Methodology and Evaluation for Machine Learning in Drug Molecule Design

Zhiyuan Yang*

School of Life Science and Technology, China Pharmaceutical University, Nanjing, 211198, China

Abstract. With the aging of world populations and the increasingly complex disease conditions, traditional drug discovery also encounters some problems such as soaring costs, extended time frames, and slim winning odds. The enormous chemical space has made laboratory screening more difficult to do. And further tangling constraints, like optimizing pharmaceutical properties and synthesizability, make it more difficult to identify lead compounds. The advances in biological data collection and high-performance computing lead this review on the topic of generative deep learning in drug molecule design. This review delineates the capabilities and limitations of linear and graph-based representations and summarizes the operating principles and strengths of commonly used machine learning architectures. In addition, this review presents a twofold evaluation process, i.e., generation performance and drug-likeness. Across recent studies, the generative approaches speed up drug-like molecule exploring and enable goal-conditioned design, although faithful 3D conformations and integration with molecular dynamics still have some problems, and a sound quantification of synthetic accessibility still is a problem. It's possible that a pipeline that is closed-loop and couples the generation of molecules with retrosynthesis and automated experimentation would be advantageous.

1 Introduction

New drug research and development face significant challenges on a global scale. Traditional pipelines typically require screening from the vast number of existing compounds, yet on average only a single agent passes clinical trials and successfully reaches the market[1]. The average time from a new drug program initiation to obtaining regulatory approval takes roughly 10–15 years, with an estimated cost of approximately $2.6 billion, while the clinical success rate for new drugs is less than 10%[2]. There is no doubt that such a low success rate makes the excessively long development cycle and high costs seem particularly awkward. Meanwhile, population aging is becoming increasingly pronounced. Since the elderly population faces a higher risk of chronic diseases, it is foreseeable that demand for therapies targeting cardiovascular disease, neurodegenerative disorders, and related conditions will surge in the future[3]. Additionally, the physiological changes in human organs caused by aging lead to alterations in pharmacokinetics and pharmacodynamics, presenting new challenges for drug development in addressing the specific medication needs of the elderly population[4]. The contradiction between the low efficiency of new drug development and the growing demand for medical care underscores the urgency of accelerating drug discovery and improving the success rates.

In recent years, propelled by the accumulation of biomedical data and advances in high-performance computing, artificial intelligence (AI) technology has been rapidly adopted by medicinal chemistry researchers. In 2020, MIT Technology Review ranked the application of AI in active small molecule discovery as one of the major breakthroughs of the year[5]. Deep learning methods show promising applications; all stages of the process are transformed by deep learning methods. On one hand, deep learning can automatically learn from vast amounts of compounds and massive biological activity data for the prediction of various drug indicators including the physical and chemical properties, efficacies, and safeties. Studies have shown that artificial intelligence (AI) models can be used for the prediction of ADMET properties of candidate molecules and help the compound have better drug-like properties[6]. Based on the learning of the distribution status of existing molecules, generative models based on deep learning have achieved remarkable progress in the design of new molecules. They will be able to make structurally different but bioactive molecules for much less cost. Compared with traditional medicinal chemistry methods that rely heavily on expert experience, generative models can better explore a large chemical space and greatly reduce trial-and-error work at the beginning of screening. A study used generative AI to find preclinical candidate molecules for new drugs in around 18 months, and researchers finished everything from discovering the target to proving its safety in phase I trials within about 30 months, a much faster timeline than conventional new drug R&D[7]. In short, the above advances collectively show the great capability of artificial intelligence in accelerating lead compound exploration and improvement.

* Corresponding author: yangzy@stu.cpu.edu.cn

This review presents the landscape of generative methods for small molecule design and introduces two major molecular representations used for generating molecules, which are linear representation and graph representation. Based on the foundation, the review lists a number of state-of-the-art representative deep generative model frameworks and the emerging diffusion model. The principles and performance of each approach are also discussed. Although deep generative models have brought new opportunities for drug finding, these models still have many problems such as the synthetic feasibility of generated molecules, multi-objective optimization, and the balance between novelty and drug activity. Therefore, this review will also inspect scores for generative models from the perspective of how well they generate molecules and whether those molecules have plausible properties suitable for pharmaceutical use.

