Scilight

*Transactions on Graph Intelligence and Network Applications*
https://www.sciltp.com/journals/tgina



Check for updates

*Article*

# A Comprehensive Survey on Graph-Structured Small-Molecule Drug Discovery and Development with Deep Learning

Junwei Hu [1,†], Wei Zhang [1,†], Zhuohao Lin [1,†], Jiajun Yu [2], Jiameng Chen [1], Jiawei Du [3], Wenbin Hu [1,4,*], Kun Li [1,5,*] and João Gama [6]

[1] School of Computer Science, Wuhan University, Wuhan 430072, China
[2] College of Computer Science and Technology, Zhejiang University, Hangzhou 310058, China
[3] Centre for Frontier AI Research, Agency for Science, Technology and Research, Singapore 138632, Singapore
[4] Shenzhen Research Institute, Wuhan University, Shenzhen 518057, China
[5] Shanghai Artificial Intelligence Laboratory, Shanghai 200233, China
[6] Faculty of Economics, University of Porto, 4200-464 Porto, Portugal
* Correspondence: hwb@whu.edu.cn (W.H.); likun98@whu.edu.cn (K.L.)
† These authors contributed equally to this work.

**How To Cite**: Hu, J.; Zhang, W.; Lin, Z.; et al. A Comprehensive Survey on Graph-Structured Small-Molecule Drug Discovery and Development with Deep Learning. *Transactions on Graph Intelligence and Network Applications* **2026**. https://doi.org/10.53941/tgina.2026.100008

Received: 23 June 2026
Revised: 7 October 2026
Accepted: 21 October 2026
Published: 4 November 2026

**Abstract:** Small-molecule drug discovery remains a cornerstone of modern therapeutics, yet traditional development pipelines are often limited by high costs, long timelines, and low success rates. Recent advances in graph-structured deep learning have significantly transformed the modeling of molecular structures, biological interactions, and pharmacological responses. By naturally representing molecules as graphs, these approaches provide an effective framework for capturing structural, geometric, and physicochemical information across multiple biological scales. In this survey, we present a comprehensive review of recent advances in graph-structured small-molecule drug discovery and development. To provide a unified perspective, we organize the literature into a four-layer framework covering data resources, representation learning, task formulation, and downstream applications. We first summarize the evolution of molecular representations from molecular sequences and topological graphs to geometric structures, molecular dynamics, and electronic field representations. We then introduce a unified task formulation covering six major tasks: molecular property prediction, drug-target interaction and affinity prediction, drug response and perturbation prediction, drug-drug interaction prediction, molecular generation, and molecular optimization. Building upon this framework, we review recent developments in graph neural networks, geometric learning, diffusion models, and foundation models. Finally, we summarize representative benchmark datasets and evaluation resources, and discuss several key challenges, such as out-of-distribution generalization and data scarcity, together with promising directions for future research.

**Keywords:** small-molecule drug discovery; deep learning; molecule representation learning; graph neural networks; foundation models

## 1. Introduction

Small-molecule drugs, typically defined as chemically synthesized compounds with molecular weights below 1000 Daltons, remain one of the most important therapeutic modalities in modern medicine [1]. Owing to their favorable pharmacokinetic properties, synthetic accessibility, and broad applicability across disease areas, small molecules continue to account for the vast majority of approved therapeutic agents [2,3]. Despite their success,

**Copyright:** © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
**Publisher's Note:** Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

traditional drug discovery pipelines remain costly, time-consuming, and highly uncertain. The development of a single approved drug often requires years of experimental screening and optimization, while only a tiny fraction of the enormous chemical space, estimated to contain more than $10^{60}$ synthesizable molecules, can be explored through conventional laboratory approaches [4,5]. These limitations have motivated the rapid adoption of deep learning-driven drug discovery, where computational models are increasingly used to prioritize candidates, predict biological activities, and guide molecular design.

Among various molecular representations, graph-structured representations have become one of the most widely adopted paradigms for modeling small molecules [6,7]. In graph formulations, atoms are naturally represented as nodes and chemical bonds as edges, providing a structured description that closely aligns with the underlying molecular topology. Compared with sequence-based representations such as SMILES, graph representations preserve connectivity patterns more explicitly and can be naturally extended to incorporate three-dimensional geometric information, physicochemical attributes, and higher-order molecular interactions. As a result, graph learning has become a central component of modern computational drug discovery. Recent years have witnessed rapid progress in graph-structured molecular learning [8–10]. Early studies primarily focused on graph neural networks operating on static molecular topologies, whereas contemporary approaches increasingly integrate three-dimensional geometric structures, physical priors, biological context information, and generative modeling techniques. Advances in geometric deep learning, diffusion models, graph Transformer frameworks, and foundation models have significantly expanded the scope of graph-structured molecular modeling. These developments have enabled learning systems that not only predict molecular properties and biological interactions but also support molecular generation, optimization, and broader drug development workflows.

As illustrated in Figure 1, graph-structured small-molecule drug discovery and development can be organized into a unified four-layer framework. At the foundation lies a diverse data ecosystem consisting of molecular databases, pharmacogenomic resources, and biological screening platforms. These resources provide the basis for multi-modal molecular representations, including molecular graphs, cellular omics profiles, and protein structural information. Building upon these representations, learning systems can be formulated through several fundamental task paradigms involving molecules, proteins, and cellular systems. Together, these components support six representative downstream tasks that span the major stages of drug discovery and development:

(1) Drug-target interaction and affinity prediction (DTI/DTA): Predicts interaction likelihoods or binding affinities between small molecules and biological targets, forming the foundation of target-based drug discovery.
(2) Drug response and perturbation prediction (DRP/DPP): Predict cellular outcomes induced by chemical perturbations, including cell viability, drug sensitivity, and transcriptomic response profiles.
(3) Molecular property prediction (MPP): Estimates absorption, distribution, metabolism, excretion, and toxicity (ADMET)-related characteristics that determine the developability of candidate molecules.
(4) Drug-drug interaction prediction (DDI): Models potential synergistic effects, adverse interactions, and safety risks arising from combinations of multiple therapeutic compounds.
(5) Molecular generation (MG): Designs novel molecular structures that satisfy predefined biological, chemical, or pharmacological objectives.
(6) Molecular optimization (MO): Improves existing lead compounds through iterative structural modification while preserving desirable properties and enhancing target objectives.

Although substantial progress has been achieved across these research directions, the literature remains fragmented across different task settings, representation paradigms, and evaluation protocols. Existing surveys often focus on specific subfields, such as molecular property prediction, drug-target interactions, or molecular generation, making it increasingly difficult to obtain a unified view of the rapidly evolving landscape. Moreover, several emerging challenges, including out-of-distribution (OOD) generalization, activity and property cliffs, data scarcity and incomplete supervision, and the lack of mechanistic interpretability, have received growing attention in recent years and deserve a systematic discussion.

To address these challenges, this survey provides a comprehensive overview of recent advances in graph-structured small-molecule drug discovery and development. Rather than organizing the literature solely according to downstream applications, we introduce a unified four-layer perspective that systematically connects data resources and benchmark ecosystems, molecular representation learning, task formulations, algorithmic paradigms, and downstream design objectives. This framework enables a clearer understanding of how modern graph-structured drug discovery systems evolve from molecular and biological data to predictive and generative decision-making processes. Building upon this perspective, we review the evolution of molecular representations from molecular sequences and topological graphs to geometric structures, trajectory-aware molecular dynamics, and electronic

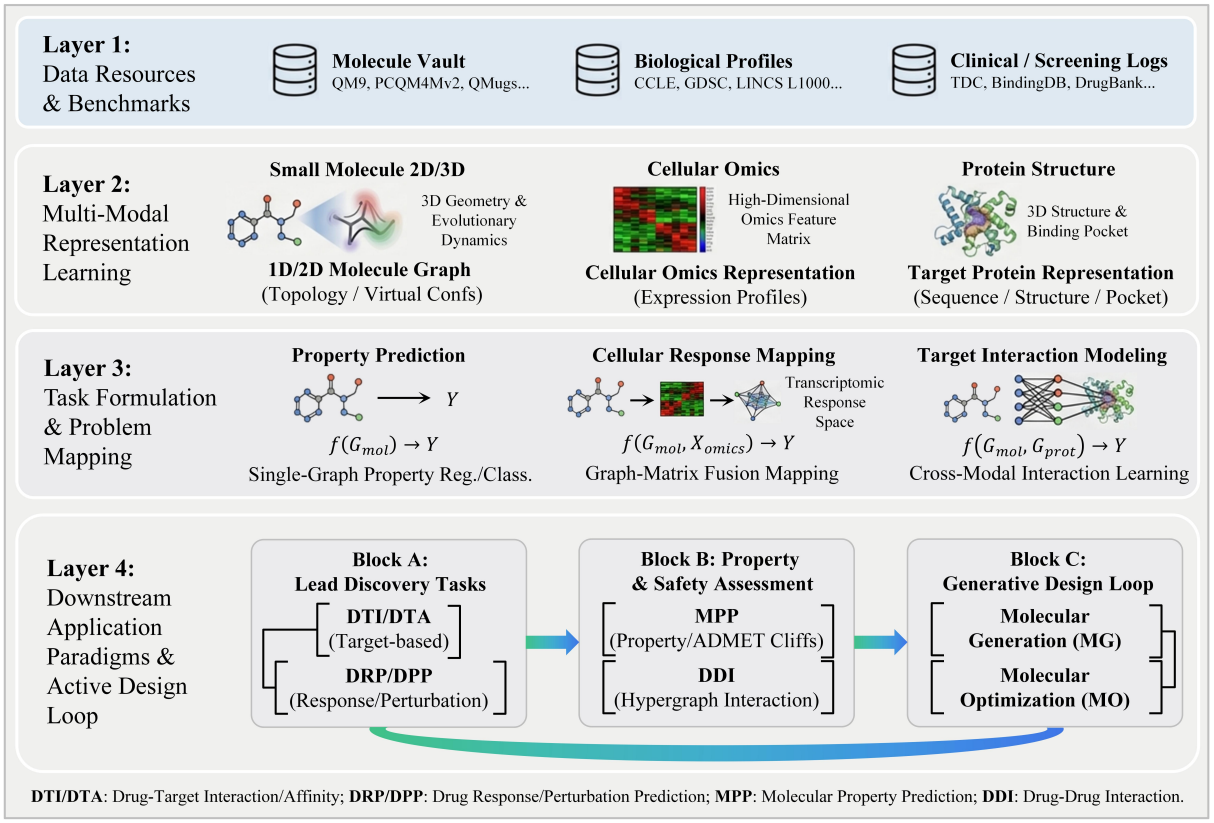



**Figure 1.** A unified four-layer framework for graph-structured small-molecule drug discovery and development.

field representations. We further establish a unified task formulation framework that links six core tasks, including DTI/DTA, DRP/DPP, MPP, DDI, MG, and MO. In addition, we systematically summarize recent methodological advances in graph neural networks, geometric deep learning, diffusion models, graph Transformer frameworks, and foundation models, while providing an updated overview of benchmark datasets, evaluation protocols, and emerging challenges. Compared with existing surveys, our goal is not only to summarize recent methodological developments but also to highlight the relationships among different research directions within a common graph-centered framework.

The remainder of this survey is organized as follows. Section 2 reviews the evolution from conventional molecular topology modeling to modern multi-modal representation learning. Section 3 introduces unified mathematical formulations for the six core tasks. Section 4 surveys recent advances in graph-structured learning paradigms across different application scenarios. Section 5 summarizes widely used benchmark datasets and discusses emerging evaluation protocols beyond conventional independent and identically distributed settings. Finally, Section 6 outlines open challenges and future research opportunities in graph-structured drug discovery and development.

## 2. Molecular Representation Learning

This section corresponds to Layer 2 in the four-layer framework shown in Figure 1. A central challenge in graph-structured drug discovery lies in how molecular structures are represented. The choice of representation determines what chemical and biological information can be captured and directly influences the effectiveness of downstream learning algorithms. Over the past decade, molecular representations have evolved from one-dimensional symbolic descriptions to graph topological structures, three-dimensional geometric conformations, and more recently, representations that incorporate molecular dynamics and electronic information [6,7,11–17]. This progression reflects an increasing effort to encode structural and physicochemical factors that govern molecular properties and biological interactions.

### *2.1. 1D and 2D Representations*

Early computational approaches commonly represented molecules as one-dimensional sequences, most notably through the Simplified Molecular Input Line Entry System (SMILES). Owing to their compact and machine-readable format, SMILES strings have been widely adopted in sequence-based machine learning frameworks and generative models. However, conventional SMILES representations are sensitive to syntactic perturbations, where small modifications may lead to chemically invalid structures. Although robust alternatives such as Self-Referencing

Embedded Strings (SELFIES) have been proposed to alleviate this limitation, graph-structured representations remain the dominant paradigm for molecular modeling.

Formally, a molecule can be represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node ($v_i \in \mathcal{V}$) corresponds to an atom and each edge ($e_{ij} \in \mathcal{E}$) corresponds to a chemical bond. Node attributes typically include atomic number, valence state, hybridization type, aromaticity, and other atom-level descriptors, while edge features encode bond types and related chemical information. Based on this formulation, graph neural networks (GNNs) and message passing neural networks (MPNNs) aggregate information from local neighborhoods to learn molecular representations. Compared with sequence-based encodings, graph representations preserve molecular connectivity more naturally and provide an effective foundation for capturing structural relationships. Nevertheless, conventional 2D graphs describe molecules primarily through topological connectivity and do not explicitly account for three-dimensional geometry or conformational flexibility.

### *2.2. 3D Geometric Representations*

Because molecular properties and intermolecular interactions are inherently dependent on spatial arrangements, three-dimensional geometry plays an important role in molecular representation learning. To incorporate structural information beyond topology, geometric deep learning extends molecular graphs into three-dimensional coordinate space by associating each atom with a spatial position $\mathbf{r}_i \in \mathbb{R}^3$. Geometry-aware models typically construct invariant or equivariant features from atomic coordinates, including interatomic distances $d_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, bond angles, and torsional angles. These geometric descriptors enable models to capture spatial patterns that are directly related to molecular conformations and physicochemical properties.

More recently, E(3)- and SE(3)-equivariant neural networks have become a major research direction in molecular machine learning [12, 18]. These architectures preserve the transformation behavior of geometric quantities under rotations, translations, and reflections, allowing directional information such as forces, velocity vectors, and dipole moments to be modeled more faithfully. Such approaches have achieved strong performance on molecular property prediction and quantum chemistry benchmarks, including QM9 and MD17 [13, 19]. Nevertheless, most geometric graph representations remain based on static molecular conformations and often neglect conformational ensembles and dynamic structural fluctuations observed under realistic biological conditions.

### *2.3. Dynamic and Trajectory-Aware Representations*

Beyond static molecular structures, recent studies have explored dynamic representations that incorporate conformational transitions, molecular trajectories, and optimization pathways. The underlying motivation is that important information regarding molecular behavior may not be fully captured by a single structural snapshot. Trajectory-aware representations seek to model how molecular structures evolve across different states, enabling learning algorithms to exploit temporal or sequential structural information. In molecular dynamics simulations, for example, a molecule is represented as a sequence of conformations rather than an isolated geometry. This perspective provides access to dynamic properties such as structural stability, flexibility, and transition patterns.

Recent studies have also investigated the use of molecular optimization trajectories as an additional source of supervision [10]. Instead of treating molecules as independent samples, these approaches model relationships among molecules connected through optimization or evolutionary pathways. Building upon this idea, path-consistent molecular representation learning frameworks have been proposed to encode structural transformation trajectories into continuous latent spaces [16]. By incorporating trajectory information into graph learning, these methods provide complementary signals that may benefit downstream tasks such as lead optimization and molecular design.

### *2.4. Electronic Structure Representations*

Beyond atomic coordinates and molecular geometry, electronic structure provides an additional level of information that governs many physicochemical and biological processes. Molecular recognition, intermolecular interactions, and quantum-mechanical properties are ultimately determined by the distribution of electrons rather than by nuclear positions alone. Traditional molecular learning models typically encode electronic effects indirectly through topology or geometry. While such representations are often sufficient for many predictive tasks, they may be less effective in capturing subtle electronic phenomena associated with electrostatic interactions, charge transfer, polarization effects, and non-covalent interactions.

To address this limitation, recent studies have explored the integration of electronic structure information into molecular representation learning [20, 21]. Instead of relying solely on discrete atomic features, these approaches incorporate continuous electronic descriptors derived from quantum chemistry calculations or electron-density approximations. For example, UniField introduces electron-density-informed representations through radial basis

function (RBF) fusion mechanisms that combine continuous electronic information with graph-structured molecular features [17]. By integrating topology, geometry, and electronic structure within a unified framework, such methods aim to improve the modeling of molecular properties and biological interactions while maintaining practical computational efficiency.

Overall, molecular representation learning has evolved from topology-centered descriptions toward increasingly rich representations that integrate geometry, dynamics, and electronic information. This evolution provides the foundation for modern graph-structured drug discovery systems and motivates the diverse learning paradigms discussed in subsequent sections.

## 3. Task Formulation

Building upon the multi-scale molecular representations introduced in Section 2, we formulate the major tasks in graph-structured small-molecule drug discovery and development within a unified learning framework, corresponding to Layer 3 in Figure 1. The major mathematical symbols used throughout these task formulations are summarized in Table 1. Although these tasks differ in their biological objectives and prediction targets, they can generally be categorized into two complementary paradigms. The first consists of discriminative learning tasks, which map molecular and biological representations to predictive outcomes. The second consists of generative learning tasks, which construct or optimize molecular structures under desired property constraints. Together, these formulations provide a common foundation for understanding the diverse methodologies discussed in subsequent sections.

### *3.1. Mathematical Entities*

To establish consistent notation across different tasks, we first define the major biological entities involved throughout this survey.

- Small Molecules ($\mathcal{M}$). A molecule is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where nodes correspond to atoms and edges correspond to chemical bonds. Each atom is associated with a feature vector $\mathbf{x}_i \in \mathbb{R}^F$, and optional three-dimensional coordinates $\mathbf{r}_i \in \mathbb{R}^3$ may be incorporated when geometric information is available.
- Proteins ($\mathcal{P}$). A target protein may be represented as an amino acid sequence $\mathcal{S}_p$, a residue-level graph, or a three-dimensional binding pocket structure depending on the task formulation.
- Cellular Systems ($\mathcal{C}$). Cellular contexts are typically represented by high-dimensional vectors $\mathbf{s} \in \mathbb{R}^D$, describing transcriptomic, genomic, proteomic, or other omics-level measurements.

### *3.2. Discriminative Learning Tasks*

Discriminative learning tasks aim to predict molecular properties or biological outcomes from task-specific molecular and biological inputs. In general, these tasks can be formulated as

$$Y = f_\theta(X), \tag{1}$$

where $X$ denotes the task-specific input and $Y$ denotes the corresponding prediction output. Depending on the task, $X$ may consist of a molecular representation, a molecule-protein pair, a molecule-cellular-state pair, or a pair of molecular representations.

MPP is one of the most fundamental tasks in molecular machine learning. Its objective is to infer intrinsic molecular properties directly from chemical structures without requiring external biological information. Depending on the application scenario, the target properties may include physicochemical descriptors, quantum-mechanical quantities, toxicity indicators, or broader ADMET-related characteristics.

$$Y_{\text{MPP}} = f_{\text{MPP}}(\mathcal{G}), \tag{2}$$

where $Y_{\text{MPP}}$ may be a scalar, vector, or multi-task prediction target. MPP serves as a foundational benchmark for evaluating the quality of molecular representations and remains one of the most widely studied tasks in graph learning.

