Multimodal Chemical Structure-Text Coreference in Intellectual Property via Rule-guided Reinforcement Learning

Hanmeng Zhong, Wentao Wu, Linqing Chen*, Peng Zhou
PatSnap Co., LTD. Suzhou, China
{zhonghanmeng, wuwentao3, chenlinqing, tonyzhoupeng}@patsnap.com

Abstract

Navigating biopharmaceutical intellectual property necessitates precisely associating visual chemical structures with their textual referents across lengthy documents. Despite its critical role in drug discovery, this multimodal coreference task remains underexplored. It presents unique challenges, including handling Markush structures and distinguishing the atom-level differences between adjacent structures. To bridge this gap, we define the multimodal **Che**mical **S**tructure-**T**ext coreference and introduce **CheST**, the first dataset explicitly designed for the task. Furthermore, to satisfy the strict logical consistency in the task, we propose **RULER**, a **RULE**-guided multimodal **R**einforcement learning framework built upon an SFT cold start. RULER utilizes rule-driven reward functions operationalizing multi-dimensional consistencies, acting as a domain-specific "verifier" to obtain the correct domain knowledge. Experimental results demonstrate that RULER achieves a 40% improvement over the strongest baseline–Gemini-2.5-Pro, demonstrating the superior efficacy. [1]

1 Introduction

Biopharmaceutical intellectual property documents are characterized by a complex integration of textual descriptions and graphical representations. Within this multimodal context, chemical structure diagrams are frequently accompanied by textual referents—located in captions, floating labels, or adjacent paragraphs. The task of chemical structure-text coreference entails the chemical structure association with reference names and structure types. For example, textual reference information of the boxed structure in Figure 1 is "[Compound 204]: [substituent, Markush structure]". Accurately resolving these coreferences is

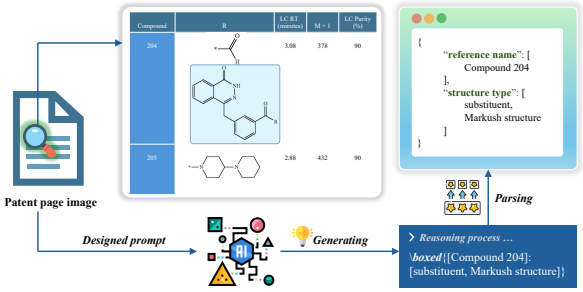


Figure 1: Example of Chemical Structure-Text Coreference Task

a prerequisite for downstream applications, such as knowledge graph construction (Jegal et al., 2023) and core structure recognition (Wang et al., 2024).

Despite its importance, no existing framework adequately addresses this multimodal challenge. Conventional Optical Chemical Structure Recognition (OCSR) tools (Rajan et al., 2020, 2023) prioritize pixel-to-molecular format conversion (e.g., SMILES) but neglect the necessary textual context. Conversely, general Document Visual Question Answering models lack the domain knowledge required to verify chemical validity. To bridge this gap, we define the **Che**mical **S**tructure-**T**ext coreference task as reference name matching and structure classification. Accordingly, we propose **CheST**, the first dataset dedicated to the multimodal coreference task, aiming to evaluate and improve the capability in this task.

Achieving robust performance on CheST is non-trivial due to three primary challenges:

Fine-Grained Visual Discrimination. Distinguishing highly similar structures requires detecting minute discrepancies, such as single atom substitutions or chirality, which is often confused.

Hierarchical Structure Complexity. Parsing Markush structures (generic scaffolds with variable substituents) requires handling nested logic where diagrams function as both local substituents

*corresponding author

[1]Our code and dataset can be seen in https://github.com/kkkeepgoing/RULER.

Findings of the Association for Computational Linguistics: ACL 2026, pages 29784–29796
July 2-7, 2026

and global scaffolds.

Layout and Reference Variability. References exhibit high syntactic irregularity, often crossing paragraph boundaries or involving many-to-one mappings that complicate alignment.

However, conventional Supervised Fine-Tuning (SFT) is often insufficient for these challenges. SFT optimizes for token-level likelihood rather than rigid scientific correctness, frequently leading to chemically plausible but factually inconsistent hallucinations. To address this, we propose **RULER**, a **RULE**-guided multimodal **R**einforcement learning method. RULER integrates a multi-objective reward system that acts as a domain-specific "verifier" for the cold-start SFT model, penalizing hallucinations and aligning the model's visual perception with strict chemical logic.

In summary, our work makes the following contributions:

- We formally define the task of multimodal chemical structure-text coreference in biopharmaceutical intellectual property and introduce CheST, the first dataset dedicated to this task.
- We conduct a systematic evaluation of general MLLMs on CheST. Our analysis reveals significant performance disparities and highlights specific deficiencies in reference matching and structure classification .
- We propose RULER, a novel method integrating rule-guided multimodal RL based on reasoning cold-start SFT, achieving state-of-the-art performance.
- We design a transferable and scalable domain-specific "verifier", which operationalizes strict chemical conventions into a multidimensional, computable reward function, explicitly penalizing hallucinations.

2 Related Work

Multimodal Large Language Models for Documents. Multimodal Large Language Models (MLLMs) for documents (Mathew et al., 2021) have evolved from foundational architectures like the LayoutLM series (Xu et al., 2020), which integrated OCR-derived spatial embeddings, to recent hybrid models like DocFormer (Appalaraju et al., 2021) and DocVLM that efficiently compress high-resolution visual features to handle "token explosion"(Nguyen et al., 2025; Nacson et al., 2025). While these excel at general document understanding, they typically lack the domain-specific granularity. In this work, we post-train Qwen-VL (Bai et al., 2023) through domain-specific rule-based verifications, enabling it to handle the rigid syntax and precise cross-modal alignment required for biopharmaceutical intellectual property analysis.

Information Extraction in Chemical Intellectual Property. Information extraction researches in chemical intellectual property focus on Named Entity Recognition (NER) and relation extraction to identify chemical compounds and their properties from unstructured text (Zhai et al., 2019; Wei et al., 2016). Advanced techniques in intellectual property analysis further resolve abbreviated references in reactions (Fang et al., 2021). However, these researches are predominantly text-centric, ignoring the visual modality entirely. Our work bridges this gap by introducing a multimodal coreference task in intellectual property, simultaneously analyzing visual regions and textual descriptions.