2 Molecule representations

In computational workflows, chemical molecules must be input into generative models in a computer-readable, standardized format, and the target molecules generated by the model need to be represented in a universal format. Such standardized representations include linear representation, graph representation, molecular fingerprint, molecular descriptor, and three-dimensional structural format. Among deep learning models for molecular generation reported in recent years, the most used are linear representation and graph representation[8].

2.1 Linear representation

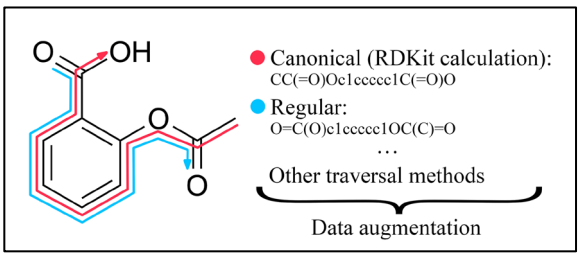

Fig. 1. SMILES representations of aspirin.

Linear representations code molecular structure into a linear string form. They stuff structural information such as atoms and chemical bonds into a line of symbols so scientists can store molecules inside words in computers. SMILES is a typical linear representation: Simplified Molecular Input Line Entry System. David Weininger proposed the SMILES in 1988. Its basic rule is to traverse the molecule and encode atoms and chemical bonds as a symbol sequence according to the traversal order[9]. In particular, the SMILES employs symbolic characters to represent symbols and bonds, applies numerical tags to mark the ring closure location and adopts parentheses to represent branched structure. This method writes in a compact form that is quick to read. SMILES is a concise and efficient format; due to this, SMILES is among the most commonly used molecular encodings[10]. However, it also has some deficiencies. As shown in Figure 1, a single molecule can have various SMILES based on starting points for traversal. To overcome the problem, canonicalization algorithms are used for delivering canonical SMILES, and randomized SMILES start atoms are sampled for data augmentation[11]. SMILES originally did not encode stereochemical information about molecules, but subsequent extensions introduced symbols for chirality and cis-trans isomerism, enabling SMILES to represent partial molecular 3D information. SMILES is widely used for molecular storage, indexing, and retrieval in chemical databases. Many chemical information platforms, such as PubChem and ChEMBL, support SMILES as input for structural searches. Its ASCII character format is compatible with text processing tools and can be directly used in text-based models and can be edited with standard text softwaree.

However, SMILES also has salient drawbacks. In generative models, SMILES often produce syntactically invalid or chemically incorrect strings that cannot be decoded into real molecules; SMILES has limited support for certain structures, such as compounds with atypical valence, hypervalent or multicenter bonding, and coordination complexes like organometallics, which are difficult to represent using conventional SMILES[12]. Additionally, SMILES is essentially a linear projection of a 2D structure; while it can express stereochemistry through symbols, it does not directly include spatial geometric information.

2.2 Graph representation

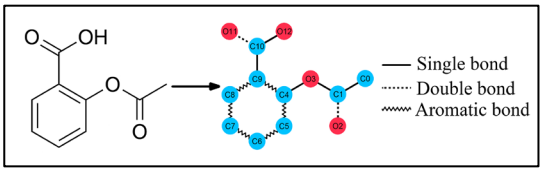

Fig. 2. Graph representation of aspirin.

A graph representation conceptualizes molecules as networks with atomic nodes and bond edges such as aspirin in Figure 2. Compared to SMILES, graph representation encodes interatomic connectivity more faithfully and more naturally captures molecular topology as well as the local chemical environments of atoms and bonds. Molecular graphs can include additional attributes: node features typically include atomic hybridization type, valence, formal charge, spin state, etc.; edge features include bond order, participation in conjugation, and stereochemical descriptors. However, a pure 2D molecular graph omits three-dimensional information like bond lengths, bond angles, and dihedral angles. For example, conventional graph representation cannot automatically encode the absolute configuration of chiral centers. If a complete representation of the three-dimensional configuration is required, atomic coordinates can be incorporated to construct a 3D molecular graph that explicitly encodes spatial geometry[13].

