

# Recent Developments in GNNs for Drug Discovery

Zhengyu Fang<sup>1</sup>, Xiaoge Zhang<sup>1</sup>, Anyin Zhao<sup>1</sup>, Xiao Li<sup>1,2,3,4</sup>, Huiyuan Chen<sup>1</sup>, and Jing Li<sup>\*1</sup>

<sup>1</sup>Department of Computer and Data Sciences, Case Western Reserve University

<sup>2</sup>Department of Biochemistry, Case Western Reserve University

<sup>3</sup>Center for RNA Science and Therapeutics, Case Western Reserve University

<sup>4</sup>Department of Biomedical Engineering, Case Western Reserve University

## Abstract

In this paper, we review recent developments and the role of Graph Neural Networks (GNNs) in computational drug discovery, including molecule generation, molecular property prediction, and drug-drug interaction prediction. By summarizing the most recent developments in this area, we underscore the capabilities of GNNs to comprehend intricate molecular patterns, while exploring both their current and prospective applications. We initiate our discussion by examining various molecular representations, followed by detailed discussions and categorization of existing GNN models based on their input types and downstream application tasks. We also collect a list of commonly used benchmark datasets for a variety of applications. We conclude the paper with brief discussions and summarize common trends in this important research area.

**Keywords:** Drug Discovery, Graph Neural Networks, Machine Learning

## 1 Introduction

It is well known that traditional drug discovery is costly, time-consuming, and with high failure rates [1]. To streamline the process of drug discovery and mitigate resource-intensive laboratory work, significant research has been dedicated to the development of computational methods. Existing literature provides some comprehensive reviews on deep learning approaches in drug discovery [2, 3, 4, 5]. In this review, we focus on the development and applications of Graph Neural Networks (GNNs) on three related areas of computational drug development, namely, *Molecule Generation*, *Molecular Property Prediction*, and *Drug-Drug Interaction Prediction*, which not only receive increasing attention but also show promising results. We will summarize some most recent developments in these research areas and focus on computational advances published since 2021.

---

\*Corresponding author. [jingli@case.edu](mailto:jingli@case.edu)

## 1.1 Molecule Generation

An earlier step in developing a drug for a certain disease usually involves the search of enormous amount of molecules to create a small subset of candidates for further laboratory testing. To reduce experimental costs and to produce drug candidates more efficiently, many computational approaches have been developed to generate novel molecules [6, 7, 8, 9, 10]. These approaches usually take some existing molecules as inputs and generate valid but different molecules with high variety. Validity basically means that the generated molecules are possible to be synthesized. At the same time, high variety will ensure that the generated molecules span over a reasonably large molecular space in hope that some of them will have desired properties. Computational approaches are usually evaluated based on these two metrics.

Generally speaking, existing computational approaches can be characterized as two different types based on whether they can incorporate certain constraints. For some approaches [6, 7], any molecules can be generated, *i.e.*, the generation of molecules has no constraints associated with them. For the second type of approaches [8, 9, 10], they only generate molecules satisfying certain constraints. For example, all the generated molecules must include some fixed substructures such as ring structures. In other cases, the constraints can be task specific, for example, many approaches [11, 12, 13, 14, 15] have been developed to generate molecules that bind to certain protein binding sites.

Early computational approaches for molecule generation were mostly applications of generative deep learning models, such as variational autoencoder [16]. More recently, there was a shift from generations of whole molecules all at once, to iterative generations of atoms and bonds [6, 8]. Many of such approaches utilized GNN modules as backbones to generate molecules with specific substructures or with high binding affinity to specific proteins [8].

## 1.2 Molecular Property Prediction

Once drug candidates are selected, researchers need to study drugable properties of candidate molecules, such as toxicity and possible adverse effects, water solubility, and binding affinity to target proteins. Experimental studies are slow and costly. With quick accumulation of high-quality training data, a great number of computational methods for molecular property prediction have been developed [17, 18, 19, 20], the goals of which are to reduce the need for extensive experimental validation and to accelerate the drug discovery process.

