

AI-Assisted QSAR and Molecular Modeling in Modern Drug Discovery

Hindavi Adhav¹, Shivani Sonar², Rupali Gite³, Dipali Bhaskar⁴, Pooja Dengale⁵

¹Asst. Professor, Department of Pharmaceutical Quality Assurance, Ashvin college of Pharmacy, Manchi Hill

^{2,3,4,5}Research Scholar, Department of Pharmacy, Ashvin college of Pharmacy, Manchi Hill

GRAPHICAL ABSTRACT

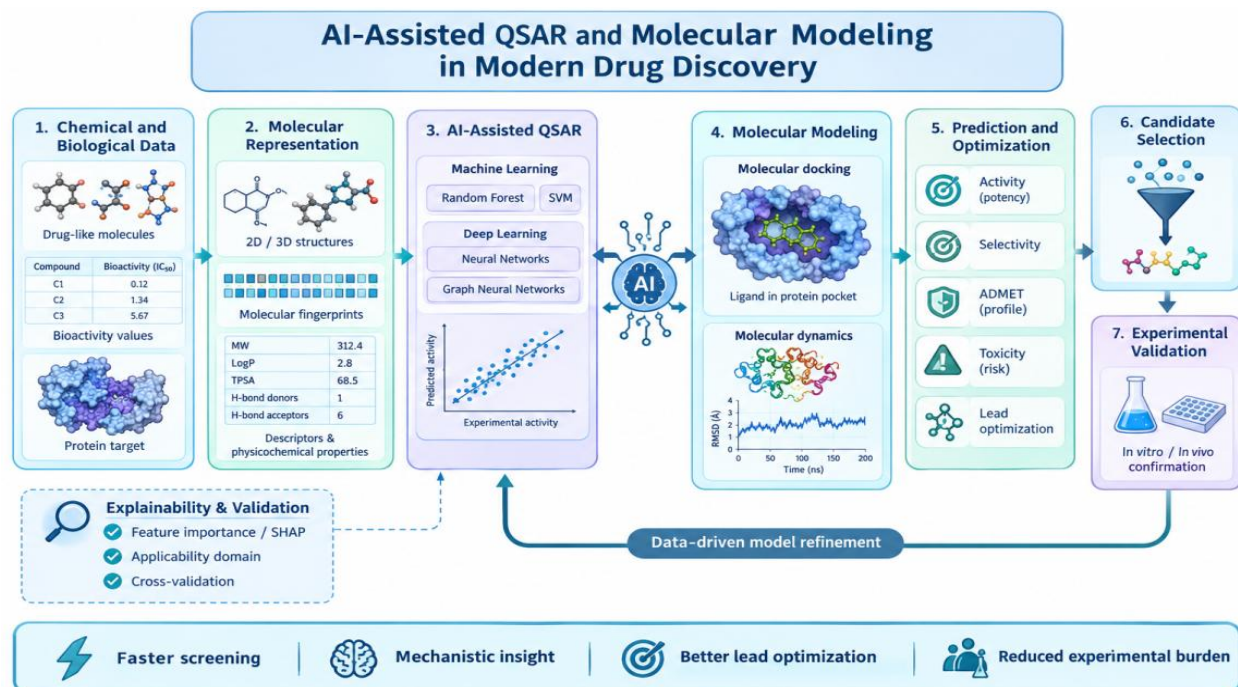


Figure 1. AI-assisted QSAR and molecular modeling workflow for modern drug discovery.[1]

Abstract—Artificial intelligence (AI) is transforming modern drug discovery by improving the speed, accuracy, and efficiency of computational approaches used to identify and optimize promising drug candidates. Quantitative structure–activity relationship (QSAR) modeling, when integrated with machine learning and deep learning, enables prediction of biological activity from molecular descriptors, fingerprints, and structural features. Molecular modeling techniques, including molecular docking, molecular dynamics, and virtual screening, provide complementary information about ligand–target interactions, binding behavior, and molecular stability. AI-assisted integration of these approaches can support rapid compound prioritization, lead optimization, selectivity assessment, and early prediction of pharmacokinetic and toxicity-related

properties. Advanced algorithms identify complex relationships within large and heterogeneous chemical datasets that may be difficult to recognize using conventional methods. However, challenges involving data quality, model interpretability, applicability domains, validation, and reproducibility remain important. This review discusses the principles, applications, advantages, limitations, and future potential of AI-assisted QSAR and molecular modeling in contemporary drug discovery.

Index Terms—Artificial Intelligence; QSAR; Molecular Modeling; Machine Learning; Drug Discovery; Molecular Docking; Virtual Screening; Lead Optimization.

I. INTRODUCTION

Drug discovery is intrinsically a multidimensional optimization problem in which potency, selectivity, physicochemical properties, pharmacokinetics, safety and synthetic accessibility must be considered simultaneously. Classical QSAR emerged from the idea that biological activity can be related quantitatively to chemical structure, creating a computational route for learning structure–activity trends before every candidate is synthesized. The historical development of substituent and physicochemical correlations established the conceptual basis for modern structure–activity modeling, while later cheminformatics representations made chemical structures machine-readable at scale. [1,2]

The growth of chemical databases and computational power has transformed QSAR from relatively small descriptor-based studies into a broader molecular machine-learning discipline. Modern models can learn from two-dimensional graphs, circular fingerprints, three-dimensional conformations, protein–ligand complexes and heterogeneous biological information. Benchmarking work such as MoleculeNet demonstrated both the promise of learnable molecular representations and the difficulty created by small, imbalanced and heterogeneous chemical datasets. More recent analyses have emphasized that deep QSAR is not simply a deeper version of classical QSAR; it incorporates representation learning, large-scale screening, generative design and structure-based prediction into a wider computational discovery framework.[9]

The importance of AI-assisted QSAR is therefore not limited to obtaining a higher coefficient of determination or a lower prediction error. A useful model must answer a medicinal-chemistry question: which compound should be made, tested or deprioritized next, and how confident should the team be in that recommendation? This requires a connection between molecular representations, predictive algorithms, structural models and experimental decision-making. Modern AI drug-discovery reviews similarly emphasize the convergence of virtual screening, molecular generation, target analysis and ADMET prediction, but also identify data quality, validation, interpretability and generalization as persistent barriers.[13]

