MotifAgent: Learning Molecular Assembly through Multi-Agent Collaboration for Chemical Language Understanding

Jinjia Feng and Wenda Wang and Zhewei Wei*
Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
{jinjia_feng, wangwenda87, zhewei}@ruc.edu.cn

Abstract

Large Language Models (LLMs) have shown great potential in molecular understanding by aligning molecular representations with text. However, existing approaches remain limited to static motif recognition without comprehending the generative principles—the connection rules governing how motifs assemble into valid topological structures. To address this challenge, we introduce **MotifAgent**, a multi-agent reinforcement learning framework inspired by emergent collective intelligence. We formulate molecular assembly as a collaborative problem where each motif is represented by an agent sharing a common LLM backbone, learning connection rules through explicit inter-motif negotiation rather than implicit sequence memorization. Key innovations include: (1) dynamic inter-agent negotiation for modeling motif connections; (2) Set-based Behavioral Cloning for learning multiple topologically equivalent assembly paths; (3) topology-aware reward shaping with MAPPO to maintain chemical validity while optimizing target properties. Extensive experiments demonstrate that MotifAgent achieves state-of-the-art performance across molecular property prediction, description generation, and reaction prediction tasks, with our generalist model surpassing specialized expert models.

1 Introduction

The computational representation and understanding of molecules represent a core challenge in modern drug discovery (Berdigaliyev and Aljofan, 2020) and materials design (Wang et al., 2019). With the remarkable success of large language models (LLMs) (Radford et al., 2018; Koroteev, 2021) in natural language processing, their application to molecular understanding and generation tasks (Bagal et al., 2021; Mazuz et al., 2023) has emerged as a prominent research direction. Current mainstream approaches represent molecules through SMILES strings (Mswahili and Jeong, 2024) and leverage Transformer architectures (Vaswani et al., 2017) to learn cross-modal alignment between molecular and textual representations (Zhao et al., 2023b; Song et al., 2024). These methods (Edwards et al., 2022; Zhang et al., 2024) have achieved significant progress in molecular property prediction, drug-target interaction modeling, and molecular description generation.

However, current LLM-based methods suffer from a fundamental limitation: they cannot comprehend the **generative principles** underlying molecular formation—specifically, the connection rules governing how motifs assemble into valid topological structures, including which sites can form bonds, what bond types are permissible, and how connections satisfy chemical constraints (Zhang et al., 2023; Geng et al., 2023). SMILES encodes topology through paired brackets and nested indices (Krenn et al., 2020), which cheminformatics tools can parse via hard-coded syntactic rules. However, this *rule-based extractability* does not imply *model-level understanding*: LLMs process SMILES as character sequences without access to explicit parsing rules, and must instead implicitly learn the complex mappings between linear notation and 2D molecular topology from data alone—a task that conflates sequential pattern matching with genuine structural comprehension (Wigh et al., 2022; Bilodeau et al., 2022).

More critically, existing approaches remain limited to static motif recognition while overlooking the dynamic connection rules that govern molecular formation (Jin et al., 2020; Bettens and Lee, 2006; Collins and Bettens, 2015). These patterns directly determine molecular properties (Zhang et al., 2021)—hydroxyl groups at ortho, meta, or para positions exhibit different biological activities, while aromatic rings connected through different

* Corresponding author.

Findings of the Association for Computational Linguistics: ACL 2026, pages 40700–40739

linkers affect molecular flexibility and target binding. Treating molecules as atomic sequences or substructure collections (Zhang et al., 2024; Luo et al., 2023a; Zhao et al., 2023a) prevents models from understanding motif interactions or generating molecules with desired properties. The fundamental issue is that a single LLM processes SMILES as character sequences, lacking the capacity to simultaneously track multiple motifs' connection states, evaluate chemical compatibility in parallel, and reason about global topological constraints—capabilities essential for understanding molecular assembly as a dynamic process rather than static pattern memorization.

To address these challenges, we draw inspiration from multi-agent systems, where collective intelligence emerges from distributed agents operating on local knowledge while their interactions produce coherent global behavior (Tran et al., 2025; Qian et al., 2024). We propose **MotifAgent**, a **Multi-Agent Collaborative** framework that **learns molecular assembly** through multi-agent reinforcement learning, enabling chemical language understanding via dynamic topology reconstruction. Each motif is represented by an agent sharing a common LLM backbone, proposing context-aware connections by considering chemical principles, local and global topological state. Through Centralized Training with Decentralized Execution (CTDE) (Lowe et al., 2017a), valid molecular topology emerges from negotiation, enabling the shared LLM to internalize connection rules through explicit inter-motif reasoning. We further introduce Set-based Behavioral Cloning (Set-BC) to learn multiple equivalent assembly pathways, while multi-level rewards guide agents to satisfy chemical validity and optimize target properties.

Our contributions are summarized as follows:

- To the best of our knowledge, MotifAgent is the first multi-agent framework that dynamically models molecular generative principles, moving beyond static pattern recognition to understand connection rules through collaborative agent negotiation with emergent collective intelligence.

- MotifAgent achieves comprehensive leading performance across molecular property prediction, molecular description generation, and chemical reaction prediction. Remarkably, using general-purpose LLMs as backbone, MotifAgent achieves or surpasses specialized expert models.

- MotifAgent provides new insights: (1) Multi-agent collaboration captures the hierarchical assembly nature of molecules while enabling controllable generation through explicit connection modeling. (2) Learned connection rules exhibit strong generalization and chemical validity. (3) Interpretable reasoning traces reveal how motif combinations produce specific properties.

2 Related Works

Molecular Generation and Understanding: Traditional molecular generation and optimization methods include VAE-based (Jin et al., 2018, 2020), autoregressive (Shi et al., 2020; Maziarz et al., 2021), and diffusion-based approaches (Xu et al., 2022). To enable LLMs (Radford et al., 2018; Raffel et al., 2023; Touvron et al., 2023) to process molecules, prior works (Edwards et al., 2022; Christofidellis et al., 2023; Liu et al., 2023b; Li et al., 2024; Zhang et al., 2024) jointly train on SMILES and text for bidirectional molecule-text conversion, with molT5 (Edwards et al., 2022) pioneering self-supervised SMILES-to-text translation. Other approaches (Su et al., 2022; Liu et al., 2023a; Luo et al., 2023a; Liu et al., 2023c; Zhao et al., 2023a) incorporate 2D graphs via multimodal contrastive learning. Recent multimodal LLMs extend to molecular images: ChemVLM (Li et al., 2025) combines vision transformers with chemistry-specialized LLMs, while ChemMLLM (Tan et al., 2025) enables bidirectional image-text generation. Our method advances this line by introducing multiple LLM agents to model motif connectivity, enabling molecular understanding through substructure assembly.

Multi-Agent Reinforcement Learning: MARL (Canese et al., 2021; Wen et al., 2022; Albrecht et al., 2024) coordinates multiple autonomous agents in shared environments, offering enhanced scalability through role-specific agents (Gao et al., 2025)—particularly valuable for molecular design where specialized agents can focus on different motifs. The CTDE framework (Lowe et al., 2017b; Sunehag et al., 2017; Rashid et al., 2020) addresses policy non-stationarity and partial observability by leveraging global information during training while maintaining decentralized execution. Methods like MAPPO (Yu et al., 2022) use shared policy networks with local observations. Our method employs CTDE with a shared LLM backbone, enabling motif agents to learn global topology during

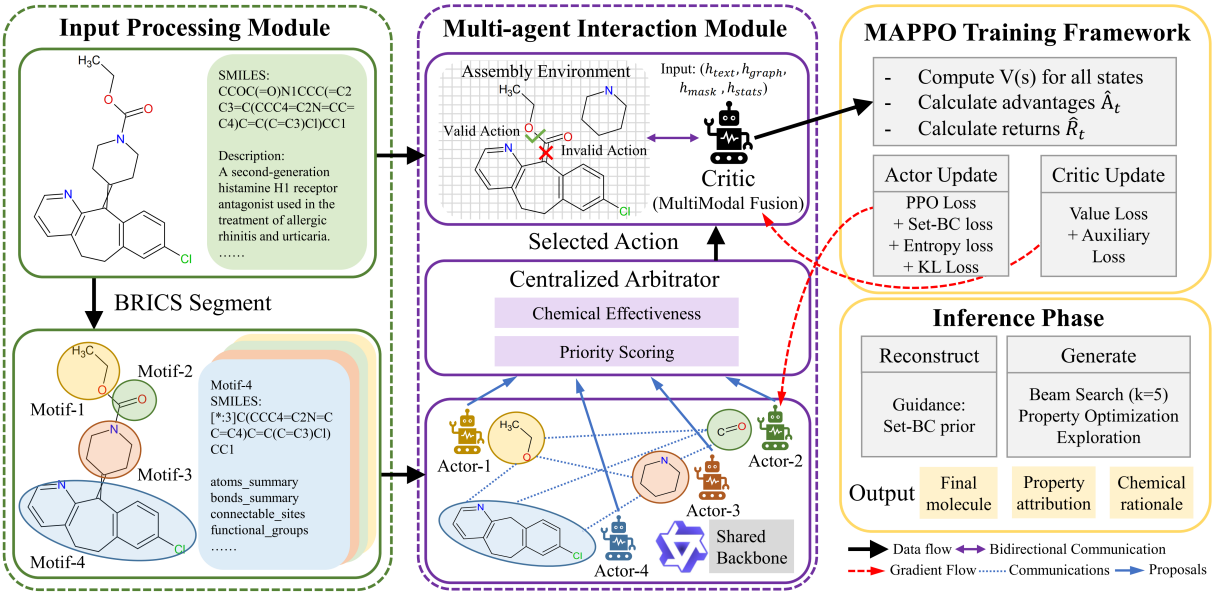


Figure 1: Overview of the proposed MotifAgent framework. MotifAgent consists of four integrated components: (1) Input Processing Module decomposes molecules into motifs using BRICS segmentation and converts them to structured text representations. (2) Multi-agent Interaction Module employs LLM-based actors (one per motif) sharing a common backbone to propose connections, with a Centralized Arbitrator selecting valid actions and a Critic evaluating assembly states through multi-modal fusion. (3) MAPPO Training Framework jointly optimizes Actor and Critic networks with separate loss functions. (4) Inference Phase supports both reconstruction mode with Set-BC guidance and generation mode for property optimization.

training while making decentralized connection decisions during execution.

3 Method

We formulate fragment-based molecular assembly as a centralized training with decentralized execution (CTDE) multi-agent reinforcement learning problem. The core idea is to employ a shared large language model (LLM) as the decentralized policy backbone, enabling each motif agent to propose connections under chemical constraints while dynamically reconstructing the molecule's 2D topology through global communication. A centralized critic receives global graph information during training to evaluate state values and assembly progress. We adopt Multi-Agent Proximal Policy Optimization (MAPPO) for policy updates, combined with Set-based Behavior Cloning (Set-BC) to handle assembly order ambiguity and potential shaping to accelerate target graph alignment.

3.1 Representation and Fragmentation

We decompose molecules into chemically meaningful bricks and linkers using the BRICS (Degen et al., 2008) algorithm with 16 bond-breaking rules, preserving complete metadata including connection sites and allowed bond types. Each motif is serialized into structured text containing its identifier, SMILES string, connectable sites (with chemical environment and allowed bond types), and property summaries (aromaticity, ring structures, functional groups). Connections follow a unified template: `CONNECTION: motif_i[site_x] -bond_type-> motif_j[site_y]`, directly encoding inter-motif adjacency relationships.

3.2 Environment Modeling

We model the assembly process as a Dec-POMDP where the state space corresponds to the molecule's 2D topological evolution. The global state at time t comprises the current assembly graph $G_t = (V, E_t)$'s textual summary, unconnected motif list, and available site topology, where V is the motif node set and E_t is the established connections. Actions are defined as $a_t = (i, s_i, j, s_j, b)$: connecting site s_i of motif i to site s_j of motif j with bond type b, plus an explicit `STOP` action for termination.

Precise termination conditions: In reconstruction mode, necessary conditions for termination are $\text{cc}(G_t) = 1$ and $E_t \supseteq E^*$, where $\text{cc}(\cdot)$ denotes the number of connected components and E^* is the target molecule's edge set. Sufficient conditions are $E_t = E^*$ or the policy selecting `STOP`, with timeout protection (steps $> 2|E^*|$) to prevent infinite loops. In generation mode, termination conditions include: (1) chemical completeness—all required valences

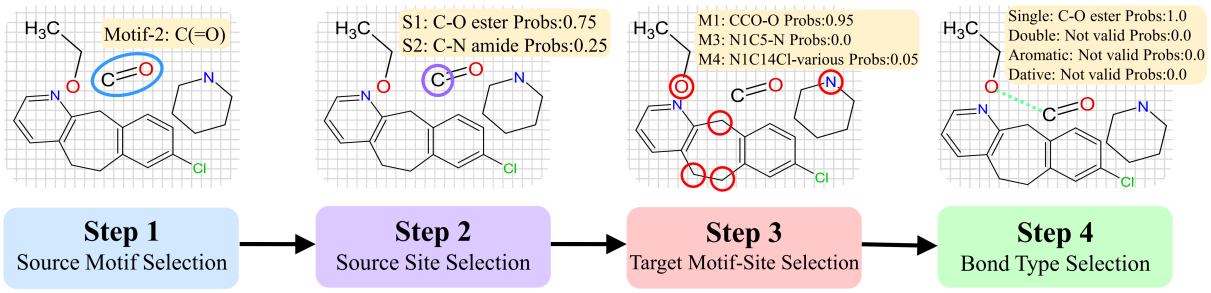


Figure 2: Hierarchical Sampling Process of MotifAgent.

saturated; (2) topological completeness—forming a connected molecular graph; (3) property convergence—improvement over k consecutive steps $< \varepsilon$, where k is the window size and ε is the convergence threshold; (4) active termination—policy outputting `STOP` probability > 0.9.

3.3 Agents and Critic

Each motif corresponds to an agent, with all agents sharing the same LLM as the policy backbone, generating connection proposals conditioned on their motif descriptions and the current global assembly's topological summary. Each agent not only knows its local structure but also perceives the molecule's 2D topological state.

LLM policy network models connection decisions: To reduce action branching, we employ hierarchical action sampling. As shown in Figure 2, during sampling, the LLM generates action distributions through specific prompt templates. The first layer evaluates each available motif's connection potential within the current topology. The second layer, conditioned on the selected motif, evaluates each site's chemical activity and topological accessibility. The third layer evaluates target compatibility based on the source's chemical environment and global topology. The fourth layer determines optimal bond type based on both ends' chemical environments. The implementation uses the LLM's last hidden layer features to generate logits through trainable action heads:

$$\pi(a|s) = \text{Softmax}(\text{MLP}(\text{LLM}_{\text{hidden}}(\text{prompt}(s)))) \quad (1)$$

where π is the policy function, a is the action, s is the state, $\text{LLM}_{\text{hidden}}$ denotes the LLM's hidden representation, and $\text{prompt}(s)$ contains the description of the current 2D topology.

Central arbitrator coordinates topological construction: The arbitrator employs two-phase coordination. Phase 1 performs validity screening: collecting all motif agents' proposals in parallel, filtering out proposals violating valence rules or destroying topological integrity through a rule engine. Phase 2 conducts priority scoring and selection:

$$\begin{aligned} S(a) = w_1 \cdot \text{ChemStability}(a) \\ + w_2 \cdot \text{TopoProgress}(a) \\ + w_3 \cdot \text{PropImprove}(a) \end{aligned} \quad (2)$$

where $S(a)$ is action a's score, w_1, w_2, w_3 are weight coefficients. Details of priority scoring are provided in Appendix H. Note that the arbitrator acts as a *safety filter* (i.e., action masking) rather than the primary decision maker: the policy itself is learned entirely inside each motif agent's network, and the arbitrator's role is to prune chemically infeasible proposals early and to break ties among valid ones. This design choice and its ablation are analyzed in detail in Appendix L.1.

Fused representation perceives global topology: The centralized critic $V_\phi(x_t)$ integrates topological information through multi-modal attention:

$$x_t = \text{MultiModalFusion}([h_{\text{text}}, h_{\text{graph}}, h_{\text{mask}}, h_{\text{topo}}]) \quad (3)$$

where x_t is the fused representation at time t, h_{text} is the LLM-encoded global state text representation, h_{graph} is the assembly graph structure, h_{mask} is the available action mask based on topological constraints, and h_{topo} encodes topological statistics. We attach auxiliary regression heads to predict remaining target edges and connected components, enhancing perception of assembly progress.

3.4 Rewards and Shaping

The reward design explicitly considers 2D topology reconstruction and optimization, comprising two complementary components: **chemical base rewards** R_{chem} that ensure molecular validity and desired properties, and **topological shaping rewards** R_{topo} that guide correct 2D structure construction. The combined single-step reward is:

$$R = R_{\text{chem}} + R_{\text{topo}} \quad (4)$$

Chemical base rewards evaluate the quality of assembled molecules from multiple chemical perspectives, including validity verification against

fundamental chemical constraints, successful formation of target functional groups and so on. These criteria collectively ensure that assembled molecules are chemically meaningful and possess desired characteristics.

Topological shaping rewards guide the assembly process toward correct 2D topology by encouraging actions that reduce molecular fragmentation and penalizing deviation from the target graph structure. We additionally apply potential-based reward shaping to accelerate learning while preserving optimal policy invariance. Details of all reward components are provided in Appendix I.

3.5 Policy Optimization and Training

We employ MAPPO (Yu et al., 2022) for policy updates, crucially enabling the policy to learn motif connection rules rather than memorizing specific sequences. During training, the Actor (policy network) and Critic are optimized separately using different loss functions.

The policy network is optimized through:

$$\begin{aligned}\mathcal{L}_{\text{actor}} = & -\mathcal{L}_{\text{clip}} - \beta\mathcal{H}(\pi_\theta) + \alpha_{\text{BC}}\mathcal{L}_{\text{BC}} \\ & + \mathcal{L}_{\text{KL}} - \sum_k \lambda_k \mathbb{E}[c_k(s)]\end{aligned} \tag{5}$$

where $\mathcal{L}_{\text{clip}}$ is the PPO clipped objective optimizing policy to maximize expected reward, $\beta\mathcal{H}(\pi_\theta)$ provides entropy regularization for exploration, $\alpha_{\text{BC}}\mathcal{L}_{\text{BC}}$ is the Set-BC supervision term learning correct assembly patterns, $\mathcal{L}_{\text{KL}}$ constrains policy change relative to reference policy, and $\sum_k \lambda_k \mathbb{E}[c_k(s)]$ represents constraint terms satisfying chemical and topological requirements.