Optical Chemical Structure Detection & Recognition. Optical Chemical Structure Detection (Chemical Layout) (Filippov and Nicklaus, 2009; Staker et al., 2019; Rajan et al., 2021) identifies the bounding box coordinates of chemical diagrams, while Optical Chemical Structure Recognition (OCSR) translates these visual depictions into machine-readable formats like SMILES or MOL (Weininger, 1988; Heller et al., 2015; Qian et al., 2023). Our goal is orthogonal to OCSR: rather than translating pixels into molecular graphs, we focus on reference matching–resolving the semantic link between the image layout and textual identity.

Reinforcement Learning in Multimodal Scientific Tasks. Reinforcement Learning has become a powerful tool for aligning LLMs with scientific goals. Methods (Christiano et al., 2017) utilizing PPO (Schulman et al., 2017) and DPO (Rafailov et al., 2023) have been widely used in drug discovery, optimizing molecular generation for physical properties such as stability, drug-likeness, or protein fitness (Bou et al., 2024; Cao and Wang, 2025; Dharuman et al., 2024). Recently, value-free algorithms like GRPO (Shao et al., 2024) have shown promise in protein sequence design by leveraging group-relative feedback (Wang et al., 2025). Distinct from previous works that use RL to optimize physical or chemical properties of generated entities, we employ algorithms based on group-level

advantages (Zheng et al., 2025a; Yu et al., 2025) to enforce multimodal coreference correctness.

3 Tasks and Data

In the RULER method, we decompose the chemical structure-text coreference task into two subtasks: Reference Name Matching (**RefMatch**) and Structure Classification (**StruCls**).

3.1 Task Formulation

Intuitively, for each boxed chemical structure detected on a intellectual property page, the system must identify *what it is called* (reference name) and *what kind of structure it is* (structure type). Formally, we consider an intellectual property document as a sequence of page images $\{I_p\}_{p=1}^{P}$ with a set of auto-synthesized candidate structure boxes $\{B_i\}_{i=1}^{S}$. For each box B_i, the model predicts a structured payload y_i:

$$y_i \triangleq \big(\mathcal{N}_i, ; \mathcal{T}_i \subset \{\textit{Mark: }, \textit{subst: }, \textit{spec: }\}\big),$$

where $\mathcal{N}_i$ is a list of reference names extracted from the image context, and $\mathcal{T}_i$ denotes the chemical structure category. The final system grounds $\mathcal{N}i$ to occurrences across $\{I_p\}_{p=1}^{P}$ via a rule-guided matcher to achieve cross-page coreference.

3.2 Reference Name Matching (RefMatch)

RefMatch aims to extract the precise textual referent(s) $\mathcal{N}_i$ strictly from the document image that refer to the boxed structure b_i. Unlike general object detection, reference names in biopharmaceutical intellectual property exhibit high layout variability: they may appear in table headers, captions, synthesis schemes, or descriptive paragraphs.

We summarize the output $\mathcal{N}_i$ into four types:

- **Serial:** Alphanumeric reference names often used in schemes or tables (e.g., "*Compound 204, 205*" in Figure 1).
- **Spec:** Specific reference chemical names usually appear in synthesize descriptions (e.g., "*3-(Tributylstannyl)pyrazine-2-carbonitrile*").
- **Multi:** Cases where a single structure corresponds to multiple references (e.g., a Markush structure may represent several compounds).
- **None:** Structures with no reference names mentioned on the intellectual property page (e.g., a page listing structures without text).

The model must handle complex visual logic in some cases. For instance, if a table contains multiple structures representing one compound, the model needs to determine which compound each structure corresponds to through visual alignment relationships, and concatenate the table header with the serial numbers to form complete referential name results (e.g., the structure in Figure 1 represents "*Compound 204*" in the first line of the table).

3.3 Structure Classification (StruCls)

StruCls predicts the structure category $\mathcal{T}_i$ of the structure b_i based on its graphical features and global context, assisting the compound merging operation in the post-processing procedure.

Specifically, $\mathcal{T}_i$ may be one of the four subsets of $\{\textit{Mark: }, \textit{subst: }, \textit{spec: }\}$, namely $\{\textit{Mark: }\}$, $\{\textit{subst: }\}$, $\{\textit{spec: }\}$, and $\{\textit{subst: }, \textit{Mark: }\}$. *Mark:* means a generic structure containing variable groups and *subst:* means a fragment or functional group that is part of a larger molecule. Normally, *spec:* represents a fully defined chemical compound with no variables. Differently, the last category is a mixed structure type, acting as a substituent but possessing a Markush structure.

overall, accurate classification requires the simultaneous possession of excellent global and local comprehension capabilities. For instance, a substituent might lack explicit unbonded electrons in certain styles; the model need to infer its role from the surrounding descriptive context rather than focusing on the local structure alone.

3.4 Data and Annotation

To support the task, we present **CheST**, the first expert-annotated dataset for chemical structure-text coreference.

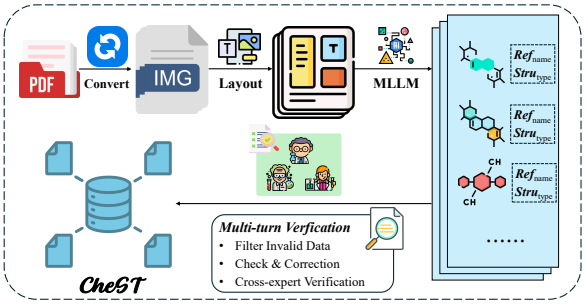


Figure 2: Construction Workflow of **CheST**

Data Collection. We collect raw PDF intellectual property from Google patents, covering the period from 2007 to 2023. To ensure domain rele-

vance, we specifically filter for biopharmaceutical patents containing complex synthesis schemes and Markush definitions. After layout analysis and initial filtering, we select pages containing at least five valid chemical structure diagrams.

Annotation Pipeline. As shown in Figure 2, the construction follows a "Model-Assisted, Expert-Refined" pipeline.

1. **Preprocessing:** PDFs are converted to high-resolution images. A layout tool [2] localizes chemical structures in each page image.
2. **LLM Pre-annotation:** A multimodal LLM (Gemini-2.5-pro) generates initial predictions for reference names and structure types to accelerate manual work.
3. **Expert Verification:** The core annotation team consists of five annotators. [3] They correct the pre-annotated labels and filtered out invalid data (e.g., the structure not been completely segmented within the layout phase).
4. **Quality Control:** To ensure consistency, all of the image-label pairs are cross-annotated. We achieve an inter-annotator agreement 99%, indicating high reliability.

