



Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



REVIEW

A review of recent advances in generative artificial intelligence models for biomolecular sciences



Jian Jiang^{a,b,*}, Daixin Li^a, Guilin Wang^a, Nicole Hayes^b, Yazhou Shi^a,
Huahai Qiu^a, Bengong Zhang^a, Tianshou Zhou^c, Guo-Wei Wei^{b,d,e,*}

^aResearch Center of Nonlinear Science, School of Mathematical and Physical Sciences, Wuhan Textile University, Wuhan 430200, China

^bDepartment of Mathematics, Michigan State University, East Lansing, MI 48824, USA

^cKey Laboratory of Computational Mathematics, Guangdong Province, and School of Mathematics, Sun Yat-sen University, Guangzhou 510006, China

^dDepartment of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA

^eDepartment of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

Received 3 April 2025; received in revised form 25 July 2025; accepted 20 October 2025

KEY WORDS

Generative artificial intelligence models;
Biomolecules;
Drug discovery;
Protein engineering;
Variational autoencoders;
Generative adversarial networks;
Diffusion models

Abstract Generative artificial intelligence (AI) models, a class of AI techniques that learn data distributions to synthesize novel samples, have emerged as impactful tools across scientific disciplines. In recent years, these models have found extensive applications in fields such as natural language processing and biomedical sciences. Despite their growing influence, comprehensive reviews on the application of generative models in biomolecular sciences remain limited. In this review, we provide a systematic overview of recent advances in generative models applied to biomolecular sciences. We discuss several prominent generative architectures, including variational autoencoders, generative adversarial networks, and diffusion models, highlighting their applications in molecular design and bioinformatics. Additionally, we examine how these models contribute to critical challenges such as molecular property prediction and molecular generation. Finally, we discuss key challenges that remain in this field, including model interpretability, scalability, and the need for high-quality molecular datasets. We highlight emerging research directions that aim to overcome these limitations and propose strategies for improving the reliability and applicability of generative models in biomolecular problems. Through this review, our objective is to provide researchers with a comprehensive understanding of the current landscape of generative modeling in biomolecular sciences and to inspire further advancements in this interdisciplinary area.

This article is part of special issue entitled: Machine Learning in Drug Discovery published in Acta Pharmaceutica Sinica B.

*Corresponding authors.

E-mail addresses: jjiang@wtu.edu.cn (Jian Jiang), weig@msu.edu (Guo-Wei Wei).

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

<https://doi.org/10.1016/j.apsb.2025.12.012>

2211-3835 © 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the past decade, machine learning (ML) has transformed biomolecular sciences, including molecular biology, biochemistry, and structural biology, allowing researchers to address complex biological questions, design novel compounds, and predict molecular properties¹. Among various ML paradigms, generative artificial intelligence (AI) models stand out for their ability to synthesize new molecular structures and properties from learned representations². This not only accelerates discovery but also enhances creativity in molecular design, reduces experimental time and costs, and facilitates deeper exploration of biomolecular sciences. Generative models learn the data distribution to generate new data similar to real data. By learning the statistical characteristics of the sample data, they can generate new data that mirrors the training set. These models are widely used in biomolecular sciences to accelerate drug discovery and protein engineering, improve experimental efficiency, and reduce development costs^{2,3}. Techniques such as generative adversarial networks (GANs)⁴, variational autoencoders (VAEs)⁵, and diffusion models (DMs)⁶ can help researchers efficiently search and optimize within a vast molecular design space, driving technological advancements in multiple fields⁷.

In biomolecular sciences, generative models are used to predict and design molecular structures and properties, especially in drug design, protein design, and chemical reaction prediction. These models, as a subset of ML, learn from existing data to generate new structures with desired properties by capturing the relationship between structure and function. Unlike other models, generative models focus on sampling data distributions to design molecules with specific properties, saving both time and cost⁸.

The field of generative models is rapidly evolving. Existing reviews on the application of generative models in biomolecular sciences are limited, and further exploration is needed. This paper examines their potential to transform biomolecular sciences by offering innovative solutions and advancing our understanding of molecular systems. Despite challenges such as data scarcity, interpretability, and computational demands, future developments should focus on improving model accuracy, interpretability, and integration with experimental research. We analyze the strengths, limitations, and prospects of generative models, demonstrating how they can accelerate scientific discovery and innovation in biomolecular sciences⁹.

The remainder of this article is organized as follows. Section 2 introduces generative technologies, methods, and theoretical foundations. Sections 2.2–2.3 provide detailed overviews of different generation methods and mathematical foundations. Section 3 discusses the applications of generative models in biomolecular sciences. Section 4 addresses current challenges and limitations, while Section 5 explores future prospects in biomolecular sciences.

2. Overview of generative models

2.1. Introduction to generative modeling techniques

Model generation is a subfield of ML that focuses on learning data distribution and generating relevant features of the original data¹⁰.

Unlike discriminative models, it focuses on understanding the data generation process and enhancing data by capturing patterns and characteristics. It is derived from statistical methods, similar to Gaussian mixture models and Markov chains¹¹. Baum et al.¹² developed hidden Markov models in the 1960s. These early models paved the way for subsequent generative models to comprehensively evaluate molecular designs.

The core goal of generative modeling is to learn the joint probability distribution $p(x)$ of the data, where x represents the data (e.g., images, texts, or audios)¹³. By learning this distribution, generative models reveal the underlying structure and generate new samples that preserve statistical properties such as texture, structure, and pattern. In biomolecular fields such as molecular biology, structural biology, and biochemistry, these models are used to learn and generate the distribution and structural properties of complex molecules, aiding in the design of new molecules and accelerating the exploration of chemical space.

Text generation models have been used in various fields, including image generation, drug discovery, protein structure prediction, and protein engineering. Some of the key advantages of generative models are the ability to synthesize new data, which plays an important role in data augmentation, and to enable probabilistic reasoning to enhance the creativity of AI systems¹⁴.

As generative models continue to demonstrate their vast potential in various fields, particularly in biomolecular sciences, understanding the strengths and limitations of different generative approaches becomes increasingly important. To better evaluate and select the most suitable generative model, the following section will provide a detailed comparison of key models, including VAEs, GANs, DMs. By examining these models, we can gain deeper insight into their performance and challenges in various applications, offering more targeted guidance for real-world use cases.

2.2. Comparison of different generative approaches

The landscape of generative modeling has evolved, with various techniques excelling in different domains. VAEs, GANs, and DMs are among the most widely adopted methods, and their flowcharts are shown in Fig. 1. Next, we compare these approaches along key dimensions, such as their architecture, training process, strengths, and weaknesses.

VAEs are probabilistic generative models that combine autoencoders with variational inference. As shown in Fig. 1a, the model consists of an encoder, which learns a function $q_\phi(z|x)$ to map input data x to a latent variable z , assumed to follow a Gaussian distribution. The decoder reconstructs the data using a conditional probability $p_\theta(x|z)$. VAEs are advantageous in providing probabilistic interpretation, enabling smooth interpolation in latent space, and effective anomaly detection. However, generated samples often appear blurry, and training can be challenging due to the need to balance reconstruction loss and regularization terms¹⁵.

GANs consist of two neural networks, a generator and a discriminator, that are trained in a competitive setting. The generator creates synthetic data, while the discriminator distinguishes between real and fake data. Both networks are trained simultaneously in a minimax game, and their specific flowchart is

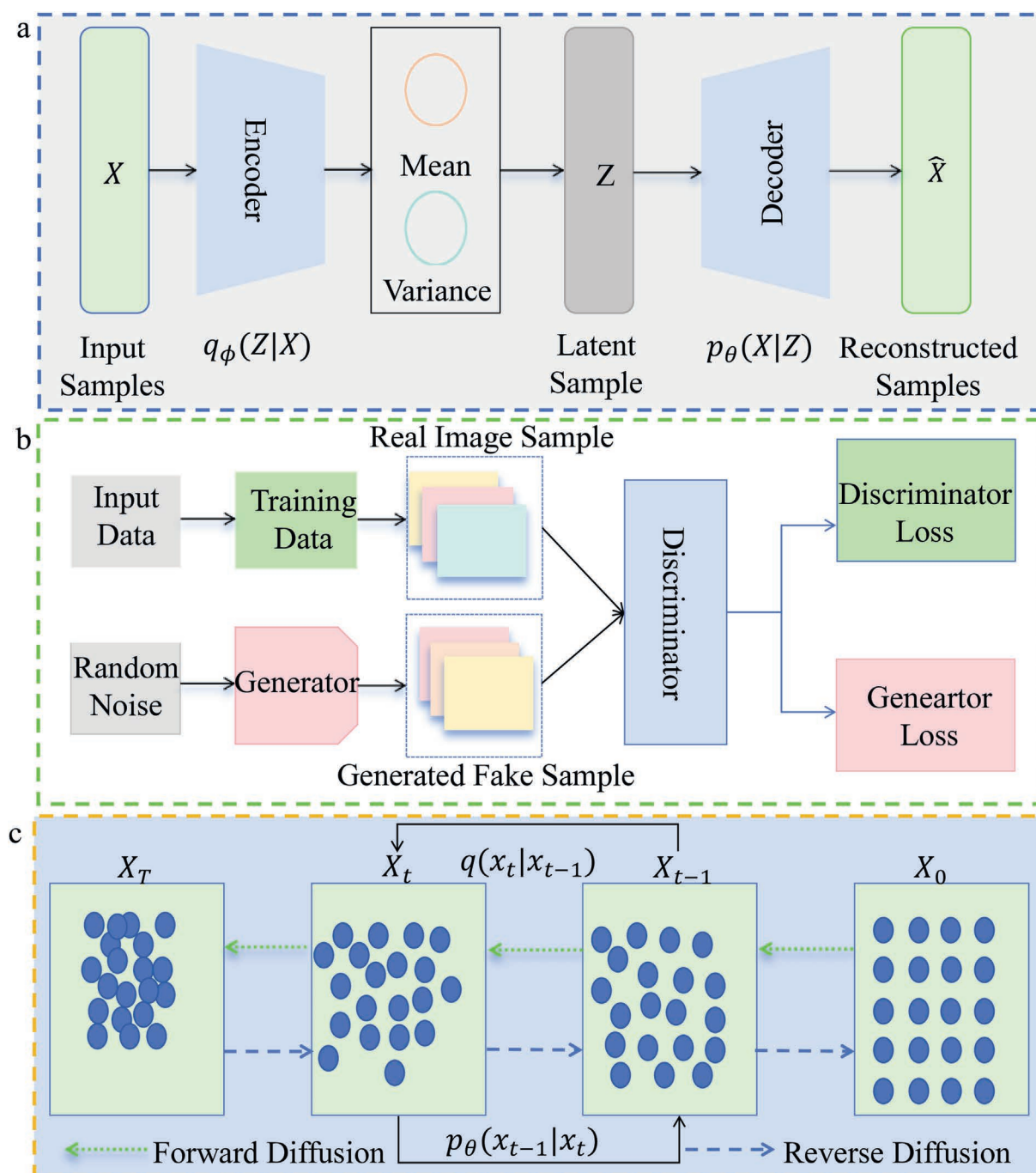


Figure 1 Flowcharts of three different generation methods for generating models. (a) Variational autoencoder extracts the encodings mainly through the encoder. In the hidden space Z , these encodings are treated as latent representations of the latent variables, and then passed to the decoder to produce an output close to the original input data. (b) The process diagram of generative adversarial network uses the generator model and the discriminator model to perform mutual game learning to generate output. (c) The diffusion model includes forward diffusion process and reverse diffusion process. The green line segment represents the forward diffusion. The blue line segment represents the reverse diffusion.

depicted in Fig. 1b. GANs excel at generating high-resolution, realistic images and are widely used in image generation, style transfer, super-resolution, and video synthesis. However, they face challenges such as unstable training, mode collapse, and vanishing gradients. Additionally, their lack of a clear probabilistic structure makes them more complex and difficult to analyze than VAEs¹⁶.

