ChemPRM: Improving Retrosynthesis by Structured Intermediate Process Reward

Lingfeng Zhang [*1] **Xianbin Ye** [*2] **Lihang Liu** [*2] **Xiaonan Zhang** [2] **Xiaomin Fang** [2] **Luca Rossi** [1]

Abstract

Retrosynthesis prediction plays a central role in computer-aided drug discovery, as it requires recall of feasible precursor molecules for a given target compound. Despite substantial progress driven by deep learning approaches, existing models often perform direct product-to-reactant mapping without explicitly encoding chemical reasoning, which limits interpretability and can result in chemically implausible predictions. In this work, we propose ChemPRM, a structured framework for single-step retrosynthesis that decomposes prediction into a sequence of chemically interpretable intermediate states. The framework introduces a structured intermediate process reward and applies supervised fine-tuning on explicit intermediate annotations to guide the model toward chemically valid reasoning trajectories. Experiments on the USPTO-50k benchmark demonstrate that ChemPRM achieves competitive performance relative to state-of-the-art methods, while substantially improving interpretability and robustness.

1. Introduction

Retrosynthesis prediction constitutes a foundational problem in computer-aided molecular design and drug discovery, as it focuses on identifying feasible precursor molecules for the synthesis of complex target compounds. The field has evolved from rule-based systems that emphasize interpretability to data-driven deep learning approaches that prioritize predictive accuracy, including graph neural networks (GNNs) (Shi et al., 2020; Somnath et al., 2021; Chen et al., 2023), Transformer-based sequence models (Zheng et al., 2019; Wan et al., 2022; Irwin et al., 2022), and large language models (LLMs) (Zhao et al., 2025b; Zhang et al., 2024). Despite this progress, several fundamental challenges remain unresolved. Contemporary data-driven models are typically designed as black-box systems that map products directly to reactants without explicitly modeling the underlying chemical mechanisms. As a result, these models often generate syntactically valid yet chemically implausible reactions, a phenomenon commonly referred to as chemical hallucination (Reed, 2025), which undermines reliability and practical applicability.

To address the gap between predictive performance and chemical interpretability, we propose **ChemPRM**, a novel retrosynthesis framework that decomposes single-step prediction into a **structured intermediate process**. Instead of formulating retrosynthesis as a direct sequence transduction problem, ChemPRM organizes the reasoning procedure into three chemically grounded stages: (1) **Reaction Site Identification**, which determines the atoms and bonds involved in the transformation; (2) **Bond Disconnection**, which generates intermediate synthons based on the identified reaction sites; and (3) **Reactant Completion**, which converts synthons into stable and chemically valid precursor molecules. This decomposition reflects the stepwise reasoning strategy employed by human chemists and anchors model predictions in interpretable chemical operations.

To further promote chemical validity throughout the reasoning procedure, we introduce a **structured intermediate process reward** together with supervised fine-tuning (SFT) on explicitly annotated intermediate states. This training strategy constrains the model to follow chemically plausible trajectories and enables fine-grained supervision at each reasoning stage, in contrast to end-to-end approaches that provide supervision only at the level of final outputs. Empirical evaluation on the USPTO-50k (Schneider et al., 2016) benchmark shows that ChemPRM attains strong performance relative to state-of-the-art retrosynthesis models, with notable advantages in complex and multi-center reactions. Furthermore, we analyze the scaling behavior of the proposed framework to investigate how model capacity affects the ability to capture structured chemical reasoning.

2. Related Work

Retrosynthesis prediction has long been a core challenge in computational chemistry, motivating the development of

[*]Equal contribution [1]Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China [2]Baidu, China. Correspondence to: Luca Rossi <luca.rossi@polyu.edu.hk>.

Published in the AI for Science workshop (ICML 2026).

a diverse set of modeling approaches. Early work relied primarily on expert-curated rule-based and template-based systems, which explicitly encode chemical knowledge but often suffer from limited generalization to unseen reaction patterns and dependence on the coverage of manually designed templates (Corey & Wipke, 1969; Segler & Waller, 2017).

