

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

Anonymous ACL submission

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; Korooteev, 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., 2020a; 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 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 as-

sembly 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, 2020b), 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 training while making decentralized connection decisions during execution.

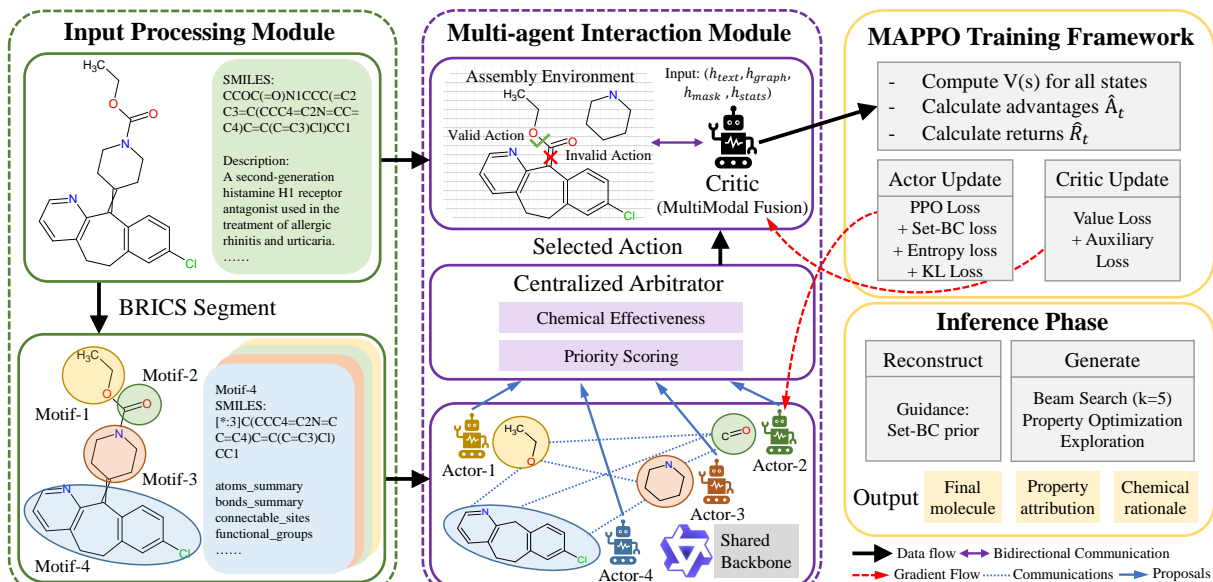


Figure 1: Overview of the proposed MotifAgent framework. MotifAgent consists of four integrated components: (1) Input Processing Module decomposes molecules into motifs using BRICS segmentation and converts them to structured text representations. (2) Multi-agent Interaction Module employs LLM-based 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.

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 $cc(G_t) = 1$ and $E_t \supseteq E^*$, where $cc(\cdot)$ 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 saturated; (2) topological completeness—forming a connected molecular graph; (3) property convergence—improvement over k consecutive steps $< \epsilon$, where k is the window size and ϵ is the convergence threshold; (4) active termination—policy outputting STOP probability > 0.9 .

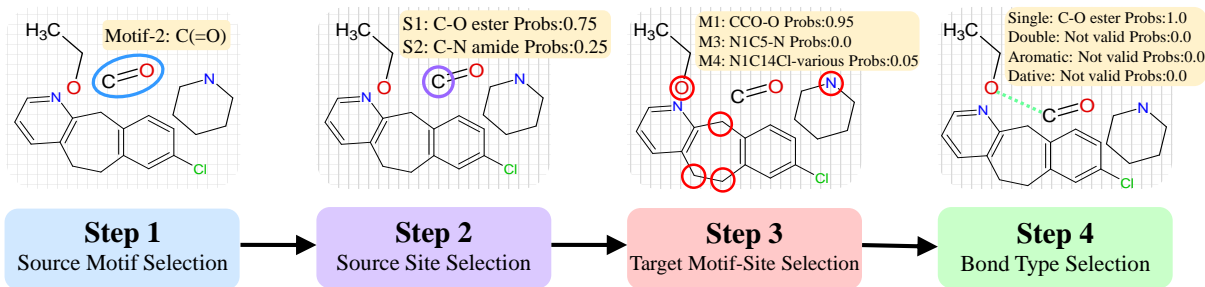


Figure 2: Hierarchical Sampling Process of MotifAgent.

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:

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

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 I.

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 J.

3.5 Policy Optimization and Training

We employ MAPPO (Yu et al., 2022) for policy updates, crucially enabling the policy to learn motif

Training Steps: 0%	50%			100%
	Stage 1: 0-25% Strict Reconstruction Pretraining	Stage 1: 25-50% Soft Constraint Reconstruction	Stage 1: 50-75% Property-Aware Reconstruction	Stage 1: 75-100% Free Generation Optimization
Masking	Hard mask on non-target edges	Soft penalty $r_{over} = -\alpha E - E^* $	Soft penalty (maintained)	Only chemical validity masks
Set-BC	$\alpha_{BC} = 1.0$ (full supervision)	$\alpha_{BC}: 1.0 \rightarrow 0.1$ (decay)	$\alpha_{BC} = 0.1$ (weak)	$\alpha_{BC} = 0.0$ (disabled)
Rewards	Only $r_{valid} = \{+1, -1\}$	Previous + $r_{conn} + r_{edge}$	Previous + r_{prop} (Weight: 0 \rightarrow 0.5)	$r_{prop} + r_{SA} + r_{novel}$

Figure 3: Curriculum Learning Training for MotifAgent.

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:

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

where \mathcal{L}_{clip} is the PPO clipped objective optimizing policy to maximize expected reward, $\beta\mathcal{H}(\pi_\theta)$ provides entropy regularization for exploration, $\alpha_{BC}\mathcal{L}_{BC}$ is the Set-BC supervision term learning correct assembly patterns, \mathcal{L}_{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:

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

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:

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

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^* :

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

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 $\alpha_{BC} = 1.0$. 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 $cc(G) = 1$). 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 interaction format while appending task-specific sections. Detailed prompt templates and training configurations are provided in Appendix K.

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
<i>Specialist Models</i>						
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	-
<i>Retrieval Based LLMs</i>						
MolReGPT	0.607	0.525	0.634	0.476	0.562	0.610
<i>LLM Based Generalist Models</i>						
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	0.007	0.160		
Baize	0.000	0.095	44.714	0.025	0.050	0.023	0.023	0.112		
ChatGLM	0.000	0.117	48.365	0.056	0.075	0.043	0.043	0.046		
LLaMA	0.000	0.036	46.844	0.018	0.029	0.017	0.017	0.010		
Vicuna	0.000	0.057	46.877	0.025	0.030	0.021	0.021	0.017		
Mol-Instruction	0.009	0.705	31.227	0.283	0.487	0.230	0.230	1.000		
LLaMA-7b (LoRA)	0.000	0.283	53.510	0.136	0.294	0.106	0.106	1.000		
InstructMol-GS	0.172	0.911	20.300	0.765	0.615	0.568	0.568	1.000		
HIGHT-GS	0.202	0.914	20.194	0.772	0.623	0.577	0.577	0.999		
MotifAgent	0.275	0.932	18.810	0.783	0.685	0.631	0.631	1.000		

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 F.

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

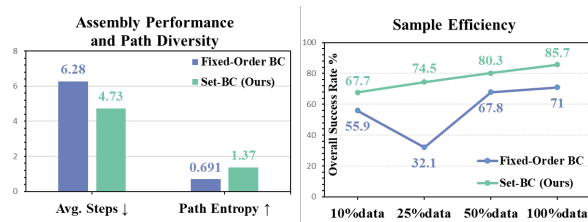


Figure 4: Ablation study on Set-BC effectiveness.

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)).

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 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.

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 G. Retrosynthesis prediction, which requires models to infer suitable reactants given target products, represents a fundamental challenge in AI-assisted synthetic route planning.

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
<i>Specialist Models</i>									
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
Atomus (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
<i>LLM Based Generalist Models</i>									
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: Ablation study on multi-agent collaboration.

Method	Validity (%)	Graph Edit Dist.↓	Morgan FTS↑	Conn. Site Acc.(%)	Bond Type Acc.(%)
Single-Agent	82.3	5.85	0.708	68.4	76.9
MotifAgent	95.6	4.31	0.792	87.6	92.3

Table 5: Ablation study for the effectiveness and time consumption of components on initial training.

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.

We employ both linguistic distance metrics (BLEU, Levenshtein) and molecular fingerprint similarities (RDKit, 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 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 excels in 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

Table 6: 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

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).

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 six critical components: (1) multi-agent collaboration, (2) computational efficiency, (3) Set-based Behavior Cloning, (4) curriculum learning strategy, (5) reward design, and (6) critic auxiliary tasks. Details of evaluation metrics are provided in Appendix ??.

Multi-Agent Collaboration. We compare MotifAgent against a single-agent baseline that concatenates all motifs into one sequence for sequential connection decisions, with identical backbones, rewards, and hyperparameters. As shown in Table 4, the multi-agent system significantly outper-

Table 7: Ablation study on reward design.

Variant	R_{chem}	R_{topo}	Validity (%) \uparrow	Levenshtein \downarrow	Morgan FTS \uparrow
Full	\checkmark	\checkmark	95.6	11.4	0.792
Chem-Only	\checkmark	\times	92.9	12.2	0.771
Topo-Only	\times	\checkmark	89.1	13.3	0.759
Minimal (r_{valid} Only)	partial	\times	85.2	15.8	0.732

Table 8: Ablation study on priority scoring $S(a)$.

Method	w_1	w_2	w_3	Validity (%) \uparrow	Levenshtein \downarrow	Morgan FTS \uparrow
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

forms the single-agent baseline across all metrics. The improvements in Graph Edit Distance and Morgan FTS demonstrate that multi-agent collaboration better preserves 2D topological structure, while the connection-level accuracy gains reveal that distributed agents with specialized local knowledge make more informed decisions than a single agent managing all motifs simultaneously.

Computational Efficiency. We further analyze the computational costs of both architectures. As shown in Table 5, the MotifAgent is surprisingly more efficient: it processes each sample 24% faster and consumes 18% fewer tokens. This counter-intuitive result arises because the single-agent must expand complete reasoning chains within one context—covering motif selection, site judgment, bond typing, and global consistency—leading to longer generations and frequent backtracking. In contrast, MotifAgent assigns these subtasks to multiple specialized agents, replacing verbose reasoning with compact structured representations for inter-agent communication. Consequently, MotifAgent reaches 75% Morgan FTS in 52.7 hours, while the single-agent never achieves this threshold within 200K steps, demonstrating that MotifAgent improves both effectiveness and efficiency.

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

Table 9: Ablation study on critic auxiliary tasks.

Variant	w_1	w_2	Conn. Site (%) \uparrow	Bond Type (%) \uparrow
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

the importance of our four-stage curriculum by comparing against: training without curriculum, a simplified 2-stage variant, and reversed curriculum order. Table 6 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 7 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 terms. Table 8 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 9 shows that removing both auxiliary tasks significantly degrades connect accuracy. Edge prediction (w_1) and connected component prediction (w_2) provide complementary benefits, with their combination yielding the best performance.

