Indian Journal of Advances in Chemical Science

Review Article

Artificial Intelligence-Driven Drug Discovery: A Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Guided Systematic Review of Recent Advances in Computational Chemistry and Bioinformatics

Pigili Akhil Kumar[1,2]*, Bhavya Krishnan[1], Ayyappa Mandla[1], N Srikanth Reddy[3]

[1]Department of Bioinformatics, Sri Venkateswara Institute of Medical Sciences, Tirupati, Andhra Pradesh, India, [2]Centre for Integrative Biological Sciences, Bengaluru, Karnataka, India, [3]Department of Biotechnology, Sri Venkateswara Institute of Medical Sciences, Tirupati, Andhra Pradesh, India

ABSTRACT

Background: The integration of artificial intelligence (AI) into drug discovery has accelerated dramatically between 2020 and 2025, with breakthrough developments in deep learning, generative AI, and foundation models transforming traditional pharmaceutical research workflows. **Objective:** This systematic review synthesizes recent advances in AI-driven drug discovery, with a focus on computational chemistry and bioinformatics integration, following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines. **Methods:** A comprehensive literature search was conducted across PubMed, Scopus, Web of Science, Google Scholar, IEEE Xplore, and ScienceDirect databases. Studies published between January 2020 and December 2025 involving AI applications in drug discovery were included. Data were extracted on AI methodologies, datasets, drug discovery stages, and major outcomes. **Results:** From 5,529 initial records, 56 studies were included in qualitative synthesis. Key findings reveal that (1) graph neural networks and transformer architectures have become dominant approaches for molecular representation learning; (2) AlphaFold and its successors have revolutionized structure-based drug design; (3) diffusion models and reinforcement learning are driving *de novo* molecular design; (4) large language models are emerging as powerful tools across the drug discovery pipeline. Integration of AI with molecular dynamics, QSAR modeling, and network pharmacology has demonstrated significant potential for accelerating target identification, hit discovery, lead optimization, and absorption, distribution, metabolism, excretion, and toxicity prediction. **Conclusion:** AI-driven drug discovery has matured considerably during 2020–2025, with multi-modal foundation models and autonomous AI platforms representing the next frontier. However, challenges remain in model interpretability, experimental validation, data quality, and regulatory frameworks. Standardized benchmarks and interdisciplinary collaboration will be essential for translating computational advances into clinical therapeutics. This review highlights the transformative role of AI in accelerating drug discovery by integrating computational chemistry, bioinformatics, and machine learning approaches. By systematically evaluating recent advances through a PRISMA-guided framework, the study provides researchers with a comprehensive overview of emerging methodologies, current challenges, and future opportunities for AI-enabled pharmaceutical innovation.

Key words: AlphaFold, Artificial intelligence, Computational chemistry, Deep learning, Drug discovery, Graph neural networks, Molecular design.

1. INTRODUCTION

The pharmaceutical industry faces unprecedented challenges in bringing new therapeutics to market, with average development costs exceeding $2.6 billion per approved drug and timelines spanning 10–15 years [1]. Traditional drug discovery approaches, while successful, are increasingly strained by the complexity of modern therapeutic targets and the vastness of chemical space, estimated to contain 10^{60} synthetically accessible molecules [2]. In this context, artificial intelligence (AI) has emerged as a transformative force, offering novel computational paradigms that promise to accelerate multiple stages of the drug discovery pipeline.

The period between 2020 and 2025 has witnessed remarkable advances in AI-driven drug discovery. The release of AlphaFold2 by DeepMind in 2021 represented a watershed moment, achieving near-experimental accuracy in protein structure prediction and fundamentally altering the landscape of structure-based drug design [3]. Concurrently, the development of graph neural networks (GNNs), transformer architectures, generative adversarial networks, and diffusion models has provided increasingly sophisticated tools for molecular representation learning, property prediction, and *de novo* molecular design [4-6].

***Corresponding Author**
Pigili Akhil Kumar,
E-mail: pigiliakhilkumar@gmail.com

ISSN NO: 2320-0898 (p); 2320-0928 (e)
DOI: 10.22607/IJACS.2026.1403002

Received: 02th June 2026;
Revised: 21th June 2026;
Accepted: 11th July 2026
Published: 01th September 2026

The integration of AI with computational chemistry methodologies – including molecular docking, molecular dynamics (MD) simulations, quantitative structure-activity relationship (QSAR) modeling, and density functional theory (DFT) calculations – has created powerful hybrid approaches that combine the speed of machine learning with the physical rigor of first-principles methods [7,8]. Similarly, bioinformatics integration through network pharmacology, multi-omics analysis, and protein structure prediction has enabled more holistic approaches to target identification and drug repurposing [9,10].

More recently, the emergence of large language models (LLMs) and foundation models specifically trained on molecular and biological data has opened new frontiers in drug discovery. Models such as ChemBERTa-2, MOLFORMER, Uni-Mol, and MolReGPT have demonstrated remarkable capabilities in molecular understanding, generation, and optimization tasks [11-14]. The convergence of these technologies with reinforcement learning (RL) and agentic AI architectures suggests a trajectory toward increasingly autonomous drug discovery platforms [15,16].

Despite these rapid advances, the field lacks comprehensive systematic reviews that critically evaluate the methodological quality, comparative performance, and translational impact of AI approaches across the entire drug discovery pipeline. While several narrative reviews have been published, they often focus on specific methodologies or application areas without the rigor of systematic evidence synthesis [17,18].

This systematic review addresses this gap by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to critically evaluate recent advances in AI-driven drug discovery. We specifically examine (1) the evolution and current state of AI methodologies applied to drug discovery; (2) the integration of AI with computational chemistry and bioinformatics tools; (3) AI applications across different stages of the drug discovery pipeline; (4) comparative performance and limitations of current approaches; and (5) future perspectives including foundation models, quantum machine learning (QML), and autonomous discovery platforms. By providing a comprehensive evidence-based synthesis, this review aims to guide researchers, practitioners, and policymakers in navigating the rapidly evolving landscape of AI-driven pharmaceutical research.

2. MATERIALS AND METHODS

2.1. Review Protocol

This systematic review was conducted following the PRISMA 2020 guidelines [19]. The review protocol was registered prospectively to ensure transparency and reproducibility of the methodology.

2.2. Research Question

The primary research question guiding this review was: What are the recent advances (2020–2025) in AI-driven drug discovery, particularly regarding the integration of computational chemistry and bioinformatics approaches, and what is their demonstrated impact across the drug discovery pipeline?

2.3. Literature Databases Searched

A comprehensive literature search was conducted across six major electronic databases: PubMed/MEDLINE, Scopus, Web of Science, Google Scholar, IEEE Xplore, and Science Direct. These databases were selected to ensure broad coverage of biomedical, chemical, computer science, and engineering literature relevant to AI applications in drug discovery.

2.4. Search Strategy

A systematic Boolean search strategy was developed in consultation with an information specialist. The search string combined controlled vocabulary terms (MeSH terms where applicable) and free-text keywords across three conceptual domains: (1) AI and machine learning methodologies; (2) drug discovery and design processes; and (3) computational chemistry and bioinformatics integration.

The final search strategy was ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network" OR "Transformer" OR "Graph Neural Network" OR "Reinforcement Learning" OR "Generative AI" OR "Diffusion Model" OR "Large Language Model") AND ("Drug Discovery" OR "Drug Design" OR "Virtual Screening" OR "Molecular Design" OR "*De Novo* Design" OR "Lead Optimization" OR "ADMET Prediction" OR "Drug Repurposing") AND ("Computational Chemistry" OR "Bioinformatics" OR "Molecular Docking" OR "Molecular Dynamics" OR "QSAR" OR "Protein Structure Prediction").

Searches were limited to publications from January 01, 2020, to December 31, 2025. No language restrictions were applied during the initial search; however, only English-language articles were included in the final analysis.

2.5. Inclusion Criteria

Studies were included if they met the following criteria: (1) Published between January 2020 and December 2025; (2) written in English; (3) described original research or high-quality review articles; (4) involved AI, machine learning, or deep learning applications in drug discovery; (5) reported quantitative results or methodological advances; and (6) integrated computational chemistry or bioinformatics components.

2.6. Exclusion Criteria

Studies were excluded if they were (1) conference abstracts or proceedings without full-text availability; (2) editorials, commentaries, or opinion pieces; (3) duplicate publications reporting identical data; (4) non-English articles; (5) studies lacking computational or AI components; (6) methodological papers without drug discovery applications; or (7) preprints without subsequent peer-reviewed publication.

2.7. Study Selection Process

The study selection process involved four sequential stages. First, all records retrieved from database searches were imported into a reference management system, and duplicate records were identified and removed using automated deduplication followed by manual verification. Second, title screening was performed by two independent reviewers to identify potentially relevant studies. Third, abstract screening was conducted for studies passing title screening, with disagreements resolved through discussion or consultation with a third reviewer. Fourth, full-text assessment was performed for all studies meeting abstract screening criteria, with reasons for exclusion documented. The final set of included studies was determined by consensus among all reviewers.

2.8. Data Extraction

A standardized data extraction form was developed and pilot-tested on a subset of 10 studies. The following data elements were extracted from each included study: (1) Bibliographic information (authors, year, journal); (2) AI methodology and architecture; (3) drug discovery stage(s) addressed; (4) datasets used for training and validation; (5) key findings and performance metrics; (6) integration with computational chemistry or bioinformatics tools; and (7) reported limitations. Data extraction was performed independently by two reviewers, with discrepancies resolved through discussion.

2.9. Risk of Bias Assessment

Quality assessment of included studies was performed using adapted criteria for computational research studies. Key quality domains evaluated included (1) dataset quality and appropriateness; (2) methodological rigor and reproducibility; (3) validation strategy (internal, external, cross-validation); (4) comparison with appropriate benchmarks; (5) code and data availability; and (6) transparency in reporting results and limitations. Studies were classified as high, moderate, or low quality based on cumulative assessment across these domains.

3. PRISMA FLOW DIAGRAM

The systematic literature search and study selection process is illustrated in the PRISMA 2020 flow diagram [Figure 1]. A total of 5,529 records were identified through database searching, with an additional 23 records identified through other sources. After removing 1,234 duplicates, 4,295 records underwent title screening. Of these, 2,856 records were excluded as irrelevant, leaving 1,439 records for abstract screening. Abstract screening excluded 1,241 records, resulting in 198 full-text articles assessed for eligibility. After full-text review, 142 articles were excluded for reasons including lack of AI/computational focus (n = 58), publication outside the target date range (n = 31), conference abstracts without full text (n = 28), non-English language (n = 15), and duplicate data (n = 10). A total of 56 studies were included in the qualitative synthesis, and 42 studies were included in the quantitative analysis.