DTI and DTA focus on modeling interactions between small molecules and biological targets. These tasks form the basis of target-based drug discovery by estimating whether a molecule can bind to a protein and, if so, how strongly the interaction occurs.

$$Y_{\text{DTI/DTA}} = f_{\text{DTI/DTA}}(\mathcal{G}_{\text{mol}}, \mathcal{G}_p), \tag{3}$$

**Table 1.** Summary of major mathematical symbols used in this survey.

| Symbol | Description |
|---|---|
| $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ | Molecular graph, where $\mathcal{V}$ and $\mathcal{E}$ denote the sets of atoms and chemical bonds, respectively. |
| $\mathcal{G}_{\mathrm{mol}}$ | Molecular representation used in downstream prediction tasks. |
| $\mathcal{G}_p$ | Protein or target representation, which may be derived from sequence, residue-level graphs, or three-dimensional structural information. |
| $\mathbf{x}_i \in \mathbb{R}^F$ | Feature vector associated with the $i$-th atom or node. |
| $\mathbf{r}_i \in \mathbb{R}^3$ | Three-dimensional coordinate of the $i$-th atom. |
| $\mathcal{S}_p$ | Amino-acid sequence of a target protein. |
| $\mathbf{s} \in \mathbb{R}^D$ | Cellular-context representation, such as transcriptomic, genomic, or other omics features. |
| $X$ | Task-specific input to a discriminative model, which may include molecular, protein, cellular, or paired-drug representations. |
| $Y_{\mathrm{MPP}}$ | Prediction target for molecular property prediction, which may be a scalar, vector, or multi-task output. |
| $Y_{\mathrm{DTI}}$ | Binary or categorical interaction target for drug-target interaction prediction. |
| $Y_{\mathrm{DTA}}$ | Continuous affinity target for drug-target affinity prediction, such as $K_d$, $K_i$, or their logarithmic transformations. |
| $Y_{\mathrm{DRP}}$ | Drug-response prediction target, such as cell viability, growth inhibition, or drug sensitivity. |
| $\mathbf{s}_{\mathrm{pert}}$ | Post-treatment cellular state or drug-induced perturbation profile. |
| $Y_{\mathrm{DDI}}$ | Prediction target for drug-drug interaction, such as an interaction label, interaction type, or continuous interaction score. |
| $\mathcal{C}_{\mathrm{prop}}$ | Property, biological, or pharmacological condition used for molecular generation or optimization. |
| $\mathcal{G}_{\mathrm{old}}$ | Original or input molecular graph before optimization. |
| $\mathcal{G}_{\mathrm{new}}$ | Generated or optimized molecular graph. |
| $\mathcal{G}_{\mathrm{seed}}$ | Seed or reference molecule used in the closed-loop molecular optimization framework. |
| $\mathrm{Sim}(\cdot, \cdot)$ | Structural similarity function between two molecules, such as the Tanimoto coefficient. |
| $f_\theta$ | Parameterized predictive, generative, or optimization mapping. |
| $\alpha_k$ | Weight assigned to the $k$-th predictive objective in multi-objective optimization. |
| $\beta$ | Weight controlling the contribution of the structural similarity term. |
| $\gamma$ | Minimum structural-similarity threshold imposed during molecular optimization. |

where $Y_{\mathrm{DTI}}$ typically denotes a binary interaction label, $\mathcal{G}_{\mathrm{mol}}$ and $\mathcal{G}_p$ denote the molecular and protein representations, respectively. In affinity prediction settings, the same formulation is extended to continuous targets such as dissociation constants ($K_d$), inhibition constants ($K_i$), or their logarithmic transformations (e.g., p$K_d$ and p$K_i$).

Unlike DTI/DTA, which focuses on molecular interactions at the target level, DRP and DPP model drug effects at the cellular level. These tasks seek to understand how a chemical perturbation alters the state of a biological system. DRP estimates cellular sensitivity to a compound:

$$Y_{\mathrm{DRP}} = f_{\mathrm{DRP}}(\mathcal{G}_{\mathrm{mol}}, \mathbf{s}), \tag{4}$$

where $Y_{\mathrm{DRP}}$ may represent viability, growth inhibition, or drug sensitivity measurements (e.g., $\mathrm{IC}_{50}$). DPP further models high-dimensional cellular state transitions:

$$\mathbf{s}_{\mathrm{pert}} = f_{\mathrm{DPP}}(\mathcal{G}_{\mathrm{mol}}, \mathbf{s}), \tag{5}$$

where $\mathbf{s}_{\mathrm{pert}}$ denotes the post-treatment cellular state, typically represented by transcriptomic profiles. Compared with DRP, DPP provides a more detailed characterization of drug-induced biological responses.

DDI prediction aims to characterize the effects of combining multiple drugs. These effects may include synergistic therapeutic benefits, antagonistic interactions, or adverse side effects that emerge only when drugs are administered together.

$$Y_{\text{DDI}} = f_{\text{DDI}}(\mathcal{G}_i, \mathcal{G}_j), \tag{6}$$

where $Y_{\text{DDI}}$ represents a predefined interaction label or interaction type, $\mathcal{G}_i$ and $\mathcal{G}_j$ denote two drug representations. DDI prediction plays an important role in drug safety assessment and combination therapy design.

### *3.3. Generative Learning Tasks*

While discriminative tasks focus on prediction, generative tasks focus on molecular design. Their objective is to learn molecular distributions or transformation functions that can generate candidate compounds satisfying predefined objectives.

MG seeks to learn a distribution over chemically valid molecular structures. The goal is to generate novel compounds that satisfy one or more desired properties.

$$\mathcal{G} \sim p_\theta(\mathcal{G} \mid \mathcal{C}_{\text{prop}}), \tag{7}$$

where $\mathcal{C}_{\text{prop}}$ specifies target conditions such as physicochemical properties, biological activities, or cellular responses. Depending on the modeling paradigm, molecular generation may be implemented using autoregressive models, variational autoencoders, diffusion models, or flow-based generative frameworks.

Rather than generating molecules from scratch, MO aims to improve existing lead compounds while preserving desirable structural characteristics. The objective is to modify a molecule so that specific properties are enhanced without completely changing its chemical scaffold.

$$\mathcal{G}_{\text{new}} = f_{\text{MO}}(\mathcal{G}_{\text{old}} \mid \mathcal{C}_{\text{prop}}), \tag{8}$$

subject to

$$\text{Sim}(\mathcal{G}_{\text{old}}, \mathcal{G}_{\text{new}}) \geq \gamma, \tag{9}$$

where $\text{Sim}(\cdot, \cdot)$ denotes a structural similarity metric such as the Tanimoto coefficient. Molecular optimization is particularly important during lead optimization stages of drug development.

### *3.4. Closed-Loop Design Framework*

The above discriminative and generative tasks naturally form a closed-loop design framework. Predictive models provide evaluation signals for generated molecules, while generative models continuously explore new regions of chemical space. Through iterative interaction, the two components guide molecular discovery toward increasingly desirable candidates. This process can be expressed as a multi-objective optimization problem:

$$\mathcal{G}^* = \arg\max_{\mathcal{G}} \left( \sum_k \alpha_k f_k\big(X_k(\mathcal{G})\big) + \beta\, \text{Sim}(\mathcal{G}, \mathcal{G}_{\text{seed}}) \right), \tag{10}$$

where $X_k(\mathcal{G})$ denotes the task-specific input associated with the $k$-th predictive objective, containing the candidate molecule $\mathcal{G}$ together with any additional protein, cellular, or partner-drug information required by the corresponding task. The coefficients $\alpha_k$ and $\beta$ control their relative importance. Here, $\mathcal{G}_{seed}$ represents an initial lead molecule or reference compound, and the similarity term encourages the optimized molecule to preserve desirable structural characteristics while improving target properties. Overall, this formulation connects molecular representation learning, predictive modeling, and generative design within a unified computational framework for graph-structured small-molecule drug discovery and development.

## 4. Algorithmic Paradigms

Building upon the task formulations introduced in Section 3, this section reviews representative algorithmic paradigms for graph-structured small-molecule drug discovery and development, corresponding to Layer 4 in Figure 1. Rather than organizing methods solely by model architecture, we categorize recent advances according to their underlying discovery objectives and biological contexts. Specifically, we focus on three major directions: lead discovery tasks, safety and ADMET assessment, and molecular generation and optimization. Together, these directions form the methodological backbone of deep learning-driven drug discovery systems.

### *4.1. Lead Discovery Tasks*

Lead discovery represents one of the primary stages of small-molecule drug development. Depending on the underlying biological hypothesis, computational approaches can generally be divided into target-based and phenotype-based paradigms. Within graph-structured molecular learning, these paradigms correspond to modeling interactions among molecules, proteins, and cellular systems. As introduced in Section 3, representative tasks include Drug-Target Interaction and Affinity Prediction (DTI/DTA), Drug Response Prediction (DRP), and Drug Perturbation Prediction (DPP). These tasks capture complementary biological scales: DTI/DTA focuses on molecule-protein interactions, DRP summarizes downstream cellular phenotypic responses, while DPP characterizes high-dimensional transcriptomic changes induced by perturbations. Together, they provide a multi-scale view from target engagement to cellular response and gene-level reprogramming. From a drug-development perspective, these tasks can also be mapped to distinct decision points in a typical workflow. DTI/DTA primarily supports target validation and early compound-target prioritization; DRP evaluates drug efficacy and sensitivity in cellular and other preclinical models; and DPP helps characterize mechanisms of action by resolving treatment-induced transcriptomic changes. This progression should not be interpreted as strictly linear, since evidence from response or perturbation studies can feed back to target selection and compound prioritization during iterative drug development. Representative methods for these tasks are summarized in Tables 2–4.

#### 4.1.1. Drug-Target Interaction and Affinity

As formulated in Section 3, DTI/DTA aims to model the interaction between a small molecule and a biological target. Depending on the prediction objective, the task may involve binary interaction classification or continuous affinity estimation. Typical inputs include molecular representations derived from SMILES strings, molecular graphs, or three-dimensional conformations, together with protein sequence representations or structural binding pocket information. As summarized in Table 2, the development of DTI/DTA models has gradually evolved from sequence-based matching frameworks toward multimodal representation learning architectures. Early methods commonly employed convolutional or recurrent neural networks to independently encode drug and protein sequences before combining them through concatenation or attention mechanisms. More recent approaches increasingly integrate heterogeneous modalities, including molecular graphs, protein language models, structural pockets, knowledge graphs, and three-dimensional molecular conformations.

A major research trend in recent years has been the development of more effective cross-modal fusion strategies. Representative methods such as R-DTI [22], DrugCMF [23], GRAM-DTI [24], and MlanDTI [25] incorporate cross-attention mechanisms, contrastive learning objectives, confidence-aware fusion, and domain adaptation strategies to improve the alignment between molecular and protein representations. Compared with earlier concatenation-based approaches, these methods better capture the complementary information contained in heterogeneous biological modalities. However, heterogeneous fusion may also suffer from modality dominance or optimization imbalance when modalities differ substantially in predictive strength, causing weaker but complementary representations to be under-optimized. Balanced multimodal learning studies have shown that a dominant modality can suppress the optimization of other branches, and adaptive gradient modulation provides an effective strategy for regulating modality-specific optimization according to their relative contributions [26]. These findings highlight the importance of explicitly balancing modality contributions during heterogeneous feature fusion.

Another important challenge is OOD generalization. In practical drug discovery scenarios, models frequently encounter unseen compounds, previously uncharacterized targets, or entirely novel drug-target combinations. As indicated in Table 2, recent studies increasingly evaluate models under Unseen-Drug (●), Unseen-Target (●), and Unseen-Both (●) settings. To address these challenges, approaches such as Otter-Knowledge [27] and MoseDTI [28] leverage biomedical knowledge graphs and mixture-of-experts architectures to improve generalization under sparse supervision. More specifically, these methods incorporate heterogeneous relational information at different levels. DTI-MGNN [29] integrates multiple drug- and protein-related networks and constructs complementary topology- and feature-based graphs, allowing graph attention and shared graph convolution to capture relational information from different views. MoseDTI [28] introduces knowledge-graph-derived representations as complementary extrinsic information and adaptively combines them with molecular and protein features. Otter-Knowledge [27] further performs relation-aware propagation over typed biomedical edges using relation-specific transformations. Such relational supervision can provide additional evidence when direct DTI observations are sparse and may facilitate transfer to previously unseen drugs or targets, although its effectiveness depends on knowledge-graph coverage and relation quality. In parallel, recent studies have explored explicit three-dimensional structural modeling. For example, LigoSpace [30] incorporates geometric information from protein binding pockets, while DrugCMF [23] and R-DTI [22] jointly model molecular topology and three-dimensional conformations to improve affinity estimation.

**Table 2.** Summary of representative methods for DTI/DTA tasks.

| Methods | Drug | Drug En. | Target | Target En. | Fusion M. | Data. & Tasks | OOD |
|---|---|---|---|---|---|---|---|
| DrugCLIP [31] | 3D | Uni-mol [32] | 3D Pocket | Uni-mol | - | : ☆ | |
| APLM [33] | MFP | MLP | Seq. | PLM | Concat | : ☆ | |
| DTI-MGNN [29] | SMILES | KGE | Seq. | KGE | Attention | : ☆ | |
| HyperAttentionDTI [34] | SMILES | CNNs | Seq. | CNNs | Attention | : ☆ | |
| Drug-VQA [35] | SMILES | LSTM | Distance Map | CNNs | Concat | : ★ | |
| MolTrans [36] | SMILES | Trans. | Seq. | Trans. | MP | : ☆ | |
| SiamDTI [37] | SMILES | Trans. | Seq. | Trans. | BAN | : ☆ | |
| DTIAM [38] | Graph | Trans. | Seq. | Trans. | Concat | : ☆★ | |
| DrugBAN [39] | Graph | GNNs | Seq. | CNNs | BAN | : ☆ | |
| PSC-CPI [40] | Graph | GNNs | Seq., Graph | HRNN, GAT | MP | : ★ | |
| Cross-interaction [41] | Graph | GNNs | Seq., Graph | HRNN, GAT | Concat | : ☆ | |
| MGNDTI [42] | SMILES, Graph | GCNs, RN | Seq. | RN | GN | : ☆ | |
| Perceiver-CPI [43] | SMILES, MFP | GNNs, MLP | Seq. | CNNs | Attention | : ★ | |
| MlanDTI [25] | SMILES | PLM | Seq. | PLM, CNNs | Attention | : ☆ | |
| GRAM-DTI [24] | SMILES | PLM | Seq. | PLM | Concat | : ☆ | |
| R-DTI [22] | SMILES, Graph | PLM, CNNs, GAT | Seq., 3D | PLM, CNNs, RN | Concat | : ☆ | - |
| MoseDTI [28] | Graph | GNNs, KGE | Seq. | PLM, CNNs, KGE | Attention | : ☆ | - |
| DrugCMF [23] | SMILES, 3D | PLM | Seq., 3D | PLM | Attention | : ☆ | |
| Otter-Knowledge [27] | SMILES, MFP | PLM, GNNs | Seq. | PLM, GNNs | Concat | : ★ | |
| LigoSpace [30] | Graph, 3D | GNNs | Graph, 3D | GNNs | MP | : ★ | - |
| BioCG [44] | SMILES | PLM | Seq. | Trans. | Gen. | : ☆ | |

Notations: ① Drug: morgan fingerprint (MFP), 3D conformation (3D). ② Techniques: Multilayer Perceptron (MLP), Convolutional Neural Networks (CNNs), Graph Convolutional Networks (GCNs), Graph Neural Networks (GNNs), Graph Attention Network (GAT), Retentive Networks (RN), Directed Message-Passing Neural Network (DMPNN); Long Short-Term Memory (LSTM), Transformer (Trans.), Protein Language Model (PLM), Hierarchical Recurrent Neural Networks (HRNN), Knowledge Graph Embedding (KGE); Manhattan Product (MP), Bilinear Attention Network (BAN), Gating Network (GN), Neural Factorization Machine (NFM), Generative decoding (Gen.). ③ Datasets (symbols below are local to this table): Davis, KIBA, DUDE, C.elegans, Human, BindingDB, BioSNAP, Metz, LIT-PCBA, Karimi. Yamanishi_08, Hetionet, DGIDB, GNBR, DrugBank, DrugCentral, DTI-DG, DTIGN, SIU. ④ Tasks: ☆ classification task, ★ regression task. ⑤ Out of distribution: Seen-Both, Unseen-Drug, Unseen-Target, Unseen-Both.

### 4.1.2. Drug Response Prediction

DRP aims to estimate the sensitivity of a cellular system to a given compound. Unlike molecular property prediction or DTI/DTA, DRP explicitly incorporates cellular context and therefore serves as an important task for precision oncology and personalized therapy design. Common prediction targets include cell viability, area under the dose-response curve (AUC), and half-maximal inhibitory concentration ($IC_{50}$).

As summarized in Table 3, DRP models increasingly combine molecular representations with transcriptomic and multi-omics profiles. Early approaches primarily relied on multilayer perceptrons or convolutional neural networks operating on predefined molecular descriptors and gene expression features. More recent methods employ graph neural networks and Transformer architectures to jointly model molecular structures and cellular systems within a unified learning framework. The availability of diverse biological profiles has also significantly influenced model development. As shown in Table 3, modern DRP systems increasingly integrate multiple omics modalities, including gene expression profiles (E), somatic mutations (M), copy number variations (C), epigenomic features (Epi), and biological interaction networks (Net). Such multi-modal integration enables a more comprehensive characterization of cellular states and improves predictive performance across heterogeneous cancer types.

A central challenge in DRP is the distribution shift between preclinical datasets and clinical applications. Models trained on cell lines frequently experience substantial performance degradation when deployed on patient-derived samples, highlighting the difficulty of generalizing across biological systems. To address this issue, recent methods increasingly employ domain adaptation, transfer learning, contrastive learning, and domain generalization strategies. Representative examples include GraphCDR [45], CLDR [46], MSDA [47], and several recent transfer-learning frameworks. More specialized approaches such as TransDRP [48], MACB-DRP [49], DeepSADR [50], WISER [51], and PREDICT-AI [52] further explore clinically motivated transfer mechanisms, interpretable interaction modeling, and domain-aware representation learning to improve robustness under realistic deployment conditions.

### 4.1.3. Drug Perturbation Prediction

DPP focuses on modeling transcriptional changes induced by chemical perturbations. Compared with DRP, which predicts a scalar drug-response outcome, DPP aims to estimate high-dimensional post-treatment cellular states and therefore provides a more detailed characterization of drug mechanisms and cellular responses. Typical outputs include gene expression profiles, differential expression signatures, and other transcriptomic representations.