DMs are generative models that involve forward and reverse diffusion processes. In the forward process, real data x_0 is

progressively corrupted by noise, becoming Gaussian noise through a Markov chain. The reverse process attempts to restore the original data distribution from the noise, with $p_\theta(x_{t-1}|x_t)$ representing the conditional probability of the previous state¹⁷. The detailed flowchart is presented in Fig. 1c. DMs are highly effective at generating high-quality, realistic images, often surpassing GANs in sample quality. Their training process is more stable, avoiding issues like unstable training. However, DMs suffer from

slow sampling speed due to multiple refinement steps, which require significant computational resources and time, especially for high-dimensional data like images.

Of course, in addition to the above-mentioned classic generative models, several novel models have emerged, such as flow matching, Bayesian flow networks (BFNs), and generative flow models. Flow matching is a generative approach based on continuous normalizing flows, which directly regresses noise onto the vector field along the data's conditional probability path, avoiding the complex reverse diffusion process and accelerating training and sampling¹⁸. BFNs, introduced by Graves et al.¹⁹ in 2023, employ Bayesian inference to update the independent distribution parameters of noise samples and then input them into a neural network to generate new distributions. Unlike DMs, BFNs does not require explicit forward diffusion, and it supports both discrete and continuous data, excelling in image and language tasks. Generative flow models refer to a class of models based on normalizing flows, which use a series of invertible, differentiable transformations to map simple initial distributions (such as standard Gaussian) to complex data distributions, offering precise probability density modeling, expressiveness, and interpretability²⁰.

Generative modeling techniques have advanced rapidly, with VAEs, GANs, and DMs demonstrating strong performance in image synthesis, drug design, and molecular generation. Recently, emerging approaches such as flow matching, BFNs, and generative flow models have been introduced, offering advantages like faster training, uncertainty modeling, and exact likelihood estimation. Their applications have also expanded into molecular dynamics (MD) modeling, drug–target interaction prediction, and biomedical data analysis. To facilitate systematic comparison, Table 1 summarizes the advantages and disadvantages of six representative generative models and their main application areas, providing guidance for model selection and application.

2.3. Key theoretical concepts and mathematical foundations

Generative models are grounded in key mathematical concepts from probability theory, optimization, and information theory. These foundations are essential for understanding how generative models work and why some models are more suitable for specific tasks. Next, we discuss the relevant concepts and mathematical foundations.

The first concept is about variational inference and evidence lower bound (ELBO). The goal of these models is to maximize data likelihood, which is often computationally intractable. To address this, VAEs introduce the ELBO as a tractable lower bound to the likelihood. The core objective is to maximize the ELBO, and its formula can be written as:

$$\mathcal{L}(\theta, \phi; x) = E_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] - D_{KL}[q_{\phi}(z|x)||p(z)] \quad (1)$$

where $E_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)]$ is the reconstruction loss, in order to measure the decoder's ability to reconstruct data. $D_{KL}[q_{\phi}(z|x)||p(z)]$ is the regularization term that ensures the latent space distribution approximates a prior, typically Gaussian. VAEs optimize training using variational inference and ELBO, mapping latent space to data space. They generate complex samples and provide training stability, making them a key method in deep generative models³³.

The next key idea is GANs and the minimax game. GANs are based on an adversarial game between two networks where the

generator G and the discriminator D interact. The generator creates realistic data to deceive the discriminator, while the discriminator distinguishes real data from fake. The objective function of GANs can be expressed as:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}}(x)[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (2)$$

in this case, $p_{data}(x)$ is the distribution of the real data, and $p_z(z)$ is the distribution of the latent variables. GANs have advanced generative modeling through the adversarial interplay between G and D . The term $E_{x \sim p_{data}}[\log D(x)]$ helps the discriminator better distinguish real data, while $E_{z \sim p_z(z)}[\log(1 - D(G(z)))]$ encourages the generator to improve the quality of generated samples. They are capable of generating highly realistic images and data with wide applications³⁴.

The following concept is the diffusion process and reverse sampling. The DMs transform data into noise through the forward process and gradually recover the real data through the reverse process. This process is described using a Markov chain, allowing the generation of data from simple to complex, demonstrating its powerful generative capability⁶.

The mathematical basis of the diffusion process is the stochastic differential equation, which is used to describe the dynamics of the forward and reverse processes³⁵. Among them, the forward process is to gradually add noise to the data through the diffusion process, and this process is formulated as:

$$dx_t = f(x_t, t)dt + g(t)dW_t, \quad (3)$$

to clarify, x_t represents the state of the data at time t , $f(x_t, t)$ is the drift term, which dictates the deterministic direction of data evolution, $g(t)$ is the diffusion term, modulating the intensity of the stochastic variation, and dW_t corresponds to the increment of standard Brownian motion, capturing the inherent randomness in the process. The reverse process aims to recover data from noise, and its equation is represented as:

$$dx_t = [f'(x_t, t) - g(t)\nabla_x \log p_t(x_t)]dt + g(t)dW_t, \quad (4)$$

specifically, $\nabla_x \log p_t(x_t)$ is the score function, guiding the denoising process. Through the reverse process, DMs recover data that closely matches the original distribution, making them powerful tools in generative modeling.

The theoretical frameworks of probabilistic inference in VAEs, adversarial optimization in GANs, and noise modeling in DMs form the foundation of generative models. They support the development of more efficient models in fields like image and text generation³⁶, and are also transforming applications in biomolecular sciences, particularly drug discovery and protein engineering².

While these theoretical constructs establish the bedrock of generative modeling, their true impact emerges when they are deployed to tackle real-world challenges. In the following section, we turn our attention to the applications of these techniques within biomolecular sciences, examining how VAEs, GANs, and DMs are being used to accelerate drug discovery, guide protein engineering, and drive innovative molecular design.

Table 1 Summary of several commonly used generative models.

Generative model	Advantages	Disadvantages	Main application domains	Ref.
Variational Autoen-coders	- Probabilistic framework with evidence lower bound optimization -Stable training -Continuous latent space	-Samples can be blurry -Difficulty capturing high-frequency details	-Small molecule and protein design -3D molecular Conformation embedding -Text generation and latent semantic Modeling	Wei et al. [15] Zhang et al. [21] Ochiai et al. [22]
Generative adversarial networks	-High-fidelity sample generation -Mature architectures Facilitate stable variants -Smooth latent-space interpolation	-Training instability -Mode collapse risks	-Molecular structure and reaction synthesis -Drug reaction outcome prediction -Text generation and style transfer	Maziarka et al. [23] Zhang et al. [24] Macedo et al. [25]
Diffusion models	-Fine, sharp sample quality -Stable optimization (no adversarial loss) -Flexible conditioning on properties	-Slow sampling due to many steps -3D molecular conformation and drug design	-Molecular graph diffusion modeling -Molecular graph diffusion design -High-quality image generation and restoration	Yang et al. [6] Lin et al. [26] Morehead et al. [27]
Flow matching	-Direct regression of data—noise vector field -Fast training convergence -Adaptive path selection	-Sensitive to path choice -Stability on complex distributions	-Conformer sampling -Molecular dynamics trajectory modeling -Inverse design of small molecules	Lipman et al. [18] Hassan et al. [28]
Bayesian flow Networks	-No explicit forward diffusion required -Support for discrete & continuous data -Unified likelihood ^{training}	-Expensive posterior computation -High sampling cost in large spaces	-Uncertainty-aware molecule generation -Bayesian estimation of binding modes -Probabilistic conformer ensembles	Wu et al. [19] Graves et al. [29] Song et al. [30]
Generative flow Models	-Invertible transforms with exact densities -Efficient sampling -High interpretability	-Challenging design of invertible layers -Limited multimodal flexibility	-Molecular density and reaction pathway Modeling -Property-to-structure inverse design in drug discovery -Speech generation and conversion	Zang et al. [31] Bengio et al. [32]

3. Applications in biomolecular sciences

3.1. Drug discovery and design: generating novel drug candidates

Drug discovery and design have traditionally been time-consuming and resource-intensive. Generative models, a branch of AI, have the potential to revolutionize this field. Techniques such as VAEs, GANs, and transformer-based architectures can design novel drug candidates with specific properties, including molecule design, property optimization, and target-specific drug discovery. These innovations accelerate and optimize drug development^{37,38}. Moreover, recent perspectives have suggested that generative AI may play an increasingly influential role throughout the drug discovery pipeline³⁹.

Scientists have combined generative model-based structure predictions with AlphaFold's capabilities to advance drug discovery. For example, in 2019, Grow et al.⁴⁰ proposed the generative network complex (GNC) platform, which utilized the SMILES autoencoder, deep neural networks, and a three-

dimensional (3D) generative network to generate new compounds, predict their properties, and screen drug candidates. The schematic diagram of this approach is presented in Fig. 2a. In 2023, Gao et al.⁴¹ advanced molecular generation by combining molecular fragments with hierarchical chemical graph representations, employing a multi-resolution graph VAEs for deep molecular generation. Their goal was to generate drug molecules with chemical diversity and potential bioactivity through deep learning. Additionally, in 2023, Wang et al.⁴² integrated deep generative models with stochastic optimization and showcased ChatGPT as a virtual assistant to enhance drug development efficiency. In 2024, Qi et al.⁴³ introduced perturbation-conditioned generative model, a perturbation-conditioned deep generative model trained on nearly 100 million high-throughput screening data. It predicted transcriptional responses of 175,549 compounds in 188 cell lines and has demonstrated effective candidate drug screening in multiple metabolic disease models. In 2024, Kanakala et al.⁴⁴ summarized recent advances in generative AI technologies, tools, and practical applications for small molecular drug design. The workflow of drug design is illustrated in Fig. 2b.