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.

6 Limitations

Our work has several limitations that should be acknowledged. First, the current multi-agent framework incurs substantial computational overhead, as each motif agent requires separate inference through the shared LLM backbone, which may limit scalability for real-time applications. Second, our method has been primarily validated on drug-like small molecules using standard benchmarks, and its generalizability to larger molecular systems such as natural products, polymers, and organometallic compounds, as well as broader real-world chemical space, remains unexplored.

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 Interdisciplinary 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 Hua-jun 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*.

715	Binyuan Hui, Jian Yang, Zeyu Cui, Jiayi Yang,	Zhiyuan Liu, Sihang Li, Yan Chen Luo, Hao Fei, Yixin	770
716	Dayiheng Liu, Lei Zhang, Tianyu Liu, Jiajun	Cao, Kenji Kawaguchi, Xiang Wang, and Tat-Seng	771
717	Zhang, Bowen Yu, Keming Lu, and 1 others. 2024.	Chua. 2023c. Molca: Molecular graph-language	772
718	Qwen2. 5-coder technical report. <i>arXiv preprint</i>	modeling with cross-modal projector and uni-modal	773
719	<i>arXiv:2409.12186</i> .	adapter. <i>arXiv preprint arXiv:2310.12798</i> .	774
720	Wengong Jin, Regina Barzilay, and Tommi Jaakkola.	Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI	775
721	2018. Junction tree variational autoencoder for	Pieter Abbeel, and Igor Mordatch. 2017a. Multi-	776
722	molecular graph generation. In <i>International confer-</i>	agent actor-critic for mixed cooperative-competitive	777
723	<i>ence on machine learning</i> , pages 2323–2332. PMLR.	environments. <i>Advances in neural information pro-</i>	778
724	Wengong Jin, Regina Barzilay, and Tommi Jaakkola.	<i>cessing systems</i> , 30.	779
725	2020a. Hierarchical generation of molecular graphs	Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI	780
726	using structural motifs. In <i>International conference</i>	Pieter Abbeel, and Igor Mordatch. 2017b. Multi-	781
727	<i>on machine learning</i> , pages 4839–4848. PMLR.	agent actor-critic for mixed cooperative-competitive	782
728	Wengong Jin, Regina Barzilay, and Tommi Jaakkola.	environments. <i>Advances in neural information pro-</i>	783
729	2020b. Hierarchical generation of molecular graphs	<i>cessing systems</i> , 30.	784
730	using structural motifs. In <i>International conference</i>	Yizhen Luo, Kai Yang, Massimo Hong, Xing Yi	785
731	<i>on machine learning</i> , pages 4839–4848. PMLR.	Liu, and Zaiqing Nie. 2023a. Molfm: A multi-	786
732	Mikhail V Koroteev. 2021. Bert: a review of applica-	modal molecular foundation model. <i>arXiv preprint</i>	787
733	tions in natural language processing and understand-	<i>arXiv:2307.09484</i> .	788
734	ing. <i>arXiv preprint arXiv:2103.11943</i> .	Yizhen Luo, Jiahuan Zhang, Siqi Fan, Kai Yang,	789
735	Mario Krenn, Florian Häse, AkshatKumar Nigam, Pas-	Yushuai Wu, Mu Qiao, and Zaiqing Nie. 2023b.	790
736	cal Friederich, and Alan Aspuru-Guzik. 2020. Self-	Biomedgpt: Open multimodal generative pre-trained	791
737	referencing embedded strings (selfies): A 100% ro-	transformer for biomedicine. <i>arXiv preprint</i>	792
738	burst molecular string representation. <i>Machine Learn-</i>	<i>arXiv:2308.09442</i> .	793
739	<i>ing: Science and Technology</i> , 1(4):045024.	Krzysztof Maziarz, Henry Jackson-Flux, Pashmina	794
740	Jiatong Li, Yunqing Liu, Wenqi Fan, Xiao-Yong Wei,	Cameron, Finton Sirockin, Nadine Schneider, Niko-	795
741	Hui Liu, Jiliang Tang, and Qing Li. 2024. Empow-	laus Stiefl, Marwin Segler, and Marc Brockschmidt.	796
742	ering molecule discovery for molecule-caption trans-	2021. Learning to extend molecular scaffolds with	797
743	lation with large language models: A chatgpt per-	structural motifs. <i>arXiv preprint arXiv:2103.03864</i> .	798
744	spective. <i>IEEE transactions on knowledge and data</i>	Eyal Mazuz, Guy Shtar, Bracha Shapira, and Lior	799
745	<i>engineering</i> , 36(11):6071–6083.	Rokach. 2023. Molecule generation using transfor-	800
746	Junxian Li, Di Zhang, Xunzhi Wang, Zeying Hao, Jingdi	mers and policy gradient reinforcement learning. <i>Sci-</i>	801
747	Lei, Qian Tan, Cai Zhou, Wei Liu, Yaotian Yang, Xin-	<i>entific Reports</i> , 13(1):8799.	802
748	rui Xiong, and 1 others. 2025. Chemvlm: Exploring	Medard Edmund Mswahili and Young-Seob Jeong.	803
749	the power of multimodal large language models in	2024. Transformer-based models for chemical smiles	804
750	chemistry area. In <i>Proceedings of the AAAI Con-</i>	representation: A comprehensive literature review.	805
751	<i>ference on Artificial Intelligence</i> , volume 39, pages	<i>Heliyon</i> , 10(20).	806
752	415–423.	Andrew Y Ng, Daishi Harada, and Stuart Russell. 1999.	807
753	Chin-Yew Lin. 2004. Rouge: A package for automatic	Policy invariance under reward transformations: The-	808
754	evaluation of summaries. In <i>Text Summarization</i>	ory and application to reward shaping. In <i>Icml</i> , vol-	809
755	<i>Branches Out</i> , pages 74–81.	ume 99, pages 278–287. Citeseer.	810
756	Pengfei Liu, Yiming Ren, Jun Tao, and Zhixiang Ren.	George Papadatos, Anna Gaulton, Anne Hersey, and	811
757	2024. Git-mol: A multi-modal large language model	John P Overington. 2015. Activity, assay and	812
758	for molecular science with graph, image, and text.	target data curation and quality in the chembl	813
759	<i>Computers in biology and medicine</i> , 171:108073.	database. <i>Journal of computer-aided molecular de-</i>	814
760	Shengchao Liu, Weili Nie, Chengpeng Wang, Jiarui	<i>sign</i> , 29(9):885–896.	815
761	Lu, Zhuoran Qiao, Ling Liu, Jian Tang, Chaowei	Kishore Papineni, Salim Roukos, Todd Ward, and Wei-	816
762	Xiao, and Animashree Anandkumar. 2023a. Multi-	Jing Zhu. 2002. Bleu: a method for automatic evalu-	817
763	modal molecule structure–text model for text-based	ation of machine translation. In <i>Proceedings of the</i>	818
764	retrieval and editing. <i>Nature Machine Intelligence</i> ,	<i>40th Annual Meeting of the Association for Compu-</i>	819
765	5(12):1447–1457.	<i>tational Linguistics</i> , pages 311–318.	820
766	Zequn Liu, Wei Zhang, Yingce Xia, Lijun Wu, Shufang	Chen Qian, Zihao Xie, Yifei Wang, Wei Liu, Kunlun	821
767	Xie, Tao Qin, Ming Zhang, and Tie-Yan Liu. 2023b.	Zhu, Hanchen Xia, Yufan Dang, Zhuoyun Du, Weize	822
768	Molxpt: Wrapping molecules with text for generative	Chen, Cheng Yang, and 1 others. 2024. Scaling	823
769	pre-training. <i>arXiv preprint arXiv:2305.10688</i> .	large language model-based multi-agent collabora-	824
		tion. <i>arXiv preprint arXiv:2406.07155</i> .	825

826	Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, and 1 others. 2018. Improving language understanding by generative pre-training.	881
827		882
828		883
829	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 . <i>Preprint</i> , arXiv:1910.10683.	884
830		885
831		886
832		887
833		888
834	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. <i>Journal of Machine Learning Research</i> , 21(178):1–51.	889
835		890
836		891
837		892
838		893
839		894
840	Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. 2020. Graphaf: a flow-based autoregressive model for molecular graph generation. <i>arXiv preprint arXiv:2001.09382</i> .	895
841		896
842		897
843		898
844	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 <i>2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)</i> , pages 1161–1168. IEEE.	899
845		900
846		901
847		902
848		903
849		904
850	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. <i>arXiv preprint arXiv:2209.05481</i> .	905
851		906
852		907
853		908
854		909
855	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. <i>arXiv preprint arXiv:1706.05296</i> .	910
856		911
857		912
858		913
859		914
860		915
861	Qian Tan, Dongzhan Zhou, Peng Xia, Wanhao Liu, Wanli Ouyang, Lei Bai, Yuqiang Li, and Tianfan Fu. 2025. Chemmlm: Chemical multimodal large language model. <i>arXiv preprint arXiv:2505.16326</i> .	916
862		917
863		918
864		919
865	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. <i>arXiv preprint arXiv:2302.13971</i> .	920
866		921
867		922
868		923
869		924
870		925
871	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. <i>arXiv preprint arXiv:2501.06322</i> .	926
872		927
873		928
874		929
875		930
876	Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. <i>Advances in neural information processing systems</i> , 30.	931
877		932
878		933
879		934
880		935
	William Yi Wang, Jinshan Li, Weimin Liu, and Zi-Kui Liu. 2019. Integrated computational materials engineering for advanced materials: A brief review. <i>Computational Materials Science</i> , 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. <i>Advances in Neural Information Processing Systems</i> , 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. <i>Wiley Interdisciplinary Reviews: Computational Molecular Science</i> , 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. <i>Chemical science</i> , 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. <i>arXiv preprint arXiv:2304.01196</i> .	
	Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. 2022. Geodiff: A geometric diffusion model for molecular conformation generation. <i>arXiv preprint arXiv:2203.02923</i> .	
	Chao Yu, Akash Velu, Eugene Vinitzky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. <i>Advances in neural information processing systems</i> , 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. <i>arXiv preprint arXiv:2210.02414</i> .	
	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. <i>arXiv preprint arXiv:2404.16880</i> .	
	Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. 2021. Motif-based graph self-supervised learning for molecular property prediction. <i>Advances in Neural Information Processing Systems</i> , 34:15870–15882.	
	Zaixi Zhang, Yaosen Min, Shuxin Zheng, and Qi Liu. 2023. Molecule generation for target protein binding with structural motifs. In <i>The eleventh international conference on learning representations</i> .	
	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.	

936 *Advances in neural information processing systems*,
937 36:5850–5887.

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

Table 10: Hyperparameter details for MotifAgent.