II. SCOPE AND REVIEW APPROACH

This article presents a narrative, method-focused review of AI-assisted QSAR and molecular modeling in contemporary small-molecule drug discovery. Literature was considered across foundational QSAR methodology, molecular representations, machine learning, graph neural networks, virtual screening, docking, generative design, protein structure prediction, uncertainty quantification and explainable AI. Particular emphasis was placed on primary methodological papers, established benchmark resources, regulatory or validation guidance and recent reviews that discuss practical limitations. The objective is to synthesize these areas into a coherent workflow rather than to compare individual commercial platforms or claim that one model family is universally superior.[5]

III. EVOLUTION OF QSAR TOWARD AI-ASSISTED MODELING

Traditional QSAR expresses the relationship between a molecular structure and a measured endpoint using descriptors such as molecular weight, hydrophobicity, hydrogen-bonding characteristics, topology, electronic parameters and fragment counts. Linear regression can provide an interpretable relationship, whereas nonlinear methods can capture interactions among structural features that are difficult to express analytically. The fundamental advantage is that once a model has been validated, it can estimate properties for untested molecules at a fraction of the experimental cost. However, prediction quality is strongly dependent on the quality and chemical diversity of the training set.[6]

Molecular fingerprints changed the scale at which structure–activity relationships could be represented. Extended-connectivity fingerprints encode local molecular environments in a compact form and are particularly useful for similarity searching and structure–activity modeling. Such fingerprints allow a model to recognize recurring substructures without requiring every feature to be specified manually. Nevertheless, fixed fingerprints can discard some information about three-dimensional arrangement, long-range interactions and context-dependent biological mechanisms, motivating the development of learned representations. [3,9]

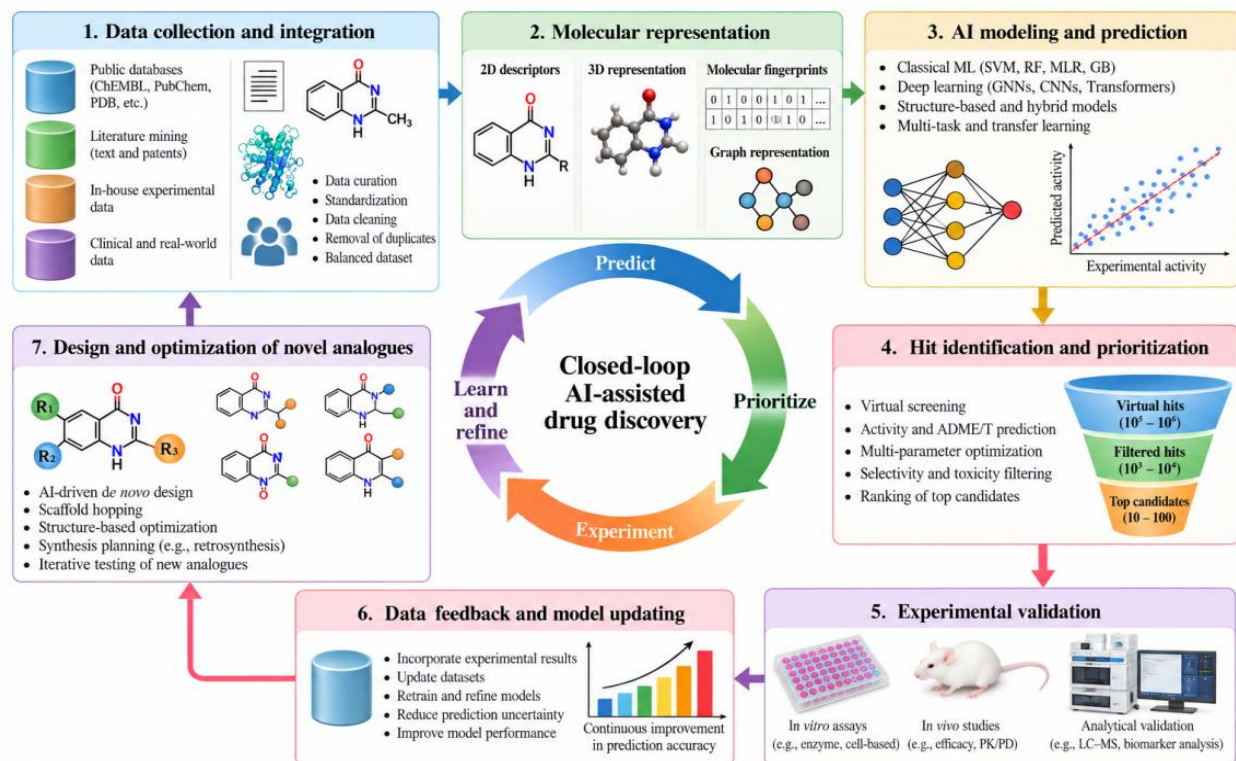


Figure 2. Closed-loop AI-assisted drug-discovery cycle. Predictions are used to prioritize experiments, while experimental observations are returned to the data and modeling layers for iterative refinement.[7]

The transition to AI-assisted QSAR has been accelerated by graph-based learning. Molecular graphs naturally describe atoms as nodes and bonds as edges, allowing convolutional and message-passing architectures to learn representations directly from chemical connectivity. Graph convolutional models and message-passing neural networks have demonstrated that molecular representations can be optimized jointly with the prediction task rather than frozen before model training. Comparative work has nevertheless shown that no representation is universally optimal, particularly when data are limited or when test compounds lie outside the chemical neighborhoods represented during training.[10]

IV. MOLECULAR REPRESENTATION: THE FOUNDATION OF AI-QSAR

A molecular AI model does not operate on a chemical structure in the abstract; it operates on a representation of that structure. SMILES provides a compact textual encoding that is widely used for storage and machine learning, while descriptor vectors translate molecular properties into numerical features. Fingerprints

represent structural motifs, molecular graphs encode connectivity explicitly, and three-dimensional representations capture conformational and spatial information. The representation determines which chemical relationships are available for learning and therefore influences both model performance and interpretability. [7,4]

