

# Biomedical Multimodal Reasoning for Molecular Understanding and Editing

Siqi Fan<sup>1</sup> Zaiqing Nie<sup>1, 2</sup>

<sup>1</sup>Institute for AI Industry Research (AIR), Tsinghua University <sup>2</sup>PharMolix Inc.

## Abstract

Molecule-centric scientific discovery relies on accurate molecular understanding and effective targeted editing. Although recent advances in large reasoning models have demonstrated significant performance advantages in natural language and vision-language tasks, existing approaches have yet to explore biomedical multimodal reasoning. In this paper, we introduce BioMedGPT-R1, a biomedical multimodal reasoning model, and investigate its capabilities in molecular understanding and editing. The model is trained using a three-stage ‘see-understand-edit’ learning strategy. This approach guides the model to first ‘see’ the molecule, then ‘understand’ it, and finally ‘edit’ it. The strategy comprises three key components: cross-modal alignment, multimodal supervised fine-tuning for efficient molecular understanding, and multimodal reinforcement learning for explainable molecular editing. To facilitate model evolution, we design specific instruction templates, a answer template, and a reward system. Experimental results demonstrate its effectiveness and highlight the power of multimodal reasoning in advancing molecular scientific discovery.

## Introduction

Molecules are the fundamental tokens of the language of chemistry, and molecule-centric scientific discovery relies on two key capabilities: (1) accurate understanding of a given molecule and (2) effective editing to achieve a specific objective. Current scientific research is predominantly conducted in an expert-in-the-loop manner. Thanks to the emergence of biomedical large language models (LLMs) (Luu and Buehler 2024; Wu et al. 2023; Fang et al. 2023), human experts can now directly discuss with them to gain domain-specific insights. Biomedical multimodal large language models (Liu et al. 2024; Li et al. 2024; Luo et al. 2024) further bridge the gap between molecular language and natural language. Cross-modal translation enables these models to: (1) more effectively describe molecules using natural language, thereby enhancing experts’ understanding of molecular structures and properties, and (2) better comprehend natural language instructions from experts, facilitating precise molecular editing, which better supports scientific discovery focused on molecules.

Inspired by the recent advancements in reasoning models (Jaech et al. 2024; Shao et al. 2024; Guo et al. 2025)

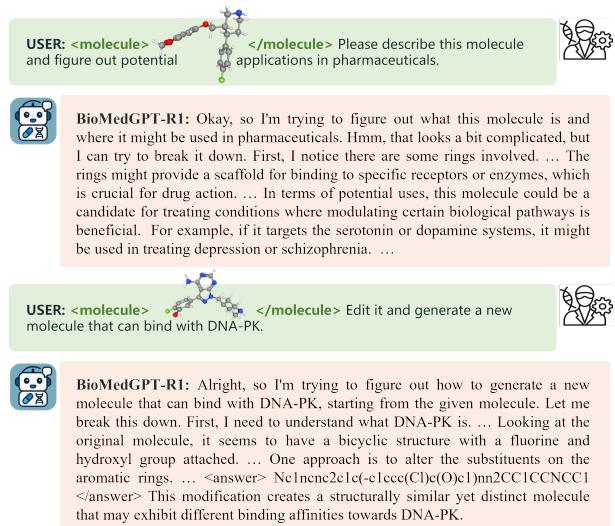


Figure 1: Chat with BioMedGPT-R1 for molecule-centric scientific discovery. Aiming to develop a biomedical research assistant with a deep and comprehensive understanding of both natural language instructions and biomolecules, we explore biomedical multimodal reasoning for molecular understanding and editing in this paper.

that have demonstrated significant performance improvements in natural language processing (NLP) tasks, a question naturally arises: Can reasoning models further facilitate molecule-centric scientific discovery? Specifically, can multimodal reasoning enhance performance in molecular understanding and targeted editing? However, existing multimodal reasoning models are predominantly focused on vision-language tasks (Chen et al. 2025; Yang et al. 2025; Huang et al. 2025; Shen et al. 2025), while biomedical multimodal reasoning remains largely unexplored. Vision, as an intuitive modality for describing the physical world, inherently possesses superior semantic properties, enabling it to align well with natural language and fully leverage the reasoning capabilities of LLMs. In contrast, the semantic information of the molecular modality is more abstract.

Therefore, biomedical multimodal reasoning faces three main challenges: (1) achieving high-level cross-modal align-

ment between molecular representations and the semantic feature space that reasoning LLMs can effectively utilize; (2) enabling efficient molecular understanding through controllable reasoning behaviors—such as implicit and explicit thinking—within an acceptable computational budget; and (3) developing explainable molecular editing that can generate valid, improved molecules with explicit explanations.

To address these challenges, we introduce BioMedGPT-R1, a biomedical multimodal reasoning model, and investigate its capabilities in molecular understanding and editing. As illustrated in Figure 1, our goal is to develop a biomedical research assistant that possesses a deep and comprehensive understanding of both instructions and biomolecules, thereby facilitating real-world applications. BioMedGPT-R1 integrates a reasoning LLM, off-the-shelf representation learning encoders for biomolecules, and modality-specific translators that convert biomedical representations into virtual tokens understandable by the LLM.