Graph representations have already shown strong performance in property prediction, activity screening, reaction prediction, molecular generation, and molecule design[14]. Xiangxiang Zeng et al. developed an unsupervised pre-training deep learning framework called

ImageMol, which leverages molecular images as feature representations for compounds, offering advantages of high accuracy and low computational cost[15]. This framework was used to identify several potential 3C-like protease inhibitors for treating COVID-19. Moreover, subgraphs in graph representations are always interpretable actual chemical fragments, whereas substrings in SMILES and other linear notations often lack chemically meaningful counterparts—giving graph representation a natural advantage for structural interpretation[16].

Nonetheless, graph representation has limitations. Molecular graphs are not directly readable by researchers and typically require visualization components. One user-friendly visualization tool is RDKit, an open-source package widely used in cheminformatics and computational chemistry, providing 2D/3D molecular visualization and integrating tightly with the Python scientific ecosystem. From a computational perspective, graphs demand more storage and computational resources than compact one-dimensional strings, and training graph-based models often entails higher complexity and greater sensitivity to hyperparameters[17].

3 Generative models

Machine-learning pipelines can be realized with several alternative architectures. For new drug researchers, molecular generative models most commonly adopt generative adversarial network (GAN), recurrent neural network (RNN), variational autoencoder (VAE), transformer model, and diffusion model[18].

3.1 Generative adversarial network

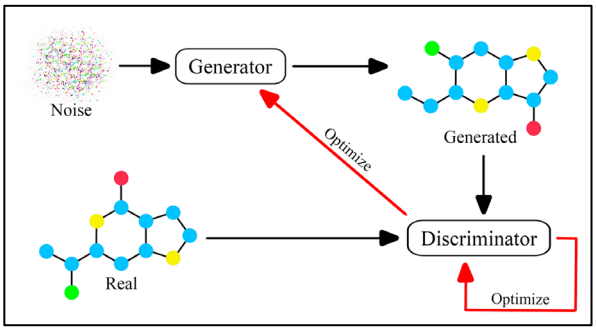

Fig. 3. Framework diagram of the GAN model.

Figure 3 shows GAN's framework. It pairs a generator with a discriminator and trains them jointly through a contest. The generator produces data-like observations from noise, whereas the discriminator evaluates their authenticity against real data. In this structural game between the two components, the generator reduces perceptual gaps to real data while the discriminator enhances classification fidelity, and the system ultimately reaches a dynamic equilibrium. Through this adversarial interplay, the generator progressively improves the realism of the generated samples, and the discriminator continuously enhances its accuracy in distinguishing between real and fake data. For molecule design, GAN offers a route to directly sample novel molecules without specifying an explicit probabilistic model, probing uncharted regions of chemical space and allowing the generator to yield plausible new structures[19]. MolGAN uses reinforcement learning by attaching a reward module to the standard GAN model. It consumes a random vector and spits out a labeled molecular graph in a single shot. The discriminator is fed both real molecules from the dataset as well as artificial molecules created by the generator, attempting to differentiate between them. Simultaneously, as for the reward network, it gives scores to the generated molecules and how well the molecule fulfills specified requirements such as physiochemical properties, thus giving metrics to provide gradient signals to the generator, making it optimized in its desired properties but also valid. Using adversarial training in combination with reinforcement learning, MolGAN creates new compounds that are valid and have target properties. This shows that GAN can sample and optimize directly in the molecular graph space and is useful for de novo drug design[20].

