Simen Owen Academic Proceedings Series

3rd International Conference on Electronics, Engineering, Computer Science and Applied Development (EESD 2026)

Review

Generative Artificial Intelligence Driven Innovation in Drug Molecule Design: Advances and Future Directions

Ruonan Wang [1,*]

[1] International School of Pharmacy and Business, China Pharmaceutical University, Nanjing, China
* Correspondence: Ruonan Wang, International School of Pharmacy and Business, China Pharmaceutical University, Nanjing, China

Abstract: The application of generative artificial intelligence (AI) has fundamentally revolutionized the early stages of modern drug discovery by enabling the rapid de novo design of novel drug-like molecules with highly specific and desired pharmacological properties. This comprehensive review systematically examines recent and significant advances in various generative models, including variational autoencoders, generative adversarial networks, autoregressive models, and state-of-the-art diffusion models. Particular emphasis is placed on their seamless integration with advanced molecular representation learning techniques and multi-objective optimization frameworks. Furthermore, key breakthroughs in generating synthetically accessible, target-specific, and pharmacokinetically favorable compounds are critically highlighted and evaluated. We also discuss emerging and transformative trends within the field, such as the deployment of large-scale pre-trained molecular language models, the utilization of reinforcement learning derived from direct chemical feedback, and the implementation of closed-loop wet-lab validation systems. Despite the significant and undeniable progress achieved thus far, substantial challenges remain prevalent in critical areas such as training data quality, practical synthetic feasibility, overall molecular diversity, and algorithmic interpretability. Finally, we outline strategic future directions toward the development of fully autonomous generative design platforms and their real-time integration with high-throughput experimentation workflows. Ultimately, these continuous advancements aim to significantly accelerate the critical transition from AI-generated molecular structures to safe, effective, and clinically viable drug candidates for complex diseases.

Keywords: generative ai; drug design; de novo generation; deep learning; computational chemistry

Received: 27 March 2026
Revised: 15 May 2026
Accepted: 29 May 2026
Published: 05 June 2026

1. Introduction

1.1. Challenges in Traditional Drug Discovery

The conventional drug discovery process is widely recognized for being exceptionally lengthy, costly, and prone to high failure rates. Traditional methodologies, such as high throughput screening and natural product extraction, often yield low success rates due to limited chemical diversity and inefficiencies in identifying viable drug candidates. Despite advancements in technology, persistent challenges remain, including poor pharmacokinetic profiles of potential compounds and the emergence of unexpected toxicity during later stages of development. Furthermore, the escalating costs of research and development, coupled with increasingly stringent regulatory requirements, have compounded these difficulties. The growing complexity of therapeutic targets, driven by advancements in understanding molecular biology and disease mechanisms, has further underscored the limitations of conventional approaches. These enduring obstacles highlight the urgent need for innovative and intelligent molecular design methodologies

that can streamline the drug discovery process, enhance efficiency, and improve the likelihood of success in developing effective therapeutic agents.

1.2. Rise of Generative AI in Molecular Design

In response to the limitations of traditional methods, generative artificial intelligence has emerged as a transformative force in drug discovery. Recent advancements in generative AI frameworks have demonstrated their ability to create novel molecular structures with optimized properties, significantly reducing the time and cost associated with hit identification and lead optimization. These models, including variational autoencoders and generative adversarial networks, have been refined to enhance molecule design by learning complex chemical distributions and producing compounds that are synthetically accessible [1, 2]. This approach represents a paradigm shift from the passive screening of existing chemical libraries to the active generation of drug-like molecules tailored to specific therapeutic needs. By leveraging these advanced algorithms, researchers can explore vast chemical spaces more efficiently, identifying potential candidates with improved pharmacological profiles. Such innovations are paving the way for a more dynamic and efficient drug discovery process, ultimately accelerating the development of new treatments for various diseases [3].