The final dataset contains 2,000 high-quality chemical structure-text pairs.

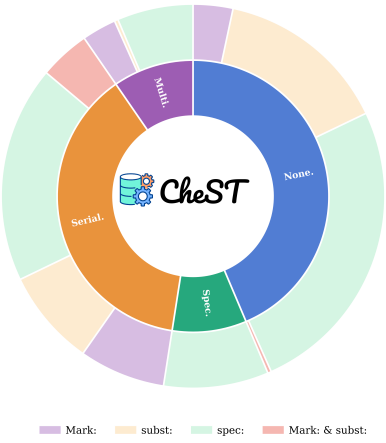


Figure 3: Type Statistics of CheST

Dataset Statistics. Figure 3 illustrates the hierarchical distribution of CheST. The inner ring represents **RefMatch** categories. We observe that **Serial.** and **None.** are the predominant types. The outer ring details the **StruCls** distribution for each reference type. Notably: (1). The **Serial.** category exhibits the highest diversity, mapping to *Mark:* , *subst:* , and *Mark: &subst:* , reflecting the complex referencing logic in reaction schemes; (2). The **Spec.** category maps almost exclusively to specific compounds; (3). The **None** category primarily consists of *spec:* intermediates and *subst:* , which are often drawn for visual aid without textual labels.

Data Split. We split **CheST** into training and test sets at a 9:1 ratio. To support the cold-start SFT phase, we sampled 10% of the training data to synthesize Chain-of-Thought (COT) reasoning paths, preventing the model from overfitting to the output format while losing reasoning capabilities. Specifically, we retain the COT reasoning paths with correct answers each time. For those with incorrect answers, we adjust the parameters and regenerate the data. If incorrect answers are generated consecutively five times, the data will be discarded.

4 Methodology

We design RULER as a comprehensive framework, integrating a cold-start SFT phase, a rule-guided multimodal RL stage to handle multimodal chemical structure-text coreference. RULER translates rigid domain constraints—specifically valid syntax, precise reference naming, and correct structure typing—into explicit, computable reward signals. Particularly, we introduce a targeted hallucination penalty to curb the tendency of MLLMs to overguess in complex visual contexts. RULER compels the model to optimize exactly valuable in chemical structure-text coreference: accuracy and strict adherence to scientific conventions. The overall training framework is illustrated in Figure 4.

4.1 Prompt Design

To enhance chemical structure-text coreference, we design a prompt that operationalizes multimodal reasoning through a precise taxonomy, visual alignment rules, and a standardized output schema. The instructions define core categories—Markush structures, substituents, and specific compounds—while accommodating dual roles to preserve granularity. Classification leverages visual cues, such as boxed regions and table layouts, alongside compositional naming rules that extract identifiers directly from image text to minimize hallucination. We require the exhaustive enumeration of all targets, utilizing

[2] https://eureka.zhihuiya.com/ls/#/ai-ls/document-agent

[3] Annotators are experts with Ph.D. degrees in chemistry. Pay $5 per data entry.

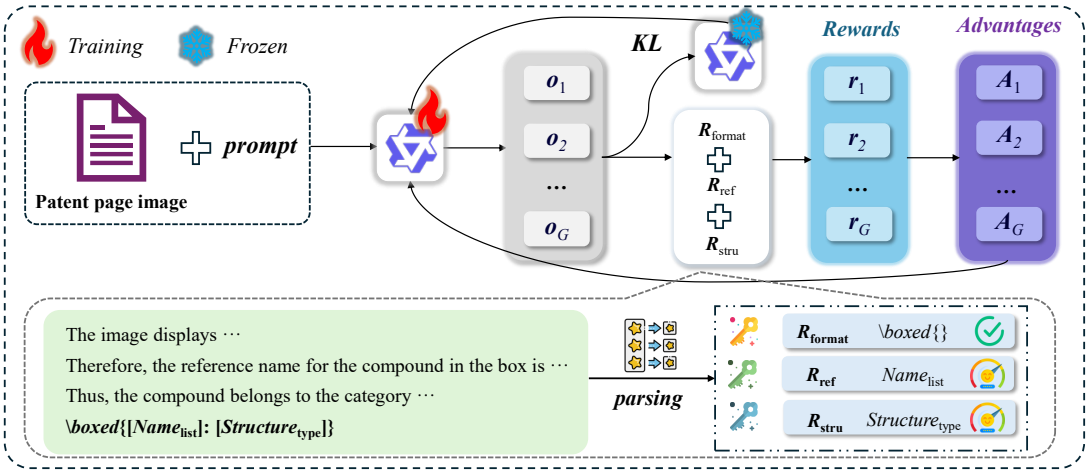


Figure 4: Training Framework of RULER (GRPO)

a "None" fallback for unnamed entities to reduce false positives. This rigorous approach ensures that open-ended visual grounding is converted into auditable, machine-parseable outputs aligned with scientific conventions. For detailed prompt specifications, please refer to Appendix A.

4.2 Cold-Start SFT

Given the model's inherently weak multimodal reasoning capability in the chemical domain, we first conduct Supervised Fine-Tuning (SFT) as a cold-start phase. This aims to guide the model to learn multimodal reasoning patterns centered on chemical structures and provide greater room for subsequent exploration of RL truncation.

During this phase, we utilize the exact same prompt designed for the subsequent RL phase. This consistency ensures that the reasoning patterns and output formats learned during the cold start are directly transferable, providing a stable policy initialization for the RL exploration.

4.3 Domain-Specific Rule-Guided Reward

To bridge the gap between open-ended generation and rigorous chemical constraints, we design a multi-dimensional reward function. Instead of a sparse binary reward, we decompose the objective into three intuitive components to guide the model's behavior explicitly:

Format Compliance (R_{format}). First and foremost, the output must be machine-readable. We incentivize the model to strictly follow the requested syntax: \boxed{[$Name_{list}$]: [$Structure_{type}$]}.

Correctness Verification (R_{ref} & R_{stru}). For the core tasks of **RefMatch** and **StruCls**, the reward directly reflects the accuracy of the extracted names and structure types against the ground truth.

Hallucination Penalty. Crucially, we introduce a penalty mechanism to discourage "guessing." A model might attempt to maximize recall by outputting excessive candidates. To counter this, we impose penalty scores (P_{ref} and P_{stru}) for incorrect predictions, forcing the model to be precise rather than verbose.