To leverage natural language reasoning capabilities and incorporate molecular multimodal knowledge, we propose a three-stage ‘see-understand-edit’ training strategy. First, we fine-tune the molecular translator using a molecule caption dataset with special tokens to achieve high-level cross-modal alignment. Next, both the translator and the reasoning LLM are trained with multimodal question-answering pairs to enhance their multimodal understanding capabilities, utilizing a pre-defined answer template to enable efficient molecular understanding. Finally, we construct a molecular editing dataset specifically designed for reinforcement learning, along with an instruction template and reward system, to encourage the model to explore multimodal reasoning with the objective of editing a given molecule to meet a specific instruction. We evaluate BioMedGPT-R1 on molecular understanding and editing tasks. The experimental results demonstrate its effectiveness and highlight the power of multimodal reasoning in advancing molecular scientific discovery. Our contributions are summarized as follows:

- We propose BioMedGPT-R1, a biomedical multimodal reasoning model designed to assist biomedical research, and investigate its capabilities in molecular understanding and editing.
- We present a three-stage training strategy to facilitate cross-modal alignment, multimodal understanding and reasoning. The model is trained to first ‘see’ the molecule, then to ‘understand’ it, and finally to ‘edit’ it.
- We validate the effectiveness of the proposed approach through comprehensive experiments and demonstrate the potential of biomedical multimodal reasoning for advancing molecular scientific discovery.

## Related Work

### Biomedical cross-modal alignment

Inspired by the success of vision-language alignment (Liu et al. 2023a; Bai et al. 2025; Chen et al. 2024; Fan et al. 2025), several studies have explored cross-modal alignment strategies to jointly model biomolecules and natural language. Three mainstream technical paradigms have

emerged: (1) Contrastive Learning. These approaches align biomolecular representations with natural language in feature space through contrastive learning techniques (Su et al. 2022; Liu et al. 2023b). (2) ‘Language’ Translation. Another line of research treats the string format of biomolecules as a unique language and translates it using a unified language model (Edwards et al. 2022a; Pei et al. 2023). (3) Representation Translation. Recent works have explored translating biomolecular representations into tokens that can be processed by LLMs. For example, GIT-Mol (Liu et al. 2024) incorporates multilevel molecular information, including 1D SMILES, 2D graphs, and images. 3D-MoLM (Li et al. 2024) connects 3D molecular conformations with natural language by integrating a 3D molecular encoder. BioMedGPT (Luo et al. 2024) employs a cross-attention projector to transform biomolecular representations into a fixed number of tokens. We leverage both SMILES strings and encoded molecular graph features as representations of the input molecule, and the cross-modal alignment is achieved in a hybrid manner that combines explicit molecular language (SMILES) translation and implicit feature translation. We introduce special tokens and corresponding instruction template to help the LLM identify and understand the molecular representations.

### Reasoning models

To endow models with System 2-type decision-making capabilities (Kahneman 2011), early attempts explored reasoning models through prompt engineering, such as chain-of-thought (CoT) (Wei et al. 2022) and Tree-of-Thought (Tot) (Yao et al. 2023). Subsequently, some works enhanced step-by-step reasoning through procedure-level supervised fine-tuning (Lightman et al. 2023; Gao et al. 2024). Recently, reinforcement learning has been adopted to encourage models to explore better outcomes via Group Relative Policy Optimization (GRPO) (Shao et al. 2024).

Leveraging the power of reasoning LLMs, several studies have explored Vision-Language reasoning models. For example, R1-V (Chen et al. 2025) introduced GRPO into VLM training for object-counting tasks. R1-OneVision (Yang et al. 2025) constructed a visual reasoning dataset, on which the VLM was first trained before the reinforcement learning stage. Similarly, Vision-R1 (Huang et al. 2025) created a multimodal CoT dataset for initial fine-tuning, followed by a GRPO training stage to further strengthen multimodal reasoning capabilities. Curr-ReFT (Deng et al. 2025) introduced a three-stage reinforcement learning approach with progressively increasing difficulty-level rewards, while VLM-R1 (Shen et al. 2025) provided a more comprehensive investigation of vision-centric multimodal reasoning. However, most of these studies have focused on vision-based mathematics and visual perception tasks. The intriguing yet challenging domain of biomedical multimodal reasoning remains largely unexplored. In this paper, we propose a three-stage training strategy to achieve multimodal reasoning in a progressive manner for molecular editing task. The model is trained to first ‘see’ the molecule, then to ‘understand’ it, and finally to ‘edit’ it. A molecular editing dataset is constructed, and we design specific instruction template and reward system to facilitate model evolution.

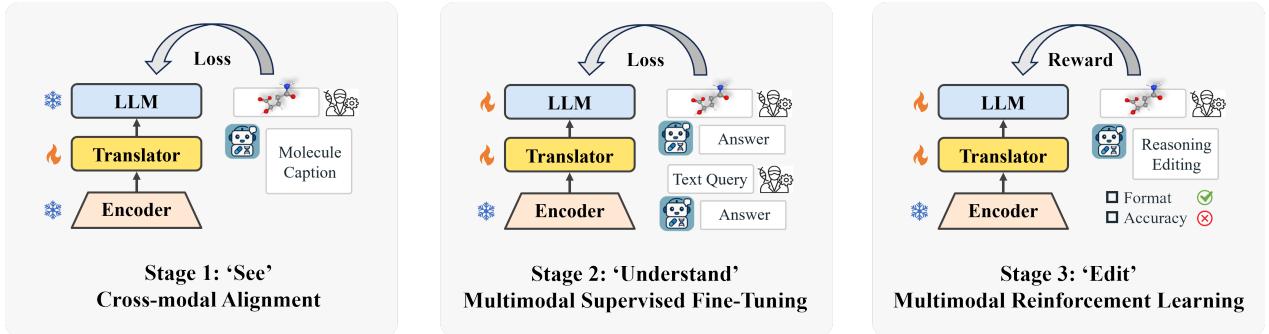


Figure 2: The three-stage ‘see-understand-edit’ training strategy for BioMedGPT-R1, including cross-modal alignment, multimodal supervised finetuning, and multimodal reinforcement learning. The model is trained to first ‘see’ the molecule, then to ‘understand’ it, and finally to ‘edit’ it.