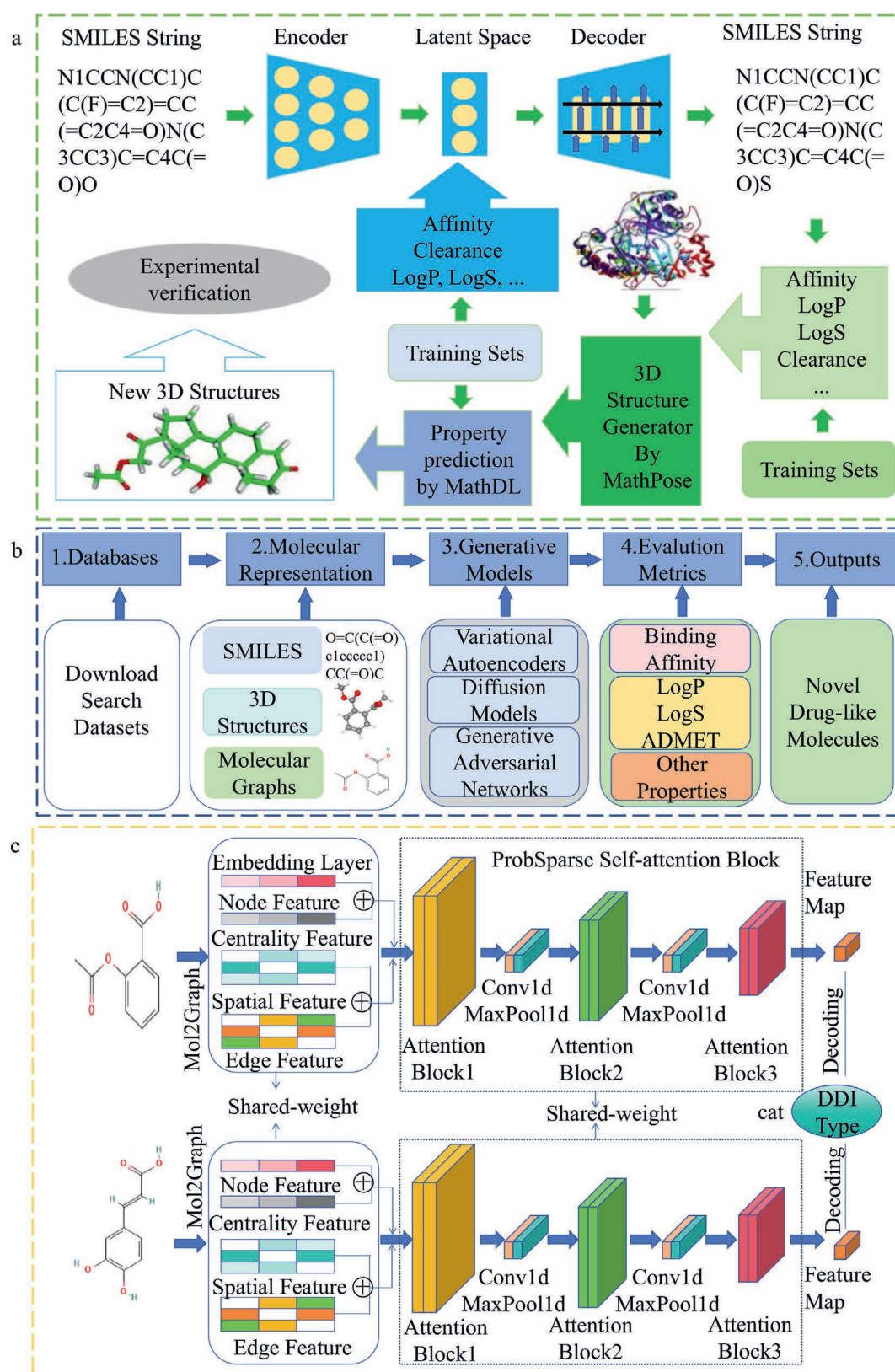


Figure 2 (a) A schematic illustration of a generative network complex. It consists of an autoencoder that takes SMILES strings into a drug-property regulated latent space, a regulated latent space, a LSTM based autoencoder, a multitask network for the evaluation of binding affinity (BA), partition coefficient (LogP), solubility (LogS), clearance, etc., a three-dimensional (3D) structure generator named MathPose, and MathDL^{45,46}, a refined 3D multitask druggable property predictor based on algebraic topology, differential geometry, and graph theory, to select new drug candidate structure⁴⁰. (b) This is an AI-driven drug discovery process diagram⁴⁴. The process begins with extracting relevant

Generative models have shown immense potential in drug discovery and design, particularly in accelerating chemical library screening, reducing costs, and exploring new chemical spaces. Despite challenges such as scaffold homogenization, inaccurate ADMET predictions for novel chemical types, conflicts in multi-objective optimization (MOO), target bias, and poor generalization to new scaffolds persist, generative models have made significant progress in generating novel drug molecules and optimizing lead compounds⁴⁴, and are expected to accelerate the discovery of new tracers and optimize their *in vivo* tracking performance and imaging contrast in the field of radio pharmaceuticals and molecular imaging⁴⁸. Looking ahead, improvements in datasets quality, the introduction of explainable AI, and the integration of multi-omics data could provide more comprehensive drug design solutions. Cross-disciplinary research and collaboration will be key to overcoming current limitations, accelerating precision drug development, and driving the revolution of AI-powered drug discovery.

3.2. Reaction prediction: learning reaction pathways and outcomes

Chemical reaction prediction is key for designing druggable molecules and macromolecules, advancing drug discovery and synthetic biology. Traditional methods rely on expert rules but struggle with the complexity of the chemical space. Recently, deep learning-based generative models have emerged, with recurrent neural networks, graph neural networks (GNNs)⁴⁹, and transformers driving rapid progress in reaction prediction⁵⁰.

The core task of generative models in reaction prediction is to learn the underlying patterns for generating reaction products. These models understand the language of chemical reactions through large-scale chemical reaction datasets and are able to predict reaction products given the reactants. For example, in 2019, Schwaller et al. developed the molecular transformer, a transformer-based model for reaction prediction. By treating reactions as a sequence-to-sequence problem, this model achieved state-of-the-art accuracy on benchmark datasets⁵¹. Additionally, in 2024, Jiang et al.⁴⁷ applied the Molorme framework to the transformer model, creating a drug–drug method interaction method that incorporates a lightweight attention mechanism, enhancing performance and reducing computational costs. The general framework is illustrated in Fig. 2c.

Generative models have shown exceptional capabilities in chemical reaction prediction, particularly in biomolecular science, such as enzyme catalysis and drug synthesis pathway design. Their advantages include scalability, flexibility (*e.g.*, GNNs for molecular graph modeling), strong generalization, and simplified prediction for non-experts. However, challenges remain in generating reaction pathways, including difficulties in capturing transition state information, prediction biases due to sensitivity to substrate stereochemistry, the tendency to produce unrealistic outcomes in multi-component complex reaction systems, and the inability to accurately model dynamic changes at catalyst active

sites. Despite these challenges, generative models are essential in modern chemical reaction prediction. Future research will focus on hybrid models, improving interpretability, and integrating with automated synthesis platforms to accelerate the discovery of new proteins and drugs, driving innovation in biomolecular research and biopharmaceuticals⁵².

3.3. Virtual screening and optimization of lead compounds

Virtual screening and lead compound optimization are key steps in drug discovery. In recent years, the introduction of generative models has significantly transformed the identification and optimization of drug candidates. While traditional virtual screening relies on methods such as molecular docking and dynamics simulations to assess binding affinity (BA), generative models enable the creation of novel molecules with desirable properties, providing a more efficient and diverse approach⁵³.

In generative models, especially those based on deep learning architectures such as VAEs, GANs, and reinforcement learning (RL) methods⁵⁴, there has been an increasing application in virtual screening and lead compound optimization⁵⁵. In 2023, Danel et al.⁵⁶ proposed a method combining generative models with molecular docking to accelerate the discovery of new drug candidates and improve virtual screening efficiency. In 2023, Chakraborty and Hasija explored how deep learning methods, particularly generative models such as VAEs and GANs, can efficiently explore potential drug molecular spaces and provide strong support for the drug discovery process⁵⁷. The structure of this model is depicted in Fig. 3a. In 2024, Lu et al.⁵⁸ proposed DynamicBind, a generative model for predicting protein-ligand complexes. By integrating MD with generative techniques, it effectively captures the flexibility and dynamics of binding conformations.

Generative models in virtual screening and lead optimization efficiently predict the activity of novel compounds, explore chemical spaces, and optimize molecular properties. However, they still face several unique challenges in virtual screening, such as generating molecules that often lack core scaffold innovation and tend to get trapped within known chemical spaces, insufficient prediction accuracy for rare structural active sites, and limited capability to assess risks related to potential off-target activities and metabolic transformation pathways. Although generative models can design molecules with strong biological activity, experimental synthesis remains challenging, and deep learning models incur high computational costs in large-scale screening⁶⁰. Future developments will improve the accuracy of generative models by integrating MD, quantum calculations, and docking techniques, while improving drug generation through diverse datasets that include biological activity and ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties⁹. Incorporating synthetic feasibility and real-time feedback optimization will drive the application of generated compounds in experimental synthesis, accelerating the drug discovery process⁶¹.

information from the collected dataset and representing molecules using SMILES, 3D structures, and molecular graphs. Then, generative models such as variational autoencoders, generative adversarial networks, and diffusion models are employed to generate candidate molecules. These candidates are evaluated based on properties such as BA, logP, logS, and ADMET (absorption, distribution, metabolism, excretion, and toxicity), ultimately leading to the identification of novel compounds with potential therapeutic value. (c) The overview of Molorme. The original drug graph goes through an embedding layer to form four feature embeddings representing the spatial structure of the drug. These features are processed by two ProbSparse self-attentive blocks, connected and fed to the decoder for final prediction. Drug–Drug Interaction (DDI) refers to the effects that occur when two or more drugs are taken together, potentially altering the effectiveness or side effects of one or both of the drugs⁴⁷.

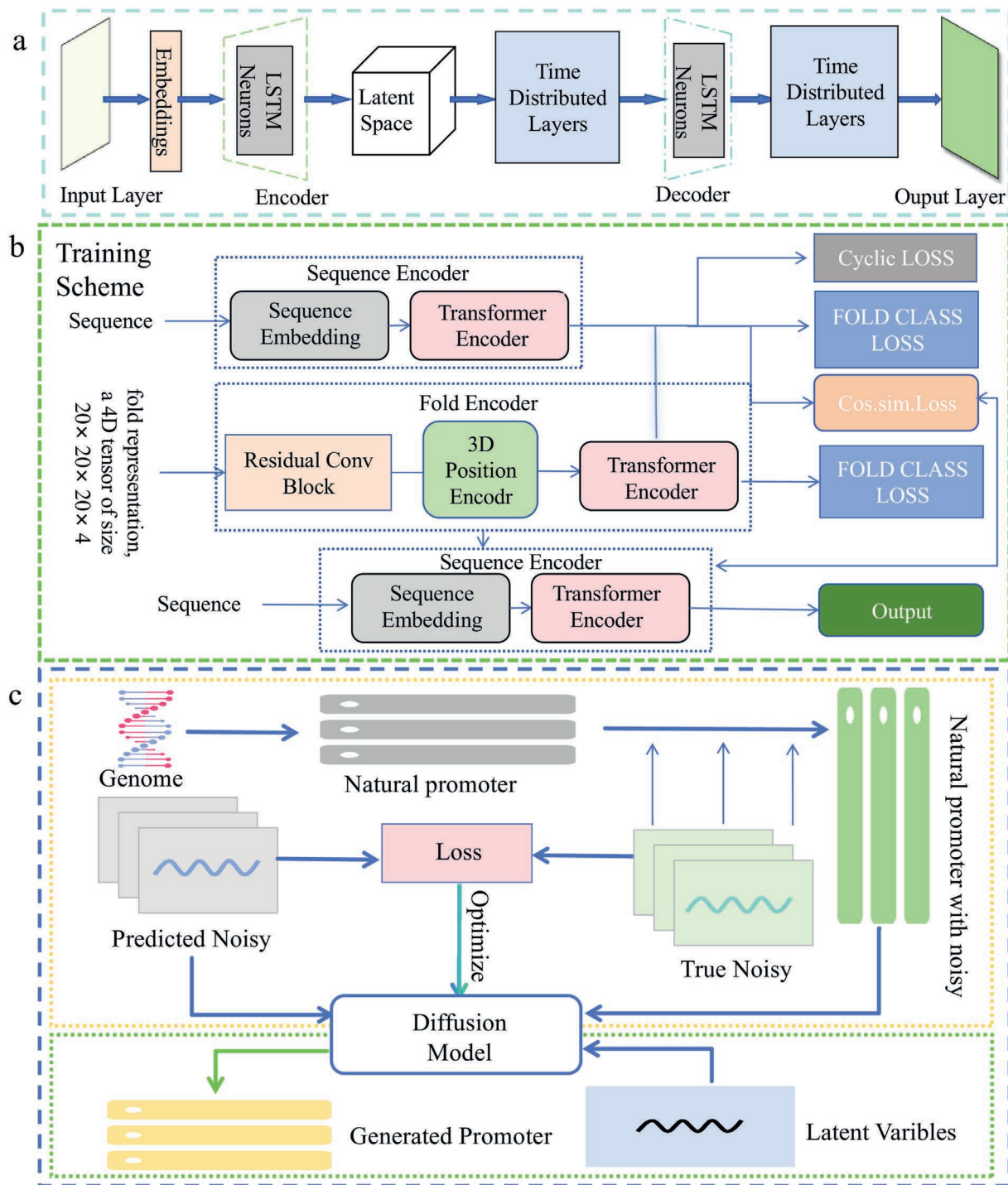


Figure 3 (a) The architecture makes use of embeddings to encode the characters within the text data. It employs a long short-term memory (LSTM) encoder to process the embedded data and map it to a latent space. Additionally, it uses an LSTM decoder to reconstruct the original text data from the latent space representation⁵⁷. (b) The training framework of the Fold2Seq model comprises three key components: the Sequence Encoder, the Fold Encoder, and the Sequence Decoder. The Sequence Encoder transforms the input sequence data into a high-dimensional representation. The Fold Encoder processes this sequence through folding to capture its structural information. Finally, the Sequence Decoder generates the output sequence based on the encoded representation. Together, these components optimize the model's sequence generation capability⁵⁹. (c) Workflow for designing synthetic promoters using a diffusion model. The yellow box represents the training phase. In this phase, random noise is progressively introduced into the natural promoter sequence until the sequence expression matrix is entirely transformed into a random matrix. The diffusion model is employed at each step to predict the noise added, based on natural promoters with noise, thereby facilitating the gradual generation of noise. The green box represents the generation phase. In this phase, latent variables with the same dimensions as the natural promoters are input. The trained diffusion model then denoises these inputs to generate synthetic promoters²⁶.