Descriptor-based models remain valuable because descriptors have direct chemical meaning. Molecular weight, calculated logP, topological polar surface area, rotatable-bond count, hydrogen-bond donors and acceptors, aromatic proportion and charge-related descriptors can explain important aspects of solubility, permeability and developability. The ESOL model illustrates how a relatively compact set of structure-derived variables can provide useful aqueous-solubility estimates. Such descriptors are also useful as complementary inputs to deep models because they can provide global physicochemical information that is not always captured efficiently by local graph operations.[4]

Fingerprints are particularly effective for similarity-driven workflows and analog prioritization. ECFP-type representations encode circular atom

neighborhoods and can capture medicinal-chemistry motifs associated with activity. However, fingerprint similarity is not equivalent to biological equivalence. Small structural changes can generate activity cliffs in which apparently similar compounds have very different potencies because of stereochemistry, conformational changes, binding-mode switches, induced fit or assay-specific mechanisms. QSAR systems therefore benefit from combining similarity information with explicit predictive uncertainty and applicability-domain analysis.[26]

Graph neural networks provide a more adaptive representation. Graph convolutional and message-

passing models repeatedly aggregate information from neighboring atoms and bonds, allowing learned vectors to reflect local and progressively larger molecular environments. Directed message passing further refines this idea by controlling how information moves along bonds and can perform strongly across diverse molecular-property tasks. The resulting representation can support multitask prediction, transfer learning and end-to-end optimization, although model complexity can make mechanistic interpretation more difficult.[14]

Table 1. Major molecular representations used in AI-assisted QSAR. [8,14]

Representation	Information encoded	Typical applications	Key limitation
Physicochemical descriptors	Global chemical properties and topology	Solubility, permeability, developability, classical QSAR	May omit detailed local or 3D context
ECFP/Morgan fingerprints	Circular local substructures	Similarity, activity classification, virtual screening	Sensitive to fingerprint settings; limited explicit 3D information
Molecular graphs	Atoms, bonds and local connectivity	GCN, GAT, MPNN property prediction	Can require more data and careful architecture design
SMILES/token sequences	Linearized molecular syntax	Transformers, generative modeling, reaction prediction	Representation depends on serialization and tokenization
3D conformations/complexes	Spatial arrangement and protein-ligand geometry	Docking, affinity modeling, 3D deep learning	Conformational and structural uncertainty

V. AI/ML ALGORITHMS USED IN QSAR

Classical machine-learning methods remain competitive in many QSAR applications, particularly when datasets are moderate in size and descriptors or fingerprints provide informative features. Support-vector machines can model nonlinear decision boundaries through kernel functions, while tree ensembles can accommodate nonlinear interactions and mixed feature scales. Gradient boosting methods are also useful for tabular chemical descriptors. These approaches often train quickly, are relatively robust with limited data and can serve as essential baselines against which deep models should be evaluated.[12] Deep neural networks become more attractive as the number and diversity of molecular examples increase. Feed-forward networks can learn nonlinear

relationships from descriptors, convolutional models can process structured representations, and graph neural networks can learn directly from molecular connectivity. Message-passing architectures extend graph learning by iteratively propagating information across molecular neighborhoods before generating a molecule-level representation. These models reduce the need for manual descriptor selection but introduce hyperparameter sensitivity, higher computational requirements and greater dependence on dataset size and curation quality.[11]

Transformers and other attention-based architectures extend AI-QSAR toward sequence and multimodal learning. Molecular strings can be treated as chemical language, allowing models to learn contextual relationships between tokens, while attention mechanisms can integrate information from different

molecular or biological sources. In reaction prediction and synthesis planning, transformer architectures have demonstrated that chemical transformations can be represented as sequence-to-sequence problems. For

drug discovery, the broader significance is that predictive modeling can be connected with the design and synthesis stages rather than remaining an isolated property-prediction step.[22]

Table 2. AI/ML model families and their roles in drug discovery.[15]

Model family	Representative role	Strength	Main caution
Linear/QSAR regression	Interpretable property/activity relationships	Simple, transparent and useful for smaller datasets	Limited nonlinear capacity
SVM	Classification and nonlinear regression	Effective in moderate-dimensional chemical spaces	Kernel and hyperparameter dependence
Random forest / boosting	Descriptor and fingerprint prediction	Strong tabular baselines; nonlinear interactions	Can be biased by data leakage or correlated features
CNN / graph convolution	Learned molecular representation	Learns task-specific structural features	Data and architecture sensitivity
MPNN / D-MPNN	Molecular property and activity prediction	Direct learning from molecular graphs	Interpretability and extrapolation remain challenging
Transformer / generative models	Molecular generation and reaction modeling	Flexible sequence/context learning	Data-hungry and can generate chemically or biologically unsuitable candidates
Hybrid AI + docking	Virtual screening and rescoring	Combines ligand and structure information	Docking and scoring-function biases remain

VI. DATA SOURCES AND DATA CURATION

AI-QSAR performance is fundamentally constrained by the quality of experimental data. Public resources such as ChEMBL provide curated bioactivity measurements together with chemical structures and target annotations, while benchmark collections such as MoleculeNet organize datasets across physicochemical, quantum, biophysical and physiological endpoints. Protein–ligand binding-affinity resources such as PDBbind connect experimental affinities with three-dimensional complexes and therefore support structure-based machine learning. These databases are valuable, but they contain measurements generated using different assays, protocols, laboratories, target constructs and reporting conventions. [12,18]

Curation should begin with chemical-structure standardization, duplicate handling, salt and solvent removal where scientifically justified, stereochemical normalization, unit harmonization and endpoint definition. Biological labels require equal attention because IC₅₀, EC₅₀, K_i, K_d and percentage inhibition are not interchangeable quantities. Measurements from different assay formats may represent different biological mechanisms even when the target name is

identical. A model trained on uncurated heterogeneous labels can therefore learn experimental artifacts rather than generalizable chemical relationships.[34]