The PPO clipped objective takes the form:

$$\begin{aligned}\mathcal{L}_{\text{clip}}(\theta) = \mathbb{E}_{i,t}\Big[& \min\left(r_t^i(\theta)\hat{A}_t,\right. \\ & \left.\text{clip}(r_t^i(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\Big]\end{aligned} \tag{6}$$

where $r_t^i(\theta)$ is the importance sampling ratio, and $\hat{A}_t$ is the advantage estimate computed using Generalized Advantage Estimation (GAE).

The critic network is optimized with:

$$\begin{aligned}\mathcal{L}_{\text{critic}} = & \mathbb{E}_t[(V_\phi(x_t) - \hat{R}_t)^2] + w_1(V_{\text{edges}}(x_t) \\ & - |E^* \setminus E_t|)^2 + w_2(V_{\text{cc}}(x_t) - \text{cc}(G_t))^2\end{aligned} \tag{7}$$

The main value loss $(V_\phi(x_t) - \hat{R}_t)^2$ uses Monte Carlo return estimates, while auxiliary heads V_{edges} and V_{cc} predict remaining target edges and connected components respectively.

Set-BC learns topologically equivalent paths: To address multiple assembly sequences reaching the same topology, Set-BC avoids enforcing specific orders by maximizing policy probability over the entire correct action set $\mathcal{A}_t^*$:

$$\mathcal{L}_{\text{BC}} = -\mathbb{E}_t\left[\log \sum_{a \in \mathcal{A}_t^*} \pi_\theta(a|s_t)\right] \tag{8}$$

where $\mathcal{A}_t^*$ contains all actions preserving topological correctness at state s_t, $\pi_\theta(a|s_t)$ is the policy probability for action a, and t is the time steps.

Curriculum learning progressively masters topological construction: As shown in Figure 3, training proceeds through four phases, transitioning from simple topological constraints to complex optimization. Phase 1 (first 25%) focuses on strict topological reconstruction with hard masks and Set-BC weight $\alpha_{\text{BC}} = 1.0$. Phase 2 (25%-50%) introduces soft-constraint learning with α_{BC} decaying to 0.1. Phase 3 (50%-75%) combines topology-aware property optimization. Phase 4 (final 25%) enables free exploration by removing hard constraints.

3.6 Inference and Downstream Applications

During inference, the policy assembles motifs based on learned rules, with hierarchical masks ensuring topological validity. Reconstruction tasks succeed when the target topology is fully recovered ($E = E^*$ and $\text{cc}(G) = 1$). Generation tasks optimize properties within topological constraints after forming a connected topology, exploring multiple paths via beam search (beam width $k = 5$).

To improve efficiency, we employ: (1) topological pruning—early elimination of infeasible paths based on learned patterns; (2) topology-guided sampling—prioritizing actions that quickly form stable topologies; (3) topological checkpoints—saving key states for backtracking and branching.

For downstream task adaptation, we adopt a two-stage training paradigm: initial training establishes molecular reconstruction capabilities, followed by task-specific fine-tuning. Due to computational constraints with large LLM backbones, we employ parameter-efficient strategies. Specifically, molecular property prediction freezes the multi-agent modules and trains only topology-aware classification heads. Molecular description generation and chemical reaction prediction employ LoRA fine-tuning combined with task-specific output heads, where the multi-agent assembly process first reconstructs molecular topology before generating task outputs. All downstream tasks share the unified prompt template that preserves the multi-agent

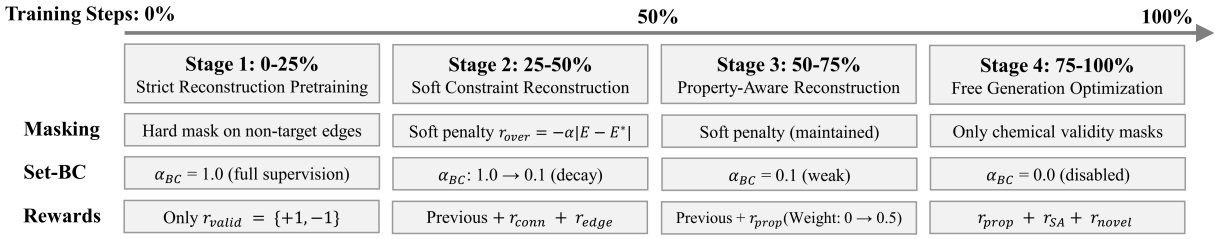


Figure 3: Curriculum Learning Training for MotifAgent.

interaction format while appending task-specific sections. Detailed prompt templates and training configurations are provided in Appendix J.

4 Experiments

4.1 Initial Training

Datasets and Training Details: We utilize the same molecular SMILES-text pairs dataset as MoleculeSTM (Liu et al., 2023a), collected from the PubChem website. Following their preprocessing pipeline, pairs with identical PubChem IDs and descriptions shorter than 18 characters are merged, with duplicates removed from downstream task datasets to prevent data leakage. This yields 51,340 unique high-quality pairs for initial training. For each molecule, we apply our improved BRICS fragmentation to generate motif sets, with molecules containing 2-15 motifs selected for training. The model employs Qwen2.5-7B as the shared policy backbone and MolT5-base as the centralized critic. All other training configurations and hyperparameters are detailed in Appendix D. Evaluation metrics are detailed in Appendix E; beyond topology-level metrics (Validity, GED, Morgan FTS, Conn. Site, Bond Type), the scalability analysis in Section 4.6 additionally reports *QED* (a 0–1 drug-likeness score) and *FG-PR* (functional-group preservation rate via RDKit fragment-descriptor recall).

4.2 Molecular Description Generation

To evaluate MotifAgent's performance on molecular description generation, we adopt the widely-used ChEBI-20 benchmark dataset (Edwards et al., 2021), which requires generating natural language descriptions of molecular properties given structures. We employ metrics following standard protocols. Baselines include specialist models (MoT5 (Edwards et al., 2022), MoMu (Su et al., 2022), MolFM (Luo et al., 2023a), MolXPT (Liu et al., 2023b), GIT-Mol (Liu et al., 2024), MolCA (Liu et al., 2023c), Text+Chem T5 (Christofidellis et al., 2023), Atomas (Zhang et al., 2024)), retrieval-based MolReGPT (Li et al., 2024), and LLM-based generalist models (BioMedGPT (Luo et al., 2023b), InstructMol (Cao et al., 2023), Mol-Instruction (Fang et al., 2023), HIGHT (Chen et al., 2025)).

Table 1: Performance comparison on molecular description generation task. The top **1st** and 2nd results are highlighted. **All metrics: higher is better.**

Method	BLEU-2	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L	METEOR
Specialist Models						
MoT5-base	0.540	0.457	0.634	0.485	0.568	0.569
MoMu	0.549	0.462	-	-	-	0.576
MolFM	0.585	0.498	0.653	0.508	0.594	0.607
MolXPT	0.594	0.505	0.660	0.511	0.597	0.626
GIT-Mol	0.352	0.263	0.575	0.485	0.560	0.430
MolCA	0.620	0.531	0.681	0.537	0.618	-
Text+Chem T5	0.625	0.542	0.682	0.543	0.622	0.648
Atomas-base	0.632	**0.549**	0.685	0.545	0.626	-
Retrieval Based LLMs						
MolReGPT	0.607	0.525	0.634	0.476	0.562	0.610
LLM Based Generalist Models						
BioMedGPT-10B	0.234	0.141	0.386	0.206	0.332	0.308
InstructMol-GS	0.453	0.349	0.546	0.372	0.482	0.483
Mol-Instruction	0.249	0.171	0.331	0.203	0.289	0.271
HIGHT-GS	0.498	0.397	0.582	0.414	0.518	0.525
Qwen2.5-3b	0.516	0.392	0.591	0.405	0.521	0.540
Qwen2.5-7b	0.544	0.431	0.610	0.446	0.543	0.571
MotifAgent-small	0.571	0.474	0.635	0.481	0.570	0.602
MotifAgent	**0.642**	0.545	**0.686**	**0.557**	**0.633**	**0.651**

Table 2: Performance comparison on retrosynthesis prediction tasks. **All metrics are higher-is-better, except Levenshtein, where lower is better.**

Method	Exact	BLEU	Levenshtein	RDK FTS	MACCS FTS	Morgan FTS	Validity
Alpaca	0.000	0.063	46.915	0.005	0.023	0.007	0.160
Baize	0.000	0.095	44.714	0.025	0.050	0.023	0.112
ChatGLM	0.000	0.117	48.365	0.056	0.075	0.043	0.046
LLaMA	0.000	0.036	46.844	0.018	0.029	0.017	0.010
Vicuna	0.000	0.057	46.877	0.025	0.030	0.021	0.017
Mol-Instruction	0.009	0.705	31.227	0.283	0.487	0.230	**1.000**
LLaMA-7b (LoRA)	0.000	0.283	53.510	0.136	0.294	0.106	**1.000**
InstructMol-GS	0.172	0.911	20.300	0.765	0.615	0.568	**1.000**
HIGHT-GS	0.202	0.914	20.194	0.772	0.623	0.577	0.999
MotifAgent	**0.275**	**0.932**	**18.810**	**0.783**	**0.685**	**0.631**	**1.000**

Experimental results in Table 1 demonstrate that MotifAgent achieves state-of-the-art performance among LLM-based generalist models, substantially outperforming existing general-purpose methods across all metrics. Compared to HIGHT-GS, MotifAgent delivers an average performance improvement of 22.5%. More notably, MotifAgent exhibits strong competitiveness against specialist models specifically designed for molecule-text tasks, surpassing all specialist baselines on the majority of metrics. To isolate framework contributions from backbone capacity, we compare MotifAgent variants against their backbones: MotifAgent yields consistent improvements over Qwen2.5-7b across all metrics, while MotifAgent-small (based on

Table 3: Performance comparison on molecular classification tasks (ROC-AUC %). We repeat MotifAgent 3 times and report the average with a 95% confidence interval. The top **1st** and 2nd results are highlighted.

Method	BBBP	Tox21	ToxCast	Sider	ClinTox	MUV	HIV	Bace	Avg
Specialist Models									
MoleculeSTM-SMILES (Liu et al., 2023a)	70.75±1.9	75.7±0.9	65.3±0.37	63.7±0.81	86.6±2.28	65.7±1.46	77.0±0.4	81.9±0.4	73.33
MolFM (Luo et al., 2023a)	72.9±0.1	77.2±0.7	64.4±0.2	64.2±0.9	79.7±1.6	76.0±0.8	78.8±1.1	83.9±1.1	74.62
MoMu (Su et al., 2022)	70.5±2.0	75.6±0.3	63.4±0.5	60.5±0.9	79.9±4.1	70.5±1.4	75.9±0.8	76.7±2.1	71.63
MolCA-SMILES (Liu et al., 2023c)	70.8±0.6	76.0±0.5	56.2±0.7	61.1±1.2	89.0±1.7	-	-	79.3±0.8	72.1
Atomas (Zhang et al., 2024)	**73.7±1.7**	77.8±0.4	66.9±0.9	64.4±1.9	**93.1±0.5**	76.3±0.7	80.5±0.43	83.1±1.7	77.01
LLM Based Generalist Models									
Qwen2.5-7b (Hui et al., 2024)	59.7±0.7	62.7±0.5	57.3±1.1	52.9±0.9	71.0±1.8	60.9±1.5	61.1±0.9	70.3±0.8	62.05
InstructMol (Cao et al., 2023)	55.4	-	-	-	-	-	57.5	63.2	58.70
HIGHT (Chen et al., 2025)	59.4	-	-	-	-	-	58.6	68.4	62.13
MotifAgent	73.4±0.8	**78.5±0.4**	**67.6±0.8**	**65.1±1.3**	90.9±0.7	**77.4±0.6**	**80.6±0.4**	**84.0±1.2**	**77.19**

Table 4: Progressive ablation across BRICS input, multi-agent architecture, and RL training. **MotifAgent-Lite** strips curriculum learning and the auxiliary critic, and reduces priority scoring to w_1 only; MotifAgent (Full) is our default configuration.

Method	BRICS	Multi-Agent	RL	Validity (%)	GED ↓	Morgan FTS↑	Conn. Site (%)	Bond Type (%)
Qwen2.5-7b (vanilla)	✗	✗	✗	67.6	11.21	0.584	—	—
Qwen2.5-7b + BRICS	✓	✗	✗	72.3	9.81	0.603	58.2	71.9
MotifAgent (w/o RL)	✓	✓	✗	77.9	7.82	0.653	60.1	73.7
Single-Agent (w/ RL)	✓	✗	✓	82.3	5.85	0.708	68.4	76.9
MotifAgent-Lite	✓	✓	✓	86.8	5.04	0.727	76.0	81.9
MotifAgent (Full)	✓	✓	✓	**95.6**	**4.31**	**0.792**	**87.6**	**92.3**

Table 5: Ablation study on curriculum learning strategy.

Method	Validity (%)↑	Levenshtein ↓	Morgan FTS↑	Conn. Site (%)↑	Bond Type (%)↑
MotifAgent (Full)	**95.6**	**11.4**	**0.792**	**87.6**	**92.3**
w/o Curriculum	89.2	13.5	0.751	81.3	85.0
2-Stage	92.8	12.3	0.769	84.1	88.4
Reverse Curriculum	90.1	13.4	0.748	81.5	84.7

Qwen2.5-3b) even outperforms the larger Qwen2.5-7b. These results confirm that the performance improvements stem from our multi-agent framework, and demonstrate the consistent effectiveness of MotifAgent across different model scales. MotifAgent also surpasses the "external graph encoder" baselines present in Table 1 (MoMu, MolFM, GIT-Mol, MolCA); a detailed comparison is given in Appendix L.4.

4.3 Chemical Reaction Prediction

To evaluate MotifAgent's capability in chemical reaction prediction tasks, we conduct comprehensive experiments on the Mol-Instructions dataset. We present the most challenging retrosynthesis prediction task here, with complete results for reagent prediction and forward reaction prediction available in Appendix F. Retrosynthesis prediction, which requires models to infer suitable reactants given target products, represents a fundamental challenge in AI-assisted synthetic route planning. We employ both linguistic distance metrics (BLEU, Levenshtein) and molecular fingerprint similarities (RDK, MACCS, MORGAN FTS) for comprehensive evaluation, along with chemical validity. We compare against general-purpose LLMs including Alpaca (Dubois et al., 2023), Baize (Xu et al., 2023), ChatGLM (Zeng et al., 2022), LLaMA (Touvron et al., 2023), and Vicuna (Chiang et al., 2023), as well as molecule-specific methods including Mol-Instruction (Fang et al., 2023), InstructMol (Cao et al., 2023), and HIGHT (Chen et al., 2025).

Table 6: Ablation study on reward design.

Variant	R_{chem}	R_{topo}	Validity (%)↑	Levenshtein ↓	Morgan FTS↑
Full	✓	✓	**95.6**	**11.4**	**0.792**
Chem-Only	✓	✗	92.9	12.2	0.771
Topo-Only	✗	✓	89.1	13.3	0.759
Minimal (r_{valid} only)	partial	✗	85.2	15.8	0.732

Table 2 presents the retrosynthesis prediction results, where MotifAgent outperforms all baseline methods across key metrics. It demonstrates superior exact match accuracy (+36.1% over HIGHT-GS), sequence generation quality, and molecular structure similarity. MotifAgent modeling molecular formation through effective multi-agent collaboration, identifying and tracking structural transformations at the motif level, which is essential for recognizing reaction centers. These results establish MotifAgent as a powerful tool for AI-driven retrosynthetic analysis.

4.4 Molecular Property Prediction

We evaluate MotifAgent on 8 benchmark datasets from MoleculeNet (Wu et al., 2018) for molecular property classification. We adopt scaffold split following MoleculeSTM (Liu et al., 2023a) to ensure rigorous evaluation. Molecules are decomposed into motif sets via the BRICS algorithm with corresponding textual descriptions constructed. We compare against specialist models (Liu et al., 2023a; Luo et al., 2023a; Su et al., 2022; Liu et al., 2023c; Zhang et al., 2024), as well as LLM-based generalist models (Cao et al., 2023; Chen et al., 2025).

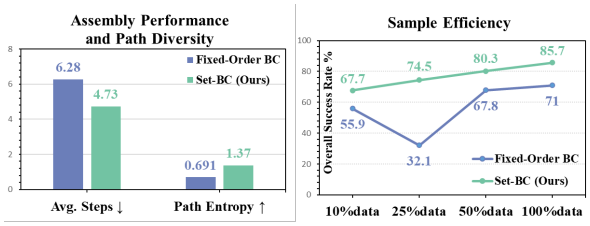


Figure 4: Ablation study on Set-BC effectiveness.

Table 3 shows MotifAgent's superior performance, with an average ROC-AUC of 77.2%, significantly outperforming existing LLM-based models. Comparing against its backbone Qwen2.5-7b, MotifAgent achieves a 24.3% relative improvement, validating the effectiveness of our multi-agent framework in reconstructing molecular structures and learning motif connection rules. MotifAgent excels as a generalist framework, surpassing specialist models on tasks like Tox21, ToxCast, Sider, MUV, HIV, and Bace, bridging the gap between generalist and specialist approaches. This demonstrates the potential of general LLM-based molecular-text representation frameworks.

4.5 Ablation Studies

To validate the key design choices in MotifAgent, we conduct comprehensive ablation studies on five critical components: (1) multi-agent collaboration, (2) Set-based Behavior Cloning, (3) curriculum learning strategy, (4) reward design, and (5) critic auxiliary tasks. Details of evaluation metrics are provided in Appendix E.4.

Multi-Agent Collaboration and Progressive Ablation. Table 4 disentangles BRICS input, the multi-agent architecture, and MAPPO training under identical backbones and hyperparameters. Each component alone yields only moderate gains over vanilla Qwen2.5-7B (67.6% Validity): +BRICS adds 4.7 points, multi-agent without RL adds another 5.6 (77.9%), and a single-agent baseline trained with MAPPO reaches 82.3%. Their combination is super-linear — *MotifAgent-Lite* (multi-agent + RL, no curriculum / no auxiliary critic / scoring restricted to w_1) reaches 86.8%, and the Full model 95.6%, with parallel gains in GED, Morgan FTS, and connection-level accuracies. On Morgan FTS, the $+0.139$ RL gain on top of the multi-agent system far exceeds the $+0.105$ on the single-agent baseline, confirming that the advantage stems from the coupling of multi-agent collaboration and RL; the +8.8-point Lite→Full gap quantifies the cumulative contribution of curriculum learning, auxiliary critic, and richer priority scoring, which matter most in the complex regime analyzed in Section 4.6.

Table 7: Ablation study on priority scoring $S(a)$.

Method	w_1	w_2	w_3	Validity (%)↑	Levenshtein ↓	Morgan FTS↑
Default	1.0	1.0	1.0	**95.6**	11.4	**0.792**
w/o Chem	0.0	1.0	1.0	92.3	12.1	0.758
w/o Topo	1.0	0.0	1.0	92.7	11.8	0.764
w/o Prop	1.0	1.0	0.0	93.4	**11.1**	0.783

Table 8: Ablation study on critic auxiliary tasks.