Hyperparameter	Value
<i>Model Architecture</i>	
policy backbone	Qwen2.5-7B
critic network	MolT5-base
LoRA rank	16
action head hidden dim	512
<i>Training Configuration</i>	
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
<i>Optimization</i>	
actor learning rate	1e-5
critic learning rate	1e-4
gradient clip norm	0.5
optimizer	AdamW
warmup steps	10000
<i>PPO Parameters</i>	
clipping ϵ	0.2
GAE λ	0.95
discount γ	0.99
value loss coefficient	0.5
<i>Regularization</i>	
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
<i>Auxiliary Tasks</i>	
edge prediction weight w_1	0.1
component prediction weight w_2	0.1

942 A Potential Risks

943 While MotifAgent is designed for beneficial ap-
944 plications in drug discovery and molecular under-
945 standing, we acknowledge potential dual-use con-
946 cerns. The model’s capability to understand molec-
947 ular assembly rules could theoretically be misused
948 to design harmful substances. However, the frame-
949 work focuses on learning general chemical con-
950 nection principles rather than optimizing for spe-
951 cific harmful properties, and all training data are

derived from publicly available databases. We rec-
ommend 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 con-
tain 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 describ-
ing chemical and biological characteristics. We
have verified that our datasets do not contain of-
fensive content, as they consist solely of standard-
ized chemical nomenclature and objective prop-
erty descriptions. No anonymization was required
since the data pertains exclusively to chemical com-
pounds rather than individuals.

C LLM Usage

Large language models (LLMs) were used for refin-
ing sentence structure, improving grammatical ac-
curacy, and enhancing the clarity of the manuscript
text. A supporting role was played by the LLMs in
the manuscript’s language polishing, but no scien-
tific 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 Ta-
ble 10. 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 envi-
ronments with a rollout length of 32 steps, result-
ing in 256 transitions per update. The actor learn-
ing rate is set to 5e-5 with LoRA (rank=16), while
the critic uses 3e-4. The training employs curricu-
lum learning with four phases, automatically transi-
tioning based on performance metrics to progres-
sively master topological construction from strict
reconstruction to free exploration. For curricu-
lum 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

999	model’s capabilities: Phase 1 (0-25%): Strict re-	(‘{motif_smiles}’,	{n_atoms}	1044
1000	construction with hard masks, focusing on learn-	atoms)		1045
1001	ing valid chemical connections; Phase 2 (25-50%):	...		1046
1002	Soft constraints with Set-BC weight decay, allow-	-	**Motif_n**:	1047
1003	ing exploration of equivalent paths; Phase 3 (50-	{motif_description}		1048
1004	75%): Property-aware reconstruction, introducing	(‘{motif_smiles}’,	{n_atoms}	1049
1005	target property rewards; Phase 4 (75-100%): Free	atoms)		1050
1006	exploration for generation, removing hard topolog-			
1007	ical constraints. The automatic phase transitions	E.1.2 Episode Start Section		1051
1008	ensure the model has sufficiently mastered each	The episode begins with the initial environment		1052
1009	level before progressing, preventing premature ex-	state and global coordinator instruction:		1053
1010	ploration that could lead to unstable training.	### Episode Start: Multi-Agent		1054
1011	E Prompt Templates and Downstream	Dialogue		1055
1012	Task Training	**Environment State (t=0)**:		1056
1013	This appendix provides the unified prompt tem-	Current Assembly: Empty graph		1057
1014	plates used in MotifAgent for both initial training	Connected Components: 0		1058
1015	and downstream task fine-tuning, along with de-	Available Motifs: [Motif_1,		1059
1016	tailed descriptions of the training procedures.	Motif_2, ..., Motif_n]		1060
1017	E.1 Initial Training Prompt Template	Target Edges: {n_edges}		1061
1018	During the initial training phase, MotifAgent learns	({edge_descriptions})		1062
1019	molecular assembly through multi-agent collabora-	Current Properties: None		1063
1020	tion. The following template defines the structured	**Global Coordinator**:		1064
1021	interaction format used throughout the training pro-	“Agents, we need to assemble		1065
1022	cess.	{molecule_name}. The target has		1066
1023	E.1.1 Initial Setup Section	{property_description} with LogP		1067
1024	The prompt begins with comprehensive molecular	around {logp_value}. Current		1068
1025	information:	topology is empty. Please propose		1069
1026	### Initial Setup	your initial connections.”		1070
1027	**Target Molecule**:	E.1.3 Round Structure		1071
1028	{molecule_name}	Each assembly round follows a consistent structure		1072
1029	- SMILES: ‘{smiles_string}’	with a descriptive title:		1073
1030	- Target Properties:	### Round {N}: {Round_Title}		1074
1031	{property_description}, LogP	For subsequent rounds (t > 0), the environment		1075
1032	≈ {logp_value}, MW =	state is updated:		1076
1033	{molecular_weight}	**Environment State		1077
1034	- PubChem CID: {cid} (if	(t={step})**:		1078
1035	available)	Current Assembly:		1079
1036	**Fragmentation Results** (via	{current_graph_description}		1080
1037	BRICS):	Connected Components:		1081
1038	- **Motif_1**:	{num_components}		1082
1039	{motif_description}	Remaining Motifs:		1083
1040	(‘{motif_smiles}’, {n_atoms}	[{remaining_motif_list}]		1084
1041	atoms)	Remaining Target Edges:		1085
1042	- **Motif_2**:	{num_remaining_edges}		1086
1043	{motif_description}	Current Properties:		1087
		{current_property_estimates}		1088

1089	E.1.4 Multi-Agent Proposal Section	- {connection_2}: ✓ Valid,	1129
1090	Each motif agent generates proposals. For the initial	{validity_reason}	1130
1091	round (t=0), agents use the full proposal format:	Topological Scoring:	1131
1092	**Agent_{i} ({motif_name})**:	S({connection_1}) = {v1}	1132
1093	My Structure:	(stability) + {v2} (progress) +	1133
1094	{structural_description}	{v3} (property) = {total_1}	1134
1095	Available Sites: [{site_list}]	S({connection_2}) = {v1}	1135
1096	Chemical Context:	(stability) + {v2} (progress) +	1136
1097	{chemical_environment_description}	{v3} (property) = {total_2}	1137
1098	Proposal: {connection_proposal}	Decision: Execute	1138
1099	Reasoning: {chemical_reasoning}	{selected_connection} connection	1139
1100	Priority: {HIGH/MEDIUM/LOW} -	first	1140
1101	{priority_explanation}		
1102	For subsequent rounds, agents may include status	E.1.6 Action Execution and Reward Section	1141
1103	updates and modified fields:	Each executed action follows the format:	1142
1104	**Agent_{i} ({motif_name})**:	**Action Executed**:	1143
1105	Status Update:	'Connect: Motif_{i}[{site_x}]	1144
1106	{occupied_sites_description}	-{bond_type}->	1145
1107	Available Sites:	Motif_{j}[{site_y}]'	1146
1108	[{remaining_site_list}]	**Reward Calculation**:	1147
1109	Proposal: {connection_proposal}	r_valid = {value} ({explanation})	1148
1110	Reasoning: {chemical_reasoning}	r_stable = {value}	1149
1111	Chemical Insight:	({explanation})	1150
1112	{additional_chemical_insight}	r_func = {value} ({explanation})	1151
1113	Alternatively, agents waiting for dependencies	r_conn = {value} ({explanation})	1152
1114	use:	r_edge = {value} ({explanation})	1153
1115	**Agent_{i} ({motif_name})**:	Total: R = {total_reward}	1154
1116	Waiting State:	For terminal states, additional rewards are in-	1155
1117	{dependency_description}	cluded:	1156
1118	Future Plan: {planned_action}	r_prop = {value}	1157
1119	Property Prediction:	({property_match_explanation})	1158
1120	{expected_property_impact}	Terminal Bonus = {value}	1159
1121	E.1.5 Central Arbitrator Section	(successful reconstruction)	1160
1122	The central arbitrator evaluates all proposals and	Total: R = {total_reward}	1161
1123	makes decisions:	E.1.7 Termination Section	1162
1124	**Central Arbitrator	Upon successful assembly:	1163
1125	Evaluation**:	### Round {N}: Termination	1164
1126	Chemical Validity Check:	Decision	1165
1127	- {connection_1}: ✓ Valid,	**Environment State	1166
1128	{validity_reason}	(t={final_step})**:	1167
		Current Assembly: Complete	1168
		{molecule_name} molecule	1169
		Connected Components: 1	1170

1171	Target Edges Matched: {n}/{n} ✓	—	1215
1172	Properties: LogP = {value}, MW =	### Episode Start: Multi-Agent	1216
1173	{value}, {additional_properties}	Analysis	1217
1174	**Global Coordinator**:	[Standard environment state and	1218
1175	“Assembly complete. All target	global coordinator instruction]	1219
1176	edges matched. Properties within	[Multi-agent interaction rounds	1220
1177	specifications.”	for topology reconstruction]	1221
1178	**Agent_1 ({motif_name})*:	—	1222
1179	“{confirmation_statement}”		
1180	**Agent_2 ({motif_name})*:	### Property Prediction Phase	1223
1181	“{confirmation_statement}”	**Environment State (Final)**:	1224
1182	...	Current Assembly: Complete	1225
1183	**Policy Decision**:	molecule	1226
1184	STOP (probability = {prob})‘	Connected Components: 1	1227
1185	E.1.8 Chemical Explanation Section	Target Edges Matched: {n}/{n} ✓	1228
1186	The episode concludes with a generated chemical	Reconstructed Topology:	1229
1187	explanation:	{topology_summary}	1230
1188	**Chemical Explanation	**Global Property Analyzer**:	1231
1189	Generated**:	Topology-Property Mapping:	1232
1190	“{detailed_chemical_rationale_	- Identified pharmacophores:	1233
1191	“explaining_assembly_strategy_	{pharmacophore_list}	1234
1192	“and_property_relationships}”	- Key structural alerts:	1235
1193	E.2 Downstream Task Prompt Templates	{alert_list}	1236
1194	For downstream tasks, we extend the base template	- Electronic effects:	1237
1195	with task-specific components. All downstream	{electronic_description}	1238
1196	tasks share the same initial setup and multi-agent	- Steric factors:	1239
1197	interaction format, with task-specific output sec-	{steric_description}	1240
1198	tions appended.	Property-Contributing Features:	1241
1199	E.2.1 Molecular Property Prediction	1. {feature_1}:	1242
1200	For property prediction tasks, the prompt includes	{contribution_to_property}	1243
1201	property-specific analysis phases:	2. {feature_2}:	1244
1202	### Initial Setup	{contribution_to_property}	1245
1203	**Task**:	Molecular Property	1246
1204	Prediction	**Prediction Output**:	1247
1205	**Property**:	Property: {property_name}	1248
1206	{property_name}	Prediction: {0/1 or value}	1249
1207	(e.g., Blood-Brain Barrier	Confidence: {score}	1250
1208	Penetration)	Key Evidence:	1251
1209	**Target Molecule**:	- {evidence_1}	1252
1210	- SMILES: ‘{smiles}‘	- {evidence_2}	
1211	- PubChem CID: {cid} (if		
1212	available)		
1213	**Fragmentation Results** (via	E.2.2 Molecular Description Generation	1253
1214	BRICS):	For description generation, the output section cap-	1254
	[Same format as initial training]	tures the assembly trace:	1255