Dataset splitting is another critical issue. Random splitting can place highly similar analogues in both training and test sets, producing optimistic estimates of performance. Scaffold splitting, cluster-based splitting and temporal splitting provide more demanding tests of chemical generalization. MoleculeNet highlighted the importance of appropriate split strategies, while later benchmarking studies showed that model rankings can change substantially when scaffold or time-based evaluation replaces random partitioning. Consequently, a credible AI-QSAR study should report the splitting strategy explicitly and should use an external test set that represents the intended deployment scenario.[15]

VII. QSAR MODEL DEVELOPMENT AND VALIDATION

A reliable QSAR workflow requires more than a high training-set fit. Internal validation may include cross-validation and bootstrap procedures, whereas external validation evaluates predictions for compounds that were not used during model development. Regression

models are commonly evaluated with RMSE, MAE and R^2 , while classification models may use ROC-AUC, PR-AUC, balanced accuracy, sensitivity, specificity and calibration measures. Because a single metric can conceal clinically important errors, model assessment should be aligned with the decision for which the model will be used.[23]

The applicability domain is equally important. A prediction for a molecule that lies near the chemical space represented in the training set is fundamentally different from a prediction for an unprecedented scaffold. Applicability-domain methods can use similarity, descriptor-space distance, leverage, density or model-based uncertainty to identify regions in which predictions are less reliable. The OECD framework emphasizes defined applicability and transparent assessment of model predictions, reinforcing the principle that a QSAR result should be accompanied by evidence about whether the model is qualified to make that prediction.[25]

Y-randomization and related negative-control tests can help determine whether apparent performance depends on meaningful structure–activity relationships rather than accidental correlations. External validation should be accompanied by inspection of prediction errors, chemical-space coverage and possible outliers. Importantly, data leakage can occur even without direct duplication when measurements for close analogues, temporal records or preprocessing steps are allowed to cross dataset boundaries. Rigorous validation is therefore a scientific control rather than a formatting requirement.[24]

VIII. AI-ASSISTED MOLECULAR PROPERTY PREDICTION

Property prediction is one of the most mature applications of AI in drug discovery. Models can estimate solubility, lipophilicity, permeability, pKa, melting behavior, plasma protein binding and other physicochemical characteristics, while biological models can estimate potency, selectivity, toxicity or target engagement. Multi-task learning can exploit correlations among endpoints, potentially improving performance when individual tasks have limited data. However, correlated endpoints can also introduce

misleading shortcuts if measurements share experimental artifacts rather than biological causality.[31]

ADMET prediction is especially important because many drug candidates fail for reasons that are not visible in a single potency measurement. AI models can prioritize molecules with more favorable absorption, distribution, metabolism, excretion and toxicity profiles before synthesis or advanced biological testing. The most useful role of these models is usually triage rather than definitive exclusion: a predicted liability should trigger experimental clarification, structural redesign or additional modeling. Combining several ADMET endpoints with uncertainty estimates can help medicinal chemists balance potency against developability instead of optimizing one property in isolation. [29,33]

IX. MOLECULAR DOCKING AND STRUCTURE-BASED MODELING

QSAR is ligand-centered, whereas molecular docking provides a structure-centered perspective by estimating how a candidate ligand may fit within a protein binding site. Classical docking searches over ligand orientations and conformations and applies scoring functions that approximate favorable and unfavorable interactions. AutoDock Vina demonstrated that efficient optimization and an improved scoring formulation could substantially increase docking speed while maintaining useful pose-prediction performance, supporting large virtual screens.[16]

AI can improve structure-based modeling at several points. Machine-learning scoring functions can rescore docking poses using interaction patterns learned from experimentally characterized complexes, while neural networks can evaluate three-dimensional grids or molecular graphs. PDBbind and DUD-E have supported benchmarking of affinity prediction and virtual screening, although benchmark performance can overestimate real-world utility when ligand similarity, target leakage or decoy construction creates favorable conditions. Consensus approaches and experimental validation therefore remain important for high-confidence hit selection.[38]