Variant	w_1	w_2	Conn. Site (%)↑	Bond Type (%)↑
Default	0.1	0.1	**87.6**	**92.3**
w/o Auxiliary	0.0	0.0	82.1	86.3
w/o Edge Pred	0.0	0.1	84.0	88.7
w/o CC Pred	0.1	0.0	84.6	89.1
Edge-Heavy	0.5	0.1	87.2	90.5

Set-BC vs. Fixed-Order Supervision. Since molecules can be correctly assembled through multiple equivalent paths, forcing models to learn single arbitrary sequences may hinder learning efficiency. We compare Set-BC against a fixed-order baseline using traditional behavior cloning with BFS-determined sequences. As shown in Figure 4, Set-BC achieves more efficient assembly with fewer steps and higher path entropy, confirming successful learning of diverse assembly strategies. Most significantly, Set-BC maintains superior sample efficiency across all data scales.

Curriculum Learning Strategy. We examine the importance of our four-stage curriculum by comparing against: training without curriculum, a simplified 2-stage variant, and reversed curriculum order. Table 5 shows that the full curriculum achieves the best performance. Removing curriculum learning causes notable degradation across all metrics, while the reversed curriculum performs even worse than no curriculum, confirming that progressive skill acquisition is essential.

Reward Design. We ablate the two reward categories introduced in Section 3.4. Table 6 shows that both chemical and topological rewards contribute substantially: removing topological rewards causes the largest validity drop (to 89.1%), while removing chemical rewards degrades structural similarity most significantly. The minimal baseline using only validity reward performs worst, confirming the necessity of comprehensive reward design.

Priority Scoring. We examine the arbitrator's scoring function by zeroing individual weight

Table 9: Quality and computational cost of MotifAgent stratified by motif count. Time and peak memory are measured on a single NVIDIA A100 (80 GB) under BFloat16 inference.

Motif Count	Validity (%)↑	Morgan FTS↑	QED↑	FG-PR (%)↑	Conn. Site (%)↑	Bond Type (%)↑	Avg. Tokens	Time (s/mol)↓	Peak Memory (GB)
2–4 (Simple)	96.3	0.798	0.560	93.1	88.5	92.8	1,601	3.0	24.2
5–7 (Moderate)	95.4	0.787	0.557	92.0	86.2	92.1	2,243	4.3	27.4
8–10 (Complex)	94.1	0.781	0.552	91.5	87.3	91.0	3,363	6.4	32.7
11–15 (Highly Complex)	92.7	0.778	0.546	90.8	84.9	90.1	5,380	10.3	42.1

Table 10: Training-phase efficiency: MotifAgent versus the single-agent baseline under identical backbones and hyperparameters.

Method	Time (s/sample)	Tokens (/sample)	Steps to 75% FTS	Total Time to 75% FTS (h)	Final FTS↑
Single-Agent	5.4	2625.8	N/A†	N/A†	0.708
MotifAgent	**4.1**	**2146.4**	**163K**	**52.7**	**0.792**

† Single-Agent never reaches 75% Morgan FTS within 200K steps.

terms. Table 7 indicates that all three components contribute positively, with chemical stability (w_1) showing the largest impact on validity and topological progress (w_2) most affecting structural metrics.

Critic Auxiliary Tasks. We ablate the auxiliary prediction heads in the critic network (Eq. 7). Table 8 shows that removing both auxiliary tasks significantly degrades connection accuracy. Edge prediction (w_1) and connected component prediction (w_2) provide complementary benefits, with their combination yielding the best performance.

4.6 Scalability and Computational Cost

We now examine MotifAgent's computational behavior along two complementary axes: training-phase efficiency relative to a single-agent baseline of the same capacity, and inference-phase scalability as the target molecule's motif count grows.

Training-phase efficiency. Counter-intuitively, the multi-agent architecture is more efficient than a single-agent baseline at the same backbone scale. As shown in Table 10, MotifAgent processes each sample 24% faster and consumes 18% fewer tokens. The reason is that the single-agent must unfold a complete reasoning chain within one context—covering motif selection, site judgment, bond typing, and global consistency—leading to long generations and frequent backtracking, whereas MotifAgent distributes these sub-decisions across specialized agents that communicate through compact structured messages rather than free-form reasoning. End-to-end, MotifAgent reaches 75% Morgan FTS in 52.7 hours, while the single-agent never crosses this threshold within 200K training steps, confirming that the multi-agent decomposition delivers higher final quality and shorter time-to-target.

Inference-phase scalability. Table 9 stratifies the PubChem test split by motif count and reports both quality and per-molecule compute. Quality degrades gracefully: from Simple (2–4) to Highly Complex (11–15), Validity drops only 3.6 points, Morgan FTS 0.020, QED 0.014, and FG-PR 2.3 points, with every metric remaining above 90% at the hardest tier, evidence that the learned connection rules transfer to larger assembly graphs and that functional-group semantics (captured by FG-PR but not by Morgan FTS) are preserved end-to-end. Cost grows sub-linearly: over a 3.75× increase in motif count, per-molecule inference time rises by only 3.4× (3.0s → 10.3s) and peak GPU memory by 1.7× (24.2 GB → 42.1 GB), both well below the linear baseline and well inside a single A100 (80 GB). Two factors explain the sub-linear growth: hierarchical action sampling decomposes the combinatorial connection choice into $O(N)$ per-layer decisions rather than an $O(N^2)$ joint enumeration, and the arbitrator's legality filter prunes infeasible proposals before they reach the policy update. A residual failure-mode analysis is deferred to the Limitations section.

5 Conclusion

We presented MotifAgent, a multi-agent reinforcement learning framework that effectively addresses LLMs' limitations in understanding molecular generation principles. Our approach explicitly learns motif connection rules governing molecular topology, and leverages the CTDE framework combined with Set-BC to learn from multiple equivalent assembly paths. Experiments demonstrate that MotifAgent achieves state-of-the-art performance on molecular property prediction, description generation, and chemical reaction prediction tasks, proving its generalization and scalability.

Acknowledgments

This research was supported in part by National Natural Science Foundation of China (No. 92470128, No. U2241212). We also wish to

acknowledge the support provided by the fund for building world-class universities (disciplines) of Renmin University of China, by Engineering Research Center of Next-Generation Intelligent Search and Recommendation, Ministry of Education, by Intelligent Social Governance Interdisciplinary Platform, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative, Public Policy and Decision-making Research Lab, and Public Computing Cloud, Renmin University of China.

Limitations

Our work has several limitations that should be acknowledged.

Computational overhead. The multi-agent framework requires separate inference through the shared LLM backbone for each motif agent, which may limit scalability for real-time applications. Section 4.6 shows that while per-molecule inference cost grows sub-linearly with motif count (3.4× time and 1.7× peak memory across a 3.75× range in motif count), molecules beyond 15 motifs remain outside our verified regime. Future directions include batched multi-agent inference and a coarse-then-local hierarchical decomposition scheme that keeps the effective motif count within the validated range.

Scope of validation. The framework has been primarily validated on drug-like small molecules using ChEBI-20, MoleculeNet, and Mol-Instructions. Generalization to natural products, polymers, and organometallic compounds is unexplored; doing so will likely require swapping BRICS for RECAP/MMPA-style decompositions, treating repeat units as motifs, and extending the bond-breaking rule set with coordination bonds.

Stereochemistry. Our current action space (i, s_i, j, s_j, b) focuses on 2D topological connection rules and does not explicitly model stereochemistry. Motif-internal chirality is preserved through SMILES `@/@@` annotations, and fewer than 4% of molecules in our training data contain a BRICS cut point that coincides with a chiral center; together with the fact that ChEBI-20, MoleculeNet, and Mol-Instructions all evaluate 2D structural and property endpoints, the omission has a minor effect on the metrics we report. A natural extension is to expand the action tuple to $(i, s_i, j, s_j, b, \text{stereo})$ with stereo $\in \{\text{none}, R, S, E, Z\}$ as a fifth layer of our hierarchical sampling; this requires only an additional action head, no architectural change, and enables future applications to enantioselective reaction prediction.

Failure modes. On the PubChem test split we observe four recurring error categories: *(1) Positional Isomer Confusion*, where two candidate sites share nearly identical local chemical environments — this is the most common category and is illustrated by the vanillin → isovanillin case (Morgan FTS 0.72, GED 2, Conn. Site 1/3, Bond Type 3/3, chemically valid) in which the policy margin at the decisive step is only $\pi(\text{C3}) = 0.47$ vs. $\pi(\text{C4}) = 0.53$; *(2) Bond-Type Misassignment*, predominantly in partially aromatic systems where BRICS leaves an aromaticity hint but not a fully constrained bond type; *(3) Incomplete Assembly* caused by premature `STOP` under a too-aggressive convergence threshold ε; and *(4) Chemical Validity Violations* from rare BRICS fragmentation patterns unseen in training. The shared root cause of (1) is that the textual state cannot disambiguate locally equivalent sites; a full taxonomy and walk-through is given in Appendix L.2, and the mitigation direction is to augment the state with explicit positional encodings or a graph-based site-disambiguation head.

References

Stefano V Albrecht, Filippos Christianos, and Lukas Schäfer. 2024. *Multi-agent reinforcement learning: Foundations and modern approaches*. MIT Press.

Viraj Bagal, Rishal Aggarwal, PK Vinod, and U Deva Priyakumar. 2021. Molgpt: molecular generation using a transformer-decoder model. *Journal of chemical information and modeling*, 62(9):2064–2076.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Nurken Berdigaliyev and Mohamad Aljofan. 2020. An overview of drug discovery and development. *Future medicinal chemistry*, 12(10):939–947.

Ryan PA Bettens and Adrian M Lee. 2006. A new algorithm for molecular fragmentation in quantum chemical calculations. *The Journal of Physical Chemistry A*, 110(28):8777–8785.

Camille Bilodeau, Wengong Jin, Tommi Jaakkola, Regina Barzilay, and Klavs F Jensen. 2022. Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdisci-*

plinary Reviews: Computational Molecular Science, 12(5):e1608.

Lorenzo Canese, Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Daniele Giardino, Marco Re, and Sergio Spanò. 2021. Multi-agent reinforcement learning: A review of challenges and applications. *Applied Sciences*, 11(11):4948.

He Cao, Zijing Liu, Xingyu Lu, Yuan Yao, and Yu Li. 2023. Instructmol: Multi-modal integration for building a versatile and reliable molecular assistant in drug discovery. *arXiv preprint arXiv:2311.16208*.

Yongqiang Chen, Quanming Yao, Juzheng Zhang, James Cheng, and Yatao Bian. 2025. Hight: Hierarchical graph tokenization for molecule-language alignment. *Preprint*, arXiv:2406.14021.

Wei-Lin Chiang, Zhuohan Li, Ziqing Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, and 1 others. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, 2(3):6.

Dimitrios Christofidellis, Giorgio Giannone, Jannis Born, Ole Winther, Teodoro Laino, and Matteo Manica. 2023. Unifying molecular and textual representations via multi-task language modelling. In *International Conference on Machine Learning*, pages 6140–6157. PMLR.

Michael A Collins and Ryan PA Bettens. 2015. Energy-based molecular fragmentation methods. *Chemical reviews*, 115(12):5607–5642.

Jorg Degen, Christof Wegscheid-Gerlach, Andrea Zaliani, and Matthias Rarey. 2008. On the art of compiling and using'drug-like'chemical fragment spaces. *ChemMedChem*, 3(10):1503.

Yann Dubois, Chen Xuechen Li, Rohan Taori, Tianyi Zhang, Ishaan Gulrajani, Jimmy Ba, Carlos Guestrin, Percy S Liang, and Tatsunori B Hashimoto. 2023. Alpacafarm: A simulation framework for methods that learn from human feedback. *Advances in Neural Information Processing Systems*, 36:30039–30069.

Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. 2022. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817*.

Carl Edwards, ChengXiang Zhai, and Heng Ji. 2021. Text2mol: Cross-modal molecule retrieval with natural language queries. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 595–607.

Yin Fang, Xiaozhuan Liang, Ningyu Zhang, Kangwei Liu, Rui Huang, Zhuo Chen, Xiaohui Fan, and Huajun Chen. 2023. Mol-instructions: A large-scale biomolecular instruction dataset for large language models. *arXiv preprint arXiv:2306.08018*.

Mingyan Gao, Yanzi Li, Banruo Liu, Yifan Yu, Phillip Wang, Ching-Yu Lin, and Fan Lai. 2025. Single-agent or multi-agent systems? why not both? *arXiv preprint arXiv:2505.18286*.

Zijie Geng, Shufang Xie, Yingce Xia, Lijun Wu, Tao Qin, Jie Wang, Yongdong Zhang, Feng Wu, and Tie-Yan Liu. 2023. De novo molecular generation via connection-aware motif mining. *arXiv preprint arXiv:2302.01129*.

Binyuan Hui, Jian Yang, Zeyu Cui, Jiaxi Yang, Dayiheng Liu, Lei Zhang, Tianyu Liu, Jiajun Zhang, Bowen Yu, Keming Lu, and 1 others. 2024. Qwen2. 5-coder technical report. *arXiv preprint arXiv:2409.12186*.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. 2018. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pages 2323–2332. PMLR.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. 2020. Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning*, pages 4839–4848. PMLR.

Mikhail V Koroteev. 2021. Bert: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*.

Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. 2020. Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024.

Jiatong Li, Yunqing Liu, Wenqi Fan, Xiao-Yong Wei, Hui Liu, Jiliang Tang, and Qing Li. 2024. Empowering molecule discovery for molecule-caption translation with large language models: A chatgpt perspective. *IEEE transactions on knowledge and data engineering*, 36(11):6071–6083.

Junxian Li, Di Zhang, Xunzhi Wang, Zeying Hao, Jingdi Lei, Qian Tan, Cai Zhou, Wei Liu, Yaotian Yang, Xinrui Xiong, and 1 others. 2025. Chemvlm: Exploring the power of multimodal large language models in chemistry area. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 415–423.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81.

Pengfei Liu, Yiming Ren, Jun Tao, and Zhixiang Ren. 2024. Git-mol: A multi-modal large language model for molecular science with graph, image, and text. *Computers in biology and medicine*, 171:108073.

Shengchao Liu, Weili Nie, Chengpeng Wang, Jiarui Lu, Zhuoran Qiao, Ling Liu, Jian Tang, Chaowei Xiao, and Animashree Anandkumar. 2023a. Multi-modal molecule structure–text model for text-based retrieval and editing. *Nature Machine Intelligence*, 5(12):1447–1457.

Zequn Liu, Wei Zhang, Yingce Xia, Lijun Wu, Shufang Xie, Tao Qin, Ming Zhang, and Tie-Yan Liu. 2023b. Molxpt: Wrapping molecules with text for generative pre-training. *arXiv preprint arXiv:2305.10688*.

Zhiyuan Liu, Sihang Li, Yanchen Luo, Hao Fei, Yixin Cao, Kenji Kawaguchi, Xiang Wang, and Tat-Seng Chua. 2023c. Molca: Molecular graph-language modeling with cross-modal projector and uni-modal adapter. *arXiv preprint arXiv:2310.12798*.

Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017a. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30.

Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. 2017b. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30.

Yizhen Luo, Kai Yang, Massimo Hong, Xing Yi Liu, and Zaiqing Nie. 2023a. Molfm: A multimodal molecular foundation model. *arXiv preprint arXiv:2307.09484*.

Yizhen Luo, Jiahuan Zhang, Siqi Fan, Kai Yang, Yushuai Wu, Mu Qiao, and Zaiqing Nie. 2023b. Biomedgpt: Open multimodal generative pre-trained transformer for biomedicine. *arXiv preprint arXiv:2308.09442*.

Krzysztof Maziarz, Henry Jackson-Flux, Pashmina Cameron, Finton Sirockin, Nadine Schneider, Nikolaus Stiefl, Marwin Segler, and Marc Brockschmidt. 2021. Learning to extend molecular scaffolds with structural motifs. *arXiv preprint arXiv:2103.03864*.

Eyal Mazuz, Guy Shtar, Bracha Shapira, and Lior Rokach. 2023. Molecule generation using transformers and policy gradient reinforcement learning. *Scientific Reports*, 13(1):8799.

Medard Edmund Mswahili and Young-Seob Jeong. 2024. Transformer-based models for chemical smiles representation: A comprehensive literature review. *Heliyon*, 10(20).

Andrew Y Ng, Daishi Harada, and Stuart Russell. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. In *Icml*, volume 99, pages 278–287. Citeseer.

George Papadatos, Anna Gaulton, Anne Hersey, and John P Overington. 2015. Activity, assay and target data curation and quality in the chembl database. *Journal of computer-aided molecular design*, 29(9):885–896.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318.

Chen Qian, Zihao Xie, Yifei Wang, Wei Liu, Kunlun Zhu, Hanchen Xia, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, and 1 others. 2024. Scaling large language model-based multi-agent collaboration. *arXiv preprint arXiv:2406.07155*.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, and 1 others. 2018. Improving language understanding by generative pre-training.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2023. Exploring the limits of transfer learning with a unified text-to-text transformer. *Preprint*, arXiv:1910.10683.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2020. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178):1–51.

Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. 2020. Graphaf: a flow-based autoregressive model for molecular graph generation. *arXiv preprint arXiv:2001.09382*.

Jia Song, Wanru Zhuang, Yujie Lin, Liang Zhang, Chunyan Li, Jinsong Su, Song He, and Xiaochen Bo. 2024. Towards cross-modal text-molecule retrieval with better modality alignment. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1161–1168. IEEE.

Bing Su, Dazhao Du, Zhao Yang, Yujie Zhou, Jiangmeng Li, Anyi Rao, Hao Sun, Zhiwu Lu, and Ji-Rong Wen. 2022. A molecular multimodal foundation model associating molecule graphs with natural language. *arXiv preprint arXiv:2209.05481*.

Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, and 1 others. 2017. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*.

Qian Tan, Dongzhan Zhou, Peng Xia, Wanhao Liu, Wanli Ouyang, Lei Bai, Yuqiang Li, and Tianfan Fu. 2025. Chemmllm: Chemical multimodal large language model. *arXiv preprint arXiv:2505.16326*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D Nguyen. 2025. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

William Yi Wang, Jinshan Li, Weimin Liu, and Zi-Kui Liu. 2019. Integrated computational materials engineering for advanced materials: A brief review. *Computational Materials Science*, 158:42–48.

Muning Wen, Jakub Kuba, Runji Lin, Weinan Zhang, Ying Wen, Jun Wang, and Yaodong Yang. 2022. Multi-agent reinforcement learning is a sequence modeling problem. *Advances in Neural Information Processing Systems*, 35:16509–16521.