1256	### Initial Setup	**Input** (Forward):	1300
1257	**Task**:	- Reactants: '{reactant_smiles}'	1301
1258	Molecular Description Generation	- Reagents: {reagent_info}	1302
1259	**Target Molecule**:	**Input** (Retrosynthesis):	1303
1260	- SMILES: '{smiles}'	- Product: '{product_smiles}'	1304
1261	- Known Properties:	**Fragmentation Results** (via BRICS):	1305
1262	{known_properties}	[Motif decomposition of reactants/products]	1306
1263	- PubChem CID: {cid}	—	1309
1264	**Fragmentation Results** (via BRICS):	### Multi-Agent Reaction Analysis	1310
1265	[Same format as initial training]	[Agents analyze reaction centers, bond changes, electron flow]	1311
1266	—	[Each agent identifies its role in the transformation]	1312
1267	### Episode Start: Multi-Agent Assembly for Description	—	1315
1268	[Standard environment state]	### Reaction Prediction Phase	1316
1269	**Global Coordinator**:	**Mechanism Analysis**:	1317
1270	“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.”	- Reaction center: {identified_centers}	1318
1271	[Multi-agent interaction rounds]	- Bond breaking: {bonds_broken}	1320
1272	—	- Bond forming: {bonds_formed}	1321
1273	### Description Generation Phase	- Electron transfer pathway: {electron_flow_description}	1322
1274	**Trajectory Summary**:	**Predicted Output**:	1324
1275	Step 1: {action_description} → {chemical_insight}	'{SELFIES_representation}'	1325
1276	Step 2: {action_description} → {chemical_insight}	**Chemical Explanation**:	1326
1277	...	"{mechanistic_rationale}"	1327
1278	**Generated Description**:	E.3 Downstream Task Training Procedures	1328
1279	"{natural_language_description_	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.	1329
1280	covering_chemical_structure_		1330
1281	functional_groups_physicochemical_		1331
1282	properties_and_biological_activities}”		1332
1283		E.3.1 Training Strategy Overview	1335
1284	E.2.3 Chemical Reaction Prediction	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.	1336
1285	For reaction prediction tasks (forward reaction and retrosynthesis):		1337
1286			1338
1287	### Initial Setup		1339
1288	**Task**:		1340
1289	{Forward Reaction Prediction / Retrosynthesis}		1341
1290			1342

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.

E.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.

E.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.

F Evaluation Metrics

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

F.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:

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

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.

F.2 Chemical Reaction Prediction

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

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

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:

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

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.

F.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, containing 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.

F.4 Ablation Study Metrics

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

F.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 mo-

tifs 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.

F.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.

G 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 11 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 (RDKit 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 transfer 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.

H Algorithm

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

I 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.

I.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 $[0, 3]$.

I.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:

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

I.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

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

```
1: function FRAGMENTMOLECULE( $M$ )
2:    $\mathcal{M}, E^* \leftarrow \text{BRICS}(M)$ ;  $\mathcal{T} \leftarrow \text{TextSerialize}(\mathcal{M})$ 
3:   return  $\mathcal{M}, \mathcal{T}, E^*$ 
4: end function
5: function MULTIAGENTPROPOSAL( $\mathcal{A}, G_t$ )
6:    $\mathcal{P} \leftarrow \emptyset$ 
7:   for each agent  $a_i \in \mathcal{A}$  do
8:      $o_i \leftarrow \text{LocalObs}(a_i, G_t)$ ;  $h_{\text{topo}} \leftarrow \text{GlobalTopo}(G_t)$ 
9:      $\mathcal{P} \leftarrow \mathcal{P} \cup \{\pi_\theta(o_i, h_{\text{topo}})\}$  {Hierarchical sampling}
10:  end for
11:  return  $\mathcal{P}$ 
12: end function
13: function CENTRALARBITRATION( $\mathcal{P}, E^*, G_t, \text{stage}$ )
14:    $\mathcal{P}_{\text{valid}} \leftarrow \text{ChemFilter}(\mathcal{P})$ 
15:   if  $\text{stage} \leq 2$  then
16:      $\mathcal{P}_{\text{valid}} \leftarrow \text{TargetMask}(\mathcal{P}_{\text{valid}}, E^*)$ 
17:   end if
18:   return  $\arg \max_{a \in \mathcal{P}_{\text{valid}}} [w_1 \cdot \text{Chem}(a) + w_2 \cdot \text{Topo}(a) + w_3 \cdot \text{Prop}(a)]$ 
19: end function
20: function OPTIMIZE( $\mathcal{B}, \pi_\theta, V_\phi, \pi_{\text{ref}}, E^*$ )
21:   // Compute advantages and returns
22:    $\hat{A} \leftarrow \sum_{l=0}^{\infty} (\gamma \lambda)^l \delta_{t+l}$ ;  $\hat{R} \leftarrow \text{MonteCarloReturn}(\mathcal{B})$ 
23:   // Actor loss with Set-BC
24:    $\mathcal{A}^* \leftarrow \{a: \text{preserves target edges from } E^*\}$ 
25:    $\mathcal{L}_{\text{actor}} \leftarrow -\mathcal{L}_{\text{PPO}} - \beta \mathcal{H}(\pi_\theta) + \alpha_{BC} \log \sum_{a \in \mathcal{A}^*} \pi_\theta(a|s) + \beta_{KL} \text{KL}(\pi_\theta || \pi_{\text{ref}})$ 
26:   // Critic loss with auxiliary tasks
27:    $\mathcal{L}_{\text{critic}} \leftarrow (V_\phi - \hat{R})^2 + w_1 (V_{\text{edges}} - |E^* \setminus E_t|)^2 + w_2 (V_{\text{cc}} - \text{cc}(G_t))^2$ 
28:   Update  $\theta \leftarrow \theta - \eta_\pi \nabla_\theta \mathcal{L}_{\text{actor}}$ ; Update  $\phi \leftarrow \phi - \eta_V \nabla_\phi \mathcal{L}_{\text{critic}}$ 
29: end function
30: // Main Training Loop
31:  $\pi_\theta \leftarrow \text{SharedLLM}()$ ;  $V_\phi \leftarrow \text{Critic}()$ ;  $\pi_{\text{ref}} \leftarrow \pi_\theta$ ;  $\text{stage} \leftarrow 1$ 
32: for episode = 1 to  $N$  do
33:    $M \sim \mathcal{D}$ ;  $\mathcal{M}, \mathcal{T}, E^* \leftarrow \text{FRAGMENTMOLECULE}(M)$ 
34:    $\mathcal{A} \leftarrow [\text{MotifAgent}(m, \pi_\theta) \text{ for } m \in \mathcal{M}]$ ;  $G_t \leftarrow \emptyset$ 
35:   while  $\neg \text{Terminal}(G_t, E^*)$  do
36:      $\mathcal{P} \leftarrow \text{MULTIAGENTPROPOSAL}(\mathcal{A}, G_t)$ 
37:      $a_t \leftarrow \text{CENTRALARBITRATION}(\mathcal{P}, E^*, G_t, \text{stage})$ 
38:      $G_{t+1}, r_t \leftarrow \text{Execute}(G_t, a_t)$ ;  $r_t \leftarrow r_t + \gamma \Phi(G_{t+1}) - \Phi(G_t)$ 
39:      $\mathcal{B} \leftarrow \mathcal{B} \cup \{(G_t, a_t, r_t, G_{t+1})\}$ ;  $G_t \leftarrow G_{t+1}$ 
40:   end while
41:   if  $|\mathcal{B}| \geq \text{batch\_size}$  then
42:     OPTIMIZE( $\mathcal{B}, \pi_\theta, V_\phi, \pi_{\text{ref}}, E^*$ );  $\mathcal{B} \leftarrow \emptyset$ 
43:     if performance meets criteria then  $\text{stage} \leftarrow \text{stage} + 1$ ; Adjust( $\alpha_{BC}, w_i$ )
44:   end if
45: end for
```

1711 as the relative reduction in connected components: 1754
1712 if the action merges the last two components into 1755
1713 one, this term contributes maximally. 1756

1714 **Ring Formation** rewards completion of ring sys- 1757
1715 tems that match the target molecule. Completing 1758
1716 a target ring system scores 1.0, contributing to a 1759
1717 partial ring scores 0.5, and other actions score 0.0. 1760

1718 **Bridge Contribution** rewards key bridging con- 1761
1719 nections between major substructures. Connecting 1762
1720 two scaffold motifs scores 1.0, connecting a func- 1763
1721 tional group to a scaffold scores 0.5, and peripheral 1764
1722 connections score 0.0. 1765

1723 The overall topological progress score sums 1766
1724 these components and normalizes to $[0, 1]$: 1767

$$1725 \text{TopoProgress}(a) = \frac{1}{4} \left(P_{\text{edge}} + P_{\text{conn}} \right. \quad (18) \\ \left. + P_{\text{ring}} + P_{\text{bridge}} \right) \quad 1770$$

1726 I.4 Property Improvement Score

1727 The property improvement score 1772
1728 $\text{PropImprove}(a) \in [0, 1]$ estimates how the 1773
1729 action affects target molecular properties. 1774

1730 For each target property p in the property set \mathcal{P} 1775
1731 (e.g., LogP, molecular weight, pKa), we compute 1776
1732 the reduction in deviation from the target value: 1777
1733 the absolute difference before the action minus the 1778
1734 absolute difference after. Positive values indicate 1779
1735 improvement toward the target. These improve- 1780
1736 ments are summed across all properties and passed 1781
1737 through a sigmoid function for normalization: 1782

$$1738 \text{PropImprove}(a) = \sigma \sum_{p \in \mathcal{P}} (|y_p^{\text{current}} - y_p^{\text{target}}| \quad (19) \\ - |y_p^{\text{after}} - y_p^{\text{target}}|) \quad 1787$$

1739 where σ is the sigmoid function, y_p^{target} is the target 1788
1740 value for property p , and y_p^{current} , y_p^{after} are predicted 1789
1741 values before and after the action. 1790

1742 Property predictions use lightweight estimators: 1791
1743 LogP is estimated via incremental atom/fragment 1792
1744 contributions, molecular weight is calculated ex- 1793
1745 actly from added atoms, and functional group prop- 1794
1746 erties use a lookup table for known pharmacophore 1795
1747 contributions. 1796

1748 I.5 Scoring Example

1749 We illustrate the scoring function using the aspirin 1797
1750 assembly case, comparing two candidate actions in 1798
1751 Round 1. 1799

1752 **Candidate A: Benzene**→**Carboxyl**. This con- 1801
1753 nection establishes the salicylic acid core. The 1802

1754 chemical stability score is high (0.95) because the 1755
1756 connection satisfies valence rules, preserves ben- 1757
1758 zene aromaticity, introduces minimal strain, and 1758
1759 matches common benzoic acid patterns. The topo- 1759
1760 logical progress score is 0.75: the edge matches the 1761
1762 target, it creates the first connected component, no 1762
1763 ring is formed, and it connects a functional group 1763
1764 to the scaffold. The property improvement score 1764
1765 is 0.65 as it moves LogP and acidity toward target 1765
1766 values. Total score: $0.95 + 0.75 + 0.65 = 2.35$. 1766