3.4. Protein structure prediction and design

Prediction and design of protein structures are crucial for understanding protein function and mechanisms, with applications in drug design, protein engineering, and directed evolution^{62,63}. Generative models have shown great potential in this field, as they can directly generate or optimize 3D protein structures from raw sequences, unlike traditional methods such as homology modeling and MD, which rely on existing data.

In generative models, particularly VAEs, GANs and deep neural networks, have been applied in protein structure prediction and design. These models not only predict the 3D structures of unknown proteins, but also generate entirely new protein molecules and optimize their structures to improve functional performance⁶². In 2021, Cao et al.⁵⁹ proposed Fold2Seq, a transformer-based generative model for the design of protein sequences based on target folds. It jointly learned protein sequences and 3D structures, capturing complex sequence-structure relationships, and showed strong performance in handling low-quality or incomplete input structures. The training scheme framework of this model is demonstrated in Fig. 3b.

Generative models in protein structure prediction offer advantages such as efficiency, diversity, and functionality-driven design. They can identify key factors affecting protein folding and function, generating a variety of protein structures. However, they also face several challenges, including difficulties in modeling transient intermediate conformations along folding pathways, limited accuracy in predicting the effects of post-translational modifications, and difficulty in capturing the conformational dynamics of membrane proteins and large multi-domain proteins. Future research should focus on building comprehensive databases for disease-related and species-specific proteins to enhance data support, and improving protein synthesis practicality for experimental validation. Integrating multi-task learning (*e.g.*, protein folding, function prediction, and drug design) will improve model accuracy, while incorporating biophysical and chemical principles will further enhance prediction reliability. Generative models have made significant progress in protein structure prediction, showing great potential in drug discovery and bioengineering².

3.5. Synthetic biology: designing genetic circuits and pathways

A key task in synthetic biology is to design genetic circuits and metabolic pathways for precise cellular control, with applications in biofuel production, drug development, and environmental sensors. Traditional methods are slow and costly, relying on iterative experimentation. Recently, generative models such as VAEs, GANs, DMs, and RL have been applied to automate the generation of new genetic circuits or pathways, accelerating design and reducing costs, thus advancing synthetic biology towards more efficient solutions⁶⁴.

Generative models can learn from large amounts of known gene sequences and metabolic network data to generate new design solutions. For example, GANs can be used to design new promoter sequences. In 2020, Zhang et al.²⁴ proposed HpGAN, a GAN-based model to generate specific sequences, while VAEs optimize key enzymes in metabolic pathways, facilitating biological sequence design and gene circuit optimization. In 2024, Lin et al.²⁶ introduced a DMs-based generative method for *de novo* design of synthetic promoter sequences, and the method's flowchart is available in Fig. 3c. This method used a stepwise denoising process to generate and optimize high-dimensional

promoter sequences, producing several synthetic promoters with high expression efficiency, which showed superior performance compared to traditional designs in experimental validation.

Generative models in synthetic biology offer efficiency, diversity, and customization by rapidly exploring design options for gene circuits and metabolic pathways. For example, HpGAN and DMs generate efficient promoter sequences using large-scale data²⁴. However, several challenges remain, such as the unpredictability of *in vivo* behavior due to context-dependent regulation, and difficulties in modeling nonlinear metabolic fluxes under fluctuating environments or cellular states. Future developments will focus on building comprehensive gene sequence and metabolic pathway databases, integrating data from various species and environments, and combining them with automated platforms to accelerate validation. Improving the model interpretability and incorporating MOO will balance factors such as stability, yield, and tolerance, driving the design of more practical gene circuits and pathways⁶⁵.

3.6. Understanding biological networks and interactions

Biological networks, including protein-protein interaction networks, gene regulatory networks, and metabolic pathways, are key to systems biology, helping to understand biological functions and responses to environmental conditions. However, their complexity and dynamic nature present challenges in analysis and modeling. Recently, deep learning-based generative models have shown great potential in modeling networks, predicting interactions, and generating novel hypotheses. These models capture intricate relationships and can generate new network structures, predicting previously undiscovered phenomena, even without comprehensive experimental data, offering powerful tools to advance systems biology research⁶⁶.

Generative models have increasingly emerged at the forefront of biological network research, with several pivotal studies demonstrating their effectiveness in network prediction and highlighting their potential to discover unknown interactions and generate hypothetical networks. In 2022, Gomari et al.⁶⁷ showed that VAEs can learn nonlinear transferable latent representations from metabolomics data, providing a useful tool for analyzing metabolic networks and disease-related changes. In 2024, Zhang et al.²¹ proposed a VAE-based model with an attention mechanism to predict drug-protein interactions, using deep learning to identify interaction patterns from large-scale drug and protein data, thus accelerating drug discovery and target identification. The schematic diagram of this model is provided in Fig. 4.

Generative models hold great potential in biological network research, enabling faster network design, structural optimization, and prediction of unknown biological relationships, thus reducing experimental burdens. They are particularly useful in drug discovery, precision medicine, and synthetic biology, where they can identify new drug targets, predict drug interactions, customize treatment plans, and design biological networks, accelerating bioengineering advancements. However, their broader application is hindered by challenges such as high uncertainty due to sparse or incomplete interactome maps, and limited capacity to model dynamic, context-dependent network changes. Despite these obstacles, ongoing technological progress will establish generative models as essential tools in systems biology, advancing the study of complex biological interactions and driving breakthroughs in related fields^{66,68}.

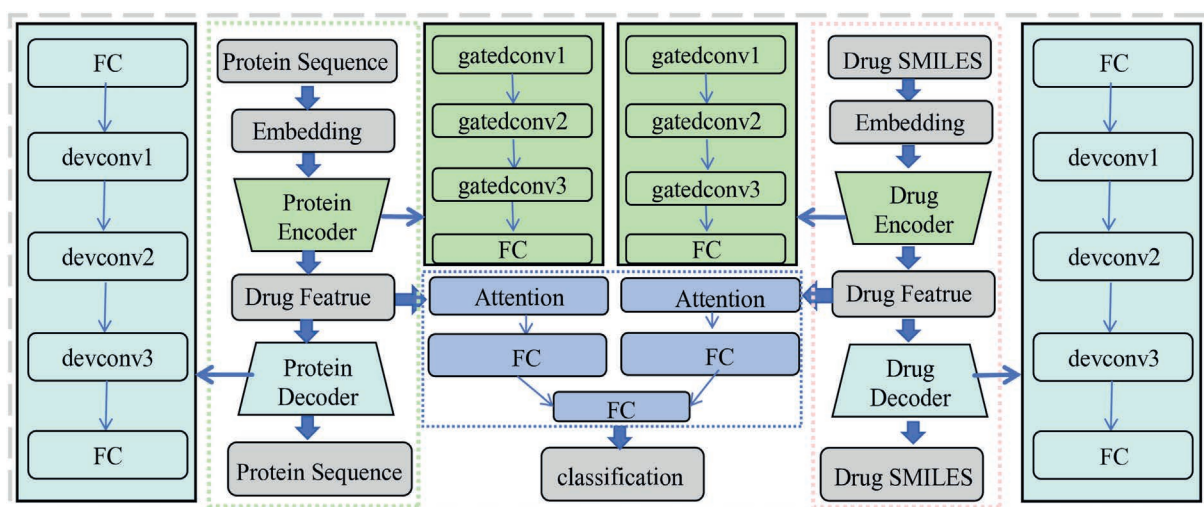


Figure 4 The model comprises three key components: the encoder, the decoder, and the prediction module. Both the encoder and decoder are used to predict drug-protein interactions (DPI). The architecture extracts features from drugs and proteins, applies an attention block to highlight important parts, and passes the output to a fully connected layer for DPI prediction. The feature extraction for drugs and proteins is similar. Text sequences representing drugs and proteins are converted into numerical vectors and then transformed into embedding matrices. The encoding layer contains three Gated Convolutional Neural Networks (GatedCNNs), with Rectified Linear Unit activation after each layer, and a max-pooling layer follows the third GatedCNN to compress the extracted features²¹.

3.7. The integration of generative models with quantum mechanics and molecular dynamics

The integration of generative models with quantum mechanism and MD holds great potential in biomolecular sciences. They accelerate molecular design and property prediction, reducing the computational cost of quantum calculations, and provide efficient initial structures or conformational evolution pathways for MD simulations⁶⁹. This combination drives molecular discovery, enabling the design of new molecules with specific functionalities.

Quantum mechanism methods play a crucial role in predicting molecular structures and properties, but their high computational costs and complexity limit their application. Generative models can reduce computational costs and improve predictive accuracy through adaptive basis set generation and uncertainty quantification⁷⁰. In 2021, Li et al.⁷¹ proposed a new quantum GAN mechanism with a hybrid generator and classical discriminator for drug discovery, which supported various number of qubits and quantum circuit layers. Their model with less than 20% of the original parameters can learn molecular distributions as efficiently as its classical counterpart.

In MD simulations, generative models can quickly generate initial conditions and simulate molecular conformation evolution, reducing computational time and enhancing efficiency. They can predict molecular behavior under various conditions and help simulate complex systems. The integration of generative models with quantum mechanics and MD enables efficient conformational sampling but faces domain-specific challenges including potential energy surface inaccuracies and timescale mismatches between MD simulations and generative sampling timescale mismatches between MD simulations and generative sampling⁶⁹.

3.8. Development of hybrid models combining generative techniques with reinforcement learning

The integration of generative models and RL offers an efficient solution for molecular design and optimization. Generative

models learn chemical space distributions to generate candidate molecules, but struggle to directly optimize target properties⁶⁸. RL, on the other hand, optimizes multi-objective properties through reward mechanisms but faces challenges in exploring high-dimensional and discontinuous spaces. Combining these approaches allows generative models to generate initial molecular structures while RL optimizes target properties, achieving a synergistic effect in molecular generation and optimization³⁷, with significant potential in drug design and protein engineering.