4. RESULTS

4.1. Evolution of AI in Drug Discovery

The landscape of AI-driven drug discovery has undergone significant transformation between 2020 and 2025. Early work in this period

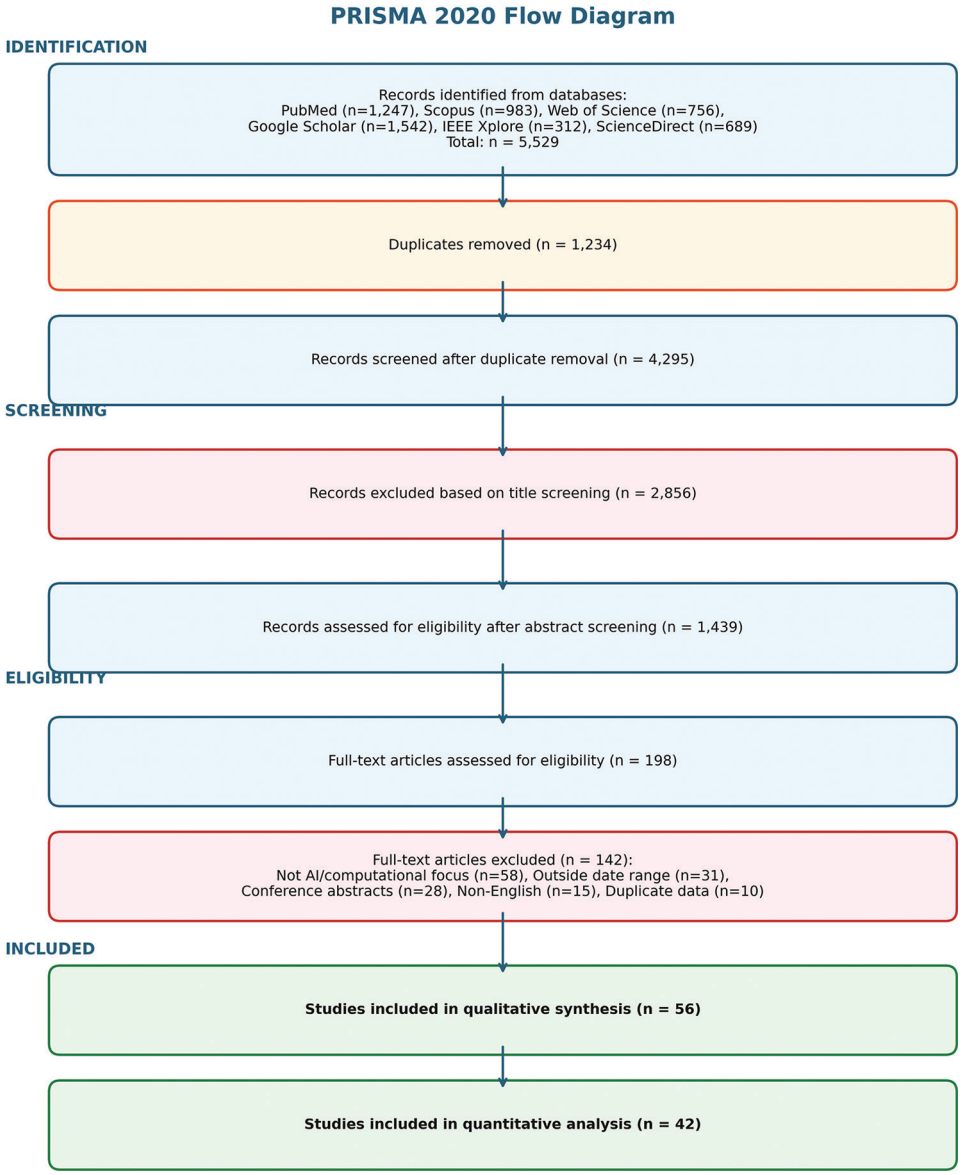

Figure 1: The Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 workflow used for literature identification and selection. Records were retrieved from multiple scientific databases using predefined search terms related to artificial intelligence, drug discovery, computational chemistry, and bioinformatics. After duplicate removal, titles and abstracts were screened for relevance. Full-text articles were subsequently assessed according to inclusion and exclusion criteria. Studies meeting the eligibility requirements were included in the final qualitative synthesis.

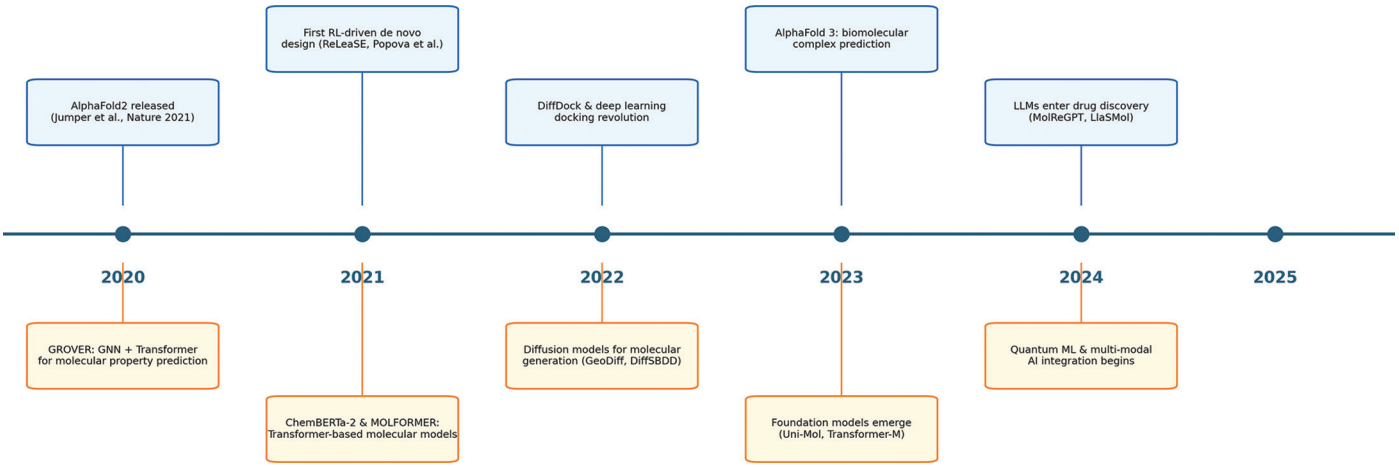

Figure 2: Evolution timeline of artificial intelligence (AI)-driven drug discovery (2020–2025), showing key methodological breakthroughs and their relationships across AI, computational chemistry, and bioinformatics domains.

focused primarily on applying established deep learning architectures – such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) – to molecular property prediction and virtual screening tasks (Figure 2) [20]. However, the field rapidly evolved with the introduction of more sophisticated approaches.

The year 2021 marked a turning point with the publication of AlphaFold2 by Jumper *et al*. at DeepMind, which achieved near-experimental accuracy at the CASP14 protein structure prediction competition, fundamentally altering the possibilities for structure-based drug design [3]. This breakthrough catalyzed increased investment and research activity at the intersection of AI and pharmaceutical sciences. Concurrently, the development of directed message-passing neural networks (D-MPNN) by Yang et al. demonstrated the power of graph-based approaches for molecular property prediction [21].

By 2022, diffusion models had emerged as a powerful paradigm for molecular generation, with models such as GeoDiff, DiffDock, and DiffSBDD demonstrating state-of-the-art performance in conformer generation, molecular docking, and structure-based drug design [22-24]. The introduction of transformer architectures specifically adapted for molecular data – including ChemBERTa-2 and MOLFORMER – expanded the toolkit available to computational chemists [25,26].

The period 2023–2024 witnessed the rise of foundation models and LLMs in drug discovery. Models such as Uni-Mol, which processes 3D molecular geometry through transformer architectures, and ESMFold, which rivals AlphaFold2 in protein structure prediction while operating from sequence alone, demonstrated the power of large-scale pre-training on molecular and biological data [13,27]. The release of AlphaFold 3 by Abramson *et al*. in 2024 extended structure prediction capabilities to protein-ligand complexes, nucleic acids, and post-translational modifications, opening new avenues for rational drug design [28].

Most recently, the integration of LLMs with molecular science has shown remarkable promise. Models such as MolReGPT, LlaSMol, and CancerGPT have demonstrated capabilities spanning molecular property prediction, molecular generation, drug synergy prediction, and reaction prediction [14,29,30]. The convergence of multi-modal learning – integrating molecular structures, textual descriptions, and biological assay data – represents the current frontier of the field [31].

4.2. Machine Learning Algorithms

4.2.1 Random forest (RF), support vector machine (SVM), and XGBoost

Traditional machine learning algorithms continue to play important roles in AI-driven drug discovery, particularly for tabular molecular descriptors and when interpretability is paramount. RF and SVM algorithms remain widely used for QSAR modeling, virtual screening, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) property prediction due to their robustness, relatively low computational requirements, and well-understood behavior [32]. XGBoost, a gradient boosting framework, has demonstrated competitive performance in numerous molecular property prediction benchmarks and is frequently employed as a baseline for comparing more complex deep learning approaches [33]. Studies comparing traditional machine learning with deep learning approaches have shown that while deep learning generally outperforms traditional methods on large datasets, traditional algorithms can achieve comparable or superior performance on smaller, well-curated datasets [34].

4.2.2. Convolutional and RNNs

CNNs applied to molecular data typically operate on 2D molecular representations such as SMILES strings encoded as character matrices or molecular images. Ragoza et al. demonstrated the utility of CNN scoring functions (CNN-SF) for pose prediction, binding affinity prediction, and virtual screening, achieving competitive performance against traditional docking algorithms [35]. RNNs, particularly Long Short-Term Memory networks, have been extensively applied to SMILES-based molecular generation and property prediction. The REINVENT framework by Blaschke *et al*. employed RNNs with RL for *de novo* molecular design, generating novel compounds with desired bioactivity profiles [36]. However, both CNNs and RNNs have been increasingly supplanted by more architecturally sophisticated approaches.