1767 **Candidate B: Benzene**→**Oxygen**. This con- 1767
1768 nection would attach the oxygen linker directly. 1768
1769 The chemical stability score is 0.84: valence is sat- 1769
1770 isfied, but connecting to an aromatic system via 1770
1771 ether linkage scores lower on aromaticity preserva- 1771
1772 tion. The topological progress score is 0.70: edge 1772
1773 matches target, creates a connected component, but 1773
1774 no ring or scaffold-scaffold bridge. The property 1774
1775 improvement is modest (0.52) as it has less im- 1775
1776 pact on target properties at this stage. Total score: 1776
1777 $0.84 + 0.70 + 0.52 = 2.06$. 1777

1778 The arbitrator selects Candidate A 1778
1779 (Benzene→Carboxyl) with the higher score, 1779
1780 prioritizing establishment of the pharmacophore 1780
1781 core before adding the ester linkage. 1781

1779 I.6 Curriculum Learning Behavior

1782 While the weights remain fixed at 1.0 throughout 1780
1783 training, the effective contribution of each score 1781
1784 component naturally shifts across curriculum learn- 1782
1785 ing phases. In Phase 1 (Strict Reconstruction), 1783
1786 TopoProgress dominates because hard masks re- 1784
1787 strict actions to target edges, and PropImprove has 1785
1788 limited signal. In Phase 2 (Soft Constraint), all 1786
1789 three components contribute meaningfully as the 1787
1790 model explores beyond exact target edges. In Phase 1788
1791 3 (Property-Aware), PropImprove gains importance 1789
1792 as property predictors become more reliable. In 1790
1793 Phase 4 (Free Generation), the model leverages all 1791
1794 three dimensions in full balance for novel molecule 1792
1795 generation. This design allows the model to learn 1793
1796 appropriate trade-offs through experience rather 1794
1797 than through manually tuned weight schedules. 1795

1796 J Reward Function Details

1797 This section provides the complete mathematical 1797
1798 formulations of reward components introduced in 1798
1799 Section 3.4. 1799

1800 J.1 Chemical Base Rewards

1801 The chemical base reward R_{chem} provides a 1801
1802 comprehensive assessment of molecular quality 1802

1803 through the following evaluation rules:

1804 **C1. Validity Rule:** Assigns +1 for chemically
1805 valid molecules satisfying valence and aromaticity constraints, -1 otherwise.
1806

1807 **C2. Stability Rule:** Penalizes high-strain configurations by $-E_{\text{strain}}$, where E_{strain} is the strain
1808 energy calculated using force field methods.
1809

1810 **C3. Functional Group Rule:** Rewards successful formation of target functional groups via
1811 $\sum_g w_g \mathbb{I}\{\text{form } g\}$, where w_g weights functional group g .
1812
1813

1814 **C4. Property Alignment Rule:** Measures deviation from target properties as $-|y_{\text{pred}} - y_{\text{target}}|$.
1815

1816 **C5. Synthetic Accessibility Rule:** Incorporates the SA score from RDKit, favoring easily synthesizable structures.
1817
1818

1819 **C6. Novelty Rule:** Encourages structural diversity via $1 - \max_{\text{ref}} \text{Tanimoto}(\text{mol}, \text{ref})$.
1820

1821 J.2 Topological Shaping Rewards

1822 The topological shaping reward R_{topo} guides correct 2D topology construction through the following shaping rules:
1823
1824

1825 **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.
1826
1827
1828

1829 **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.
1830
1831

1832 **T3. Topological Distance Rule:** Penalizes structural deviation via $-\text{GraphEditDistance}(G_{t+1}, G^*)$.
1833
1834

1835 **T4. Over-connection Rule:** Prevents excessive edges by $-\alpha \cdot \max(0, |E_{t+1}| - |E^*|)$ with penalty coefficient α .
1836
1837

1838 J.3 Potential-Based Reward Shaping

1839 To accelerate learning while preserving optimal policy invariance (Ng et al., 1999), we apply potential-based shaping with potential function:
1840
1841

$$1842 \Phi(s) = -|E^* \setminus E(s)| - \beta \cdot \text{cc}(G_s) \quad (20)$$

1843 where β weights the connectivity term. The shaped reward is computed as $r'(s_t, a_t, s_{t+1}) = r(s_t, a_t, s_{t+1}) + \gamma \Phi(s_{t+1}) - \Phi(s_t)$.
1844
1845

1846 K Prompt Templates and Downstream Task Training 1847

1848 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.
1849
1850
1851

1852 K.1 Initial Training Prompt Template

1853 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.
1854
1855
1856
1857

1858 K.1.1 Initial Setup Section

1859 The prompt begins with comprehensive molecular information:
1860

```
1861 ### Initial Setup
1862 **Target Molecule**:
```

1863 {molecule_name}

1864 - SMILES: '{smiles_string}'

1865 - Target Properties:

1866 {property_description}, LogP

1867 \approx {logp_value}, MW =

1868 {molecular_weight}

1869 - PubChem CID: {cid} (if available)

1870

1871 **Fragmentation Results** (via BRICS):

1872

1873 - **Motif_1**:

1874 {motif_description}

1875 ('{motif_smiles}', {n_atoms} atoms)

1876

1877 - **Motif_2**:

1878 {motif_description}

1879 ('{motif_smiles}', {n_atoms} atoms)

1880

1881 ...

1882 - **Motif_n**:

1883 {motif_description}

1884 ('{motif_smiles}', {n_atoms} atoms)

1885

1886 K.1.2 Episode Start Section

1887 The episode begins with the initial environment state and global coordinator instruction:
1888

1889	### Episode Start: Multi-Agent	Chemical	Context:	1931
1890	Dialogue	{chemical_environment_description}		1932
1891	**Environment State (t=0)**:	Proposal: {connection_proposal}		1933
1892	Current Assembly: Empty graph	Reasoning: {chemical_reasoning}		1934
1893	Connected Components: 0	Priority: {HIGH/MEDIUM/LOW} -		1935
1894	Available Motifs: [Motif_1,	{priority_explanation}		1936
1895	Motif_2, ..., Motif_n]	For subsequent rounds, agents may include status updates and modified fields:		1937
1896	Target Edges: {n_edges}	**Agent_{i} ({motif_name})*:		1938
1897	{edge_descriptions}	Status Update:		1939
1898	Current Properties: None	{occupied_sites_description}		1940
1899	**Global Coordinator**:	Available Sites:		1941
1900	“Agents, we need to assemble	[{remaining_site_list}]		1942
1901	{molecule_name}. The target has	Proposal: {connection_proposal}		1943
1902	{property_description} with LogP	Reasoning: {chemical_reasoning}		1944
1903	around {logp_value}. Current	Chemical Insight:		1945
1904	topology is empty. Please propose	{additional_chemical_insight}		1946
1905	your initial connections.”	Alternatively, agents waiting for dependencies use:		1947
1906	K.1.3 Round Structure	**Agent_{i} ({motif_name})*:		1948
1907	Each assembly round follows a consistent structure	Waiting State:		1949
1908	with a descriptive title:	{dependency_description}		1950
1909	### Round {N}: {Round_Title}	Future Plan: {planned_action}		1951
1910	For subsequent rounds (t > 0), the environment	Property Prediction:		1952
1911	state is updated:	{expected_property_impact}		1953
1912	**Environment State	K.1.5 Central Arbitrator Section		1954
1913	(t={step})*:	The central arbitrator evaluates all proposals and makes decisions:		1955
1914	Current Assembly:	**Central Arbitrator		1956
1915	{current_graph_description}	Evaluation**:		1957
1916	Connected Components:	Chemical Validity Check:		1958
1917	{num_components}	- {connection_1}: ✓ Valid,		1959
1918	Remaining Motifs:	{validity_reason}		1960
1919	[{remaining_motif_list}]	- {connection_2}: ✓ Valid,		1961
1920	Remaining Target Edges:	{validity_reason}		1962
1921	{num_remaining_edges}	Topological Scoring:		1963
1922	Current Properties:	S({connection_1}) = {v1}		1964
1923	{current_property_estimates}	(stability) + {v2} (progress) +		1965
1924	K.1.4 Multi-Agent Proposal Section	{v3} (property) = {total_1}		1966
1925	Each motif agent generates proposals. For the initial round (t=0), agents use the full proposal format:	S({connection_2}) = {v1}		1967
1926		(stability) + {v2} (progress) +		1968
1927	**Agent_{i} ({motif_name})*:	{v3} (property) = {total_2}		1969
1928	My Structure:	Decision: Execute		1970
1929	{structural_description}	{selected_connection} connection		1971
1930	Available Sites: [{site_list}]	first		1972

1976	K.1.6 Action Execution and Reward Section	K.1.8 Chemical Explanation Section	2020
1977	Each executed action follows the format:	The episode concludes with a generated chemical explanation:	2021
1978	**Action Executed**:	**Chemical Explanation	2023
1979	‘Connect: Motif_{i}[{site_x}]	Generated**:	2024
1980	-{bond_type}->	“{detailed_chemical_rationale_explaining_assembly	
1981	Motif_{j}[{site_y}]’		
1982	**Reward Calculation**:		
1983	r_valid = {value} ({explanation})	K.2 Downstream Task Prompt Templates	2026
1984	r_stable = {value}	For downstream tasks, we extend the base template	2027
1985	({explanation})	with task-specific components. All downstream	2028
1986	r_func = {value} ({explanation})	tasks share the same initial setup and multi-agent	2029
1987	r_conn = {value} ({explanation})	interaction format, with task-specific output sections	2030
1988	r_edge = {value} ({explanation})	appended.	2031
1989	Total: R = {total_reward}	K.2.1 Molecular Property Prediction	2032
1990	For terminal states, additional rewards are included:	For property prediction tasks, the prompt includes	2033
1991		property-specific analysis phases:	2034
1992	r_prop = {value}	### Initial Setup	2035
1993	({property_match_explanation})	**Task**: Molecular Property Prediction	2036
1994	Terminal Bonus = {value}		2037
1995	(successful reconstruction)	**Property**: {property_name}	2038
1996	Total: R = {total_reward}	(e.g., Blood-Brain Barrier Penetration)	2039
1997	K.1.7 Termination Section		2040
1998	Upon successful assembly:	**Target Molecule**:	2041
1999	### Round {N}: Termination Decision	- SMILES: ‘{smiles}’	2042
2000		- PubChem CID: {cid} (if available)	2043
2001	**Environment State		2044
2002	(t={final_step})**:	**Fragmentation Results** (via BRICS):	2045
2003	Current Assembly: Complete	[Same format as initial training]	2046
2004	{molecule_name} molecule	—	2047
2005	Connected Components: 1		2048
2006	Target Edges Matched: {n}/{n} ✓	### Episode Start: Multi-Agent Analysis	2049
2007	Properties: LogP = {value}, MW =	[Standard environment state and	2050
2008	{value}, {additional_properties}	global coordinator instruction]	2051
2009	**Global Coordinator**:	[Multi-agent interaction rounds	2052
2010	“Assembly complete. All target	for topology reconstruction]	2053
2011	edges matched. Properties within	—	2054
2012	specifications.”		2055
2013	**Agent_1 ({motif_name})**:	### Property Prediction Phase	2056
2014	“{confirmation_statement}”	**Environment State (Final)**:	2057
2015	**Agent_2 ({motif_name})**:	Current Assembly: Complete	2058
2016	“{confirmation_statement}”	molecule	2059
2017	...	Connected Components: 1	2060
2018	**Policy Decision**: ‘Action:	Target Edges Matched: {n}/{n} ✓	2061
2019	STOP (probability = {prob})’		