For example, in 2021, Hu⁷² applied molecular deep Q-networks to optimize molecular structures for multi-objective problems by treating the molecular construction process as a Markov decision process and using deep Q-learning to guide step-wise modifications. This approach enables the agent to sequentially add atoms or bonds to maximize long-term rewards, demonstrating flexible control over molecule generation. In another study, Koshino et al.⁷³ developed a molecular GANs framework that integrates GANs with RL to generate molecules with specific properties. By combining adversarial training with property-based reward signals, this model facilitates goal-directed molecule generation, optimizing features like drug-likeness and synthetic accessibility. Additionally, methods using pretrained generative models and RL mechanisms have shown success in drug design. For example, variational graph autoencoder (VGAE) with Monte Carlo tree search (MCTS), proposed by Iwata et al.⁷⁴ in 2023, combines a VGAE-MCTS to explore the chemical space and optimize molecular properties. The VGAE captures latent molecular representations, while the MCTS navigates the space efficiently through reward-guided search. Similarly, graph-based autoregressive flow, introduced by Han et al.⁷⁵, employs a sequential generation strategy where molecular graphs are built node by node. RL is used to fine-tune the generation process based on the reward of stability and chemical property. Key challenges for hybrid models include discrete action space limitations in molecular graph construction and delayed reward signals in multi-step generation⁷⁴, requiring specialized optimization approaches.

Future research could address these issues through efficient search and optimization strategies, automated reward function design, and collaborative training methods. Incorporating synthetic feasibility prediction models and Pareto-based optimization methods could enhance MOO capabilities⁷⁶. Cross-disciplinary collaboration and quantum-assisted computation could further improve efficiency and applicability in drug discovery and biological science.

3.9. Other applications

Although generative models have made significant progress in drug discovery and protein structure prediction, their applications extend far beyond these areas. With ongoing research, generative models demonstrate enormous potential in various biomedical fields, particularly in epigenetics⁷⁷, microbiome modeling⁷⁸, clinical practice⁷⁹, as well as mRNA vaccine design and nucleotide modification optimization⁸⁰.

Generative models are increasingly being applied in biomedicine, showing significant potential in epigenetics, microbiome modeling, clinical practice, as well as mRNA vaccine design and nucleotide modification optimization. In 2025, Mandal et al.⁷⁷ proposed the supervised GANs method, which effectively combines supervised learning and GANs to apply to labeled genomic or epigenomic data, providing an innovative solution for labeled data analysis in genomics and epigenomics. In 2023, Choi et al.⁷⁸ employed a deep learning method based on GANs to propose DeepMicroGen, a tool to impute missing values in longitudinal microbiome data while preserving the diversity of microbial communities and time-series characteristics. In 2025, Wang et al.⁷⁹ proposed a self-improving generative model that enhances the quality and diversity of generated images. It was successfully applied to clinical decision support, disease prediction, and personalized treatment, providing new solutions for medical imaging and precision medicine. In 2024, Tang et al.⁸⁰ presented the first deep generative framework specifically designed for m1 Ψ -modified mRNAs, termed smart 5 untranslated region (UTR). The model accurately predicted and generated optimized 5'UTR sequences, enhancing mRNA translation efficiency. These advancements highlight the transformative potential of generative models in biomedicine, paving the way for more accurate data analysis, improved medical imaging, and better personalized healthcare solutions. As research continues, the scope of their applications is likely to expand, offering innovative tools for precision medicine and disease management.

Despite the significant potential of generative models in biomedicine, their clinical applications still face some challenges, including the inability to capture patient-specific heterogeneity and inter-individual variability. There is also a risk of generating biologically implausible or unsafe candidates, along with difficulties in designing mechanistic validation and conducting prospective clinical trials. Moreover, ensuring the reliability and accuracy of these models in real-world applications remains a critical challenge.

4. Challenges and limitations

4.1. Data quality and availability: addressing biases in training datasets

The application of generative models in biomolecular sciences relies on high-quality training data, but such data often suffer from

uneven distribution, noise, and biases. The complexity of molecular structures, especially large macromolecules, makes it difficult to standardize molecular features, affecting data quality. Additionally, the limited availability of molecular samples challenges the development of accurate ML models. Researchers often use physical models and theoretical methods (*e.g.*, molecular mechanics, dynamics, and quantum mechanics) to generate molecular properties and expand datasets⁷. However, these methods have limitations, and biases in simulated data can still affect model accuracy and generalizability. Therefore, ensuring data quality, diversity, and addressing biases are key challenges in applying generative models in biomolecular sciences⁶⁸. Publicly available datasets such as MoleculeNet and therapeutics data commons offer standardized benchmarks and curated molecular data, helping alleviate issues of datasets bias and scarcity. These data sets support tasks such as prediction of molecular properties, estimation of toxicity and modeling of drug–target interaction, making them essential resources for training robust generative models.

Data biases during training can significantly impact model performance and output quality. These biases arise from limitations in training datasets, often dominated by certain molecule categories, restricting the model's ability to generate novel molecules⁸¹. Strategies like data augmentation, diversity-focused objective functions, and fairness constraints can help improve data diversity and balance. These approaches aim to improve model generalization and improve the quality of generative models in biomolecular sciences⁹.

Balancing innovation and feasibility is crucial, as generated molecules must meet drug design requirements and be experimentally validated. Ensuring the diversity and comprehensiveness of the training datasets is essential for the success of generative models in drug discovery⁸².

4.2. Interpretability of generative models in biological contexts

As the interpretability of generative models in ML gains increasing attention, it is crucial to explore how these models can be made understandable, especially in complex fields such as biology⁸³. Currently, there is no unified standard, and traditional methods like disentanglement have limitations in practical applications. This highlights the need to evaluate the interpretability of generative models in biological contexts, as understanding the decision-making process can provide valuable insights for applications such as drug discovery.

One method to enhance interpretability is through the use of attention mechanisms, which have been widely applied in natural language processing⁸⁴. In drug discovery, attention mechanisms help identify key features of molecular structures, improving the transparency and interpretability of generative models. Visualization techniques also help improve interpretability, allowing researchers to intuitively understand the decision making process of generative models in drug design⁸⁵, revealing how the model generates drug-like molecules while focusing on key biological concepts during the process.

Integrating generative models with quantum mechanism and MD further enhances interpretability⁶⁸. Several explainable AI frameworks have emerged for generative modeling, such as feature attribution methods and graph explanation tools, which help clarify how molecular features influence generation outcomes⁸⁶. For instance, explainable VAE-based models have been applied in identifying latent variables related to toxicity or

solubility, facilitating property-driven molecule design. This integration not only optimizes molecular structures but also helps researchers better understand the decision-making process behind generation, improving transparency and verifiability in drug discovery⁸⁷.

4.3. Resource limitation

In recent years, generative models have experienced a leap from hundreds of millions to trillions of parameters (*e.g.*, OpenAI's GPT-3, GPT-4, Google's Gemini series, DeepSeek's DeepSeek series, etc.), significantly improving generation quality and broadening their range of applications through the expansion of parameter sizes and training data⁸⁸. Beyond traditional text generation, multimodal models for drug design, protein design, and protein structure prediction (such as ChemBERTa, MolGPT, GraphDF, AlphaFold, etc.) showcase the broad applicability of generative technology across various domains. Architectural innovations like transformers, DMs, and GANs have further improved performance, but this comes with increased computational demands and more complex training processes.

Resource demands vary significantly by model type, with VAEs being memory-efficient but requiring extended training periods⁸⁹, while DMs incur high FLOPs per sampling step⁹⁰, creating distinct optimization requirements. For example, training GPT-3 requires millions of dollars and spans weeks or even months⁹¹. In fields like biological and molecular generation, DMs (*e.g.*, BioDiffusion or MatDiff) necessitate multiple iterations in latent space, resulting in exponential increases in flops requirements, further escalating computational demands. Consequently, the multi-step iteration in DMs not only

exacerbates computational burdens but also challenges real-time performance and system response speed. To address this, more efficient optimization strategies and computational architectures are needed to balance performance with resource consumption⁹⁰.

5. Future directions

5.1. Large language model assisted generation

Rapid advancements in AI are transforming pharmaceutical research. Foundation models, including large language models (LLMs) like GPT-4, can analyze vast literature, propose hypotheses, optimize molecular structures, and support complex decision-making. Their use in drug discovery aims to improve efficiency, reduce costs, and speed up the identification of new therapeutic compounds. Unlike the lengthy and expensive traditional pipeline, AI-driven methods, particularly generative models, offer a data-centric framework. They complement existing computational and experimental approaches and reshape pharmaceutical innovation⁹².

LLMs, as a subclass of foundation models specialized in processing textual and symbolic information, play an important role in drug development. For example, in 2023, Wang et al.⁴² incorporated GPT-4 into the full drug design workflow. The system facilitated idea generation, methodological analysis, and programming support. They developed a randomly generated network complex using AI-driven molecular design, combined with Langevin equation simulations and DMs analysis. Debugging and optimization were implemented throughout the process. The framework integrated multiple components including a seq2seq

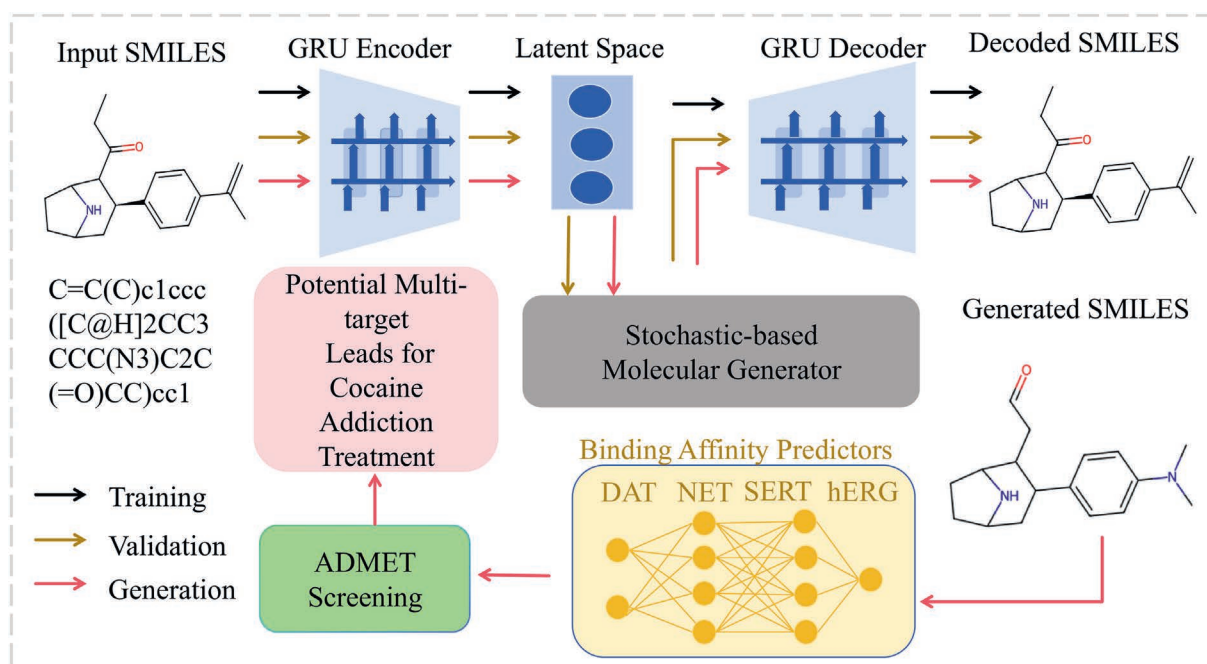


Figure 5 Workflow of the stochastic-based generative network complex (SGNC)⁴². ChatGPT was extensively involved in the building process of the SGNC. Dark arrows show the training process, brown arrows indicate the validation process, and red arrows show the generation process. The SGNC comprises five primary structures: a sequence-to-sequence AutoEncoder in blue, binding affinity predictors in yellow, a stochastic-based molecular generator in grey, ADMET (absorption, distribution, metabolism, excretion, and toxicity) screening *via* ADMETlab 2.0 in green, and potential multiple targets for cocaine addiction treatment in pink.

autoencoder, affinity prediction models, a Langevin dynamics-based molecule generator, and ADMET screening protocols. Ultimately, this approach yielded 15 promising drug candidates for combating cocaine addiction. The workflow for generating a network complex is visualized in Fig. 5.