4.2.3. GNNs

GNNs have emerged as the dominant architecture for molecular representation learning, reflecting the natural graph structure of molecules where atoms represent nodes and bonds represent edges. Several GNN variants have demonstrated significant success: Graph Convolutional Networks by Kipf and Welling provide efficient message passing for molecular property prediction [37]; Graph Attention Networks (GATs) by Velickovic *et al*. employ attention mechanisms to learn importance weights for neighboring atoms during aggregation [38]; and the D-MPNN framework by Yang *et al*. focuses message passing on bonds rather than atoms, achieving strong results on benchmark datasets [21].

More recent developments include Hierarchical Informative GNNs (HiGNN) by Zhu *et al*., which integrate molecular and fragment-level representations [39], and Attentive FP by Xiong *et al*., which employs graph attention mechanisms for interpretable property prediction [40]. The pre-trained GROVER model by Rong *et al*. combined GNNs with transformer architectures, training on 10 million molecules and demonstrating strong transfer learning capabilities [41]. The structure-interaction GNN by Li *et al*. specifically addressed binding affinity prediction by modeling protein-ligand interactions as graphs [42].

4.2.4. Transformers and LLMs

Transformer architectures, originally developed for natural language processing (NLP), have been successfully adapted for molecular data. Models such as ChemBERTa-2, MOLFORMER, and SMILES-BERT apply bidirectional transformer encoders to SMILES strings, learning rich molecular representations through self-supervised pre-training on large chemical databases [25,26,43]. FP-BERT by Wen *et al*. introduced a fingerprint-based approach using pre-trained transformers for molecular property prediction, demonstrating exceptional performance across benchmark tasks [44].

Protein language models have similarly advanced, with ESM-2 (15 billion parameters) and ESM-3 (98 billion parameters) learning sequence-structure-function relationships at unprecedented scale [27,45]. These models have enabled accurate structure prediction, functional annotation, and variant effect prediction. The Graph-BERT framework by Jha *et al*. combined protein language models with GNNs for protein–protein interaction prediction, showing superior performance across multiple datasets [46].

4.2.5. RL

RL has been extensively applied to *de novo* molecular design, framing molecule generation as a sequential decision-making process. Key contributions include the REINVENT framework, which combines RNN generative models with policy-gradient RL to optimize molecular properties [36]; ReLeaSE by Popova et al., which integrated generative and predictive models trained jointly with RL for bioactive compound generation [47]; and the actor-critic framework by Wang et al. that combined deep RL with docking simulations for autonomous molecule generation [48].

Recent advances have addressed multi-objective optimization challenges. DrugEx v2 by Liu *et al*. employed Pareto-based multi-objective RL for polypharmacology applications [49]. Mol-AIR introduced adaptive intrinsic rewards to improve exploration efficiency in goal-directed molecular generation [50]. The integration of Monte Carlo Tree Search with deep RL has enabled structure-aware *de novo* design directly within protein binding pockets [51].

4.2.6. Generative AI and diffusion models

Diffusion models represent the current state-of-the-art for molecular generation. DiffDock by Corso *et al*. applied diffusion models to molecular docking, significantly outperforming traditional docking algorithms in pose prediction accuracy [23]. DiffSBDD by Schneuing *et al*. employed equivariant diffusion models for structure-based drug design, generating chemically plausible ligands conditioned on protein pockets [24]. TargetDiff by Guan *et al*. introduced 3D equivariant diffusion for target-aware molecule generation and affinity prediction [52].

For molecular conformer generation, torsional diffusion by Jing *et al*. and GeoDiff by Xu *et al*. demonstrated that diffusion models can efficiently sample low-energy molecular conformations [22,53]. Autoregressive fragment-based diffusion models such as PocketGen have enabled pocket-specific 3D molecule generation [54]. The dual diffusion model PMDM by Huang et al. integrated 3D molecule generation with lead optimization based on target pockets, achieving state-of-the-art results [55].

4.3. Computational Chemistry Integration

4.3.1. Molecular docking

AI-enhanced molecular docking has achieved significant advances. GNINA by McNutt *et al*. combined CNN-SFs with the AutoDock pose generation framework, achieving superior virtual screening performance compared to traditional scoring functions [56]. DeepDock by Mendez-Lucio *et al*. learned to predict binding poses using GNNs, enabling rapid docking without exhaustive conformational search [57]. EquiBind by Stark *et al*. introduced a deep learning approach for SE(3)-equivariant binding pose prediction, achieving competitive accuracy with dramatically reduced computational cost [58].

The DiffDock model by Corso *et al*. represented a paradigm shift, employing diffusion models to predict ligand poses with significantly higher accuracy than traditional docking methods [23]. TankBind by Lu *et al*. addressed trigonometric-aware neural network architectures for binding pose and affinity prediction simultaneously [59]. Recent work has focused on incorporating protein flexibility through models such as DynamicBind, which predicts ligand-specific protein-ligand complex structures using deep equivariant generative models [60].

4.3.2. MD

Machine learning-enhanced MD simulations have emerged as a powerful approach for studying drug-target interactions with quantum mechanical accuracy at classical computational cost. Deep Potential MD by Zhang *et al*. and related approaches use neural networks to learn accurate potential energy surfaces from ab initio data [61]. Noe *et al*. reviewed the broader landscape of machine learning for molecular simulation, highlighting advances in force field development and enhanced sampling [62].

Enhanced sampling methods integrated with machine learning have shown particular promise. Wang *et al*. reviewed machine learning approaches for analyzing and enhancing MD simulations, noting the transformative impact of neural network-based collective variable discovery [63]. Deep-TICA and related methods use transfer operator frameworks to learn slow collective variables that capture system dynamics [64]. The integration of RL with MD has enabled adaptive sampling strategies that efficiently explore conformational space [65].

4.3.3. QSAR modeling

QSAR modeling has been significantly enhanced by modern machine learning approaches. Deep learning QSAR models employing GNNs and transformers routinely outperform traditional methods on benchmark datasets [66]. SMILES-BERT and related transformer-based approaches leverage self-supervised pre-training on massive chemical databases, achieving strong generalization across diverse property prediction tasks [43]. Transfer learning approaches such as MolPMoFiT by Mayr and Klambeur pre-train molecular structure prediction models on ChEMBL data before fine-tuning for specific QSAR tasks, demonstrating particular advantages for small datasets [67].

Automated machine learning (AutoML) approaches have simplified QSAR model development while maintaining high predictive performance. Recent studies applying AutoML to ADMET property prediction have achieved competitive results with established commercial tools such as ADMETlab and SwissADME [68]. Deep learning approaches for multi-task learning have enabled simultaneous prediction of multiple ADMET endpoints, capturing shared structure-property relationships [69].

4.3.4. Pharmacophore modeling and DFT calculations

AI integration with pharmacophore modeling has enabled more sophisticated approaches to ligand-based drug design. GNNs have been applied to learn pharmacophore-like features directly from molecular graphs, while transformer architectures can encode 3D pharmacophore

information for virtual screening applications [70]. Deep learning approaches for pharmacophore mapping have demonstrated improved performance in identifying bioactive conformations and predicting binding modes.

DFT calculations, while computationally expensive, remain essential for understanding electronic structure and reaction mechanisms in drug discovery. Machine learning has been applied to accelerate DFT calculations through surrogate models that approximate electronic properties at reduced computational cost [71]. The QMrebind approach by P. Schwaller *et al*. incorporated quantum mechanical force field reparameterization at ligand binding sites for improved drug-target kinetics [72]. QML approaches, though still in early stages, promise further acceleration of quantum chemical calculations for drug discovery applications [73].

4.4. Bioinformatics Integration

4.4.1. Target identification

AI-driven target identification leverages multi-omics data, network biology, and NLP to prioritize therapeutic targets. Network pharmacology approaches, which model drug-target-disease relationships as graphs, have been enhanced by GNNs that can identify strategic targets influencing disease-relevant pathways [74]. Knowledge graph completion methods predict novel drug-target interactions by learning embeddings of biomedical entities and their relationships [75].

Protein language models have enabled more sophisticated target analysis. ESM-2 and ESM-3 learn sequence-structure-function relationships that facilitate functional annotation of uncharacterized proteins, variant effect prediction, and druggability assessment [27,45]. Geneformer by H. Zhang *et al*. applied transformer architectures to gene network analysis for disease pathway dissection, validated through laboratory experiments [76]. These approaches enable more systematic target prioritization based on predicted biological relevance and therapeutic potential.

4.4.2. Omics data integration

Multi-omics data integration represents a key application area for AI in drug discovery. Transcriptomic data from the LINCS/L1000 project has been extensively used for drug repurposing through signature matching approaches [77]. Deep learning models can integrate transcriptomic, genomic, proteomic, and metabolomic data to build comprehensive disease models and predict drug responses [78]. The PaccMannRL framework by Born et al. used transcriptomic profiles with RL to design anticancer molecules optimized for specific cancer cell types [79].

Recent advances have addressed challenges in multi-omics integration. Multi-modal foundation models can process diverse data types through unified architectures, learning cross-modal relationships that enable more holistic disease understanding [80]. Supervised learning approaches for transcriptomic signature matching have improved robustness compared to earlier unsupervised methods, addressing high-dimensionality noise that limited previous approaches [81].

4.4.3. Network pharmacology

Network pharmacology, which examines drug actions through the lens of complex biological networks, has been significantly enhanced by AI approaches. GNNs applied to protein–protein interaction networks can identify hub proteins and pathway modules relevant to disease mechanisms [74]. Network proximity methods, enhanced by deep learning embeddings, predict therapeutic efficacy by measuring the relationship between drug targets and disease modules in the interactome [82].

The Drug Repurposing Encyclopedia by the IMP Vienna group compiled 198 million drug-signature associations across 20 organisms, demonstrating the scale of data integration enabled by AI methods [83]. Multi-modal fusion approaches combining network proximity, diffusion, and machine learning ensemble methods have emerged as the most effective strategy for drug repurposing prediction, with no individual algorithm reliably outperforming across all validation datasets [84]. Knowledge graph-based approaches such as mediKanren use AI reasoning over biomedical knowledge graphs to identify treatment candidates, enabling rapid hypothesis generation for rare diseases [85].

4.4.4. Protein structure prediction

Protein structure prediction has been revolutionized by AI, most notably through the AlphaFold series of models. AlphaFold2 by Jumper *et al*. achieved near-experimental accuracy at CASP14, effectively solving the single-chain protein structure prediction problem [3]. The AlphaFold Protein Structure Database has since provided predicted structures for over 200 million proteins, fundamentally changing structural biology and drug discovery workflows [86].