Formally, the rewards are calculated as follows:

$$R_{format} = \begin{cases} 1, & \text{if Regex(pred)} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$\text{Acc}(\text{pred}, \text{gold}) = \frac{|\text{pred} \cap \text{gold}|}{|\text{gold}|} \tag{2}$$

$$R_{ref} = \text{Acc}(\mathcal{N}_{pred}, \mathcal{N}_{gold}) + P_{ref} \cdot \text{errors}(\mathcal{N}_{pred}) \tag{3}$$

$$R_{stru} = \text{Acc}(\mathcal{T}_{pred}, \mathcal{T}_{gold}) + P_{stru} \cdot \text{errors}(\mathcal{T}_{pred}) \tag{4}$$

where $\text{Regex}(\cdot)$ parses the content enclosed in \boxed{...}, and $\text{errors}(\cdot)$ counts the number of incorrect items in the predicted content.

4.4 Multi-Task RL Training

We apply the designed rule-guide "verifier" on the algorithms based on group-level advantages (GRPO (Shao et al., 2024)). Unlike standard PPO (Schulman et al., 2017) which relies on a separate value network, GRPO estimates the baseline by computing the mean score of a group of outputs generated from the same input. This group-relative ranking stabilizes training and reduces the

variance of reward scales. To verify the transferability and effectiveness of the designed "verifier", we also conduct validation experiments based on two group-level advantage algorithms optimized by GRPO–**GSPO** (sequence-level optimization based on GRPO)(Zheng et al., 2025a) and **DAPO** (sampling and clip optimization based on GRPO)(Yu et al., 2025). We find that the "verifier" in the two algorithms can indeed improve the task performance compared to conventional GRPO.

The final total reward $\boldsymbol{R}$ is a weighted sum of the three components, ensuring the optimization covers format, reference matching, and structure classification simultaneously:

$$\boldsymbol{R} = \lambda_{\text{format}} \cdot \boldsymbol{R}_{\text{format}} + \lambda_{\text{ref}} \cdot \boldsymbol{R}_{\text{ref}} + \lambda_{\text{stru}} \cdot \boldsymbol{R}_{\text{stru}} \quad (5)$$

where weights λ_i are utilized to control the importance of rewards and to ensure the maximum total reward does not exceed 1.

5 Experiments

Tasks We evaluate the following tasks: (i) **RefMatch**: reference names matching, (ii) **StruCls**: structure types classification, (iii) **All**: considering both **RefMatch** and **StruCls** for each sample.

Metrics For the above three tasks, we evaluated two set-based metrics, Pass@1 and Pass@all, respectively.

$$\text{Pass@1} = \mathbb{I}(|Pred_i \cap Gold_i| > 0) \quad i \in \{ref, stru, all\} \quad (6)$$

$$\text{Pass@all} = \mathbb{I}(Pred_i = Gold_i) \quad i \in \{ref, stru, all\} \quad (7)$$

where $\mathbb{I}(\cdot)$ is a standard indicator function. As shown in the Equations, Pass@1 refers to the scenario where there is one or more intersecting elements between the predicted set and the correct set, while Pass@all means that the elements of the predicted set exactly match those of the correct set.

5.1 Baselines

To verify the necessity of further research on the chemical structure-text coreference task, we compared the current general multimodal large language models (GPT-5, Gemini-2.5-Pro, Claude-4.5-sonnet, GLM-4.5V, Qwen-VL-max) as well as Qwen3-vl-8B, the base model of RULER.

Discussion on Domain-Specific Baselines We deliberately exclude existing domain-specific models from our baselines due to fundamental architectural mismatches with the multimodal coreference task. Current Optical Chemical Structure Recognition (OCSR) tools, such as Decimer (Rajan et al., 2023) and MolScribe (Qian et al., 2023), are strictly designed for pixel-to-graph (e.g., SMILES) generation; they lack the capability to process bounding boxes, comprehend complex table layouts, or extract textual labels (e.g., "Compound 204") from images. Conversely, existing chemical NLP models (e.g., ChemNER) typically accept only textual or 1D SMILES inputs, inherently lacking the visual layout comprehension required to navigate the complex graphical representations in documents.

5.2 Training Details.

We adopt the multimodal reinforcement learning framework EasyR1 (Zheng et al., 2025b) for RL training and LLaMAFactory (Zheng et al., 2024) for SFT training. We apply RULER to three RL algorithms that supports rule-guided rewards (GRPO, GSPO, DAPO) and verify the effectiveness of RULER. During RL training, the number of samples (n) is set to 4, $\{\lambda_{\text{format}}, \lambda_{\text{ref}}, \lambda_{\text{stru}}\}$ are set to $\{0.2, 0.4, 0.4\}$, $\{\boldsymbol{P}_{\text{ref}}, \boldsymbol{P}_{\text{stru}}\}$ are all set to -0.1, the context length is set to 8192, and the number of epochs is set to 20.

Computational Cost and Efficiency All experiments are conducted on a computing cluster equipped with 8 × NVIDIA A800 (80GB) GPUs. The SFT phase requires approximately 8 GPU hours to converge. The RL training takes approximately 240 GPU hours.

5.3 Results

We perform all baselines and RULER on the test set of CheST. The experimental results are presented in Table 1. Among the general MLLMs, **Gemini-2.5-pro** emerges as the strongest competitor, achieving the best performance in the All category. It achieves a Pass@1 (All) score of 73.23% and a Pass@all (All) score of 63.13%. Notably, **GPT-5** demonstrates exceptional capability in **StruCls**, achieving a Pass@1 score of 87.88%.

However, the results reveal significant limitations in current general MLLMs regarding chemical reference matching tasks:

Inconsistency across sub-tasks. There is a noticeable performance gap between StruCls and Ref-

Models	Pass@1			Pass@all		
	RefMatch	**StruCls**	**All**	**RefMatch**	**StruCls**	**All**
GPT-5	63.13	87.88	59.60	53.54	86.36	50.51
Claude-4.5-sonnet	57.07	78.79	48.99	45.96	73.74	36.87
Gemini-2.5-pro	75.25	83.84	73.23	66.16	82.32	63.13
GLM-4.5V	56.06	81.31	49.49	53.54	76.77	44.95
Qwen-vl-max	47.98	76.77	43.43	44.44	66.16	33.84
Qwen3-vl-8B	44.95	34.34	17.17	40.40	29.29	10.61
Qwen3-vl-8B + Cold-Start	47.98	86.87	45.96	43.94	83.84	41.41
+ RULER (GRPO)	90.40	**98.48**	88.89	81.31	95.96	78.28
+ RULER (GSPO)	92.42	**98.48**	90.91	83.84	**97.98**	81.82
+ RULER (DAPO)	**93.43**	**98.48**	**91.92**	**90.40**	**97.98**	**88.38**

Table 1: Main Results of ChemRefMatch. We compare general MLLMs with our Qwen3-vl-8B fine-tuned variants. The best results are highlighted in **bold**.

