AI IN CHEMISTRY: DRUG DISCOVERY, REACTION PREDICTION, AND MOLECULAR SIMULATION

Mahinder Singh
AP Chemistry, VGC Mandi (H.P)
mah.snr@gmail.com

Abstract
Artificial Intelligence (AI) is profoundly transforming modern chemistry, particularly in drug discovery, chemical reaction prediction, and molecular simulation. In drug discovery, AI has accelerated hit identification, virtual screening, molecular generation, and prediction of ADMET properties via models like Graph Neural Networks (GNNs), generative adversarial networks (GANs), and reinforcement learning (RL) frameworks. For reaction prediction, sequence-to-sequence models such as the Molecular Transformer excel at forward and retrosynthesis planning, often achieving >90 % accuracy and even yield prediction. In molecular simulation, AI—exemplified by DeepMind's AlphaFold—has revolutionized structure prediction and docking, while ML interatomic potentials have begun integrating high-order data (e.g., Hessians) to enhance dynamics modeling . Despite remarkable progress, challenges persist in interpretability, data quality, experimental validation, and the long-standing gap between computational discovery and clinical success. This paper provides a comprehensive review and outlook on AI's pivotal role across these domains.

1.Introduction:

AI has increasingly become a driver of innovation in chemistry, enabling predictive modeling and automation across multiple domains. Modern deep learning, generative models, and graph-based architectures empower chemists and pharmacologists to navigate vast chemical spaces, plan synthetic routes, and simulate molecular interactions with unprecedented speed and precision.

2. AI in Drug Discovery:

2.1 Virtual Screening & Property Prediction: Traditional QSAR models leveraged ML (e.g., decision trees, SVMs, random forests), achieving accuracies above 85 % for binder/non-binder classification.Platforms like DeepChem offer open-source frameworks with datasets (e.g., MoleculeNet, millions of compounds), enabling virtual screening, lead optimization, and predictive modeling of drug properties.

2.2 Generative Models for De Novo Design: ORGANIC combines GANs and RL to generate molecules with desired pharmacological properties. Chemistry42, developed by Insilico Medicine, automates de novo molecular design within an AI-integrated drug discovery suite.Other generative approaches, including RANC, ReLeaSE, and RNN-based strategies, target specific therapeutic areas (e.g., Staphylococcus aureus, dopamine receptors), showing promising activity.

2.3 Graph Neural Networks (GNNs): GNNs analyze molecular topology directly and have become powerful in molecular property prediction, virtual screening, generation, and synthesis planning. Recent reviews highlight techniques such as geometric GNNs, uncertainty quantification, and self-supervised learning in AIDD (AI-aided drug discovery).

2.4 ADME–Tox & Safety Prediction: AI models—like multitask GNNs and integrated platforms such as CLARITY_{PV}—assist in early prediction of ADME (Absorption, Distribution, Metabolism, and Excretion) and toxicity, reducing late-stage attrition .

2.5 Drug Repurposing & Drug–Target Interaction:Tools such as DeepDTI, DeepConv-DTI, and Ligand Express predict novel drug–target interactions, thereby facilitating drug repurposing .

2.6 Industrial Integration & Clinical Outlook: Industries like Recursion and Insilico are advancing AI-derived candidates into clinical trials—though none have yet reached commercialization—highlighting both promise and caution in translation.

3. AI for Reaction Prediction and Synthesis Planning:

3.1 Forward Reaction Prediction: Sequence-to-sequence models and graph-based convolutional neural networks predict reaction outcomes—typically from reactant SMILES inputs—with accuracy around 80–90 % without template dependence.

3.2 Retrosynthesis & Route Planning: IBM's RXN for Chemistry, using Molecular Transformer models trained on millions of reactions, predicts retrosynthetic pathways. Hybrid models further improve predictions by combining data-driven models with chemical knowledge.

3.3 Yield and Classification Prediction: Transformer models can predict reaction yields and classify reaction types using SMILES-based

regression; k-NN fingerprint models can sometimes match them with lower complexity.

3.4 Autonomous Synthesis & Multi-Objective Optimization:AI-enabled systems are beginning to integrate synthesis planning with automated laboratory experimentation, optimizing for yield, cost, sustainability, and scalability in a closed loop.

4. AI-Assisted Molecular Simulation:

4.1 Protein Structure Prediction & Docking: DeepMind's AlphaFold uses attention-based deep learning to accurately predict protein structures and is now expanding to DNA, RNA, and ligand interactions with transformative potential for drug development.

4.2 Molecular Docking: Flexible docking tools like FlexAID outperform traditional methods (e.g., AutoDock Vina) by modeling ligand and side-chain flexibility, enhancing pose accuracy in structure-based drug design.

4.3 ML-Enhanced Molecular Dynamics: Inclusion of Hessian (second derivative) data in training machine learning interatomic potentials significantly improves predictions of energies, forces, and reaction paths, enabling more accurate dynamics simulations. AI also enhances simulation of binding affinities and kinetics through differentiable modeling .

5. Challenges and Future Directions:

Data Quality & Interpretability: Many AI models are "black boxes," and their predictions require rigorous validation.Experimental Validation Gap: To date, AI-generated drug candidates have yet to reach the market—clinical translation remains uncertain .Multi-Parameter Decision Making: Holistic design must consider cost, yield, sustainability, and scalability—not just chemical reactivity .Human Collaboration: Creative synergy between AI tools and expert chemists is essential for designing feasible molecules and reactions .Autonomous Labs & Digital Twins: Advancements in AI-driven, automated experimentation and virtual "digital twin" systems promise accelerated discovery while mitigating resource constraints .

6. Conclusion:

AI is catalyzing a paradigm shift in chemistry—transforming drug discovery, reaction planning, and molecular simulation via powerful predictive and generative models. While breakthroughs like AlphaFold and GNNs showcase AI's capabilities, substantial challenges remain in bridging computational predictions with real-world outcomes. The future lies in integrating AI with experimental validation, interpretability, and human creativity, fostering a collaborative ecosystem toward faster, safer, and more effective chemical innovation.

References:

1. Wallach, I., Dzamba, M., & Heifets, A. (2015). AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. arXiv preprint.
2. Schwaller, P. et al. (2018). Molecular Transformer – A Model for Uncertainty-Calibrated Chemical Reaction Prediction. arXiv preprint.
3. Zhang, O. et al. (2025). Graph Neural Networks in Modern AI-aided Drug Discovery. arXiv preprint.
4. Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine. MDPI.
5. The Role of AI in Drug Discovery – Abbas (2024). ChemBioChem .
6. AI in drug discovery and its clinical relevance: Heliyon .
7. Chemoinformatics and AI colloquium: progress and challenges. Journal of Cheminformatics
8. RSC "physics-AI dialogue in drug design" (2025) RSC Medicinal Chemistry.
9. Artificial intelligence in drug discovery – Discover AI (2025).
10. Virtual screening – Wikipedia.
11. Forward reaction prediction review – MDPI.
12. AI-assisted reaction prediction and synthesis – ScienceDirect.
13. Autonomous synthesis and retrosynthesis trends – RSC Chemical Communications.
14. AlphaFold – Wikipedia.
15. Nobel Prize coverage of AI in protein folding – Financial Times (2024).
16. DeepMind announces AlphaFold 3 – Time.
17. AI-designed drugs entering clinical trials – Wired (2025).
18. FlexAID docking tool – Wikipedia.
19. Hessian-augmented ML interatomic potentials – Wikipedia (Mendoza-Cortés) (2025) .
20. Computational simulations enhanced by AI – Patsnap Synapse.