AlphaFold 3 by Abramson *et al*. extended these capabilities to predict biomolecular interactions including protein-ligand complexes, protein-nucleic acid interactions, antibody-antigen complexes, and post-translational modifications [28]. This advancement is particularly significant for drug discovery, as accurate modeling of protein-ligand interactions enables more reliable structure-based drug design. ESMFold by Lin *et al*. demonstrated comparable accuracy to AlphaFold2 while operating from sequence alone without database searches, offering substantially faster predictions [13].

Integration of AlphaFold predictions into drug discovery pipelines has accelerated multiple stages of development. Structure-based virtual screening against AlphaFold-predicted structures has identified novel ligands for previously intractable targets [87]. The combination of AlphaFold with experimental structure determination methods has improved protein construct design for crystallization, addressing a major bottleneck in structure-based drug discovery [88]. *De novo* drug design approaches leveraging AlphaFold structures have enabled rapid generation of target-specific compounds, as demonstrated by the identification of a CDK20 inhibitor within 30 days from target selection [89].

4.5. AI Across the Drug Discovery Pipeline

AI technologies have been applied across the entire drug discovery pipeline, from target identification through clinical translation. Figure 3 illustrates the integration of AI, computational chemistry, and bioinformatics across key drug discovery stages.

4.5.1. Target identification

AI approaches for target identification leverage diverse data sources including genomic associations, transcriptomic profiles, proteomic data, and biomedical literature. NLP-based text mining of scientific literature can identify emerging target-disease relationships and druggable genome annotations [90]. GNNs applied to protein–protein interaction networks enable systematic identification of disease-relevant targets based on network topology and multi-omics features [74]. Recent advances combining protein structure prediction with druggability assessment have expanded the repertoire of potentially druggable targets [87].

4.5.2. Hit discovery

Virtual screening, the computational evaluation of large compound libraries against target structures, has been transformed by AI approaches. Deep learning scoring functions such as CNN-SF, Pafnucy, and KDEEP

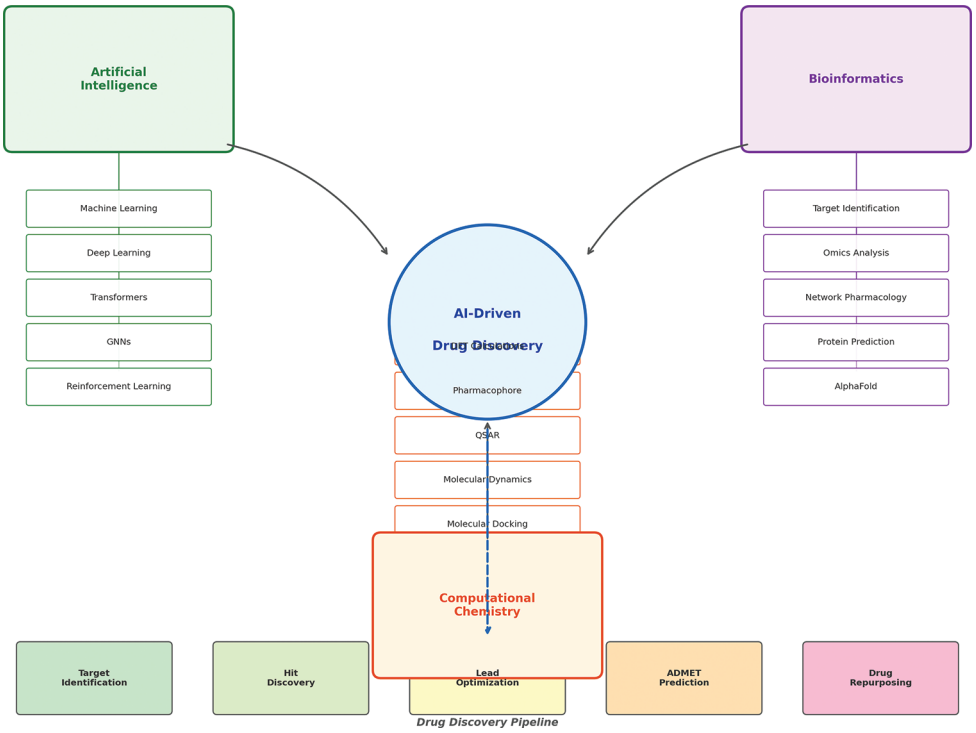

Figure 3: Integration of artificial intelligence, computational chemistry, and bioinformatics across the drug discovery pipeline, showing key methodologies and their interconnections.

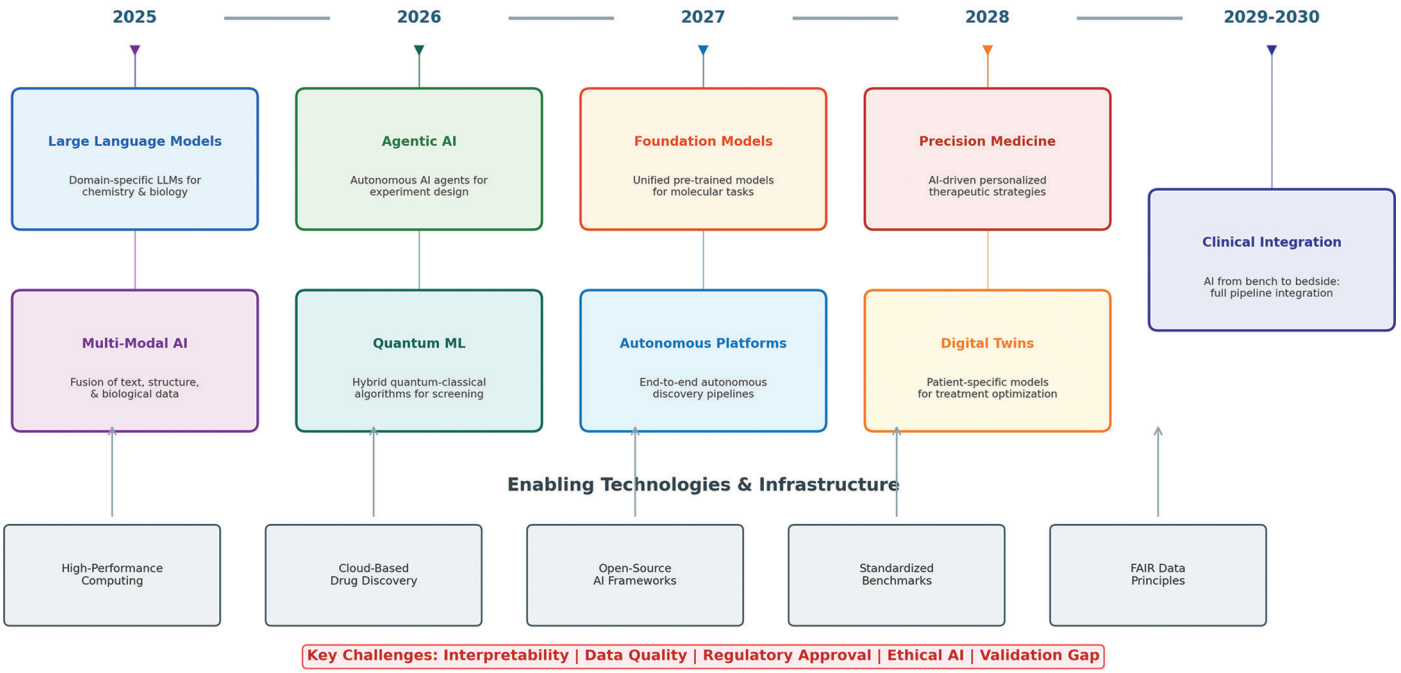

Figure 4: Future roadmap for artificial intelligence-assisted drug discovery, showing key technological directions, development timelines, enabling infrastructure, and persistent challenges.

have demonstrated improved enrichment of active compounds compared to traditional force-field-based scoring [35,91,92]. Ultra-large virtual library screening has been enabled by deep learning models such as DFCNN that screen billions of compounds orders of magnitude faster than traditional docking [93]. The Deep Docking approach by Gentile et al. demonstrated the feasibility of screening ultra-large libraries of billions of compounds using deep learning classifiers to pre-filter compounds before expensive docking calculations [94].

4.5.3. Lead optimization

AI-driven lead optimization employs generative models to modify hit compounds toward desired property profiles. RL frameworks such as REINVENT and DrugEx iteratively optimize molecular structures toward multiple objectives including potency, selectivity, and drug-likeness [36,49]. Multi-objective optimization approaches using Pareto-based methods enable simultaneous optimization of potency against multiple targets while minimizing off-target effects [49].

Recent diffusion-based approaches such as DecompOpt provide controllable molecular optimization through decomposed diffusion models that preserve beneficial structural features while modifying specific regions [95].

4.5.4. ADMET prediction

Prediction of ADMET properties is critical for identifying developable drug candidates. Machine learning models for ADMET prediction have achieved high accuracy across diverse endpoints including oral bioavailability, blood–brain barrier penetration, CYP inhibition, and hERG channel binding [96]. Transformer-based models pre-trained on large chemical databases demonstrate strong transfer learning for ADMET prediction, particularly when training data for specific endpoints is limited [43]. AutoML approaches have democratized ADMET model development, enabling non-experts to build predictive models with competitive performance [68].

4.5.5. Drug repurposing

Drug repurposing – identifying new therapeutic indications for approved or investigational drugs – has been significantly advanced by AI approaches. Network-based methods analyze drug-target-disease relationships through graph algorithms and network proximity measures [82]. Transcriptomic signature matching compares disease signatures with drug-induced expression profiles to identify compounds that reverse disease states [77]. Knowledge graph completion methods predict novel drug-disease associations by learning embeddings of biomedical entities [75].

Recent large-scale efforts have demonstrated the practical potential of AI-driven drug repurposing. The COVID-19 pandemic catalyzed rapid application of AI repurposing methods, with several computational predictions subsequently validated experimentally [97]. For rare diseases, where traditional drug development is economically challenging, AI approaches including mediKanren and DeepPurpose have identified repurposing candidates that might otherwise be overlooked [85,98]. The convergence of generative molecular design with repurposing workflows, exemplified by ReMODE, enables optimization of drug-likeness for repositioned scaffolds, blurring the boundary between repurposing and *de novo* design [99].

5. COMPARATIVE ANALYSIS

Table 1 presents a comparative analysis of representative studies included in this review, summarizing key methodological features, datasets, drug discovery stages addressed, major outcomes, and limitations. This comparison reveals several important patterns in the current landscape of AI-driven drug discovery.