With the emergence of machine learning, data-driven methods have gained prominence. GNNs (Kipf, 2016; Hamilton et al., 2017; Xu et al., 2018) directly operate on molecular graphs and enable more effective extraction of structural chemical features (Dai et al., 2019; Somnath et al., 2021). Sequence-to-sequence models based on SMILES representations, particularly Transformer architectures, further advanced the field by framing retrosynthesis as a translation task from product to reactant strings (Schwaller et al., 2019; Irwin et al., 2022; Zheng et al., 2019; Wan et al., 2022). More recently, LLMs have been applied to single-step retrosynthesis, leveraging their strong generative capacity and contextual modeling ability (Zhang et al., 2024; Zhao et al., 2025b). However, such end-to-end models typically lack interpretability and remain prone to chemically invalid predictions.

Several recent studies introduce reaction centers or synthon-based intermediates to improve interpretability and chemical validity (Shi et al., 2020; Zhao et al., 2025a). These approaches generally rely on pipeline-style training and do not explicitly supervise the entire reasoning trajectory. In contrast, the proposed framework formulates retrosynthesis as a structured sequence of intermediate states and applies supervised learning directly to this decomposition, enabling explicit control over each stage of chemical reasoning.

3. Methods

3.1. Problem Formulation

Single-step retrosynthesis is formulated as a structured sequence generation problem. Given a target product molecule $\mathcal{P}$ represented as a SMILES string, the objective is to predict a set of reactant molecules $\mathcal{R} = \{R_1, R_2, \ldots, R_n\}$ that can synthesize $\mathcal{P}$ via a single chemical reaction. Unlike conventional approaches that directly learn a mapping $\mathcal{P} \rightarrow \mathcal{R}$, the proposed framework decomposes the prediction into a sequence of chemically interpretable intermediate states as illustrated in Figure 1, *i.e.*,

$$\mathcal{P} \xrightarrow{\text{Stage 1}} \mathcal{S}_{\text{site}} \xrightarrow{\text{Stage 2}} \mathcal{S}_{\text{synthon}} \xrightarrow{\text{Stage 3}} \mathcal{R}. \tag{1}$$

Here, $\mathcal{S}_{\text{site}}$ denotes the identified reaction sites, and $\mathcal{S}_{\text{synthon}}$ represents the intermediate synthons obtained after bond disconnection.

3.2. Structured Intermediate Process

Stage 1: Reaction Site Identification. The model first predicts the atoms and bonds involved in the retrosynthetic transformation. Given the molecular graph $G_{\mathcal{P}} = (V, E)$, where vertices V correspond to atoms and edges E correspond to bonds, the objective is to identify a set of reaction centers $\mathcal{C} \subseteq V$ and a set of active bonds $\mathcal{B} \subseteq E$. These elements are encoded as structured annotations within the SMILES sequence by marking the relevant atoms and bonds with special tokens.

Stage 2: Bond Disconnection. Conditioned on the predicted reaction sites $\mathcal{S}_{\text{site}} = (\mathcal{C}, \mathcal{B})$, the model performs bond cleavage to generate intermediate synthons, *i.e.*,

$$\mathcal{S}_{\text{synthon}} = \text{Disconnect}(G_{\mathcal{P}}, \mathcal{B}) = \{S_1, S_2, \ldots, S_k\}. \tag{2}$$

Each synthon corresponds to a molecular fragment that may exhibit unstable valence states, which are preserved using an extended SMILES representation.

Stage 3: Reactant Completion. In the final stage, each synthon is transformed into a chemically stable reactant by adding appropriate leaving groups, protecting groups, or functional moieties, *i.e.*,

$$\mathcal{R} = \{R_1, R_2, \ldots, R_n\} = \text{Complete}(\mathcal{S}_{\text{synthon}}). \tag{3}$$

This stage ensures that the predicted reactants are chemically valid and synthetically accessible.