Daniel S Wigh, Jonathan M Goodman, and Alexei A Lapkin. 2022. A review of molecular representation in the age of machine learning. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5):e1603.

Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. 2018. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530.

Canwen Xu, Daya Guo, Nan Duan, and Julian McAuley. 2023. Baize: An open-source chat model with parameter-efficient tuning on self-chat data. *arXiv preprint arXiv:2304.01196*.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. 2022. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in neural information processing systems*, 35:24611–24624.

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, and 1 others. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*.

Yikun Zhang, Geyan Ye, Chaohao Yuan, Bo Han, Long-Kai Huang, Jianhua Yao, Wei Liu, and Yu Rong. 2024. Atomas: Hierarchical alignment on molecule-text for unified molecule understanding and generation. *arXiv preprint arXiv:2404.16880*.

Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. 2021. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34:15870–15882.

Zaixi Zhang, Yaosen Min, Shuxin Zheng, and Qi Liu. 2023. Molecule generation for target protein binding with structural motifs. In *The eleventh international conference on learning representations*.

Haiteng Zhao, Shengchao Liu, Ma Chang, Hannan Xu, Jie Fu, Zhihong Deng, Lingpeng Kong, and Qi Liu. 2023a. Gimlet: A unified graph-text model for instruction-based molecule zero-shot learning. *Advances in neural information processing systems*, 36:5850–5887.

Wenyu Zhao, Dong Zhou, Buqing Cao, Kai Zhang, and Jinjun Chen. 2023b. Adversarial modality alignment network for cross-modal molecule retrieval. *IEEE Transactions on Artificial Intelligence*, 5(1):278–289.

Table 11: Hyperparameter details for MotifAgent.

Hyperparameter	**Value**
Model Architecture	
policy backbone	Qwen2.5-7B
critic network	MolT5-base
LoRA rank	16
action head hidden dim	512
Training Configuration	
parallel environments	8
rollout length	32
batch size (transitions)	256
PPO epochs	10
max training steps	200K
max sequence length	512
precision	BFloat16 Automatic Mixed Precision
Optimization	
actor learning rate	1e-5
critic learning rate	1e-4
gradient clip norm	0.5
optimizer	AdamW
warmup steps	10000
PPO Parameters	
clipping ϵ	0.2
GAE λ	0.95
discount γ	0.99
value loss coefficient	0.5
Regularization	
entropy β (initial)	0.01
entropy β (final)	0.001
KL penalty β_{KL}	0.1
Set-BC α_{BC} (initial)	1.0
Set-BC α_{BC} (final)	0.1
Set-BC decay steps	50% of training
Auxiliary Tasks	
edge prediction weight w_1	0.1
component prediction weight w_2	0.1

A Potential Risks

While MotifAgent is designed for beneficial applications in drug discovery and molecular understanding, we acknowledge potential dual-use concerns. The model's capability to understand molecular assembly rules could theoretically be misused

to design harmful substances. However, the framework focuses on learning general chemical connection principles rather than optimizing for specific harmful properties, and all training data are derived from publicly available databases. We recommend responsible deployment with appropriate safeguards when applying this technology to real-world molecular design scenarios.

B Data Privacy and Content Review

All data used in this work are derived from publicly available chemical databases, including PubChem and MoleculeNet benchmarks. These datasets contain only molecular structures (SMILES strings), chemical properties, and scientific descriptions, without any personally identifiable information or user-generated content. The molecular descriptions are professionally curated scientific text describing chemical and biological characteristics. We have verified that our datasets do not contain offensive content, as they consist solely of standardized chemical nomenclature and objective property descriptions. No anonymization was required since the data pertains exclusively to chemical compounds rather than individuals.

C LLM Usage

Large language models (LLMs) were used for refining sentence structure, improving grammatical accuracy, and enhancing the clarity of the manuscript text. A supporting role was played by the LLMs in the manuscript's language polishing, but no scientific content, data analysis, or experimental design was generated by the LLMs.

D Initial Training Details

We provide comprehensive training configuration and hyperparameter settings for MotifAgent in Table 11. The model is trained with 8 NVIDIA Tesla A100 GPUs (80GB RAM/GPU) with Qwen2.5-7B as the shared policy backbone and MolT5-base as the centralized critic. We employ 8 parallel environments with a rollout length of 32 steps, resulting in 256 transitions per update. The actor learning rate is set to 5e-5 with LoRA (rank=16), while the critic uses 3e-4. The training employs curriculum learning with four phases, automatically transitioning based on performance metrics to progressively master topological construction from strict reconstruction to free exploration. For curriculum learning, Set-BC weight α_{BC} decays from 1.0 to 0.1 over the first 50% of training (steps).

Curriculum Learning Details: The four-phase curriculum is designed to progressively build the model's capabilities: Phase 1 (0-25%): Strict reconstruction with hard masks, focusing on learning valid chemical connections; Phase 2 (25-50%): Soft constraints with Set-BC weight decay, allowing exploration of equivalent paths; Phase 3 (50-75%): Property-aware reconstruction, introducing target property rewards; Phase 4 (75-100%): Free exploration for generation, removing hard topological constraints. The automatic phase transitions ensure the model has sufficiently mastered each level before progressing, preventing premature exploration that could lead to unstable training.

E Evaluation Metrics

In this section, we introduce the evaluation metrics used in the experiments presented in the main text and ablation studies.

E.1 Molecular Description Generation

BLEU-N: BLEU (Bilingual Evaluation Understudy) (Papineni et al., 2002) measures the N-gram overlap between the generated text $\hat{y}$ and the reference text y. BLEU-1 focuses on unigram matching, BLEU-2 considers bigrams, and BLEU-4 requires longer phrase matching, meaning that not only vocabulary but also expression patterns should be similar. The BLEU-N score is computed as:

$$\text{BLEU-N} = BP \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right), \quad (9)$$

where p_n is the modified N-gram precision, $BP = \min(1, e^{1-|y|/|\hat{y}|})$ is the brevity penalty, and w_n is typically set to $\frac{1}{N}$ for uniform weighting.

ROUGE-N: ROUGE (Recall-Oriented Understudy for Gisting Evaluation) (Lin, 2004) emphasizes recall by measuring how much of the reference text is captured in the generated text. ROUGE-N (based on N-grams) is computed as:

$$\text{ROUGE-N} = \frac{\sum_{g \in y} \min(C_{\hat{y}}(g), C_y(g))}{\sum_{g \in y} C_y(g)}, \quad (10)$$

where g denotes an n-gram, and $C_{\hat{y}}(g)$, $C_y(g)$ represent the count of g in the generated text $\hat{y}$ and reference text y, respectively.

ROUGE-L: ROUGE-L is based on the Longest Common Subsequence (LCS) and is computed as:

$$\text{ROUGE-L} = F_{\text{lcs}} = \frac{(1+\beta^2) \cdot P_{\text{lcs}} \cdot R_{\text{lcs}}}{R_{\text{lcs}} + \beta^2 \cdot P_{\text{lcs}}}, \quad (11)$$

where $P_{\text{lcs}} = \frac{\text{LCS}(\hat{y},y)}{|\hat{y}|}$, $R_{\text{lcs}} = \frac{\text{LCS}(\hat{y},y)}{|y|}$, LCS denotes the length of the longest common subsequence, and β is typically set to a large value to favor recall.

METEOR: METEOR (Metric for Evaluation of Translation with Explicit ORdering) (Banerjee and Lavie, 2005) considers both precision and recall while incorporating synonym matching and word order penalty:

$$\text{METEOR} = F_{\text{mean}} \cdot (1 - \text{Penalty}), \tag{12}$$

where $F_{\text{mean}} = \frac{10 \cdot P \cdot R}{R + 9 \cdot P}$, with P and R denoting the unigram precision and recall, respectively. The Penalty term reflects the degree of word order disruption through the number of contiguous chunks of matched words.

E.2 Chemical Reaction Prediction

EXACT: Exact match accuracy measures whether the generated reactant SMILES is identical to the ground truth:

$$\text{EXACT} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}[\hat{y}_i = y_i], \tag{13}$$

where $\mathbf{1}[\cdot]$ is the indicator function. This is the strictest metric, requiring the generated result to be exactly the same as the reference.

LEVENSHTEIN: Edit distance measures the minimum number of single-character operations (insertions, deletions, substitutions) required to transform the generated sequence into the reference sequence.

Fingerprint-based Tanimoto Similarity (FTS): We employ three molecular fingerprints to measure structural similarity between generated and reference molecules via Tanimoto similarity:

$$\text{Tanimoto}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \tag{14}$$

where A and B are the fingerprint bit vectors of two molecules. The Tanimoto coefficient is computed as the ratio of the intersection to the union of fingerprint bits, ranging from 0 (completely dissimilar) to 1 (identical). Specifically, we report:

- **RDK FTS:** RDKit topological fingerprints encode subgraph patterns by enumerating all linear and branched substructures up to a specified path length (default 7 bonds). This fingerprint captures the presence of specific atomic connectivity patterns, providing a comprehensive representation of molecular topology.
- **MACCS FTS:** MACCS (Molecular ACCess System) structural keys consist of 166 predefined binary features representing the presence or absence of specific structural fragments, functional groups, and atom configurations. This fingerprint is particularly effective for detecting pharmacophore-relevant substructures.
- **MORGAN FTS:** Morgan circular fingerprints (also known as Extended-Connectivity Fingerprints, ECFP) encode circular atomic neighborhoods with radius 2 and 2048 bits. Each atom's local chemical environment is hashed into the fingerprint, capturing both atom types and their connectivity within the specified radius.

VALIDITY: Validity measures the proportion of generated SMILES strings that can be parsed into chemically valid molecular structures. For each molecule, we verify valence rules for all atoms, check for proper aromaticity preservation, ensure no sterically impossible connections exist, and validate that all formed rings are chemically reasonable.

E.3 Molecular Property Prediction

BBBP (Blood-Brain Barrier Penetration): This dataset contains binary labels for over 2,000 compounds indicating whether they can penetrate the blood-brain barrier, a membrane that blocks most drugs from reaching the central nervous system.

Tox21 (Toxicology in the 21st Century): A public toxicity dataset from the Tox21 Data Challenge, containing qualitative toxicity measurements for 8,014 compounds across 12 biological targets, including nuclear receptors and stress response pathways.

ToxCast: A large-scale toxicology dataset from the same initiative as Tox21, providing *in vitro* high-throughput screening data for thousands of compounds across over 600 toxicity-related tasks.

SIDER (Side Effect Resource): A database of marketed drugs and their recorded adverse drug reactions (ADR), grouped into 27 system organ classes.

ClinTox: This dataset compares FDA-approved drugs with drugs that failed clinical trials due to toxicity reasons, compiled from the SWEETLEAD and AACT databases.

MUV (Maximum Unbiased Validation): A benchmark dataset selected from PubChem BioAssay through refined nearest neighbor analysis, con-

taining 17 challenging tasks specifically designed for validating virtual screening techniques.

HIV: This dataset contains experimentally measured abilities of compounds to inhibit HIV replication.

BACE: This dataset provides binary classification labels for inhibitors of human β-secretase 1 (BACE-1), a key target for Alzheimer's disease therapeutics.

E.4 Ablation Study Metrics

This subsection describes the evaluation metrics used in our ablation studies for analyzing multi-agent collaboration and Set-BC effectiveness.

E.4.1 Multi-Agent vs. Single-Agent Architecture

Chemical Validity (%): This metric evaluates the percentage of assembled molecules that satisfy fundamental chemical constraints, including valence rules, aromaticity preservation, steric feasibility, and ring validity. It is computed as:

$$\text{Chem. Val.} = \frac{\#\text{chemically valid assemblies}}{\#\text{total assembly attempts}} \times 100\%. \quad (15)$$

Graph Edit Distance (GED): GED quantifies the structural difference between the assembled molecule and the target molecule using the minimum number of graph edit operations (node/edge insertion and deletion) required for transformation. We use the Hungarian algorithm for optimal node matching, considering both node labels (motif types) and edge labels (bond types). Lower values indicate better topological reconstruction.

Connection Site Accuracy (%): This metric evaluates the model's ability to identify correct connection points on motifs. For each connection, we verify whether the chosen sites on both motifs match the target molecule's sites, regardless of bond type.

Bond Type Accuracy (%): This metric measures the correctness of bond type selection (single, double, triple, aromatic) given correctly identified connection sites, isolating the model's understanding of chemical bonding rules.

E.4.2 Set-BC vs. Fixed-Order Supervision

Average Assembly Steps: This metric measures the mean number of connection actions required to successfully complete molecular assembly, excluding failed attempts. Lower values indicate more efficient assembly strategies.

Path Entropy: Path entropy quantifies the diversity of assembly strategies learned by the model. For each target molecule, we generate 100 independent assembly trajectories and compute the Shannon entropy over the distribution of unique paths:

$$H = -\sum_i p_i \log p_i, \quad (16)$$

where p_i is the frequency of the i-th unique path. Higher entropy indicates successful learning of multiple valid assembly strategies.

Sample Efficiency: This metric evaluates model robustness under data-scarce conditions by training on randomly sampled subsets (10%, 25%, 50%, 100%) of the training data. We report the Overall Success Rate, combining reconstruction accuracy and chemical validity. Larger performance gaps in low-data regimes indicate superior sample efficiency.

F Chemical Reaction Prediction

We evaluated MotifAgent on three chemical reaction prediction tasks from the Mol-Instructions dataset (Fang et al., 2023): reagent prediction, forward reaction prediction, and retrosynthesis. These tasks are crucial for AI-assisted drug discovery. All inputs and outputs adopt SELFIES representation. Evaluation metrics include linguistic distance measures (BLEU, Levenshtein distance) and molecular fingerprint similarities (RDK FTS, MACCS FTS, MORGAN FTS) computed via RDKit.

Table 12 shows that MotifAgent achieves state-of-the-art performance across all three tasks. For reagent prediction, MotifAgent attains 8.5% exact match rate (26.9% improvement over HIGHT-GS), 0.516 BLEU score, and 22.571 Levenshtein distance, outperforming all baselines including Mol-Instruction which uses Llama-2 (Touvron et al., 2023) backbone. In forward reaction prediction, MotifAgent achieves 31.5% exact match rate and 0.937 BLEU score, with molecular fingerprint similarities reaching 0.806 (RDK FTS), 0.669 (MACCS FTS), and 0.582 (MORGAN FTS), all setting new records. For the most challenging retrosynthesis task, MotifAgent reaches 27.5% exact match rate (36.1% relative improvement over HIGHT-GS), with MORGAN FTS achieving 0.631, significantly higher than other methods.

MotifAgent's superior performance stems from its multi-agent collaborative mechanism that understands chemical reactions at the motif level—modeling functional group transformations, reaction center identification, and electron trans-

fer paths. Each motif agent encodes local chemical environments while perceiving global reaction changes through communication. The Set-BC mechanism enables learning multiple equivalent reaction pathways, crucial for reactions with multiple mechanisms. All tasks achieve 100% chemical validity, demonstrating the effectiveness of our chemical constraints and topological consistency checks. These results establish MotifAgent as a new benchmark for AI-assisted reaction prediction, providing a novel technical pathway for computational chemistry applications.

G Algorithm

In this section, we present the detailed procedure of MotifAgent. The complete workflow is summarized in Algorithm 1.

H Priority Scoring Details

This section provides the detailed formulations of the priority scoring function used by the central arbitrator (Equation 2 in the main text). The scoring function evaluates candidate actions across three dimensions: chemical stability, topological progress, and property improvement.

H.1 Scoring Function Overview

We adopt equal weights ($w_1 = w_2 = w_3 = 1.0$) to treat all three dimensions as equally important, allowing the model to learn the appropriate balance through training rather than imposing manual preferences. Each component score is normalized to $[0, 1]$, yielding a total score in $[0, 3]$.

H.2 Chemical Stability Score

The chemical stability score $\text{ChemStability}(a) \in [0, 1]$ evaluates whether the proposed connection forms a chemically favorable configuration by examining four aspects.

Valence Satisfaction checks whether the connection respects atomic valence constraints. A score of 1.0 is assigned if both atoms have available valence for the proposed bond type, 0.5 if the connection requires implicit hydrogen removal, and 0.0 if valence would be violated.

Aromaticity Preservation ensures aromatic systems remain intact after the connection. Actions that preserve or enhance aromaticity receive 1.0, appropriate connections to aromatic systems receive 0.7, and actions that would disrupt aromaticity receive 0.0.

Strain Energy penalizes high-strain configurations. We compute the estimated strain energy increment $E_{\text{strain}}(a)$ using force field calculations and convert it to a score via exponential decay with temperature parameter $\tau = 5.0$ kcal/mol.

Chemical Environment Compatibility assesses whether the connection matches typical chemical environments observed in training data. This is computed by an MLP that takes concatenated site embeddings (encoding hybridization, neighboring atoms, functional group membership) and bond type embedding as input.

The overall chemical stability score averages these four components:

$$\text{ChemStability}(a) = \frac{1}{4}\Big(S_{\text{valence}} + S_{\text{arom}} + S_{\text{strain}} + S_{\text{env}}\Big) \tag{17}$$

H.3 Topological Progress Score

The topological progress score $\text{TopoProgress}(a) \in [0, 1]$ measures how much the action advances toward the target molecular topology by considering four factors.

Target Edge Match rewards actions that create edges present in the target topology. If the proposed connection (i, j, b) exactly matches an edge in the target edge set E^* (including bond type), the score is 1.0. If the motif pair (i, j) matches but with a different bond type, the score is 0.2. Otherwise, the score is 0.0.

Connectivity Improvement rewards actions that reduce molecular fragmentation by merging disconnected components. The score is computed as the relative reduction in connected components: if the action merges the last two components into one, this term contributes maximally.

Ring Formation rewards completion of ring systems that match the target molecule. Completing a target ring system scores 1.0, contributing to a partial ring scores 0.5, and other actions score 0.0.

Bridge Contribution rewards key bridging connections between major substructures. Connecting two scaffold motifs scores 1.0, connecting a functional group to a scaffold scores 0.5, and peripheral connections score 0.0.