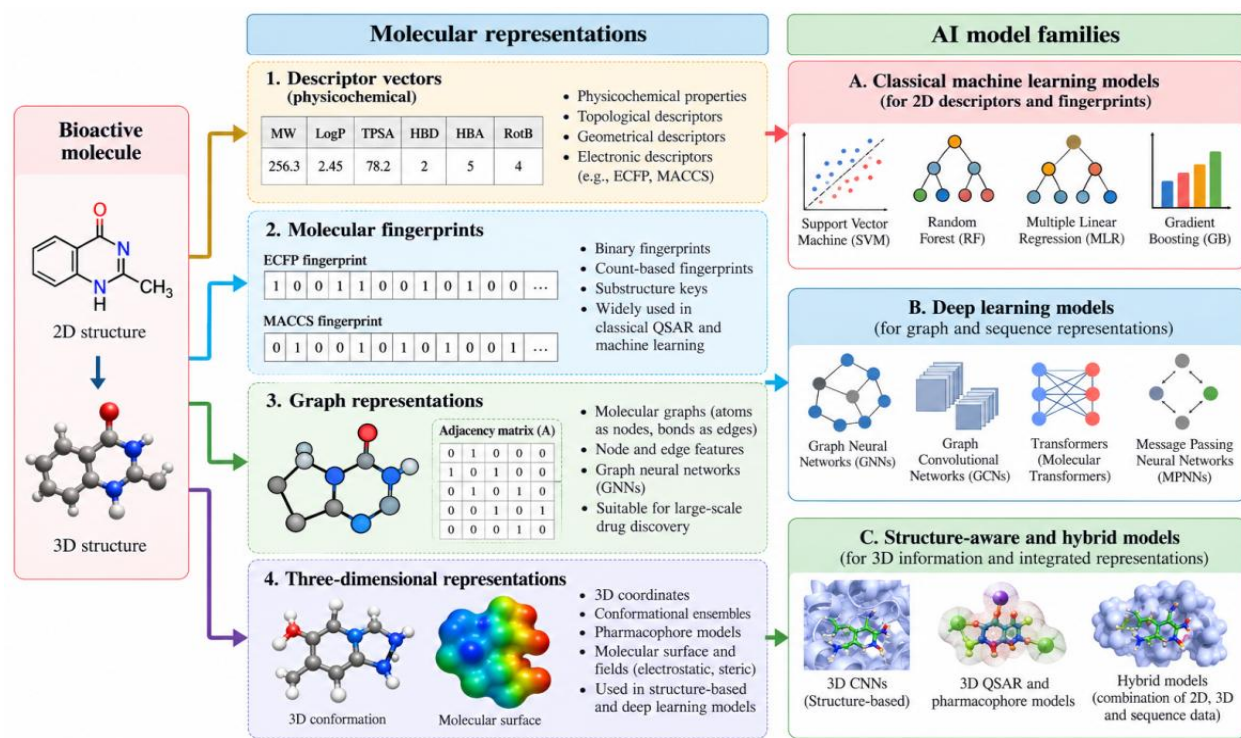


Figure 3. Molecular representations and model families used in AI-assisted drug discovery. Descriptor vectors and fingerprints remain useful for classical QSAR, while graph and three-dimensional representations support learned molecular and structure-aware models.[18]

Protein structure prediction has further expanded the accessible structural space. AlphaFold2 demonstrated highly accurate protein-structure prediction across many targets, and large-scale structure resources subsequently increased structural coverage for proteins that lacked experimentally determined models. AlphaFold3 extended this direction toward complexes containing proteins, nucleic acids, small molecules, ions and modified residues. These advances can support target assessment and structure-based modeling, but predicted structures should not be treated as experimentally equivalent in every binding-site context, particularly for flexible, disordered or induced-fit systems.[21]

X. AI-ENHANCED VIRTUAL SCREENING

Virtual screening combines chemical libraries with computational filters to reduce the number of compounds requiring experimental testing. A practical AI-assisted screen may begin with physicochemical and structural filters, followed by QSAR or graph-based activity prediction, applicability-domain analysis, docking, rescoring and ADMET

prioritization. This layered strategy is often more efficient than applying a single expensive method to millions of molecules because inexpensive models can remove candidates before structure-based calculations are performed.[28]

The most valuable screening pipelines are therefore hierarchical rather than algorithmically monolithic. Ligand-based QSAR can identify molecules that resemble known active chemical space, whereas docking can test whether candidates are structurally compatible with a target pocket. Generative models can propose novel structures, but generated molecules should be evaluated for predicted activity, physicochemical properties, toxicity, synthetic accessibility and structural plausibility before experimental selection. The combination of complementary models reduces dependence on any one imperfect scoring function.[17]

XI. GENERATIVE AI AND MOLECULAR DESIGN

Generative models change the computational question from “which existing compound is promising?” to

“which new molecular structures should be considered?” Variational autoencoders, graph-generative models, reinforcement learning and transformer-based systems can generate candidates while optimizing one or more molecular objectives. Junction-tree approaches demonstrated that molecular graphs can be generated through chemically meaningful substructures, while modern generative AI reviews describe transformers, diffusion models and reinforcement learning as major paradigms for de novo design.[8]

Generative design is most useful when coupled to validated property predictors. A generated molecule may appear novel and chemically valid but still have poor potency, unacceptable toxicity, difficult synthesis or undesirable physicochemical characteristics. Multi-objective optimization therefore becomes essential. A practical design system can use QSAR predictions as reward functions, structural constraints as feasibility filters and synthetic planning models as downstream checks. This creates a design loop in which generation and prediction continually constrain each other.[28]

Experimental evidence illustrates the potential and limitations of this strategy. Deep-learning approaches have been used to prioritize antibacterial compounds from large chemical spaces, demonstrating that AI can identify biologically active molecules outside conventional screening workflows. At the same time, subsequent analyses have shown that reported performance and apparent scaffold novelty can depend strongly on dataset composition, preprocessing and molecular representation. Such findings reinforce the need to treat AI-generated hits as hypotheses requiring experimental confirmation rather than as validated drug candidates.[35]

XII. EXPLAINABILITY AND MECHANISTIC INTERPRETATION

Interpretability is especially important in drug discovery because medicinal chemists need to understand why a model recommends a compound and how a structural modification might change the prediction. Descriptor-based QSAR can offer direct coefficient-level interpretations, whereas tree ensembles can provide feature-importance estimates. Deep models require additional tools, including atom-level attribution, substructure explanations, attention analysis and counterfactual design. Explainable AI is

increasingly viewed as a bridge between model predictions and human chemical reasoning.[33]

Explainability should nevertheless be interpreted cautiously. An attribution map identifies features associated with a model output; it does not automatically establish a biological mechanism. A model may assign importance to a substructure because it correlates with an unmeasured property, assay artifact or chemical-series membership. Mechanistic credibility improves when explanations are chemically plausible, stable across perturbations, consistent with known biology and supported by experimental analogs. Explainability is therefore best used as an evidence layer rather than as proof of causality.[32]

XIII. UNCERTAINTY, APPLICABILITY AND TRUSTWORTHY AI-QSAR

Prediction uncertainty is central to trustworthy computational drug discovery. A model can be highly accurate on average while producing occasional high-confidence failures on novel scaffolds. Uncertainty can arise from limited training data, noisy measurements, model uncertainty, distribution shift and extrapolation beyond the applicability domain. Ensemble methods, Bayesian approaches, conformal methods and distance-based approaches can provide complementary estimates of prediction reliability.[27]

Uncertainty should influence experimental prioritization. A high predicted potency with low uncertainty may be suitable for immediate synthesis, whereas a high predicted potency with high uncertainty may be better treated as an information-gain experiment.