Molecular properties can be classified into different categories based on their levels, from molecular level (e.g., quantum mechanics) to physiological level [21]. Computationally, the tasks of molecular property prediction can be categorized into two broad classes based on their inputs: predictions based on single molecules (*i.e.*, drug candidates) [17, 22, 23], and predictions of relationships based on two or more molecules (e.g., drug-target or protein-ligand) [24, 25, 26]. The predictions based on single molecules can be either classification-based or regression-based. The goal of classification tasks is to predict the presence/absence of certain properties. The regression tasks focus on the prediction of the quantitative value of a particular property of a molecule. Similarly, relationships among molecules [27, 28] can be represented by a binary value or a numerical value. Examples of the former include whether a protein and a target interact. The affinity score between two molecules is an example of the latter.

Traditionally, drug property prediction heavily relied on classical machine learning techniques such as random forests or support vector machines [2], necessitating deep domain knowledge for feature engineering. With the emergence of deep learning, there was a significant shift towards neural network based models [2, 29]. This transition marked a move from one-dimensional string representations of drugs to richer two-dimensional graph and three-dimensional conformation models, broadening the scope and accuracy of drug property prediction. In particular, GNNs, which take in a molecular graph representation and directly

predict molecular properties, have attracted substantial attention in cheminformatics and bioinformatics (*e.g.*, [30]).

### 1.3 Drug-Drug Interaction Prediction

In the treatment of complex diseases such as cancer and neurological disorders, combinational drug therapies have shown promise in improving treatment efficacy by targeting multiple biological pathways simultaneously [4, 31, 32]. However, as the number of possible drug combinations grows with the increasing number of available drugs, the potential for undetected and unexpected drug-drug interaction (DDIs) also rises [33]. Such interactions can lead to reduction of therapeutic effectiveness, adverse side effects [34], and sometimes increased hospitalization [35], posing significant challenges to the screening and optimization of combinational drug therapies. Therefore, it is crucial to develop methods that can precisely predict the drug-drug synergies and adverse interactions to ensure safer and more effective treatments.

Strictly speaking, drug combination therapies, which rely on synergistic or additive interactions of drugs to increase efficacy, reduce toxicity, and/or prevent drug resistance, can be treated as one type of drug-drug interactions. In this review, we use the broad definition of drug-drug interactions, which include both synergistic/additive as well as antagonistic/adverse interactions. To predict drug-drug interactions computationally, the inputs usually include representations of drugs and types of interactions. In the simplest case, methods may only focus on the prediction of the existence of a particular type of interaction [36]. The problem is simply formulated as a binary classification problem. In other cases (*e.g.*, data from cell-based assays), the relationships can be quantified based on one or multiple measures of individual drugs and/or drug pairs (*e.g.*, half maximal inhibitory concentration or IC50), which can be viewed as regression tasks. Yet in some other cases (data from patient-based studies), different criteria or biomarkers (*e.g.*, blood pressure, blood sugar) are considered and only a few categorical values (*e.g.*, increasing, no change, decreasing) are considered for each criterion/biomarker. The problem can be formulated as a multi-class classification problem [37].

Over the years, many computational approaches have been developed for drug-drug interaction predictions [37, 36, 38, 39]. Earlier approaches were mostly based on traditional machine learning algorithms and/or matrix decomposition framework [38]. Some methods [40, 41, 42] can even take into considerations of patients’ medical history and recommend drugs that are safe to use for specific patients. More recently, more and more approaches based on deep learning models have drawn increasing attention, and have achieved better performance [38]. In particular, with advancements in the development of GNNs, methods that adapt GNNs as backbones to incorporate interactions as networks have achieved state-of-the-art results [42].