1.3. Scope and Structure of This Review

This review provides a comprehensive and systematic examination of the role of generative AI in driving innovation within the field of drug molecule design. The structure of this review is carefully organized to facilitate a clear understanding of the topic. Section 2 delves into the foundational principles and methodologies underlying generative models, offering insights into their theoretical and practical frameworks. Section 3 explores molecular representations and the diverse learning strategies employed to enhance the design process. Section 4 focuses on significant advancements, including multi-objective generation, target-specific molecule design, synthetic accessibility considerations, and the integration of closed-loop validation techniques. Section 5 addresses the major challenges and unresolved issues that persist in this domain. Section 6 outlines potential future directions for research and development, while Section 7 provides a concise conclusion [4]. Figure 1 illustrates the general pipeline of generative AI for drug molecule design, serving as a visual guide to the discussed concepts.

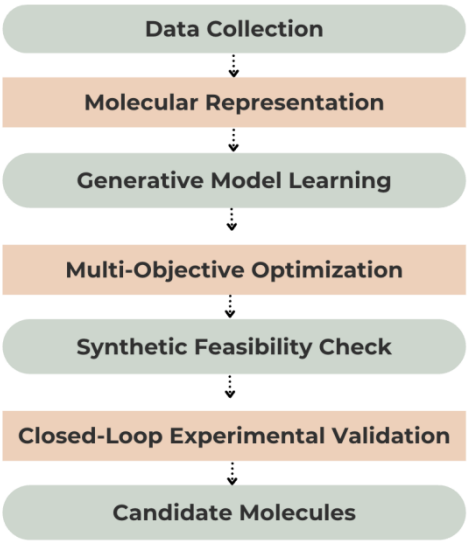

Figure 1. Pipeline of Generative AI-Driven Drug Molecule Design

2. Foundational Generative Models for Drug Design

2.1. Variational Autoencoders (VAEs) and Latent Space Exploration

Variational autoencoders have emerged as a fundamental tool in molecular generation, offering the ability to learn continuous latent representations of discrete chemical structures. These models enable the exploration of chemical space in a more structured and efficient manner. A geometry-based molecular generation framework has demonstrated the capability to effectively capture three-dimensional molecular geometries while adhering to chemical valence constraints, ensuring the generated molecules are both realistic and chemically valid. Additionally, scaffold-based approaches have been developed to enhance the controllability and interpretability of molecule design. For instance, a two-step molecule generation process integrates a variational autoencoder with a transformer model. This method first generates molecular scaffolds, which serve as the core structural framework, and subsequently decorates them with substituents. This layered approach allows for a more targeted exploration of chemical diversity, facilitating the design of molecules with specific desired properties [5].

2.2. Generative Adversarial Networks (GANs) for De Novo Generation

Generative adversarial networks represent a transformative approach for de novo molecular generation by employing a competitive framework between a generator and a discriminator [2, 6]. This dynamic interaction enables the creation of highly realistic molecular structures. Recent advancements have demonstrated the potential of these networks in designing peptides and proteins with specific structural and functional attributes. Furthermore, innovative methodologies, such as graph transformer-based generative adversarial networks, have been developed to enhance the precision of target-specific drug design. These approaches have shown remarkable success in generating candidate molecules with optimized binding affinities to predefined protein targets, thereby advancing the field of computational drug discovery [7]. The integration of such advanced algorithms into molecular design workflows holds significant promise for accelerating the identification of novel therapeutic compounds with desired properties.

2.3. Autoregressive Models (RNNs, Transformers)

Autoregressive models are designed to generate molecules incrementally, treating molecular strings or graphs as sequences of tokens. These models leverage advanced architectures to enable the stepwise construction of molecular structures, ensuring that each step is informed by the preceding context. For instance, one approach utilizes a conditional generative pre-trained transformer to produce SMILES strings that are specifically conditioned on protein target information, thereby tailoring molecular generation to specific biological contexts [8]. Another innovative method combines reinforcement learning with a transformer-based generative adversarial network (GAN) hybrid model. This integration allows for the sequential generation capabilities of transformers to be enhanced by adversarial training, optimizing multiple molecular properties simultaneously. Such advancements highlight the potential of autoregressive models in addressing complex challenges in molecular design and property optimization.