Match. For instance, while GPT-5 excels at structure classification, its performance drops significantly to 63.13% on the RefMatch metric (Pass@1). This suggests that while MLLMs possess strong general capabilities for recognizing structure types, they struggle with the precise comprehension and matching required for chemical references.

Strict constraint satisfaction. The decline in scores from the Pass@1 to the Pass@all setting across all baseline models highlights the difficulty of satisfying multiple constraints simultaneously. For example, **Claude-4.5-sonnet** sees its overall performance drop to 36.87% in the Pass@all (All) metric, indicating a lack of robustness in handling complex, multi-step chemical reasoning tasks without specific optimization.

Our proposed method, **RULER**, outperforms all general MLLMs across all metrics:

Overall Improvement: In the most challenging "All" category, our best model achieves a **Pass@1 score of 91.92%** and a **Pass@all score of 88.38%**. This represents a substantial improvement over the strong general MLLM, with gains of approximately 18.69% and 25.25% respectively.

Robustness in Sub-tasks: RULER demonstrates balanced superiority in both sub-tasks. The best model achieves a remarkable **97.98%** in Pass@all **StruCls**, surpassing GPT-5, while simultaneously leading the **RefMatch** metric with **90.40%**.

These results validate the effectiveness of our RL approach. By specifically optimizing for the alignment between chemical structures and their references, RULER not only improves individual task accuracy but also significantly enhances the model's ability to satisfy strict, holistic evaluation criteria compared to general MLLMs.

6 Analysis

To better understand the chemical structure-text coreference task and the RULER method, we conduct a detailed analysis. We compare the performance of GRPO trained with rewards calculated based on token-level F1 scores. Furthermore, we analyzed the high-frequency bad cases of all evaluated models in the chemical structure-text coreference task.

6.1 Ablations

To verify the effectiveness of RULER, we design the following ablation methods: (1). SFT: supervised fine-tuning; (2). - *Cold-Start*: only RL training; (3). - *Penalty*: remove penalty in the reward; The ablation results are presented in Table 2.

Models	Pass@all		
	RefMatch	**StruCls**	**All**
SFT	78.79	88.38	69.70
RULER	81.31	**95.96**	**78.28**
- *Cold-Start*	73.74	88.38	69.70
- *Penalty*	**84.34**	87.37	72.73

Table 2: Ablation Results. (RULER is based on GRPO.)

As shown in the table, the full RULER method achieves the best performance across all metrics, significantly outperforming the SFT baseline with an overall Pass@all score of 78.28% compared to 69.70%. This confirms the superiority of our proposed rule-guided reinforcement learning method. When analyzing the specific components, the removal of the cold-start mechanism ("- Cold-Start") leads to a substantial drop in the RefMatch score (73.74%), indicating that a solid supervised initialization is crucial for effective RL training. Furthermore, the ablation of the penalty term ("- Penalty") reveals an important trade-off: while removing the penalty yields a higher RefMatch score (84.34%), it negatively impacts StruCls and the overall performance (72.73%). This suggests that the penalty is essential for ensuring the structural integrity of the generated outputs.

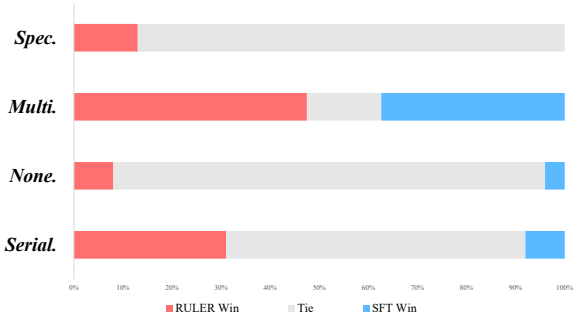


Figure 5: Head-to-head Comparison between RULER and SFT

To further evaluate the robustness of RULER, Figure 5 illustrates the head-to-head comparison between RULER and SFT across different task categories (Spec., Multi., None., and Serial.). RULER consistently matches or surpasses the SFT baseline. Notably, in the Multi. category, RULER demonstrates a significant advantage with a much larger win rate compared to SFT, highlighting its capability in handling complex, multi-constraint scenarios. In the Spec. and Serial. categories, RULER maintains a high tie rate while securing more wins than SFT, proving that it enhances performance.

6.2 Case Study

To more intuitively demonstrate the difficulty of the chemical structure-text coreference task and the effectiveness of the RULER method, we conduct specific case studies. In detail, we analyze the cases that MLLMs currently fail to solve and the cases improved by the RULER method.

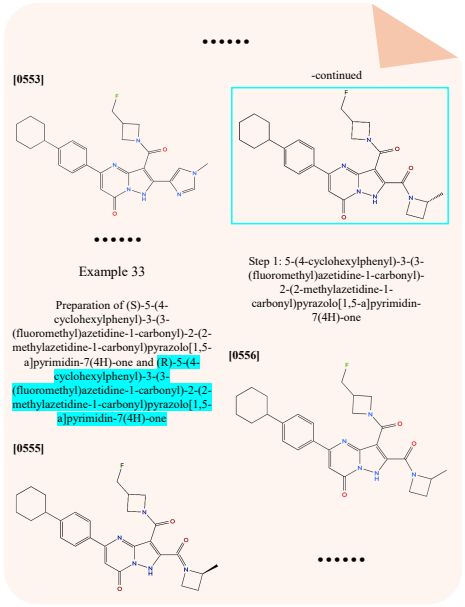


Figure 6: Difficult Case of General MLLMs

As shown in Figure 6, the coreference text name of the chemical structure in the blue box in the figure is "*(R)-5-(4-cyclohexylphenyl)-3-(3-(fluoromethyl)azetidine-1-carbonyl)-2-(2-methylazetidine-1-carbonyl)pyrazolo[1,5-a]pyrimidin-7(4H)-one*" (marked with a blue background), but it is difficult for general MLLMs to recognize this correctly. Specifically, the chemical structures in the images are extremely similar, with the only difference lying in a single atom or the stereochemical configuration of a chiral center. However, general MLLMs are quite confused about this and cannot correctly distinguish such important professional details. For example, Gemini-2.5-pro, the general MLLMs with the highest performance in this task, confused several chemical structures in the figure, treated their names as those of the chemical structure in the box, and incorrectly returned a list containing multiple names.