The overall topological progress score sums these components and normalizes to $[0, 1]$:

$$\text{TopoProgress}(a) = \frac{1}{4}\Big(P_{\text{edge}} + P_{\text{conn}} + P_{\text{ring}} + P_{\text{bridge}}\Big) \tag{18}$$

Algorithm 1 MotifAgent: Multi-Agent Molecular Assembly with Topological Learning

Require: Molecule dataset $\mathcal{D}$, Target properties $\mathcal{Y}$
Ensure: Trained policy π_θ, Critic V_ϕ

function FRAGMENTMOLECULE(M)
  $\mathcal{M}, E^* \leftarrow$ BRICS(M); $\mathcal{T} \leftarrow$ TextSerialize($\mathcal{M}$)
  return $\mathcal{M}, \mathcal{T}, E^*$
end function
function MULTIAGENTPROPOSAL($\mathcal{A}, G_t$)
  $\mathcal{P} \leftarrow \emptyset$
for each agent $a_i \in \mathcal{A}$ **do**
    $o_i \leftarrow$ LocalObs(a_i, G_t); $h_{\text{topo}} \leftarrow$ GlobalTopo(G_t)
    $\mathcal{P} \leftarrow \mathcal{P} \cup \{\pi_\theta(o_i, h_{\text{topo}})\}$ {Hierarchical sampling}
end for
  return $\mathcal{P}$
end function
function CENTRALARBITRATION($\mathcal{P}, E^*, G_t$, stage)
  $\mathcal{P}_{\text{valid}} \leftarrow$ ChemFilter($\mathcal{P}$)
if stage ≤ 2 **then**
    $\mathcal{P}_{\text{valid}} \leftarrow$ TargetMask($\mathcal{P}_{\text{valid}}, E^*$)
end if
  return $\arg\max_{a \in \mathcal{P}_{\text{valid}}} [w_1 \cdot \text{Chem}(a) + w_2 \cdot \text{Topo}(a) + w_3 \cdot \text{Prop}(a)]$
end function
function OPTIMIZE($\mathcal{B}, \pi_\theta, V_\phi, \pi_{\text{ref}}, E^*$)
  // Compute advantages and returns
  $\hat{A} \leftarrow \sum_{l=0}^{\infty} (\gamma\lambda)^l \delta_{t+l}$; $\hat{R} \leftarrow$ MonteCarloReturn($\mathcal{B}$)
  // Actor loss with Set-BC
  $\mathcal{A}^* \leftarrow \{a$: preserves target edges from $E^*\}$
  $\mathcal{L}_{\text{actor}} \leftarrow -\mathcal{L}_{\text{PPO}} - \beta\mathcal{H}(\pi_\theta) + \alpha_{BC} \log \sum_{a \in \mathcal{A}^*} \pi_\theta(a|s) + \beta_{KL}\text{KL}(\pi_\theta || \pi_{\text{ref}})$
  // Critic loss with auxiliary tasks
  $\mathcal{L}_{\text{critic}} \leftarrow (V_\phi - \hat{R})^2 + w_1(V_{\text{edges}} - |E^* \setminus E_t|)^2 + w_2(V_{\text{cc}} - \text{cc}(G_t))^2$
  Update $\theta \leftarrow \theta - \eta_\pi \nabla_\theta \mathcal{L}_{\text{actor}}$; Update $\phi \leftarrow \phi - \eta_V \nabla_\phi \mathcal{L}_{\text{critic}}$
end function
// **Main Training Loop**
$\pi_\theta \leftarrow$ SharedLLM(); $V_\phi \leftarrow$ Critic(); $\pi_{\text{ref}} \leftarrow \pi_\theta$; stage $\leftarrow 1$
for episode $= 1$ to N **do**
    $M \sim \mathcal{D}$; $\mathcal{M}, \mathcal{T}, E^* \leftarrow$ FRAGMENTMOLECULE(M)
    $\mathcal{A} \leftarrow$ [MotifAgent(m, π_θ) for $m \in \mathcal{M}$]; $G_t \leftarrow \emptyset$
  while $\neg$Terminal(G_t, E^*) **do**
      $\mathcal{P} \leftarrow$ MULTIAGENTPROPOSAL($\mathcal{A}, G_t$)
      $a_t \leftarrow$ CENTRALARBITRATION($\mathcal{P}, E^*, G_t$, stage)
      $G_{t+1}, r_t \leftarrow$ Execute(G_t, a_t); $r_t \leftarrow r_t + \gamma\Phi(G_{t+1}) - \Phi(G_t)$
      $\mathcal{B} \leftarrow \mathcal{B} \cup \{(G_t, a_t, r_t, G_{t+1})\}$; $G_t \leftarrow G_{t+1}$
  end while
  if $|\mathcal{B}| \geq$ batch_size **then**
      OPTIMIZE($\mathcal{B}, \pi_\theta, V_\phi, \pi_{\text{ref}}, E^*$); $\mathcal{B} \leftarrow \emptyset$
      if performance meets criteria **then** stage $\leftarrow$ stage $+1$; Adjust(α_{BC}, w_i)
  end if
end for

H.4 Property Improvement Score

The property improvement score $\text{PropImprove}(a) \in [0, 1]$ estimates how the action affects target molecular properties.

For each target property p in the property set $\mathcal{P}$ (e.g., LogP, molecular weight, pKa), we compute the reduction in deviation from the target value: the absolute difference before the action minus the absolute difference after. Positive values indicate improvement toward the target. These improvements are summed across all properties and passed through a sigmoid function for normalization:

$$\text{PropImprove}(a) = \sigma \sum_{p \in \mathcal{P}} \big(|y_p^{\text{current}} - y_p^{\text{target}}| - |y_p^{\text{after}} - y_p^{\text{target}}| \big) \tag{19}$$

where σ is the sigmoid function, y_p^{target} is the target value for property p, and y_p^{current}, y_p^{after} are predicted values before and after the action.

Property predictions use lightweight estimators: LogP is estimated via incremental atom/fragment contributions, molecular weight is calculated exactly from added atoms, and functional group properties use a lookup table for known pharmacophore contributions.

H.5 Scoring Example

We illustrate the scoring function using the aspirin assembly case, comparing two candidate actions in Round 1.

Candidate A: Benzene→Carboxyl. This connection establishes the salicylic acid core. The chemical stability score is high (0.95) because the connection satisfies valence rules, preserves benzene aromaticity, introduces minimal strain, and matches common benzoic acid patterns. The topological progress score is 0.75: the edge matches the target, it creates the first connected component, no ring is formed, and it connects a functional group to the scaffold. The property improvement score is 0.65 as it moves LogP and acidity toward target values. Total score: $0.95 + 0.75 + 0.65 = \mathbf{2.35}$.

Candidate B: Benzene→Oxygen. This connection would attach the oxygen linker directly. The chemical stability score is 0.84: valence is satisfied, but connecting to an aromatic system via ether linkage scores lower on aromaticity preservation. The topological progress score is 0.70: edge matches target, creates a connected component, but no ring or scaffold-scaffold bridge. The property improvement is modest (0.52) as it has less impact on target properties at this stage. Total score: $0.84 + 0.70 + 0.52 = \mathbf{2.06}$.

The arbitrator selects Candidate A (Benzene→Carboxyl) with the higher score, prioritizing establishment of the pharmacophore core before adding the ester linkage.

H.6 Curriculum Learning Behavior

While the weights remain fixed at 1.0 throughout training, the effective contribution of each score component naturally shifts across curriculum learning phases. In Phase 1 (Strict Reconstruction), TopoProgress dominates because hard masks restrict actions to target edges, and PropImprove has limited signal. In Phase 2 (Soft Constraint), all three components contribute meaningfully as the model explores beyond exact target edges. In Phase 3 (Property-Aware), PropImprove gains importance as property predictors become more reliable. In Phase 4 (Free Generation), the model leverages all three dimensions in full balance for novel molecule generation. This design allows the model to learn appropriate trade-offs through experience rather than through manually tuned weight schedules.

I Reward Function Details

This section provides the complete mathematical formulations of reward components introduced in Section 3.4.

I.1 Chemical Base Rewards

The chemical base reward R_{chem} provides a comprehensive assessment of molecular quality through the following evaluation rules:

C1. Validity Rule: Assigns $+1$ for chemically valid molecules satisfying valence and aromaticity constraints, -1 otherwise.

C2. Stability Rule: Penalizes high-strain configurations by $-E_{\text{strain}}$, where E_{strain} is the strain energy calculated using force field methods.

C3. Functional Group Rule: Rewards successful formation of target functional groups via $\sum_g w_g \mathbb{I}\{\text{form } g\}$, where w_g weights functional group g.

C4. Property Alignment Rule: Measures deviation from target properties as $-|y_{\text{pred}} - y_{\text{target}}|$.

C5. Synthetic Accessibility Rule: Incorporates the SA score from RDKit, favoring easily synthesizable structures.

C6. Novelty Rule: Encourages structural diversity via $1 - \max_{\text{ref}} \text{Tanimoto}(\text{mol}, \text{ref})$.

I.2 Topological Shaping Rewards

The topological shaping reward R_{topo} guides correct 2D topology construction through the following shaping rules:

T1. Connectivity Rule: Rewards reduction in molecular fragmentation by $(\text{cc}(G_t) - 1) - (\text{cc}(G_{t+1}) - 1)$, where $\text{cc}(\cdot)$ counts connected components.

T2. Edge Progression Rule: Rewards addition of target-matching edges by $|E^* \cap E_{t+1}| - |E^* \cap E_t|$, where E^* is the target edge set.

T3. Topological Distance Rule: Penalizes structural deviation via $-\text{GraphEditDistance}(G_{t+1}, G^*)$.

T4. Over-connection Rule: Prevents excessive edges by $-\alpha \cdot \max(0, |E_{t+1}| - |E^*|)$ with penalty coefficient α.

I.3 Potential-Based Reward Shaping

To accelerate learning while preserving optimal policy invariance (Ng et al., 1999), we apply potential-based shaping with potential function:

$$\Phi(s) = -|E^* \setminus E(s)| - \beta \cdot \text{cc}(G_s) \tag{20}$$

where β weights the connectivity term. The shaped reward is computed as $r'(s_t, a_t, s_{t+1}) = r(s_t, a_t, s_{t+1}) + \gamma\Phi(s_{t+1}) - \Phi(s_t)$.

J Prompt Templates and Downstream Task Training

This appendix provides the unified prompt templates used in MotifAgent for both initial training and downstream task fine-tuning, along with detailed descriptions of the training procedures.

J.1 Initial Training Prompt Template

During the initial training phase, MotifAgent learns molecular assembly through multi-agent collaboration. The following template defines the structured interaction format used throughout the training process.

J.1.1 Initial Setup Section

The prompt begins with comprehensive molecular information:

```
### Initial Setup
**Target Molecule**: {molecule_name}
- SMILES: `{smiles_string}`
- Target Properties: {property_description}, LogP ≈ {logp_value}, MW = {molecular_weight}
- PubChem CID: {cid} (if available)
**Fragmentation Results** (via BRICS):
- **Motif_1**: {motif_description} (`{motif_smiles}`, {n_atoms} atoms)
- **Motif_2**: {motif_description} (`{motif_smiles}`, {n_atoms} atoms)
...
- **Motif_n**: {motif_description} (`{motif_smiles}`, {n_atoms} atoms)
```

J.1.2 Episode Start Section

The episode begins with the initial environment state and global coordinator instruction:

```
### Episode Start: Multi-Agent Dialogue
**Environment State (t=0)**:
Current Assembly: Empty graph
Connected Components: 0
Available Motifs: [Motif_1, Motif_2, ..., Motif_n]
Target Edges: {n_edges} ({edge_descriptions})
Current Properties: None
**Global Coordinator**: "Agents, we need to assemble {molecule_name}. The target has
```

```
{property_description} with LogP around {logp_value}. Current topology is empty. Please propose your initial connections."
```

J.1.3 Round Structure

Each assembly round follows a consistent structure with a descriptive title:

```
### Round {N}: {Round_Title}
```

For subsequent rounds (t > 0), the environment state is updated:

```
**Environment State (t={step})**:
Current Assembly: {current_graph_description}
Connected Components: {num_components}
Remaining Motifs: [{remaining_motif_list}]
Remaining Target Edges: {num_remaining_edges}
Current Properties: {current_property_estimates}
```

J.1.4 Multi-Agent Proposal Section

Each motif agent generates proposals. For the initial round (t=0), agents use the full proposal format:

```
**Agent_{i} ({motif_name})**:
My Structure: {structural_description}
Available Sites: [{site_list}]
Chemical Context: {chemical_environment_description}
Proposal: {connection_proposal}
Reasoning: {chemical_reasoning}
Priority: {HIGH/MEDIUM/LOW} - {priority_explanation}
```

For subsequent rounds, agents may include status updates and modified fields:

```
**Agent_{i} ({motif_name})**:
Status Update: {occupied_sites_description}
Available Sites: [{remaining_site_list}]
Proposal: {connection_proposal}
Reasoning: {chemical_reasoning}
Chemical Insight: {additional_chemical_insight}
```

Alternatively, agents waiting for dependencies use:

```
**Agent_{i} ({motif_name})**:
Waiting State: {dependency_description}
Future Plan: {planned_action}
Property Prediction: {expected_property_impact}
```

J.1.5 Central Arbitrator Section

The central arbitrator evaluates all proposals and makes decisions:

```
**Central Arbitrator Evaluation**:
Chemical Validity Check:
- {connection_1}: ✓ Valid, {validity_reason}
- {connection_2}: ✓ Valid, {validity_reason}
Topological Scoring:
S({connection_1}) = {v1} (stability) + {v2} (progress) + {v3} (property) = {total_1}
S({connection_2}) = {v1} (stability) + {v2} (progress) + {v3} (property) = {total_2}
Decision: Execute {selected_connection} connection first
```

J.1.6 Action Execution and Reward Section

Each executed action follows the format:

```
**Action Executed**:
‘Connect: Motif_{i}[{site_x}] –{bond_type}-> Motif_{j}[{site_y}]‘
**Reward Calculation**:
r_valid = {value} ({explanation})
r_stable = {value} ({explanation})
```

```
r_func = {value} ({explanation})
r_conn = {value} ({explanation})
r_edge = {value} ({explanation})
Total: R = {total_reward}
```

For terminal states, additional rewards are included:

```
r_prop = {value} ({property_match_explanation})
Terminal Bonus = {value} (successful reconstruction)
Total: R = {total_reward}
```

J.1.7 Termination Section

Upon successful assembly:

```
### Round {N}: Termination Decision
**Environment State (t={final_step})**:
Current Assembly: Complete {molecule_name} molecule
Connected Components: 1
Target Edges Matched: {n}/{n} ✓
Properties: LogP = {value}, MW = {value}, {additional_properties}
**Global Coordinator**: "Assembly complete. All target edges matched. Properties within specifications."
**Agent_1 ({motif_name})**: "{confirmation_statement}"
**Agent_2 ({motif_name})**: "{confirmation_statement}"
...
**Policy Decision**: 'Action: STOP (probability = {prob})'
```

J.1.8 Chemical Explanation Section

The episode concludes with a generated chemical explanation:

```
**Chemical Explanation Generated**:
"{detailed_chemical_rationale_explaining_assembly_strategy_and_property_relationships}"
```

J.2 Downstream Task Prompt Templates

For downstream tasks, we extend the base template with task-specific components. All downstream tasks share the same initial setup and multi-agent interaction format, with task-specific output sections appended.

J.2.1 Molecular Property Prediction

For property prediction tasks, the prompt includes property-specific analysis phases:

```
### Initial Setup
**Task**: Molecular Property Prediction
**Property**: {property_name} (e.g., Blood-Brain Barrier Penetration)
**Target Molecule**:
- SMILES: '{smiles}'
- PubChem CID: {cid} (if available)
**Fragmentation Results** (via BRICS):
[Same format as initial training]
—
### Episode Start: Multi-Agent Analysis
[Standard environment state and global coordinator instruction]
[Multi-agent interaction rounds for topology reconstruction]
—
### Property Prediction Phase
**Environment State (Final)**:
Current Assembly: Complete molecule
Connected Components: 1
Target Edges Matched: {n}/{n} ✓
Reconstructed Topology: {topology_summary}
**Global Property Analyzer**:
Topology-Property Mapping:
- Identified pharmacophores: {pharmacophore_list}
```

- Key structural alerts: {alert_list}

- Electronic effects: {electronic_description}

- Steric factors: {steric_description}

Property-Contributing Features:

1. {feature_1}: {contribution_to_property}

2. {feature_2}: {contribution_to_property}

Prediction Output:

Property: {property_name}

Prediction: {0/1 or value}

Confidence: {score}

Key Evidence:

- {evidence_1}

- {evidence_2}

J.2.2 Molecular Description Generation

For description generation, the output section captures the assembly trace:

Initial Setup

Task: Molecular Description Generation

Target Molecule:

- SMILES: '{smiles}'

- Known Properties: {known_properties}

- PubChem CID: {cid}

Fragmentation Results (via BRICS):

[Same format as initial training]

—

Episode Start: Multi-Agent Assembly for Description

[Standard environment state]

Global Coordinator: "Agents, reconstruct this molecule step by step. Each connection should reveal structural insights for description generation. Focus on functional groups, pharmacophores, and structure-activity relationships."

[Multi-agent interaction rounds]

—

Description Generation Phase

Trajectory Summary:

Step 1: {action_description} → {chemical_insight}

Step 2: {action_description} → {chemical_insight}

...

Generated Description: "{natural_language_description_covering_chemical_structure_functional_groups_physicochemical_properties_and_biological_activities}"

J.2.3 Chemical Reaction Prediction

For reaction prediction tasks (forward reaction and retrosynthesis):

Initial Setup

Task: {Forward Reaction Prediction / Retrosynthesis}

Input (Forward):

- Reactants: '{reactant_smiles}'

- Reagents: {reagent_info}

Input (Retrosynthesis):

- Product: '{product_smiles}'

Fragmentation Results (via BRICS):

[Motif decomposition of reactants/products]

—

Multi-Agent Reaction Analysis

[Agents analyze reaction centers, bond changes, electron flow]

[Each agent identifies its role in the transformation]

—

Reaction Prediction Phase

Mechanism Analysis:

- Reaction center: {identified_centers}

```
- Bond breaking: {bonds_broken}
- Bond forming: {bonds_formed}
- Electron transfer pathway: {electron_flow_description}
**Predicted Output**: '{SELFIES_representation}'
**Chemical Explanation**: "{mechanistic_rationale}"
```

J.3 Downstream Task Training Procedures

We adopt a two-stage training paradigm: initial training on molecular reconstruction followed by task-specific fine-tuning. Due to computational constraints with the Qwen2.5-7B backbone, we employ parameter-efficient fine-tuning strategies rather than full model fine-tuning.

J.3.1 Training Strategy Overview

For all downstream tasks, we load the pre-trained MotifAgent checkpoint and apply task-specific adaptations. The initial training phase uses LoRA (rank=16) to train the shared LLM backbone, and this approach is maintained during downstream fine-tuning to preserve the learned chemical knowledge while adapting to specific tasks.

Molecular Property Prediction. This task employs supervised learning with frozen multi-agent interaction modules. We add a topology-aware classification head that pools LLM hidden states weighted by motif importance. Only the classification head parameters are trained, while the LLM backbone with LoRA adapters remains frozen. This preserves the pre-trained chemical understanding while learning task-specific decision boundaries. The training uses cross-entropy loss with auxiliary topology consistency regularization.