## Methodology

In this section, we first introduce the model architecture of BioMedGPT-R1, followed by a detailed presentation of the three-stage training strategy.

### Model Architecture

BioMedGPT-R1 is a multimodal reasoning LLM that aims at comprehending molecules and natural language texts. The molecule is formulated as a 1D SMILES string  $S$  and a 2D graph  $G = \{\mathcal{V}, \mathcal{E}\}$ , where  $\mathcal{V}$  represents atoms and  $\mathcal{E}$  represents bonds. The natural language text is formulated as a series of tokens  $T$ .

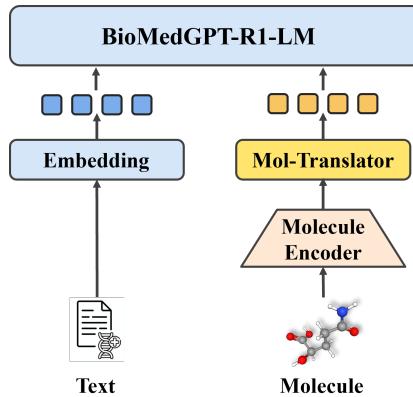


Figure 3: Model architecture of BioMedGPT-R1, comprising a molecule encoder, a translator module, and an LLM.

As illustrated in Figure 3, BioMedGPT-R1 comprises three main components: a molecule encoder, a molecular translator, and a large language model. Specifically, the molecule encoder is a 5-layer GIN (Xu et al. 2018) initialized with the pre-trained checkpoint of GraphMVP (Liu et al. 2021). It encodes the molecular graph  $G$  into a dense representation  $H_m \in \mathbb{R}^{|\mathcal{V}| \times d_m}$ , where  $d_m$  denotes the hidden dimension. The LLM, named BioMedGPT-R1-LM, adopts the same architecture as Qwen2-14B (Team 2024; Guo et al. 2025), a decoder-based Transformer model. To

bridge the modality gap, the molecular translator is implemented as a multi-layer perceptron, which aligns the molecular representation with the feature space of the LLM’s input embeddings. This architecture, though simple, proves effective. We enhance the multimodal reasoning capability in a progressive manner via the three-stage training strategy.

### Three-stage ‘see-understand-edit’ training strategy

Drawing inspiration from the cognitive process of human experts, we propose a three-stage training strategy that guides the model to first ‘see’ the molecule, then ‘understand’ it, and finally ‘edit’ it, as illustrated in Figure 2.

**Cross-modal alignment** To enable BioMedGPT-R1-LM to ‘see’ the molecular representation encoded by the molecule encoder, we first focus on cross-modal alignment. We fine-tune the molecule translator while keeping the other modules frozen.

#### Instruction Template for Cross-modal Alignment

```
<molecule> {molRepHere} </molecule>
<SMILES> {molSmilesHere} </SMILES>
{question}.
```

Specifically, we perform fine-tuning using the molecule caption dataset ChEBI-20 (Edwards et al. 2022b) to align the molecular dense representation  $H_m$  with the high-level semantic feature space of the LLM’s input embeddings. To mitigate the catastrophic forgetting problem that can occur when aligning the LLM with molecular representations, we design a instruction template that organizes the aligned virtual token embeddings and the natural language question in an instruction-following manner. We introduce special tokens  $\langle\text{molecule}\rangle$ ,  $\langle/\text{molecule}\rangle$ ,  $\langle\text{SMILES}\rangle$ ,  $\langle/\text{SMILES}\rangle$  to help the LLM identify the start and end of the molecular representation or the 1D SMILES representation. The prompt template is structured as shown in the box above. Here,  $\{molRepHere\}$  is populated by the aligned virtual tokens generated by the translator,  $\{molSmilesHere\}$  is populated by the SMILES strings,

and  $\{question\}$  is populated by the natural language instruction. This approach ensures that the LLM can effectively integrate molecular information with natural language instructions.

**Multimodal supervised fine-tuning for molecular understanding** To equip BioMedGPT-R1 with a deep and holistic understanding of molecules and enable it to generate insightful descriptions of a given molecule, we perform supervised fine-tuning on the molecule understanding task. Both the molecular translator and the LLM are trained using a molecule question-answering dataset.

#### Answer Template for Multimodal SFT

```
<think> Okay, the user provided both molecular representation and SMILES notation. SMILES is a way to represent chemical structures using text strings. Now, the user is asking me to describe this drug. To do that, I need to figure out its structure, properties, uses, and perhaps its role in the body or in medical applications. </think> Here is a brief description of this drug. {answer}
```

After the initial cross-modal alignment training, the model can already describe the given molecule in an explicit thinking manner. However, considering the high computational cost associated with the reasoning-for-understanding setting, we design a pre-defined answer template to enable an implicit thinking mode for the understanding task, as illustrated in the box above. Here,  $\{answer\}$  is populated by the expected answer. By simulating a brief reasoning process, we guide the model to describe the given molecule more efficiently. The combination of the instruction and the answer template allows the LLM to generate descriptions through implicit reasoning over the molecular representation in an auto-regressive manner. This approach not only maintains the model's original explicit reasoning capability but also enables developers to manage and control reasoning behaviors within an acceptable computational budget.