Table 3. Validation and trustworthiness controls for AI-QSAR.[30]

Control	Purpose	Recommended interpretation
External test set	Measures prospective-like generalization	Primary evidence beyond training performance
Scaffold split	Tests transfer to new chemotypes	More demanding than random splitting for many medicinal-chemistry tasks
Temporal split	Mimics future discovery campaigns	Useful when assay dates are available

Applicability domain	Identifies predictions outside learned chemical space	Predictions outside the domain require additional caution
Uncertainty estimation	Quantifies confidence or prediction variability	High uncertainty can prioritize confirmatory experiments
Y-randomization	Tests whether performance depends on real labels	Strong randomized performance suggests leakage or methodological problems
Calibration / class imbalance checks	Assesses probability quality and rare-event behavior	Especially important for safety and toxicity endpoints
Prospective experimental validation	Tests real-world utility	Most important evidence for a discovery workflow

Similarly, a predicted safety liability outside the model's domain should not automatically eliminate a candidate. Instead, uncertainty-aware decision rules

can allocate experimental resources toward compounds that are both promising and informative, moving the workflow from passive prediction toward active learning.[27]

XIV. INTEGRATION OF QSAR, DOCKING, ADMET AND EXPERIMENTAL FEEDBACK

The strongest conceptual model for modern AI-assisted discovery is a closed loop. Chemical data are curated, molecules are represented using complementary encodings, QSAR models prioritize candidates, structural modeling examines plausible binding modes, ADMET models estimate developability, and experimental results are returned to the database. New measurements then improve subsequent model training. This cycle resembles a design–make–test–analyze process in which computation does not replace experimentation but determines which experiments are most informative.[19]

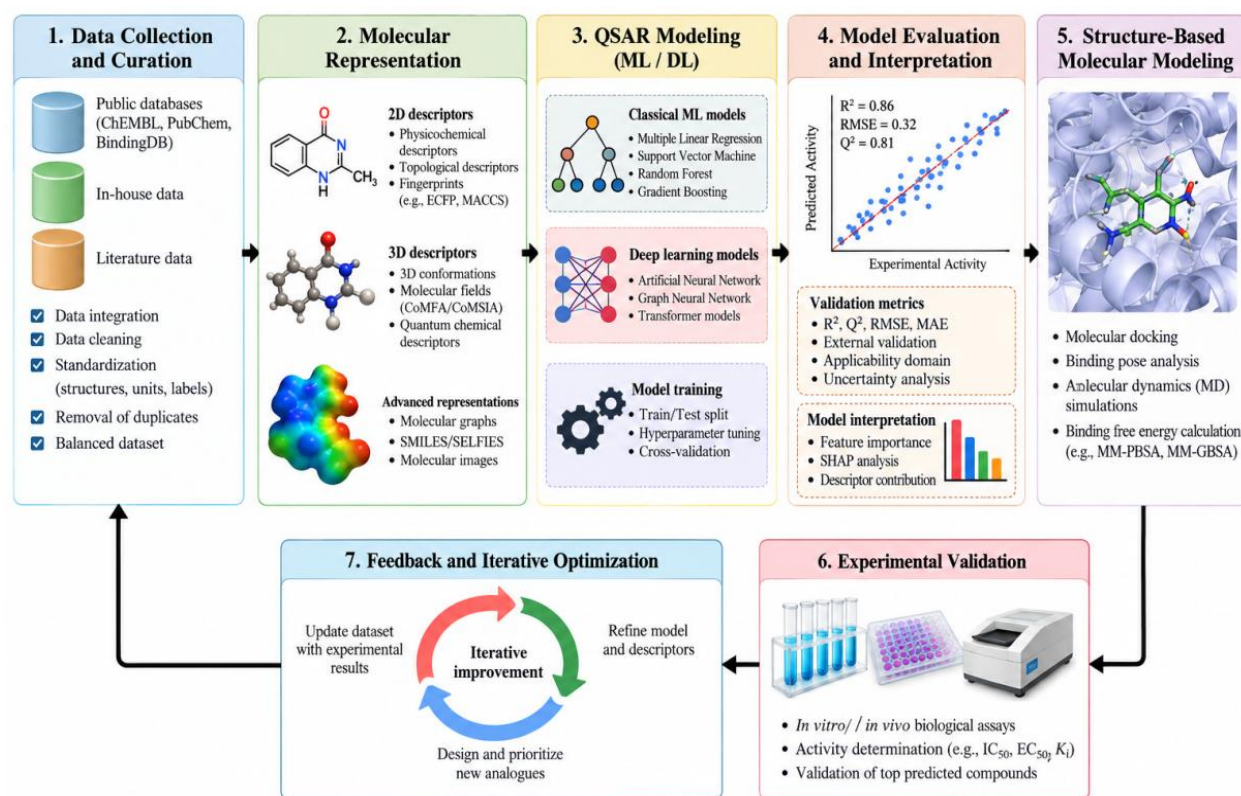


Figure 4. AI-assisted QSAR and molecular-modeling workflow. Chemical and biological data are curated, converted into complementary molecular representations, modeled with classical or deep-learning algorithms, evaluated with external validation and uncertainty analysis, and integrated with structure-based modeling and experimental feedback.[19]

Integration also enables staged decision-making. Early filters can remove obvious physicochemical liabilities; ligand-based QSAR can reduce chemical-library size; docking can focus on structural compatibility; machine-learning scoring can refine rankings; and experimental assays can resolve uncertainties that computation cannot reliably settle. Such an architecture can reduce computational cost and improve interpretability because each stage answers a different question. The resulting workflow is more defensible than treating a single AI score as a surrogate for the complete drug-discovery process.[20]

XV. MAJOR CHALLENGES AND LIMITATIONS

The first major challenge is data quality. Experimental activity values can vary with assay format, target construct, incubation time, cell system and reporting unit. Chemical structures may contain salts, mixtures, incorrect stereochemistry or inconsistent tautomeric states. These inconsistencies can introduce noise that no algorithm can fully remove. Data curation should therefore be treated as part of model development rather than as a preliminary clerical step.[18]

The second challenge is chemical-space extrapolation. Deep models can interpolate extremely well within a familiar chemical neighborhood while failing on novel scaffolds. Activity cliffs further complicate prediction because structurally small changes may cause large biological effects. Scaffold-based and temporal evaluation, applicability-domain analysis and external experimental validation are therefore necessary to determine whether an apparently strong model is genuinely useful for prospective discovery.[25]