2.4. Diffusion Models for Molecular Graphs and 3D Structures

Diffusion models have recently gained prominence as robust generative frameworks capable of transforming random noise into coherent and valid molecular structures [9]. These models operate by iteratively refining noisy inputs, ultimately producing outputs that adhere to specific chemical and structural constraints. A notable advancement in this domain involves the development of conditional diffusion models tailored for discrete graph structures, which enable the generation of diverse molecular graphs while maintaining adherence to predefined chemical rules. Additionally, the integration of transformer architectures with diffusion processes has introduced a new level of precision and control. This approach facilitates multi-conditional molecular generation, allowing for fine-tuned manipulation of molecular properties and structural features. Such advancements underscore the potential of diffusion models in addressing complex challenges in molecular design. A detailed comparison of the performance of mainstream

generative models is provided in Table 1, highlighting their relative strengths and applications.

Table 1. Performance Comparison of Generative Models

Model	Strength	Limitation	Typical Use
VAE	Stable training, interpretable latent space	Limited molecular diversity	Scaffold-based generation
GAN	High structural realism	Training instability	Target-specific generation
Autoregressive	Strong sequence modeling	Slow generation	SMILES-based design
Diffusion	High validity & diversity	High computation cost	3D molecular generation

3. Molecular Representations and Learning Strategies

3.1. Smiles Strings: Strengths and Limitations

SMILES (Simplified Molecular Input Line Entry System) strings offer a text-based approach to representing molecular structures, making them a widely utilized tool in generative molecular design. This format enables sequence-based generation, which aligns well with the capabilities of modern language models. However, SMILES strings exhibit certain limitations that can hinder their effectiveness in capturing the full complexity of molecular structures. For instance, their linear string format, while convenient, struggles to accurately represent intricate molecular topologies and three-dimensional spatial arrangements. Additionally, SMILES strings are syntactically vulnerable, meaning that even minor errors in the sequence can render the representation invalid [4, 10]. These challenges have prompted comparisons with alternative encoding schemes that aim to address these shortcomings [7, 11]. Despite these limitations, SMILES remains a foundational tool in molecular representation, particularly in applications where simplicity and compatibility with machine learning models are prioritized. Further advancements in encoding methods may help overcome these inherent constraints.

3.2. Molecular Graphs and Graph Neural Networks (GNNs)

Molecular graphs, where atoms serve as nodes and bonds act as edges, provide a highly intuitive and detailed representation of chemical structures. These graphs enable the modeling of molecular interactions and properties in a manner that aligns closely with their physical and chemical behavior [12, 13]. Graph neural networks (GNNs) have emerged as powerful tools for analyzing molecular graphs, leveraging techniques such as message passing and neighborhood aggregation to learn intricate molecular representations. However, applying GNNs to large-scale molecular datasets presents significant computational challenges. Strategies for efficient training and inference have been proposed to address these issues, ensuring scalability and robustness in handling extensive datasets [14, 15]. This approach facilitates advancements in computational chemistry and drug discovery, where accurate molecular modeling is crucial.

3.3. 3D Representations and Geometric Deep Learning

Three-dimensional molecular representations are crucial for capturing the spatial arrangements and conformational details necessary for structure-based drug design. These representations provide a comprehensive understanding of molecular geometry, enabling researchers to analyze how molecules interact within three-dimensional space. Advanced methods in geometric deep learning have been developed to address these challenges, including architectures that inherently respect rotational and translational symmetries [16]. Such approaches are particularly effective in predicting binding affinities

and estimating poses for protein-ligand complexes, which are critical tasks in drug discovery. By leveraging three-dimensional data, these methods enhance the accuracy and reliability of computational predictions, offering significant improvements over traditional two-dimensional approaches. This progress underscores the transformative potential of geometric deep learning in modern pharmaceutical research.