## 2 Representations of Molecules

In general, molecules can be represented via fingerprints, the Simplified Molecular Input Line Entry System (SMILES) strings, or 2D-/3D-graphs in Fig. 1. Binary fingerprints representing molecular substructure or topology allows efficient computation and database search [43]. However, they cannot easily encode global features of molecules such as size and shape. SMILES is the most widely used linear representation for describing chemical structures since its invention [44], and is superior to other one-dimensional representation schemes such as binary fingerprints. However, there are innate limitations associated with the internal structure of SMILES representations when used in Natural Language Processing (NLP) algorithms.

Molecules can be represented as 2D graphs, where nodes represent atoms and edges represent chemical bonds, or as 3D graphs that also incorporate 3D coordinates, providing detailed spatial information. In

both cases, both nodes and edges can have their own unique properties or features. While the 2D graph representation is simpler, 3D representation can better capture interactions based on distances and angles between atoms [45, 46], thus providing a more comprehensive view that is crucial for modeling molecular dynamics and bindings.

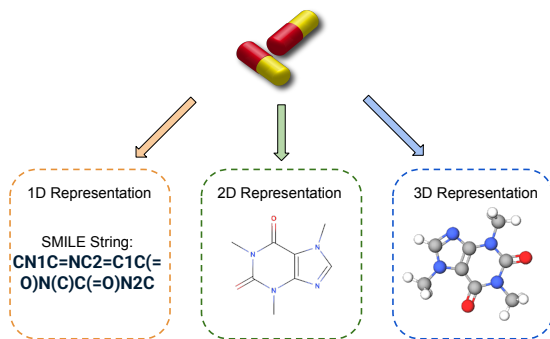


Figure 1: Various molecular representations commonly employed in computational drug discovery models: 1D SMILES strings, 2D molecular graphs, and 3D molecular graphs.

With molecules intuitively represented as graphs, GNNs offer a natural framework for handling and analyzing molecular data. This synergy has sparked extensive research, positioning GNNs at the forefront of innovation in molecular property and interaction prediction. GNNs allow nodes to aggregate information through their edges, creating comprehensive graph representations. Furthermore, by combining graph structures with neural networks, GNNs can readily handle both classification and regression tasks [47, 48, 49]. Therefore, many innovations in GNNs focus on graph representation learning, rather than specific prediction tasks. Most of the approaches to be discussed in this survey thus can be utilized in many downstream applications.

### 3 Molecule Generation

One of the initial steps in drug discovery involves selecting a set of candidate molecules for early-stage screening. Traditionally, this process has been conducted in laboratories. However, recent advances in computational methods have expanded the toolkit available to researchers, including the use of GNN models for molecule generation. These models leverage 2D and 3D graph representations of molecules to efficiently generate novel compounds within defined constraints. The development of GNN-based molecule generation methods can be broadly categorized into three types, as illustrated in Fig. 2: unconstrained generation [6, 7, 50, 51], constrained generation with targeted substructures [8, 10, 9], and ligand-protein-based generation [11, 12, 13, 14, 15]. Unconstrained methods prioritize structural diversity, constrained approaches focus on generating molecules containing specific functional groups or motifs relevant to desired chemical or biological properties, and ligand-protein-based strategies are designed to produce molecules that interact with specific protein targets. These advances demonstrate the versatility and promise of GNNs in accelerating drug discovery and highlight their pivotal role in identifying novel therapeutic candidates [52, 53, 54, 55].