As summarized in Table 4, the development of DPP methods has gradually progressed from latent-space reconstruction models toward increasingly expressive generative frameworks. Early approaches, including

**Table 3.** Summary of representative methods for DRP.

| Methods | Cell Profiles | Drug | Cell Line Arch. | Drug Arch. | Tech. | Datasets & Tasks |
|---|---|---|---|---|---|---|
| Dr. VAE [53] | E | No constr. | VAE | Not req. | Supervised Learning | ■■: ★ |
| DeepTTA [54] | E | Subcomponent | MLP | Trans. | Supervised Learning | ■: ☆ |
| SubCDR [55] | E, R | Subcomponent | CNNs | CNNs | Supervised Learning | ■■■: ☆★ |
| TGSA [56] | M, E, C, Net | Graph | GAT | GIN | Supervised Learning | ■: ★ |
| DIPK [57] | E, Net | Graph | GAE | GNNs, Trans. | Pre-training | ■■: ★ |
| GraphCDR [45] | M, E, C | Graph | DNNs | GNNs | Contrastive Learning | ■■: ☆ |
| CLDR [46] | No constr. | No constr. | Trans. | Trans. | Contrastive Learning | ■: ★ |
| MSDA [47] | No constr. | No constr. | No constr. | No constr. | Domain Generalization | ■■: ★ |
| CODE-AE [58] | R | Not req. | AE | Not req. | Domain Generalization | ■■■: ☆ |
| scDEAL [59] | R, S | Not req. | AE | Not req. | Transfer Learning | ■■■: ☆ |
| WISER [51] | E | No constr. | MLP | Not req. | Weak Supervision Learning | ■■: ☆ |
| drGAT [60] | E, Net | Graph | GAT | GAT | Supervised Learning | ■: ☆ |
| XGDP [61] | E | Graph | CNNs | GAT | Supervised Learning | ■: ★ |
| PASO [62] | M, E, C, Path | Subcomponent | GAT | CNNs | Supervised Learning | ■■: ★ |
| PAM-CDR [63] | M, E | Graph | DNNs | GCN | Supervised Learning | ■: ★ |
| MTEGDRP [64] | M, E, C, Epi | Graph | Trans. | GNNs | Supervised Learning | ■: ★ |
| TransDRP [48] | E | Graph | MLP | GNNs | Domain Adaptation | ■■■: ☆ |
| MACB-DRP [49] | E | No constr. | MLP | Not req. | Domain Adaptation | ■■■: ☆ |
| DeepSADR [50] | R | Graph | GAE | GNNs | Domain Adaptation | ■■■: ☆ |
| PREDICT-AI [52] | M | No constr. | Trans. | Not req. | Domain Adaptation | ■■■: ☆★ |

Notations: ① Datasets: ■ GDSCv1/2, ■ CCLE, ■ TCGA, ■ NCI-60, ■ COSMIC, ■ PubChem, ■ PDTC, ■ DepMap, ■ BioSNAP. ② Profiles: mutation status (M), gene expression profiles (E), copy number variation (C), RNA sequence (R), single-cell RNA sequence (S), epigenomics (Epi), pathway difference features (Path), gene interaction network/protein-protein association network from STRING dataset (Net), molecular graphs (Graph), subcomponent/SMILES fragment (Subcomponent), no need for special module or data (Not req.), no restrictions on the type of module or data (No constr.). ③ Techniques: Multilayer Perceptron (MLP), Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), Transformer (Trans.), Graph Neural Networks (GNNs), Graph Attention Network (GAT), Graph Isomorphism Network (GIN), Graph Convolutional Network (GCN), Autoencoder (AE), Variational Autoencoder (VAE), Graph Autoencoder (GAE). ④ Tasks: ☆ classification task, ★ regression task.

scGen [65] and DeepCellState [66], relied on autoencoder-based architectures to learn perturbation representations. Subsequent methods such as CPA [67] and ChemCPA [68] introduced explicit disentanglement mechanisms that separate basal cellular states from perturbation effects, improving compositional generalization across unseen experimental conditions. Recent studies have increasingly adopted foundation-model and generative-learning paradigms. Methods such as CRISP [69] incorporate single-cell foundation models to improve biological representation learning, while Squidiff [70] and PertDit [71] employ diffusion-based architectures to model complex cellular response distributions. These approaches provide substantially greater flexibility for modeling nonlinear and stochastic perturbation effects than earlier latent-variable frameworks.

Generalization to unseen compounds remains a key challenge in DPP. Early methods were largely restricted to interpolation settings, where compounds or perturbation conditions had already been partially observed during training. To enable prediction for previously unseen molecules, recent frameworks increasingly incorporate explicit molecular encoders. For example, ChemCPA, DeepICER, and XPert represent compounds using molecular graphs or SMILES-derived substructures, allowing learned chemical representations to be transferred into transcriptomic response spaces. This design enables more effective zero-shot or few-shot generalization to novel compounds.

From a data perspective, DPP research is primarily evaluated on two complementary experimental platforms. Bulk transcriptomic datasets, including Connectivity Map (CMap) and LINCS L1000, provide large-scale perturbation measurements across many compounds and cell lines. In contrast, recent single-cell perturbation datasets, such as Sci-Plex, provide cell-level resolution that captures biological heterogeneity hidden in bulk measurements. Consequently, recent DPP models increasingly emphasize cross-platform transferability and multi-scale biological representation learning.

### *4.2. Property and Safety Assessment*

Following lead discovery, candidate compounds must undergo comprehensive evaluation of their physicochemical properties, pharmacokinetic behavior, toxicity risks, and potential adverse interactions. While strong target affinity is a prerequisite for efficacy, successful drug development additionally requires favorable absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles, as well as acceptable safety characteristics under realistic clinical conditions. Within graph-structured molecular learning, Molecular Property Prediction (MPP) and Drug-Drug Interaction Prediction (DDI) constitute two representative tasks for assessing molecular quality and pharmacological safety. Representative methods for these tasks are summarized in Tables 5–7.

**Table 4.** Summary of representative methods for DPP.

| Methods | Cell Profiles | Drug | Cell State Arch. | Drug Arch. | Tech. | Datasets & Tasks |
|---|---|---|---|---|---|---|
| scGen [65] | S | No constr. | VAE | Not req. | Unsupervised Learning | ■: ★ |
| DeepCellState [66] | E | No constr. | AE | Not req. | Unsupervised Learning | ■■: ★ |
| CPA [67] | S | No constr. | AE | Not req. | Supervised Learning | ■: ★ |
| ChemCPA [68] | S | Graph | AE | GNNs | Supervised Learning | ■: ★ |
| CellOT [72] | S | No constr. | Neural OT | Not req. | Unsupervised Learning | ■: ★ |
| CPA-Screen [73] | S | No constr. | AE | Not req. | Supervised Learning | ■: ★ |
| CycleCDR [74] | S | No constr. | GAN | Not req. | Cycle Consistency Learning | ■: ★ |
| PRNet [75] | E | Graph | VAE | GNNs | Supervised Learning | ■■: ★ |
| DeepICER [76] | E | Subcomponent | MLP | Trans. | Supervised Learning | ■: ★ |
| XPert [77] | S, E | Subcomponent | Trans. | Trans. | Supervised Learning | ■■: ★ |
| CRISP [69] | S | No constr. | scFM, VAE | Not req. | Transfer Learning | ■: ☆★ |
| Squidiff [70] | S | No constr. | Diffusion | Not req. | Supervised Learning | ■: ★ |
| PertDit [71] | E | Subcomponent | Diffusion | Trans. | Supervised Learning | ■: ★ |

Notations: ① Datasets & Tasks: ■ LINCS L1000, ■ Connectivity Map (CMap), ■ Single-cell RNA-seq perturbation datasets (i.e., Sci-Plex). ② Profiles: single-cell RNA sequence (S), bulk gene expression profiles (E), molecular graphs (Graph), subcomponent/SMILES fragment (Subcomponent), no need for special module or data (Not req.), No restrictions on the type of module or data (No constr.). ③ Techniques: Autoencoder (AE), Variational Autoencoder (VAE), Neural Optimal Transport (Neural OT), Generative Adversarial Network (GAN), Diffusion Model/Transformer (Diffusion), Single-cell Foundation Model (scFM), Transformer (Trans.). ④ Tasks: ☆ classification task, ★ regression task.

#### 4.2.1. Molecular Property Prediction

As introduced in Section 3, MPP aims to infer physicochemical, quantum-mechanical, pharmacokinetic, or toxicity-related properties from molecular representations. Because these properties often determine downstream drug efficacy and safety, MPP serves as one of the most fundamental tasks in molecular machine learning. A long-standing challenge in MPP arises from the existence of activity cliffs and property cliffs, where structurally similar molecules exhibit substantially different biological or physicochemical properties. Such discontinuities make generalization difficult and limit the effectiveness of purely topology-based molecular representations. Consequently, recent research has increasingly explored richer molecular representations that incorporate geometric and physical information beyond two-dimensional molecular graphs.

As summarized in Tables 5 and 6, the development of MPP models has broadly followed three directions. First, geometric learning methods such as DimeNet [19], PaiNN [13], SphereNet [78], Equiformer [79], ViSNet [80], and GotenNet [81] incorporate three-dimensional molecular conformations and geometric inductive biases to better capture spatial interactions. These approaches have substantially improved performance on quantum chemistry and molecular dynamics benchmarks, including QM9, MD17, MD22, and Open Catalyst datasets. Second, large-scale self-supervised pre-training has become an increasingly important paradigm for molecular representation learning. Representative frameworks such as Uni-Mol [32], Transformer-M [82], MOL-AE [83], Uni-Mol2 [84], JMP [85], MOLEBLEND [86], and SliDe [87] learn transferable molecular representations from large unlabeled molecular corpora and subsequently adapt them to downstream property prediction tasks. Compared with task-specific supervised training, these approaches often exhibit stronger transferability and robustness across diverse molecular benchmarks.

However, the benefit of pre-training is not universal and depends strongly on the relationship between the pre-training and downstream settings. Its relative advantage is generally greater when labeled downstream data are limited and the pre-training corpus provides transferable molecular priors, whereas the benefit may become smaller when sufficient task-specific supervision is already available. Moreover, mismatches in data distributions, pre-training objectives, or representation scales can lead to limited transfer or even negative transfer. Previous studies have shown that naive graph pre-training may provide only marginal gains and can even degrade performance on some downstream molecular tasks [88,89]. Cross-scale mismatch is another important consideration: GraphMVP and 3D Infomax explicitly align 2D molecular topology with 3D geometric information and demonstrate that carefully designed cross-view objectives can improve downstream transfer [90,91]. These observations highlight the importance of task-aware pre-training, distribution alignment, and appropriate cross-scale adaptation rather than direct reuse of pretrained representations. Third, recent studies have begun incorporating electronic structure information into molecular representations. Rather than relying solely on atomic connectivity and geometric coordinates, methods such as UniField [17] and EDG [92] integrate electron-density-related information to provide a more physically grounded description of molecular interactions. This direction aligns closely with the transition from topology-centered representations toward unified molecular representations that jointly encode structural, geometric, and electronic information.

Overall, recent progress in MPP reflects a gradual movement toward more comprehensive molecular representations capable of capturing multiple levels of chemical information simultaneously. This trend is

**Table 5.** Summary of representative methods for MPP without pre-training.

| Methods | Rep. | Arch. | Dataset | Model Name | Rep. | Arch. | Dataset |
|---|---|---|---|---|---|---|---|
| DimeNet [19] | 3D graph | GNNs | ●$^{3}$ ■$^{1}$ | MoleOOD [93] | 2D graph | GNNs | ◆$^{3,4,6,8}$ |
| PaiNN [13] | 3D graph | MPNN | ●$^{3}$ ■$^{1}$ | GraphSAM [94] | 2D graph | GT | ◆$^{2,3,5,8}$ ▼$^{1,3}$ |
| SphereNet [78] | 3D graph | GNNs | ●$^{3}$ ■$^{1}$ ▲$^{3}$ | HDGNN [95] | 3D graph | GNNs | ●$^{3}$ ▲$^{1,3}$ |
| TorchMD-Net [96] | 3D graph | Trans. | ●$^{3}$ ■$^{1}$ ★$^{5}$ | ViSNet [80] | 3D graph | Trans. | ●$^{3}$ ■$^{1,2,3}$ ★$^{8}$ |
| Allegro [97] | 1D desc. | MLP | ●$^{3}$ ■$^{2}$ ★$^{7}$ | Polymer Walk [98] | 2D graph | GNNs | ★$^{9}$ |
| Geo-DEG [99] | 2D graph | GNNs | ◆$^{2}$ ▼$^{2,3}$ ★$^{9}$ | QTAIM [100] | 3D graph | GNNs | ●$^{2,3}$ ◆$^{5}$ |
| Equiformer [79] | 3D graph | GT | ●$^{3}$ ■$^{1}$ ▲$^{1}$ | SCHull [101] | 3D graph | GNNs | ■$^{1}$ |
| Molformer [102] | 2/3D graph | Trans. | ●$^{1,2,3}$ ◆$^{2,6,8}$ | E2Former [103] | 2/3D graph | GT | ▲$^{1,2}$ ★$^{4}$ |
| GotenNet [81] | 3D graph | GNNs | ●$^{3}$ ■$^{2,3}$ ★$^{8}$ | UniField [17] | 3D/ED | Trans. | ●$^{3}$ ★$^{1,6}$ |

Notations: ① Datasets: ●$^{n}$ QM Benchmarks ($n$ indicates specific sets: 1: QM7; 2: QM8; 3: QM9); ■$^{n}$ MD Benchmarks ($n$ indicates specific sets: 1: MD17; 2: rMD17; 3: MD22); ▲$^{n}$ OC Benchmarks ($n$ indicates specific sets: 1: OC20; 2: OC22; 3: IS2RE); ◆$^{n}$ Classification benchmarks ($n$ indicates specific sets: 1: MUV; 2: ClinTox; 3: SIDER; 4: HIV; 5: Tox21; 6: BACE; 7: ToxCast; 8: BBBP); ▼$^{n}$ Regression benchmarks ($n$ indicates specific sets: 1: ESOL; 2: FreeSolv; 3: Lipophilicity); ★$^{n}$ Other specific/auxiliary benchmarks ($n$ indicates specific sets: 1: QMugs and GEOM-Drugs; 2: COMPAS-1D; 3: PCQM4Mv2; 4: MatBench, QMOF, SPICE, ANI-1x, and Transition-1x; 5: ANI-1x; 6: EDBench and PCQM4Mv2; 7: 3BPA, $Li_3PO_4$, and Ag bulk crystal; 8: Molecule3D; 9: Polymer properties & specific sets, e.g., CROW, Permeability, Group, HOPV, PTC, and DILI). ② Representations (Rep.): 1D desc.: 1D descriptors with position; 2D/3D: Joint use of 2D molecular topology and 3D geometry; 3D: 3D molecular conformations; 3D/ED: Molecular geometry enhanced by electron-density-related information. ③ Architectures (Arch.): Graph Neural Networks (GNNs), Message Passing Neural Networks (MPNN), Graph Transformer (GT), Transformer (Trans.).

**Table 6.** Summary of representative methods for MPP with pre-training.

| Methods | Rep. | Arch. | Dataset | Methods | Rep. | Arch. | Dataset |
|---|---|---|---|---|---|---|---|
| InfoGraph [104] | 2D graph | GIN | ●$^{3}$ | MOL-AE [83] | 3D graph | Trans. | ●$^{1-3}$ ◆$^{1-8}$ ▼$^{1-3}$ ▲$^{1}$ |
| 3D Infomax [91] | 2/3D graph | PNA | ●$^{3}$ ★$^{1}$ | Uni-Mol2 [84] | 2D graph | Trans. | ●$^{3}$ ★$^{2}$ |
| GeomGCL [105] | 2/3D graph | GNNs | ◆$^{2,3,6,8}$ ▼$^{1-3}$ | UniCorn [106] | 2/3D graph | Trans. | ●$^{3}$ ■$^{2}$ |
| GEM [107] | 2/3D graph | GeoGNN | ●$^{1-3}$ ◆$^{1-8}$ ▼$^{1-3}$ | Uni-Mol+ [108] | 3D graph | Trans. | ▲$^{1}$ ★$^{3}$ |
| GraphMAE [109] | 2D graph | GNNs | ◆$^{1-8}$ | JMP [85] | 3D graph | GNNs | ●$^{3}$ ■$^{2}$ ▲$^{1,2}$ ★$^{4}$ |
| MolCLR [110] | 2D graph | GNNs | ◆$^{2-8}$ ▼$^{1-3}$ | MOLEBLEND [86] | 2/3D graph | Trans. | ●$^{3}$ ◆$^{1-8}$ ★$^{3}$ |
| GraphMVP [90] | 2/3D graph | GNNs | ◆$^{1-8}$ | SliDe [87] | 3D graph | Trans. | ●$^{3}$ ■$^{1}$ ★$^{5}$ |
| Uni-Mol [32] | 3D graph | Trans. | ●$^{1-3}$ ◆$^{1-8}$ ▼$^{1-3}$ | MolMCL-GCN [111] | 2D graph | GT | ◆$^{2-8}$ |
| Transformer-M [82] | 2/3D graph | Trans. | ●$^{3}$ ■$^{1}$ ★$^{3}$ | EDG [92] | 3D/ED | KD/GNNs | ●$^{3}$ ■$^{1}$ ★$^{6}$ |

Notations: ① Datasets: ●$^{n}$ QM Benchmarks ($n$ indicates specific sets: 1: QM7; 2: QM8; 3: QM9); ■$^{n}$ MD Benchmarks ($n$ indicates specific sets: 1: MD17; 2: rMD17; 3: MD22); ▲$^{n}$ OC Benchmarks ($n$ indicates specific sets: 1: OC20; 2: OC22; 3: IS2RE); ◆$^{n}$ Classification benchmarks ($n$ indicates specific sets: 1: MUV; 2: ClinTox; 3: SIDER; 4: HIV; 5: Tox21; 6: BACE; 7: ToxCast; 8: BBBP); ▼$^{n}$ Regression benchmarks ($n$ indicates specific sets: 1: ESOL; 2: FreeSolv; 3: Lipophilicity); ★$^{n}$ Other specific/auxiliary benchmarks ($n$ indicates specific sets: 1: QMugs and GEOM-Drugs; 2: COMPAS-1D; 3: PCQM4Mv2; 4: MatBench, QMOF, SPICE, ANI-1x, and Transition-1x; 5: ANI-1x; 6: EDBench and PCQM4Mv2; 7: 3BPA, $Li_3PO_4$, and Ag bulk crystal; 8: Molecule3D; 9: Polymer properties & specific sets, e.g., CROW, Permeability, Group, HOPV, PTC, and DILI). ② Representations (Rep.): 2D/3D: Joint use of 2D molecular topology and 3D geometry; 3D: 3D molecular conformations; 3D/ED: Molecular geometry enhanced by electron-density-related information. ③ Architectures (Arch.): Graph Neural Networks (GNNs), Graph Transformer (GT), Transformer (Trans.), Graph Isomorphism Network (GIN), Principal Neighbourhood Aggregation (PNA), Geometry-based GNN (GeoGNN), Knowledge Distillation applied to GNNs (KD/GNNs).

particularly important for improving robustness on challenging molecular property benchmarks and enhancing generalization across diverse chemical domains.

#### 4.2.2. Drug-Drug Interaction

DDI prediction aims to identify potential interactions that arise when multiple drugs are administered simultaneously. Depending on the application scenario, the task may involve binary interaction detection, multi-class interaction classification, interaction recommendation, or severity/mechanism prediction. As discussed in Section 3, DDI can be viewed as a relational learning problem over molecular pairs and their associated biological contexts.

Graph neural networks have become one of the dominant modeling paradigms for DDI prediction because they naturally represent molecular structures and pairwise interaction patterns. As summarized in Table 7, recent studies further incorporate knowledge graphs, contrastive learning, information bottleneck objectives, causal inference, reinforcement learning, neural architecture search, and language-based explanation generation to improve representation quality, generalization performance, and interpretability.

One major research direction focuses on identifying biologically meaningful substructures responsible for drug interactions. Early graph-structured models primarily relied on global molecular representations, whereas more recent approaches increasingly model interaction-relevant motifs, pharmacophores, and adaptive subgraphs. For example, MOTOR [112] introduces motif-level representations that explicitly capture hierarchical structural information across multiple granularities. PC-DDI [113] further constructs pharmacophore-aware representations and incorporates causal inference to reduce shortcut correlations between molecular scaffolds and interaction labels. RISE-DDI [114] formulates adaptive subgraph extraction as a reinforcement learning problem, while CSSE-DDI [115] employs neural architecture search to jointly optimize subgraph sampling and graph encoding strategies.