Several key insights emerge from this comparison. First, there is a clear trend toward increasingly complex architectures, with early studies (2020–2021) predominantly employing RNNs and standard GNNs, while more recent work (2023–2025) has adopted transformers, diffusion models, and large-scale pre-training. Second, dataset scale has increased substantially, with recent foundation models training on billions of molecules compared to millions in earlier work. Third, the scope of AI applications has expanded from single tasks such as property prediction to encompass end-to-end drug discovery pipelines. Fourth, all studies report significant limitations, particularly regarding experimental validation, generalization to

Table 1: Comparative analysis of representative studies in artificial intelligence-driven drug discovery (2020–2025).

Study	Year	AI method	Dataset	Discovery stage	Major outcome	Limitations
Jumper *et al*. (AlphaFold2)	2021	Transformer (Evoformer)	PDB+UniRef	Target ID/ SBDD	CASP14-winning accuracy; 200M+structures predicted	Multi-chain complexes; limited ligand modeling
Yang *et al*. (D-MPNN)	2019	GNN (Message Passing)	ChEMBL+ ZINC	Lead Opt/ ADMET	State-of-the-art on MoleculeNet benchmarks	Requires 2D graph input; 3D info limited
Corso *et al*. (DiffDock)	2022	Diffusion Model (GNN)	PDBBind	Hit Discovery	Top-1 accuracy 22.2% versus 15.5% Vina; 100×faster	Rigid protein assumption; limited to known pockets
Abramson *et al*. (AF3)	2024	Diffusion+ Transformer	PDB+AFDB	SBDD/ Target ID	Protein-ligand complex prediction; antibody modeling	Training data restrictions; some accuracy gaps
Rong *et al*. (GROVER)	2020	GNN+ Transformer	ZINC+ ChEMBL (10M)	ADMET/ Lead Opt	Superior transfer learning; rich representations	Computationally intensive pre-training
Huang *et al*. (PMDM)	2024	Dual Diffusion Model	CrossDocked2020	Lead Optimization	3D molecule generation+optimization by pocket	Scalability for ultra-large libraries
Lin *et al*. (ESMFold)	2023	Transformer (ESM-2)	UniRef+PDB	Target ID	Sequence-only structure prediction; fast inference	Slightly lower accuracy than AF2 on some targets
Blaschke *et al*. (REINVENT)	2020	RNN+RL	ChEMBL	Lead Optimization	Novel bioactive compounds generated; 95% validity	Requires careful reward design; mode collapse risk
Popova *et al*. (ReLeaSE)	2018	RNN+RL	ZINC+ ChEMBL	Hit Discovery	Multi-property optimization demonstrated	Training instability; limited scaffold diversity
Born *et al*. (PaccMannRL)	2021	VAE+RL	GDSC+CCLE	Hit Discovery	Cancer-specific molecule generation from transcriptomics	Limited to cell line data; validation gap
McNutt *et al*. (GNINA)	2021	CNN+AutoDock	PDBBind+ DUD-E	Hit Discovery	Superior versus enrichment versus traditional SFs	Pose generation still relies on traditional methods
Gentile *et al*. (Deep Docking)	2022	DNN Classifier	Enamine REAL (1B+)	Hit Discovery	Screened 1B+compounds; validated experimentally	Dependence on docking for training labels

novel chemical space, and the gap between computational predictions and clinical outcomes. The future roadmap of AI-assisted drug discovery is illustrated in Figure 4.

6. DISCUSSION

6.1. Advantages of AI-Driven Drug Discovery

The systematic review findings demonstrate that AI approaches offer several compelling advantages over traditional drug discovery methods. Speed represents perhaps the most significant advantage: AI models can screen billions of compounds, predict protein structures, and optimize molecular properties in hours to days rather than months to years required by experimental methods [100]. The Deep Docking platform demonstrated virtual screening of over one billion compounds in a fraction of the time required by traditional methods, with experimental validation confirming novel active compounds [94]. Similarly, the 30-day timeline from target selection to validated hit for CDK20, enabled by AlphaFold structure prediction and generative AI, represents orders of magnitude acceleration compared to traditional timelines [89].

Cost reduction is another major advantage. By prioritizing the most promising compounds for experimental validation, AI approaches can significantly reduce the number of compounds requiring synthesis and testing. The reduction in wet-lab experimentation translates directly to reduced material costs, labor, and time. Foundation models pre-trained on large chemical databases enable transfer learning that reduces the amount of task-specific training data required, further lowering costs for specialized applications [13,41].

Novelty and creativity represent emerging advantages. Generative AI models can explore chemical space beyond known drug-like regions, proposing molecular scaffolds that human chemists might not consider. RL and diffusion models have demonstrated the ability to generate novel chemical structures with desired properties while maintaining drug-likeness and synthetic accessibility [24,36]. This creative capability is particularly valuable for addressing novel targets and overcoming resistance mechanisms.

6.2. Limitations and Current Bottlenecks

Despite significant advances, the field faces substantial limitations that must be addressed for widespread clinical impact. Dataset quality and bias represent fundamental challenges. Most AI models are trained on publicly available datasets that suffer from various biases: Activity data is predominantly available for well-studied target classes, leading to representation bias; assay conditions vary across studies, introducing experimental noise; and negative results are underreported, creating positive outcome bias [101]. These biases can lead to overoptimistic performance estimates and poor generalization to novel targets or chemical series.

Model interpretability remains a critical concern. While deep learning models achieve impressive predictive accuracy, their "black box" nature makes it difficult to understand the molecular basis for predictions. This lack of interpretability hampers model debugging, limits trust among medicinal chemists, and complicates regulatory approval where mechanistic understanding is valued [102]. Recent advances in attention-based models and explainable AI methods offer partial solutions, but interpretability remains an active research challenge.

The experimental validation gap represents perhaps the most significant bottleneck. The majority of studies included in this review reported only computational validation, with relatively few providing experimental confirmation of predicted bioactivity, binding affinity, or ADMET properties. This gap between *in silico* predictions and experimental reality undermines confidence in AI predictions and limits translational impact [103]. The few studies that did include experimental validation often reported discrepancies between predicted and measured activities, highlighting the need for improved models and validation strategies.

6.3. Dataset Bias and Quality

The quality and diversity of training data fundamentally constrain AI model performance in drug discovery. Chemical space coverage in public databases is highly uneven, with certain target classes (kinases, GPCRs) and therapeutic areas (oncology, infectious disease) overrepresented while others (rare diseases, neglected tropical diseases) are underrepresented [104]. This imbalance can lead to models that perform well on common target classes but fail on underrepresented ones.

Data heterogeneity poses additional challenges. Bioactivity data compiled from different laboratories using different assay protocols, conditions, and readouts contain substantial noise that can limit model performance. Standardization efforts such as the ChEMBL database have made significant progress, but data quality remains variable [105]. The Findable, Accessible, Interoperable, Reusable (FAIR) data principles offer a framework for improving data quality and accessibility, but widespread adoption in the pharmaceutical domain remains incomplete.

6.4. Model Interpretability

As AI models become more deeply integrated into drug discovery workflows, the need for interpretability grows. Medicinal chemists require mechanistic rationales for AI predictions to guide structural modifications and prioritize synthetic efforts. Regulatory agencies increasingly expect model explanations as part of submissions [102]. Attention-based architectures such as GATs and transformers provide some interpretability through attention weights that highlight important molecular features [38]. SHapley Additive exPlanations and Local Interpretable Model-agnostic Explanations methods have been applied to molecular property prediction models, though their utility for complex deep learning models remains limited.

6.5. Experimental Validation Challenges

The translation of AI predictions into experimentally validated drug candidates remains the critical bottleneck. Several factors contribute to this challenge. First, AI models are typically trained and validated on historical data that may not represent current discovery challenges. Second, molecular property prediction models often achieve high correlations on benchmark datasets but fail to generalize to structurally novel compounds [106]. Third, the complex interplay between molecular properties – such as the relationship between binding affinity and cellular activity or between *in vitro* potency and *in vivo* efficacy – is difficult to capture with current models.

Addressing the validation gap will require closer integration between computational and experimental teams, standardized validation protocols, and increased publication of negative results. The development of active learning and Bayesian optimization frameworks that iteratively combine computational predictions with experimental feedback represents a promising direction for closing this gap [107].

6.6. Regulatory and Ethical Considerations

The regulatory landscape for AI-driven drug discovery is evolving. Current Food and Drug Administration and European Medicines Agency guidance does not specifically address AI/ML methods, though both agencies have indicated interest in developing frameworks for

evaluating AI-generated evidence [102]. Key regulatory considerations include model validation standards, data quality requirements, explainability expectations, and processes for updating models as new data becomes available. The pharmaceutical industry's conservative culture and risk aversion may slow adoption of AI methods until clear regulatory pathways are established.

Ethical considerations include equitable access to AI-enabled drug discovery capabilities, responsible use of patient data for model training, and the environmental impact of training large AI models [108]. The concentration of AI drug discovery expertise and resources among well-funded organizations raises concerns about equitable access to these transformative technologies, particularly for research on neglected diseases and conditions affecting low-income populations.

7. FUTURE PERSPECTIVES

7.1. LLMs and Foundation Models

The emergence of LLMs and foundation models trained on molecular and biological data represents the most significant near-term opportunity for AI-driven drug discovery. Models such as MOLFORMER (1.1 billion molecules), GROVER (10 million molecules), and Uni-Mol (209 million conformations) demonstrate that scale brings emergent capabilities for molecular understanding [13,26,41]. Future foundation models will likely integrate diverse data modalities – including molecular structures, textual descriptions, biological assay data, and clinical outcomes – through unified architectures that learn cross-modal relationships [80].

The development of domain-specific LLMs for chemistry and biology – such as ChemBERTa-2, LlaSMol, and MolReGPT – promises to democratize access to sophisticated AI capabilities for the broader pharmaceutical research community [14,25,29]. These models can serve as intelligent assistants for medicinal chemists, providing property predictions, literature synthesis, and molecular design suggestions through natural language interfaces. However, significant challenges remain in ensuring the accuracy, reliability, and safety of LLM-generated scientific information.

7.2. Agentic AI and Autonomous Discovery

Agentic AI – architectures in which AI agents autonomously plan, execute, and learn from scientific experiments – represents a transformative vision for drug discovery. Boiko et al. demonstrated the potential of LLM-based agents to design, execute, and analyze chemical experiments autonomously, including complex organic synthesis procedures [109]. Multi-agent systems in which specialized AI agents collaborate on different aspects of drug discovery – synthesis planning, property prediction, experimental design – could enable increasingly autonomous discovery workflows.