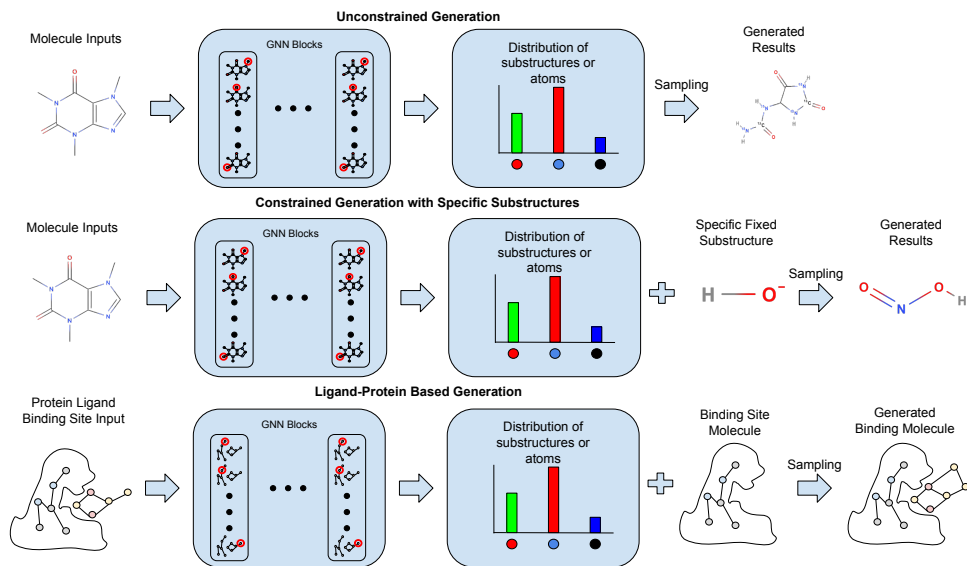


Figure 2: The general framework of three different types of molecule generation processes. Molecular graphs and protein-ligand complexes are fed into GNN backbone models, which output the probability distribution of molecular substructures to be sampled, based on which the models select substructures and assemble the resulting molecules.

### 3.1 Unconstrained Generation

For computational methods to generate promising drug candidates, a natural approach is to generate chemically valid molecules in the vicinity of existing drugs and drug candidates (*i.e.*, known training data). In other words, computational models are typically trained on “hit molecules”—compounds that already exhibit desirable properties for drug development. Because there are no additional constraints on molecular substructures or chemical properties (*e.g.*, binding affinity) imposed explicitly, such methods are classified as unconstrained. The primary goal of such approaches is to generate structurally valid and chemically plausible molecules with high diversity, guided by the distribution of the training data.

GraphINVENT [56] is one of the initial explorations into unconstrained molecular graph generation, which explored various GNN architectures—such as gated GNNs and attention-based GNNs—to learn an “action” probability distribution from training molecules. The model then iteratively sampled one action at a time, *e.g.*, adding a bond or an atom, until a “terminate generation” action was reached. Doing so allowed the approach to construct a molecule step by step. GraphINVENT is a representative among a broader class of models [57, 58, 59, 60] that all adopt such an iterative sampling strategy that build new graphs by repeatedly sampling components or actions from learned distributions.

Beyond this one-piece-at-a-time paradigm, another type of method adopted generative frameworks that produce complete molecules (graphs) in a single pass. For instance, ConfVAE [61] integrated both 2D molecular graphs and 3D conformations, ensuring rotational and translational invariance. It employed a Conditional Variational Autoencoder (CVAE) framework, with Message Passing Neural Networks (MPNNs) for graph encoding, enabling end-to-end conditional molecule generation. While ConfVAE leveraged GNNs within a generative framework, VonMisesNet [7] took a different approach by focusing on capturing the re-

alistic distribution of torsional angles in molecules. It introduced a novel GNN architecture that sampled torsion angles from the Von Mises distribution, which better reflected the physical constraints of molecular geometry. Moreover, VonMisesNet addressed key challenges such as chirality inversion of atoms and supports molecules with a large number of rotatable bonds-enhancing the chemical accuracy and diversity of its outputs.

Overall, unconstrained generation models are increasingly integrating with generative architectures such as VAEs and GANs, aiming to better approximate the true data distribution. Nonetheless, as the need grows to reduce the cost of candidate selection, improve the quality of generated molecules, and generate molecules that could be synthesized, research is gradually shifting toward generation models that incorporate explicit constraints on substructures and target properties.