3.3. Supervised Fine-Tuning

Model Architecture. A pre-trained large language model, Qwen2.5-3B, is adopted as the backbone. The model is fine-tuned to autoregressively generate the complete structured sequence encompassing all intermediate stages. In this setting, the probability that the model outputs a specific sequence $\mathcal{S}_{\text{site}}, \mathcal{S}_{\text{synthon}}, \mathcal{R}$ given the target product molecule $\mathcal{P}$ is

$$p_\theta(\mathcal{S}_{\text{site}}, \mathcal{S}_{\text{synthon}}, \mathcal{R} \mid \mathcal{P}) = \prod_{t=1}^{T} p_\theta(y_t \mid y_{<t}, \mathcal{P}), \tag{4}$$

where y_t denotes the t-th output token and θ represents the model parameters.

Training Objective. Supervised fine-tuning is performed on a curated dataset $\mathcal{D} = \{(\mathcal{P}_i, \mathcal{S}_i, \mathcal{R}_i)\}_{i=1}^{N}$ that includes explicit intermediate annotations. The objective minimizes the negative log-likelihood of the ground-truth sequences, *i.e.*,

$$\mathcal{L}_{\text{SFT}}(\theta) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \log p_\theta(y_t^{(i)} \mid y_{<t}^{(i)}, \mathcal{P}_i). \tag{5}$$

This formulation enables the model to learn chemically valid transformations at each stage of the reasoning process.

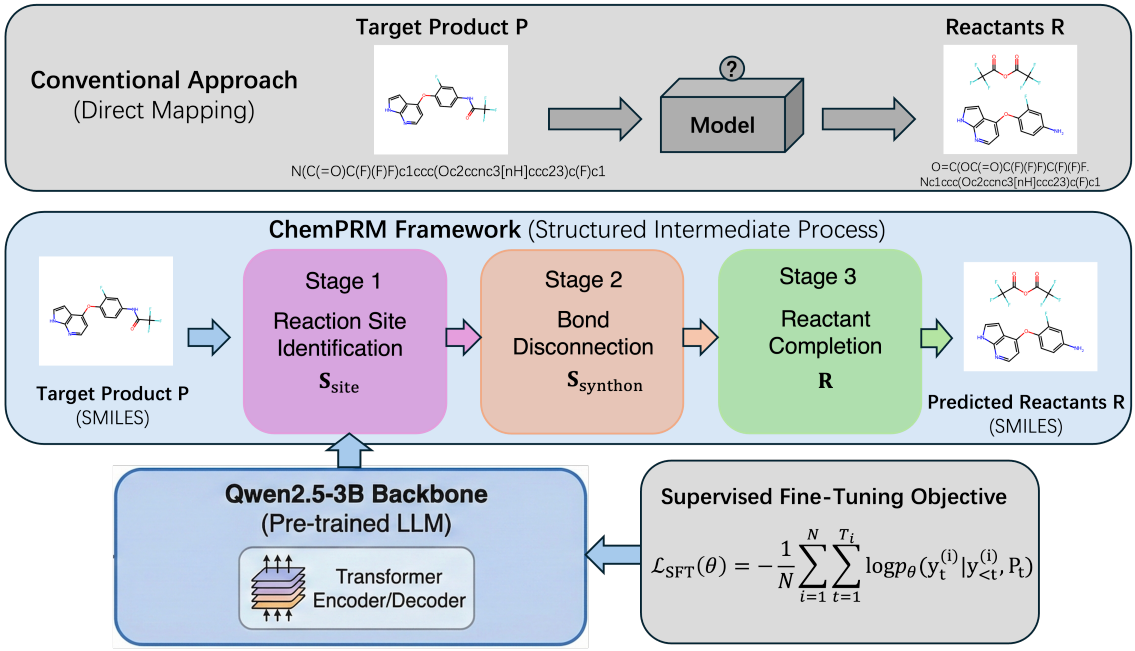

Figure 1. ChemPRM: Structured Framework for Interpretable Single-Step Retrosynthesis.