Molecular Description Generation. This task requires generating natural language outputs, necessitating adaptation of the language generation capabilities. We fine-tune the model using LoRA adapters (rank=8) on the LLM backbone combined with a task-specific generation head. The multi-agent assembly process first reconstructs the molecular topology, then the generation head produces descriptions conditioned on the assembly trace. Training uses teacher forcing with cross-entropy loss on the description tokens.

Chemical Reaction Prediction. Both forward reaction and retrosynthesis prediction require understanding molecular transformations at the motif level. We employ LoRA fine-tuning (rank=8) combined with reaction-specific output heads. The multi-agent framework identifies reaction centers by analyzing motif-level changes between reactants and products. For retrosynthesis, agents propose bond disconnections that reverse the assembly process. Training combines supervised learning on reaction outcomes with the MARL framework for exploring chemically valid transformation pathways.

J.3.2 Fine-tuning Configuration

All downstream tasks share the following base configuration with task-specific modifications:

- **Backbone**: Pre-trained MotifAgent (Qwen2.5-7B with LoRA)
- **LoRA Configuration**: rank=8, alpha=16, target modules: `[q_proj, v_proj]`
- **Optimizer**: AdamW with weight decay 0.01
- **Learning Rate**: 2e-5 for task heads, 1e-5 for LoRA parameters
- **Warmup**: 10% of total training steps
- **Precision**: BFloat16 mixed precision

For property prediction, we use batch size 32 and train for 50 epochs with early stopping (patience=10) based on validation ROC-AUC. For description generation and reaction prediction, we use batch size 16 and train for 30 epochs, monitoring BLEU scores and exact match rates respectively.

J.3.3 Inference Procedure

During inference, the model operates in reconstruction mode for understanding tasks (property prediction, description generation) and generation mode for synthesis tasks. In reconstruction mode, Set-BC priors guide the assembly to recover the target topology before applying the task-specific head. In generation mode for reaction prediction, beam search (k=5) explores multiple chemically valid pathways, with the central arbitrator ensuring chemical validity at each step.

The three-level explanation framework (topology construction, connection mechanisms, and topology-property relationships) is generated alongside task outputs, providing interpretable reasoning traces that connect molecular assembly decisions to final predictions.

J.4 Dataset Descriptions

This section provides detailed descriptions of all datasets used in our experiments, including statistics and data splitting strategies.

J.4.1 Initial Training Dataset

For initial training, we utilize the molecular SMILES-text pairs dataset from MoleculeSTM (Liu et al., 2023a), which is collected from the PubChem database. Following the preprocessing pipeline of MoleculeSTM, pairs with identical PubChem IDs are merged, descriptions shorter than 18 characters are filtered out, and duplicates from downstream task datasets are removed to prevent data leakage. This yields **51,340 unique high-quality molecule-description pairs** for initial training.

For each molecule, we apply our improved BRICS fragmentation algorithm with 16 chemical environment-specific bond-breaking rules. Molecules containing 2-15 motifs are selected for training to ensure appropriate complexity for the multi-agent assembly task. The dataset covers diverse chemical space including drug-like molecules, natural products, and synthetic compounds, providing comprehensive coverage of common motif connection patterns.

J.4.2 Molecular Property Prediction Datasets

We evaluate MotifAgent on 8 benchmark classification datasets from MoleculeNet (Wu et al., 2018). All datasets use **scaffold splitting** with an 80/10/10 train/validation/test ratio, which provides a more rigorous test of model generalizability by ensuring that molecules in different splits have distinct molecular scaffolds. Table 13 summarizes the dataset statistics.

BBBP (Blood-Brain Barrier Penetration): Contains binary labels indicating whether compounds can penetrate the blood-brain barrier, which is crucial for central nervous system drug development.

Tox21: Created by the "Toxicology in the 21st Century" initiative, this dataset contains qualitative toxicity measurements across 12 different targets including nuclear receptors and stress response pathways.

ToxCast: A larger toxicology dataset from the same initiative as Tox21, providing toxicity data across 617 high-throughput assays.

SIDER: Contains information about marketed drugs and their recorded adverse drug reactions, organized into 27 system organ classes.

ClinTox: Compares FDA-approved drugs with drugs that failed clinical trials for toxicity reasons, comprising two binary classification tasks.

MUV: Maximum Unbiased Validation dataset designed for virtual screening benchmarking, featuring challenging classification tasks with high class imbalance.

HIV: Contains experimental results on the ability of compounds to inhibit HIV replication.

BACE: Provides quantitative binding results for inhibitors of human β-secretase 1 (BACE-1), converted to binary classification.

J.4.3 Molecular Description Generation Dataset

For molecular description generation, we use the **ChEBI-20** benchmark dataset (Papadatos et al., 2015), which is derived from the Chemical Entities of Biological Interest (ChEBI) database combined with PubChem annotations. Table 14 shows the dataset statistics.

The ChEBI-20 dataset contains molecule-description pairs where each description has more than 20 words, providing detailed textual annotations covering chemical structures, functional groups, physicochemical properties, and biological activities. We follow the standard train/validation/test split provided by the original dataset.

J.4.4 Chemical Reaction Prediction Datasets

For chemical reaction prediction tasks (forward reaction prediction, retrosynthesis, and reagent prediction), we use the **Mol-Instructions** dataset (Fang et al., 2023). This dataset is part of a large-scale biomolecular instruction dataset designed for large language models, containing molecule-oriented instructions across six tasks.

All reaction data are represented using SELFIES (Self-Referencing Embedded Strings) format to ensure 100% validity of molecular representations. The dataset follows the default train/validation/test split provided by Mol-Instructions, with an approximate 80/10/10 ratio.

Forward Reaction Prediction: Given reactants and reagents, predict the products of the chemical reaction.

Retrosynthesis: Given a target product molecule, predict suitable reactants that could synthesize the target. This is the inverse problem of forward reaction prediction and represents a fundamental challenge in AI-assisted synthetic route

planning.

Reagent Prediction: Given reactants and products, predict the reagents or conditions required to facilitate the reaction.

J.4.5 Data Preprocessing for MotifAgent

For all datasets, molecules are processed through our unified preprocessing pipeline:

1. **BRICS Decomposition**: Each molecule is decomposed into chemically meaningful motifs using our improved BRICS algorithm with 16 bond-breaking rules. Motifs containing at least 2 non-hydrogen atoms are retained.

2. **Structured Text Serialization**: Each motif is serialized into structured text containing: unique identifier, SMILES string, atom and bond information, connectable sites with chemical environment annotations, and property summaries (aromaticity, ring structures, functional groups).

3. **Connection Graph Construction**: The molecular graph's connectivity information is preserved through explicit connection templates encoding adjacency relationships between motifs.

4. **Instruction Format Conversion**: Data is converted to task-specific instruction formats following the templates described in Appendix J.

Molecules that cannot be properly decomposed (e.g., those resulting in fewer than 2 or more than 15 motifs) are excluded from the respective datasets to ensure compatibility with the multi-agent framework.

K Cases Studies

In this section, we demonstrate the conversational and decision-making processes of MotifAgent through four representative cases. We first present a standard Algorithm Demonstration using the aspirin molecule, followed by detailed assembly workflows on three molecules of varying complexity: a simple molecule (Paracetamol), a complex molecule (Ibuprofen), and a complex heterocyclic molecule (Omeprazole).

Given the input molecular SMILES, target properties, and fragmentation results, each motif establishes its individual profile and analyzes its current state and relationships with other motifs. Through chemical validity assessment and topological scoring, agents evaluate the benefits of establishing inter-motif connections, execute connection actions, and compute corresponding rewards. After multiple rounds of negotiation, all motifs are successfully connected, followed by a comprehensive validation to ensure molecular validity and property satisfaction. The process concludes with the output of the final assembled molecule.

L Further Discussion

L.1 Central Arbitrator as a Safety Filter

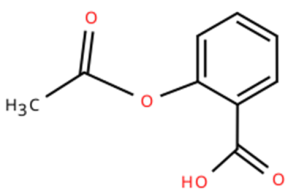
A possible misreading of Section 3.4 (and the priority-scoring function in Eq. 2) is that the central arbitrator carries the bulk of MotifAgent's decision-making capacity. This appendix clarifies why the arbitrator is more accurately characterized as a *safety filter* that operates at the level of action masking, analogous to constraint-based masking widely used in MARL systems, while the learned policy lives entirely inside each motif agent's network.

Policy learning is internal. The policy $\pi(a \mid s)$ (Eq. 1) is parameterized by the shared LLM backbone and trainable action heads, and it is updated by MAPPO with the full gradient signal. The arbitrator does not itself hold any trainable parameters: it (i) rejects proposals that violate valence or topology constraints using a deterministic rule engine, and (ii) applies the scoring function $S(a)$ of Eq. 2 only to break ties among the remaining valid proposals. Its effect on the gradient is therefore a shift in the action distribution rather than a change in the objective.

Fixed weights, adaptive effective contribution. We keep the priority-scoring weights fixed at $w_1 = w_2 = w_3 = 1.0$ throughout training. Despite the fixed weights, the effective contribution of each component varies naturally along the curriculum. In Phase 1, hard action masks constrain selections to the target-edge set, making TopoProgress the dominant signal. In Phase 3, the property predictor becomes sufficiently reliable for PropImprove to carry a larger effective weight. This is why we did not find it necessary to schedule the weights by hand.

Empirical decoupling. The per-component ablation in Table 7 removes one scoring component at a time and observes only a 2–3% drop in Validity. If the arbitrator carried the bulk of the learning signal, removing any single component would be expected to produce much larger regressions.

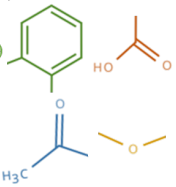
Case 1: Algorithm Demonstration: Multi-Agent Assembly of Aspirin



Input Molecule

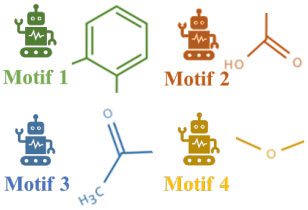
Initial Setup

Target Molecule : Aspirin ((Acetylsalicylic acid))
SMILES : CC(=O)Oc1ccccc1C(=O)O
Target Properties: Anti-inflammatory (COX inhibition), LogP ≈ 1.2, MW = 180.16
Fragmentation Results:
Motif 1: Benzene ring with two substituent sites (**c1ccccc1**, 6 atoms)
Motif 2: Carboxyl group (**C(=O)O**, 3 atoms)
Motif 3: Acetyl group (**CC(=O)**, 3 atoms)
Motif 4: Oxygen linker (**O**, 1 atom, acts as bridge)



Episode Start: Multi-Agent Dialogue

Environment State (t=0):
Current Assembly: Empty graph
Connected Components: 0
Available Motifs: [Motif_1, Motif_2, Motif_3, Motif_4]
Target Edges: 3 (benzene-carboxyl, benzene-oxygen, oxygen-acetyl)
Current Properties: None



Global Coordinator

Agents, we need to assemble **aspirin**. The target has **anti-inflammatory** properties with **LogP around 1.2**. **Current topology is empty.** Please propose your initial connections.

Round 1: Initial Core Selection

Motif 1 (Benzene)

My Structure: Aromatic ring, 6 carbons
Available Sites: [ortho_1, ortho_2, meta_1, meta_2, para]
Chemical Context: I'm a stable aromatic core, ideal as scaffold
Proposal: I should be the starting point. Ortho substitution pattern typically yields aspirin-like properties.
Priority: HIGH - I'm the largest motif and central scaffold

Motif 2 (Carboxyl)

My Structure: C(=O)O, acidic functional group
Available Sites: [carbon_connection]
Chemical Context: I provide acidity (pKa ~3-4) crucial for COX binding
Proposal: Connect to Agent_1's ortho position for salicylic acid core
Reasoning: Ortho-hydroxybenzoic acid is the pharmacophore base
Priority: HIGH - Essential for biological activity

Motif 3 (Acetyl)

My Structure: CC(=O), ester-forming group
Available Sites: [carbonyl_carbon]
Chemical Context: I can form esters, improving stability over free phenol
Proposal: Wait for oxygen bridge, then form ester linkage
Reasoning: Direct connection to benzene would create ketone, not ester
Priority: MEDIUM - Need linker first

Motif 4 (Oxygen)

My Structure: Single oxygen atom, bridging unit
Available Sites: [connection_1, connection_2]
Chemical Context: I enable ester formation, crucial for aspirin vs salicylic acid
Proposal: Bridge between Agent_1 and Agent_3 after carboxyl attachment
Reasoning: Ester linkage masks the phenol, reducing GI irritation
Priority: MEDIUM - Sequential dependency

Central Arbitrator Evaluation

Chemical Validity Check:
- Benzene-Carboxyl (ortho): ✓ Valid, maintains aromaticity
- Benzene-Oxygen (direct): ✓ Valid, forms phenolic ether
- Acetyl-Oxygen: ✓ Valid, forms ester

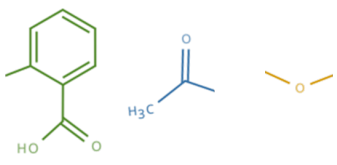
Topological Scoring:
S(benzene→carboxyl) = 0.8 (stability) + 0.9 (progress) + 0.7 (property) = 2.4
S(benzene→oxygen) = 0.7 (stability) + 0.6 (progress) + 0.3 (property) = 1.6

Decision: Execute benzene-carboxyl connection first

Action Executed: Connect: Motif_1[ortho_1] --single--> Motif_2[carbon]

Reward Calculation:

r_valid = +1.0 (valid connection)
r_stable = -0.1 (minimal strain)
r_func = +0.5 (formed benzoic acid substructure)
r_conn = +1.0 (created first connected component)
r_edge = +1.0 (matched target edge)
Total: R = 3.4



Round 2: Completing the Pharmacophore

Motif 1 (Benzene)

Status Update: Ortho_1 occupied by carboxyl
Available Sites: [ortho_2, meta_1, meta_2, para]
Proposal: Accept oxygen at ortho_2 for ester formation
Reasoning: Ortho relationship between COOH and OR creates aspirin scaffold
Chemical Insight: This positioning enables intramolecular H-bonding

Environment State (t=1):
Current Assembly: Benzene-COOH (salicylic acid core)
Connected Components: 1
Remaining Motifs: [Motif_3, Motif_4]
Remaining Target Edges: 2
Current Properties: pKa ≈ 3.0, partial COX affinity

Motif 3 (Acetyl)

Waiting State: Need oxygen bridge established first
Future Plan: Will form ester bond through oxygen
Property Prediction: Ester will increase LogP to ~1.2 (target range)

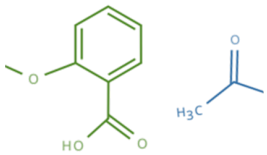
Motif 4 (Oxygen)

Analysis: Benzene has ortho_2 available, perfect for aspirin pattern
Proposal: Connect to Motif_1[ortho_2]
Reasoning: Sets up ester formation with acetyl group
Compatibility: Maintains sp2 oxygen, allows resonance with benzene

Action Executed: Connect: Motif_1[ortho_2] --single--> Motif_4[connection_1]

Reward Calculation:

r_valid = +1.0
r_stable = -0.2 (slight ortho crowding)
r_topo = +0.8 (approaching target topology)
r_edge = +1.0 (matched target edge)
Total: R = 2.6



Round 3: Final Assembly

Motif 4 (Oxygen)

Status: One connection used (to benzene)
Available Sites: [connection_2]
Proposal: Ready to accept acetyl group
Chemical Check: Ester formation favorable (ΔG < 0)

Environment State (t=2):
Current Assembly: ortho-hydroxyl benzoic acid
Connected Components: 1
Remaining Motifs: [Motif_3]
Remaining Target Edges: 1
Current Properties: LogP ≈ 0.8, incomplete pharmacophore

Motif 3 (Acetyl)

Final Move: Connect to oxygen to complete aspirin
Proposal: Motif_3[carbonyl] --> Motif_4[connection_2]
Property Impact:
- Masks phenolic OH (reduces GI irritation)
- Increases LogP to target range
- Maintains COX inhibition through hydrolysis

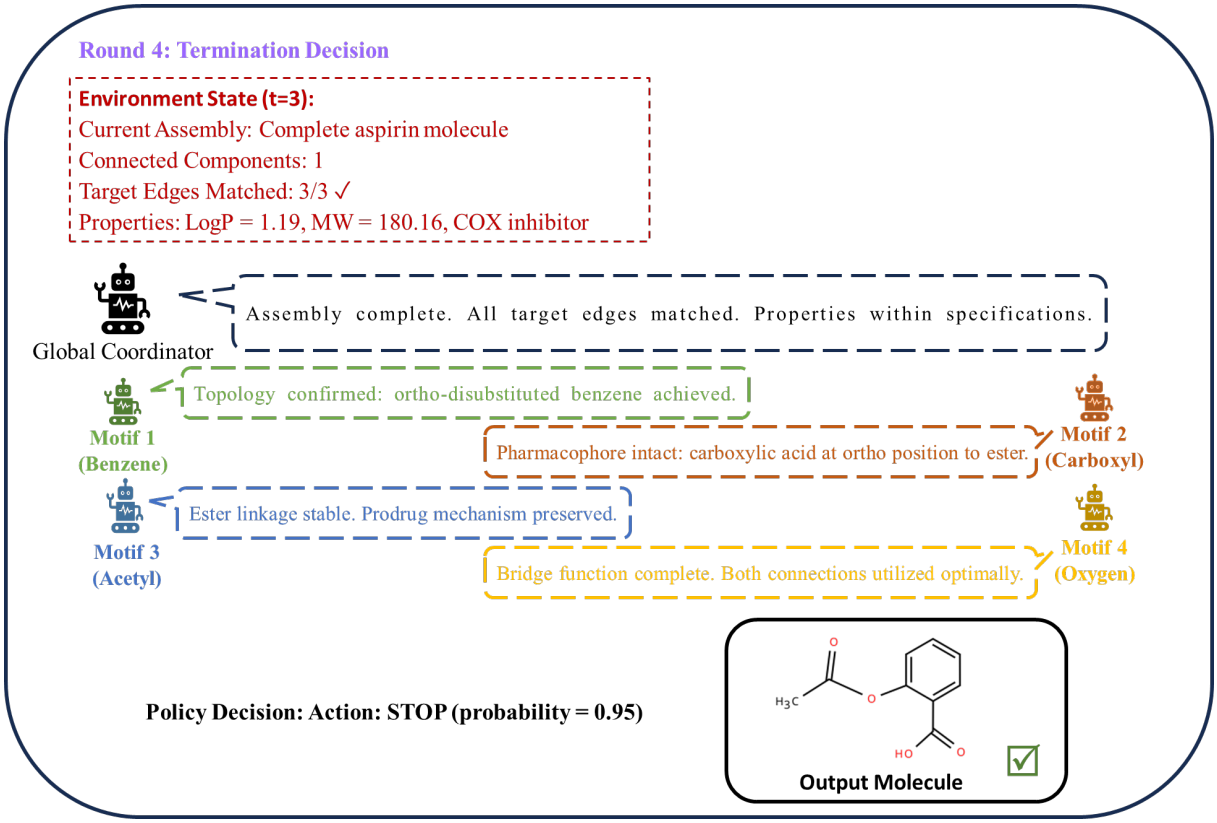
Global Property Predictor

If connection completed:
- MW: 180.16 ✓ (matches target)
- LogP: ~1.19 ✓ (within range)
- H-bond donors: 1 ✓ (COOH only)
- H-bond acceptors: 4 ✓ (typical for aspirin)
- Bioactivity: High confidence for COX inhibition

Final Reward Calculation:

r_valid = +1.0
r_stable = -0.05 (stable ester)
r_func = +1.0 (completed aspirin structure)
r_prop = +0.9 (LogP within 0.01 of target)
r_edge = +1.0 (final target edge)
r_conn = 0 (already connected)
Terminal Bonus = +5.0 (successful reconstruction)
Total: R = 8.85

Action Executed: Connect: Motif_4[connection_2] --single--> Motif_3[carbonyl]



Episode Summary

Trajectory Interpretation:

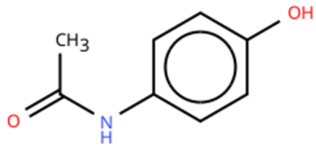
Step 1: Established salicylic acid core (benzene-COOH)
→ Created primary pharmacophore
Step 2: Added oxygen at ortho position
→ Prepared ester formation site
Step 3: Completed acetyl ester
→ Achieved target molecule with desired properties

Key Insights Learned:

- Ortho substitution pattern critical for aspirin activity
- Assembly order: scaffold → functional groups → modifiers
- Ester formation requires bridging oxygen (not direct connection)
- Properties emerge from specific topology (ortho relationship)

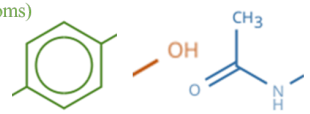
Chemical Explanation Generated: The assembly successfully reconstructed aspirin through strategic ortho-substitution on benzene. The carboxyl group provides COX binding affinity while the acetyl ester masks the phenolic hydroxyl, reducing gastric irritation while maintaining efficacy through in vivo hydrolysis. The learned assembly pattern prioritizes pharmacophore establishment before property optimization.