### 3.2 Constrained Generation with Specific Substructures

In drug development, generating molecules with specific substructures and targeted chemical properties is often more desirable than unconstrained molecule generation. MoLeR, introduced by Maziarz et al. [8], demonstrated the capability to perform both constrained and unconstrained molecule generation. It utilized motifs—common chemical substructures—within an encoder-decoder framework. By combining GNN and Multilayer Perceptron (MLP) blocks, MoLeR carefully constructed molecules one motif at a time, sequentially selecting motifs or atoms, determining attachment points, and assigning bond types, each step optimized for molecular validity and functionality.

Building on the success of fragment-based approaches like MoLeR, newer models have further advanced constrained molecule generation. One such model, GEAM [10], introduced the Graph Information Bottleneck (GIB) principle to identify substructures most relevant to specific drug properties. GEAM first extracted a vocabulary of meaningful substructures and then assembled molecules from this learned vocabulary. A soft actor-critic (SAC) reinforcement learning algorithm was used to identify high-quality samples, which were subsequently mutated through a genetic algorithm (GA) to produce final molecules that were chemically valid and aligned with the desired drug properties.

While models like MoLeR and GEAM focused primarily on the generation process, MiCam [9] proposed a novel strategy for building a chemically “reasonable” motif vocabulary. MiCam addressed the limitation of previous fragment-based methods, which often failed to identify appropriate motifs for molecule generation. Its vocabulary construction involved two phases: in the merging-operation learning phase, the model iteratively merged the most frequent atomic patterns found across molecules to form a preliminary set of motifs. In the motif-vocabulary construction phase, the model disconnected fragments at learned attachment points, marking these connection sites with special tokens to preserve the information necessary for molecule assembly. This approach allowed MiCam to flexibly generate molecules either by adding known motifs or by extending partially generated structures based on the connection history.

Overall, these models share a common strategy of using substructures as modular building blocks, aligning generation objectives with the training loss, and constructing molecules in a stepwise manner. Among the three, GEAM and MiCam offer greater flexibility in incorporating specific constraints on both substructures and chemical properties, as they allow the use of both atoms and motifs during generation. In contrast, MoLeR primarily relies on starting from a predefined scaffold.

### 3.3 Protein-Ligand based Generation

In addition to generating molecules based solely on training data of individual compounds, researchers have developed models that focus on protein binding sites and their associated ligands, addressing a new

set of challenges in drug discovery. Recent advances in GNN-based molecule generation have enabled the creation of molecules specifically tailored for target proteins. These models employ GNN blocks to maintain structural consistency—ensuring robustness against flips, shifts, and rotations—while processing attributes and 3D coordinates of protein binding sites. Approaches such as the AR model [11] and GraphBP [12] introduced distinct strategies for representing atoms and binding environments.

These models adopted various techniques to prioritize contextual representation and resilience to rigid transformations. For instance, AR combined MLP blocks with an auxiliary network to guide atom generation and bonding decisions, whereas GraphBP utilized spherical coordinates alongside MLPs for sequential atom-by-atom construction. Other notable methods, including Pocket2Mol [13] and FLAG [14], incorporated auxiliary MLP classifiers and predictors to optimize atom positioning and motif attachment. Collectively, these strategies significantly improved model robustness and adaptability, representing critical progress toward customizing molecules for specific protein targets—a key advancement in drug discovery.

Each model has distinct strengths: the AR model emphasized specificity and binding affinity optimization for particular protein site structures; GraphBP introduced a dual-diffusion architecture to enhance flexibility; Pocket2Mol achieved greater computational efficiency through conditional 3D coordinate sampling; and FLAG leveraged motif-based generation to improve structural realism and diversity.

Beyond generating molecules that bind to specific protein sites, there is also a growing need to generate molecules constrained by a desired 3D binding conformation. In many cases, experimental data reveal that certain binding postures are particularly effective for interacting with specific proteins, and generating candidate molecules that adopt these conformations can greatly accelerate screening. SQUID [15] was the first model designed to address this challenge. Given a target 3D shape, SQUID encodes the input conformation—treated as an unordered point cloud—into hidden features using GNN layers, and then iteratively generates 3D molecular fragments that reconstruct the desired shape fragment-by-fragment.