3.4. Pre-Trained Molecular Language Models

Pre-trained language models have been adapted from natural language processing to learn rich molecular representations from large unlabeled chemical datasets. These models leverage advanced architectures, such as BERT and GPT variants, to capture chemical semantics and structural patterns effectively. One innovative approach involves random functional group masking during pre-training, which strategically obscures chemical functional groups to enhance the model's ability to learn chemically meaningful representations [17]. This technique has been shown to improve performance in downstream tasks, such as molecular property prediction and drug discovery. By focusing on the structural and functional aspects of molecules, these models provide a robust framework for analyzing complex chemical data. Table 2 compares four widely used molecular representations, highlighting their strengths and applications in computational chemistry and molecular design.

Table 2. Comparison of Molecular Representations

Representation	Encoding Type	Merit	Demerit
SMILES	Sequential text	Simple, compatible with LLMs	Syntax sensitive
Molecular Graph	Topological structure	Accurate topology	Complex computation
3D Geometry	Spatial coordinates	Binding-aware	Demands 3D data
Pre-trained LLM	Semantic embedding	Transferable knowledge	Data-hungry

4. Key Advances in Generative AI Driven Drug Discovery

4.1. Multi Objective Molecular Generation

Multi-objective molecular generation focuses on optimizing multiple desired properties of molecules simultaneously, such as bioactivity, solubility, and toxicity. This approach is particularly valuable in drug discovery, where balancing these often-conflicting properties is critical for developing effective and safe pharmaceuticals. Advanced methodologies, such as those leveraging Pareto algorithms and Monte Carlo tree search, have been employed to explore chemical space efficiently. These techniques enable researchers to identify molecules that achieve an optimal balance between competing objectives. Additionally, reinforcement learning methods have been utilized to enhance the generation of diverse molecular structures that meet multiple drug-like criteria [18]. Such advancements significantly improve the ability to design molecules with tailored properties, accelerating the drug discovery process and expanding the potential for innovative therapeutic solutions.

4.2. Target Specific and Structure Based Design

Target-specific molecular generation emphasizes the creation of compounds that exhibit strong binding affinity to designated protein targets. This approach leverages advanced computational techniques to design molecules that interact optimally with specific biological structures. A notable development in this field involves deep generative models tailored for structure-based drug design, which integrate three-dimensional protein-ligand interaction data to produce highly selective molecules. By incorporating detailed protein structural information, researchers can achieve rational

molecular designs that enhance target specificity and therapeutic potential [19]. Such methodologies represent a significant advancement in drug discovery, enabling the generation of compounds with improved efficacy and reduced off-target effects. This integration of structural data into molecular design underscores the importance of precision in modern pharmaceutical research.

4.3. Synthetic Accessibility and Reaction Aware Generation

Ensuring that AI-generated molecules can be feasibly synthesized in the laboratory remains a critical challenge in the field of drug discovery. One approach involves integrating synthetic accessibility scoring with AI-driven retrosynthesis analysis to evaluate the practical feasibility of synthesizing AI-designed drug molecules. This method allows researchers to assess whether the proposed molecular structures can be realistically produced using available chemical reactions and resources. Additionally, frameworks that combine synthetic accessibility with AI-based generative drug design have demonstrated significant potential. By employing reaction-aware generation strategies, these frameworks ensure that the molecules produced are not only novel but also practically synthesizable, bridging the gap between theoretical design and experimental implementation in a laboratory setting.

4.4. Closed Loop Generation with Experimental Validation

Closed loop systems represent an advanced integration of generative artificial intelligence with automated experimental validation, enabling iterative improvements in molecular design. These systems are particularly transformative in fields such as drug discovery, where they facilitate the rapid generation and testing of hypotheses. By automating both the design and validation processes, closed loop approaches significantly reduce the time and resources required for experimental therapeutics. For example, in the context of brain disease, these systems can integrate drug discovery with experimental validation to optimize therapeutic outcomes. Table 3 highlights the core technical advances driving this field, including innovations in algorithmic efficiency, data integration, and experimental automation. Such advancements underscore the potential of closed loop systems to revolutionize scientific research and practical applications.