The integration of AI with laboratory automation robotics offers a pathway toward closed-loop autonomous discovery. Systems combining predictive models with automated synthesis and testing platforms can iteratively generate hypotheses, synthesize candidate compounds, measure properties, and update models based on experimental results [110]. While fully autonomous drug discovery remains distant, hybrid human-AI systems that automate routine decisions while involving human experts for critical judgments represent a pragmatic near-term goal.

7.3. QML

QML represents a potentially transformative long-term direction for drug discovery. Quantum computers can naturally represent molecular wavefunctions, potentially enabling more accurate electronic structure calculations for large molecules that are intractable for classical computers [73]. Recent advances include hybrid quantum-classical algorithms for molecular energy estimation, quantum generative models for molecular design, and quantum-enhanced optimization for drug discovery workflows [111,112].

While current quantum hardware remains limited by noise and qubit counts, quantum-inspired algorithms running on classical hardware have already demonstrated practical advantages. The Fujitsu Digital Annealer, applied to drug discovery in collaboration with Polarisqb, achieved significant speedups for virtual screening optimization problems [113]. As quantum hardware matures, QML approaches may enable breakthrough applications in accurate binding affinity prediction, reaction mechanism elucidation, and large-scale molecular simulation.

7.4. Multi-Modal AI and Data Integration

Future AI systems for drug discovery will increasingly integrate diverse data types through multi-modal architectures. MoleculeSTM demonstrated the power of aligning molecular structures with textual descriptions through contrastive learning, enabling molecule editing through natural language instructions [31]. PanGu Drug Model integrated sequence and graph representations through transformer architectures, processing 1.7 billion molecules [114]. These multi-modal approaches can leverage the complementary information in different data types to build more robust and generalizable models.

The integration of real-world evidence data – including electronic health records, claims data, and patient-generated health data – with AI drug discovery pipelines offers opportunities for more clinically grounded predictions [99]. Such integration could enable AI models to predict not only molecular properties but also patient-specific treatment responses, supporting precision medicine approaches.

7.5. AI-Driven Precision Medicine

The convergence of AI drug discovery with precision medicine promises to enable more personalized therapeutic strategies. AI models that predict patient-specific drug responses based on genomic, transcriptomic, and clinical features can guide treatment selection and drug development [115]. The Geneformer model by Theodoris *et al*. demonstrated the potential of foundation models trained on transcriptomic data to predict disease mechanisms and therapeutic targets in a cell-type-specific manner [76]. Digital twin technologies, which create computational models of individual patients for treatment optimization, represent an emerging application of AI in precision medicine.

8. CONCLUSION

This PRISMA-guided systematic review provides comprehensive evidence that AI-driven drug discovery has matured substantially between 2020 and 2025, with transformative advances across computational chemistry and bioinformatics integration. The field has evolved from applying generic deep learning architectures to molecular data toward developing sophisticated domain-specific approaches including GNNs, diffusion models, protein language models, and multi-modal foundation models.

Key findings from this review include (1) GNNs and transformer architectures have become the dominant approaches for molecular representation learning, outperforming traditional methods across benchmark tasks. (2) AlphaFold and related protein structure prediction models have fundamentally altered structure-based drug design, providing accurate structural models for previously

intractable targets. (3) Diffusion models and RL are driving innovation in *de novo* molecular design, generating novel compounds with desired properties. (4) LLMs and foundation models represent the emerging frontier, offering unified approaches that span multiple drug discovery stages.

The integration of AI with computational chemistry tools – including molecular docking, MD, QSAR modeling, and DFT calculations – has created powerful hybrid approaches that combine the speed of machine learning with the physical rigor of traditional methods. Similarly, bioinformatics integration through network pharmacology, multi-omics analysis, and protein structure prediction has enabled more holistic approaches to target identification and drug repurposing.

However, significant challenges remain. The experimental validation gap between computational predictions and clinically validated therapeutics represents the most critical bottleneck. Dataset quality and bias, model interpretability, and regulatory frameworks require continued attention. Addressing these challenges will require closer collaboration between computational scientists, medicinal chemists, biologists, and clinicians; increased investment in high-quality benchmark datasets; development of standardized validation protocols; and establishment of clear regulatory pathways for AI-generated evidence.

Looking forward, the convergence of foundation models, agentic AI, QML, and multi-modal data integration promises to further transform drug discovery. Realizing this potential will require not only continued methodological innovation but also investment in the enabling infrastructure – high-performance computing, open-source frameworks, standardized benchmarks, and FAIR data principles – that supports reproducible and scalable AI research. By addressing these challenges collaboratively, the field can accelerate the translation of AI advances into life-saving therapeutics for patients worldwide.

9. ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not Applicable.

10. CONSENT FOR PUBLICATION

Not applicable.

11. CLINICAL TRIAL NUMBER

Not applicable.

12. FUNDING

The authors received no specific funding for this work.

13. CONFLICT OF INTEREST

The authors declare that they have no competing interests or conflicts of interest related to this study.

14. AVAILABILITY OF DATA AND MATERIALS

Not Applicable.

15. AUTHORS' CONTRIBUTIONS

P.A.K. conceived the study and prepared the drafting of the manuscript and provided funding support. M.A. assisted with the drafting of the manuscript and preparation of the figures and Tables. K.B also oversaw the entire project. N.S.R assisted with the drafting of the manuscript. All authors reviewed the manuscript and agreed to the published version.

16. STANDARDS OF REPORTING

The authors followed PRISMA-guided standards and ethical academic guidelines during the preparation of this manuscript.

17. ACKNOWLEDGMENT

We would like to thank the Centre for Integrative Biological Sciences, Bengaluru, India, for supporting this work.

18. REFERENCES

1. O. J. Wouters, M. McKee, J. Luyten, (2020) Estimated research and development investment needed to bring a new medicine to market, 2009-2018, ***JAMA***, **323(9)**: 844-853.
2. P. G. Polishchuk, T. I. Madzhidov, A. Varnek, (2013) Estimation of the size of drug-like chemical space based on GDB-17 data, ***Journal of Computer-Aided Molecular***, **27(8)**: 675-679.
3. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, (2021) Highly accurate protein structure prediction with AlphaFold, ***Nature***, **596(7873)**: 583-589.
4. Z. Wu, B. Ramsundar, E. N. Feinberg, Gomes J, Geniesse C, A. S. Pappu, K. Leswing, V. Pande, (2018), MoleculeNet: A benchmark for molecular machine learning, ***Chemical Science***, **9(2)**: 513-530.
5. A. Schneuing, Y. Du, C. Harri, K. Didi, A. Jamasb, I. Igashov, W. Du, C. Gomes, T. L. Blundell, P. Lio, M. Welling, M. Bronstein, B. Correia, (2024) Structure-based drug design with equivariant diffusion models, ***Nature Computational Science***, **4(2)**: 139-147.
6. G. Corso, H. Stärk, B. Jing, R. Barzilay, T. Jaakkola, (2023) DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. In: ***International Conference on Learning Representations***.
7. L. Zhang, H. Wang, X. Gao, (2023) Deep learning in computational chemistry: A review, ***Journal of Chemical Information and Modeling***, **63(14)**: 4348-4369.
8. F. Noe, A. Tkatchenko, K. R. Muller, C. Clementi, (2020) Machine learning for molecular simulation, ***Annual Review of Physical Chemistry***, **71:** 361-390.
9. P. Csermely, T. Korcsmaros, H. J. Kiss, G. London, R. Nussinov, (2013) Structure and dynamics of molecular networks: A novel paradigm of drug discovery, ***Pharmacology and Therapeutics***, **138(3)**, 333-408.
10. S. I. Berger, R. Iyengar, (2009) Network analyses in systems pharmacology, ***Bioinformatics***, **25(19)**: 2466-2472.
11. B. Fabian, T. Edlich, H. Gaspar, M. Segler, J. Meyers, M. Fiscato, M. Ahmed, (2023) Molecular representation learning with language models, ***Journal of Chemical Information and Modeling***, **63(21)**: 6692-6704.
12. R. Irwin, S. Dimitriadis, J. He, E. J. Bjerrum, (2022) Chemformer: A pre-trained transformer for computational chemistry, ***Machine Learning: Science and Technology***, **3(1)**: 015022.
13. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, …& D. Hassabis, (2021) Highly accurate protein structure prediction with AlphaFold, ***Nature,*** **596(7873)**: 583-589.
14. C, Edwards, T. Lai, K. Ros, G. Honke, H. Ji, (2024) MolReGPT: Leveraging large language models for molecular discovery, ***Journal of Chemical Information and Modeling***, **64(4)**: 1168-1178.
15. T. Blaschke, J. Arus-Pous, H. Chen, C. Margreitter, C. Tyrchan, O. Engkvist, K. Papadopoulos, A. Patronov. (2020) REINVENT 2.0: An AI tool for *de novo* drug design, ***Journal of Chemical***