**Table 7.** Summary of representative methods for DDI.

| Methods | Drug | Tech. | Data. & Tasks | Methods | Drug | Tech. | Data. & Tasks |
|---|---|---|---|---|---|---|---|
| ZeroDDI [116] | SMILES | SEL | ■: ✪ | DeepDDI [117] | SMILES, Text | SSP | ■: ☆ |
| MKG-FENN [118] | SMILES | KG | ■: ☆ | DDKG [119] | SMILES, Graph | KG | ■■: ☆ |
| BI-GNN [120] | Graph | GNN | ■■: ☆ | MOTOR [112] | Graph | GNNs, Motif | ■■■: ☆ |
| CGIB [121] | Graph | GIB | ■■: ☆★ | PC-DDI [113] | Graph | GNNs, Causal | ■■: ☆ |
| IGIB-ISE [121] | Graph | IGIB | ■■: ☆★ | RISE-DDI [114] | Graph | GNNs, RL | ■■: ☆ |
| PHGL-DDI [122] | Graph | CL | ■■: ☆ | S2VM [123] | 2D Image | Trans., SSL | ■■: ☆ |
| DANN-DDI [124] | Graph | Attention | ■■: ☆ | CSSE-DDI [115] | Graph | GNNs, NAS | ■■: ☆ |
| TIGER [125] | Graph | GNN | ■■■: ☆ | ExDDI [126] | SMILES, Text | Trans., NLG | ■■: ☆, ★ |
| GoGNN [127] | Graph | GNN | ■■■: ☆ | MolecBioNet [128] | Graph | GNNs, KG, Pooling, MI | ■■: ☆ |
| CSGNN [129] | Graph | CL | ■: ☆ | K-Paths [130] | KG Entity | Path, LLMs, GNNs | ■■: ☆ |

Notations: ① Techniques: Enhanced Learning (SEL), Knowledge Graph (KG), Graph Neural Networks (GNNs), Graph Information Bottleneck (GIB), Interactive Graph Information Bottleneck (IGIB), Contrastive Learning (CL), Structural Similarity Profile (SSP), Transformer (Trans.), Reinforcement Learning (RL), Neural Architecture Search (NAS), Motif-oriented representation (Motif), Causal inference (Causal), Self-supervised Learning (SSL), Natural Language Generation (NLG), Mutual Information (MI), Large Language Models (LLMs), Path reasoning (Path). ② Datasets (symbols below are local to this table): ■ DrugBank, ■ KEGG, ■ SIDER, ■ ChCh-Miner, ■ ZhangDDI, ■ DRUGCOMBO, ■ OGB-biokg, ■ TWOSIDES, ■ Ryu et al.'s dataset, ■ DDInter. ③ Tasks: ☆ multi-class classification task, ☆ classification task, ★ regression task, ✪ recommendation task, ☆ generation task.

Another emerging direction explores alternative molecular modalities and self-supervised learning objectives. S2VM [123], for example, represents molecular structures through two-dimensional molecular images and learns interaction-aware visual representations via self-supervised reconstruction objectives. Such approaches reduce dependence on large labeled interaction datasets and demonstrate promising generalization under unseen-drug and unseen-pair settings. Beyond predictive accuracy, recent DDI studies have increasingly emphasized interpretability and reasoning-oriented prediction. MolecBioNet [128] integrates molecular-level structural information with biomedical knowledge graphs by treating each drug pair as a unified entity, and introduces context-aware subgraph pooling and attention-guided influence pooling to provide both network-level and molecular-level explanations. K-Paths [130] approaches interpretability from a knowledge-graph reasoning perspective: it retrieves diverse multi-hop paths from biomedical KGs and converts them into structured reasoning evidence that can be used by LLMs or GNNs for interaction prediction, particularly under inductive settings involving emerging entities. Complementarily, ExDDI [126] reformulates DDI prediction as a natural-language explanation generation problem, aiming to reveal pharmacodynamic and pharmacokinetic mechanisms rather than only outputting interaction labels. Together, these methods suggest that DDI research is moving beyond black-box prediction toward models that expose molecular substructures, biomedical reasoning paths, or textual mechanistic rationales, thereby improving the transparency of computational drug safety assessment.

From a data perspective, DDI research relies on several complementary categories of resources. Drug-centric databases such as DrugBank, KEGG, PubChem, and DrugCentral provide molecular and pharmacological knowledge. Adverse-effect resources, including TWOSIDES and SIDER, offer large-scale records of clinically observed interaction outcomes. In addition, biomedical knowledge graphs such as DRKG and Bio2RDF provide heterogeneous relational information linking drugs, proteins, diseases, pathways, and biological processes. The integration of these complementary data sources continues to play a central role in improving both predictive accuracy and biological interpretability.

### *4.3. Molecular Generation and Optimization*

Modern drug discovery increasingly relies on generative modeling to explore and optimize molecular structures beyond the limits of exhaustive screening. Within graph-structured molecular learning, generative approaches are typically divided into two complementary tasks: Molecular Generation (MG), which aims to design novel molecules from desired conditions, and Molecular Optimization (MO), which seeks to improve existing lead compounds while preserving favorable structural characteristics. Representative methods for these tasks are summarized in Tables 8 and 9.

#### 4.3.1. Molecular Generation

As introduced in Section 3, MG aims to learn a distribution over chemically valid molecular structures and generate novel compounds that satisfy desired design objectives. Depending on the application scenario, generation may be unconditional or guided by molecular properties, protein targets, biological phenotypes, natural-language descriptions, or synthetic constraints.

Early molecular generation methods primarily focused on learning valid molecular representations through variational autoencoders (VAEs), generative adversarial networks (GANs), and normalizing flows. Representative

**Table 8.** Summary of molecular generation methods.

| Models | Category | Backbones | Conditions |
|---|---|---|---|
| 3DLinker [131] | VAE | $p(\mathcal{G}', \mathcal{M}' \mid \mathcal{G}_{\text{frag}}, \mathcal{M}_{\text{frag}})$ | – |
| CGCF [132] | Flow | $p_\theta(\mathbf{R} \mid \mathcal{G}) = \int p(\mathbf{R} \mid \mathbf{d}, \mathcal{G}) p_\theta(\mathbf{d} \mid \mathcal{G})\, d\mathbf{d}$ | – |
| LatentGAN [133] | GAN | $p(M' \mid \mathbf{z}_M, \mathbf{z}_{\text{fake}})$ | – |
| GDSS [134] | Diffusion | $\begin{cases} d\mathcal{X}_t = f_{1,t}(\mathcal{X}_t)dt + g_{1,t} d\mathbf{w}_1 - g_{1,t}^2 s_{\theta,t}(\mathcal{X}_t, \mathcal{A}_t)dt, \\ d\mathcal{A}_t = f_{2,t}(\mathcal{A}_t)dt + g_{2,t} d\mathbf{w}_2 - g_{2,t}^2 s_{\theta,t}(\mathcal{X}_t, \mathcal{A}_t)dt. \end{cases}$ | – |
| GeoLDM [135] | Diffusion | $\begin{cases} p_\theta(\mathbf{z}_{\mathcal{X},t-1}, \mathbf{z}_{\mathbf{R},t-1} \mid \mathbf{z}_{\mathcal{X},t}, \mathbf{z}_{\mathbf{R},t}) \\ \quad = p_\theta(C\mathbf{z}_{\mathcal{X},t-1}, \mathbf{z}_{\mathbf{R},t-1} \mid C\mathbf{z}_{\mathcal{X},t}, \mathbf{z}_{\mathbf{R},t}), \\ C\mathbf{z}_{\mathcal{X},t-1} + \mathbf{t}, \mathbf{z}_{\mathbf{R},t-1} = \epsilon_\theta(C\mathbf{z}_{\mathcal{X},t} + \mathbf{t}, \mathbf{z}_{\mathbf{R},t}, t). \end{cases}$ | – |
| GF-VAE [136] | Flow | $p(\mathcal{X}' \mid \mathbf{z}_{\mathcal{X}}, \mathcal{A}) p(\mathcal{A}' \mid \mathbf{z}_{\mathcal{A}}) = p(\mathbf{z}_{\mathcal{X}}) \left\lvert \det \left( \frac{\partial f_\omega^{\text{atom}}(\mathcal{X}', \mathcal{A})}{\partial \mathcal{X}'} \right) \right\rvert p(\mathbf{z}_{\mathcal{A}}) \left\lvert \det \left( \frac{\partial f_\theta^{\text{bond}}(\mathcal{A}')}{\partial \mathcal{A}'} \right) \right\rvert.$ | Property |
| MolHF [137] | Flow | $\mathcal{G}' = (\mathcal{X}', \mathcal{A}') = f_{\mathcal{X}\mid\mathcal{A}}^{-1}\left(\mathbf{z}_{\mathcal{X}}, f_{\mathcal{A}}^{-1}(\mathbf{z}_{\mathcal{A}})\right)$ | Property |
| TransORGAN [138] | GAN | $\begin{cases} M' = \mathbf{G}_\theta(\tilde{M}), \\ \min\limits_\theta \max\limits_\phi \tilde{V}(\mathbf{G}_\theta, \mathbf{D}_\phi) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log \mathbf{D}_\phi(x)] \\ \qquad + \mathbb{E}_{z \sim p_z(z)}[\log(1 - \mathbf{D}_\phi(\mathbf{G}_\theta(z)))]. \end{cases}$ | Property |
| MOOD [139] | Diffusion | $d\mathcal{G}_t = \left[ f_t(\mathcal{G}_t) - g_t^2 \nabla_{\mathcal{G}_t} \log p_t(\mathcal{G}_t \mid \mathbf{y}_o = \lambda_o) \right] dt + g_t d\mathbf{w}$ | Property |
| Graph DiT [140] | Diffusion | $\hat{p}_\theta(\mathcal{G}_{t-1} \mid \mathcal{G}_t, \mathcal{C}_{\text{prop}}) = \log p_\theta(\mathcal{G}_{t-1} \mid \mathcal{G}_t) + s\Big( \log p_\theta(\mathcal{G}_{t-1} \mid \mathcal{G}_t, \mathcal{C}_{\text{prop}}) - \log p_\theta(\mathcal{G}_{t-1} \mid \mathcal{G}_t) \Big).$ | Property |
| LDMol [141] | Diffusion | $\begin{aligned} \mathbf{z}_M &= \text{Enc}_\phi(M), \\ p_\theta(\mathbf{z}_0 \mid \mathbf{z}_t, \mathbf{c}_{\text{text}}) &= \text{Denoise}_\theta(\mathbf{z}_t, t, \mathbf{c}_{\text{text}}), \\ M' &= \text{Dec}_\psi(\mathbf{z}_0). \end{aligned}$ | Text |
| DecompDiff [142] | Diffusion | $\begin{cases} p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t, \mathbf{x}_0, \mathcal{P}) = \prod\limits_{i=1}^{\lvert\mathcal{X}\rvert} \sum\limits_{k=1}^{\mathcal{K}} \eta_{ik} \mathcal{N}\left( \tilde{\mathbf{x}}_{t-1,k}^{(i)}; \tilde{\mu}_t, \tilde{\beta}_t \Sigma_k \right), \\ \tilde{\mu}_t = \frac{\sqrt{\alpha_t}(1 - \overline{\alpha}_{t-1})}{1 - \overline{\alpha}_t} \tilde{\mathbf{x}}_{t,k}^{(i)} + \frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1 - \overline{\alpha}_t} \tilde{\mathbf{x}}_{0,k}^{(i)}. \end{cases}$ | Protein |
| Apo2Mol [143] | Diffusion | $\begin{aligned} p_\theta(\mathcal{L}', \mathcal{P}_{\text{holo}} \mid \mathcal{P}_{\text{apo}}) &= \prod_{t=1}^{T} p_\theta(\mathcal{L}_{t-1}, \mathcal{P}_{\text{holo}}^{t-1} \mid \mathcal{L}_t, \mathcal{P}_{\text{holo}}^t, \mathcal{P}_{\text{apo}}), \\ \mathcal{L}' &= \text{GenerateLigand}(\mathcal{P}_{\text{apo}}). \end{aligned}$ | Protein |
| RxnFlow [144] | GFlowNet | $\begin{aligned} p_\theta(\tau \mid \mathcal{P}, \mathcal{B}, \mathcal{T}_{\text{rxn}}) &= \prod_{t=1}^{T} p_\theta(a_t \mid s_t, \mathcal{P}, \mathcal{B}, \mathcal{T}_{\text{rxn}}), \\ M' &= \text{Assemble}(\tau; \mathcal{B}, \mathcal{T}_{\text{rxn}}). \end{aligned}$ | Protein<br>Synthesis |
| SynFlowNet [145] | GFlowNet | $\begin{aligned} p_\theta(\tau \mid \mathcal{B}, \mathcal{T}_{\text{rxn}}, \mathbf{y}) &= \prod_{t=1}^{T} p_\theta(a_t \mid s_t, \mathcal{B}, \mathcal{T}_{\text{rxn}}, \mathbf{y}), \\ M' &= \text{Execute}(\tau), \qquad \tau = (a_1, a_2, \ldots, a_T). \end{aligned}$ | Synthesis |

Notations: The global symbols follow Table 1. $\mathcal{G}$ denotes a molecular graph, while $M$ denotes a molecule represented in a non-graph form. $\mathcal{X}$ and $\mathcal{A}$ denote atom features and adjacency/bond information, respectively, and $\mathbf{R}$ denotes atomic coordinates. $\mathbf{z}$ denotes latent representations and $\mathcal{C}_{\text{prop}}$ denotes a property condition. $\mathbf{d}$ denotes pairwise atomic distances; $\mathcal{G}_{\text{frag}}$ and $\mathcal{M}_{\text{frag}}$ denote fragment-level graph and geometry representations. $f_\omega^{\text{atom}}$ and $f_\theta^{\text{bond}}$ denote atom- and bond-level coupling layers. $\mathbf{G}(\cdot)$ and $\mathbf{D}(\cdot)$ denote the generator and discriminator, respectively; $\tilde{M}$ denotes a variant SMILES representation and $\tilde{V}$ denotes the adversarial objective. $f_{1,t}(\cdot)$ and $f_{2,t}(\cdot)$ are drift functions, $g_{1,t}$ and $g_{2,t}$ are diffusion coefficients, and $\mathbf{w}_1$ and $\mathbf{w}_2$ are Wiener processes. $\mathbf{y}_o$ denotes an OOD condition and $\lambda_o$ controls OOD guidance. $\mathcal{P}$ denotes a protein binding-site condition, $\mathcal{K}$ denotes the fragment set, $\Sigma_k$ denotes the prior covariance matrix, and $\eta_{ik}$ specifies atom-to-prior correspondence. $\alpha_t$, $\beta_t$, $\overline{\alpha}_t$, and $\tilde{\beta}_t$ are diffusion noise parameters. $s$ denotes the conditional-guidance scale. $\mathcal{B}$ denotes purchasable building blocks, $\mathcal{T}_{\text{rxn}}$ denotes reaction templates, $\tau$ denotes a generation trajectory, and $a_t$ denotes the action at step $t$. $\mathbf{c}_{\text{text}}$ denotes a text condition; $\mathcal{P}_{\text{apo}}$ and $\mathcal{P}_{\text{holo}}$ denote ligand-free and ligand-bound pockets, respectively; and $\mathcal{L}$ denotes a ligand.

examples include 3DLinker [131], LatentGAN [133], CGCF [132], and GF-VAE [136]. These approaches demonstrated that deep generative models can capture the underlying distribution of molecular structures and generate chemically valid candidates beyond the training set. Recent studies increasingly adopt diffusion-based generative frameworks. Methods such as GDSS [134], GeoLDM [135], MOOD [139], GraphDiT [140], and LDMol [141] formulate molecular generation as a progressive denoising process. Compared with earlier generative paradigms, diffusion models have often shown stable optimization and flexible conditioning in reported molecular-generation studies; however, their iterative denoising procedure can substantially increase inference cost. Furthermore, geometric diffusion models explicitly operate on three-dimensional molecular structures, enabling the generation of spatially realistic conformations. From a cross-paradigm perspective, different generative models exhibit distinct trade-offs. VAEs and normalizing flows

**Table 9.** Summary of molecular optimization methods.

| Models | Category | Backbones | Conditions |
|---|---|---|---|
| DecompOpt [146] | Diffusion | $\begin{cases} p_\theta(M_{0:T-1} \mid M_T, \mathcal{P}, \{A_k\}) = \prod_{t=1}^{T} p_\theta(M_{t-1} \mid M_t, \mathcal{P}, \{A_k\}), \\ M_0 \sim q(M' \mid \mathcal{P}, \{A_k\}). \end{cases}$ | Protein |
| PMDM [147] | Diffusion | $\begin{cases} p_\theta(\mathcal{G}_{t-1} \mid \mathcal{G}_t) = \mathcal{N}\left(\mathcal{G}_{t-1}; \mu_\theta(\mathcal{G}_t, t), \sigma_t^2 I\right), \\ p_\theta(\mathcal{G}_{0:T-1} \mid \mathcal{G}_T), \\ \mu_\theta(\mathcal{G}_t, t) = \frac{1}{\sqrt{1-\beta_t}}\left(\mathcal{G}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathcal{G}_t, t)\right). \end{cases}$ | Protein |
| Lee et al. [148] | Reward Optimization | $\mathcal{R}(M, \mathcal{P}) = \sum_{k=1}^{K} w_k \widehat{r}_k(M, \mathcal{P}),$ $\max_\theta \ \mathbb{E}_{M \sim p_\theta(\cdot \mid \mathcal{P})}[\mathcal{R}(M, \mathcal{P})] - \lambda D_{\mathrm{KL}}\left(p_\theta \| p_{\mathrm{ref}}\right).$ | Protein |
| VOS [149] | Schedule Optimization | $\sigma^* = \arg\min_\sigma \int_0^T \mathcal{L}_{\mathrm{VLB}}\big(\mathcal{X}_t, \mathcal{A}_t, \mathcal{P}; \sigma(t)\big) dt,$ $p_\theta(M' \mid \mathcal{P}) = p_\theta(\mathcal{X}', \mathcal{A}' \mid \mathcal{P}, \sigma^*).$ | Protein |
| MolJO [150] | Gradient | $\tilde{\theta}_t = \theta_t + \eta \nabla_{\theta_t} \mathcal{S}(M_t, \mathcal{P}),$ $p_\phi(M_{t-1} \mid M_t, \mathcal{P}) = \mathrm{BFN}_\phi(M_{t-1} \mid \tilde{\theta}_t, \mathcal{P}),$ $M^* = \mathrm{BackwardCorrection}(M_{t:t-h}, \mathcal{P}).$ | Protein |
| CombiMOTS [151] | Pareto MCTS | $a^* = \arg\max_{a \in \Omega(s)} \left[Q(s,a) + c\sqrt{\frac{\log N(s)}{N(s,a)}}\right],$ $\mathbf{r}(M) = [r_{\mathrm{target1}}(M), r_{\mathrm{target2}}(M), r_{\mathrm{prop}}(M), r_{\mathrm{syn}}(M)],$ $M^* \in \mathrm{ParetoFront}\left(\mathbf{r}(M)\right).$ | Proteins |
| FMOP [152] | Diffusion | $\begin{cases} \mathcal{A}_{t-1} = W M^{\mathcal{A}} \odot \mathcal{A}_{t-1}^{\mathrm{ukn}} + (1 - M^{\mathcal{A}}) \odot \mathcal{A}_{t-1}^{\mathrm{kn}}, \\ \mathcal{X}_{t-1} = W M^{\mathcal{X}} \odot \mathcal{X}_{t-1}^{\mathrm{ukn}} + (1 - M^{\mathcal{X}}) \odot \mathcal{X}_{t-1}^{\mathrm{kn}}, \\ \tilde{\epsilon}_\theta(\mathcal{G}_{t-1}, \mathcal{C}) = \tilde{w}\epsilon_\theta(\mathcal{G}_t + \epsilon, \mathcal{C}) + (1 - \tilde{w})\epsilon_\theta(\mathcal{G}_t + \epsilon, \emptyset). \end{cases}$ | Cell |
| JTVAE [153] | VAE | $p(\mathcal{G}' \mid \mathbf{z}_\mathcal{T}, \mathbf{z}_\mathcal{G})$ | Property |
| Prompt-MolOpt [154] | AutoReg | $p(M_{\mathrm{tag}_{t+1}} \mid M_{\mathrm{tag}_{1:t}}, \mathcal{C}_{\mathrm{prop}}) = \mathrm{Trans}\left(\mathrm{Emb} + \sum \mathbf{z}_\mathcal{C}\right)$ | Property |
| Mol-CycleGAN [155] | GAN | $\begin{cases} M' = \mathbf{G}_1(\mathbf{G}_2(M)), \\ \mathbf{G}_1^*, \mathbf{G}_2^* = \arg\min_{\mathbf{G}_1, \mathbf{G}_2} \max_{\mathbf{D}_1, \mathbf{D}_2} \tilde{V}(\mathbf{G}_1, \mathbf{G}_2, \mathbf{D}_1, \mathbf{D}_2). \end{cases}$ | Property |
| MolEvol [156] | Search | $p_\theta(M) = \int_{s \in \mathcal{S}} p(s) p_\theta(M \mid s)\, ds$ | Property |
| MolSearch [157] | Search | $\begin{cases} \mathbf{U}_k = \frac{\mathcal{F}_k}{n_k} + \lambda\sqrt{\frac{4\ln n + \ln d}{2n_k}}, \quad k = 1, \ldots, K, \\ V_p = \mathrm{P}(\mathbf{U}_1, \ldots, \mathbf{U}_K). \end{cases}$ | Property |
| HN-GFN [158] | Search | $\mathcal{S}_{i+1} = \mathcal{S}_i \cup \left\{(M_j^i, f(M_j^i))\right\}_{j=1}^{b}$ | Property |
| DyMol [159] | Search | $\begin{cases} \mathbf{F}(\mathbf{z}_x, t) = \{f_1(\mathbf{z}_x), f_2(\mathbf{z}_x), \ldots, f_t(\mathbf{z}_x)\}, \\ \mathbf{R}(x, t) = \sum_{i=1}^{t} w_i(t) f_i(\mathbf{z}_x). \end{cases}$ | Property |
| MolEditRL [160] | Editing | $\begin{cases} a_t \in \Omega(M_t), \quad M_{t+1} = a_t(M_t), \\ a_t \sim \pi_\theta(\cdot \mid M_t, \mathcal{C}_{\mathrm{prop}}), \\ M^* = \arg\max_{M'} \left[\mathcal{R}(M', \mathcal{C}_{\mathrm{prop}}) - \eta \mathrm{dist}(M', M_0)\right]. \end{cases}$ | Property |
| SMER-Opt [161] | Editing | $\begin{cases} a_t^* = \arg\max_{a \in \Omega(M_t)} \left[Q_\theta(M_t, a; \Delta y) + c\sqrt{\frac{\log N(M_t)}{N(M_t, a) + 1}}\right], \\ M_{t+1} = a_t^*(M_t), \\ M^* = \arg\max_{M_T} \sum_{t=0}^{T-1} Q_\theta(M_t, a_t; \Delta y). \end{cases}$ | Property |