The third challenge is structural uncertainty. Docking scores are approximations, protein conformations are dynamic, binding pockets can rearrange and solvent effects are difficult to model accurately at screening scale. AI scoring functions may inherit biases from their training complexes. Predicted protein structures add valuable coverage but do not remove uncertainty in ligand-induced conformational changes. Consequently, docking should be interpreted as a structural hypothesis generator and ranking tool rather than as a direct measurement of experimental binding free energy.[36]

The fourth challenge is reproducibility and interpretability. Different preprocessing choices, random seeds, molecular standardization rules, split

strategies and hyperparameters can change model performance. Proprietary datasets may prevent independent verification. Open benchmarks and transparent reporting help, but benchmark scores should not be treated as universal indicators of prospective performance. A credible study should report data provenance, structure standardization, split design, model settings, evaluation metrics, applicability domain, uncertainty and prospective validation wherever possible.[37]

XVI. FUTURE PERSPECTIVES

Future AI-QSAR systems are likely to become increasingly multimodal. Instead of predicting activity from a ligand alone, models may jointly represent molecular structure, protein sequence and structure, assay context, cell phenotype, omics measurements and historical experimental outcomes. Such systems could better distinguish a genuine mechanistic signal from a correlation that only exists within one dataset. The emergence of foundation-style chemical models and multimodal generative architectures suggests that the boundary between QSAR, molecular modeling and synthesis planning will continue to narrow.[30]

A second direction is physics-informed and mechanism-aware learning. Purely data-driven models can fit patterns without explicitly respecting chemical constraints, whereas hybrid approaches can combine learned representations with physical descriptors, docking information, molecular mechanics, quantum calculations or experimentally derived interaction features. Such hybridization may improve generalization when labeled data are scarce and can make model outputs more scientifically interpretable. The goal is not to choose between physics and AI but to encode useful physical knowledge where it improves predictive reliability. [21]

A third direction is autonomous or semi-autonomous experimentation. In a mature closed-loop system, uncertainty-aware models would select informative compounds, generative systems would propose structures, synthesis planners would evaluate routes, automated platforms would perform experiments and analytical results would update the models. However, autonomy must remain bounded by chemical safety, experimental controls, data provenance and human scientific oversight. The most realistic near-term

model is therefore human-centered automation in which AI accelerates repetitive prediction and prioritization while researchers retain responsibility for experimental interpretation and critical decisions.[26]

XVII. CONCLUSION

AI-assisted QSAR and molecular modeling have evolved from isolated computational techniques into interconnected components of modern drug discovery. Classical descriptors and fingerprints remain useful, but graph neural networks, message-passing architectures, generative models, machine-learning docking scores and AI-based protein structure prediction now permit richer representations of molecular and biological information. The greatest value arises when these methods are combined rather than evaluated as competing technologies.[39]

The reliability of an AI-assisted drug-discovery workflow ultimately depends on scientific discipline: curated data, chemically appropriate representations, leakage-resistant validation, defined applicability domains, uncertainty estimation, interpretable evidence and experimental testing. Models should guide decisions, not replace evidence. When QSAR, molecular modeling and experimental feedback are integrated into a closed loop, AI can make chemical-space exploration more systematic, reduce unnecessary synthesis and testing, and increase the probability that computationally prioritized molecules become experimentally meaningful starting points for drug discovery.[34]

REFERENCES

- [1] C. Hansch and T. Fujita, "p- σ - π analysis: A method for the correlation of biological activity and chemical structure," *J. Am. Chem. Soc.*, vol. 86, no. 8, pp. 1616–1626, 1964, doi: 10.1021/ja01062a035.
- [2] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," *J. Chem. Inf. Comput. Sci.*, vol. 28, no. 1, pp. 31–36, 1988, doi: 10.1021/ci00057a005.
- [3] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.*, vol. 46, nos. 1–3, pp. 3–26, 2001, doi: 10.1016/S0169-409X(00)00129-0.
- [4] J. S. Delaney, "ESOL: A simple model for aqueous solubility based on molecular properties," *J. Chem. Inf. Comput. Sci.*, vol. 44, no. 3, pp. 1000–1005, 2004, doi: 10.1021/ci034243x.
- [5] A. Tropsha, "The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models," *QSAR Comb. Sci.*, vol. 22, no. 1, pp. 69–77, 2003, doi: 10.1002/qsar.200390007.
- [6] A. Tropsha, "Best practices for QSAR model development, validation, and exploitation," *Mol. Inform.*, vol. 29, nos. 6–7, pp. 476–488, 2010, doi: 10.1002/minf.201000061.
- [7] D. Rogers and M. Hahn, "Extended-connectivity fingerprints," *J. Chem. Inf. Model.*, vol. 50, no. 5, pp. 742–754, 2010, doi: 10.1021/ci100050t.
- [8] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995, doi: 10.1007/BF00994018.
- [9] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, et al., "Convolutional networks on graphs for learning molecular fingerprints," in *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.
- [10] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: Moving beyond fingerprints," *J. Comput.-Aided Mol. Des.*, vol. 30, no. 8, pp. 595–608, 2016, doi: 10.1007/s10822-016-9938-8.
- [11] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, PMLR, vol. 70, 2017, pp. 1263–1272.
- [12] Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, et al., "MoleculeNet: A benchmark for molecular machine learning," *Chem. Sci.*, vol. 9, pp. 513–530, 2018, doi: 10.1039/C7SC02664A.
- [13] W. Jin, R. Barzilay, and T. Jaakkola, "Junction tree variational autoencoder for molecular graph generation," in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, PMLR, vol. 80, 2018, pp. 2323–2332.

