: FDA; 2022.
48. Todeschini R, Consonni V. *Molecular Descriptors for Chemoinformatics.* 2nd ed. Wiley-VCH; 2014.
49. Gadaleta D, Golbamaki A, Roncaglioni A. Improving interpretability in ADMET modeling. *Mol Inform.* 2021;40(8):e2000155.
50. Gao H, et al. Benchmarking ML algorithms for ADMET prediction. *J Chem Inf Model.* 2019;59(7):3056–3065.

51. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. Deep learning in toxicology: the DeepTox approach. *Front Environ Sci.* 2016;3:80.
52. Gilmer J, Schoenholz SS, Riley PF, et al. Neural message passing for quantum chemistry. *Proc Int Conf Mach Learn (ICML).* 2017;1263–1272.
53. Riniker S, Landrum GA. Benchmarking molecular machine learning models. *J Chem Inf Model.* 2021;61(7):3141–3152.
54. Samek W, Müller K-R. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning.* Springer; 2019.
55. Pires DEV, Blundell TL, Ascher DB. pkCSM: graph-based predictive modeling. *J Med Chem.* 2015;58(9):4066–4072.
56. Yang H, Lou C, Sun L, Li J, et al. admetSAR 2.0: web-service for prediction and optimization of ADMET properties. *Nucleic Acids Res.* 2019;47(W1):W537–W544.
57. Xiong G, Wu Z, Yi J, Fu L, et al. ADMETlab 2.0: integrated online platform for accurate and comprehensive ADMET predictions. *Nucleic Acids Res.* 2021;49(W1):W5–W14.
58. Schrödinger LLC. *QikProp User Manual, Version 2023-3.* New York: Schrödinger; 2023.
59. Richard AM, Judson RS, Houck KA, Grulke CM, et al. ToxCast chemical landscape revisited. *Chem Res Toxicol.* 2021;34(4):861–879.
60. Huang K, Xiao C, Glass LM, Zitnik M. Therapeutics Data Commons: machine learning datasets and tasks for drug discovery. *Nat Comput Sci.* 2022;2:687–694.
61. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov.* 2019;18(1):41–58.
62. Rodríguez-Pérez R, Bajorath J. Interpretable ML for drug discovery. *J Med Chem.* 2021;64(20):15026–15038.
63. OECD. *Guidance Document on the Validation of (Q)SAR Models.* ENV/JM/MONO(2007)2. Paris: OECD; 2007.
64. Cronin MTD, Madden JC. *In Silico Toxicology: Principles and Applications.* Cambridge: Royal Society of Chemistry; 2020.
65. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable AI. *Nat Mach Intell.* 2022;4(12):1075–1085.
66. Kaushik S, Fuchs M, Dünser MW, et al. Digital twins in pharmacology for personalized therapeutic simulation. *Front Pharmacol.* 2024;15:1389295.
67. Wang J, Peng Y, Dong C, et al. Multi-omics data integration in toxicogenomics: recent advances. *Front Genet.* 2023;14:1173389.
68. Liu X, Chen Y, Xu H, et al. Federated learning in pharmaceutical research and healthcare. *Drug Discov Today.* 2023;28(3):103627.
69. Batra R, Tran H, Kim C, et al. Quantum machine learning models for drug discovery. *npj Comput Mater.* 2023;9:114.