3.4. Training Details

We construct our training dataset by augmenting the USPTO-50k benchmark (Schneider et al., 2016) with explicit intermediate state annotations, as shown in Fig. 2. Each training sample is formatted as a SMILES sequence, *i.e.*, `<SMILES><Product>[PREDICT]<Reactants></SMILES>`, where the intermediate stages are embedded within the prediction sequence. The final dataset contains approximately 900K training examples after data augmentation and preprocessing following R-SMILES (Zhong et al., 2022).

We fine-tune the Qwen2.5-3B model using DeepSpeed with ZeRO stage-1 optimization on 8 NVIDIA L40 GPUs (48 GB memory each). The maximum sequence length is set to 8192 tokens, with a global batch size of 1024. Training is performed for 2 epochs using a learning rate of 1×10^{-5}. To support long-context training, we employ gradient checkpointing and FlashAttention for improved memory and computational efficiency. All experiments use mixed-precision training (bfloat16) and the AdamW optimizer with default hyperparameters.

4. Experiments

4.1. Main Results on USPTO-50k

Table 1 presents a comprehensive comparison between ChemPRM and representative retrosynthesis models on the USPTO-50k benchmark. The ChemPRM-3B model demonstrates consistently strong performance across all Top-k metrics, which provides empirical support for the effectiveness of the structured intermediate process formulation. Specifically, ChemPRM-3B attains a Top-1 accuracy of 58.5%, a Top-3 accuracy of 77.5%, a Top-5 accuracy of 83.2%, and a Top-10 accuracy of 86.9%.

When compared with template-based methods, ChemPRM-3B substantially outperforms early approaches such as RetroSim and NeuralSym, yielding absolute improvements of 21.2 and 14.1 percentage points in Top-1 accuracy, respectively. Although LocalRetro achieves a higher Top-10 accuracy through local template matching, the proposed framework attains competitive overall performance while offering greater modeling flexibility by learning representations rather than relying on predefined reaction templates.

Relative to graph-based methods, ChemPRM-3B surpasses several competitive baselines. Improvements of 9.6, 10.4, 4.8, and 4.4 percentage points in Top-1 accuracy are observed over G2Gs, MEGAN, GraphRetro, and G2Retro, respectively. Graph2Edits exhibits a slightly lower Top-1 accu-

Table 1. Performance comparison on USPTO-50k dataset. Top-k accuracy (%) is reported for single-step retrosynthesis prediction.

Model	Top-1	Top-3	Top-5	Top-10
Template-Based Methods				
RetroSim (Coley et al., 2017)	37.3	54.7	63.3	74.1
NeuralSym (Segler & Waller, 2017)	44.4	65.3	72.4	78.9
LocalRetro (Chen et al., 2020)	53.4	77.5	85.9	92.4
Graph-Based Methods				
G2Gs (Shi et al., 2020)	48.9	67.6	72.5	75.5
MEGAN (Sacha et al., 2021)	48.1	70.7	78.4	86.1
GraphRetro (Somnath et al., 2021)	53.7	68.3	72.2	75.5
G2Retro (Chen et al., 2023)	54.1	74.1	81.2	86.7
Graph2Edits (Zhong et al., 2023)	55.1	77.3	83.4	89.4
Sequence-Based Methods				
SCROP (Zheng et al., 2019)	43.7	60.0	65.2	68.7
Retroformer (Wan et al., 2022)	53.2	71.1	76.6	82.1
R-SMILES (Zhong et al., 2022)	56.3	79.2	86.2	91.0
EditRetro (Han et al., 2024)	60.8	80.6	86.0	90.3
ChemPRM (3B)	58.5	77.5	83.2	86.9

racy but achieves a higher Top-10 accuracy than ChemPRM, which suggests that Graph2Edits may generate a broader set of high-ranked candidates. In contrast, ChemPRM emphasizes interpretable predictions that are grounded in explicitly modeled intermediate states.