The training objective is calculated as follows:

$$\mathcal{L} = \sum_{i=1}^N \mathcal{L}_{CE}(f([H_m; \mathcal{S}; \mathcal{I}; \mathcal{A}_p; \mathcal{A}])_{l+i}, \mathcal{A}_i) \quad (1)$$

where  $\mathcal{I}$  and  $\mathcal{A}_p$  are the instructions and pre-defined answer template,  $N$  is the length of the answer, and  $l$  is position where the answer starts.

**Dataset construction for molecule editing** To explore the multimodal reasoning capability in the context of molecular editing tasks, we have constructed a dataset specifically designed for reinforcement learning. Molecular editing for a given objective is inherently a one-to-N task, where multiple solutions may exist for the same goal. This characteristic makes reinforcement learning a more suitable approach compared to supervised fine-tuning. Specifically, if the model's predicted editing result matches any of the potential acceptable answers, it should receive a reward.

We built the dataset based on FS-Mol (Stanley et al. 2021), and the editing tasks encompass five major categories: ADME (absorption, distribution, metabolism, and excretion), binding, functional, physicochemical property, and toxicity. Each data pair in the dataset consists of a source molecule, an instruction, and a list of acceptable editing results. An example is provided as follows.

#### Data Example of Molecular Editing Task

**Source molecule:** COc1cccc(Nc2ncc3nc(-c4ccc(F)cc4)c(=O)n(CCC#N)c3n2)c1

**Instruction:** Edit it and generate a new molecule that has activity in a qHTS assay for Anthrax Lethal Toxin Internalization.

**Acceptable results:** COc1cccc(Nc2ncc3nc(-c4ccc(F)cc4)c(=O)n(C)c3n2)c1, COc1cccc(Nc2ncc3nc(-c4cc(F)cc(F)c4)c(=O)n(CCC#N)c3n2)c1

**Multimodal reinforcement learning for molecular editing** We employ the GRPO algorithm (Shao et al. 2024) to update the model parameters and design a specific instruction template and reward system tailored for the molecular editing task.

**Instruction template.** We design an instruction template incorporating the aforementioned special tokens and providing a clear description of the requirements for molecular editing, as shown in the following box.

#### Instruction Template for Multimodal RL

```
<molecule> {molRepHere} </molecule>
<SMILES> {molSmilesHere} </SMILES>
{question}. The original molecule is given between <molecule> and </molecule>, and the SMILES of it is between <SMILES> and </SMILES>. {editInstruction}. The generated molecule should NOT be the same as the original one. Please first thinks about the reasoning process in the mind and then give the SMILES of the generated molecule as your answer between <answer> and </answer>, i.e., <think> reasoning process here </think><answer> SMILES here </answer>.
```

Here,  $\{editInstruction\}$  is populated by the specific editing instruction. The model is tasked with editing the source molecule to generate a new one, while explicitly detailing the reasoning process. For ease of post-processing, the editing result should be enclosed between  $\langle answer \rangle$  and  $\langle /answer \rangle$ .

**Format reward** checks whether the response follows the specified format that requires the model to output the reasoning process in the  $\langle think \rangle$  tag and the editing result in the  $\langle answer \rangle$  tag, returning 0.5 for each compliance.

**Accuracy reward** evaluates the acceptability of the model's editing results from multiple perspectives. If an  $\langle answer \rangle$  tag is present in the response, the strings enclosed within the tag are parsed. The accuracy reward then

Table 1: **Performance on Molecular Understanding.** All metrics are in %.

Methods	BLEU-2	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L	METEOR
3D-MoLM (Li et al. 2024)	6.7	3.0	12.1	4.2	8.6	18.3
Mol-Instructions (Fang et al. 2023)	24.9	17.1	33.1	20.3	28.9	27.1
GIT-Mol (Liu et al. 2024)	35.2	26.3	57.5	48.5	56.0	43.0
BioMedGPT (Luo et al. 2024)	30.6	19.8	44.7	25.7	38.3	35.0
BioMedGPT-R1	38.3	27.3	51.7	33.1	44.8	43.1

includes three checks: validity, improvement, and exact matching. If no `<answer>` tag is found, the entire response is used to calculate the Levenshtein distance with the first molecular SMILES in the acceptable results list. This approach encourages the model to output results in the required format. **Validity check** assesses whether the editing result is structurally valid using RDKit. A valid result returns a score of 0.5. For valid editing results, we determine the groundtruth molecule based on fingerprint similarity. The most similar molecule in the acceptable results list is considered the groundtruth. If the predicted molecule is more similar to the groundtruth than the source molecule, it indicates effective editing, and the model receives an 0.25 score for passing **improvement check**. If the prediction exactly matches the groundtruth, an additional 0.25 reward is awarded for passing **exact matching check**.

The multi-level accuracy reward system is designed to encourage BioMedGPT-R1 to output valid, improved molecules while adhering to the required format.

## Experiments

In this section, we evaluate BioMedGPT-R1 on molecular understanding and editing tasks and present an in-depth analysis of the experimental results.