2062	Reconstructed	Topology:	**Global Coordinator**:	“Agents,	2104
2063	{topology_summary}		reconstruct this molecule step	2105	
2064	**Global Property Analyzer**:		by step. Each connection	2106	
2065	Topology-Property Mapping:		should reveal structural	2107	
2066	- Identified pharmacophores:		insights for description	2108	
2067	{pharmacophore_list}		generation. Focus on functional	2109	
2068	- Key structural alerts:		groups, pharmacophores,	2110	
2069	{alert_list}		and structure-activity	2111	
2070	- Electronic effects:		relationships.”	2112	
2071	{electronic_description}		[Multi-agent interaction rounds]	2113	
2072	- Steric factors:		—	2114	
2073	{steric_description}		### Description Generation Phase	2115	
2074	Property-Contributing Features:		**Trajectory Summary**:	2116	
2075	1. {feature_1}:		Step 1: {action_description} →	2117	
2076	{contribution_to_property}		{chemical_insight}	2118	
2077	2. {feature_2}:		Step 2: {action_description} →	2119	
2078	{contribution_to_property}		{chemical_insight}	2120	
2079	**Prediction Output**:		...	2121	
2080	Property: {property_name}		**Generated Description**:	2122	
2081	Prediction: {0/1 or value}		“{natural_language_description_covering_chemical_	2123	
2082	Confidence: {score}		K.2.3 Chemical Reaction Prediction	2124	
2083	Key Evidence:		For reaction prediction tasks (forward reaction and	2125	
2084	- {evidence_1}		retrosynthesis):	2126	
2085	- {evidence_2}		### Initial Setup	2127	
2086	K.2.2 Molecular Description Generation		**Task**:	2128	
2087	For description generation, the output section cap-		{Forward Reaction	2129	
2088	tures the assembly trace:		Prediction / Retrosynthesis}		
2089	### Initial Setup		**Input** (Forward):	2130	
2090	**Task**:	Molecular Description	- Reactants: ‘{reactant_smiles}’	2131	
2091	Generation		- Reagents: {reagent_info}	2132	
2092	**Target Molecule**:		**Input** (Retrosynthesis):	2133	
2093	- SMILES: ‘{smiles}’		- Product: ‘{product_smiles}’	2134	
2094	- Known Properties:		**Fragmentation Results** (via	2135	
2095	{known_properties}		BRICS):	2136	
2096	- PubChem CID: {cid}		[Motif decomposition of	2137	
2097	**Fragmentation Results** (via		reactants/products]	2138	
2098	BRICS):		—	2139	
2099	[Same format as initial training]		### Multi-Agent Reaction Analysis	2140	
2100	—		[Agents analyze reaction centers,	2141	
2101	### Episode Start: Multi-Agent		bond changes, electron flow]	2142	
2102	Assembly for Description		[Each agent identifies its role	2143	
2103	[Standard environment state]		in the transformation]	2144	
			—	2145	
			### Reaction Prediction Phase	2146	

2147	**Mechanism Analysis**:		
2148	- Reaction center:		
2149	{identified_centers}		
2150	- Bond breaking: {bonds_broken}		
2151	- Bond forming: {bonds_formed}		
2152	- Electron transfer pathway:		
2153	{electron_flow_description}		
2154	**Predicted Output**:		
2155	{SELFIES_representation}		
2156	**Chemical Explanation**:		
2157	{mechanistic_rationale}		
2158	K.3 Downstream Task Training Procedures		
2159	We adopt a two-stage training paradigm: initial		
2160	training on molecular reconstruction followed by		
2161	task-specific fine-tuning. Due to computational		
2162	constraints with the Qwen2.5-7B backbone, we		
2163	employ parameter-efficient fine-tuning strategies		
2164	rather than full model fine-tuning.		
2165	K.3.1 Training Strategy Overview		
2166	For all downstream tasks, we load the pre-trained		
2167	MotifAgent checkpoint and apply task-specific		
2168	adaptations. The initial training phase uses LoRA		
2169	(rank=16) to train the shared LLM backbone, and		
2170	this approach is maintained during downstream		
2171	fine-tuning to preserve the learned chemical knowl-		
2172	edge while adapting to specific tasks.		
2173	Molecular Property Prediction. This task		
2174	employs supervised learning with frozen multi-		
2175	agent interaction modules. We add a topology-		
2176	aware classification head that pools LLM hidden		
2177	states weighted by motif importance. Only the		
2178	classification head parameters are trained, while		
2179	the LLM backbone with LoRA adapters remains		
2180	frozen. This preserves the pre-trained chemical un-		
2181	derstanding while learning task-specific decision		
2182	boundaries. The training uses cross-entropy loss		
2183	with auxiliary topology consistency regularization.		
2184	Molecular Description Generation. This task		
2185	requires generating natural language outputs, ne-		
2186	cessitating adaptation of the language generation		
2187	capabilities. We fine-tune the model using LoRA		
2188	adapters (rank=8) on the LLM backbone combined		
2189	with a task-specific generation head. The multi-		
2190	agent assembly process first reconstructs the molec-		
2191	ular topology, then the generation head produces		
2192	descriptions conditioned on the assembly trace.		
2193	Training uses teacher forcing with cross-entropy		
2194	loss on the description tokens.		
	Chemical Reaction Prediction. Both forward		2195
	reaction and retrosynthesis prediction require un-		2196
	derstanding molecular transformations at the motif		2197
	level. We employ LoRA fine-tuning (rank=8)		2198
	combined with reaction-specific output heads. The		2199
	multi-agent framework identifies reaction centers		2200
	by analyzing motif-level changes between reac-		2201
	tants and products. For retrosynthesis, agents pro-		2202
	pose bond disconnections that reverse the assembly		2203
	process. Training combines supervised learning		2204
	on reaction outcomes with the MARL framework		2205
	for exploring chemically valid transformation path-		2206
	ways.		2207
	K.3.2 Fine-tuning Configuration		2208
	All downstream tasks share the following base con-		2209
	figuration with task-specific modifications:		2210
	• Backbone: Pre-trained MotifAgent		2211
	(Qwen2.5-7B with LoRA)		2212
	• LoRA Configuration: rank=8, alpha=16, tar-		2213
	get modules: [q_proj, v_proj]		2214
	• Optimizer: AdamW with weight decay 0.01		2215
	• Learning Rate: 2e-5 for task heads, 1e-5 for		2216
	LoRA parameters		2217
	• Warmup: 10% of total training steps		2218
	• Precision: BFloat16 mixed precision		2219
	For property prediction, we use batch size 32		2220
	and train for 50 epochs with early stopping (pa-		2221
	tience=10) based on validation ROC-AUC. For de-		2222
	scription generation and reaction prediction, we use		2223
	batch size 16 and train for 30 epochs, monitoring		2224
	BLEU scores and exact match rates respectively.		2225
	K.3.3 Inference Procedure		2226
	During inference, the model operates in reconstruc-		2227
	tion mode for understanding tasks (property predic-		2228
	tion, description generation) and generation mode		2229
	for synthesis tasks. In reconstruction mode, Set-		2230
	BC priors guide the assembly to recover the target		2231
	topology before applying the task-specific head.		2232
	In generation mode for reaction prediction, beam		2233
	search (k=5) explores multiple chemically valid		2234
	pathways, with the central arbitrator ensuring chem-		2235
	ical validity at each step.		2236
	The three-level explanation framework (topol-		2237
	ogy construction, connection mechanisms, and		2238
	topology-property relationships) is generated		2239

2240 alongside task outputs, providing interpretable rea- 2289
2241 soning traces that connect molecular assembly de- 2290
2242 cisions to final predictions. 2291

2243 K.4 Dataset Descriptions 2292

2244 This section provides detailed descriptions of all 2293
2245 datasets used in our experiments, including statis- 2294
2246 tics and data splitting strategies. 2295

2247 K.4.1 Initial Training Dataset 2296

2248 For initial training, we utilize the molec- 2297
2249 ular SMILES-text pairs dataset from 2298
2250 MoleculeSTM (Liu et al., 2023a), which is 2299
2251 collected from the PubChem database. Following 2300
2252 the preprocessing pipeline of MoleculeSTM, pairs 2301
2253 with identical PubChem IDs are merged, descrip- 2302
2254 tions shorter than 18 characters are filtered out, and 2303
2255 duplicates from downstream task datasets are re- 2304
2256 moved to prevent data leakage. This yields **51,340** 2305
2257 **unique high-quality molecule-description pairs** 2306
2258 for initial training. 2307

2259 For each molecule, we apply our improved 2308
2260 BRICS fragmentation algorithm with 16 chem- 2309
2261 ical environment-specific bond-breaking rules. 2310
2262 Molecules containing 2-15 motifs are selected 2311
2263 for training to ensure appropriate complexity for 2312
2264 the multi-agent assembly task. The dataset cov- 2313
2265 ers diverse chemical space including drug-like 2314
2266 molecules, natural products, and synthetic com- 2315
2267 pounds, providing comprehensive coverage of com- 2316
2268 mon motif connection patterns. 2317

2269 K.4.2 Molecular Property Prediction Datasets 2318

2270 We evaluate MotifAgent on 8 benchmark classifica- 2319
2271 tion datasets from MoleculeNet (Wu et al., 2018). 2320
2272 All datasets use **scaffold splitting** with an 80/10/10 2321
2273 train/validation/test ratio, which provides a more 2322
2274 rigorous test of model generalizability by ensur- 2323
2275 ing that molecules in different splits have distinct 2324
2276 molecular scaffolds. Table 12 summarizes the 2325
2277 dataset statistics. 2326

2278 **BBBP** (Blood-Brain Barrier Penetration): Con- 2327
2279 tains binary labels indicating whether compounds 2328
2280 can penetrate the blood-brain barrier, which is cru- 2329
2281 cial for central nervous system drug development. 2330

2282 **Tox21**: Created by the “Toxicology in the 21st 2331
2283 Century” initiative, this dataset contains qualita- 2332
2284 tive toxicity measurements across 12 different tar- 2333
2285 gets including nuclear receptors and stress response 2334
2286 pathways. 2335

2287 **ToxCast**: A larger toxicology dataset from the 2336
2288 same initiative as Tox21, providing toxicity data 2337

2289 across 617 high-throughput assays.