Table 3. Core Technical Advances

Advance	Core Idea	Improvement
Multi-objective generation	Balance conflicting properties	Better drug-likeness
Target-specific design	Integrate protein information	Higher binding affinity
Synthesis-aware generation	Predict synthesizability	Reduced experimental failure
Closed-loop validation	AI experiment iteration	Faster optimization

5. Major Challenges and Open Problems

5.1. Data Scarcity, Quality, and Bias

The performance of generative AI models is profoundly influenced by the availability of high-quality training data. Challenges such as data scarcity, inconsistent data quality, and inherent biases in chemical and biological datasets significantly hinder advancements in artificial intelligence applications, particularly in pharmacological research. Limited experimental data for certain target classes often restricts the scope of AI-driven drug discovery, while biased chemical libraries can result in skewed molecular generation and reduced model generalizability. These issues underscore the need for more robust data collection methodologies and the development of unbiased datasets to enhance the reliability and applicability of generative AI models [20]. Addressing these

challenges is critical for ensuring that AI systems can produce accurate predictions and facilitate meaningful scientific breakthroughs in drug discovery and related fields.

5.2. Synthetic Feasibility and Patent Novelty

While generative models have the capability to produce an extensive array of novel molecular structures, ensuring that these molecules are both synthetically feasible and patentable remains a significant challenge. Recent advancements have introduced frameworks that integrate generative molecular design with synthetic viability prediction, emphasizing the ongoing gap between computationally generated molecules and those that can be practically synthesized in laboratory settings. Furthermore, critical evaluations have highlighted the difficulties in achieving synthetic accessibility for these molecules, as well as the potential risks of generating structures that either infringe upon existing patents or fail to demonstrate genuine novelty. Addressing these issues requires a more robust alignment between computational design methodologies and practical chemical synthesis processes, ensuring that the generated molecules meet both scientific and legal standards.

5.3. Generalization Beyond Training Distributions

Generative models frequently encounter difficulties when attempting to generalize beyond the chemical distributions present in their training datasets [21–23]. This limitation arises because the models are often optimized for specific patterns and structures within the training data, making it challenging to adapt to unseen distributions. To address these challenges, researchers have explored innovative approaches such as molecular graph-based invariant representation learning, which incorporates environmental inference and subgraph generation techniques. These methods aim to enhance the model's ability to generate molecular structures that are robust to variations in test distributions. Additionally, frameworks designed for retrosynthesis prediction have provided valuable insights into handling out-of-distribution scenarios, offering transferable strategies for molecular generation tasks where the test data significantly diverges from the training data. Such advancements highlight the importance of developing methodologies that prioritize adaptability and robustness in generative modeling.

5.4. Lack of Standardized Evaluation Benchmarks

The absence of standardized benchmarks poses significant challenges in fairly comparing different generative models and accurately assessing their progress. This lack of uniformity hinders the ability to evaluate advancements in a consistent and reliable manner [24, 25]. To address this issue, some frameworks have been developed to provide standardized datasets, metrics, and evaluation protocols. For instance, one benchmark focuses on the comprehensive evaluation of molecular generation models in de novo drug design, offering tools to assess performance across various chemical and biological properties. Another framework introduces a scoring and benchmarking system that enables consistent assessment of generative models, ensuring that their performance can be measured across multiple dimensions with greater reliability and precision.

6. Future Directions

6.1. Toward Autonomous Generative Design Platforms

The integration of generative AI with automated experimentation holds significant potential for the development of fully autonomous drug design platforms. This approach envisions a future where advanced systems can seamlessly address the complexities of molecular design and optimization [26]. By combining generative capabilities with predictive modeling and experimental validation, these platforms aim to eliminate the need for human intervention in critical stages of drug discovery. Such advancements could revolutionize biopharmaceutical research, enabling faster and more efficient identification of therapeutic candidates. Furthermore, the incorporation of iterative feedback loops between AI-driven generation and experimental testing could enhance the

precision and reliability of these systems, paving the way for transformative breakthroughs in drug development.