Among sequence-based methods, ChemPRM-3B also demonstrates strong competitiveness. The model outperforms SCROP, Retroformer, and R-SMILES by 14.8, 5.3, and 2.2 percentage points in Top-1 accuracy, respectively. EditRetro achieves a higher Top-1 accuracy, exceeding ChemPRM-3B by 2.3 percentage points; however, the performance gap narrows at Top-3 and Top-5, indicating that ChemPRM generates high-quality alternative predictions beyond the top-ranked result. Moreover, explicit modeling of intermediate states, from reaction sites to synthons and subsequently to reactants, yields substantially improved interpretability compared with direct product-to-reactant mappings. These results support the hypothesis that decomposing retrosynthesis prediction into structured intermediate stages enables accurate predictions while providing actionable chemical insights and reducing reliance on purely black-box inference.

As illustrated in Fig. 2, ChemPRM performs retrosynthesis through a sequence of explicit intermediate steps, including reaction center identification, synthon generation, and reactant reconstruction. Notably, although the predicted synthons in Fig. 2 differ from the ground-truth synthons in their SMILES representations, they are chemically equivalent in terms of molecular structure and functional groups. This discrepancy reflects the non-uniqueness of SMILES rather than an error in chemical reasoning. The model's ability to produce structurally correct synthons despite such representational differences indicates that ChemPRM is not merely optimizing for character-level prediction. Instead, it demonstrates an understanding of molecular structure and reaction logic, capturing chemically meaningful fragments that support accurate downstream reactant prediction.

Table 2. Ablation study on model scaling. Top-k accuracy (%) for different model sizes.

Model Variant	Top-1	Top-3	Top-5	Top-10
ChemPRM-1.5B	54.9	75.9	82.0	86.2
ChemPRM-3B	58.5	77.5	**83.2**	**86.9**
ChemPRM-7B	**58.6**	**77.6**	82.0	85.0

4.2. Model Scaling Analysis

Table 2 reports an ablation study that examines the impact of model size on retrosynthesis performance. Three variants, ChemPRM-1.5B, ChemPRM-3B, and ChemPRM-7B, are evaluated while keeping the architecture and training procedure identical across all settings.

Increasing the number of parameters from 1.5B to 3B yields consistent improvements across most evaluation metrics. Top-1 accuracy increases by 3.6% (54.9% $\rightarrow$ 58.5%), indicating that a larger model capacity improves the ability of the model to capture complex patterns in retrosynthetic transformations. Performance gains are also observed for Top-3 (+1.6%), Top-5 (+1.2%), and Top-10 (+0.7%), although the magnitude of improvement decreases as k increases.