3.2 Recurrent neural network

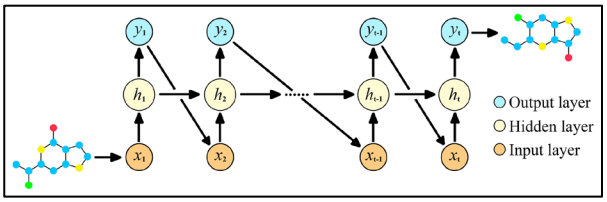

Fig. 4. Framework diagram of the RNN model.

As illustrated in Figure 4, recurrent neural networks can utilize the state from the previous step to influence the output of the current step, thereby capturing temporal dependencies within sequences. RNN architecture is good at handling order-dependent data. RNN-type generative models can learn existing compound libraries' SMILES syntax and efficiently generate many novel compounds like those from the training set. Anvita Gupta et al. build an RNN-driven model to do high-throughput screening and optimize lead compounds, which shows the power of the RNN in building a large number of diverse structures for de novo drug design[21]. RNN is often used with RL to place constraints on the desired physicochemical properties or biological activity. In the study by Marcus Olivecrona et al., they first trained an RNN to learn the prior distribution of molecular syntax based on a database and then fine-tuned it with a policy gradient algorithm. This fine-tuning made the RNN tend to generate molecules meeting certain property requirements. This strategy greatly boosts goal-directed optimization, letting researchers fix many things at once, such as bioactivity improvement and how good a compound is for medicine, without changing the molecule diversity[22].

3.3 Variational autoencoder

VAE is an architecture that combines molecular representation learning with generation through an

encoder–decoder architecture and a shared latent space (Figure 5). The encoder maps discrete molecular representations to latent vectors, and the decoder reconstructs molecules from these continuous vectors. By embedding molecules in a continuous space where each point corresponds to a structure, the VAE enables smooth interpolation and stochastic sampling to create new compounds. A key advantage lies in the ability to search latent space efficiently with gradient-based methods, thereby discovering molecules optimized for specified properties. Gómez-Bombarelli et al. demonstrated this by training a VAE on drug-like molecules and adjusting latent vectors along gradient directions to identify compounds with improved target attributes[23].

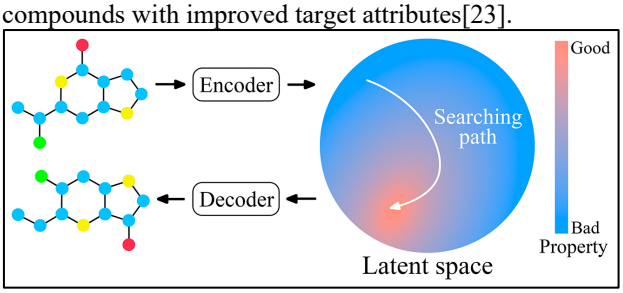

Fig. 5. Framework diagram of the VAE model.

3.4 Transformer model

Transformer model is a deep learning architecture built on self-attention. Originally developed for natural language processing, in recent years, it has also been adopted into the field of molecular generation. Unlike RNN, which processes data sequentially, transformer model can simultaneously focus on the global relationships between elements, making it well-suited to capture long-range dependencies between remote structures in SMILES or molecule graphs. Transformer model often follows an encoder–decoder paradigm: the encoder ingests conditioning signals (e.g., a given scaffold or fragment constraints), while the decoder generates molecular sequences based on internal self-attention mechanisms[24]. For example, Chemformer is pretrained on large-scale SMILES data and attains state-of-the-art accuracy on sequence-to-sequence problems such as reaction prediction and retrosynthesis, while supporting conditional, multi-objective generation of candidate molecules and demonstrating strong performance in multi-task molecular optimization[25].

3.5 Diffusion model

Diffusion model is a class of generative methods built on a forward noise addition process and a backward noise removal process. Figure 6 presents this kind of procedure. It first gradually adds noise to data until it is completely corrupted, then trains a neural network to learn how to remove the noise in reverse, ultimately enabling the neural network to map pure noise back into meaningful data structures. In molecular generation, diffusion model can learn high-dimensional features of molecular structures and, through denoising, produce molecules that satisfy specified properties or constraints. Its strength is particularly evident for three-dimensional generation: diffusion model can capture the complex distribution of different conformations while ensuring that the molecular geometry remains unchanged, thereby yielding stable and practically realistic 3D molecular structures[26]. It also makes diffusion model more beneficial in drug molecule design compared to other models because the 3D configuration of the molecule determines its interactions with biological targets (receptors, enzymes), thus directly affecting its pharmacology.