6.2. Integration with High Throughput and Automated Synthesis

Combining generative models with high throughput experimentation and automated synthesis has the potential to significantly accelerate the design-make-test cycle in scientific research and development. Advanced platforms have been developed to integrate robotics and artificial intelligence, enabling the rapid production and testing of molecules generated by AI systems. These platforms demonstrate how automation can streamline the traditionally labor-intensive processes of chemical synthesis and testing. Furthermore, innovative workflows have been proposed to enhance the synergy between computational design and automated laboratory systems, ensuring seamless transitions from virtual predictions to physical experimentation. This integration not only reduces the time required for discovery but also enhances the precision and reproducibility of experimental outcomes, paving the way for transformative advancements in fields such as drug discovery and materials science [27, 28].

6.3. Large Language Models and Multi Modal Generation

Large language models and multi-modal approaches represent a promising frontier for molecular design. Git Mol, a multi-modal large language model for molecular science, integrates graph, image, and text information to enhance molecular understanding and generation. This innovative approach leverages the strengths of diverse data modalities to provide a more comprehensive analysis of molecular structures and properties. Additionally, chain-of-thought enabled large language models have been utilized for effective and explainable molecular property prediction. By fusing multi-modal molecular information, these models demonstrate improved performance and enhanced interpretability, offering significant advancements in the field. The integration of diverse data types not only refines predictive accuracy but also facilitates deeper insights into molecular behavior, paving the way for more robust applications in molecular science and design.

6.4. Ethical and Regulatory Considerations

As generative AI becomes increasingly integrated into drug discovery, it is imperative for ethical and regulatory frameworks to adapt in tandem with these advancements [29]. Regulatory agencies are progressively addressing the implications of AI-generated molecules, focusing on how to ensure compliance with existing standards while developing new guidelines tailored to this emerging field. Ethical considerations include safeguarding data privacy, mitigating algorithmic bias, and addressing intellectual property challenges that arise from AI-driven innovations. These issues necessitate a comprehensive approach to ensure that AI technologies are utilized responsibly and equitably. Furthermore, the development of specific regulatory pathways for AI-assisted pharmaceutical research is essential to streamline the approval process while maintaining rigorous safety and efficacy standards. Key challenges and potential solutions in this domain are systematically outlined in Table 4, providing a roadmap for future advancements in the ethical and regulatory landscape of AI-driven drug development.

Table 4. Challenges and Future Directions

Challenge	Key Issue	Future Direction
Data limitation	Scarce & biased data	High-quality curated datasets
Synthesis gap	Computation–experiment mismatch	Automated synthesis platforms
Generalization	Poor OOD performance	Robust learning strategies

Evaluation	No unified standard	Standardized benchmarks

7. Conclusion

Generative artificial intelligence has fundamentally transformed the landscape of drug molecule design, shifting the paradigm from high-throughput screening to active molecular generation. This review has systematically examined the foundational generative models, including variational autoencoders, generative adversarial networks, autoregressive models, and diffusion models, each offering unique advantages for exploring chemical space. These models enable the efficient exploration of vast and complex chemical landscapes, facilitating the identification of novel molecular structures with desirable properties. Additionally, molecular representation strategies, ranging from SMILES strings to graph neural networks and geometric deep learning, have been pivotal in encoding chemical information. Pre-trained molecular language models, which capture intricate chemical semantics, further enhance the ability to predict molecular behavior and interactions, underscoring the transformative potential of these approaches in modern drug discovery pipelines.

Key advances in multi-objective generation, target-specific design, synthetic accessibility awareness, and closed-loop experimental validation have demonstrated the growing maturity of generative AI-driven drug discovery. Multi-objective generation allows for the simultaneous optimization of multiple molecular properties, such as potency, selectivity, and safety, while target-specific design ensures that molecules are tailored to interact with specific biological targets. Synthetic accessibility awareness addresses the practical feasibility of synthesizing proposed molecules, bridging the gap between theoretical design and real-world application. Closed-loop experimental validation integrates computational predictions with experimental feedback, creating a dynamic and iterative process for refining molecular candidates. Despite these advancements, significant challenges remain, including issues related to data quality and bias, synthetic feasibility, out-of-distribution generalization, and the absence of standardized evaluation benchmarks. Addressing these challenges will require the development of more robust datasets, improved algorithms for handling biased or incomplete data, and the establishment of universally accepted metrics for evaluating generative models.