LLMs show significant potential in drug development, yet challenges persist. Future efforts should focus on integrating LLMs with molecular generation models like DMs and VAEs. RL could be employed to optimize molecular structures, balancing binding scores, biological activity, and synthetic feasibility. Enhancing fact-checking capabilities through cross-validation with authoritative databases (*e.g.*, PubChem, ChEMBL, AlphaFold) will improve reliability. Robotic lab automation could streamline automated synthesis and testing workflows. Additionally, domain-specific fine-tuning of models may further enhance predictions of drug-likeness, toxicity, and ADMET properties⁹³. The case of anti-cocaine addiction drugs shows that LLM-assisted generation can significantly accelerate drug development, but fully autonomous AI-driven drug development still needs further breakthroughs in hybrid modeling, RL and experimental verification.

5.2. Generation based on foundation models

By learning complex patterns within vast datasets, generative models can generate biologically relevant structures or sequences that possess desired pharmacological activities, and combining these generative approaches with foundation models can further elevate their capabilities. Foundation models, such as large-scale

GNNs and molecular transformers pretrained on diverse chemical and biological datasets, can encode a broader range of domain-specific knowledge and effectively capture generalizable molecular representations. Integrating generative models with foundation models improves their predictive power, stability, and generalization abilities, thus producing more meaningful and experimentally viable biomolecular designs with fewer computational resources⁹⁴.

The GNC method proposed by Gao et al.⁹⁵ in 2020 employed autoencoders and deep neural networks to optimize multiple chemical properties through gradient descent to automatically generate optimized drug molecules. This method has generated more than 1000 new molecules and provided alternative molecules that were superior to the original drugs for 8 marketed drugs (such as Ceritinib and Ribociclib), which has wide applications in small molecule drug design, MOO, and personalized drug discovery. Its schematic illustration is given in Fig. 6.

In the future, generative models built up on domain-specific foundation models are expected to become even more sophisticated and integral to the drug discovery pipeline. The main directions include combining more powerful foundation models (such as molecular transformers or graph-based transformer architectures) to improve the molecular characterization ability and the smoothness of the generation process, combining experimental data to optimize the actual feasibility of molecules, developing efficient MOO and improving generation algorithms (such as explainable AI models, quantum computing acceleration)⁹⁶. Additionally, the conversion of peptides into small molecules also provides a feasible path for generative model-driven molecular

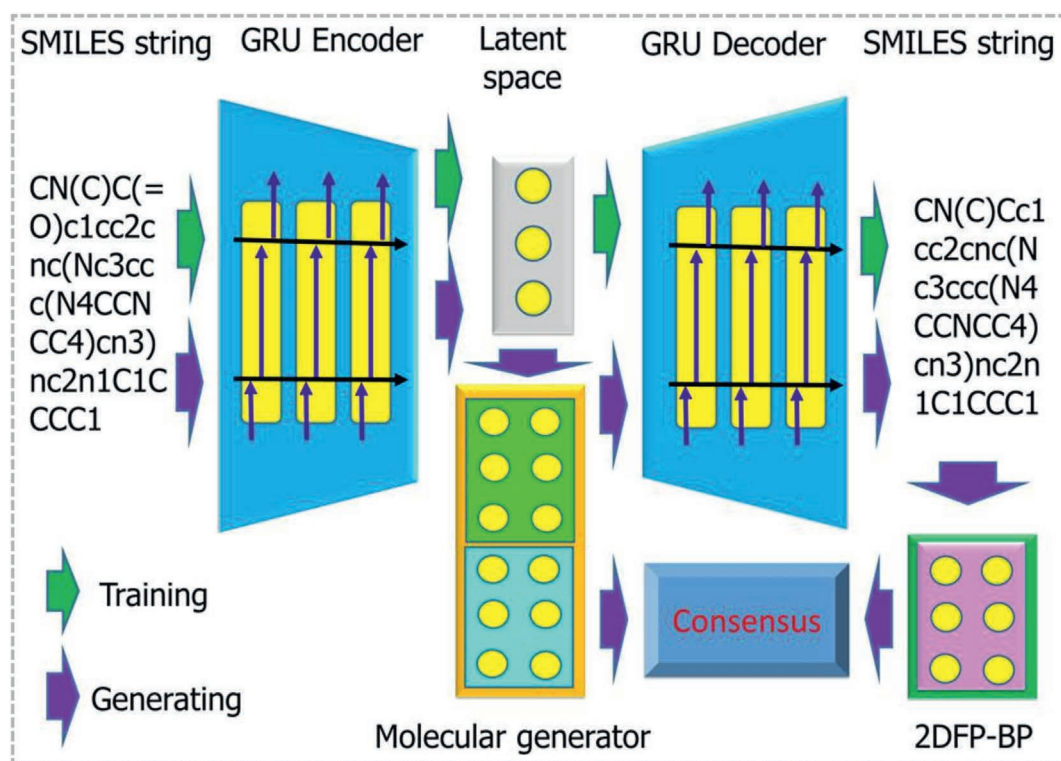


Figure 6 Schematic diagram of the generative network complex. First, the input SMILES string is encoded into the latent space. Then, the molecule generator optimizes the molecule in the latent space and reconstructs it into a new SMILES string through the decoder. Finally, the properties or activities of the generated molecules are evaluated by consensus, utilizing modules such as 2DFP-BP, thereby achieving a complete closed loop from training to molecule generation. The green arrows indicate the training process, while the purple arrows indicate the generation process⁹⁵.

design to convert candidate molecules into synthesizable and administrable drugs⁹⁷. These advances will promote computer-aided drug design based on technologies such as GNC, assist in new-drug development and precision medicine, and further expand the application potential of AI in drug discovery.

5.3. Expanding the role of generative models in multi-modal biomolecular sciences

Generative models play a key role in multi-modal biomolecular sciences by integrating diverse data types, such as molecular structures, chemical properties, and phenotypic data. These models capture relationships between data sources, facilitating data fusion and overcoming the limitations of single-modal data⁹⁸. In drug discovery and protein design, they optimize properties like drug efficacy or protein stability while generating novel molecular structures. They also enable cross-modal learning for better predictions of drug potency, safety, and efficacy⁹⁹. Additionally, generative models help fill gaps in datasets by predicting missing data, such as chemical properties or experimental outcomes. These capabilities accelerate discovery, improve predictive accuracy, and drive innovation in biomolecular sciences, aiding drug development and molecular optimization¹⁰⁰.

However, their application faces challenges, such as the heterogeneity of multi-modal data, which complicates unified representation and processing, affecting model performance¹⁰¹. Capturing complex, nonlinear relationships between modalities

requires models to integrate data while preserving important features, demanding advanced model design and training¹⁰². The quality and completeness of multi-modal data, often noisy or imbalanced, also affect model generalization, and high-quality annotated datasets are scarce in biomolecular sciences. Complex architectures for data integration increase computational costs and may lead to training instability. Furthermore, ensuring the reliability and interpretability of generated outputs, especially in MOO, adds to the challenges. Lastly, the high computational demands of training multi-modal generative models limit their accessibility to resource-limited labs and institutions.

Future progress should focus on building high-quality multi-modal datasets, innovating model architectures, and ensuring reliable validation of results¹⁰³. Developing lightweight models for resource-constrained environments will expand accessibility, accelerating breakthroughs in drug development, protein science, and precision medicine.

5.4. Multi-objective generation

Multi-objective generation is a method in drug design that optimizes multiple molecular properties simultaneously, balancing factors such as multi-target BA, pharmacokinetics, toxicological properties, ADMET, and molecular novelty and synthesizability. Recently, the integration of generative models has advanced MOO, enabling efficient chemical space exploration and automated generation of potential drug molecules that meet multiple

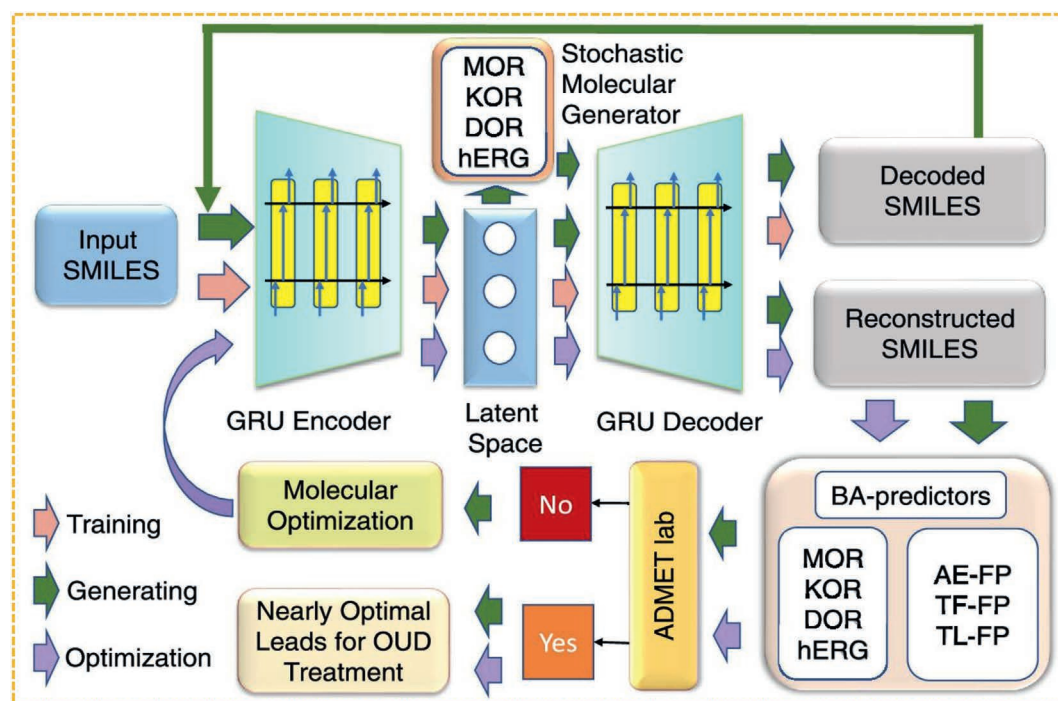


Figure 7 Schematic illustration of multitarget stochastic generative network complex, used to design novel compounds for the treatment of opioid use disorder (OUD). Three different paths, *i.e.*, model training, molecular generation, and lead optimization are colored in pink, green, and purple, respectively. The pink path represents the model training process of the pretrained autoencoder network. The blue path signifies the compound generation process: the SMILES string of a given compound undergoes encoding in the encoding network, and its latent space representation is then fed into the stochastic molecular generator. The generated compounds are subsequently processed through the decoding–encoding network. The resulting molecules that have stable SMILESs are evaluated for their BA and ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties. Molecules exhibiting desired BA and ADMET properties are regarded as nearly optimal leads. In cases where the properties are not satisfactory, molecular optimization (the purple path) is performed to generate more potential druglike compounds⁹⁹.

constraints¹⁰⁴. Unlike traditional drug discovery methods, generative models driven by MOO use deep learning to analyze existing data distributions, ensuring the chemical validity of generated compounds while speeding up the process of creating and optimizing high-quality drug candidates. This approach significantly improves the efficiency and success rates of drug development.