### Performance on Molecular Understanding

Molecular understanding aims to describe the molecule in natural language and provide comprehensive information about the given molecule for human experts. For this task, we utilize the widely adopted ChEBI-20 dataset (Edwards et al. 2022b) for both training and evaluation. BioMedGPT-R1 is fine-tuned for 50 epochs with a learning rate of  $7 \times 10^{-5}$  and a batch size of 8. To reduce training costs, we apply low-rank adaptation (LoRA) to the BioMedGPT-R1-LM. Additionally, we perform linear warmup for the first 3 epochs and a cosine annealing strategy to stabilize training.

**Baselines.** We compare the performance of BioMedGPT-R1 with several representative biomedical multimodal methods. These include GIT-Mol (Liu et al. 2024), which incorporates multilevel molecular information; 3D-MoLM (Li et al. 2024), which leverages 3D molecular conformations; and BioMedGPT (Luo et al. 2024), which aligns molecular representations with a fixed number of virtual tokens. Additionally, we report the performance of Mol-Instructions (Fang et al. 2023), a biomedical LLM baseline fine-tuned on a comprehensive instruction-tuning dataset.

**Evaluation metrics.** We adopt the evaluation metrics used in ChEBI-20 (Edwards et al. 2022b), including BLEU-2, BLEU-4, ROUGE-2, ROUGE-L, and METEOR.

**Results and analysis.** We compare BioMedGPT-R1 with the baselines in Table 1. BioMedGPT-R1 demonstrates predictive power on molecular understanding. There are two key observations. (1) Impact of input molecular information: The type of molecular information used significantly affects understanding performance. For example, 3D-MoLM (Li et al. 2024) primarily relies on 3D molecular conformations, which increases the difficulty of alignment from structural feature space to semantic one. In contrast, Mol-Instructions (Fang et al. 2023) leverages the SMILES representation widely adopted in papers and patents, facilitating a more natural translation from molecular SMILES language to the natural language. This approach results in performance advantages over 3D-MoLM. GIT-Mol (Liu et al. 2024) incorporates multiple types of molecular information, including 1D SMILES, 2D graphs, and images, allowing the model to understand molecules from various perspectives. Given that molecular images are a less detailed representation compared to molecular graphs, BioMedGPT-R1 only uses SMILES and encoded graph representations, achieving more accurate molecular descriptions (higher BLEU scores). (2) Reasoning capability enhances molecular understanding. Compared to BioMedGPT (Luo et al. 2024), the introduction of a reasoning large language model in BioMedGPT-R1 leads to significant improvements across all evaluation metrics. In turn, robust biomedical multimodal understanding serves as the foundation for more complex multimodal reasoning tasks, such as molecular editing.

### Performance on Molecular Editing

Molecular editing aims to edit and generate a new molecule based on a given natural language instruction from human experts. To facilitate this task, we constructed an editing dataset (as described in Methodology Section) and split it into training and evaluation subsets. We ensured that molecules in the acceptable list for evaluation did not appear in the training set. The test set consists of 897 molecular editing cases. When training BioMedGPT-R1 with reinforcement learning, we adopt the default GRPO parameter settings, with  $N$  set to 4 and the temperature set to 1.0. The model is trained for 4,000 steps at a learning rate of  $1 \times 10^{-5}$ . To reduce training costs, we apply low-rank adaptation (LoRA) to the BioMedGPT-R1-LM.

**Baselines.** We focus primarily on comparing the performance of biomedical multimodal reasoning with that of text-

only natural language reasoning, a comparison that has not yet been thoroughly explored in existing work. To ensure a fair comparison, we utilize BioMedGPT-R1-LM as the baseline model. In this setup, only molecular SMILES strings are fed into the reasoning language model, and the same instruction template and reward system are employed.

**Evaluation Metrics.** We evaluate the models to assess their ability to generate valid and improved molecules according to the given editing instructions. Specifically, we parse the output molecule enclosed between `<answer>` and `</answer>`, and conduct both validity and improvement checks. The groundtruth molecule is determined based on fingerprint similarity, with the most similar molecule in the provided acceptable list being regarded as the groundtruth. We calculate three key metrics: validity rate (VR), success rate (SR), and exact matching rate (EMR).

$$VR = \frac{N_{valid}}{N_{test}} \quad (2)$$

where  $N_{test} = 897$  and  $N_{valid}$  is the number of valid cases, with validity verified using RDKit.

$$SR = \frac{N_{improved}}{N_{test}} \quad (3)$$

where  $N_{improved}$  represents the number of successful improved cases. Specifically, if the output molecule is more similar to the groundtruth than the source molecule, it is classified as a successful case.

$$EMR = \frac{N_{exactMatching}}{N_{test}} \quad (4)$$

where  $N_{exactMatching}$  denotes the number of cases where the predictions exactly match the groundtruth.

**Table 2: Performance on Molecular Editing.** EMR: Exact Matching Rate. SR: Success Rate. VR: Validity Rate. All metrics are in %.