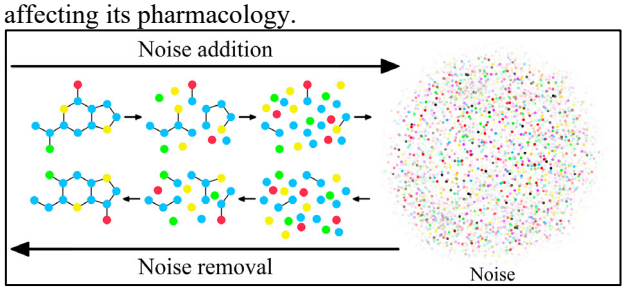

Fig. 6. Framework diagram of the diffusion model.

4 Metrics

Generative models would be liable for producing more intricate candidates, so it comes to a doubt as to how to impartially and exactly determine structural qualities such as also the developability. The traditional experimental validation takes a large amount of time and resources, so rapid screening and assessment of model outputs are necessary. The design and application of evaluation metrics are important: contemporary work tends to evaluate the models in terms of both molecular generation performance and drug-likeness, which gives a full view of how effective and useful the models are. Generation performance takes a look at the chemical structures of those outputs. It gives people some indication of how well scientists have imprinted the training distribution into the model as well as whether the model can create new molecular scaffoldings. Drug-likeness evaluation emphasizes the pharmacological prospects of a chemical compound and its biopharmaceutical characteristics. It offers a filter to focus on candidates with great potential for development.

4.1 Generation performance

Validity measures the chemical plausibility of generated molecules, typically defined as the rate at which molecules are parsed without error as structurally valid by chemistry software. The goal of assessing validity is to filter out invalid outputs, such as those with valence violations or broken bonds, enforcing basic chemical rules. As the most fundamental metric, the level of validity directly reflects the model's proficiency with SMILES grammar or molecular structural constraints. High validity ensures that generated compounds have practical significance and are meaningful precursors for subsequent evaluation and experimental testing.

Uniqueness indicates whether the generated molecules are diverse. It gauges whether the model produces diverse outcomes; high uniqueness indicates the

absence of mode collapse and broader exploration of chemical space. In practice, duplicates are removed when tallying samples, and the share of remaining molecules among all valid ones is reported. In molecular generation tasks, the uniqueness metric prevents the model from repeatedly outputting the same molecules, ensuring structural diversity in the output set. Low uniqueness suggests the model is trapped in a narrow region of the training distribution and fails to explore novel structures effectively.

Novelty counts the fraction of molecules new to the training database. It characterizes how far the designs extend beyond known molecular inventories. Novelty is crucial for discovering potential drug lead molecules, as it ensures that the generated model does not simply replicate existing compounds but can explore unknown chemical spaces. High novelty indicates that the model has internalized general chemical patterns rather than memorizing exemplars, enabling the creation of molecules outside the training corpus.

Controllability describes how well the model responds to external conditions or user instructions during the generation process. For example, meeting thresholds on properties, enforcing a given scaffold, or including particular functional groups. It is commonly quantified by the proportion of generated compounds that satisfy all given constraints. Strong controllability is essential in drug design, where candidates must simultaneously meet multiple physicochemical properties and biological activity requirements, enabling on-demand generation of molecules with prescribed attribute combinations.