In the field of molecular generation, Feng et al.⁹⁹ proposed a multi-target stochastic generation network complex to optimize molecules targeting opioid receptors (MOR, KOR, DOR). This method combined GNC with MOO, integrating stochastic differential equations and autoencoder-based latent space optimization to enhance molecular diversity. It balanced BA, ADMET, and molecular novelty, which ensured high activity, favorable pharmacokinetics, and innovative structures. Key innovations included latent space optimization with DMs, using Langevin equations for molecular evolution, and employing pretrained autoencoders for filtering unstable candidates. Multi-step optimization improved drug activity and drug-likeness, offering new insights for opioid drug design. The schematic diagram of this multi-target stochastically generated network complex is presented in Fig. 7.

Future multi-objective generation in drug discovery will advance by integrating GNNs with RL to enhance model robustness-capturing molecular topology more accurately and dynamically optimizing generation strategies¹⁰⁵. Additionally, the multi-target stochastic generation network complex method is expected to extend to drug design for neurodegenerative diseases, cancer, and infectious diseases, aligning with the trend toward polypharmacology. Improved ADMET predictions through self-supervised learning and high-throughput virtual screening will further accelerate lead compound selection and optimization, ultimately speeding up new drug development¹⁰⁶.

6. Conclusions

Generative models are expected to revolutionize biomolecular sciences by offering innovative strategies for designing, predicting, and optimizing molecular structures in biochemistry, molecular biology, and protein science¹⁰⁷. By generating new molecular entities from learned representations, these models hold significant promise in drug discovery, reaction prediction, protein structure design, and the development of molecules with customized properties. They simplify the discovery process and minimize dependence on conventional trial-and-error approaches, facilitating a more efficient exploration of extensive chemical and biological spaces.

Despite these advancements, challenges in data quality and model interpretability must be addressed to ensure the responsible use of generative models in molecular research⁹. Integrating models such as VAEs, GANs, and DMs with techniques such as quantum chemistry, MD, and RL can improve prediction accuracy.

Hybrid and multi-modal models, combined with experimental feedback, will further improve the practical utility of generative models in molecular challenges⁶⁸.

Future trends in generative modeling will increasingly draw on automated generation using LLMs to autonomously encode chemical and biological knowledge, enabling simplified molecular discovery workflows. Simultaneously, multi-objective generative approaches will gain prominence, facilitating the concurrent optimization of multiple molecular properties, thus efficiently navigating complex chemical spaces for practical applications in drug discovery, bioengineering, and advanced protein design. Another trend is to utilize advanced mathematics, such as

algebraic topology, combinatorial theory, and differential geometry, to regulate molecular generation and enhance structure prediction⁴⁶.

By overcoming current limitations and using advances in ML and computational biology, these models can drive progress in personalized medicine, synthetic biology, and sustainable protein design, ultimately reshaping biomolecular sciences and accelerating innovation in medicine, material, and biology science⁹.

This review highlights the growing importance of generative models and calls for interdisciplinary collaboration to address challenges such as data quality and model interpretability, unlocking their full potential in advancing drug discovery, protein science, and synthetic biology.

Acknowledgments

This work was supported in part by National Science Foundation grants DMS2052983 and IIS-1900473, and MSU Research Foundation (USA). The work of Huahai Qiu and Bengong Zhang was supported by the National Natural Science Foundation of China under Grant No. 12271416 and No. 12371500, respectively.

Author contributions

Jian Jiang was responsible for study design, manuscript writing, reviewing, and editing. Daixin Li and Guilin Wang collected relevant literature and prepared the initial draft. Nicole Hayes, Yazhou Shi, Huahai Qiu, Bengong Zhang reviewed and edited the manuscript. Tianshou Zhou, and GuoWei Wei supervised, conceptualized the manuscript, and revised the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Noé F, Tkatchenko A, Müller K-R, Clementi C. Machine learning for molecular simulation. *Annu Rev Phys Chem* 2020;**71**:361–90.
2. Qiu Y, Wei GW. Artificial intelligence-aided protein engineering: from topological data analysis to deep protein language models. *Briefings Bioinf* 2023;**24**:bbad289.
3. Qiu Y, Wei GW. Persistent spectral theory-guided protein engineering. *Nat Comput Sci* 2023;**3**:149–63.
4. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Commun ACM* 2020;**63**:139–44.
5. Kingma DP, Welling M. An introduction to variational autoencoders. *Found Trends Mach Learn* 2019;**12**:307–92.
6. Yang L, Zhang Z, Song Y, Hong S, Xu R, Zhao Y, et al. Diffusion models: a comprehensive survey of methods and applications. *ACM Comput Surv* 2023;**56**:1–39.
7. Dou BZ, Zhu ZL, Merkurjev E, Ke L, Chen L, Jiang J, et al. Machine learning methods for small data challenges in molecular science. *Chem Rev* 2023;**123**:8736–80.
8. Jørgensen PB, Schmidt MN, Winther O. Deep generative models for molecular science. *Mol Inform* 2018;**37**:1700133.
9. Bilodeau C, Jin W, Jaakkola T, Barzilay R, Jensen KF. Generative models for molecular discovery: recent advances and challenges. *Wiley Interdiscip Rev Comput Mol Sci* 2022;**12**:e1608.
10. Castelli M, Manzoni L. Generative models in artificial intelligence and their applications. *Appl Sci* 2022;**12**:4127.

11. Harshvardhan G, Gourisaria MK, Pandey M, Rautaray SS. A comprehensive survey and analysis of generative models in machine learning. *Comput Sci Rev* 2020;**38**:100285.
12. Baum LE, Petrie T. Statistical inference for probabilistic functions of finite state markov chains. *Ann Math Stat* 1966;**37**:1554–63.
13. Pu Y, Gan Z, Henao R, Yuan X, Li C, Stevens A, et al. Variational autoencoder for deep learning of images, labels and captions. *Adv Neural Inf Process Syst* 2016;**29**:2352–60.
14. Ooi KB, Tan GWH, Al-Emran M, Al-Sharafi MA, Capatina A, Chakraborty A, et al. The potential of generative artificial intelligence across disciplines: perspectives and future directions. *J Comput Inf Syst* 2025;**65**:76–107.
15. Wei R, Garcia C, El-Sayed A, Peterson V, Mahmood A. Variations in variational autoencoders—a comparative evaluation. *IEEE Access* 2020;**8**:153651–70.
16. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. *IEEE Signal Process Mag* 2018;**35**:53–65.
17. Kingma D, Salimans D, Poole B, Ho J. Variational diffusion models. *Adv Neural Inf Process Syst* 2021;**34**:21696–707.
18. Lipman Y, Chen RT, Ben-Hamu H, Nickel M, Le M. Flow matching for generative modeling. In: *11th international conference on learning representations, ICLR*; 2023.
19. Graves A, Srivastava RK, Atkinson T, Gomez F. Bayesian flow networks. *arXiv* 2023. Available from: <https://doi.org/10.48550/arXiv.2308.07037>.
20. Rezende D, Mohamed S. Variational inference with normalizing flows. In: *International conference on machine learning*. PMLR; 2015. p. 1530–8.
21. Zhang Y, Hu Y, Li H, Liu X. Drug-protein interaction prediction via variational autoencoders and attention mechanisms. *Front Genet* 2022;**13**:1032779.
22. Ochiai T, Inukai T, Akiyama M, Furui K, Ohue M, Matsumori N, et al. Variational autoencoder-based chemical latent space for large molecular structures with 3d complexity. *Commun Chem* 2023;**6**:249.
23. Maziarka Ł, Pocha A, Kaczmarczyk J, Rataj K, Danel T, Warchoń M. Mol-cyclegan: a generative model for molecular optimization. *J Cheminf* 2020;**12**:2.
24. Zhang M, Zhou Z, Li L, Liu Z, Yang M, Feng Y. Hpgan: sequence search with generative adversarial networks. *IEEE Transact Neural Networks Learn Syst* 2021;**34**:4944–56.
25. Macedo B, Ribeiro Vaz I, Taveira Gomes T. Medgan: optimized generative adversarial network with graph convolutional networks for novel molecule design. *Sci Rep* 2024;**14**:1212.
26. Lin J, Wang X, Liu T, Teng Y, Cui W. Diffusion-based generative network for *de novo* synthetic promoter design. *ACS Synth Biol* 2024;**13**:1513–22.
27. Morehead A, Cheng J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun Chem* 2024;**7**:150.
28. Hassan M, Shenoy N, Lee J, Stärk H, Thaler S, Beaini D. Et-flow: equivariant flow-matching for molecular conformer generation. *Adv Neural Inf Process Syst* 2024;**37**:128798–824.
29. Wu S, Lambard G, Liu C, Yamada H, Yoshida R. Iqspr in xenonpy: a bayesian molecular design algorithm. *Mol Inform* 2020;**39**:1900107.
30. Song Y, Gong J, Zhou H, Zheng M, Liu J, Ma WY. Unified generative modeling of 3d molecules with bayesian flow networks. In: *The twelfth international conference on learning representations*; 2024.
31. Zang C, Wang F. Moflow: an invertible flow model for generating molecular graphs. In: *Proceedings of the 26th ACM SIGKDD international conference on knowledge Discovery & data mining*; 2020. p. 617–26.
32. Bengio E, Jain M, Korablyov M, Precup D, Bengio Y. Flow network based generative models for non-iterative diverse candidate generation. *Adv Neural Inf Process Syst* 2021;**34**:27381–94.
33. Ding M. The road from mle to em to vae: a brief tutorial. *AI Open* 2022;**3**:29–34.
34. Mohebbi Moghaddam M, Boroomand B, Jalali M, Zareian A, Daeijavad A, Manshaei MH, et al. Games of gans: game-theoretical models for generative adversarial networks. *Artif Intell Rev* 2023;**56**: 9771–807.
35. Carrillo JA, del Pino M, Figalli A, Mingione G, Vázquez JL, Vázquez JL. The mathematical theories of diffusion: nonlinear and fractional diffusion. In: *Nonlocal and nonlinear diffusions and interactions: new methods and directions*; 2017. p. 205–78.
36. Bond-Taylor S, Leach A, Long Y, Willcocks CG. Deep generative modelling: a comparative review of vaes, gans, normalizing flows, energy-based and autoregressive models. *IEEE Trans Pattern Anal Mach Intell* 2021;**44**:7327–47.
37. Gangwal A, Ansari A, Ahmad I, Azad AK, Kumarasamy V, Subramaniyan V, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Front Pharmacol* 2024;**15**:1331062.
38. Ai R, Xiao X, Deng S, Yang N, Xing X, Watne LO, et al. Artificial intelligence in drug development for delirium and Alzheimer's disease. *Acta Pharm Sin B* 2025;**15**:4386–410.
39. Jusoh AS, Remli MA, Mohamad MS, Cazenave T, Fong CS. How generative artificial intelligence can transform drug discovery?. *Eur J Med Chem* 2025:117825.
40. Grow C, Gao K, Nguyen DD, Wei G-W. Generative network complex (GNC) for drug discovery. *Commun Inf Syst* 2019;**19**:241.
41. Gao Z, Wang X, Blumenfeld Gaines B, Shi X, Bi J, Song M. Fragment-based deep molecular generation using hierarchical chemical graph representation and multi-resolution graph variational autoencoder. *Mol Inform* 2023;**42**:2200215.
42. Wang R, Feng H, Wei GW. Chatgpt in drug discovery: a case study on anticocaine addiction drug development with chatbots. *J Chem Inf Model* 2023;**63**:7189–209.
43. Qi X, Zhao L, Tian C, Li Y, Chen ZL, Huo P, et al. Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nat Commun* 2024;**15**:9256.
44. Kanakala GC, Devata S, Chatterjee P, Priyakumar UD. Generative artificial intelligence for small molecule drug design. *Curr Opin Biotechnol* 2024;**89**:103175.
45. Nguyen DD, Gao K, Wang M, Wei G-W. Mathdl: mathematical deep learning for d3r grand challenge 4. *J Comput Aided Mol Des* 2020;**34**:131–47.
46. Nguyen DD, Cang Z, Wu K, Wang M, Cao Y, Wei GW. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R grand challenges. *J Comput Aided Mol Des* 2019;**33**: 71–82.
47. Jiang J, Ke L, Chen L, Dou BZ, Zhu YY, Liu J, et al. Transformer technology in molecular science. *Wiley Interdiscip Rev Comput Mol Sci* 2024;**14**:e1725.
48. Tao J, Liang L, Hao S, Chen Y, Yang Z, Cai Y, et al. Artificial intelligence for radiopharmaceutical and molecular imaging. *Acta Pharm Sin B* 2025. Available from: <https://doi.org/10.1016/j.apsb.2025.09.039>.
49. Wang R, Zhuang C. Graph neural networks driven acceleration in drug discovery. *Acta Pharm Sin B* 2025. Available from: <https://doi.org/10.1016/j.apsb.2025.10.011>.
50. Meuwly M. Machine learning for chemical reactions. *Chem Rev* 2021;**121**:10218–39.
51. Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS Cent Sci* 2019;**5**:1572–83.
52. Gao W, Coley CW. The synthesizability of molecules proposed by generative models. *J Chem Inf Model* 2020;**60**:5714–23.
53. Perron Q, Mirguet O, Tajmouati H, Skiredj A, Rojas A, Gohier A, et al. Deep generative models for ligand-based *de novo* design applied to multi-parametric optimization. *J Comput Chem* 2022;**43**: 692–703.
54. Wiering MA, Van Otterlo M. Reinforcement learning. *Adapt Learn Optim* 2012;**12**:729.
55. Yoshimori A, Miljković F, Bajorath J. Approach for the design of covalent protein kinase inhibitors via focused deep generative modeling. *Molecules* 2022;**27**:570.