Methods	EMR	SR	VR
BioMedGPT-R1-LM	1.8	10.8	93.1
BioMedGPT-R1 <i>improvement</i>	2.9	12.6	93.1
	+ 61.1	+ 16.6	-

**Results and analysis.** We compare BioMedGPT-R1 with BioMedGPT-R1-LM in Table. 2. The experimental results demonstrate that biomedical multimodal reasoning can significantly enhance molecular editing performance. Three key observations are highlighted. (1) Both models are capable of generating valid molecules, with no significant advantage observed for the multimodal model. This suggests that contemporary language models can already effectively model SMILES strings, likely due to their frequent presence in molecule-related papers and patents, which are leveraged as pretraining data for LLMs. (2) Biomedical multimodal reasoning significantly improves editing quality. Compared to BioMedGPT-R1-LM, BioMedGPT-R1 achieves substantial performance improvements, with a 61.1% increase in

exact matching rate and a 16.6% increase in success rate. (3) Despite these advancements, molecular editing remains a challenging task. Generating improved molecules that meet expectations is still difficult for both models, indicating that the diverse editing objectives place high demands on the LLM’s internal knowledge. We speculate that incorporating additional inputs from external environments using RAG techniques or in-domain tools could bring significant improvements to this task.

### In-depth Analysis

We present further discussions on multimodal supervised fine-tuning and multimodal reinforcement learning. Moreover, qualitative examples for molecular understanding and editing are also provided.

**Ablation study on molecular understanding.** Inspired by the indistinguishable performance difference in validity rate between the multimodal model and LLM, we are curious whether multimodal input benefits molecular understanding in the reasoning model era. To address this, we evaluate and compare their molecular understanding performance, as shown in Table 3. BioMedGPT-R1-LM is fine-tuned using the same strategy as BioMedGPT-R1, but only molecular SMILES strings are fed into the reasoning language model. The same instruction and answer templates are employed. As seen in Table 3, multimodal input significantly improves performance on this task. The fine-tuned BioMedGPT-R1-LM achieves similar performance to Mol-Instructions (Fang et al. 2023) (as shown in Table 1), an in-domain language model trained on a large-scale molecular question-answering dataset. Although both BioMedGPT-R1 and BioMedGPT-R1-LM can model SMILES language, additional multimodal input introduces molecular semantic information from other perspectives.

**Performance on out-of-domain text-based medical reasoning.** Recent work in vision multimodal reasoning has observed that multimodal fine-tuning can simultaneously improve text-based reasoning capabilities (Ji et al. 2024). To explore whether biomedical multimodal fine-tuning exhibits a similar phenomenon, we fine-tune BioMedGPT-R1 using both the ChEBI-20 molecule description dataset (Edwards et al. 2022b) and the USMLE medical QA dataset (Jin et al. 2021). We then evaluate the out-of-domain performance using the expert-level medical reasoning benchmark MedXpertQA (Zuo et al. 2025). The results are compared and reported in Table 4. Text-based fine-tuning on the USMLE dataset results in a performance improvement of 7.3%, while BioMedGPT-R1 gains an additional 0.9% performance advantage through multimodal fine-tuning. Similar to the observation in the vision-language domain, biomedical multimodal fine-tuning may also enhance related text-based reasoning capabilities.

**Learning curve of text-based RL and multimodal RL.** To further investigate the differences between text-based reinforcement learning and multimodal RL, we report the accuracy reward curve for the first 2500 training steps in Figure 4. BioMedGPT-R1-LM achieves higher accuracy re-

Table 3: **Ablation Study on Molecular Understanding.** All metrics are in %.

Methods	BLEU-2	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L	METEOR
BioMedGPT-R1-LM	22.2	11.5	35.1	16.4	29.1	28.2
BioMedGPT-R1 <i>improvement</i>	38.3	27.3	51.7	33.1	44.8	43.1
	+72.5	+137.4	+47.3	+101.8	+54.0	+52.8

Table 4: **Performance on text-based medical reasoning.**

All metrics are in %.

Methods	Acc.	Abs. Improv.
DS-R1-Qwen-14B	11.9	-
BioMedGPT-R1-LM	19.2	+ 7.3
BioMedGPT-R1	20.1	+ 0.9

wards in the early stages (around 500 steps), as text-based models are generally more adept at adhering to format requirements. Once both models generate responses that meet the format requirement (i.e., the predicted SMILES string is enclosed within the `<answer>` tag), BioMedGPT-R1 usually gains higher rewards.

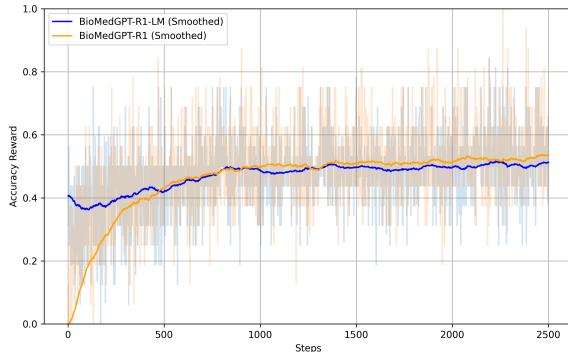


Figure 4: Learning curve of BioMedGPT-R1 (multimodal RL) and BioMedGPT-R1-LM (text-based RL).

**Qualitative example on molecular understanding and editing.** BioMedGPT-R1 is trained using the three-stage ‘see-understand-edit’ strategy. For intuitive analysis, we provide two examples: (1) an example of molecular understanding after the cross-modal alignment learning stage (Figure 5) and (2) an example of molecular editing after the multi-modal reinforcement learning stage (Figure 6). The model is capable of generating valid, improved molecules with explicit explanations, thereby achieving explainable molecular editing.