Notations: The global symbols follow Table 1. $\mathcal{G}$ denotes a molecular graph, while $M$ denotes a molecular object; $\mathcal{X}$ and $\mathcal{A}$ denote atom features and adjacency/bond information, respectively. $\mathcal{P}$ denotes a protein or binding-pocket condition, $\mathcal{C}_{\mathrm{prop}}$ denotes a property condition, and $\mathbf{z}$ denotes a latent representation. $A_k$ denotes a reference arm. $\widehat{r}_k(M, \mathcal{P})$ and $w_k$ denote the normalized reward and its weight for the $k$-th objective, respectively, and $p_{\mathrm{ref}}$ denotes the reference generative model. $\sigma(t)$ denotes the noise schedule and $\mathcal{L}_{\mathrm{VLB}}$ denotes the variational lower-bound objective. $\tilde{\theta}_t$ denotes the guided Bayesian-flow parameter, $\mathcal{S}(M_t, \mathcal{P})$ denotes the guidance score, and $h$ denotes the backward-correction horizon. $\Omega(s)$ denotes the feasible action set, $Q(s,a)$ denotes the action-value estimate, $N(s)$ and $N(s,a)$ denote visitation counts, and $c$ denotes the exploration coefficient. $\mathbf{r}(M)$ denotes a multi-objective reward vector. $\odot$ denotes element-wise multiplication; "kn" and "ukn" denote known and unknown components, respectively; $\epsilon \sim \mathcal{N}(0, I)$ denotes Gaussian noise; and $\tilde{w}$ denotes the conditional control strength. $\mathbf{z}_\mathcal{T}$ and $\mathbf{z}_\mathcal{G}$ denote latent representations of the junction tree and molecular graph. $M_{\mathrm{tag}_i}$ denotes the $i$-th molecular token/tag, $\mathrm{Trans}(\cdot)$ denotes the Transformer, and $\mathrm{Emb}$ denotes its input embedding. In MolEvol, $\mathcal{S}$ denotes the molecular search space; in HN-GFN, $\mathcal{S}_i$ denotes the candidate set at iteration $i$. $\mathcal{F}_k$, $n_k$, $n$, $d$, and $\lambda$ denote the node reward, node visits, total iterations, reward dimension, and exploration coefficient in MolSearch, respectively. $M_0$ and $M_t$ denote the initial molecule and the molecule at editing step $t$; $a_t$ denotes an editing action; $\pi_\theta$ denotes the editing policy; $\mathrm{dist}(\cdot, \cdot)$ denotes structural distance; and $\eta$ denotes the structure-preservation trade-off coefficient. $Q_\theta(M_t, a; \Delta y)$ denotes the predicted score of edit action $a$ under the desired property change $\Delta y$.

generally support efficient generation, while GANs enable fast sampling but may suffer from unstable training or mode collapse. Diffusion models provide flexible conditional generation and effective 2D/3D modeling, whereas

GFlowNets emphasize diverse and reward-guided exploration through sequential generation. These differences show that model selection should jointly consider generation quality, training stability, sampling efficiency, and task-specific requirements. Empirical benchmark studies provide further evidence for these trade-offs. On GEOM-QM9 and GEOM-Drugs, comparative evaluations of conformer-generation methods, including ConfVAE, CGCF, GeoDiff, and GeoMol, reveal substantial differences in conformational coverage and RMSD-based matching quality [162]. A recent systematic benchmark of nine diffusion-based 3D molecular generators on QM9 and GEOM-Drugs further shows that performance is generally more challenging under 3D geometric metrics than under conventional 2D molecular-quality metrics, with larger and structurally more complex GEOM-Drugs molecules posing additional difficulties [163]. These empirical observations reinforce that generative performance should be assessed jointly with the benchmark dataset, molecular complexity, and evaluation metric rather than through a single aggregate criterion.

A major trend in recent years has been the development of increasingly expressive conditional generation frameworks. As summarized in Table 8, contemporary models support multiple conditioning modalities, including molecular properties, protein structures, natural-language descriptions, and synthetic constraints. For example, MolHF [137], TransORGAN [138], and GraphDiT [140] perform property-guided molecular generation. DecompDiff [142] and Apo2Mol [143] generate molecules conditioned on protein binding environments, while LDMol [141] enables text-guided molecular generation through cross-modal alignment between language and molecular representations. Beyond conventional metrics such as chemical validity, novelty, and diversity, molecular generation should also be evaluated from the perspectives of geometric realism, physicochemical fidelity, synthesizability, OOD generalization, and computational efficiency. GEOM-Drugs [164] provides large-scale conformational ensembles of drug-like molecules for evaluating three-dimensional molecular structures and conformational quality, while QMugs [165] provides optimized molecular geometries together with quantum-mechanical properties for a large collection of drug-like molecules. These resources provide complementary geometry- and quantum-aware perspectives for evaluating molecular generative models. Meanwhile, multi-condition generation remains challenging when molecular properties, protein contexts, textual descriptions, and synthetic constraints provide heterogeneous or potentially conflicting signals, requiring models to balance conditional controllability with generation quality and computational efficiency.

Another important direction focuses on synthesizability-aware molecular generation. Traditional generative models may produce chemically valid molecules that are difficult or impossible to synthesize experimentally. To address this issue, methods such as RxnFlow [144] and SynFlowNet [145] explicitly incorporate reaction templates and purchasable building blocks into the generation process. By modeling molecular design as a sequence of chemically feasible reaction steps, these approaches better align generated molecules with practical laboratory constraints. Overall, molecular generation has evolved from unconditional graph generation toward multi-modal conditional design frameworks capable of incorporating structural, biological, textual, and synthetic information simultaneously. This trend reflects the growing demand for controllable and experimentally feasible molecular design systems.

#### 4.3.2. Molecular Optimization

While molecular generation focuses on designing compounds from scratch, MO aims to improve existing molecules with respect to desired objectives while preserving favorable structural characteristics. Typical optimization targets include binding affinity, selectivity, pharmacokinetic properties, toxicity profiles, and synthesizability.

As summarized in Table 9, existing optimization approaches can be broadly categorized into generative optimization methods, search-based strategies, and molecular editing frameworks. Early methods such as JT-VAE [153], Mol-CycleGAN [155], MolEvol [156], and MolSearch [157] explored latent-space optimization and evolutionary search strategies for molecular refinement. These approaches demonstrated the feasibility of navigating chemical space through learned molecular representations. More recent studies increasingly employ diffusion-based optimization frameworks. Representative examples include PMDM [147], DecompOpt [146], FMOP [152], VOS [149], and MolJO [150]. These methods formulate optimization as a conditional generation problem and guide molecular modifications using target proteins, cellular phenotypes, or desired molecular properties. Compared with traditional latent-space optimization approaches, diffusion-based methods generally provide finer control over generated structures and greater flexibility for incorporating multiple optimization objectives.

Another notable trend is the emergence of editing-based optimization paradigms. Rather than generating entirely new molecules, these approaches perform localized structural modifications on existing compounds. MolEditRL [160] formulates optimization as a reinforcement-learning problem over molecular editing actions, while SMER-Opt [161] performs sequential property-guided molecular editing through adaptive search strategies. Such methods are particularly attractive in medicinal chemistry because they preserve key molecular scaffolds while introducing targeted modifications that improve downstream properties. Search-based optimization remains an

important complementary direction. Methods including HN-GFN [158], DyMol [159], and CombiMOTS [151] employ Monte Carlo tree search, Pareto optimization, reinforcement learning, and generative flow networks to explore large chemical spaces efficiently. These approaches are especially useful in multi-objective optimization settings where competing design objectives must be balanced simultaneously.

From a practical perspective, molecular optimization often operates under stricter constraints than molecular generation. In many real-world drug development scenarios, preserving the biological activity, scaffold integrity, and synthetic accessibility of a lead compound is as important as improving its target properties. From a computational perspective, different generative and optimization paradigms exhibit distinct inference profiles. Latent-variable models such as VAEs avoid iterative denoising after latent sampling, although their decoders may still involve structured or autoregressive generation. Normalizing-flow models generate samples through inversion of a finite sequence of learned transformations. In contrast, diffusion models require repeated denoiser evaluations over a predefined sampling schedule, and representative molecular diffusion models may use tens to hundreds of sampling steps or substantially longer schedules depending on the formulation [135,166]. Search-, reinforcement-learning-, GFlowNet-, and sequential-editing methods instead require repeated proposal, scoring, rollout, or molecular-editing operations, so their inference cost depends strongly on the number of explored candidates or optimization steps. Peak memory should be considered separately from inference latency. For neural generators, memory consumption is mainly determined by model size and the intermediate molecular representations processed at each step. Three-dimensional equivariant models additionally maintain atom-, edge-, and coordinate-level hidden states, while search- and reinforcement-learning-based optimization may require candidate populations, search states, trajectories, or replay buffers. Therefore, both sampling latency and memory overhead should be considered alongside generation quality when evaluating molecular design methods for large-scale or high-throughput screening.

Evaluation metrics also require careful interpretation according to what they actually measure. QED evaluates drug-likeness through desirability functions derived from molecular properties [167], whereas the widely used SA score provides a heuristic estimate of synthetic accessibility based on fragment contributions and molecular complexity [168]. Consequently, neither a high QED value nor a favorable SA score establishes that a generated molecule has a practical experimental synthesis route. Practical synthesizability additionally depends on the availability of starting materials, feasible reaction sequences, reaction conditions, stereochemical constraints, and route complexity. Retrosynthesis-based evaluation provides a stronger assessment by explicitly examining whether viable synthetic routes to accessible precursors can be identified [169]. Nevertheless, such computational assessments remain proxies for practical feasibility, and experimental synthesis provides the most direct validation when available.

## 5. Datasets and Evaluation Benchmarks

This section corresponds to Layer 1 in the four-layer framework shown in Figure 1. Benchmark datasets play a central role in the development and evaluation of graph-structured drug discovery models. While advances in molecular representation learning and algorithm design have significantly improved predictive and generative capabilities, the reliability of these methods ultimately depends on the availability of diverse, high-quality, and biologically meaningful datasets. Thus, benchmark construction has become an essential component of modern molecular machine learning research. Table 10 summarizes representative datasets commonly used across major small-molecule drug discovery tasks, including molecular property prediction, drug-target interaction and affinity prediction, drug-drug interaction prediction, drug response prediction, and drug perturbation prediction. These resources span multiple biological scales, ranging from atom-level quantum-mechanical simulations to cellular response measurements and pharmacological interaction networks.

Several benchmark collections have emerged as community standards. Therapeutics Data Commons (TDC) provides a unified platform covering diverse therapeutic learning tasks, including drug-target interactions, drug responses, and drug combinations. TUDataset and MoleculeNet offer widely adopted benchmarks for molecular graph representation learning and molecular property prediction. Similarly, the Open Graph Benchmark (OGB) has established large-scale standardized datasets for evaluating graph learning algorithms under consistent experimental protocols. More recently, pharmacogenomic resources such as GDSCv1/v2 have become important benchmarks for studying drug sensitivity and precision medicine applications. From a task perspective, molecular property prediction datasets primarily focus on physicochemical, quantum-mechanical, and ADMET-related properties. Representative examples include QM7, QM7b, QM8, and QM9. These datasets provide the foundation for evaluating molecular representation quality and property prediction performance. For target-based drug discovery, datasets such as LIT-PCBA, DUD-E, and BioSNAP provide experimentally validated drug-target interactions and binding affinity

**Table 10.** Common datasets for small molecule drug discovery.

| Datasets | Data Structure | Tasks | Number of Samples | OOD |
|---|---|---|---|---|
| Benchmark 1: TDC Datasets [1] | | | | |
| BindingDB | Drug-Target | Reg. | 2,701,247 | S |
| DAVIS | Drug-Target | Reg. | 30,056 | S |
| KIBA | Drug-Target | Reg. | 118,254 | S |
| DrugBank | Drug-Drug | Cla. | 191,808 | C |
| TWOSIDES | Drug-Drug | Cla. | 4,651,131 | C |
| GDSCv1 | Drug-Cell | Reg. | 177,310 | S |
| GDSCv2 | Drug-Cell | Reg. | 92,703 | S |
| Benchmark 2: TUDataset [2] | | | | |
| QM9 | Drug (Qm., 3D) | Reg. | 133,885 | C |
| ZINC | Drug (Qm.) | Reg. | 249,456 | C |
| Benchmark 3: Open Graph Benchmark [3] | | | | |
| HIV | Drug (Bio.) | Cla. | 41,913 | C |
| Tox21 | Drug (Physio.) | Cla. | 8014 | C |
| ToxCast | Drug (Physio.) | Cla. | 8615 | C |
| BBBP | Drug (Physio.) | Cla. | 2053 | C |
| PCQM4Mv2 [5] | Drug (Qm.) | Reg. | 3,746,619 | C† |
| Benchmark 4: MoleculeNet [4] | | | | |
| PCBA | Drug (Bio.) | Cla. | 439,863 | C |
| MUV | Drug (Bio.) | Cla. | 93,127 | C |
| BACE | Drug (Bio.) | Cla. | 1522 | C |
| PDBbind | Drug (Bio.) | Reg. | 11,908 | C |
| SIDER | Drug (Physio.) | Cla. | 1427 | C |
| ClinTox | Drug (Physio.) | Cla. | 1491 | C |
| QM7 | Drug (Qm., 3D) | Reg. | 7165 | C |
| QM7b | Drug (Qm., 3D) | Reg. | 7210 | C |
| QM8 | Drug (Qm., 3D) | Reg. | 21,786 | C |
| ESOL | Drug (Phys. Chem.) | Reg. | 1128 | C |
| FreeSolv | Drug (Phys. Chem.) | Reg. | 643 | C |
| Lipophilicity | Drug (Phys. Chem.) | Reg. | 4200 | C |
| Molecule Property Prediction | | | | |
| MD17/22 [6] | Drug (Qm.) | Reg. | 99k–627k/5k–85k | L |
| OC20/22 [7] | Drug (Qm.) | Reg. | 1,281,040/62,331 | S |
| Molecule3D [8] | Drug (Qm.) | Reg. | 3,899,647 | C |
| 3BPA [9] | Drug (Qm.) | Reg. | 500 | L |
| ANI-1 [10] | Drug (Qm.) | Reg. | 24,416,306 | C |
| QM9-ED [11] | Drug (Qm.) | Reg. | 133,885 | C |
| QMugs-ED [12] | Drug (Qm.) | Reg. | 50,000 | C |
| Drug-Target Interaction and Affinity | | | | |
| LIT-PCBA [13] | Drug-Target | Cla. | 415,225/15 | C |
| C.elegans [14] | Drug-Target | Cla. | 1434/2504 | C |
| Human [15] | Drug-Target | Cla. | 1052/852 | C |
| DUD-E [16] | Drug-Target | Cla. | 22,886/102 | C |
| Metz [17] | Drug-Target | Reg. | 1423/170 | C |
| Karimi [18] | Drug-Target | Reg. | 3672/1287 | C |
| BioSNAP [19] | Drug-Target | Reg. | 4510/2481 Drugs/Targets | C |
| Drug-Drug Interaction | | | | |
| DRUGCOMBO [20] | Drug-Drug | Cla. | 3242/49,392 | C |
| KEGG [21] | Drug-Drug | Cla. | 1295/56,983 | C |
| SIDER [22] | Drug-Drug | Cla. | 1430/139,756 | C |
| ChCh-Miner [23] | Drug-Drug | Cla. | 1322/48,514 | C |
| Ogbl-biokg [24] | Drug-Drug | Cla. | 808/111,520 | C |
| ZhangDDI [25] | Drug-Drug | Reg. | 548/48,548 Drugs/Total | C |

**Table 10.** *Cont.*

| Datasets | Data Structure | Tasks | Number of Samples | OOD |
|---|---|---|---|---|
| Drug Response Prediction | | | | |
| CCLE [26] | Drug-Cell | Reg. | 24/479 | C |
| NCI-60 [27] | Drug-Cell | Reg. | 50,000/60 | C |
| CTRPv2 [28] | Drug-Cell | Reg. | 494/720 | C |
| gCSI [28] | Drug-Cell | Reg. | 16/312 | C |
| | | | Drugs/Cells | |
| Drug Perturbation Prediction | | | | |
| LINCS L1000 [29] | Drug-Cell | Reg. | 175,549/82 | C |
| Sci-Plex 3 [30] | Drug-Cell | Reg. | 188/3 | C |
| PBMCs [31] | Drug-Cell | Reg. | 144/6 | C |
| CMap [32] | Drug-Cell | Reg. | 164/5 | C |
| | | | Drugs/Cells | |

[1] Therapeutics data commons dataset; [2] The collection of benchmark datasets for learning with graphs; [3] The open graph benchmark; [4] The benchmark for molecular machine learning; [5] Molecular dynamics benchmark for evaluating force fields; [6] Two molecular dynamics datasets; [7] Two DFT-based molecular datasets for computational chemistry; [8] A 3D geometries benchmark; [9] A benchmark dataset for equivariant many-body interaction operations; [10] A dataset of 20 million conformations; [11] UniField-ED Benchmark; [12] UniField-ED Benchmark; [13] A dataset with high-confidence data; [14] A Caenorhabditis elegans database; [15] The human DTI dataset; [16] A database of useful (docking) decoys-Enhanced; [17] The G protein-coupled receptor focused dataset; [18] A DTA and side effect dataset; [19] The biomedical network dataset; [20] A drug combination synergy database; [21] The gene&genome database; [22] A drug-side effect association database; [23] A DDI network dataset; [24] The biological knowledge with gene-drug-disease triples; [25] A multi-source dataset; [26] The cancer cell line encyclopedia; [27] The NCI-60 human tumor cell lines screen; [28] The benchmark dataset for monotherapy drug response prediction; [29] The LINCS L1000 transcriptomic response data; [30] The Sci-Plex 3 screen of 188 compounds across cancer lines; [31] A novel single-cell perturbational dataset in human peripheral blood mononuclear cells; [32] A perturbational transcriptomic resource. Notations: Quantum mechanics (Qm.); Biophysics (Bio.); Physical chemistry (Phys. Chem.); Physiology (Physio.). For OOD evaluation, **S** denotes a standardized/native OOD-oriented split or a commonly standardized OOD protocol, **C** denotes that an OOD split can be constructed from available molecular, entity, or biological-context information, and **L** denotes limited or non-typical applicability to OOD evaluation. † PCQM4Mv2 uses a benchmark-defined PubChem Compound Identifier (CID)-based split that introduces a moderate distribution shift, rather than a strict scaffold-based OOD split. These labels indicate only the standardization, constructibility, or applicability of OOD-oriented splits and do not imply that structural overlap, target homology, repeated-observation leakage, or auxiliary-relation leakage has been explicitly audited or eliminated. The dataset statistics are accurate as of June, 2026.

measurements. In parallel, drug-drug interaction resources such as KEGG, SIDER, ChCh-Miner, and Ogbl-biokg support the study of adverse effects, synergistic combinations, and interaction mechanisms among therapeutic compounds.