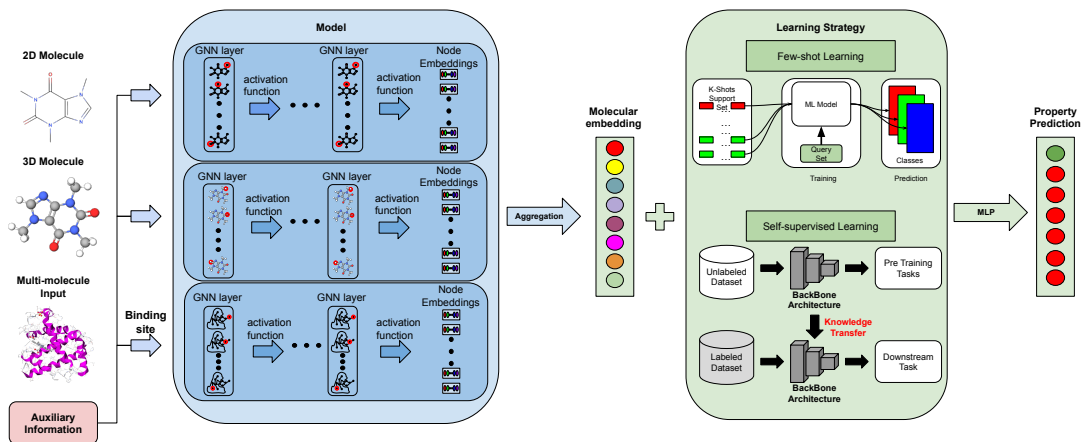


Figure 3: The general framework for GNN-based drug property and interaction prediction. Three common types of inputs are used individually or jointly: 2D molecule graphs, 3D molecule graphs, multi-molecule interaction graphs such as protein-ligand complexes. Additional auxiliary information can also be incorporated by some approaches. These inputs are then fed into GNN models, which aggregate information from neighboring nodes and produced final latent node representations. To alleviate the label sparsity issue, various learning strategies, such as few-shot learning or self-supervised learning, are widely adopted.



## 4 Prediction of Molecular Properties and Interactions

In this section, we review various GNN-based approaches for predicting molecular properties and interactions. Different algorithms have been developed for different applications, each utilizing varying types of inputs. Common types of input data include 2D molecular graphs, 3D molecular conformations, multi-molecule complexes, and potentially additional auxiliary information (Fig. 3).

A key component shared by all approaches is the learning of a latent molecular representation through GNNs, which is subsequently used for prediction tasks. To address challenges such as label sparsity, many methods also incorporate advanced learning strategies, including self-supervised (pre-training) and few-shot learning. In this section, we categorize and discuss these approaches based on their specific prediction tasks and input data types, highlighting their novel contributions and offering insights into how they advance the field.

### 4.1 Property Prediction and Molecule Representation Learning Based on 2D Graphs

Learning effective molecular representations is a fundamental step in property prediction. Generally, the goal of molecular representation learning is to embed molecules into numerical vectors in a latent space, enabling a variety of downstream tasks [17]. Therefore, much of the innovation in property prediction from 2D graphs lies in the strategies developed for learning molecular representations, which is the focus of this subsection.

When molecules are represented as 2D graphs, early studies typically employed MPNNs [62], wherein nodes (atoms) exchange messages with neighboring nodes, aggregating this information to update their respective states. However, obtaining class labels or quantitative measurements often requires costly laboratory experiments or human annotations. As a result, many datasets contain large numbers of unlabeled or imbalanced samples. To address this, newer approaches often incorporate pre-training strategies or rely on few-shot learning.