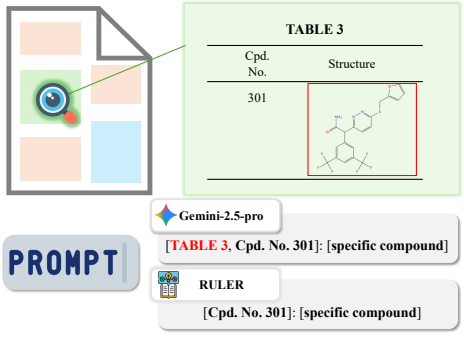


Figure 7: Improved Case of RULER

Although MLLMs still face some challenges in the chemical structure-text coreference task, RULER improves these situations to a certain extent. When identifying coreferential names, MLLMs often tend to rely on relative positional relationships, treating all referential nouns around a chemical structure as the coreferential names of that structure. For example, in the case of "TABLE 3" shown in the Figure 7, the strongest MLLM, Gemini-2.5-pro also classifies it as a coreferential name. In contrast, RULER, through reinforcement training, focuses on referential nouns with chemical semantics, thereby avoiding such situations.

7 Conclusion

In this work, we formally define the multimodal chemical structure-text coreference in biopharmaceutical intellectual property and introduce CheST, the first expert-verified dataset designed for it. Moreover, we proposed RULER, a rule-guided multimodal reinforcement learning method based on cold-start SFT, optimizing for the coreference correctness. Experimental results demonstrate that RULER significantly outperforms all the general MLLMs, effectively bridging the domain gap in chemical structure analysis and enabling more precise knowledge extraction for drug discovery.

Limitations

Our proposed method relies on upstream layout analysis; consequently, it is susceptible to cascading errors where initial failures in identifying chemical regions or recognizing text lead to inaccuracies in the final coreference resolution. Future research will focus on developing end-to-end architectures to mitigate these dependencies and further improve robustness in complex document environments.

References

Srikar Appalaraju, Bhavan Jasani, Bhargava Urala Kota, Yusheng Xie, and R Manmatha. 2021. Docformer: End-to-end transformer for document understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 993–1003.

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. 2023. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv preprint arXiv:2308.12966*.

Albert Bou, Morgan Thomas, Sebastian Dittert, Carles Navarro, Maciej Majewski, Ye Wang, Shivam Patel, Gary Tresadern, Mazen Ahmad, Vincent Moens, and 1 others. 2024. Acegen: Reinforcement learning of generative chemical agents for drug discovery. *Journal of Chemical Information and Modeling*, 64(15):5900–5911.

Zhendong Cao and Lei Wang. 2025. Crystalformer-rl: Reinforcement fine-tuning for materials design. *arXiv preprint arXiv:2504.02367*.

Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

Gautham Dharuman, Kyle Hippe, Alexander Brace, Sam Foreman, Väinö Hatanpää, Varuni K Sastry, Huihuo Zheng, Logan Ward, Servesh Muralidharan, Archit Vasan, and 1 others. 2024. Mprot-dpo: Breaking the exaflops barrier for multimodal protein design workflows with direct preference optimization. In *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–13. IEEE.

Biaoyan Fang, Christian Druckenbrodt, Saber A Akhondi, Jiayuan He, Timothy Baldwin, and Karin Verspoor. 2021. Chemu-ref: A corpus for modeling anaphora resolution in the chemical domain. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1362–1375.

Igor V Filippov and Marc C Nicklaus. 2009. Optical structure recognition software to recover chemical information: Osra, an open source solution.

Stephen R Heller, Alan McNaught, Stephen Stein, Dmitrii Tchekhovskoi, and Igor Pletnev. 2015. InChI, the iupac international chemical identifier. *Journal of Cheminformatics*, 7(1):23.

Yongseung Jegal, Jaewoong Choi, Jiho Lee, Ki-Su Park, Seyoung Lee, and Janghyeok Yoon. 2023. Learning a patent-informed biomedical knowledge graph reveals technological potential of drug repositioning candidates. *arXiv preprint arXiv:2309.03227*.

Minesh Mathew, Dimosthenis Karatzas, and CV Jawahar. 2021. Docvqa: A dataset for vqa on document images. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2200–2209.

Mor Shpigel Nacson, Aviad Aberdam, Roy Ganz, Elad Ben Avraham, Alona Golts, Yair Kittenplon, Shai Mazor, and Ron Litman. 2025. Docvlm: Make your vlm an efficient reader. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29005–29015.

Son Nguyen, Giang Nguyen, Hung Dao, Thao Do, and Daeyoung Kim. 2025. Vdinstruct: Zero-shot key information extraction via content-aware vision tokenization. *arXiv preprint arXiv:2507.09531*.

Yujie Qian, Jiang Guo, Zhengkai Tu, Zhening Li, Connor W Coley, and Regina Barzilay. 2023. Molscribe: robust molecular structure recognition with image-to-graph generation. *Journal of chemical information and modeling*, 63(7):1925–1934.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

Kiran Rajan, Achim Zielesny, and Christoph Steinbeck. 2020. Decimer: Towards deep learning for chemical image recognition. *Journal of Cheminformatics*, 12(1):65.

Kohulan Rajan, Henning Otto Brinkhaus, M Isabel Agea, Achim Zielesny, and Christoph Steinbeck. 2023. Decimer. ai: an open platform for automated optical chemical structure identification, segmentation and recognition in scientific publications. *Nature communications*, 14(1):5045.

Kohulan Rajan, Henning Otto Brinkhaus, Maria Sorokina, Achim Zielesny, and Christoph Steinbeck. 2021. Decimer-segmentation: Automated extraction of chemical structure depictions from scientific literature. *Journal of cheminformatics*, 13(1):20.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Joshua Staker, Kyle Marshall, Robert Abel, and Carolyn M McQuaw. 2019. Molecular structure extraction from documents using deep learning. *Journal of chemical information and modeling*, 59(3):1017–1029.

Xin Wang, Yifan Zhang, Xiaojing Zhang, Longhui Yu, Xinna Lin, Jindong Jiang, Bin Ma, and Kaicheng Yu. 2024. Patentagent: Intelligent agent for automated pharmaceutical patent analysis. *arXiv preprint arXiv:2410.21312*.