4.2 Drug-likeness

The physicochemical properties of a drug serve as foundational metrics for assessing its drug-likeness, including molecular weight (MW), partition coefficient (log*P*), polar surface area (PSA), counts of hydrogen-bond donors/acceptors (HBD/HBA), and number of rotatable bonds, among others. These parameters reflect key behaviors such as absorption, solubility, and membrane permeability. The classic "Lipinski's Five Rules" state that orally administered drugs typically require HBD $\leqslant$ 5, HBA $\leqslant$ 10, MW $\leqslant$ 500 Da, and log*P* $\leqslant$ 5, among other criteria, to ensure adequate absorption by the body; otherwise, adequate absorption will be unlikely[27]. Such empirical guidelines, along with the Veber rule (emphasizing restrictions on rotatable bonds and PSA), are commonly used to filter and guide molecular library design. Quantitative schemes based on molecular properties have also been proposed. For example, quantitative estimation of drug-likeness (QED) by G. Richard Bickerton et al., which combines multiple descriptors to provide a comprehensive drug potential score ranging from 0 to 1[28]. As a whole, they will generally be fairly flexible and let researchers eliminate those invalid molecules that are too large, too polar, too lipophilic, or too hydrophobic.

Molecular docking is the prediction of a small molecule's binding orientation and affinity to a druggable target. Docking algorithms are looking for preferable conformations in the binding site of molecules and using score functions to estimate binding free energy, which can help judge ligand-target complementarity. AI-assisted pipelines use docking scores, the output of AutoDock, SMINA, etc., as screening criteria[29]. Guangfeng Zhou et al. stressed the need for an accurate prediction of binding modes and affinities for virtual screen success. Their result showed that correct predictions can greatly speed up lead compound finding[30]. Molecular docking will filter out the highest affinity one, yet it is also a feasibility check of all those conformation choices.

Synthetic feasibility is about the practicality of generated molecules. Although the AI-based generation method can cover huge chemical space, it may produce things that are undesirable. Molecules can be pretty complicated so they have a very long synthesis path. Therefore, in order to avoid waste of research resources, it is very necessary to carry out synthesis feasibility evaluation in time from the process of molecular design. Common metrics are scores that describe retrosynthetic analysis. One of the most common metrics for molecular complexity is the SA score as proposed by Peter Ertl and Ansgar Schuffenhauer in 2009. The SA score mixes fragment frequency statistics and structural features like large rings and functional groups' complexity into a 1-10 score[31]. New scoring methods predict by outputting routes by automated retrosynthetic algorithms like RS score and RA score from AI-powered retrosynthetic algorithms[32]. These ones can actually give a better sense as to whether we can synthesize a molecule in rational steps, but they are computationally quite costly. Synthetic feasibility metrics help get rid of candidates that are difficult to prepare. The adoption of synthetic feasibility as an optimization objective will lead the generator to make reachable chemotypes for reality.

5 Conclusion

This review systematically investigates two molecular presentations and some commonly used molecular generation models. Linear representation and graph representation are present in detail and molecular generative frameworks built upon them are discussed. Each way has its own strength when it comes to holding information about the physical structure and chemicals. From existing research, the design of the generative model is key for drug discovery, since it speeds up new compound discovery and demonstrates the performance of AI transferred to the chemical area.

But currently, there are still some deficiencies in three-dimensional molecular conformation design and molecular dynamics simulation. Breakthroughs in these areas would enable more authentic simulations of drug target interactions and give more guidance to future research. Though some progress has been made, there is still room for improving the quality of generation of already existing design models. Many of these more current approaches chiefly learn existing, proven computer vision and NLP techniques but have few or no models, design principles, etc., specific to the pharmacy world. And synthetic accessibility evaluation still needs

improvement. Prevailing assessments use empirical metrics or simple estimations; a more accurate assessment is needed. And progress in generating molecules is creating chances in related tasks like working out retrosynthetic route plans, which signifies that a closed loop, automatically carried out process going from molecular designing to determining how to synthesize it can be realized in the future.

Computing and hardware technology are progressing and improving continuously, which means the generated molecules that behave like carefully designed drugs should also get better. Future research must combine deep generative models with automated experimental platforms to speed up the drug discovery process in a fully automatic way. In such an environment, more user-friendly tools and operating environments would allow for closer collaboration between chemists and computational scientists with a better handoff between in-silico design and experiments.