56. Danel T, Łęski J, Podlowska S, Podolak IT. Docking-based generative approaches in the search for new drug candidates. *Drug Discov Today* 2023;**28**:103439.
57. Chakraborty R, Hasija Y. Utilizing deep learning to explore chemical space for drug lead optimization. *Expert Syst Appl* 2023;**229**:120592.
58. Lu W, Zhang J, Huang W, Zhang Z, Jia X, Wang Z, et al. Dynamicbind: predicting ligand-specific protein–ligand complex structure with a deep equivariant generative model. *Nat Commun* 2024;**15**:1071.
59. Cao Y, Das P, Chenthamarakshan V, Chen P-Y, Melnyk I, Shen Y. Fold2seq: a joint sequence (1D)- fold (3D) embedding-based generative model for protein design. *Proc Mach Learn Res* 2021;**139**:1261–71.
60. Zhang O, Lin H, Zhang H, Zhao H, Huang Y, Hsieh C-Y, et al. Deep lead optimization: leveraging generative ai for structural modification. *J Am Chem Soc* 2024;**146**:31357–70.
61. Weller JA, Rohs R. Structure-based drug design with a deep hierarchical generative model. *J Chem Inf Model* 2024;**64**:6450–63.
62. Wei GW. Protein structure prediction beyond alphafold. *Nat Mach Intell* 2019;**1**:336–7.
63. Qiu Y, Hu J, Wei GW. Cluster learning-assisted directed evolution. *Nat Comput Sci* 2021;**1**:809–18.
64. Lopez R, Gayoso A, Yosef N. Enhancing scientific discoveries in molecular biology with deep generative models. *Mol Syst Biol* 2020;**16**:e9198.
65. Lee M. Recent advances in generative adversarial networks for gene expression data: a comprehensive review. *Mathematics* 2023;**11**:3055.
66. Betzel RF, Bassett DS. Generative models for network neuroscience: prospects and promise. *J R Soc Interface* 2017;**14**:20170623.
67. Gomari DP, Schweickart A, Cerchietti L, Paietta E, Fernandez H, Al-Amin H, et al. Variational autoencoders learn transferrable representations of metabolomics data. *Commun Biol* 2022;**5**:645.
68. Anstine DM, Isayev O. Generative models as an emerging paradigm in the chemical sciences. *J Am Chem Soc* 2023;**145**:8736–50.
69. Bai Q, Liu S, Tian Y, Xu T, Banegas-Luna AJ, Pérez-Sánchez H, et al. Application advances of deep learning methods for *de novo* drug design and molecular dynamics simulation. *Wiley Interdiscip Rev Comput Mol Sci* 2022;**12**:e1581.
70. Pyrkov A, Aliper A, Bezrukov D, Lin YC, Polykovskiy D, Kanya P, et al. Quantum computing for near-term applications in generative chemistry and drug discovery. *Drug Discov Today* 2023;**28**:103675.
71. Li J, Topaloglu RO, Ghosh S. Quantum generative models for small molecule drug discovery. *IEEE Trans Quantum Eng* 2021;**2**:1–8.
72. Hu W. Reinforcement learning of molecule optimization with bayesian neural networks. *Comput Mol Biosci* 2021;**11**:69–83.
73. Koshino K, Werner RA, Pomper MG, Bundschuh RA, Toriumi F, Higuchi T, et al. Narrative review of generative adversarial networks in medical and molecular imaging. *Ann Transl Med* 2021;**9**:821.
74. Iwata H, Nakai T, Koyama T, Matsumoto S, Kojima R, Okuno Y. Vgae-mcts: a new molecular generative model combining the variational graph auto-encoder and monte carlo tree search. *J Chem Inf Model* 2023;**63**:7392–400.
75. Han X, Chen X, Ruiz FJ, Liu LP. Fitting autoregressive graph generative models through maximum likelihood estimation. *J Mach Learn Res* 2023;**24**:1–30.
76. Dong J, Zhao M, Liu Y, Su Y, Zeng X. Deep learning in retrosynthesis planning: datasets, models and tools. *Briefings Bioinf* 2022;**23**:bbab391.
77. Mandal D, Yeasin M, Sarkar A, Mallik S, Tomar N. Supervised gan (sgan): method and application on labelled genomic/epigenomic data. In: *Swarm optimization for biomedical applications*. CRC Press; 2025. p. 234–51.
78. Choi JM, Ji M, Watson LT, Zhang L. Deepmicrogen: a generative adversarial network-based method for longitudinal microbiome data imputation. *Bioinformatics* 2023;**39**:btad286.
79. Wang J, Wang K, Yu Y, Lu Y, Xiao W, Sun Z, et al. Selfimproving generative foundation model for synthetic medical image generation and clinical applications. *Nat Med* 2025;**31**:609–17.
80. Tang X, Huo M, Chen Y, Huang H, Qin S, Luo J, et al. A novel deep generative model for mRNA vaccine development: designing 5' UTRs with N1-methyl-pseudouridine modification. *Acta Pharm Sin B* 2024;**14**:1814–26.
81. Gangwal A, Ansari A, Ahmad I, Azad AK, Sulaiman WMAW. Current strategies to address data scarcity in artificial intelligence-based drug discovery: a comprehensive review. *Comput Biol Med* 2024;**179**:108734.
82. Tong X, Liu X, Tan X, Li X, Jiang J, Xiong Z, et al. Generative models for *de novo* drug design. *J Med Chem* 2021;**64**:14011–27.
83. Lisboa PJ, Saralajew S, Vellido A, Fernández-Domenech R, Villmann T. The coming of age of interpretable and explainable machine learning models. *Neurocomputing* 2023;**535**:25–39.
84. Tutek M, Šnajder J. Toward practical usage of the attention mechanism as a tool for interpretability. *IEEE Access* 2022;**10**:47011–30.
85. Kumar SA, Ananda Kumar TD, Beeraka NM, Pujar GV, Singh M, Narayana Akshatha HS, et al. Machine learning and deep learning in data-driven decision making of drug discovery and challenges in high-quality data acquisition in the pharmaceutical industry. *Future Med Chem* 2022;**14**:245–70.
86. Nandan M, Mitra S, De D. Graphxai: a survey of graph neural networks (GNNs) for explainable ai (XAI). *Neural Comput Appl* 2025;**37**:10949–1000.
87. Alizadehsani R, Oyeler SS, Hussain S, Jagatheesaperumal SK, Calixto RR, Rahouti M, et al. Explainable artificial intelligence for drug discovery and development-a comprehensive survey. *IEEE Access* 2024;**12**:35796–812.
88. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst* 2020;**33**:1877–901.
89. Ruthotto L, Haber E. An introduction to deep generative modeling. *GAMM-Mitt* 2021;**44**:e202100008.
90. Zhang H, Lu Y, Alkhouri I, Ravishanker S, Song D, Qu Q. Improving training efficiency of diffusion models via multi-stage framework and tailored multi-decoder architecture. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2024. p. 7372–81.
91. Li C, Zhang M, He Y. The stability-efficiency dilemma: investigating sequence length warmup for training gpt models. *Adv Neural Inf Process Syst* 2022;**35**:26736–50.
92. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med* 2023;**29**:1930–40.
93. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;**2**:e0000198.
94. He Y, Huang F, Jiang X, Nie Y, Wang M, Wang J, et al. Foundation model for advancing healthcare: challenges, opportunities and future directions. *IEEE Rev Biomed Eng* 2024;**18**:172–91.
95. Gao K, Nguyen DD, Tu M, Wei GW. Generative network complex for the automated generation of drug-like molecules. *J Chem Inf Model* 2020;**60**:5682–98.
96. Zeng X, Wang F, Luo Y, Kang SG, Tang J, Lightstone FC, et al. Deep generative molecular design reshapes drug discovery. *Cell Rep Med* 2022;**3**:100794.
97. Han Z, Shen Z, Pei J, You Q, Zhang Q, Wang L. Transformation of peptides to small molecules in medicinal chemistry: challenges and opportunities. *Acta Pharm Sin B* 2024;**14**:4243–65.
98. Zhao F, Zhang C, Geng B. Deep multimodal data fusion. *ACM Comput Surv* 2024;**56**:1–36.
99. Feng H, Wang R, Zhan CG, Wei GW. Multiobjective molecular optimization for opioid use disorder treatment using generative network complex. *J Med Chem* 2023;**66**:12479–98.

100. Mock M, Langmead CJ, Grandsard P, Edavettal S, Russell A. Recent advances in generative biology for biotherapeutic discovery. *Trends Pharmacol Sci* 2024;**45**:255–67.
101. Tang Q, Liang J, Zhu F. A comparative review on multi-modal sensors fusion based on deep learning. *Signal Process* 2023;**213**:109165.
102. Rahate A, Walambe R, Ramanna S, Kotecha K. Multimodal co-learning: challenges, applications with datasets, recent advances and future directions. *Inf Fusion* 2022;**81**:203–39.
103. Duan J, Xiong J, Li Y, Ding W. Deep learning based multimodal biomedical data fusion: an overview and comparative review. *Inf Fusion* 2024;**112**:102536.
104. Luukkonen S, van den Maagdenberg HW, Emmerich MT, van Westen GJ. Artificial intelligence in multi-objective drug design. *Curr Opin Struct Biol* 2023;**79**:102537.
105. Angelo JS, Guedes IA, Barbosa HJ, Dardenne LE. Multi-and many-objective optimization: present and future in *de novo* drug design. *Front Chem* 2023;**11**:1288626.
106. Lambrinidis G, Tsantili-Kakoulidou A. Multi-objective optimization methods in novel drug design. *Expet Opin Drug Discov* 2021;**16**: 647–58.
107. Merz KM Jr, De Fabritiis G, Wei GW. Generative models for molecular design. *J Chem Inf Model* 2020;**60**:5635–6.