Ziwen Wang, Jiajun Fan, Ruihan Guo, Thao Nguyen, Heng Ji, and Ge Liu. 2025. Proteinzero: Self-improving protein generation via online reinforcement learning. *arXiv preprint arXiv:2506.07459*.

Chih-Hsuan Wei, Yifan Peng, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Jiao Li, Thomas C Wiegers, and Zhiyong Lu. 2016. Assessing the state of the art in biomedical relation extraction: overview of the biocreative v chemical-disease relation (cdr) task. *Database*, 2016:baw032.

David Weininger. 1988. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36.

Yiheng Xu, Minghao Li, Lei Cui, Shaohan Huang, Furu Wei, and Ming Zhou. 2020. Layoutlm: Pre-training of text and layout for document image understanding. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1192–1200.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, and 1 others. 2025. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*.

Zenan Zhai, Dat Quoc Nguyen, Saber Akhondi, Camilo Thorne, Christian Druckenbrodt, Trevor Cohn, Michelle Gregory, and Karin Verspoor. 2019. Improving chemical named entity recognition in patents with contextualized word embeddings. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 328–338.

Chujie Zheng, Shixuan Liu, Mingze Li, Xiong-Hui Chen, Bowen Yu, Chang Gao, Kai Dang, Yuqiong Liu, Rui Men, An Yang, and 1 others. 2025a. Group sequence policy optimization. *arXiv preprint arXiv:2507.18071*.

Yaowei Zheng, Junting Lu, Shenzhi Wang, Zhangchi Feng, Dongdong Kuang, and Yuwen Xiong. 2025b. Easyr1: An efficient, scalable, multi-modality rl training framework.

Yaowei Zheng, Richong Zhang, Junhao Zhang, YeYanhan YeYanhan, and Zheyan Luo. 2024. Llamafactory: Unified efficient fine-tuning of 100+ language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 400–410.

Appendix

A Prompt Design

To better guide the model in exploring and learning the chemical structure-text coreference task, we have carefully designed prompts with detailed procedural key points, making full use of the model's inherent multimodal reasoning capabilities. Building on this objective, the prompt is deliberately structured to operationalize multimodal reasoning through a precise taxonomy, explicit visual-to-text alignment rules, and a standardized output schema.

To start, it defines core categories—Markush structure as a fixed parent nucleus with bounded variable substituent, substituent as a non-standalone component, specific compound as a fully instantiated molecule—and allows a dual role "*Mark: &subst:* " when a substituent is itself a Markush pattern, preventing category drift and preserving granularity. Building on that, classification relies on visual/layout cues: boxed regions mark targets; "*R*" columns indicate substituent-bearing positions; Markush–name pairing in tables follows a vertical continuation from the structure's row until the next Markush entry, so the model follows scientific conventions rather than brittle heuristics. Next, a compositional naming rule recovers reference names only from the image—often by joining table titles with sequence identifiers such as "Formula IV," and "Compound 10,"—to curb hallucinations, ensure traceability, and standardize IDs; when sequence numbers and explicit chemical names co-occur, both are kept to reflect many-to-one Markush instantiations.

In parallel, it requires exhaustive enumeration of all boxed compounds and all valid reference names per structure, including instantiation sets from a shared Markush core, enabling coverage of combinatorial spaces and downstream one-to-many scoring. When no image-derived name exists, the fallback is "*None*", clarifying uncertainty and reducing false positives. Lastly, outputs are machine-parseable—a boxed list pairing each complete reference name with its inferred structure type—minimizing post-processing, enabling cross-page linking.

Collectively, these instructions turn an open-ended visual–chemical grounding task into auditable micro-decisions aligned with chemical notation, fostering disciplined chemical-centered image reasoning and clear, parsable results for structure–text coreference. Complete version can be seen in Figure 8 and Figure 9.

B Cross-Page Grounding via Rule Matching

Although our model training operates on individual pages, the output representation is explicitly designed to support scalability to multi-page contexts. In scenarios requiring document-level analysis, such as resolving chemical entities that span across pages, our method accommodates the aggregation of content through a document-wide grounding mechanism. The supporting post-processing procedures are as follows:

1. **Anchor:** Parse the predicted names $\mathcal{N}_i$ and structure types $\mathcal{T}_i$ from the $\backslash boxed\{\}$ payload.

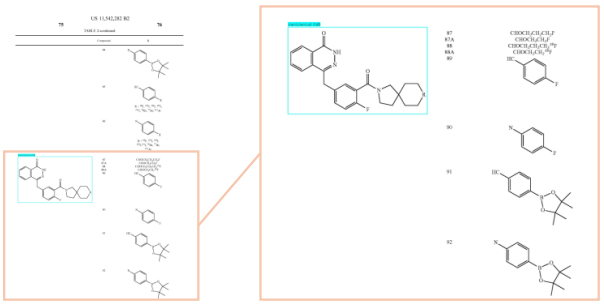
2. **Normalize:** Standardize referential names (e.g., Cpd. $\leftrightarrow$ Compound) to enable robust string matching.

3. **Scan&Merge:** We implement a proximity-based scanning rule. For a detected *Mark:* structure, the system scans a specified page range for corresponding *subst:* components with matching names, merging them into a complete compound entry.

4. **Verify:** To mitigate cascading errors from **RefMatch**, we perform an existence check against the document's OCR results. Predicted names that do not appear textually in the specified page range are filtered out, significantly reducing hallucination.

Through the above processing, the complete patent's chemical structure-textual coreference result can be obtained based on the output of our method.