## Conclusion

In this work, we introduce BioMedGPT-R1, a biomedical multimodal reasoning model, and investigate its capabilities in molecular understanding and editing. We propose a three-stage training strategy that includes cross-modal alignment, multimodal supervised fine-tuning, and multimodal reinforcement learning. To facilitate model evolution, we design an instruction template system, a reward system, and a

molecular editing dataset. The model is trained to first ‘see’ the molecule, then ‘understand’ it, and finally ‘edit’ it. Comprehensive experimental results demonstrate the effectiveness of BioMedGPT-R1. We provide an in-depth analysis of biomedical multimodal reasoning for the foundational tasks of molecular understanding and editing, highlighting the potential of multimodal reasoning to advance molecular scientific discovery. Future work will explore more valuable yet challenging practical applications.

## References

- Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; et al. 2025. Qwen2.5-VL technical report. *arXiv preprint arXiv:2502.13923*.
- Chen, L.; Li, L.; Zhao, H.; and Song, Y. 2025. R1-V: Reinforcing super generalization ability in vision-language models with less than \$3.
- Chen, Z.; Wang, W.; Cao, Y.; Liu, Y.; Gao, Z.; Cui, E.; Zhu, J.; Ye, S.; Tian, H.; Liu, Z.; et al. 2024. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*.
- Deng, H.; Zou, D.; Ma, R.; Luo, H.; Cao, Y.; and Kang, Y. 2025. Boosting the generalization and reasoning of vision language models with curriculum reinforcement learning. *arXiv preprint arXiv:2503.07065*.
- Edwards, C.; Lai, T.; Ros, K.; Honke, G.; Cho, K.; and Ji, H. 2022a. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817*.
- Edwards, C.; Lai, T.; Ros, K.; Honke, G.; Cho, K.; and Ji, H. 2022b. Translation between molecules and natural language. *arXiv preprint arXiv:2204.11817*.
- Fan, S.; Xie, Y.; Cai, B.; Xie, A.; Liu, G.; Qiao, M.; Xing, J.; and Nie, Z. 2025. OCSU: Optical Chemical Structure Understanding for Molecule-centric Scientific Discovery. *arXiv preprint arXiv:2501.15415*.
- Fang, Y.; Liang, X.; Zhang, N.; Liu, K.; Huang, R.; Chen, Z.; Fan, X.; and Chen, H. 2023. Mol-instructions: A large-scale biomolecular instruction dataset for large language models. *arXiv preprint arXiv:2306.08018*.
- Gao, Z.; Niu, B.; He, X.; Xu, H.; Liu, H.; Liu, A.; Hu, X.; and Wen, L. 2024. Interpretable contrastive monte carlo tree search reasoning. *arXiv preprint arXiv:2410.01707*.
- Guo, D.; Yang, D.; Zhang, H.; Song, J.; Zhang, R.; Xu, R.; Zhu, Q.; Ma, S.; Wang, P.; Bi, X.; et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

USER: <molecule>



</molecule> Please describe this molecule and figure out potential applications in pharmaceuticals.



**BioMedGPT-R1:** Okay, so I'm trying to figure out what this molecule is and where it might be used in pharmaceuticals. Hmm, that looks a bit complicated, but I can try to break it down.

First, I notice there are some rings involved. The "C1CNC..." part suggests a cyclic structure, maybe a six-membered ring with some substituents. The "[C@H]" and "[C@@H]" indicate stereochemistry, so this molecule has chiral centers. That's interesting because chirality is important in pharmaceuticals since it can affect how a drug interacts with the body. Then there's the "C2=CC=C(C=C2)F" part, which looks like a benzene ring with a fluorine substituent. Fluorine is a common substituent in drugs because it can make molecules more stable or increase their lipophilicity, which is important for how drugs are absorbed in the body. The "COC3=CC4=C(C=C3)OCO4" part seems to have an oxygen-containing ring, maybe an epoxide or something similar. Epoxides are reactive groups, so if this molecule has one, it might be involved in some kind of biochemical reaction once inside the body. Putting it all together, this molecule seems to have a complex structure with multiple rings, fluorine substituents, and stereochemistry. These features could make it a candidate for various pharmaceutical applications. For example, the fluorine could help in making the molecule more resistant to metabolic breakdown, increasing its half-life in the body. The rings might provide a scaffold for binding to specific receptors or enzymes, which is crucial for drug action.

I wonder if this molecule is part of a larger class of drugs. Maybe it's an antidepressant, antipsychotic, or something else. Fluorinated compounds are used in several drugs, like some antidepressants where the fluorine helps in modifying the molecule's properties to enhance efficacy or reduce side effects. The stereochemistry also plays a big role. Different enantiomers (mirror image forms) of a molecule can have different effects in the body. So, if this molecule has specific stereochemistry, it might be designed to target certain receptors more effectively or avoid binding to off-target sites, which is important for minimizing side effects. I'm also thinking about how this molecule might be synthesized. ....

In terms of potential uses, this molecule could be a candidate for treating conditions where modulating certain biological pathways is beneficial. For example, if it targets the serotonin or dopamine systems, it might be used in treating depression or schizophrenia. Alternatively, if it interacts with other receptors, it could be useful in pain management or as an anti-inflammatory agent. I should also consider the pharmacokinetics of this molecule. How is it absorbed, distributed, metabolized, and excreted? The fluorine and oxygen groups might influence these processes. For instance, fluorine can increase the molecule's lipophilicity, aiding in absorption, while the oxygen-containing rings might affect how the molecule is metabolized in the liver. ....