At the systems biology level, pharmacogenomic datasets including CCLE, CTRPv2, gCSI, and NCI-60 associate chemical perturbations with cellular sensitivity measurements, enabling the development of drug response prediction models. Meanwhile, large-scale perturbation resources, such as LINCS L1000 and Sci-Plex 3, provide transcriptomic responses following drug treatment and have become standard benchmarks for drug perturbation prediction. In addition to dataset selection, evaluation protocols determine what type of generalization a benchmark actually measures. Random or independent and identically distributed (IID) splits mainly assess interpolation within the observed data distribution, whereas scaffold splits separate molecules according to their core chemical structures and provide a more stringent test of structural generalization. Temporal splits train models on earlier records and evaluate them on later observations, better approximating prospective prediction. Cross-domain evaluation further holds out biological contexts, such as cell lines, tissues, or assay conditions, to assess transferability across distinct experimental settings. The choice of OOD evaluation protocol should also reflect the downstream task and intended deployment scenario. For MPP, scaffold splits are particularly informative because they test generalization to unseen chemical scaffolds rather than interpolation within structurally similar molecular families. For DTI/DTA, temporal splits can better approximate prospective prediction when timestamped interaction or assay records are available, while unseen-drug and unseen-target settings directly assess cold-start generalization. For DRP, cross-domain evaluation across cell lines, tissues, cohorts, or screening datasets is particularly relevant for assessing transfer across biological contexts. For DDI, unseen-drug and unseen-pair settings evaluate whether interaction models can generalize to novel compounds or previously unobserved combinations. Activity-cliff evaluation provides an additional challenge, particularly for molecular property prediction, by testing whether models can distinguish structurally similar molecules with substantially different activities. These protocols are complementary rather than mutually exclusive and should be selected according to the generalization scenario of interest. Table 10 further indicates whether representative datasets can support OOD-oriented evaluation.

Beyond discriminative learning tasks, many of these datasets also provide important supervision signals and biological constraints for molecular generation and molecular optimization. Property annotations, interaction labels, cellular responses, and perturbation profiles are frequently incorporated as optimization objectives or conditional

signals in modern generative frameworks. Overall, the current benchmark ecosystem covers a broad spectrum of drug discovery scenarios, ranging from molecular property estimation and target interaction modeling to cellular response prediction and generative molecular design. The increasing integration of molecular structures, biological networks, transcriptomic measurements, and pharmacological annotations continues to provide a comprehensive foundation for developing and evaluating graph-structured drug discovery models.

## 6. Challenges and Opportunities

Despite substantial progress in graph-structured molecular learning, several fundamental challenges continue to limit the translation of computational advances into practical drug discovery workflows. Many recent models demonstrate impressive performance on established benchmarks, yet their deployment in real-world pharmaceutical settings remains constrained by issues related to generalization, data quality, interpretability, and biological realism. At the same time, emerging developments in geometric learning, foundation models, and generative models are creating new opportunities for addressing these limitations. This section discusses several important challenges and outlines promising directions for future research.

### *6.1. Out-of-Distribution Generalization*

A central challenge in molecular machine learning is the discrepancy between benchmark evaluation and real-world deployment. Most existing models are trained and evaluated under relatively homogeneous data distributions, where training and test samples often share similar molecular scaffolds, protein families, or experimental conditions. Under such settings, models may achieve strong predictive performance while relying heavily on statistical patterns that do not generalize beyond the benchmark environment [170–172]. In practical drug discovery, however, models are frequently required to predict the behavior of previously unseen molecules, novel protein targets, rare disease contexts, or newly generated compounds. Distribution shifts can arise from differences in assay protocols, batch effects, molecular diversity, species-specific biology, and temporal changes in experimental data collection. As a result, performance often deteriorates substantially when models are evaluated under realistic deployment scenarios.

Recent benchmark initiatives have introduced scaffold splits, temporal splits, cold-start settings, and unseen-entity evaluation protocols to better assess model robustness [170,171]. Nevertheless, reliable out-of-distribution generalization remains largely unsolved. However, OOD evaluation can still be biased by information leakage across training and test sets. Structural leakage may arise from overlapping or highly similar molecular structures, while target homology can introduce closely related proteins into different splits and inflate apparent generalization performance [173,174]. Similar issues may also occur through repeated or highly related experimental observations, highly correlated biological contexts, or auxiliary relational information. Therefore, reliable OOD evaluation should explicitly audit structural, biological, experimental, and relational overlap across split boundaries rather than relying solely on nominal split labels [174]. Several emerging research directions may provide promising solutions. First, causal representation learning offers a principled framework for disentangling biologically meaningful factors from spurious dataset-specific correlations. By explicitly modeling causal relationships among molecular structures, targets, and phenotypic outcomes, future models may achieve improved robustness across experimental environments [175,176]. Second, invariant learning and domain generalization techniques can be incorporated into graph neural networks to identify features that remain stable across different datasets, cell lines, or assay conditions. Such approaches may enable the transfer of learned knowledge to previously unseen biological domains. Third, recent advances in foundation models also provide new opportunities for improving out-of-distribution robustness. Large-scale molecular foundation models trained on billions of molecular graphs, protein sequences, and biomedical texts may learn more transferable biochemical priors than task-specific models. In particular, representation alignment across molecular, protein, cellular, and textual modalities could help bridge distribution gaps that arise when transferring knowledge between different biological systems.

### *6.2. Activity and Property Cliffs*

The molecular similarity principle assumes that structurally similar molecules tend to exhibit similar properties. While this assumption provides the foundation for many molecular learning methods, numerous exceptions exist in practical drug discovery. Small structural modifications can sometimes lead to disproportionately large changes in biological activity, toxicity, binding affinity, or pharmacokinetic behavior [177]. These phenomena are commonly referred to as activity cliffs or property cliffs.

Property cliffs are particularly challenging because many graph learning models implicitly assume smooth mappings between molecular structures and target properties. Consequently, models often perform well on average while failing to capture subtle local structural variations that determine critical biological outcomes [178].

Such failures can significantly affect lead optimization, where minor molecular modifications often dictate the success or failure of a candidate compound. Recent studies have proposed cliff-aware benchmarks, contrastive learning objectives, and geometry-enhanced molecular representations to improve sensitivity to local structural variations [17,179,180]. However, accurately modeling discontinuous structure-property relationships remains an open problem.

One promising direction is the development of multi-scale molecular representations that jointly model local functional groups, global molecular topology, and three-dimensional conformational structures. Since many activity cliffs originate from subtle modifications around specific pharmacophores or binding motifs, hierarchical graph architectures capable of simultaneously capturing atom-level and substructure-level information may offer improved sensitivity to critical structural changes. Another important opportunity lies in incorporating richer geometric and electronic information into molecular learning frameworks. Recent studies suggest that many cliff phenomena are not solely determined by molecular topology, but also by changes in conformational preferences, electrostatic interactions, charge distributions, and non-covalent binding effects. Future models that integrate geometric deep learning, electron-density representations, and quantum-informed features may better capture the physical mechanisms underlying abrupt property transitions. Beyond representation learning, new training paradigms specifically designed for cliff scenarios are also emerging. For example, local contrastive learning, metric learning, and hard-pair mining strategies can explicitly emphasize structurally similar molecules with substantially different properties during optimization. Such approaches encourage models to learn more discriminative local decision boundaries rather than relying on global structural similarity alone.

### *6.3. Data Scarcity and Incomplete Supervision*

Although large-scale molecular datasets have become increasingly available, data scarcity remains a persistent challenge across many drug discovery tasks. Experimentally validated drug-target interactions, drug combinations, perturbation profiles, and clinical outcomes represent only a small fraction of the enormous biological search space [181]. In addition, available data are often biased toward well-studied proteins, diseases, and molecular scaffolds. A related challenge is the lack of reliable supervision signals. Most public databases primarily record successful experiments or biologically meaningful observations, whereas inactive compounds and failed experiments are frequently underreported [182]. Therefore, many learning frameworks implicitly treat unobserved interactions as negatives, even though these interactions may simply be untested. Such assumptions introduce label noise and complicate model calibration.

Addressing these limitations will require a shift toward more data-centric learning paradigms. One promising direction is active learning and closed-loop experimentation, where predictive models iteratively identify the most informative molecules, targets, or perturbations for experimental validation [183–185]. By coupling machine learning models with automated design-build-test-learn cycles, future platforms may substantially improve data efficiency while reducing experimental costs. Another important opportunity lies in improving supervision quality through synthetic data generation and hard negative construction. Molecular docking, molecular dynamics simulations, counterfactual molecular generation, and structure-based filtering techniques may help construct biologically plausible negative samples that are currently missing from public datasets. Such approaches could significantly improve model calibration and robustness in tasks such as DTI, DDI, and DPP. Knowledge-enhanced learning offers another promising avenue. Biomedical knowledge graphs that integrate drugs, genes, proteins, diseases, pathways, and clinical evidence can provide complementary supervision beyond experimentally validated observations. Combined with graph foundation models and large language models, future systems may be capable of automatically inferring missing biological relationships and prioritizing high-confidence hypotheses for experimental verification.

### *6.4. Interpretability and Mechanistic Understanding*

Despite impressive predictive capabilities, many graph neural networks and foundation models remain difficult to interpret [186]. In pharmaceutical research, predictive performance alone is rarely sufficient. Medicinal chemists and biologists often require mechanistic explanations that connect model predictions to molecular structures, protein interactions, signaling pathways, or biological processes [187]. The lack of interpretability creates a trust gap between computational predictions and experimental validation. A model may accurately predict binding affinity or toxicity while providing little insight into which molecular substructures, interaction motifs, or physicochemical mechanisms drive the prediction. Recent research has explored attention-based explanations, subgraph attribution methods, counterfactual molecular editing, and causal reasoning frameworks [126,188]. However, many existing explanation techniques remain qualitative and difficult to validate experimentally.

One promising direction is the development of inherently interpretable graph learning architectures. Instead of generating explanations after prediction, future models may explicitly incorporate chemically meaningful entities such as functional groups, pharmacophores, reaction centers, and molecular motifs into the learning process. Such designs could provide explanations that are more directly aligned with domain knowledge and medicinal chemistry practice. Another important opportunity lies in integrating graph learning with structural biology and molecular simulation. Combining graph neural networks with protein docking, molecular dynamics simulations, and quantum-chemical calculations may help bridge the gap between statistical prediction and physical mechanism. Rather than identifying important atoms alone, future systems may explain predictions through binding interactions, conformational changes, energy landscapes, and molecular recognition processes. Recent advances in causal learning also provide a promising framework for mechanistic understanding. By distinguishing causal biological factors from spurious correlations, causal graph models may reveal how molecular modifications propagate through protein interactions, signaling pathways, and cellular networks to produce downstream phenotypic effects. Such approaches have the potential to move explainability beyond descriptive attribution toward mechanistic reasoning [189,190]. These interpretability routes offer complementary trade-offs. Post-hoc attribution is efficient for identifying influential atoms or substructures but does not establish causality; inherently interpretable models provide more structured explanations but often require task-specific inductive biases; structure- and simulation-based approaches offer stronger physical grounding at higher computational cost; and causal methods can support mechanistic hypotheses but depend on assumptions and supervision that are difficult to validate in complex biological systems.

### *6.5. Foundation Models in Drug Discovery*

Foundation models are increasingly being adopted across molecular, protein, cellular, and biomedical-text representations, offering a common pre-training paradigm for transferring knowledge across multiple drug-discovery tasks [191–195]. However, their increasing scale and breadth also introduce challenges that extend beyond conventional task-specific learning. First, large-scale pre-training may inherit and amplify biases present in heterogeneous biomedical corpora. Molecular databases are often dominated by well-studied chemical scaffolds, protein families, disease areas, and assay types, while negative results and under-represented biological contexts remain comparatively sparse [196]. As a result, scaling the pre-training corpus does not necessarily guarantee uniform downstream reliability and may instead reinforce imbalances in chemical and biological coverage. Future work should therefore place greater emphasis on balanced data curation, bias auditing, and representation quality across under-represented regimes. Second, standardized evaluation of foundation models remains limited. Existing studies often differ in downstream tasks, data splits, adaptation strategies, parameter budgets, and model-access assumptions, making direct comparison difficult. Moreover, large pre-training corpora increase the risk of hidden overlap between pre-training and downstream evaluation data [197]. Reliable assessment should therefore incorporate leakage-aware OOD splits, few-shot and zero-shot transfer, calibration analysis, and reproducible fine-tuning or adaptation protocols to distinguish genuine transferability from benchmark-specific advantages.

A further challenge is the integration of explicit domain knowledge. Large-scale pre-training alone does not guarantee chemically or biologically valid reasoning, particularly when molecular topology, three-dimensional geometry, protein context, cellular measurements, and biomedical text provide heterogeneous forms of evidence [192]. Promising directions include knowledge-graph augmentation, physics- and geometry-informed objectives, retrieval from curated biomedical resources, and tool-augmented reasoning. Knowledge-enhanced foundation models such as TxGNN further demonstrate the potential of incorporating structured biomedical relations into transferable representations [198]. These mechanisms may enable foundation models to combine broad transferable priors with explicit chemical constraints, biological mechanisms, and experimentally grounded evidence. Overall, foundation models for drug discovery should be evaluated not only by scale and benchmark performance, but also by robustness to data bias and distribution shifts, reproducibility, and the ability to incorporate verifiable domain knowledge.

## Author Contributions

J.H.: data curation, writing—original draft preparation; W.Z.: data curation, writing—original draft preparation; Z.L.: data curation, writing—original draft preparation; J.Y.: conceptualization, methodology, software; J.C.: conceptualization, methodology, software; J.D.: conceptualization, methodology, software; W.H.: supervision; K.L.: writing—reviewing and editing; J.G.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded in part by the Natural Science Foundation of China (No. 62476203), Key Project of Traditional Chinese Medicine Joint Fund of Hubei Provincial Natural Science Foundation (No. 2025AFD47), Hubei Province Science and Technology Innovation Plan Project (No. 2025BCB035), the Guangdong Provincial Natural Science Foundation General Project (No. 2025A1515012155), the Shenzhen Natural Science Foundation Project (Nos. JCYJ20250604122534006, JCYJ20230807090211021), and the Hubei Jiangxia laboratory development project (No. 2025JXKF034).

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

Not applicable.

## Conflicts of Interest

The authors declare no conflict of interest.

## Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

## References

1. Li, K.; Xiong, Y.; Zhang, H.; et al. Graph-Structured Small Molecule Drug Discovery Through Deep Learning: Progress, Challenges, and Opportunities. In Proceedings of the 2025 IEEE International Conference on Web Services (ICWS), Helsinki, Finland, 7–12 July 2025; pp. 1033–1042. https://doi.org/10.1109/icws67624.2025.00135.
2. Hopkins, A.L.; Groom, C.R. The druggable genome. *Nat. Rev. Drug Discov.* **2002**, *1*, 727–730. https://doi.org/10.1038/nrd892.
3. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25. https://doi.org/10.1016/s0169-409x(96)00423-1.
4. Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432*, 823. https://doi.org/10.1038/432823a.
5. Reymond, J.L. The Chemical Space Project. *Acc. Chem. Res.* **2015**, *48*, 722–730. https://doi.org/10.1021/ar500432k.
6. Gilmer, J.; Schoenholz, S.S.; Riley, P.F.; et al. Neural Message Passing for Quantum Chemistry. In Proceedings of the International Conference on Machine Learning, Sydney, NSW, Australia, 6–11 August 2017; pp. 1263–1272.
7. Duvenaud, D.K.; Maclaurin, D.; Iparraguirre, J.; et al. Convolutional networks on graphs for learning molecular fingerprints. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 7–12 December 2015; Volume 28.
8. Zheng, S.; He, J.; Liu, C.; et al. Predicting equilibrium distributions for molecular systems with deep learning. *Nat. Mach. Intell.* **2024**, *6*, 558–567. https://doi.org/10.1038/s42256-024-00837-3. Correction in *Nat. Mach. Intell.* **2024**, *6*, 1626.
9. Sheshanarayana, R.; You, F. Molecular representation learning: cross-domain foundations and future Frontiers. *Digit. Discov.* **2025**, *4*, 2298–2335. https://doi.org/10.1039/d5dd00170f.
10. Li, K.; Hu, L.; Chen, J.; et al. Can molecular evolution mechanism enhance molecular representation? In Proceedings of the AAAI Conference on Artificial Intelligence, Singapore, 20–27 January 2026; Volume 40, pp. 15108–15116.
11. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36. https://doi.org/10.1021/ci00057a005.
12. Satorras, V.G.; Hoogeboom, E.; Welling, M. E(n) Equivariant Graph Neural Networks. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 9323–9332.
13. Schütt, K.T.; Unke, O.T.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; Volume 139, pp. 9377–9388.
14. Xiong, Y.; Li, K.; Chen, J.; et al. Text-guided multi-property molecular optimization with a diffusion language model. *Inf. Fusion* **2026**, *127*, 103907. https://doi.org/10.1016/j.inffus.2025.103907.
15. Xiong, Y.; Chen, J.; Gong, X.; et al. Variational Bayesian Flow Network for Graph Generation. *arXiv* **2026**, arXiv:2601.22524.