Despite remaining obstacles, the convergence of generative AI with automated experimentation and increasing regulatory clarity is poised to accelerate the transition from AI-generated molecules to clinically viable drug candidates. This integration has the potential to significantly reduce the time and cost associated with drug development, enabling faster delivery of new medicines to patients. Autonomous design platforms, which combine generative AI with high-throughput and automated synthesis, represent a promising avenue for streamlining the drug discovery process. Furthermore, the incorporation of large language models and multi-modal generation techniques could enhance the ability to predict complex molecular interactions and design multifunctional compounds. The establishment of robust ethical and regulatory frameworks will be critical to ensuring the responsible application of these technologies, addressing concerns related to safety, transparency, and equitable access. As these innovations continue to mature, they hold the promise of transforming the pharmaceutical industry, paving the way for a new era of precision medicine and personalized therapeutics.

References

1. Y. Wang, Z. Li, and A. Barati Farimani, "Graph neural networks for molecules," in *Machine Learning in Molecular Sciences*, Cham: Springer International Publishing, 2023, pp. 21-66.
2. E. Zhang, Z. Pan, Z. Yao, T. Dong, G. Chen, T. Deng, ... and C. Y. C. Chen, "Pre-trained language models for protein and molecular design," *Physical Chemistry Chemical Physics*, vol. 27, no. 27, pp. 14189-14216, 2025.
3. E. Lin, C. H. Lin, and H. Y. Lane, "De novo peptide and protein design using generative adversarial networks: an update," *Journal of Chemical Information and Modeling*, vol. 62, no. 4, pp. 761-774, 2022.