2290 **SIDER**: Contains information about marketed 2291
2292 drugs and their recorded adverse drug reactions, 2292
2293 organized into 27 system organ classes. 2292

2293 **ClinTox**: Compares FDA-approved drugs with 2293
2294 drugs that failed clinical trials for toxicity reasons, 2294
2295 comprising two binary classification tasks. 2295

2296 **MUV**: Maximum Unbiased Validation dataset 2296
2297 designed for virtual screening benchmarking, fea- 2297
2298 turing challenging classification tasks with high 2298
2299 class imbalance. 2299

2300 **HIV**: Contains experimental results on the abil- 2300
2301 ity of compounds to inhibit HIV replication. 2301

2302 **BACE**: Provides quantitative binding results for 2302
2303 inhibitors of human β -secretase 1 (BACE-1), con- 2303
2304 verted to binary classification. 2304

2305 K.4.3 Molecular Description Generation 2305 2306 Dataset 2306

2307 For molecular description generation, we use the 2307
2308 **ChEBI-20** benchmark dataset (Papadatos et al., 2308
2309 2015), which is derived from the Chemical Entities 2309
2310 of Biological Interest (ChEBI) database combined 2310
2311 with PubChem annotations. Table 13 shows the 2311
2312 dataset statistics. 2312

2313 The ChEBI-20 dataset contains molecule- 2313
2314 description pairs where each description has 2314
2315 more than 20 words, providing detailed textual 2315
2316 annotations covering chemical structures, func- 2316
2317 tional groups, physicochemical properties, and 2317
2318 biological activities. We follow the standard 2318
2319 train/validation/test split provided by the original 2319
2320 dataset. 2320

2321 K.4.4 Chemical Reaction Prediction Datasets 2321

2322 For chemical reaction prediction tasks (for- 2322
2323 ward reaction prediction, retrosynthesis, and 2323
2324 reagent prediction), we use the **Mol-Instructions** 2324
2325 dataset (Fang et al., 2023). This dataset is part 2325
2326 of a large-scale biomolecular instruction dataset 2326
2327 designed for large language models, containing 2327
2328 molecule-oriented instructions across six tasks. 2328

2329 All reaction data are represented using SELFIES 2329
2330 (Self-Referencing Embedded Strings) format to en- 2330
2331 sure 100% validity of molecular representations. 2331
2332 The dataset follows the default train/validation/test 2332
2333 split provided by Mol-Instructions, with an approx- 2333
2334 imate 80/10/10 ratio. 2334

2335 **Forward Reaction Prediction**: Given reactants 2335
2336 and reagents, predict the products of the chemical 2336
2337 reaction. 2337

2338 **Retrosynthesis:** Given a target product
2339 molecule, predict suitable reactants that could syn-
2340 thesimize the target. This is the inverse problem of
2341 forward reaction prediction and represents a fun-
2342 damental challenge in AI-assisted synthetic route
2343 planning.

2344 **Reagent Prediction:** Given reactants and prod-
2345 ucts, predict the reagents or conditions required to
2346 facilitate the reaction.

2347 **K.4.5 Data Preprocessing for MotifAgent**

2348 For all datasets, molecules are processed through
2349 our unified preprocessing pipeline:

- 2350 1. **BRICS Decomposition:** Each molecule is de-
2351 composed into chemically meaningful motifs
2352 using our improved BRICS algorithm with
2353 16 bond-breaking rules. Motifs containing at
2354 least 2 non-hydrogen atoms are retained.
- 2355 2. **Structured Text Serialization:** Each mot-
2356 if is serialized into structured text contain-
2357 ing: unique identifier, SMILES string, atom
2358 and bond information, connectable sites with
2359 chemical environment annotations, and prop-
2360 erty summaries (aromaticity, ring structures,
2361 functional groups).
- 2362 3. **Connection Graph Construction:** The
2363 molecular graph’s connectivity information
2364 is preserved through explicit connection tem-
2365 plates encoding adjacency relationships be-
2366 tween motifs.
- 2367 4. **Instruction Format Conversion:** Data is
2368 converted to task-specific instruction formats
2369 following the templates described in Ap-
2370 pendix K.

2371 Molecules that cannot be properly decomposed
2372 (e.g., those resulting in fewer than 2 or more than
2373 15 motifs) are excluded from the respective datasets
2374 to ensure compatibility with the multi-agent frame-
2375 work.

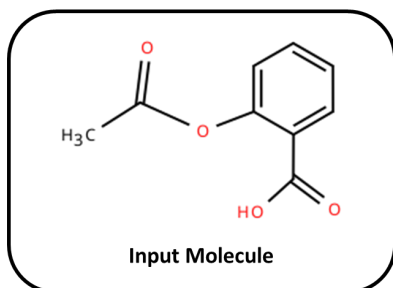
2376 **L Cases Studies**

2377 In this section, we demonstrate the conversational
2378 and decision-making processes of MotifAgent
2379 through four representative cases. We first present
2380 a standard Algorithm Demonstration using the
2381 aspirin molecule, followed by detailed assembly
2382 workflows on three molecules of varying complex-
2383 ity: a simple molecule (Paracetamol), a complex

2384 molecule (Ibuprofen), and a complex heterocyclic
2385 molecule (Omeprazole).

2386 Given the input molecular SMILES, target prop-
2387 erties, and fragmentation results, each motif estab-
2388 lishes its individual profile and analyzes its current
2389 state and relationships with other motifs. Through
2390 chemical validity assessment and topological scor-
2391 ing, agents evaluate the benefits of establishing
2392 inter-motif connections, execute connection ac-
2393 tions, and compute corresponding rewards. After
2394 multiple rounds of negotiation, all motifs are suc-
2395 cessfully connected, followed by a comprehensive
2396 validation to ensure molecular validity and prop-
2397 erty satisfaction. The process concludes with the
2398 output of the final assembled molecule.

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



Initial Setup

Target Molecule : Aspirin ((Acetylsalicylic acid))

SMILES : CC(=O)Oc1ccccc1C(=O)O

Target Properties: Anti-inflammatory (COX inhibition), $\text{LogP} \approx 1.2$, $\text{MW} = 180.16$

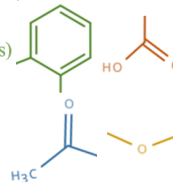
Fragmentation Results:

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):

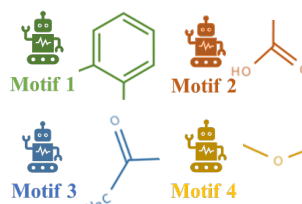
Current Assembly: Empty graph

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.

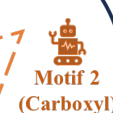
Round 1: Initial Core Selection



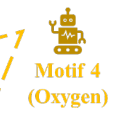
My Structure: Aromatic ring, 6 carbons
Available Sites: [ortho_1, ortho_2, meta_1, meta_2, para]
Chemical Context: I'm a stable aromatic core, ideal as scaffold
Proposal: I should be the starting point. Ortho substitution pattern typically yields aspirin-like properties.
Priority: HIGH - I'm the largest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



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



My Structure: C(=O)O, acidic functional group
Available Sites: [carbon_connection]
Chemical Context: I provide acidity ($\text{pKa} \sim 3-4$) 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



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): \checkmark Valid, maintains aromaticity
- Benzene-Oxygen (direct): \checkmark Valid, forms phenolic ether
- Acetyl-Oxygen: \checkmark Valid, forms ester

Topological Scoring:

$S(\text{benzene} \rightarrow \text{carboxyl}) = 0.8$ (stability) + 0.9 (progress) + 0.7 (property) = 2.4

$S(\text{benzene} \rightarrow \text{oxygen}) = 0.7$ (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:

$r_{\text{valid}} = +1.0$ (valid connection)

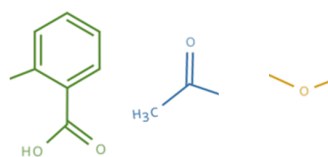
$r_{\text{stable}} = -0.1$ (minimal strain)

$r_{\text{func}} = +0.5$ (formed benzoic acid substructure)

$r_{\text{conn}} = +1.0$ (created first connected component)

$r_{\text{edge}} = +1.0$ (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 \approx 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 \sim 1.2 (target range)

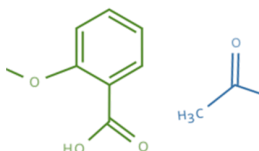
Analysis: Benzene has ortho_2 available, perfect for aspirin pattern
Proposal: Connect to Motif_1[ortho_2]
Reasoning: Sets up ester formation with acetyl group
Compatibility: Maintains sp² oxygen, allows resonance with benzene


Motif 4
(Oxygen)

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 ($\Delta G < 0$)

Environment State (t=2):

Current Assembly: ortho-hydroxyl benzoic acid
Connected Components: 1
Remaining Motifs: [Motif_3]
Remaining Target Edges: 1
Current Properties: LogP \approx 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 \checkmark (matches target)
 - LogP: \sim 1.19 \checkmark (within range)
 - H-bond donors: 1 \checkmark (COOH only)
 - H-bond acceptors: 4 \checkmark (typical for aspirin)
 - Bioactivity: High confidence for COX inhibition

Final Reward Calculation:

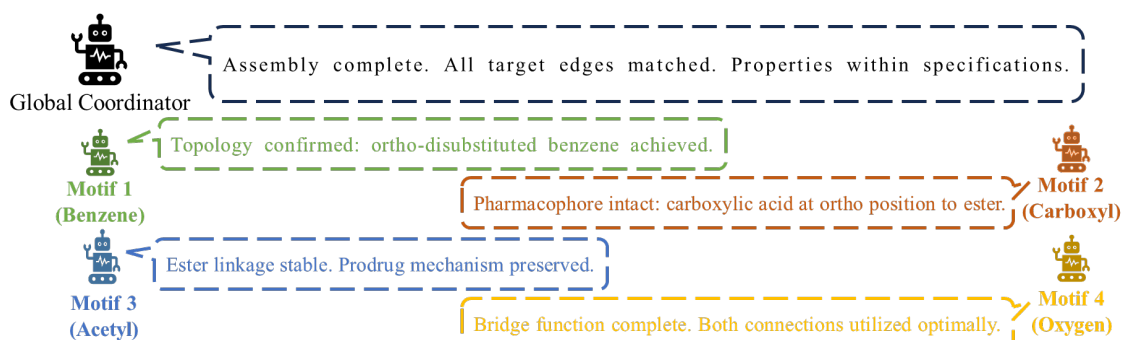
r_valid = +1.0
 r_stable = -0.05 (stable ester)
 r_func = +1.0 (completed aspirin structure)
 r_prop = +0.9 (LogP within 0.01 of target)
 r_edge = +1.0 (final target edge)
 r_conn = 0 (already connected)
 Terminal Bonus = +5.0 (successful reconstruction)
Total: R = 8.85

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

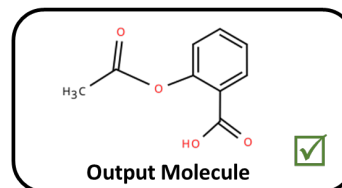
Round 4: Termination Decision

Environment State (t=3):

Current Assembly: Complete aspirin molecule
Connected Components: 1
Target Edges Matched: 3/3 ✓
Properties: LogP = 1.19, MW = 180.16, COX inhibitor