16. Li, K.; Hu, L.; Xiong, Y.; et al. PCEvo: Path-Consistent Molecular Representation via Virtual Evolutionary. *arXiv* **2026**, arXiv:2601.19257.
17. Zhang, W.; Li, K.; Chen, J.; et al. UniField: RBF-Guided Electron Density Fusion for Enhanced Molecular Representations. *arXiv* **2026**, arXiv:2605.24013.
18. Du, W.; Zhang, H.; Du, Y.; et al. SE(3) Equivariant Graph Neural Networks with Complete Local Frames. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 5583–5608. https://proceedings.mlr.press/v162/du22e.html.
19. Gasteiger, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
20. Na, G.S.; Park, C. Electron-Informed Coarse-Graining Molecular Representation Learning for Real-World Molecular Physics. In Proceedings of the KDD'25: The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Toronto, ON, Canada, 3–7 August 2025; Volume 1, pp. 1080–1091. https://doi.org/10.1145/3690624.3709270.
21. Na, G.S.; Park, C. Electron-Derived Molecular Representation Learning for Real-World Molecular Physics. In Proceedings of the NeurIPS 2023 AI for Science Workshop, New Orleans, LA, USA, 16 December 2023.
22. Hua, Y.; Xu, T.; Song, X.; et al. R-DTI: drug target interaction prediction based on second-order relevance exploration. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 17368–17376.
23. Yang, J.; Zhang, J.; Qian, K.; et al. Bridging the Modality Reliability Gap in Drug-Target Interaction Prediction via a Confidence-aware Multimodal Fusion Framework. In Proceedings of the AAAI Conference on Artificial Intelligence, Singapore, 20–27 January 2026; Volume 40, pp. 27529–27537.
24. Jiang, F.; Aminanmu, M.; Ma, H.; et al. GRAM-DTI: Adaptive Multimodal Representation Learning for Drug–Target Interaction Prediction. In Proceedings of the International Conference on Learning Representations, Rio de Janeiro, Brazil, 23–27 April 2026; Volume 2026.
25. Xie, Z.; Tu, S.; Xu, L. Multilevel attention network with semi-supervised domain adaptation for drug-target prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 329–337.
26. Peng, X.; Wei, Y.; Deng, A.; et al. Balanced Multimodal Learning via On-the-fly Gradient Modulation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 8228–8237. https://doi.org/10.1109/cvpr52688.2022.00806.
27. Hoang, T.L.; Sbodio, M.L.; Martinez Galindo, M.; et al. Knowledge Enhanced Representation Learning for Drug Discovery. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 10544–10552. https://doi.org/10.1609/aaai.v38i9.28924.
28. Zhai, X.; Wang, C.; Wang, R.; et al. Blend the Separated: Mixture of Synergistic Experts for Data-Scarcity Drug-Target Interaction Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 22336–22344.
29. Li, Y.; Qiao, G.; Wang, K.; et al. Drug–target interaction predication via multi-channel graph neural networks. *Brief. Bioinform.* **2022**, *23*, bbab346. https://doi.org/10.1093/bib/bbab346.
30. Zhou, Z.; Yin, Y.; Yang, Y.; et al. Enhancing bioactivity prediction via spatial emptiness representation of protein-ligand complex and union of multiple pockets. In Proceedings of the Advances in Neural Information Processing Systems, San Diego, CA, USA, 2–7 December 2025; Volume 38, pp. 106162–106181.
31. Gao, B.; Qiang, B.; Tan, H.; et al. Drugclip: Contrastive protein-molecule representation learning for virtual screening. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023; Volume 36; pp. 44595–44614.
32. Zhou, G.; Gao, Z.; Ding, Q.; et al. Uni-Mol: A Universal 3D Molecular Representation Learning Framework. In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
33. Sledzieski, S.; Singh, R.; Cowen, L.; et al. Adapting protein language models for rapid DTI prediction. *bioRxiv* **2022**. https://doi.org/10.1101/2022.11.03.515084.
34. Zhao, Q.; Zhao, H.; Zheng, K.; et al. HyperAttentionDTI: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics* **2022**, *38*, 655–662. https://doi.org/10.1093/bioinformatics/btab715.
35. Zheng, S.; Li, Y.; Chen, S.; et al. Predicting drug–protein interaction using quasi-visual question answering system. *Nat. Mach. Intell.* **2020**, *2*, 134–140. https://doi.org/10.1038/s42256-020-0152-y. Correction in *Nat. Mach. Intell.* **2020**, *2*, 551.
36. Huang, K.; Xiao, C.; Glass, L.M.; et al. MolTrans: Molecular Interaction Transformer for drug–target interaction prediction. *Bioinformatics* **2021**, *37*, 830–836. https://doi.org/10.1093/bioinformatics/btaa880.
37. Zhang, H.; Gong, X.; Pan, S.; et al. A Cross-Field Fusion Strategy for Drug-Target Interaction Prediction. *arXiv* **2024**, arXiv:2405.14545.
38. Lu, Z.; Song, G.; Zhu, H.; et al. DTIAM: a unified framework for predicting drug-target interactions, binding affinities and drug mechanisms. *Nat. Commun.* **2025**, *16*, 2548. https://doi.org/10.1038/s41467-025-57828-0.
39. Bai, P.; Miljković, F.; John, B.; et al. Interpretable bilinear attention network with domain adaptation improves drug–target prediction. *Nat. Mach. Intell.* **2023**, *5*, 126–136. https://doi.org/10.1038/s42256-022-00605-1.

40. Wu, L.; Huang, Y.; Tan, C.; et al. Psc-cpi: Multi-scale protein sequence-structure contrasting for efficient and generalizable compound-protein interaction prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 310–319.
41. You, Y.; Shen, Y. Cross-modality and self-supervised protein embedding for compound–protein affinity and contact prediction. *Bioinformatics* **2022**, *38*, ii68–ii74. https://doi.org/10.1093/bioinformatics/btac470.
42. Peng, L.; Liu, X.; Chen, M.; et al. MGNDTI: A Drug-Target Interaction Prediction Framework Based on Multimodal Representation Learning and the Gating Mechanism. *J. Chem. Inf. Model.* **2024**, *64*, 6684–6698. https://doi.org/10.1021/acs.jcim.4c00957.
43. Nguyen, N.Q.; Jang, G.; Kim, H.; et al. Perceiver CPI: a nested cross-attention network for compound–protein interaction prediction. *Bioinformatics* **2023**, *39*, btac731. https://doi.org/10.1093/bioinformatics/btac731.
44. Sicherman, A.; Radinsky, K. BioCG: Constrained Generative Modeling for Biochemical Interaction Prediction. In Proceedings of the Advances in Neural Information Processing Systems, San Diego, CA, USA, 2–7 December 2025; Volume 38, pp. 4266–4285.
45. Liu, X.; Song, C.; Huang, F.; et al. GraphCDR: a graph neural network method with contrastive learning for cancer drug response prediction. *Brief. Bioinform.* **2022**, *23*, bbab457. https://doi.org/10.1093/bib/bbab457.
46. Li, K.; Gong, X.; Wu, J.; et al. Contrastive Learning Drug Response Models from Natural Language Supervision. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), Jeju, Republic of Korea, 3–9 August 2024; pp. 2126–2134. https://doi.org/10.24963/ijcai.2024/235.
47. Li, K.; Liu, W.; Luo, Y.; et al. Zero-shot Learning for Preclinical Drug Screening. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), Jeju, Republic of Korea, 3–9 August 2024; pp. 2117–2125. https://doi.org/10.24963/ijcai.2024/234.
48. Liu, X.; Li, M. Knowledge-guided domain adaptation model for transferring drug response prediction from cell lines to patients. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 523–531.
49. Bai, Y.; Lv, H.; Shi, W.; et al. Multi-Level Domain Adaptation and Contrastive Domain Isolation with Bilinear Fusion for Patient Drug Response Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Singapore, 20–27 January 2026; Volume 40, pp. 19675–19683.
50. Zhang, Y.; Huang, Z.; Fan, Z.; et al. DeepSADR: Deep Transfer Learning with Subsequence Interaction and Adaptive Readout for Cancer Drug Response Prediction. In Proceedings of the International Conference on Learning Representations, Rio de Janeiro, Brazil, 23–27 April 2026; Volume 2026, pp. 116485–116513.
51. Shubham, K.; Jayagopal, A.; Danish, S.M.; et al. WISER: Weak Supervision and Supervised Representation Learning to Improve Drug Response Prediction in Cancer. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024.
52. Jayagopal, A.; Xue, H.; He, Z.; et al. Personalised Drug Identifier for Cancer Treatment with Transformers using Auxiliary Information. In Proceedings of the KDD'24: The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Barcelona, Spain, 25–29 August 2024; pp. 5138–5149. https://doi.org/10.1145/3637528.3671652.
53. Rampášek, L.; Hidru, D.; Smirnov, P.; et al. Dr.VAE: improving drug response prediction via modeling of drug perturbation effects. *Bioinformatics* **2019**, *35*, 3743–3751. https://doi.org/10.1093/bioinformatics/btz158.
54. Jiang, L.; Jiang, C.; Yu, X.; et al. DeepTTA: a transformer-based model for predicting cancer drug response. *Brief. Bioinform.* **2022**, *23*, bbac100. https://doi.org/10.1093/bib/bbac100.
55. Liu, X.; Zhang, W. A subcomponent-guided deep learning method for interpretable cancer drug response prediction. *PLoS Comput. Biol.* **2023**, *19*, e1011382. https://doi.org/10.1371/journal.pcbi.1011382.
56. Zhu, Y.; Ouyang, Z.; Chen, W.; et al. TGSA: protein–protein association-based twin graph neural networks for drug response prediction with similarity augmentation. *Bioinformatics* **2022**, *38*, 461–468. https://doi.org/10.1093/bioinformatics/btab650.
57. Li, P.; Jiang, Z.; Liu, T.; et al. Improving drug response prediction via integrating gene relationships with deep learning. *Brief. Bioinform.* **2024**, *25*, bbae153. https://doi.org/10.1093/bib/bbae153.
58. He, D.; Liu, Q.; Wu, Y.; et al. A context-aware deconfounding autoencoder for robust prediction of personalized clinical drug response from cell-line compound screening. *Nat. Mach. Intell.* **2022**, *4*, 879–892. https://doi.org/10.1038/s42256-022-00541-0.
59. Chen, J.; Wang, X.; Ma, A.; et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. *Nat. Commun.* **2022**, *13*, 6494. https://doi.org/10.1038/s41467-022-34277-7.
60. Sefer, E. DRGAT: Predicting Drug Responses Via Diffusion-Based Graph Attention Network. *J. Comput. Biol.* **2025**, *32*, 330–350. https://doi.org/10.1089/cmb.2024.0807.
61. Wang, C.; Kumar, G.A.; Rajapakse, J.C. Drug discovery and mechanism prediction with explainable graph neural networks. *Sci. Rep.* **2025**, *15*, 179. https://doi.org/10.1038/s41598-024-83090-3.
62. Wu, Y.; Chen, M.; Qin, Y. Anticancer drug response prediction integrating multi-omics pathway-based difference features and multiple deep learning techniques. *PLoS Comput. Biol.* **2025**, *21*, e1012905. https://doi.org/10.1371/journal.pcbi.1012905.
63. Li, Y.; Liu, C.; Cui, H.; et al. PAM-CDR: Property-Aware Multi-Modal Drug Representation Learning for Accurate Cancer Drug Response Prediction. *IEEE J. Biomed. Health Inform.* **2026**, 1–14. https://doi.org/10.1109/jbhi.2026.3658090.

64. Liu, Z.; Lyu, K.; Li, Y.; et al. MTEGDRP: Interpretable Molecular Self-Attention Transformer and Equivariant Graph Neural Network Based on Multi-Omics Fusion for Drug Response Prediction in Cancer Cell Lines. *J. Med. Chem.* **2026**, *69*, 4771–4788. https://doi.org/10.1021/acs.jmedchem.5c03438.
65. Lotfollahi, M.; Wolf, F.A.; Theis, F.J. scGen predicts single-cell perturbation responses. *Nat. Methods* **2019**, *16*, 715–721. https://doi.org/10.1038/s41592-019-0494-8.
66. Umarov, R.; Li, Y.; Arner, E. DeepCellState: An autoencoder-based framework for predicting cell type specific transcriptional states induced by drug treatment. *PLoS Comput. Biol.* **2021**, *17*, e1009465. https://doi.org/10.1371/journal.pcbi.1009465.
67. Lotfollahi, M.; Susmelj, A.K.; De Donno, C.; et al. Compositional perturbation autoencoder for single-cell response modeling. *bioRxiv* **2021**. https://doi.org/10.1101/2021.04.14.439903.
68. Hetzel, L.; Boehm, S.; Kilbertus, N.; et al. Predicting cellular responses to novel drug perturbations at a single-cell resolution. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 26711–26722.
69. Wang, Y.; Liu, X.; Fan, Y.; et al. Predicting drug responses of unseen cell types through transfer learning with foundation models. *Nat. Comput. Sci.* **2026**, *6*, 39–52. https://doi.org/10.1038/s43588-025-00887-6.
70. He, S.; Zhu, Y.; Tavakol, D.N.; et al. Squidiff: predicting cellular development and responses to perturbations using a diffusion model. *Nat. Methods* **2026**, *23*, 65–77. https://doi.org/10.1038/s41592-025-02877-y.
71. Hu, Q.; Chen, Z.; Gu, J. Predicting drug-perturbed transcriptional responses using multi-conditional diffusion transformer. *Quant. Biol.* **2026**, *14*, e70016. https://doi.org/10.1002/qub2.70016.
72. Bunne, C.; Stark, S.G.; Gut, G.; et al. Learning single-cell perturbation responses using neural optimal transport. *Nat. Methods* **2023**, *20*, 1759–1768. https://doi.org/10.1038/s41592-023-01969-x. Correction in *Nat. Methods* **2023**, *20*, 1830.
73. Lotfollahi, M.; Klimovskaia Susmelj, A.; De Donno, C.; et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Mol. Syst. Biol.* **2023**, *19*, MSB202211517. https://doi.org/10.15252/msb.202211517.
74. Huang, W.; Liu, H. Predicting single-cell cellular responses to perturbations using cycle consistency learning. *Bioinformatics* **2024**, *40*, i462–i470. https://doi.org/10.1093/bioinformatics/btae248.
75. Qi, X.; Zhao, L.; Tian, C.; et al. Predicting transcriptional responses to novel chemical perturbations using deep generative model for drug discovery. *Nat. Commun.* **2024**, *15*, 9256. https://doi.org/10.1038/s41467-024-53457-1.
76. Meng, F.; Wang, C.; Lin, Y.; et al. DeepICER: A deep learning framework for predicting compound-induced gene expression profiles. *Acta Pharm. Sin. B* **2026**, *16*, 2947–2963. https://doi.org/10.1016/j.apsb.2026.01.046.
77. Guo, Y.; Zhang, H.; Hu, H.; et al. Modelling drug-induced cellular perturbation responses with a biologically informed dual-branch transformer. *Nat. Mach. Intell.* **2026**, *8*, 96–112. https://doi.org/10.1038/s42256-025-01165-w.
78. Liu, Y.; Wang, L.; Liu, M.; et al. Spherical message passing for 3d molecular graphs. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022.
79. Liao, Y.L.; Smidt, T. Equiformer: Equivariant Graph Attention Transformer for 3D Atomistic Graphs. In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
80. Wang, Y.; Wang, T.; Li, S.; et al. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nat. Commun.* **2024**, *15*, 313. https://doi.org/10.1038/s41467-023-43720-2.
81. Aykent, S.; Xia, T. GotenNet: Rethinking Efficient 3D Equivariant Graph Neural Networks. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025; Volume 2025, pp. 40622–40646.
82. Luo, S.; Chen, T.; Xu, Y.; et al. One Transformer Can Understand Both 2D & 3D Molecular Data. In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
83. Yang, J.; Zheng, K.; Long, S.; et al. MOL-AE: Auto-encoder based molecular representation learning with 3D cloze test objective. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024.
84. Ji, X.; Wang, Z.; Gao, Z.; et al. Exploring Molecular Pretraining Model at Scale. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 10–15 December 2024; Volume 37, pp. 46956–46978. https://doi.org/10.52202/079017-1489.
85. Shoghi, N.; Kolluru, A.; Kitchin, J.R.; et al. From Molecules to Materials: Pre-training Large Generalizable Models for Atomic Property Prediction. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
86. Yu, Q.; Zhang, Y.; Ni, Y.; et al. Multimodal Molecular Pretraining via Modality Blending. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
87. Ni, Y.; Feng, S.; Ma, W.Y.; et al. Sliced Denoising: A Physics-Informed Molecular Pre-Training Method. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024; Volume 2024, pp. 17064–17088.
88. Hu, W.; Liu, B.; Gomes, J.; et al. Strategies for Pre-training Graph Neural Networks. *arXiv* **2019**, arXiv:1905.12265.
89. Sun, R.; Dai, H.; Yu, A.W. Does GNN Pretraining Help Molecular Representation? In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 12096–12109. https://doi.org/10.52202/068431-0879.
90. Liu, S.; Wang, H.; Liu, W.; et al. Pre-training Molecular Graph Representation with 3D Geometry. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022.

91. Stärk, H.; Beaini, D.; Corso, G.; et al. 3D Infomax improves GNNs for Molecular Property Prediction. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Volume 162, pp. 20479–20502.
92. Xiang, H.; Xia, J.; Jin, X.; et al. Electron Density-enhanced Molecular Geometry Learning. In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25), Montreal, QC, Canada, 16–22 August 2025; pp. 7840–7848. https://doi.org/10.24963/ijcai.2025/872.
93. Yang, N.; Zeng, K.; Wu, Q.; et al. Learning Substructure Invariance for Out-of-Distribution Molecular Representations. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 12964–12978. https://doi.org/10.52202/068431-0942.
94. Wang, Y.; Zhou, K.; Liu, N.; et al. Efficient Sharpness-Aware Minimization for Molecular Graph Transformer Models. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024; Volume 2024, pp. 44737–44759.
95. An, J.; Qu, C.; Zhou, Z.; et al. Hybrid Directional Graph Neural Network for Molecules. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
96. Thölke, P.; Fabritiis, G.D. Equivariant Transformers for Neural Network based Molecular Potentials. In Proceedings of the International Conference on Learning Representations, Virtual, 25–29 April 2022.
97. Musaelian, A.; Batzner, S.; Johansson, A.; et al. Learning local equivariant representations for large-scale atomistic dynamics. *Nat. Commun.* **2023**, *14*, 579. https://doi.org/10.1038/s41467-023-36329-y.
98. Sun, M.; Guo, M.; Yuan, W.; et al. Representing Molecules as Random Walks Over Interpretable Grammars. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024; pp. 46988–47016.
99. Guo, M.; Thost, V.; Song, S.W.; et al. Hierarchical Grammar-Induced Geometry for Data-Efficient Molecular Property Prediction. In Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 12055–12076.
100. Vargas, S.; Gee, W.; Alexandrova, A. High-throughput quantum theory of atoms in molecules (QTAIM) for geometric deep learning of molecular and reaction properties. *Digit. Discov.* **2024**, *3*, 987–998. https://doi.org/10.1039/d4dd00057a.
101. Wang, S.H.; Huang, Y.; Baker, J.; et al. A Theoretically-Principled Sparse, Connected, and Rigid Graph Representation of Molecules. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025; Volume 2025, pp. 36269–36294.
102. Wu, F.; Radev, D.; Li, S.Z. Molformer: Motif-based transformer on 3d heterogeneous molecular graphs. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 5312–5320.
103. Li, Y.; Huang, L.; Ding, Z.; et al. E2Former: An Efficient and Equivariant Transformer with Linear-Scaling Tensor Products. In Proceedings of the Thirty-Ninth Annual Conference on Neural Information Processing Systems, San Diego, CA, USA, 2–7 December 2025; Volume 38, pp. 26777–26814. https://doi.org/10.52202/085713-0797.
104. Sun, F.Y.; Hoffman, J.; Verma, V.; et al. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
105. Li, S.; Zhou, J.; Xu, T.; et al. Geomgcl: Geometric graph contrastive learning for molecular property prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 4541–4549.
106. Feng, S.; Ni, Y.; Li, M.; et al. UniCorn: A Unified Contrastive Learning Approach for Multi-view Molecular Representation Learning. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 21–27 July 2024; Volume 235, pp. 13256–13277.
107. Fang, X.; Liu, L.; Lei, J.; et al. Geometry-enhanced molecular representation learning for property prediction. *Nat. Mach. Intell.* **2022**, *4*, 127–134. https://doi.org/10.1038/s42256-021-00438-4.
108. Lu, S.; Gao, Z.; He, D.; et al. Data-driven quantum chemical property prediction leveraging 3D conformations with Uni-Mol+. *Nat. Commun.* **2024**, *15*, 7104. https://doi.org/10.1038/s41467-024-51321-w. Correction in *Nat. Commun.* **2024**, *15*, 8548.
109. Hou, Z.; Liu, X.; Cen, Y.; et al. GraphMAE: Self-Supervised Masked Graph Autoencoders. In Proceedings of the KDD'22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 594–604. https://doi.org/10.1145/3534678.3539321.
110. Wang, Y.; Wang, J.; Cao, Z.; et al. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.* **2022**, *4*, 279–287. https://doi.org/10.1038/s42256-022-00447-x.
111. Wan, Y.; Wu, J.; Hou, T.; et al. Multi-channel learning for integrating structural hierarchies into context-dependent molecular representation. *Nat. Commun.* **2025**, *16*, 413. https://doi.org/10.1038/s41467-024-55082-4.
112. Zhang, R.; Wang, X.; Liu, G.; et al. Motif-oriented representation learning with topology refinement for drug-drug interaction prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 1102–1110.
113. Dou, M.; Wen, L.; Xie, J.; et al. Closer to Biological Mechanism: Drug-Drug Interaction Prediction from the Perspective of Pharmacophore. In Proceedings of the AAAI Conference on Artificial Intelligence, Singapore, 20–27 January 2026; Volume 40, pp. 20887–20895.