#### 4.1.1 Pre-training

Pre-training<sup>1</sup> has emerged as a promising technique to mitigate label scarcity and is widely adopted in modern GNN models for molecular property prediction. For example, MGSSL [63] introduced a motif-based graph self-supervised learning framework that exploited rich information in subgraphs often overlooked at the node level. In this setting, the choice of molecular fragmentation method is critical, as poor fragmentation can lead to suboptimal motif generation and degraded model performance.

Many other methods adopt contrastive graph learning as their pre-training strategy. MoCL [18], for instance, utilized knowledge-aware contrastive learning informed by local and global domain knowledge. Local domain knowledge ensured semantic invariance during augmentation, while global domain knowledge infused structural similarity across the learning process. Nevertheless, designing augmentation schemes that generalize across diverse molecular structures remains challenging. KCL [64] combined contrastive learning with domain-specific knowledge graphs, offering tailored augmentations at the cost of generalizability. MCHNN [65] applied multi-view contrastive learning with task-specific augmentations, enhancing the expressiveness of molecular representations. HiMol [66] advanced self-supervised learning by proposing a hierarchical GNN that captured node, motif, and graph-level representations. However, constructing motif dictionaries can be computationally intensive, especially for large molecular databases.

---

<sup>1</sup>Here, pre-training refers to self-supervised learning on the input data itself or its augmented versions, differing slightly from the conventional notion in transfer learning where models are first trained on large independent datasets.



### 4.1.2 Few-shot Learning

Few-shot learning approaches aim to predict molecular properties with minimal labeled data [67]. HSL-RG [19] explored both global and local structural semantics for few-shot learning: global information was captured via molecular relation graphs built from graph kernels, while local information was learned through transformation-invariant representations. Similarly, MHNfs [68] employed a context module that retrieved and enriched molecule representations from a large pool of reference molecules, although the requirement for such large reference sets may limit its practical use.

GS-Meta [69] extended few-shot learning to simultaneously handle multiple properties or labels. PACIA [20] introduced hypernetworks to generate adaptive parameters for modulating the GNN encoder, reducing overfitting while maintaining flexibility. However, designing effective hypernetworks demands significant domain expertise and may constrain the method’s generalizability across tasks. An alternative strategy to combat label scarcity involves grammar-based generation, as exemplified by Geo-DEG [70], which employed a hierarchical molecular grammar to create molecular graphs, using production paths as informative priors for structural similarity.

### 4.1.3 Incorporating Auxiliary Information

Beyond the standard 2D graph representation, researchers have explored integrating additional molecular information during learning. For example, PhysChem [22] showed that incorporating physical and chemical properties improved molecular representations, though its cooperation mechanism was less interpretable than traditional ensemble strategies. O-GNN [23] incorporated ring priors into the modeling, leveraging their importance in determining molecular properties. MoleOOD [71] introduced invariant substructure learning to better handle distribution shifts across environments. However, in datasets with large environmental variations, MoleOOD’s advantage over simpler methods like ERM [72] may diminish.

Some recent methods have also integrated both one-dimensional sequential encodings (*e.g.*, SMILES strings) and 2D graphs to jointly leverage information from both views. DVMP [73], for instance, encoded molecular graphs via GNNs and SMILES sequences via Transformers, using a dual-view consistency loss to maintain semantic coherence. However, involving both Transformer and GNN branches significantly increases training costs compared to single-branch models.

## 4.2 Property Prediction based on 3D-graphs

Recent advancements in deep learning and molecular science, driven by the increasing availability of large-scale 3D molecular datasets, have significantly advanced property prediction based on 3D molecular graphs. SphereNet [45] introduced an innovative approach for 3D molecular representation learning, proposing a spherical message-passing scheme that explicitly incorporates 3D spatial information. While SphereNet demonstrated strong predictive performance, it lacked transparency and interpretability, which hinders the understanding of its decision-making processes. MolKGNN [74] addressed this limitation in the context of quantitative structure–activity relationship (QSAR) modeling. It enhanced 3D molecular representation learning by employing molecular graph convolution with learnable molecular kernels, effectively capturing chemical patterns. Importantly, MolKGNN incorporated molecular chirality, a critical aspect often neglected in previous models. However, the method emphasizes distinguishing specific molecular substructures, which may limit its generalizability across diverse chemical variations encountered in practical drug discovery scenarios.