- [14] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, et al., "Analyzing learned molecular representations for property prediction," *J. Chem. Inf. Model.*, vol. 59, no. 8, pp. 3370–3388, 2019, doi: 10.1021/acs.jcim.9b00237.
- [15] M. M. Mysinger, M. Carchia, J. J. Irwin, and B. K. Shoichet, "Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking," *J. Med. Chem.*, vol. 55, no. 14, pp. 6582–6594, 2012, doi: 10.1021/jm300687e.
- [16] O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading," *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010, doi: 10.1002/jcc.21334.
- [17] J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, et al., "A deep learning approach to antibiotic discovery," *Cell*, vol. 180, no. 4, pp. 688–702.e13, 2020, doi: 10.1016/j.cell.2020.01.021.
- [18] D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, et al., "ChEMBL: Towards direct deposition of bioassay data," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D930–D940, 2019, doi: 10.1093/nar/gky1075.
- [19] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, et al., "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.
- [20] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Židek, et al., "Highly accurate protein structure prediction for the human proteome," *Nature*, vol. 596, pp. 590–596, 2021, doi: 10.1038/s41586-021-03828-1.
- [21] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, et al., "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, pp. 493–500, 2024, doi: 10.1038/s41586-024-07487-w.
- [22] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, et al., "Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction," *ACS Cent. Sci.*, vol. 5, no. 9, pp. 1572–1583, 2019, doi: 10.1021/acscentsci.9b00576.
- [23] Organisation for Economic Co-operation and Development, *Guidance Document on the Validation of (Quantitative) Structure–Activity Relationship [(Q)SAR] Models*, OECD Series on Testing and Assessment, No. 69. Paris, France: OECD Publishing, 2014, doi: 10.1787/9789264085442-en.
- [24] Organisation for Economic Co-operation and Development, *(Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure Activity Relationship Models and Predictions*, OECD Series on Testing and Assessment, No. 386. Paris, France: OECD Publishing, 2023, doi: 10.1787/d96118f6-en.
- [25] S. Kar, K. Roy, and J. Leszczynski, "Applicability domain: A step toward confident predictions and decidability for QSAR modeling," in *Advances in QSAR Modeling*, K. Roy, Ed. Cham, Switzerland: Springer, 2018, pp. 141–169, doi: 10.1007/978-1-4939-7899-1_6.
- [26] E. B. Lenselink, N. ten Dijke, B. Bongers, G. Papadatos, H. van Vlijmen, J. Kowalski, et al., "Beyond the hype: Deep neural networks outperform established methods for QSAR and ADME prediction," *J. Cheminform.*, vol. 9, Art. no. 45, 2017, doi: 10.1186/s13321-017-0232-0.
- [27] G. Scalia, C. A. Grambow, B. Pernici, Y. P. Li, and W. H. Green, "Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction," *J. Chem. Inf. Model.*, vol. 60, no. 6, pp. 2697–2717, 2020.
- [28] F. J. N. Ferreira and A. S. Carneiro, "AI-driven drug discovery: A comprehensive review," *ACS Omega*, vol. 10, no. 23, pp. 23889–23903, 2025, doi: 10.1021/acsomega.5c00549.
- [29] A. Gangwal, A. Ansari, I. Ahmad, A. K. Azad, V. Kumarasamy, V. Subramanian, et al., "Generative artificial intelligence in drug discovery: Basic framework, recent advances, challenges, and opportunities," *Front. Pharmacol.*, vol. 15, Art. no. 1331062, 2024, doi: 10.3389/fphar.2024.1331062.
- [30] A. Tropsha, O. Isayev, A. Varnek, G. Schneider, and A. Cherkasov, "Integrating QSAR modelling and deep learning in drug discovery: The emergence of deep QSAR," *Nat. Rev. Drug Discov.*, vol. 23, pp. 141–155, 2024, doi: 10.1038/s41573-023-00832-0.

- [31] E. N. Muratov, J. Bajorath, R. P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, et al., “QSAR without borders,” *Chem. Soc. Rev.*, vol. 49, pp. 3525–3564, 2020, doi: 10.1039/D0CS00098A.
- [32] H. Li, L. Zou, J. A. H. Kowah, D. He, Z. Liu, X. Ding, et al., “A compact review of progress and prospects of deep learning in drug discovery,” *J. Mol. Model.*, vol. 29, Art. no. 117, 2023, doi: 10.1007/s00894-023-05492-w.
- [33] H. Askr, E. Elgeldawi, H. Aboul Ella, Y. A. M. M. Elshaier, M. M. Gomaa, and A. E. Hassanien, “Deep learning in drug discovery: An integrative review and future challenges,” *Artif. Intell. Rev.*, vol. 56, pp. 5975–6037, 2023, doi: 10.1007/s10462-022-10306-1.
- [34] Y. A. Qadri, S. Shaikh, K. Ahmad, I. Choi, S. W. Kim, and A. V. Vasilakos, “Explainable artificial intelligence: A perspective on drug discovery,” *Pharmaceutics*, vol. 17, no. 9, Art. no. 1119, 2025, doi: 10.3390/pharmaceutics17091119.
- [35] Q. Lv, F. Zhou, X. Liu, and L. Zhi, “Artificial intelligence in small molecule drug discovery from 2018 to 2023: Does it really work?” *Bioorg. Chem.*, vol. 141, Art. no. 106894, 2023, doi: 10.1016/j.bioorg.2023.106894.
- [36] R. Wang, X. Fang, Y. Lu, C. Y. Yang, and S. Wang, “The PDBbind database: Methodologies and updates,” *J. Med. Chem.*, vol. 48, no. 12, pp. 4111–4119, 2005, doi: 10.1021/jm048957q.
- [37] M. Wójcikowski, P. J. Ballester, and P. Siedlecki, “Performance of machine-learning scoring functions in structure-based virtual screening,” *Sci. Rep.*, vol. 7, Art. no. 46710, 2017, doi: 10.1038/srep46710.
- [38] Y. Li et al., “Improving AutoDock Vina using random forest: The growing accuracy of binding affinity prediction by the effective exploitation of larger data sets,” *Mol. Inform.*, vol. 35, no. 12, pp. 598–609, 2016, doi: 10.1002/minf.201400132.
- [39] “Chemprop v2: An efficient, modular machine learning package for chemical property prediction,” *J. Chem. Inf. Model.*, 2025, doi: 10.1021/acs.jcim.5c02332.