References

1. Torjesen, I. (2015, May 12). Drug development: the journey of a medicine from lab to shelf. the Pharmaceutical Journal. https://pharmaceutical-journal.com/article/feature/drug-development-the-journey-of-a-medicine-from-lab-to-shelf
2. Rousseaux, C. G., Bracken, W. M., & Guionaud, S. (2023). Chapter 1 - Overview of Drug Development. In W. M. Haschek, C. G. Rousseaux, M. A. Wallig, B. Bolon, B. Bolon, K. M. Heinz-Taheny, D. G. Rudmann, & B. W. Mahler (Eds.), Haschek and Rousseaux' s Handbook of Toxicologic Pathology (Fourth Edition) (pp. 3-48). Academic Press. https://doi.org/10.1016/B978-0-12-821047-5.00016-6
3. Fermini, B., & Bell, D. C. (2022). On the perspective of an aging population and its potential impact on drug attrition and pre-clinical cardiovascular safety assessment. J Pharmacol Toxicol Methods, 117, 107184. https://doi.org/10.1016/j.vascn.2022.107184
4. Drenth-van Maanen, A. C., Wilting, I., & Jansen, P. A. F. (2020). Prescribing medicines to older people-How to consider the impact of ageing on human organ and body functions. Br J Clin Pharmacol, 86(10), 1921-1930. https://doi.org/10.1111/bcp.14094
5. Temple, J. (2020, February 26). 10 Breakthrough Technologies 2020. MIT Technology Review. https://www.technologyreview.com/10-breakthrough-technologies/2020/
6. Nag, S., Baidya, A. T. K., Mandal, A., Mathew, A. T., Das, B., Devi, B., & Kumar, R. (2022). Deep learning tools for advancing drug discovery and development. 3 Biotech, 12(5), 110. https://doi.org/10.1007/s13205-022-03165-8
7. Xu, Z., Ren, F., Wang, P. et al. (2025). A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial. Nat Med 31, 2602–2610. https://doi.org/10.1038/s41591-025-03743-2
8. Pang, C., Qiao, J. B., Zeng, X. X., Zou, Q., & Wei, L. Y. (2023). Deep Generative Models in De Novo Drug Molecule Generation [Review]. Journal of Chemical Information and Modeling, 64(7), 2174-2194. https://doi.org/10.1021/acs.jcim.3c01496
9. Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences, 28(1), 31-36. https://doi.org/10.1021/ci00057a005
10. Wigh, D. S., Goodman, J. M., & Lapkin, A. A. (2022). A review of molecular representation in the age of machine learning. WIREs Computational Molecular Science, 12(5). https://doi.org/10.1002/wcms.1603
11. Bjerrum, E. J. (2017). SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules. arXiv preprint arXiv:1703.07076. https://doi.org/10.48550/arXiv.1703.07076
12. Leon, M., Perezhohin, Y., Peres, F., Popovic, A., & Castelli, M. (2024). Comparing SMILES and SELFIES tokenization for enhanced chemical language modeling. Sci Rep, 14(1), 25016. https://doi.org/10.1038/s41598-024-76440-8
13. Guo, Z., Guo, K., Nan, B., Tian, Y., Iyer, R. G., Ma, Y., Wiest, O., Zhang, X., Wang, W., Zhang, C., & Chawla, N. V. (2023). Graph-based Molecular Representation Learning. arXiv preprint arXiv:2207.04869. https://doi.org/10.48550/arXiv.2207.04869
14. David, L., Thakkar, A., Mercado, R., & Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. J Cheminform, 12(1), 56. https://doi.org/10.1186/s13321-020-00460-5
15. Zeng, X., Xiang, H., Yu, L., Wang, J., Li, K., Nussinov, R., & Cheng, F. (2022). Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. Nature Machine Intelligence, 4(11), 1004-1016. https://doi.org/10.1038/s42256-022-00557-6
16. Chen, X., & Qian, Q. (2023). subGE: Enhancing the subgraph representation of molecular compounds structure–activity relationship discovery. Engineering Applications of Artificial Intelligence, 119. https://doi.org/10.1016/j.engappai.2022.105727
17. Wang, S., Zhang, R., Li, X., Cai, F., Ma, X., Tang, Y., Xu, C., Wang, L., Ren, P., Liu, L., Wu, S., Qian, Q., & Bai, F. (2025). Recent advances in molecular representation methods and their applications in scaffold hopping. npj Drug Discovery, 2(1). https://doi.org/10.1038/s44386-025-00017-2
18. Tong, X., Liu, X., Tan, X., Li, X., Jiang, J., Xiong, Z., Xu, T., Jiang, H., Qiao, N., & Zheng, M. (2021). Generative Models for De Novo Drug Design. J Med Chem, 64(19), 14011-14027. https://doi.org/10.1021/acs.jmedchem.1c00927