Case 2: Paracetamol (Acetaminophen) - Simple Molecule



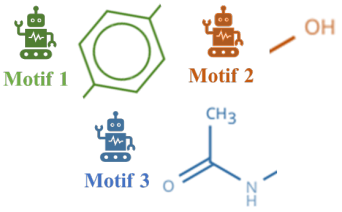
Input Molecule

Initial Setup
Target Molecule : Paracetamol
SMILES : CC(=O)Nc1ccc(O)cc1
Target Properties: Analgesic/antipyretic, LogP ≈ 0.46, MW = 151.16
PubChem CID: 1983
Fragmentation Results:
Motif 1: para-disubstituted benzene (c1ccc(-)cc1, 6 atoms)
Motif 2: Hydroxyl group (O, 1 atom)
Motif 3: Acetamide group (CC(=O)N, 4 atoms)



Episode Start: Multi-Agent Dialogue

Environment State (t=0):
Current Assembly: Empty graph
Connected Components: 0
Available Motifs: [Motif_1, Motif_2, Motif_3]
Target Edges: 2 (benzene-hydroxyl, benzene-acetamide)
Current Properties: None
Target: Analgesic with low LogP for CNS penetration



Global Coordinator

Agents, we need to assemble **paracetamol**. The target has **analgesic properties** with **LogP around 0.46**. Current topology is empty. Please propose your initial connections.

Round 1: Initial Core Selection

Motif 1 (Benzene)

My Structure: Aromatic ring, 6 carbons, para-substitutable
Available Sites: [para_1, para_2, ortho_1, ortho_2, meta_1, meta_2]
Chemical Context: I'm the scaffold for most analgesics
Proposal: I should be the starting point. Para substitution reduces toxicity compared to ortho/meta patterns
Priority: HIGH - I'm the core scaffold

Motif 3 (Acetamide)

My Structure: CC(=O)N, metabolically stable group
Available Sites: [nitrogen_connection]
Chemical Context: I provide metabolic stability vs acetyl ester
Proposal: Connect to Agent_1's other para position
Reasoning: Amide more stable than ester in vivo
Priority: HIGH - Prevents toxic metabolite formation

Motif 2 (Hydroxyl)

My Structure: Single OH group
Available Sites: [oxygen_connection]
Chemical Context: I provide H-bonding for receptor interaction
Proposal: Connect to Agent_1's para position for phenolic structure
Reasoning: Phenol group essential for COX interaction
Priority: HIGH - Key pharmacophore element

Central Arbitrator Evaluation

Chemical Validity Check:
- Benzene-Hydroxyl (para): ✓ Valid, forms p-hydroxyphenyl
- Benzene-Acetamide (para): ✓ Valid, forms p-acetamidophenyl

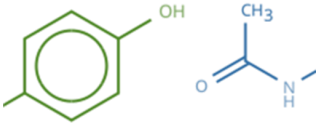
Topological Scoring:
S(benzene→hydroxyl) = 0.9 (stability) + 0.8 (progress) + 0.9 (property) = 2.6
S(benzene→acetamide) = 0.8 (stability) + 0.8 (progress) + 0.7 (property) = 2.3

Decision: Execute benzene-hydroxyl connection first

Action Executed: Connect: Motif_1[para_1] --single--> Motif_2[oxygen]

Reward Calculation:

r_valid = +1.0 (valid connection)
r_stable = -0.05 (minimal strain, para position)
r_func = +0.8 (formed phenolic group)
r_conn = +1.0 (created first connected component)
r_edge = +1.0 (matched target edge)
Total: R = 3.75



Round 2: Completing the Structure

Environment State (t=1):
Current Assembly: p-hydroxyphenyl
Connected Components: 1
Remaining Motifs: [Motif_3]
Remaining Target Edges: 1
Current Properties: pKa ≈ 10 (phenolic), partial analgesic activity

Motif 1 (Benzene)

Status Update: para_1 occupied by hydroxyl
Available Sites: [para_2, ortho_1, ortho_2, meta_1, meta_2]
Proposal: Accept acetamide at para_2
Reasoning: Creates p-hydroxyacetanilide (paracetamol)
Chemical Insight: Electron donation from both groups activates ring safely

Motif 3 (Acetamide)

Analysis: Para position to hydroxyl is optimal
Proposal: Connect to Motif_1[para_2]
Property Prediction: Will achieve target LogP ~0.46
Mechanism: Amide prevents formation of toxic NAPQI at normal doses

Action Executed: Connect: Motif_1[para_2] --single--> Motif_3[nitrogen]

Reward Calculation:
r_valid = +1.0
r_stable = -0.05 (no steric clash in para)
r_func = +1.0 (completed paracetamol structure)
r_prop = +0.9 (LogP = 0.47, within target)
r_edge = +1.0 (matched final target edge)
Terminal Bonus = +5.0 (successful reconstruction)
Total: R = 8.85

Round 3: Termination Decision

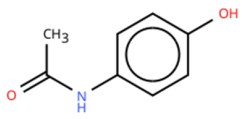
Environment State (t=2):
Current Assembly: Complete paracetamol molecule
Connected Components: 1
Target Edges Matched: 2/2 ✓
Properties: LogP = 0.47, MW = 151.16, pKa = 9.38

Global Coordinator

Assembly complete. All target edges matched. Properties within specifications.

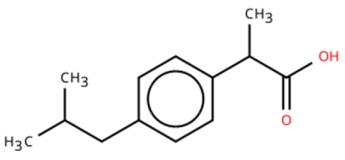
Policy Decision: Action: STOP (probability = 0.96)

Chemical Explanation Generated: Paracetamol assembled through para-substitution pattern. The phenolic hydroxyl and acetamide groups in para positions provide optimal balance of activity and safety, avoiding toxic quinone imine formation seen with other substitution patterns.



Output Molecule

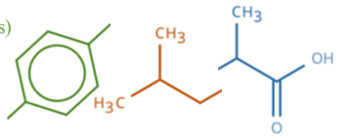
Case 3: Ibuprofen - Complex Molecule



Input Molecule

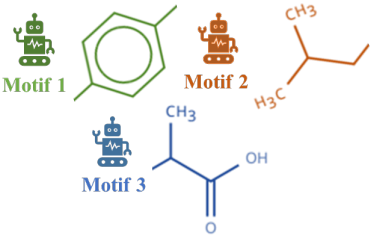
Initial Setup

Target Molecule : Ibuprofen
SMILES : CC(C)Cc1ccc(C(C)C(=O)O)cc1
Target Properties: NSAID, LogP ≈ 3.97, MW = 206.28
PubChem CID: 3672
Fragmentation Results:
Motif 1: para-disubstituted benzene (c1ccc(-)cc1, 6 atoms)
Motif 2: Isobutyl group (CC(C)C, 4 atoms)
Motif 3: α-methylpropionic acid (C(C)C(=O)O, 4 atoms)



Episode Start: Multi-Agent Dialogue

Environment State (t=0):
Current Assembly: Empty graph
Connected Components: 0
Available Motifs: [Motif_1, Motif_2, Motif_3]
Target Edges: 2 (benzene-isobutyl, benzene-propionic)
Current Properties: None
Target: NSAID with high LogP for membrane penetration



Global Coordinator

Agents, we need to assemble **ibuprofen**. The target is an **NSAID** with **LogP around 3.97** for **optimal membrane penetration**. Please propose your initial connections.

Round 1: Initial Core Selection

Motif 1 (Benzene)

My Structure: Aromatic ring with para substitution sites
Available Sites: [para_1, para_2, ortho_1, ortho_2, meta_1, meta_2]
Chemical Context: Central scaffold for propionic acid NSAIDs
Proposal: I should anchor the assembly
Priority: HIGH - I'm the core connecting both alkyl groups

Motif 3 (α-Methy-lpropionic acid)

My Structure: C(C)C(=O)O, chiral center at α-position
Available Sites: [alpha_carbon]
Chemical Context: I'm the primary pharmacophore for COX inhibition
Proposal: Connect to benzene para (opposite to isobutyl)
Reasoning: α-methyl provides COX-2 selectivity
Priority: HIGH - Essential for NSAID activity

Motif 2 (Isobutyl)

My Structure: CC(C)C, branched alkyl chain
Available Sites: [primary_carbon]
Chemical Context: I increase lipophilicity and membrane permeability
Proposal: Connect to benzene's para position
Reasoning: Creates hydrophobic pocket for COX active site
Priority: MEDIUM - Important for selectivity but not primary pharmacophore

Central Arbitrator Evaluation

Chemical Validity Check:
- Benzene-Isobutyl: ✓ Valid, increases lipophilicity
- Benzene-Propionic: ✓ Valid, forms arylpropionic acid

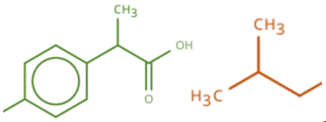
Topological Scoring:
S(benzene→propionic) = 0.8 (stability) + 0.9 (progress) + 1.0 (property) = 2.7
S(benzene→isobutyl) = 0.9 (stability) + 0.7 (progress) + 0.6 (property) = 2.2

Decision: Execute benzene-propionic connection first

Reward Calculation:

r_valid = +1.0 (valid connection)
r_stable = -0.1 (slight steric from α-methyl)
r_func = +1.0 (formed arylpropionic acid pharmacophore)
r_conn = +1.0 (created first connected component)
r_edge = +1.0 (matched target edge)
Total: R = 3.9

Action Executed: Connect: Motif_1[para_1] --single--> Motif_3[alpha_carbon]



Round 2: Completing Lipophilic Modification

Environment State (t=1):
Current Assembly: p-methylphenylpropionic acid
Connected Components: 1
Remaining Motifs: [Motif_2]
Remaining Target Edges: 1
Current Properties: pKa ≈ 4.5, LogP ≈ 2.1, partial COX activity

Motif 1 (Benzene)

Status Update: para_1 occupied by propionic acid
Available Sites: [para_2, ortho_1, ortho_2, meta_1, meta_2]
Proposal: Accept isobutyl at para_2
Reasoning: Para relationship minimizes steric clash
Chemical Insight: This completes the "profen" scaffold

Analysis: Para to propionic acid is ideal positioning
Proposal: Connect to Motif_1[para_2]
Property Impact: Will increase LogP to ~3.97 (target)
COX Interaction: Forms hydrophobic channel interaction

Motif 2 (Isobutyl)

Global Property Predictor

If connection completed:
- MW: 206.28 ✓ (matches target)
- LogP: ~3.96 ✓ (optimal for oral absorption)
- pKa: 4.4 ✓ (good tissue penetration)
- Rotatable bonds: 4 ✓ (acceptable flexibility)

Reward Calculation:
r_valid = +1.0
r_stable = -0.05 (minimal strain in para)
r_func = +0.5 (enhanced lipophilicity)
r_prop = +0.95 (LogP = 3.96, nearly perfect)
r_edge = +1.0 (matched final target edge)
Terminal Bonus = +5.0 (successful reconstruction)
Total: R = 8.4

Action Executed: Connect: Motif_1[para_2] --single--> Motif_2[primary_carbon]

Round 3: Termination Decision

Environment State (t=2):
Current Assembly: Complete ibuprofen molecule
Connected Components: 1
Target Edges Matched: 2/2 ✓
Properties: LogP = 3.96, MW = 206.28, COX-2 selective

Global Coordinator

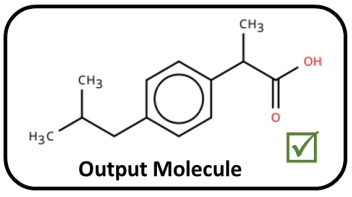
Assembly complete. Ibuprofen structure achieved with optimal lipophilicity.

Motif 3 (Propionic acid)

The α-methyl substitution pattern is crucial for reduced GI side effects compared to non-selective NSAIDs.

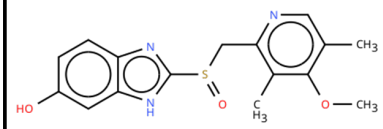
Policy Decision: Action: STOP (probability = 0.94)

Chemical Explanation Generated: Ibuprofen assembled with para-disubstituted benzene pattern. The α-methylpropionic acid provides COX inhibition while the isobutyl group enhances membrane penetration and selectivity. The high LogP enables effective tissue distribution.



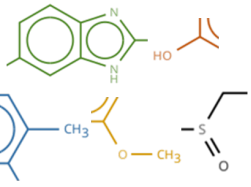
Output Molecule

Case 4: Omeprazole - Complex Heterocyclic Molecule



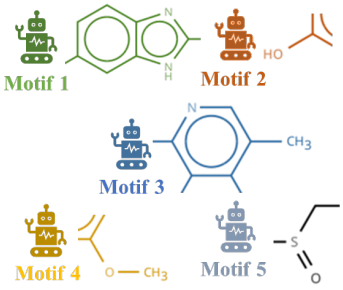
Input Molecule

Initial Setup
Target Molecule : Omeprazole
SMILES : Oc1ccc2nc(S(=O)Cc3ncc(C)c(OC)c3C)[nH]c2c1
Target Properties: Proton pump inhibitor, LogP ≈ 2.23, MW = 345.42
PubChem CID: 4594
Fragmentation Results:
Motif 1: Benzimidazole core (c1ccc2[nH]c(-)nc2c1, 9 atoms)
Motif 2: Methoxy group (OC, 2 atoms)
Motif 3: 3-methylpyridine (c1ncc(C)cc1, 7 atoms)
Motif 4: Methoxy-methyl fragment (COC, 3 atoms)
Motif 5: Sulfinylmethyl bridge (S(=O)C, 3 atoms)



Episode Start: Multi-Agent Dialogue

Environment State (t=0):
Current Assembly: Empty graph
Connected Components: 0
Available Motifs: [Motif_1, Motif_2, Motif_3, Motif_4, Motif_5]
Target Edges: 4 (benzimidazole-methoxy, benzimidazole-sulfoxide, sulfoxide-pyridine, pyridine-methoxymethyl)
Current Properties: None
Target: PPI requiring acid activation, LogP ≈ 2.23



Global Coordinator

Agents, we need to assemble omeprazole with 5 motifs. This proton pump inhibitor requires precise heterocycle arrangement with a sulfoxide bridge. Please propose your initial connections.