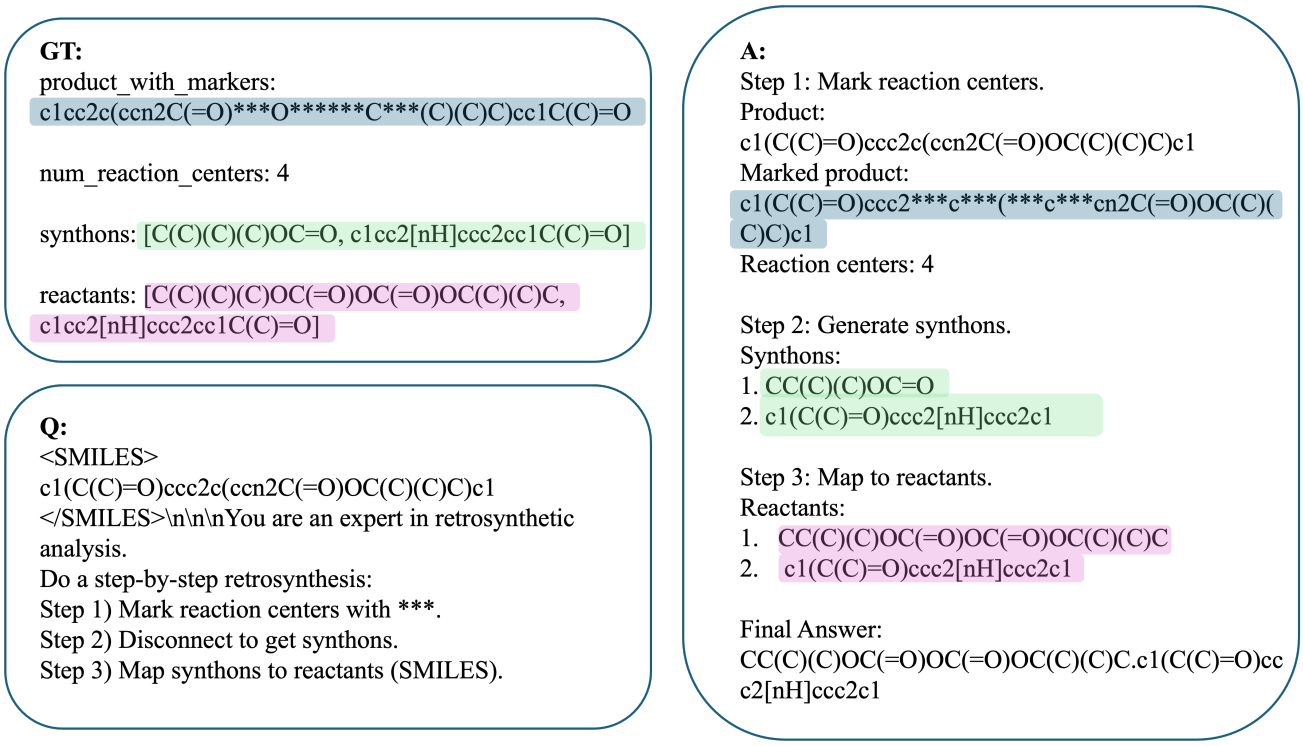

Figure 2. Example of step-by-step retrosynthesis predicted by ChemPRM.

In contrast, further scaling from 3B to 7B parameters leads to mixed outcomes. Top-1 accuracy shows a marginal increase of 0.1% to 58.6%, and Top-3 accuracy remains nearly unchanged with a 0.1% increase to 77.6%. However, Top-5 and Top-10 accuracy decrease by 1.2% and 1.9%, reaching 82.0% and 85.0%, respectively. This non-monotonic behavior with respect to model size suggests the presence of overfitting or optimization difficulties at larger scales under the current training regime. Given the dataset size of approximately 900K examples, the 7B model may require more diverse data, longer training schedules, or stronger regularization to effectively exploit its increased capacity.

5. Limitations and Future Directions

While ChemPRM demonstrates strong performance, several limitations warrant discussion. First, the staged approach introduces additional sequence length, increasing computational costs during inference. Second, the model currently relies on ground-truth intermediate annotations during training, which require expert labeling or algorithmic extraction. Third, the non-monotonic scaling behavior beyond 3B parameters suggests that larger models may require substantially more training data or improved optimization techniques. Future work will explore: (1) semi-supervised learning to leverage unlabeled reaction data, (2) reinforcement learning to refine intermediate state generation without explicit supervision, (3) incorporating chemical constraints as hard rules to guarantee validity, and (4) extending the framework to multi-step retrosynthesis planning for complete synthetic route design.

6. Discussion

In summary, ChemPRM presents a preliminary investigation into incorporating structured chemical reasoning within a large language model framework. While our current results do not yet establish a new state-of-the-art—partially due to the simplified training regime and the modest scale of the experimental setup—the empirical evidence clearly demonstrates that explicitly supervising intermediate reasoning states (reaction sites and synthons) significantly enhances the model's predictive accuracy and interpretability. This work serves as a proof-of-concept for the "process-oriented" retrosynthesis paradigm. By mimicking the stepwise logic of human chemists, we bridge the gap between black-box sequence mapping and formal chemical theory. These initial findings suggest that the integration of intermediate process rewards and structured annotations is a promising direction. Future exploration could involve scaling the dataset, refining the granularity of intermediate states, or employing reinforcement learning to further optimize the reasoning trajectories beyond supervised labels

References

Chen, X., Dai, H., Li, Y., Gao, X., and Song, L. Learning to stop while learning to predict. In *International Conference on Machine Learning*, pp. 1520–1530. PMLR, 2020.

Chen, Z., Ayinde, O. R., Fuchs, J. R., Sun, H., and Ning, X. G 2 retro as a two-step graph generative models for retrosynthesis prediction. *Communications Chemistry*, 6 (1):102, 2023.

Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H., and Jensen, K. F. Prediction of organic reaction outcomes using machine learning. *ACS central science*, 3(5):434–443, 2017.

Corey, E. J. and Wipke, W. T. Computer-assisted design of complex organic syntheses: Pathways for molecular synthesis can be devised with a computer and equipment for graphical communication. *Science*, 166(3902):178–192, 1969.

Dai, H., Li, C., Coley, C., Dai, B., and Song, L. Retrosynthesis prediction with conditional graph logic network. *Advances in Neural Information Processing Systems*, 32, 2019.

Hamilton, W., Ying, Z., and Leskovec, J. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Han, Y., Xu, X., Hsieh, C.-Y., Ding, K., Xu, H., Xu, R., Hou, T., Zhang, Q., and Chen, H. Retrosynthesis prediction with an iterative string editing model. *Nature Communications*, 15(1):6404, 2024.

Irwin, R., Dimitriadis, S., He, J., and Bjerrum, E. J. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology*, 3(1):015022, 2022.

Kipf, T. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Reed, S. M. Augmented and programmatically optimized llm prompts reduce chemical hallucinations. *Journal of Chemical Information and Modeling*, 65(9):4274–4280, 2025.

Sacha, M., Błaz, M., Byrski, P., Dabrowski-Tumanski, P., Chrominski, M., Loska, R., Włodarczyk-Pruszynski, P., and Jastrzebski, S. Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *Journal of Chemical Information and Modeling*, 61(7): 3273–3284, 2021.

Schneider, N., Stiefl, N., and Landrum, G. A. What's what: The (nearly) definitive guide to reaction role assignment. *Journal of chemical information and modeling*, 56(12): 2336–2346, 2016.

Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., and Lee, A. A. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science*, 5(9):1572–1583, 2019.

Segler, M. H. and Waller, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry–A European Journal*, 23(25):5966–5971, 2017.

Shi, C., Xu, M., Guo, H., Zhang, M., and Tang, J. A graph to graphs framework for retrosynthesis prediction. In *International conference on machine learning*, pp. 8818–8827. PMLR, 2020.

Somnath, V. R., Bunne, C., Coley, C., Krause, A., and Barzilay, R. Learning graph models for retrosynthesis prediction. *Advances in Neural Information Processing Systems*, 34:9405–9415, 2021.

Wan, Y., Hsieh, C.-Y., Liao, B., and Zhang, S. Retroformer: Pushing the limits of end-to-end retrosynthesis transformer. In *International Conference on Machine Learning*, pp. 22475–22490. PMLR, 2022.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Zhang, D., Liu, W., Tan, Q., Chen, J., Yan, H., Yan, Y., Li, J., Huang, W., Yue, X., Ouyang, W., et al. Chemllm: A chemical large language model. *arXiv preprint arXiv:2402.06852*, 2024.

Zhao, P.-C., Wei, X.-X., Wang, Q., Wang, Q.-H., Li, J.-N., Shang, J., Lu, C., and Shi, J.-Y. Single-step retrosynthesis prediction via multitask graph representation learning. *Nature Communications*, 16(1):814, 2025a.

Zhao, Z., Ma, D., Chen, L., Sun, L., Li, Z., Xia, Y., Chen, B., Xu, H., Zhu, Z., Zhu, S., et al. Developing chemdfm as a large language foundation model for chemistry. *Cell Reports Physical Science*, 6(4), 2025b.

Zheng, S., Rao, J., Zhang, Z., Xu, J., and Yang, Y. Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of chemical information and modeling*, 60(1):47–55, 2019.

Zhong, W., Yang, Z., and Chen, C. Y.-C. Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing. *Nature Communications*, 14 (1):3009, 2023.

Zhong, Z., Song, J., Feng, Z., Liu, T., Jia, L., Yao, S., Wu, M., Hou, T., and Song, M. Root-aligned smiles: a tight representation for chemical reaction prediction. *Chemical Science*, 13(31):9023–9034, 2022.