19. Blanchard, A. E., Stanley, C., & Bhowmik, D. (2021). Using GANs with adaptive training data to search for new molecules. J Cheminform, 13(1), 14. https://doi.org/10.1186/s13321-021-00494-3
20. Cao, N. D., & Kip, T. (2022). MolGAN: An implicit generative model for small molecular graphs. arXiv preprint arXiv:1805.11973. https://doi.org/10.48550/arXiv.1805.11973
21. Gupta, A., Muller, A. T., Huisman, B. J. H., Fuchs, J. A., Schneider, P., & Schneider, G. (2018). Generative Recurrent Networks for De Novo Drug Design. Mol Inform, 37(1-2). https://doi.org/10.1002/minf.201700111
22. Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. J Cheminform, 9(1), 48. https://doi.org/10.1186/s13321-017-0235-x
23. Gomez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernandez-Lobato, J. M., Sanchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., & Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent Sci, 4(2), 268-276. https://doi.org/10.1021/acscentsci.7b00572
24. Mswahili, M. E., & Jeong, Y. S. (2024). Transformer-based models for chemical SMILES representation: A comprehensive literature review. Heliyon, 10(20), e39038. https://doi.org/10.1016/j.heliyon.2024.e39038
25. Irwin, R., Dimitriadis, S., He, J., & Bjerrum, E. J. (2022). Chemformer: a pre-trained transformer for computational chemistry. Machine Learning: Science and Technology, 3(1). https://doi.org/10.1088/2632-2153/ac3ffb
26. Alakhdar, A., Poczos, B., & Washburn, N. (2024). Diffusion Models in De Novo Drug Design. J Chem Inf Model, 64(19), 7238-7256. https://doi.org/10.1021/acs.jcim.4c01107
27. Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. Advanced Drug Delivery Reviews, 46(1), 3-26. https://doi.org/10.1016/S0169-409X(00)00129-0
28. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. Nat Chem, 4(2), 90-98. https://doi.org/10.1038/nchem.1243
29. Cieplinski, T., Danel, T., Podlewska, S., & Jastrzebski, S. (2023). Generative Models Should at Least Be Able to Design Molecules That Dock Well: A New Benchmark. J Chem Inf Model, 63(11), 3238-3247. https://doi.org/10.1021/acs.jcim.2c01355
30. Zhou, G., Rusnac, D. V., Park, H., Canzani, D., Nguyen, H. M., Stewart, L., Bush, M. F., Nguyen, P. T., Wulff, H., Yarov-Yarovoy, V., Zheng, N., & DiMaio, F. (2024). An artificial intelligence accelerated virtual screening platform for drug discovery. Nat Commun, 15(1), 7761. https://doi.org/10.1038/s41467-024-52061-7
31. Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform, 1(1), 8. https://doi.org/10.1186/1758-2946-1-8
32. Skoraczynski, G., Kitlas, M., Miasojedow, B., & Gambin, A. (2023). Critical assessment of synthetic accessibility scores in computer-assisted synthesis planning. J Cheminform, 15(1), 6. https://doi.org/10.1186/s13321-023-00678-z