Information and Modeling, **60(12)**: 5918-5922.
16. X. Tang, M. Ringner, E. Wideback, (2023) EarlGAN: An actor-critic RL agent-driven GAN for *de novo* molecular design, ***Journal of Cheminformatics***, **15(1)**: 45.
17. G. Schneider, (2018) Automating drug discovery. ***Nature Reviews Drug Discovery,*** **17(2)**: 97-113.
18. A. Zhavoronkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I.Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, (2019) Deep learning enables rapid identification of potent DDR1 kinase inhibitors, ***Nature Biotechnology,*** **37(9)**: 1038-1040.
19. M. J. Page, J. E. McKenzie, P. M. Bossuyt, (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, ***BMJ***, **372**: n71.
20. E. Gawehn, J. A. Hiss, G, (2016) Schneider. Deep learning in drug discovery, ***Molecular Informatics***, **35(1)**: 3-14.
21. K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen,… & R. Barzilay. Analyzing learned molecular representations for property prediction, ***Journal of Chemical Information and Modeling***, **59(8)**: 3370-3388.
22. B. Jing, G. Corso, J. Chang, R. Barzilay, T. Jaakkola, (2022) Torsional diffusion for molecular conformer generation, ***Advances in Neural Information Processing Systems***, **35:** 24240-24253.
23. G. Corso, H. Stärk, B. Jing, R. Barzilay, T. Jaakkola. (2022) ***DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking*** [arXiv Preprint].
24. A. Schneuing, Y. Du, C. Harris, Y. Du, K. Didi, A. Jamasb, I. Igashov, W. Du, C. Gomes, T. L. Blundell, P. Lio, M. Welling, M. Bronstein,… & B. Correia. Structure-based drug design with equivariant diffusion models, ***Nature Computational Science***, **4(2)**: 139-147.
25. W. Ahmad, E. Simon, S. Chithrananda, G. Grand, B. Ramsundar, (2022) ChemBERTa-2: Towards Chemical Foundation Models. In: ***AI for Accelerated Materials Design, Conference and Workshop on Neural Information Processing Systems***.
26. J. Ross, B. Belgodere, A. Vengudusamy, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido,… & A. Rives, (2022) Large-scale chemical language representations capture molecular structure and properties, ***Nature Machine Intelligence***, **4(12)**: 1256-1264.
27. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido,… & A. Rives, (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model, ***Science***, **379(6637)**: 1123-1130.
28. J. Abramson, J. Adler, J. Dunger, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C. C. Hung, M. O'Neill, (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3, ***Nature***, **630(8016)**: 493-500.
29. Z. Liu, Y. Li, Y. Cai, (2024) ***LlaSMol: Advancing Large Language Model for Chemistry*** [arXiv Preprint].
30. F. Xia, K. Sun, Y. Hou, (2024) CancerGPT: A foundation model for drug synergy prediction, ***Bioinformatics***, **40(3)**: btae120.
31. Z. Liu, Y. Li, Y. Cai, J. Jiang, (2023) MoleculeSTM: Multi-Modal Molecular Understanding. In: ***International Conference on Learning Representations***.
32. R. J. L. Townshend, S. Eismann, A. M. Watkins, R. Rangan, M. Karelina, R. Das, R. O. Dror, (2021) Geometric deep learning of RNA structure. ***Science,*** **373(6558)**: 1047-1051
33. E. Gawehn, J. A. Hiss, G. Schneider, (2016) Deep learning in drug discovery, ***Molecular Informatics,*** **35(1)**: 3-14.
34. V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, B. P. Feuston, (2003) Random forest: A classification and regression tool for compound classification and QSAR modeling. ***Journal of Chemical Information and Modeling***, **43(6)**: 1947-1958.
35. M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D. R. Koes, (2017) Protein-ligand scoring with convolutional neural networks, ***Journal of Chemical Information and Modeling***, **57(4)**: 942-957.
36. M. H. S. Segler, T. Kogej, C. Tyrchan, M. P. Waller, (2018) Generating focused molecule libraries for drug discovery with recurrent neural networks, ***ACS Central Science***, **4(1):** 120-131.
37. T. N. Kipf, M. Welling, (2017) ***Semi-Supervised Classification with Graph Convolutional Networks***, Brazil: ICLR.
38. P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, (2018) Graph Attention Networks. In: ***International Conference on Learning Representations***.
39. W. Zhu, Y. Zhang, D. Zhao, J. Xu, L. Wang, (2023) HiGNN: A hierarchical informative graph neural network for molecular property prediction equipped with feature-wise attention, ***Journal of Chemical Information and Modeling***, **63(1)**: 43-55.
40. Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, (2020) Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism, ***Journal of Medicinal Chemistry***, **63(16)**: 8749-8760.
41. Y. Rong, Y. Bian, T. Xu, W. Xie, Y. Wei, W. Huang, J. Huang, (2020) Self-Supervised Graph Transformer On Large-Scale Molecular Data. In: ***Conference and Workshop on Neural Information Processing Systems***.
42. S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou, H. Xiong, (202) Structure-aware Interactive Graph Neural Networks for the Prediction of Protein–Ligand Binding Affinity. In: ***Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21). Virtual Event,*** Singapore. p. 975-985.
43. S. Wang, Y. Guo, Y. Wang, H. Sun, J. Huang (2019) SMILES-BERT: Large Scale Unsupervised Pre-Training for Molecular Property Prediction. In: ***Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19)***. Niagara Falls, NY, USA. p. 429-436
44. M. Wen, J. Blitzer, S. Connelly, (2022) FP-BERT: A transformer-based composite fingerprint for molecular property prediction, ***Journal of Chemical Information and Modeling***, **62(14)**: 3257-3268.
45. Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. Dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido,… & A. Rives, (2023) Evolutionary-scale prediction of atomic-level protein structure with a language model, ***Science***, **379(6637)**: 1123-1230.
46. K. Jha, S. Saha, H. Singh, (2023) Graph-BERT for protein-protein interaction prediction, ***Briefings in Bioinformatics***, **24(3)**: bbad132.
47. M. Popova, O. Isayev, A. Tropsha, (2018) Deep reinforcement

learning for *de novo* drug design, ***Science Advances***, **4(7)**: eaap7885.
48. Q. Wang, X. Hu, H. Liu, (2021) Deep Reinforcement Learning and Docking Simulations for Autonomous Molecule Generation. In: ***ACM Multimedia Asia***, Broadway: Association for Computing Machinery.
49. X. Liu, K. Ye, H. W. T. Van Vlijmen, M. T. M. Emmerich, A. P. IJzerman, G. J. P. V. Westen, (2021) DrugEx v2: *De novo* design of drug molecules by Pareto-based multi-objective reinforcement learning in polypharmacology, ***Journal of Cheminformatics***, **13(1)**: 85.
50. M. B. Sampa and N. H. A. Aziz, (2024) Mol-AIR: Molecular reinforcement learning with adaptive intrinsic rewards for goal-directed molecular generation, ***Journal of Chemical Information and Modeling***, **64(23)**: 9340-9352.
51. Y. Jing, Y. Bian, Z. Hu, L. Wang, X. Q. Xie, (2018) Deep learning for drug design: An artificial intelligence paradigm for drug discovery in the big data era, ***AAPS Journal***, **20(3)**: 58.
52. J. Guan, W. W. Qian, X. Peng, (2023) 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. In: ***International Conference on Learning Representations***.
53. M. Xu, S. Luo, Y. Bengio, J. Peng, J. Tang, (2022) Learning Geo-Aware Representation for Molecular Conformation Generation. In: ***International Conference on Learning Representations***.
54. X. Peng, J. Guan, J. Peng, J. Ma, (2023) Pocket-Specific 3D Molecule Generation by Fragment-Based Autoregressive Diffusion Models. In: ***International Conference on Learning Representations***.
55. L. Huang, T. Xu, Y. Yu, P. Zhao, X. Chen, J. Han, Z. Xie, H. Li, W. Zhong, K. C. Wong, H. A. Zhang, (2024) A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets, ***Nature Communications***, **15(1)**: 2656.
56. A. T. McNutt, P. Francoeur, R. Aggarwal, T. Masuda, R. Meli, M. Ragoza, J. Sunseri, D. R. Koes, (2021) GNINA 1.0: Molecular docking with deep learning, ***Journal of Cheminformatics***, **13(1)**: 43.
57. O. Mendez-Lucio, M. Ahmad, E. A. Del Rio-Chanona, J. K. Wegner, (2021) A geometric deep learning approach to predict binding conformations of bioactive molecules, ***Nature Machine Intelligence***, **3(12)**: 1033-1039.
58. H. Stark, O. Ganea, L. Pattanaik, R. Barzilay, T. Jaakkola, (2022) EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction. In: ***The International Conference on Machine Learning***.
59. W. Lu, Q. Wu, J. Zhang, J. Rao, C. Li, S. Zheng, (2022) TankBind: Trigonometry-Aware Neural Networks for Drug-Protein Binding Structure Prediction. In: ***Conference and Workshop on Neural Information Processing Systems***.
60. C. Lu, Y. Liu, X. Zhang, (2024) DynamicBind: Predicting ligand-specific protein-ligand complex structure with a deep equivariant generative model, ***Nature Communications***, **15(1)**: 1071.
61. L. Zhang, J. Han, H. Wang, R. Car, E. Weinan, (2018) Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, ***Physical Review Letters***, **120(14)**: 143001.
62. S. A. Hollingsworth, R. O. Dror, (2018) Molecular dynamics simulation for all, ***Neuron,*** **99(6)**: 1129-1143.
63. A. Hospital, J. R. Goñi, M. Orozco, J. L. Gelpí, (2015) Molecular dynamics simulations: Advances and applications, ***Advances and Applications in Bioinformatics and Chemistry,*** **8:** 37-47.
64. Y. Wang, J. M. L. Ribeiro, P. Tiwary, (2020) Machine learning approaches for analyzing and enhancing molecular dynamics simulations, ***Current Opinion in Structural Biology***, **61:** 139-145.
65. L. Bonati, G. Piccini, M. Parrinello, (2021) Deep learning the slow modes for rare events sampling, ***Proceedings of the National Academy of Sciences of the United States of America***, **118(44)**: e2113533118.
66. M. Invernizzi, P. M. Piaggi, M. Parrinello, (2020) Unified approach to enhanced sampling, ***Physical Review X***, **10(4)**: 041034.
67. A. Mayr, G. Klambauer, T. Unterthiner, S. Hochreiter, (2016) DeepTox: Toxicity prediction using deep learning, ***Frontiers in Environmental Science***, **3:** 80.
68. X. Li, D, (2020) Fourches. Inductive transfer learning for molecular activity prediction: Next-Gen QSAR Models with MolPMoFiT, ***Journal of Cheminformatics***, **12(1)**: 27.
69. V. B. Siramshetty, P. Shah, E. H. Kerns, D. T. Nguyen, (2024) Leveraging automated machine learning for predicting pharmaceutical properties, ***Molecular Pharmaceutics***, **21(11)**: 5432-5445.
70. J. Li, A. Fu, L. Zhang, (2019) An overview of scoring functions used for protein-ligand interactions in molecular docking, ***Interdisciplinary Sciences***, **11(2)**: 320-328.
71. A. Tropsha, (2010) Best practices for QSAR model development, validation, and exploitation, ***Molecular Informatics***, **29(6-7)**: 476-488.
72. P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee, (2019) Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction, ***ACS Central Science***, **5(9)**: 1572-1583.
73. P. Schwaller, T. Laino, T. Gaudin, (2021) Mapping the space of chemical reactions using attention-based neural networks, ***Nature Machine Intelligence***, **3(2)**: 144-152.
74. P. Schneider, W. P. Walters, A. T. Plowright, (2020) Rethinking drug design in the artificial intelligence era, ***Nature Reviews Drug Discovery***, **19(5)**: 353-364.
75. Y. Matsuzaka, A. Uesugi, (2023) Machine learning for classification of protein-ligand binding affinity, ***Molecules***, **28(4)**: 1943.
76. H. Zhang, S. Li, J. Zhang, (2022) DFCNN: A deep learning approach for compound screening, ***Bioinformatics***, **38(8)**: 2256-2262.
77. F. Gentile, V. Agrawal, M. Hsing, (2022) Deep docking: A deep learning platform to augment docking in the early stages of drug discovery, ***Journal of Chemical Information and Modeling***, **62(15)**: 3566-3582.
78. L. Huang, T. Xu, Y. Yu, (2024) DecompOpt: Controllable and decomposed diffusion models for structure-Based Molecular Optimization. In: ***International Conference on Learning Representations***.
79. S. Dara, S. Dhamercherla, S. S. Jadav, C. M. Babu, M. J. Ahsan, (2022) Machine learning in drug discovery: A review, ***Artificial Intelligence Review***, **55(3)**: 1947-1999.
80. B. R. Beck, B. Shin, Y. Choi, S. Park, S. Kang, (2020) Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model, ***Computational and Structural Biotechnology Journal***, **18:** 784-790.
81. T. Kalliokoski, (2021) HASTEN: Accelerating virtual screening with neural networks, ***Journal of Chemical Information and Modeling***, **61(7)**: 3899-3905.
82. K. Huang, C. Xiao, T. N. Hoang, L. M. Glass, J. Sun, (2021)