4. M. Sypetkowski, F. Wenkel, F. Poursafaei, N. Dickson, K. Suri, P. Fradkin, and D. Beaini, "On the scalability of GNNs for molecular graphs," *Advances in Neural Information Processing Systems*, vol. 37, pp. 19870-19906, 2024.
5. M. Liu, C. Li, R. Chen, D. Cao, and X. Zeng, "Geometric deep learning for drug discovery," *Expert Systems with Applications*, vol. 240, p. 122498, 2024.
6. D. B. Singh, R. K. Pathak, and D. Rai, "From traditional herbal medicine to rational drug discovery: strategies, challenges, and future perspectives," *Revista Brasileira de Farmacognosia*, vol. 32, no. 2, pp. 147-159, 2022.
7. C. Isert, K. Atz, and G. Schneider, "Structure-based drug design with geometric deep learning," *Current Opinion in Structural Biology*, vol. 79, p. 102548, 2023.
8. C. Li, J. Yao, W. Wei, Z. Niu, X. Zeng, J. Li, and J. Wang, "Geometry-based molecular generation with deep constrained variational autoencoder," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 4852-4861, 2022.
9. I. Bano, U. D. Butt, and S. A. H. Mohsan, "New challenges in drug discovery," in *Novel Platforms for Drug Delivery Applications*, Woodhead Publishing, 2023, pp. 619-643.
10. T. Peng, Y. Li, X. Li, J. Bian, Z. Xie, N. Sui, ... and H. Xiong, "Pre-trained molecular language models with random functional group masking," *npj Artificial Intelligence*, vol. 1, no. 1, p. 28, 2025.
11. D. S. Wigh, J. M. Goodman, and A. A. Lapkin, "A review of molecular representation in the age of machine learning," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 12, no. 5, p. e1603, 2022.
12. Y. Liu, Y. Zhu, J. Wang, R. Hu, C. Shen, W. Qu, ... and T. Hou, "A multi-objective molecular generation method based on Pareto algorithm and Monte Carlo tree search," *Advanced Science*, vol. 12, no. 20, p. 2410640, 2025.
13. J. Wang and F. Zhu, "Multi-objective molecular generation via clustered Pareto-based reinforcement learning," *Neural Networks*, vol. 179, p. 106596, 2024.
14. F. J. Ferreira and A. S. Carneiro, "AI-driven drug discovery: a comprehensive review," *ACS Omega*, vol. 10, no. 23, pp. 23889-23903, 2025.
15. S. Singh, R. Kumar, S. Payra, and S. K. Singh, "Artificial intelligence and machine learning in pharmacological research: bridging the gap between data and drug discovery," *Cureus*, vol. 15, no. 8, 2023.
16. F. Urbina, C. T. Lowden, J. C. Culberson, and S. Ekins, "MegaSyn: integrating generative molecular design, automated analog designer, and synthetic viability prediction," *ACS Omega*, vol. 7, no. 22, pp. 18699-18713, 2022.
17. Y. Yu, L. Yuan, Y. Wei, H. Gao, F. Wu, Z. Wang, and X. Ye, "RetroOOD: Understanding out-of-distribution generalization in retrosynthesis prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 1, pp. 374-382, Mar. 2024.
18. M. Stanley and M. Segler, "Fake it until you make it? Generative de novo design and virtual screening of synthesizable molecules," *Current Opinion in Structural Biology*, vol. 82, p. 102658, 2023.
19. M. Wang, C. Y. Hsieh, J. Wang, D. Wang, G. Weng, C. Shen, ... and T. Hou, "Relation: A deep generative model for structure-based de novo drug design," *Journal of Medicinal Chemistry*, vol. 65, no. 13, pp. 9478-9492, 2022.
20. M. Thomas, N. M. O'Boyle, A. Bender, and C. De Graaf, "MolScore: a scoring, evaluation and benchmarking framework for generative models in de novo drug design," *Journal of Cheminformatics*, vol. 16, no. 1, p. 64, 2024.
21. A. Bettanti, A. R. Beccari, and M. Biccarino, "Exploring the future of biopharmaceutical drug discovery: can advanced AI platforms overcome current challenges?," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 102, 2024.
22. S. Zhang, H. Zhu, R. Zhou, and M. Li, "MoGE: A Benchmark for Comprehensive Evaluation of Molecular Generation Models in De Novo Drug Design," in *International Symposium on Bioinformatics Research and Applications*, Singapore: Springer Nature Singapore, 2025, pp. 21-33.
23. K. Zhang, X. Yang, Y. Wang, Y. Yu, N. Huang, G. Li, ... and S. Yang, "Artificial intelligence in drug development," *Nature Medicine*, vol. 31, no. 1, pp. 45-59, 2025.
24. J. M. Lu, J. Z. Pan, Y. M. Mo, and Q. Fang, "Automated Intelligent Platforms for High-Throughput Chemical Synthesis," *Artificial Intelligence Chemistry*, vol. 2, no. 1, p. 100057, 2024.
25. P. Liu, Y. Ren, J. Tao, and Z. Ren, "Git-mol: A multi-modal large language model for molecular science with graph, image, and text," *Computers in Biology and Medicine*, vol. 171, p. 108073, 2024.
26. S. Balaguru and A. Gandra, "Unleashing Molecular Potential: A Process Discovery and Automation Workflow for Generative AI in Accelerating Drug Discovery," *International Journal of Innovative Science and Research Technology*, vol. 9, pp. 1235-1241, 2024.
27. C. Jin, S. Guo, S. Zhou, and J. Guan, "Effective and explainable molecular property prediction by chain-of-thought enabled large language models and multi-modal molecular information fusion," *Journal of Chemical Information and Modeling*, vol. 65, no. 11, pp. 5438-5455, 2025.
28. P. Warokar and S. Lote, "Ethical and regulatory consideration in AI-assisted drug development," in *2024 2nd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIEI)*, Nov. 2024, pp. 1-6.
29. F. Mirakhori and S. K. Niazi, "Harnessing the AI/ML in drug and biological products discovery and development: the regulatory perspective," *Pharmaceuticals*, vol. 18, no. 1, p. 47, 2025.