114. Xie, J.; Wei, W.; Zhang, C.; et al. Informative Subgraph Extraction with Deep Reinforcement Learning for Drug-Drug Interaction Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Singapore, 20–27 January 2026; Volume 40, pp. 1319–1327.
115. Du, H.; Yao, Q.; Zhang, J.; et al. Customized subgraph selection and encoding for drug-drug interaction prediction. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 10–15 December 2024; Volume 37, pp. 109582–109608.
116. Wang, Z.; Xiong, Z.; Huang, F.; et al. ZeroDDI: A Zero-Shot Drug-Drug Interaction Event Prediction Method with Semantic Enhanced Learning and Dual-Modal Uniform Alignment. *arXiv* **2024**, arXiv:2407.00891.
117. Ryu, J.Y.; Kim, H.U.; Lee, S.Y. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4304–E4311. https://doi.org/10.1073/pnas.1803294115.
118. Wu, D.; Sun, W.; He, Y.; et al. MKG-FENN: A Multimodal Knowledge Graph Fused End-to-End Neural Network for Accurate Drug–Drug Interaction Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 10216–10224.
119. Su, X.; Hu, L.; You, Z.; et al. Attention-based Knowledge Graph Representation Learning for Predicting Drug-Drug Interactions. *Brief. Bioinform.* **2022**, *23*, bbac140. https://doi.org/10.1093/bib/bbac140.
120. Bai, Y.; Gu, K.; Sun, Y.; et al. Bi-Level Graph Neural Networks for Drug-Drug Interaction Prediction. *arXiv* **2020**, arXiv:2006.14002.
121. Lee, N.; Hyun, D.; Na, G.S.; et al. Conditional Graph Information Bottleneck for Molecular Relational Learning. In Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 18852–18871.
122. Yuan, Y.; Yue, J.; Zhang, R.; et al. PHGL-DDI: A pre-training based hierarchical graph learning framework for drug-drug interaction prediction. *Expert Syst. Appl.* **2025**, *270*, 126408. https://doi.org/10.1016/j.eswa.2025.126408.
123. Ma, T.; Chen, K.; Zang, Y.; et al. Self-supervised Blending Structural Context of Visual Molecules for Robust Drug Interaction Prediction. In Proceedings of the Advances in Neural Information Processing Systems, San Diego, CA, USA, 2–7 December 2025; Volume 38, pp. 133133–133161.
124. Liu, S.; Zhang, Y.; Cui, Y.; et al. Enhancing Drug-Drug Interaction Prediction Using Deep Attention Neural Networks. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 976–985. https://doi.org/10.1109/tcbb.2022.3172421.
125. Su, X.; Hu, P.; You, Z.H.; et al. Dual-Channel Learning Framework for Drug-Drug Interaction Prediction via Relation-Aware Heterogeneous Graph Transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 249–256.
126. Sun, Z.; Li, J.; Pergola, G.; et al. Exddi: Explaining drug-drug interaction predictions with natural language. In Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, PA, USA, 25 February–4 March 2025; Volume 39, pp. 25228–25236.
127. Wang, H.; Lian, D.; Zhang, Y.; et al. GoGNN: Graph of Graphs Neural Network for Predicting Structured Entity Interactions. *arXiv* **2020**, arXiv:2005.05537.
128. Chen, M.; Zhang, M.; Qu, C. Towards Interpretable Drug-Drug Interaction Prediction: A Graph-Based Approach with Molecular and Network-Level Explanations. In Proceedings of the KDD'25: The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Toronto, ON, Canada, 3–7 August 2025; pp. 203–214. https://doi.org/10.1145/3711896.3737163.
129. Zhao, C.; Liu, S.; Huang, F.; et al. CSGNN: Contrastive Self-Supervised Graph Neural Network for Molecular Interaction Prediction. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21), Montreal, QC, Canada, 19–27 August 2021; pp. 3756–3763. https://doi.org/10.24963/ijcai.2021/517.
130. Abdullahi, T.; Gemou, I.; Nayak, N.V.; et al. K-Paths: Reasoning over Graph Paths for Drug Repurposing and Drug Interaction Prediction. In Proceedings of the KDD'25: The 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Toronto, ON, Canada, 3–7 August 2025; pp. 5–16. https://doi.org/10.1145/3711896.3737011.
131. Huang, Y.; Peng, X.; Ma, J.; et al. 3DLinker: An E(3) Equivariant Variational Autoencoder for Molecular Linker Design. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 9280–9294.
132. Xu, M.; Luo, S.; Bengio, Y.; et al. Learning Neural Generative Dynamics for Molecular Conformation Generation. *arXiv* **2021**, arXiv:2102.10240.
133. Prykhodko, O.; Johansson, S.V.; Kotsias, P.C.; et al. A de novo molecular generation method using latent vector based generative adversarial network. *J. Cheminform.* **2019**, *11*, 74. https://doi.org/10.1186/s13321-019-0397-9.
134. Jo, J.; Lee, S.; Hwang, S.J. Score-based Generative Modeling of Graphs via the System of Stochastic Differential Equations. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 10362–10383.
135. Xu, M.; Powers, A.S.; Dror, R.O.; et al. Geometric Latent Diffusion Models for 3D Molecule Generation. In Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 38592–38610.
136. Ma, C.; Zhang, X. GF-VAE: A Flow-based Variational Autoencoder for Molecule Generation. In Proceedings of the CIKM'21: The 30th ACM International Conference on Information and Knowledge Management, Virtual, 1–5 November 2021; pp. 1181–1190. https://doi.org/10.1145/3459637.3482260.

137. Zhu, Y.; Ouyang, Z.; Liao, B.; et al. MolHF: A Hierarchical Normalizing Flow for Molecular Graph Generation. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23), Macau, China, 19–25 August 2023; pp. 5002–5010. https://doi.org/10.24963/ijcai.2023/556.
138. Li, C.; Yamanaka, C.; Kaitoh, K.; et al. Transformer-based Objective-reinforced Generative Adversarial Network to Generate Desired Molecules. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22), Vienna, Austria, 23–29 July 2022; pp. 3884–3890. https://doi.org/10.24963/ijcai.2022/539.
139. Lee, S.; Jo, J.; Hwang, S.J. Exploring chemical space with score-based out-of-distribution generation. In Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 18872–18892.
140. Liu, G.; Xu, J.; Luo, T.; et al. Graph Diffusion Transformers for Multi-Conditional Molecular Generation. In Proceedings of the Advances in Neural Information Processing Systems 37, Vancouver, BC, Canada, 10–15 December 2024; pp. 8065–8092. https://doi.org/10.52202/079017-0260.
141. Chang, J.; Ye, J.C. LDMol: A Text-to-Molecule Diffusion Model with Structurally Informative Latent Space Surpasses AR Models. In Proceedings of the International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 7374–7392.
142. Guan, J.; Zhou, X.; Yang, Y.; et al. DecompDiff: Diffusion Models with Decomposed Priors for Structure-Based Drug Design. In Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23–29 July 2023; pp. 11827–11846.
143. Zheng, X.; Jiang, S.; Seabra, G.; et al. Apo2Mol: 3D Molecule Generation via Dynamic Pocket-Aware Diffusion Models. In Proceedings of the AAAI Conference on Artificial Intelligence, Singapore, 20–27 January 2026; Volume 40, pp. 1614–1622.
144. Seo, S.; Kim, M.; Shen, T.; et al. Generative Flows on Synthetic Pathway for Drug Design. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025; Volume 2025, pp. 14029–14060.
145. Cretu, M.; Harris, C.; Igashov, I.; et al. SynFlowNet: Design of Diverse and Novel Molecules with Synthesis Constraints. In Proceedings of the International Conference on Learning Representations, Singapore, 24–28 April 2025; Volume 2025, pp. 47101–47126.
146. Zhou, X.; Cheng, X.; Yang, Y.; et al. DecompOpt: Controllable and Decomposed Diffusion Models for Structure-based Molecular Optimization. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
147. Huang, L.; Xu, T.; Yu, Y.; et al. A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. *Nat. Commun.* **2024**, *15*, 2657. https://doi.org/10.1038/s41467-024-46569-1.
148. Lee, S.; Jo, M.; Ok, J.; et al. Enhancing Ligand Validity and Affinity in Structure-Based Drug Design with Multi-Reward Optimization. In Proceedings of the International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 33129–33142.
149. Qiu, K.; Song, Y.; Fan, Z.; et al. Piloting Structure-Based Drug Design via Modality-Specific Optimal Schedule. In Proceedings of the International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 50619–50644.
150. Qiu, K.; Song, Y.; Yu, J.; et al. Empower Structure-Based Molecule Optimization with Gradient Guided Bayesian Flow Networks. In Proceedings of the International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 50645–50671.
151. Southiratn, T.; Koo, B.; Lu, Y.; et al. CombiMOTS: Combinatorial Multi-Objective Tree Search for Dual-Target Molecule Generation. In Proceedings of the International Conference on Machine Learning, Vancouver, BC, Canada, 13–19 July 2025; Volume 267, pp. 56650–56691.
152. Li, K.; Cai, X.; Wu, J.; et al. Fragment-Masked Diffusion for Molecular Optimization. *arXiv* **2024**, arXiv:2408.09106.
153. Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 2323–2332.
154. Wu, Z.; Zhang, O.; Wang, X.; et al. Leveraging language model for advanced multiproperty molecular optimization via prompt engineering. *Nat. Mach. Intell.* **2024**, *6*, 1359–1369. https://doi.org/10.1038/s42256-024-00916-5.
155. Maziarka, Ł.; Pocha, A.; Kaczmarczyk, J.; et al. Mol-CycleGAN: a generative model for molecular optimization. *J. Cheminform.* **2020**, *12*, 2. https://doi.org/10.1186/s13321-019-0404-1.
156. Chen, B.; Wang, T.; Li, C.; et al. Molecule Optimization by Explainable Evolution. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021.
157. Sun, M.; Xing, J.; Meng, H.; et al. MolSearch: Search-based Multi-objective Molecular Generation and Property Optimization. In Proceedings of the KDD'22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 4724–4732. https://doi.org/10.1145/3534678.3542676.
158. Zhu, Y.; Wu, J.; Hu, C.; et al. Sample-efficient multi-objective molecular optimization with gflownets. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023; Volume 36; pp. 79667–79684.
159. Shin, D.H.; Son, Y.H.; Lee, D.J.; et al. Dynamic Many-Objective Molecular Optimization: Unfolding Complexity with Objective Decomposition and Progressive Optimization. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24), Jeju, Republic of Korea, 3–9 August 2024; pp. 6026–6034. https://doi.org/10.24963/ijcai.2024/666.

160. Zhuang, Y.; Shen, D.; Sun, Y. MolEditRL: Structure-Preserving Molecular Editing via Discrete Diffusion and Reinforcement Learning. *arXiv* **2025**, arXiv:2505.20131.
161. Rao, H.; Li, K.; Xiong, Y.; et al. From Single-Step Edit Response to Multi-Step Molecular Optimization. *arXiv* **2026**, arXiv:2605.10035.
162. Reidenbach, D.; Krishnapriyan, A.S. CoarsenConf: Equivariant Coarsening with Aggregated Attention for Molecular Conformer Generation. *J. Chem. Inf. Model.* **2025**, *65*, 22–30. https://doi.org/10.1021/acs.jcim.4c01001.
163. Qin, Y.; Wei, X.; Xu, M.; et al. Comprehensive Benchmark Study of Diffusion-Based 3D Molecular Generation Models. *ACS Omega* **2025**, *10*, 42760–42775. https://doi.org/10.1021/acsomega.5c05077.
164. Axelrod, S.; Gómez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data* **2022**, *9*, 185. https://doi.org/10.1038/s41597-022-01288-4.
165. Isert, C.; Atz, K.; Jiménez-Luna, J.; et al. QMugs, quantum mechanical properties of drug-like molecules. *Sci. Data* **2022**, *9*, 273. https://doi.org/10.1038/s41597-022-01390-7.
166. Hoogeboom, E.; Satorras, V.G.; Vignac, C.; et al. Equivariant diffusion for molecule generation in 3D. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 8867–8887.
167. Bickerton, G.R.; Paolini, G.V.; Besnard, J.; et al. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4*, 90–98. https://doi.org/10.1038/nchem.1243.
168. Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **2009**, *1*, 8. https://doi.org/10.1186/1758-2946-1-8.
169. Parrot, M.; Tajmouati, H.; da Silva, V.B.R.; et al. Integrating synthetic accessibility with AI-based generative drug design. *J. Cheminform.* **2023**, *15*, 83. https://doi.org/10.1186/s13321-023-00742-8.
170. Gui, S.; Li, X.; Wang, L.; et al. Good: A graph out-of-distribution benchmark. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 2059–2073.
171. Ji, Y.; Zhang, L.; Wu, J.; et al. Drugood: Out-of-distribution dataset curator and benchmark for ai-aided drug discovery—A focus on affinity prediction problems with noise annotations. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 8023–8031.
172. Lin, Z.; Li, K.; Chen, J.; et al. Learning to Select Source Domains: Proxy-Rewarded Policy Optimization for Molecular OOD Generalization. *arXiv* **2026**, arXiv:2605.13932v2.
173. Li, J.; Guan, X.; Zhang, O.; et al. Leak Proof PDBBind: A Reorganized Data Set of Protein–Ligand Complexes for More Generalizable Binding Affinity Prediction. *J. Phys. Chem. B* **2026**, *130*, 730–740. https://doi.org/10.1021/acs.jpcb.5c08598.
174. Joeres, R.; Blumenthal, D.B.; Kalinina, O.V. Data splitting to avoid information leakage with DataSAIL. *Nat. Commun.* **2025**, *16*, 3337. https://doi.org/10.1038/s41467-025-58606-8.
175. Liu, G.; Zhao, T.; Xu, J.; et al. Graph Rationalization with Environment-based Augmentations. In Proceedings of the KDD'22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 14–18 August 2022; pp. 1069–1078. https://doi.org/10.1145/3534678.3539347.
176. Li, H.; Zhang, Z.; Wang, X.; et al. Learning invariant graph representations for out-of-distribution generalization. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 11828–11841.
177. Stumpfe, D.; Bajorath, J. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55*, 2932–2942. https://doi.org/10.1021/jm201706b.
178. Cui, C.; Su, X.; Zhang, Z.; et al. Activity-Cliff Awareness Enables Robust Graph Learning for Molecular Property Prediction. *Nat. Commun.* **2026**, *17*, 9376. https://doi.org/10.1038/s41467-026-75713-2.
179. Chen, Y.; Wu, T.; Jiang, Y.; et al. MAPCliff-WMGR: Exploring Activity Cliffs in Molecular Activity Prediction Enhanced by Weighted Molecular Graph Representations. *J. Chem. Inf. Model.* **2025**, *65*, 12978–12989. https://doi.org/10.1021/acs.jcim.5c01915.
180. Hu, D.; Li, K.; Rao, H.; et al. When Molecular Similarity Works: Property Cliffs Reveal Hidden Errors. *arXiv* **2026**, arXiv:2605.17265v1.
181. Bekker, J.; Davis, J. Learning from positive and unlabeled data: a survey. *Mach. Learn.* **2020**, *109*, 719–760. https://doi.org/10.1007/s10994-020-05877-5.
182. Li, P.; Shao, B.; Zhao, G.; et al. Negative sampling strategies impact the prediction of scale-free biomolecular network interactions with machine learning. *BMC Biol.* **2025**, *23*, 123. https://doi.org/10.1186/s12915-025-02231-w.
183. Graff, D.E.; Shakhnovich, E.I.; Coley, C.W. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* **2021**, *12*, 7866–7881. https://doi.org/10.1039/d0sc06805e.
184. van Tilborg, D.; Grisoni, F. Traversing chemical space with active deep learning for low-data drug discovery. *Nat. Comput. Sci.* **2024**, *4*, 786–796. https://doi.org/10.1038/s43588-024-00697-2.
185. Hao, M.; Gong, J.; Zeng, X.; et al. Large-scale foundation model on single-cell transcriptomics. *Nat. Methods* **2024**, *21*, 1481–1491. https://doi.org/10.1038/s41592-024-02305-7.
186. Yuan, H.; Yu, H.; Gui, S.; et al. Explainability in Graph Neural Networks: A Taxonomic Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 5782–5799. https://doi.org/10.1109/tpami.2022.3204236.

187. Jiménez-Luna, J.; Grisoni, F.; Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2020**, *2*, 573–584. https://doi.org/10.1038/s42256-020-00236-4.
188. Wellawatte, G.P.; Seshadri, A.; White, A.D. Model agnostic generation of counterfactual explanations for molecules. *Chem. Sci.* **2022**, *13*, 3697–3705. https://doi.org/10.1039/d1sc05259d.
189. Alizadehsani, R.; Oyelere, S.S.; Hussain, S.; et al. Explainable Artificial Intelligence for Drug Discovery and Development: A Comprehensive Survey. *IEEE Access* **2024**, *12*, 35796–35812. https://doi.org/10.1109/access.2024.3373195.
190. Ding, Q.; Yao, R.; Bai, Y.; et al. Explainable Artificial Intelligence in the Field of Drug Research. *Drug Des. Dev. Ther.* **2025**, *19*, 4501–4516. https://doi.org/10.2147/dddt.s525171.
191. Delile, J.; Mukherjee, S.; Mueller, J.; et al. Foundation models in drug discovery: Phenomenal growth today, transformative potential tomorrow? *Drug Discov. Today* **2025**, *30*, 104518. https://doi.org/10.1016/j.drudis.2025.104518.
192. Choi, J.; Nam, G.; Choi, J.; et al. A Perspective on Foundation Models in Chemistry. *JACS Au* **2025**, *5*, 1499–1518. https://doi.org/10.1021/jacsau.4c01160.
193. Li, K.; Wu, Z.; Wang, S.; et al. DrugPilot: LLM-based Parameterized Reasoning Agent for Drug Discovery. *arXiv* **2025**, arXiv:2505.13940v2.
194. Li, K.; Wu, Z.; Xiong, Y.; et al. BSL: A Unified and Generalizable Multitask Learning Platform for Virtual Drug Discovery from Design to Synthesis. *arXiv* **2025**, arXiv:2508.01195v1.
195. Yu, J.; Li, K.; Wu, Z.; et al. SkillMol: Skill-Conditioned Action Editing for Molecular Optimization. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2026, Budapest, Hungary, 24–29 October 2026; pp. 19428–19445.
196. Kretschmer, F.; Seipp, J.; Ludwig, M.; et al. Coverage bias in small molecule machine learning. *Nat. Commun.* **2025**, *16*, 554. https://doi.org/10.1038/s41467-024-55462-w.
197. Pastore, E.P.; De Rango, F. Foundation and Multimodal Models for Drug Discovery in Molecular Informatics: Principles, Evaluation, and Practical Guidance. *Mol. Inform.* **2026**, *45*, e70027. https://doi.org/10.1002/minf.70027.
198. Huang, K.; Chandak, P.; Wang, Q.; et al. A foundation model for clinician-centered drug repurposing. *Nat. Med.* **2024**, *30*, 3601–3613. https://doi.org/10.1038/s41591-024-03233-x.