Round 1: Initial Core Selection

Motif 1 (Benzimidazole)
My Structure: Fused heterocycle, 9 atoms
Available Sites: [C2_imidazole, C5_position]
Chemical Context: I'm the primary PPI pharmacophore, pKa ~6
Proposal: Start with me as foundation - need C5 methoxy for activity
Reasoning: My C2 position will later accept sulfoxide bridge
Priority: HIGH - Core pharmacophore

Motif 2 (Methoxy)
My Structure: OC, electron-donating group
Available Sites: [oxygen]
Chemical Context: I modulate benzimidazole electronics
Proposal: Attach to Motif_1 at C5 position first
Reasoning: Increases electron density for stability
Priority: HIGH - Essential for benzimidazole activation

Motif 3 (3-Methylpyridine)
My Structure: Pyridine with methyl at position 3
Available Sites: [C2, C4, C5]
Chemical Context: I need further decoration for full activity
Proposal: Wait - need my additional substituents
Priority: LOW - Require complete substitution pattern

Motif 4 (Methoxy-methyl)
My Structure: COC fragment
Available Sites: [carbon, oxygen]
Chemical Context: I complete pyridine substitution
Proposal: Will attach to pyridine later
Priority: MEDIUM - Needed for complete structure

Motif 5 Sulfinyl methyl Bridge
My Structure: S(=O)C, chiral sulfoxide with methyl
Available Sites: [sulfur, carbon]
Chemical Context: I'm the critical bridge and activation site
Proposal: Connect to benzimidazole after it's activated
Reasoning: Sulfoxide enables acid-triggered mechanism
Priority: HIGH - Essential bridge element

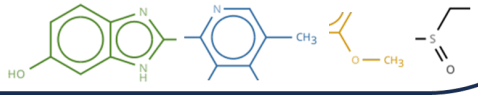
Central Arbitrator Evaluation

Chemical Validity Check:
- Benzimidazole-Methoxy (C5): ✓ Valid, enhances basicity
- Benzimidazole-Sulfoxide (C2): ✓ Valid, typical PPI pattern

Topological Scoring:
S(benzimidazole→methoxy) = 0.9 (stability) + 0.8 (progress) + 0.9 (property) = 2.6
S(benzimidazole alone) = 0.8 (stability) + 0.7 (progress) + 0.6 (property) = 2.1
Decision: Connect benzimidazole with methoxy first

Reward Calculation:
r_valid = +1.0 (valid connection)
r_stable = -0.05 (minimal strain)
r_func = +0.8 (enhanced benzimidazole basicity)
r_conn = +1.0 (created first connected component)
r_edge = +1.0 (matched target edge)
Total: R = 3.75

Action Executed: Connect: Motif_1[C5_position] --single--> Motif_2[oxygen]



Round 2: Sulfoxide Bridge Formation

Environment State (t=1):
Current Assembly: 5-methoxybenzimidazole
Connected Components: 1
Remaining Motifs: [Motif_3, Motif_4, Motif_5]
Remaining Target Edges: 3
Current Properties: pKa ~6.2, partial PPI activity

Motif 1 (Benzimidazole)

Status Update: C5 occupied by methoxy
Available Sites: [C2_imidazole]
Proposal: Accept sulfoxide bridge at C2
Reasoning: This is the key PPI connection
Chemical Insight: Electron flow from methoxy stabilizes sulfoxide

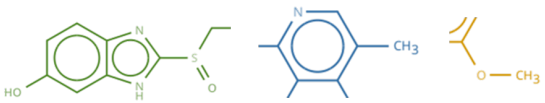
Motif 5 Sulfinylmethyl Bridge

Analysis: Benzimidazole C2 properly activated
Proposal: Connect my sulfur to Motif_1[C2]
Mechanism: My sulfoxide will undergo Pummerer rearrangement in acid
Property Impact: Enables tissue-selective activation

Action Executed: Connect: Motif_1[C2_imidazole] --single--> Motif_5[sulfur]

Reward Calculation:

r_valid = +1.0
r_stable = -0.15 (sulfoxide somewhat reactive)
r_func = +1.1 (formed key PPI intermediate)
r_conn = 0 (already connected)
r_edge = +1.0 (matched target edge)
Total: R = 2.95



Round 3: Pyridine Decoration

Motif 3 (3-Methylpyridine)

Status: Need decoration before main chain connection
Available Sites: [C2, C4, C5]
Proposal: Accept Motif_4 at my C4/C5 positions
Reasoning: Creates 3,5-dimethyl-4-methoxy pattern

Motif 4 (Methoxy-methyl)

Analysis: Should create branched substitution on pyridine
Proposal: Connect to Motif_3 - my carbon to C5, oxygen bridges to C4
Chemical Impact: Completes electron-rich pyridine
Note: This creates the specific substitution needed for activity

Environment State (t=2):
Current Assembly: Benzimidazole-methoxy with sulfinylmethyl attached
Connected Components: 1
Remaining Motifs: [Motif_3, Motif_4]
Remaining Target Edges: 2
Current Properties: Basic scaffold formed, needs completion

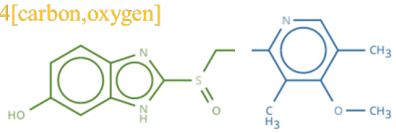
Central Arbitrator

Special Case: Motif_4 bridges two positions on pyridine
Creating: 3,5-dimethyl-4-methoxy substitution pattern
Decision: Execute complex bridging connection

Reward Calculation:

r_valid = +1.0
r_stable = -0.1 (some steric crowding)
r_func = +0.9 (completed pyridine substitution)
r_conn = 0 (new component formed)
r_edge = +1.0 (matched target pattern)
Total: R = 2.8

Action Executed: Connect: Motif_3[C5,C4] --bridge--> Motif_4[carbon,oxygen]



Round 4: Final System Unification

Motif 3 (3-Methylpyridine)

Status: Fully decorated with dimethyl-methoxy pattern
Available Sites: [C2]
Proposal: Accept bridge at C2 position
Mechanism Impact: Creates full activation cascade:
- Pyridine protonation (pKa ~4)
- Sulfoxide rearrangement
- Covalent pump inhibition

Motif 5 (Sulfinyl methyl Bridge)

Status: Sulfur connected to benzimidazole
Available Sites: [methyl_carbon]
Proposal: Connect my carbon to pyridine C2
Function: Complete the heterocycle bridge

Environment State (t=2):
Current Assembly: Two separate systems
System 1: 5-methoxybenzimidazole-S(=O)C
System 2: 3,5-dimethyl-4-methoxypyridine
Connected Components: 2
Remaining Target Edges: 1 (bridge connection)
Current Properties: Both pharmacophores ready

Global Property Predictor

If connection completed:
- MW: 345.42 ✓ (matches target)
- LogP: ~2.23 ✓ (optimal oral absorption)
- Dual basic sites for accumulation
- Sulfoxide bridge for activation
- Complete PPI pharmacophore

Action Executed: Connect: Motif_5[carbon] --single--> Motif_3[C2]

Reward Calculation:

r_valid = +1.0
r_stable = -0.2 (complex but stable system)
r_func = +1.5 (completed PPI structure)
r_prop = +0.95 (LogP = 2.24, excellent match)
r_conn = +2.0 (unified into single component)
r_edge = +1.0 (final target edge)
r_topo = +1.0 (correct topology achieved)
Terminal Bonus = +5.0 (successful reconstruction)
Total: R = 12.25

Round 5: Termination Decision

Environment State (t=3):
Current Assembly: Complete omeprazole molecule
Connected Components: 1
Target Edges Matched: 4/4 ✓
Properties: LogP = 2.24, MW = 345.42, Prodrug mechanism intact
Substitution Pattern: 5-methoxybenzimidazole, 3,5-dimethyl-4-methoxypyridine

Global Coordinator: Assembly complete. All 5 motifs successfully connected. Omeprazole structure achieved with correct substitution patterns.

Motif 1 (Benzimidazole): 5-methoxy substitution enhances my electron density, stabilizing the sulfoxide for controlled activation.

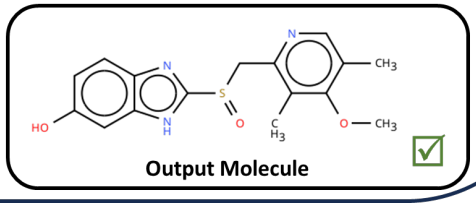
Motif 2 (Methoxy): Electron donation to benzimidazole optimized for pKa ~6.

Motif 3 (Pyridine): Complete 3,5-dimethyl-4-methoxy pattern achieved. This substitution is critical for both stability and selectivity.

Motif 4 (Methoxy-methyl): Successfully decorated pyridine with required substitution pattern.

Motif 5 Sulfinylmethyl: Bridge formed correctly. My chiral sulfoxide center will undergo acid-catalyzed rearrangement to form the active sulfenamide.

Policy Decision: Action: STOP (probability = 0.97)



Episode Summary

Step 1: Benzimidazole + methoxy (electron activation)
Step 2: Added sulfinylmethyl bridge to C2
Step 3: Decorated pyridine with methoxy-methyl fragment
Step 4: United systems via bridge completion
Step 5: Termination
Total Reward: 22.75 (across 4 connection steps)
Assembly Efficiency: 100% (all motifs correctly incorporated)
Property Achievement: LogP 2.24 vs 2.23 target (99.6% accuracy)
Key Insights:
- 5-methoxy on benzimidazole crucial for stability
- Sulfoxide bridge enables pH-dependent activation
- 3,5-dimethyl-4-methoxy pyridine pattern essential for selectivity
- Assembly order: core activation → bridge formation → decoration → unification

Chemical Explanation Generated: Omeprazole assembled through strategic 5-motif construction. Starting from benzimidazole core, we added 5-methoxy for electron activation, attached the sulfinylmethyl bridge at C2, separately decorated the pyridine with methoxy-methyl to achieve 3,5-dimethyl-4-methoxy pattern, then united both systems through the bridge. The assembly demonstrates how PPIs are built from heterocyclic cores with precise substitution patterns: benzimidazole for enzyme binding, sulfoxide for acid-activated prodrug mechanism, and electron-rich pyridine for selective accumulation in acidic compartments. The multi-step assembly preserves the critical structural features while building complexity systematically.

The observed insensitivity is consistent with our characterization: the policy network internalizes the connection rules, and the arbitrator's role is to keep the rollout trajectories chemically feasible so that the MAPPO update is well-conditioned, not to supply the chemistry itself. This also explains why the progressive ablation in Table 4 shows the Single-Agent-with-RL variant reaching only 82.3% — without multi-agent decomposition, the arbitrator has nothing to coordinate and the policy alone is forced to balance all sub-decisions in one context.

L.2 Failure Mode Analysis

This appendix expands the failure taxonomy introduced in Section 4.6. We examine the test-set errors of the Full MotifAgent across the four categories, report their approximate frequencies, and walk through a representative case for the most common one.

(1) Positional Isomer Confusion. Two candidate sites that share nearly indistinguishable local chemical environments (e.g., two symmetric positions on an aromatic ring relative to a distant functional group) cause the policy to pick the wrong one. The reward signal is weak because the resulting molecule is typically chemically valid and has similar bulk properties. This is the most common category in our error analysis.

(2) Bond-Type Misassignment. Site selection is correct but the bond head assigns the wrong bond order (e.g., single versus aromatic for a partially aromatic ring closure). These errors frequently occur in hetero-aromatic systems where BRICS leaves an aromaticity hint but not a fully constrained bond type.

(3) Incomplete Assembly / Premature Termination. The policy emits `STOP` before all required connections are made. This typically happens when the PropImprove score plateaus for k steps and the convergence termination condition fires early; a less aggressive ε mitigates the issue but slows training.

(4) Chemical Validity Violations. Residual valence or aromaticity violations that slip past the arbitrator's rule engine, mostly caused by rare BRICS fragmentation patterns unseen during training.

Representative case: vanillin. The target molecule is vanillin (4-hydroxy-3-methoxybenzaldehyde, `O=Cc1ccc(O)c(OC)c1`); MotifAgent produces isovanillin (3-hydroxy-4-methoxybenzaldehyde), which swaps the `OH` and `OMe` groups. Inspecting the policy logits at the decisive step shows $\pi(\text{C3}) = 0.47$ vs. $\pi(\text{C4}) = 0.53$, a margin of only 0.06; the priority-scoring function returns nearly identical ChemStability, TopoProgress, and PropImprove scores for the two candidate sites because the sole discriminating context — the distal aldehyde at C1 — fails to break the local symmetry of the aromatic ring. The error metrics are: Exact Match ✗;

Morgan FTS 0.72; GED 2; Conn. Site Acc. 1/3; Bond Type Acc. 3/3; chemical validity ✓. Because isovanillin itself is a valid molecule with properties close to the target (LogP ≈ 1.18 vs. target 1.21), the RL signal can only weakly penalize this failure mode.

Mitigation. The common thread across category (1) cases is that the current textual state representation cannot disambiguate positions that are equivalent under local chemical environment. We therefore identify two complementary extensions as future work: (a) augment the state with explicit positional encodings (e.g., ring-position indices relative to a global canonical ordering) and (b) use a lightweight graph-based site disambiguation head that scores candidate sites by their global structural context rather than their local chemical environment.

L.3 Role of the LLM Backbone

This appendix elaborates on why using an LLM — rather than a graph-native encoder — as the policy backbone yields MotifAgent's observed advantages. Our view is that the LLM contributes four distinct affordances, corresponding to four empirically observable effects.

(1) Encoded chemical priors. The vanilla Qwen2.5-7B baseline in Table 4 reaches 67.6% Validity with no BRICS, no multi-agent structure, and no RL, confirming that pre-training already encodes non-trivial chemistry (amide formation, ester hydrolysis, aromaticity maintenance). A graph-native encoder has to learn these priors from scratch, typically through task-specific supervision.

(2) Cross-task unification. A single LLM backbone serves property prediction, description generation, and reaction prediction under a shared interface — arriving at each task only by swapping the task head and the decoding template. Graph-native approaches generally require per-task architectures (e.g., contrastive pre-training for retrieval, autoregressive decoders for generation, classification heads for property prediction), which fragments the engineering surface.

(3) Interpretable reasoning chains. Because each motif agent's decision is conditioned on a textual prompt, the same LLM forward pass can emit a human-readable chemical explanation of why a given connection was proposed. Such explanations are critical for medicinal chemists who need to audit assembly decisions before acting on them; GNN-based scoring functions, by contrast, do not expose intermediate chemical reasoning.

(4) Activation of latent knowledge. Table 1 shows that MotifAgent-small (3B parameters) outperforms vanilla Qwen2.5-7B. Since both share the same chemical knowledge distribution in their pre-training corpus, the difference cannot be attributed to scale. We read this as evidence that MotifAgent's multi-agent interaction surface and MAPPO training do not merely teach new chemistry — they recall and organize chemistry that is already latent in the LLM but underutilized in a single-pass decoder setting.

L.4 Comparison with Graph-Native Baselines

Our main experiments (Tables 1, 2, 3) include strong representatives of the "external graph encoder" paradigm that injects explicit 2D graph representations into a language model: MoMu (GNN encoder + text encoder with contrastive alignment), MolFM (joint 2D graph + knowledge graph + text), GIT-Mol (graph + image + text multimodal LLM), and MolCA (GNN encoder + Galactica LLM via a Q-Former cross-modal projector). On description generation, property prediction, and reaction prediction, MotifAgent outperforms all four across essentially every reported metric.

We interpret this as evidence for a methodological contrast rather than a parameter-count contrast. An external graph encoder supplies the LLM with a pre-computed topological summary, but the topology is produced *before* the LLM reasons about it; the LLM itself never acts on the graph. In MotifAgent, the topology is *constructed* by the LLM through a sequence of motif-level decisions, and the construction process is itself what the model is trained to reason about. The resulting representations are therefore aligned with the generative principles underlying molecules, not merely with their final structures. Two additional observations support this reading. First, MotifAgent-small (3B) already surpasses vanilla Qwen2.5-7B on description generation, so the improvement is not attributable to backbone scale. Second, the progressive ablation in Table 4 shows that neither BRICS input nor multi-agent structure nor RL alone is sufficient — the gain is produced by the coupling of all three, which is precisely the coupling absent from the external-encoder paradigm.

Table 12: Performance comparison on reaction prediction tasks.

Method	EXACT↑	BLEU↑	LEVENSHTEIN↓	RDK FTS↑	MACCS FTS↑	MORGAN FTS↑	VALIDITY↑
Reagent Prediction							
Alpaca	0.000	0.026	29.037	0.029	0.016	0.001	0.186
Baize	0.000	0.051	30.628	0.022	0.018	0.004	0.099
ChatGLM	0.000	0.019	29.169	0.017	0.006	0.002	0.074
LLama	0.000	0.003	28.040	0.037	0.001	0.001	0.001
Vicuna	0.000	0.010	27.948	0.038	0.002	0.001	0.007
Mol-Instruction	0.044	0.224	23.167	0.237	0.364	0.213	1.000
Llama-7b (LoRA)	0.000	0.283	53.510	0.136	0.294	0.106	1.000
InstructMol-G	0.031	0.429	31.447	0.389	0.249	0.220	1.000
InstructMol-GS	0.057	0.439	29.757	0.437	0.314	0.271	0.999
HIGHT-G	0.050	0.462	28.970	0.441	0.314	0.275	1.000
HIGHT-GS	0.067	0.482	27.167	0.462	0.346	0.303	1.000
MotifAgent	**0.085**	**0.516**	**22.571**	**0.502**	**0.376**	**0.379**	**1.000**
Forward Reaction Prediction							
Alpaca	0.000	0.065	41.989	0.004	0.024	0.008	0.138
Baize	0.000	0.044	41.500	0.004	0.025	0.009	0.097
ChatGLM	0.000	0.183	40.008	0.050	0.100	0.044	0.108
LLama	0.000	0.020	42.002	0.001	0.002	0.001	0.039
Vicuna	0.000	0.057	41.690	0.007	0.016	0.006	0.059
Mol-Instruction	0.045	0.654	27.262	0.313	0.509	0.262	1.000
Llama-7b (LoRA)	0.012	0.804	29.947	0.499	0.649	0.407	1.000
InstructMol-G	0.031	0.853	24.790	0.512	0.362	0.303	0.993
InstructMol-GS	0.252	0.926	17.773	0.755	0.599	0.543	1.000
HIGHT-G	0.037	0.869	23.759	0.590	0.394	0.340	0.993
HIGHT-GS	0.293	0.935	16.687	0.774	0.618	0.566	1.000
MotifAgent	**0.315**	**0.937**	**15.127**	**0.806**	**0.669**	**0.582**	**1.000**
Retrosynthesis							
Alpaca	0.000	0.063	46.915	0.005	0.023	0.007	0.160
Baize	0.000	0.095	44.714	0.025	0.050	0.023	0.112
ChatGLM	0.000	0.117	48.365	0.056	0.075	0.043	0.046
LLama	0.000	0.036	46.844	0.018	0.029	0.017	0.010
Vicuna	0.000	0.057	46.877	0.025	0.030	0.021	0.017
Mol-Instruction	0.009	0.705	31.227	0.283	0.487	0.230	1.000
Llama-7b (LoRA)	0.000	0.283	53.510	0.136	0.294	0.106	1.000
InstructMol-G	0.001	0.835	31.359	0.447	0.277	0.241	0.996
InstructMol-GS	0.172	0.911	20.300	0.765	0.615	0.568	1.000
HIGHT-G	0.008	0.863	28.912	0.564	0.340	0.309	1.000
HIGHT-GS	0.202	0.914	20.194	0.772	0.623	0.577	0.999
MotifAgent	**0.275**	**0.932**	**18.810**	**0.783**	**0.685**	**0.631**	**1.000**

Table 13: Statistics of MoleculeNet classification datasets used for molecular property prediction.

Dataset	#Molecules	#Tasks	Split	Description
BBBP	2,039	1	Scaffold	Blood-brain barrier penetration
Tox21	7,831	12	Scaffold	Toxicity on 12 biological targets
ToxCast	8,576	617	Scaffold	Toxicology data from ToxCast
SIDER	1,427	27	Scaffold	Drug side effects (27 categories)
ClinTox	1,478	2	Scaffold	Clinical trial toxicity outcomes
MUV	93,087	17	Scaffold	PubChem bioactivity (virtual screening)
HIV	41,127	1	Scaffold	HIV replication inhibition
BACE	1,513	1	Scaffold	BACE-1 inhibitor activity

Table 14: Statistics of the ChEBI-20 dataset for molecular description generation.

Split	#Samples	Avg. Words/Description	Avg. Sentences
Train	26,407	43.4	3.3
Validation	3,301	43.4	3.3
Test	3,300	43.4	3.3
Total	33,008	43.4	3.3

Table 15: Statistics of Mol-Instructions dataset for chemical reaction prediction tasks.

Task	Train	Validation	Test
Forward Reaction Prediction	83,488	10,436	10,436
Retrosynthesis	83,488	10,436	10,436
Reagent Prediction	83,488	10,436	10,436