Image



Prompt

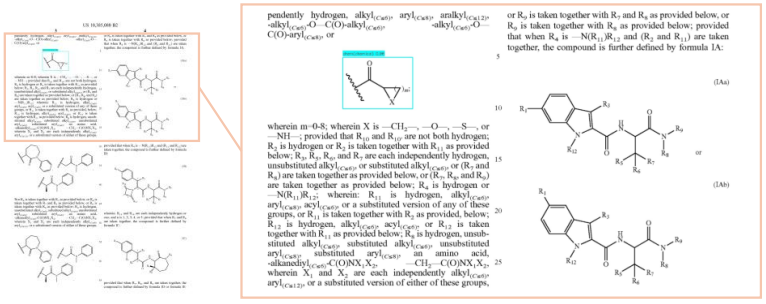
Please identify the reference names of the chemical molecule shown in the box in the image. Please understand the image content and determine which category the compound in the box belongs to: [Markush structure], [substituent], [specific compound], [Markush structure, substituent].
- The key characteristic of a Markush structure is that it contains a “fixed parent nucleus + a variable substituent with a defined range” (e.g., R_1).
- The core characteristic of a substituent is that it is a “component” of the complete molecule (e.g., -OH, -CH_3, benzene ring) and cannot form a complete molecule independently.
- Please classify the compounds based on the image content. For example, if a compound is listed in column “R” of the table, it is a substituent.
- Note: A compound may be a Markush structure itself when it acts as a substituent; that is, it appears as a substituent globally, but is a Markush structure when viewed alone. In this case, its structure type should be [Markush structure, substituent].
- When analyzing tables listing compounds and their structures, note that Markush structures correspond to a set of compound names, starting with the name at the same vertical height as the structure and proceeding sequentially downwards until the next Markush structure appears in the table.
- Append the results after the complete reference name. Note that the reference name must be obtained from the image, which often requires concatenation (e.g., concatenating the table title with the sequence number, such as “Formula IV”). When the compound to be identified is in the table, its sequence number must be concatenated with the title of the sequence number column to form the complete reference name.
- If multiple boxed compounds in an image have reference names, provide all reference names in a list. This includes, but is not limited to: multiple compounds consisting of the same Markush structure and different substituents, where each compound's name is a reference name for that Markush structure; and compounds whose sequence number and specific name are both indicated in the image, such as Example 10 and (S)-2-[1-(4-bromophenyl)-2,2,2-trifluoroethylamino]-4-methylpentane-1-ol.
- If a reference name cannot be obtained from the image, return “None”.
Output Format:
Wrap a list of all reference names with `\boxed{}`, for example, `\boxed{[Compound No. 55]: [substituents, Markush structure]}`, `\boxed{[Compound 1, Compound 2]: [Markush structure]}`, `\boxed{[3,6-Dichloropyridazine]: [specific compound]}`, `\boxed{[None]: [specific Compound]}`.

Response

Based on the provided image and the instructions:
Identify the molecule in the box: The image displays a chemical structure inside a cyan box. This structure contains a fixed scaffold (a phthalazinone-based core with a spiro-cyclic amine) and a variable substituent labeled "R".
Determine the category: Since the structure contains a variable group "R" defined by a table, it represents a generic formula for a class of compounds. Therefore, it is a [Markush structure].
Identify the reference names:
The structure is part of "TABLE 2-continued".
To the right of the structure, there is a column titled "Compound" and a column titled "R".
The instructions state that for Markush structures in a table, the corresponding names start at the same vertical height as the structure and proceed downwards.
The top of the boxed structure aligns horizontally with Compound 87.
The structure spans vertically alongside the rows for Compound 87, 87A, 88, 88A, 89, 90, 91, and 92.
Compounds 84, 85, and 86 are listed above the structure and are not vertically aligned with it.
Therefore, the reference names are the compound identifiers from 87 to 92.
Following the naming convention "Title + Number", the names are "Compound 87", "Compound 87A", etc.
Result:
\boxed{[Compound 87, Compound 87A, Compound 88, Compound 88A, Compound 89, Compound 90, Compound 91, Compound 92]: [Markush structure]}

Figure 8: Example 1 of complete input and output.

Image



Prompt

Please identify the reference names of the chemical molecule shown in the box in the image. Please understand the image content and determine which category the compound in the box belongs to: [Markush structure], [substituent], [specific compound], [Markush structure, substituent].

- The key characteristic of a Markush structure is that it contains a “fixed parent nucleus + a variable substituent with a defined range” (e.g., R_1).
- The core characteristic of a substituent is that it is a “component” of the complete molecule (e.g., -OH, -CH_3, benzene ring) and cannot form a complete molecule independently.
- Please classify the compounds based on the image content. For example, if a compound is listed in column “R” of the table, it is a substituent.
- Note: A compound may be a Markush structure itself when it acts as a substituent; that is, it appears as a substituent globally, but is a Markush structure when viewed alone. In this case, its structure type should be [Markush structure, substituent].
- When analyzing tables listing compounds and their structures, note that Markush structures correspond to a set of compound names, starting with the name at the same vertical height as the structure and proceeding sequentially downwards until the next Markush structure appears in the table.
- Append the results after the complete reference name. Note that the reference name must be obtained from the image, which often requires concatenation (e.g., concatenating the table title with the sequence number, such as “Formula IV”). When the compound to be identified is in the table, its sequence number must be concatenated with the title of the sequence number column to form the complete reference name.
- If multiple boxed compounds in an image have reference names, provide all reference names in a list. This includes, but is not limited to: multiple compounds consisting of the same Markush structure and different substituents, where each compound's name is a reference name for that Markush structure; and compounds whose sequence number and specific name are both indicated in the image, such as Example 10 and (S)-2-[1-(4-bromophenyl)-2,2,2-trifluoroethylamino]-4-methylpentane-1-ol.
- If a reference name cannot be obtained from the image, return “None”.

Output Format:

Wrap a list of all reference names with `\boxed{}`, for example, `\boxed{[Compound No. 55]: [substituents, Markush structure]}`, `\boxed{[Compound 1, Compound 2]: [Markush structure]}`, `\boxed{[3,6-Dichloropyridazine]: [specific compound]}`, `\boxed{[None]: [specific Compound]}`.

Response

Based on the image content, the chemical structure in the cyan box is a generic chemical fragment used to define a substituent in a patent claim or description.

Reference Name Identification:

The structure appears within a text paragraph listing possible definitions for a variable (implied by the word "pendently" and the list of groups like "alkyl", "aryl", etc.).

There is no specific label, figure number, or compound number (e.g., "Formula A", "Compound 1") directly attached to or associated with the boxed structure in the image.

Although the text below mentions "R10 and R10'", this is a proviso and not a direct label for the structure itself. The text defining the variable name is cut off at the beginning of the column (starts with "pendently...").

Therefore, no specific reference name can be obtained directly from the image.

Category Determination:

Substituent: The structure has a wiggly line crossing a bond, which indicates the point of attachment to a larger parent molecule. It is listed as an option in a list of substituent groups.

Markush structure: The structure contains variables m (an integer from 0-8) and X (defined as -CH2-, -O-, -S-, or -NH-), representing a class of related substructures rather than a single specific chemical entity.

Combining these, the structure is a substituent that is defined as a Markush structure.

Result: \boxed{[None]: [Markush structure, substituent]}

Figure 9: Example 2 of complete input and output.