Policy Decision: Action: STOP (probability = 0.95)



Episode Summary

Trajectory Interpretation:

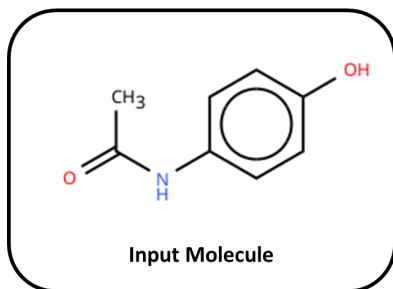
- Step 1: Established salicylic acid core (benzene-COOH)
 - Created primary pharmacophore
- Step 2: Added oxygen at ortho position
 - Prepared ester formation site
- Step 3: Completed acetyl ester
 - Achieved target molecule with desired properties

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



Initial Setup

Target Molecule : Paracetamol

SMILES : CC(=O)Nc1ccc(O)cc1

Target Properties: Analgesic/antipyretic, $\text{LogP} \approx 0.46$, $\text{MW} = 151.16$

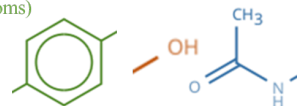
PubChem CID: 1983

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

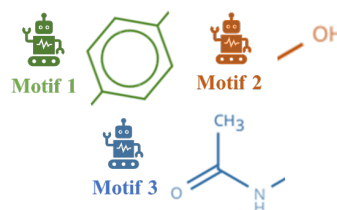
Connected Components: 0

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): \checkmark Valid, forms p-hydroxyphenyl
- Benzene-Acetamide (para): \checkmark Valid, forms p-acetamidophenyl

Topological Scoring:

$S(\text{benzene} \rightarrow \text{hydroxyl}) = 0.9$ (stability) + 0.8 (progress) + 0.9 (property) = 2.6

$S(\text{benzene} \rightarrow \text{acetamide}) = 0.8$ (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:

$r_{\text{valid}} = +1.0$ (valid connection)

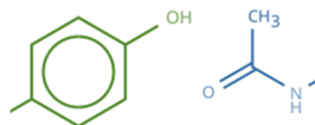
$r_{\text{stable}} = -0.05$ (minimal strain, para position)

$r_{\text{func}} = +0.8$ (formed phenolic group)

$r_{\text{conn}} = +1.0$ (created first connected component)

$r_{\text{edge}} = +1.0$ (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 \approx 10 (phenolic), partial analgesic activity



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

Analysis: Para position to hydroxyl is optimal
Proposal: Connect to Motif_1[para_2]
Property Prediction: Will achieve target LogP \sim 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):

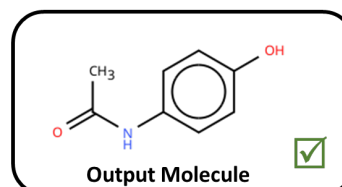
Current Assembly: Complete paracetamol molecule
Connected Components: 1
Target Edges Matched: 2/2 ✓
Properties: LogP = 0.47, MW = 151.16, pKa = 9.38



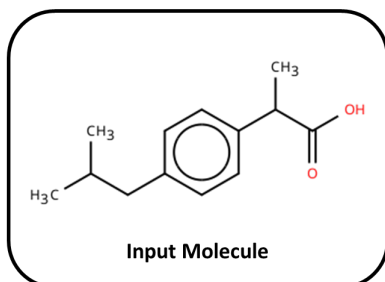
Assembly complete. All target edges matched. Properties within specifications.

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.



Case 3: Ibuprofen - Complex Molecule



Initial Setup

Target Molecule : Ibuprofen

SMILES : CC(C)Cc1ccc(C(C)C(=O)O)cc1

Target Properties: NSAID, LogP \approx 3.97, MW = 206.28

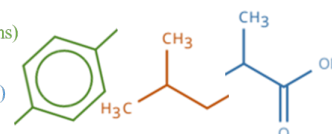
PubChem CID: 3672

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

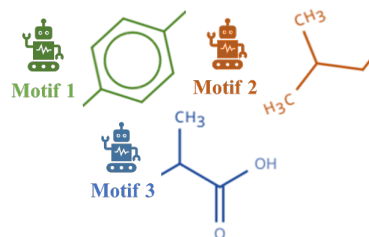
Connected Components: 0

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 -Methylpropionic 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: \checkmark Valid, increases lipophilicity
- Benzene-Propionic: \checkmark Valid, forms arylpropionic acid

Topological Scoring:

$S(\text{benzene} \rightarrow \text{propionic}) = 0.8$ (stability) + 0.9 (progress) + 1.0 (property) = **2.7**

$S(\text{benzene} \rightarrow \text{isobutyl}) = 0.9$ (stability) + 0.7 (progress) + 0.6 (property) = **2.2**

Decision: Execute benzene-propionic connection first

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

Reward Calculation:

$r_{\text{valid}} = +1.0$ (valid connection)

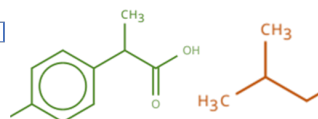
$r_{\text{stable}} = -0.1$ (slight steric from α -methyl)

$r_{\text{func}} = +1.0$ (formed arylpropionic acid pharmacophore)

$r_{\text{conn}} = +1.0$ (created first connected component)

$r_{\text{edge}} = +1.0$ (matched target edge)

Total: R = 3.9



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 \approx 4.5, LogP \approx 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 \sim 3.97 (target)
COX Interaction: Forms hydrophobic channel interaction


Motif 2
(Isobutyl)

Global Property Predictor

If connection completed:

- MW: 206.28 \checkmark (matches target)
- LogP: \sim 3.96 \checkmark (optimal for oral absorption)
- pKa: 4.4 \checkmark (good tissue penetration)
- Rotatable bonds: 4 \checkmark (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 \checkmark
Properties: LogP = 3.96, MW = 206.28, COX-2 selective


Global Coordinator

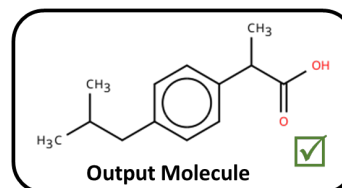
Assembly complete. Ibuprofen structure achieved with optimal lipophilicity.


Motif 3
(Propionic acid)

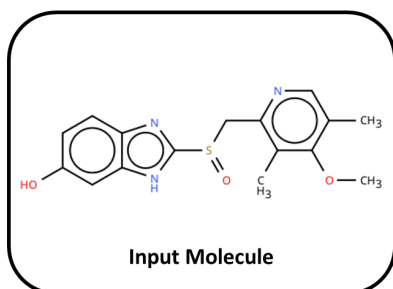
The α -methyl substitution pattern is crucial for reduced GI side effects compared to non-selective NSAIDs.

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.



Case 4: Omeprazole - Complex Heterocyclic Molecule



Initial Setup

Target Molecule : Omeprazole

SMILES : Oc1ccc2nc(S(=O)C)c3ncc(C)c(OC)c3C[nH]c2c1

Target Properties: Proton pump inhibitor, $\text{LogP} \approx 2.23$, $\text{MW} = 345.42$

PubChem CID: 4594

Fragmentation Results:

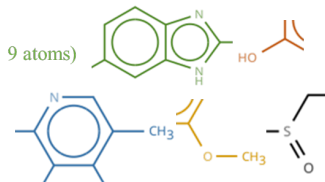
Motif 1: Benzimidazole core (c1ccc2[nH]c(-)nc2c1, 9 atoms)

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

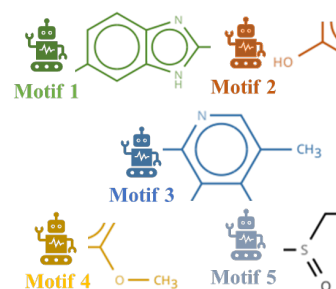
Connected Components: 0

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, $\text{LogP} \approx 2.23$



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, $\text{pKa} \sim 6$
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 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

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 2
(Methoxy)

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 4
(Methoxy-methyl)

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

Motif 5
Sulfinyl methyl Bridge

Central Arbitrator Evaluation

Chemical Validity Check:

- Benzimidazole-Methoxy (C5): \checkmark Valid, enhances basicity
- Benzimidazole-Sulfoxide (C2): \checkmark Valid, typical PPI pattern

Topological Scoring:

$$S(\text{benzimidazole} \rightarrow \text{methoxy}) = 0.9 (\text{stability}) + 0.8 (\text{progress}) + 0.9 (\text{property}) = 2.6$$

$$S(\text{benzimidazole alone}) = 0.8 (\text{stability}) + 0.7 (\text{progress}) + 0.6 (\text{property}) = 2.1$$

Decision: Connect benzimidazole with methoxy first

Reward Calculation:

$$r_{\text{valid}} = +1.0 (\text{valid connection})$$

$$r_{\text{stable}} = -0.05 (\text{minimal strain})$$

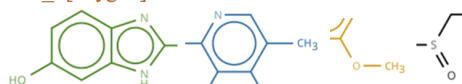
$$r_{\text{func}} = +0.8 (\text{enhanced benzimidazole basicity})$$

$$r_{\text{conn}} = +1.0 (\text{created first connected component})$$

$$r_{\text{edge}} = +1.0 (\text{matched target edge})$$

$$\text{Total: } R = 3.75$$

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

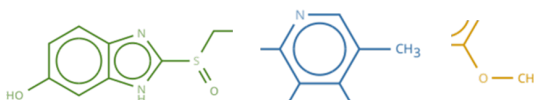
Analysis: Benzimidazole C2 properly activated
Proposal: Connect my sulfur to Motif_1[C2]
Mechanism: My sulfoxide will undergo Pummerer rearrangement in acid
Property Impact: Enables tissue-selective activation

 Motif 5
Sulfinyl methyl
Bridge

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

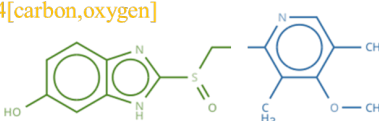
Central Arbitrator

Special Case: Motif_4 bridges two positions on pyridine

Creating: 3,5-dimethyl-4-methoxy substitution pattern

Decision: Execute complex bridging connection

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



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

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 (Sulfinylmethyl 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.

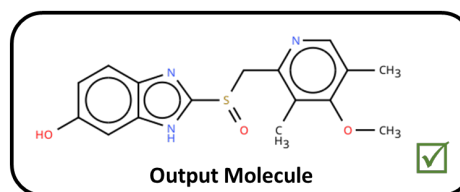
Motif 2 (Methoxy): Electron donation to benzimidazole optimized for pKa ~6.

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.

Table 11: Performance comparison on reaction prediction tasks.

Method	EXACT \uparrow	BLEU \uparrow	LEVENSHTEIN \downarrow	RDK FTS \uparrow	MACCS FTS \uparrow	MORGAN FTS \uparrow	VALIDITY \uparrow
<i>Reagent Prediction</i>							
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
<i>Forward Reaction Prediction</i>							
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
<i>Retrosynthesis</i>							
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 12: 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 13: 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 14: 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