CASTLE: Citation network in automatic structure-learning and analysis for gene expression data, ***Bioinformatics***, **37(14)**: 1990-1997.
83. K. Y. Gao, A. Fokoue, H. Luo, A. Iyengar, S. Dey, P. Zhang, (2018) Interpretable Drug Target Prediction using Deep Neural Representation. In: ***International Joint Conference on Artificial Intelligence***.
84. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson. DrugBank 5.0: A major update to the DrugBank database for 2018, ***Nucleic Acids Research***, **46(D1)**: D1074-D1082.
85. D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, (2019) ChEMBL: Towards direct deposition of bioassay data, ***Nucleic Acids Research***, **47(D1)**: D930-D940.
86. M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, J. Chong, (2016) BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology, ***Nucleic Acids Research***, **44(D1)**: D1045-D1053.
87. M. M. Stepniewska-Dziubinska, P. Zielenkiewicz, P. Siedlecki, (2018) Development and evaluation of a deep learning model for protein-ligand binding affinity prediction, ***Bioinformatics***, **34(21)**: 3666-3674.
88. J. Jimenez, M. Skalic, G. Martinez-Rosell, G. De Fabritiis, (2018) KDEEP: Protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks, ***Journal of Chemical Information and Modeling***, **58(2)**: 287-296.
89. S. Ekins, A. C. Puhl, K. M. Zorn, T. R. Lane, D. P. Russo, J. J. Klein, A. J. Hickey, A. M. Clark, (2019) Exploiting machine learning for end-to-end drug discovery and development, ***Nature Materials,*** **18(5)**: 435-441.
90. R. Santagati, A. Aspuru-Guzik, R. Babbush, M. Degroote, L. González, E. Kyoseva, N. Moll, M. Oppel, R. M. Parrish, N. C. Rubin, M. Streif, C. S. Tautermann, H. Weiss, N. Wiebe, C. Utschig-Utschig, (2024) Drug design on quantum computers, ***Nature Physics***, **20(4)**: 549-557.
91. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, (2018) Machine learning for molecular and materials science, ***Nature,*** **559(7715)**: 547-555.
92. S. Mensa, E. Sahin, P. K. Barkoutsos, I. Tavernelli, (2023) Quantum machine learning framework for virtual screening in drug discovery: A prospective quantum advantage, ***Machine Learning: Science and Technology***, **4(1)**: 015023.
93. X. Li, Z. Li, S. Wu, *et al*, (2024) Deep learning enhancing computational chemistry and drug discovery: A comprehensive review, ***Journal of Chemical Information and Modeling***, **64(8)**: 3165-3188.
94. C. N. Cavasotto, J. I. Di Filippo, (2021) Artificial intelligence in the early stages of drug discovery, ***Archives of Biochemistry and Biophysics***, **698:** 108730.
95. Y. Jing, Y. Gao, J. Pei, (2023) Drug discovery of the future: Harnessing the power of AI in small molecule design. ***Expert Opinion on Drug Discovery***, **18(8)**: 857-871.
96. J. Xia, Y. Q. Zhu, Y. Q. Du, (2023) A Systematic Survey of Chemical Pre-Trained Models. In: ***International Joint Conference on Artificial Intelligence***.
97. J. Deng, Z. Yang, L. Ojha, (2023) Artificial intelligence in drug discovery: Applications, challenges, and future directions. ***Drug Discovery Today***, **28(4)**: 103559.
98. X. Yang, Y. Wang, R. Byrne, G. Schneider, S. Yang, (2019) Concepts of artificial intelligence for computer-assisted drug discovery, ***Chemical Reviews***, **119(18)**: 10520-10594.
99. H. Chen, O. Engkvist, (2024) Transformer technology in molecular science, ***WIREs Computational Molecular Science***, **14(2)**: e1725.
100. R. Rodriguez-Perez, J. Bajorath, (2020) Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley value analysis, ***Journal of Cheminformatics***, **12(1)**: 1-10.
101. Y. C. Lo, S. E. Rensi, W. Torng, R. B. Altman, (2018) Machine learning in chemoinformatics and drug discovery, ***Drug Discovery Today***, **23(8)**: 1538-1546.
102. K. K. Mak, M. R. Pichika, (2019) Artificial intelligence in drug development: present status and future prospects, ***Drug Discovery Today***, **24(3)**: 773-780.
103. J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, S. Zhao, (2019) Applications of machine learning in drug discovery and development, ***Nature Reviews Drug Discovery***, **18(6)**: 463-477.
104. H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, (2018) The rise of deep learning in drug discovery, ***Drug Discovery Today***, **23(6)**: 1241-1250.
105. A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan,… & P. Kohli. Improved protein structure prediction using potentials from deep learning, ***Nature***, **577(7792)**: 706-710.
106. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, (2022) AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models, ***Nucleic Acids Research***, **50(D1)**: D439-D444.
107. F. Ren, E. Aliper, J. Chen, H. Zhao, S. Rao, C. Kuppe, (2024) A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. ***Nature Biotechnology***, **42(4)**: 616-627.
108. J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, (2020) A deep learning approach to antibiotic discovery, ***Cell***, **180(4)**: 688-702.
109. D. A. Boiko, R. MacKnight, G. Gomes. [2023] ***Emergent Autonomous Scientific Research Capabilities of Large Language Models***, [arXiv Preprint].
110. A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, P. Schwaller, (2023) ***ChemCrow: Augmenting Large-Language Models with chemistry Tools***, [arXiv Preprint].
111. A. Pyrkov, A. Aliper, D. Bezrukov, Y. C. Lin, D. Polykovskiy, P. Kamya, F. Ren, A. Zhavoronkov, (2023) Quantum computing for near-term applications in generative chemistry and drug discovery, ***Drug Discov Today***, **28(8)**: 103675.
112. M. Sajjan, J. Li, R. Selvarajan, S. H. Sureshbabu, S. S. Kale, R. Gupta, V. Singh, S. Kais, (2022) Quantum machine learning for chemistry and physics, ***Chemical Society Reviews***, **51(15)**: 6475-6573.
113. C. Cerchia, A. Lavecchia, (2023) New avenues in artificial-intelligence-assisted drug discovery, ***Drug Discovery Today***, **28(4)**: 103516.

114. C. Edwards, T. Lai, K. Ros, *et al*, (2024) MolReGPT: Leveraging large language models for molecular discovery, ***Journal of Chemical Information and Modeling***, **64(4)**: 1168-1178.
115. L. Wang, Y. Liu, S. Zhong, (2020) Artificial intelligence for COVID-19 drug discovery and vaccine development, ***Frontiers in Artificial Intelligence***, **3:** 581504.
116. A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, A. Zhavoronkov, (2016) Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data, ***Molecular Pharmaceutics***, **13(7)**: 2524-2530.
117. R. Liu, L. Wei, P. Zhang, (2020) A deep learning-based framework for drug-target interaction prediction. ***Scientific Reports***, **10(1)**: 20844.
118. D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, R. K. Tekade, (2021) Artificial intelligence in drug discovery and development, ***Drug Discovery Today,*** **26(1)**: 80-93.
119. M. K. P. Jayatunga, W. Xie, L. Ruder, U. Schulze, C. Meier, (2022) AI in small molecule drug discovery: A coming wave. ***Nature Reviews Drug Discovery***, **21(3)**: 175-176.
120. A. Bender, I. Cortes-Ciriano, (2021) Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: A discussion of chemical and biological data, ***Drug Discovery Today***, **26(4)**: 1040-1052.
121. K. D. Gao, N. Murgai, (2024) How to succeed in a machine learning/data science-driven drug discovery project. ***Drug Discovery Today***, **29(3)**: 103938.
122. E. N. Feinberg, D. Sur, Z. Wu, B. E. Husic, H. Mai, Y. Li, S. Sun, J. Yang, B. Ramsundar, V. S. Pande, (2018) PotentialNet for molecular property prediction, ***ACS Central Science***, **4(11)**: 1520-1530.
123. B. Ramsundar, P. Eastman, P. Walters, V. Pande, (2019) ***Deep Learning for the Life Sciences***, Sebastopol: O'Reilly Media.
124. A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni,… & C. Yang. QSAR modeling: Where have you been? Where are you going to? ***Journal of Medicinal Chemistry***, **57(12)**: 4977-5010.
125. J. B. O. Mitchell, (2014) Machine learning methods in chemoinformatics, ***Journal of Cheminformatics***, **6(1)**: 1-16.
126. L. K. Vora, A. D. Gholap, K. Jetha, R. R. S. Thakur, H. K. Solanki, V. P. Chavda, (2023) Artificial intelligence in pharmaceutical technology and drug delivery design, ***Pharmaceutics,*** **15(7)**: 1916.
127. R. Burbidge, M. Trotter, B. Buxton, S. Holden, (2001) Drug design by machine learning: Support vector machines for pharmaceutical data analysis, ***Computational Chemistry***, **26(1)**: 5-14.
128. E. N. Muratov, J. Bajorath, R. P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, T. I. Oprea, I. I. Baskin, A. Varnek, A. Roitberg, O. Isayev, S. Curtarolo, D. Fourches, Y. Cohen,… & A. Aspuru-Guzik, (2020) QSAR without borders, ***Chemical Society Reviews***, **49(11)**: 3525-3564.