Several studies have explored integrating both 2D and 3D information for property prediction. For instance, GraphMVP [75] developed a 2D graph encoder enriched by discriminative 3D geometric informa-

tion. It employed a self-supervised pre-training strategy that leveraged the correspondence and consistency between 2D topologies and 3D conformations. Similarly, 3D-Informax [30] proposed a transfer learning framework that pre-trained on molecules with both 2D and 3D data and then transferred the learned knowledge to molecules with only 2D structures. However, such approaches may risk overfitting, as evidenced by performance variability across different datasets.

UnifiedPML [76] further improved representation learning by jointly considering 2D and 3D information in its pre-training scheme. The framework employed three complementary tasks: reconstruction of masked atoms and coordinates, generation of 3D conformations conditioned on 2D graphs, and generation of 2D graphs conditioned on 3D conformations. GeomGCL [77] adopted a dual-channel message-passing neural network to effectively capture both topological and geometric features of molecular graphs.

MoleculeSDE [78] unified 2D and 3D molecular representations by treating them as separate modalities in a multi-modal pre-training framework. 3D-PGT [79] proposed a generative pre-training approach on 3D molecular graphs, which was subsequently fine-tuned on molecules lacking 3D structural data. It employed a multi-task learning strategy based on three geometric descriptors—bond lengths, bond angles, and dihedral angles—and used total molecular energy as an optimization target. While promising, the effectiveness of this framework remains to be validated on larger and more structurally complex molecules, as current evaluations have primarily focused on small molecules.

### 4.3 Interaction Prediction

Beyond property prediction, interaction prediction has been extensively explored, especially for drug-target or drug-disease interaction predictions. Many researchers (e.g., NeurTN [17]) have utilized drug-target interaction networks as input, where the nodes are drugs and targets and links are known drug-target relationship. These models typically infer new interactions based on the guilt-by-association principle and are fundamentally different from the methods discussed in this work, which rely primarily on drugs’ structure information represented as molecular graphs. Therefore, interaction network based approaches are excluded from further discussion in this section.

In drug discovery, one of the most critical and extensively studied relationships is the interaction between drugs (or chemical compounds) and their protein targets. In the literature, various terms are used to describe these interactions, each emphasizing different aspects. These include drug-target interaction, protein-ligand interaction, drug-target binding affinity, protein-ligand binding affinity, molecular docking. Computationally, given the 2D or 3D structures of two molecules, the interaction can be studied at three levels: (1) binary interaction (*i.e.*, whether an interaction occurs), (2) binding affinity (a numerical value, typically reflecting binding free energy), and (3) docking or protein-ligand binding dynamics.

GNN-based approaches have been proposed to predict drug-target interactions based on their 2D structure. For example, CGIB [28] predicted interactions primarily using substructure information from paired graphs. MGraphDTA [27] predicted drug-target binding affinities based on 2D compound graphs and protein sequences. It utilized a deep GNN to capture both local and global molecular structures and a multi-scale convolutional neural network (CNN) to extract features from protein sequences. However, capturing long-range dependencies within complex molecular graphs remains a challenge for such models. Given the superior performance generally observed when utilizing 3D molecular geometries, the trend shows that more approaches incorporate 3D information for interaction prediction, especially for binding affinity prediction and docking.

For binding affinity prediction, the inputs are usually protein-ligand complexes, and the objective is to predict a binding score that reflects the strength of interaction, typically in terms of free energy. Recent developments have leveraged 3D graph representation learning to tackle this problem. For instance, Jones