Figure 5: Qualitative example on molecular understanding.

Huang, W.; Jia, B.; Zhai, Z.; Cao, S.; Ye, Z.; Zhao, F.; Xu, Z.; Hu, Y.; and Lin, S. 2025. Vision-R1: Incentivizing reasoning capability in multimodal large language models. *arXiv preprint arXiv:2503.06749*.

Jaech, A.; Kalai, A.; Lerer, A.; Richardson, A.; El-Kishky, A.; Low, A.; Helyar, A.; Madry, A.; Beutel, A.; Carney, A.; et al. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Ji, J.; Zhou, J.; Lou, H.; Chen, B.; Hong, D.; Wang, X.; Chen, W.; Wang, K.; Pan, R.; Li, J.; Wang, M.; Dai, J.; Qiu, T.; Xu, H.; Li, D.; Chen, W.; Song, J.; Zheng, B.; and Yang, Y. 2024. Align Anything: Training All-Modality Models to Follow Instructions with Language Feedback. *arXiv preprint arXiv:2412.15838*.

Jin, D.; Pan, E.; Oufattolle, N.; Weng, W.-H.; Fang, H.; and Szolovits, P. 2021. What disease does this patient have? a large-scale open domain question answering dataset from medical exams. *Applied Sciences*, 11(14): 6421.

Kahneman, D. 2011. Thinking, fast and slow. *Farrar, Straus and Giroux*.

Li, S.; Liu, Z.; Luo, Y.; Wang, X.; He, X.; Kawaguchi, K.; Chua, T.-S.; and Tian, Q. 2024. Towards 3d molecule-text interpretation in language models. *arXiv preprint arXiv:2401.13923*.

Lightman, H.; Kosaraju, V.; Burda, Y.; Edwards, H.; Baker, B.; Lee, T.; Leike, J.; Schulman, J.; Sutskever, I.; and Cobbe, K. 2023. Let's verify step by step. In *The Twelfth International Conference on Learning Representations*.

Liu, H.; Li, C.; Wu, Q.; and Lee, Y. J. 2023a. Visual instruction tuning. *Advances in neural information processing systems*, 36: 34892–34916.

Liu, P.; Ren, Y.; Tao, J.; and Ren, Z. 2024. GIT-Mol: A multi-modal large language model for molecular science with graph, image, and text. *Computers in biology and medicine*, 171: 108073.

Liu, S.; Nie, W.; Wang, C.; Lu, J.; Qiao, Z.; Liu, L.; Tang, J.; Xiao, C.; and Anandkumar, A. 2023b. Multi-modal molecule structure–text model for text-based retrieval and editing. *Nature Machine Intelligence*, 5(12): 1447–1457.

Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; and Tang, J. 2021. Pre-training molecular graph representation with 3d geometry. *arXiv preprint arXiv:2110.07728*.

Luo, Y.; Zhang, J.; Fan, S.; Yang, K.; Hong, M.; Wu, Y.; Qiao, M.; and Nie, Z. 2024. BioMedGPT: An open multi-modal large language model for biomedicine. *IEEE Journal of Biomedical and Health Informatics*, 1–12.

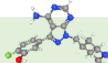
Luu, R. K.; and Buehler, M. J. 2024. BioinspiredLLM: Conversational large language model for the mechanics of biological and bio-inspired materials. *Advanced Science*, 11(10): 2306724.

Pei, Q.; Zhang, W.; Zhu, J.; Wu, K.; Gao, K.; Wu, L.; Xia, Y.; and Yan, R. 2023. BioT5: Enriching cross-modal integration in biology with chemical knowledge and natural language associations. *arXiv preprint arXiv:2310.07276*.

Shao, Z.; Wang, P.; Zhu, Q.; Xu, R.; Song, J.; Bi, X.; Zhang, H.; Zhang, M.; Li, Y.; Wu, Y.; et al. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Shen, H.; Liu, P.; Li, J.; Fang, C.; Ma, Y.; Liao, J.; Shen, Q.; Zhang, Z.; Zhao, K.; Zhang, Q.; et al. 2025. VLM-R1: A stable and generalizable r1-style large vision-language model. *arXiv preprint arXiv:2504.07615*.

USER: <molecule>



</molecule> Edit it and generate a new molecule that can bind with DNA-PK.



**BioMedGPT-R1:** Alright, so I'm trying to figure out how to generate a new molecule that can bind with DNA-PK, starting from the given molecule. First, I need to understand what DNA-PK is. DNA-PK stands for DNA-dependent protein kinase. It's an enzyme involved in DNA repair, particularly in the non-homologous end joining (NHEJ) pathway. .... I should think about the structural features that are known to interact with DNA-PK. From what I remember, DNA-PK has a catalytic subunit (PRKDC) which is a serine/threonine kinase. Inhibitors often target the ATP-binding pocket or other critical regions involved in kinase activity. Looking at the original molecule, it seems to have a bicyclic structure with a fluorine and hydroxyl group attached. There's also a piperazine ring. .... To create a new molecule, I should consider modifying parts of the structure that could affect its binding to DNA-PK. One approach is to alter the substituents on the aromatic rings. Fluorine is an electron-withdrawing group, and replacing it with something else might change the electronic environment, potentially affecting binding. Alternatively, I could add or remove groups that might influence the molecule's ability to fit into the binding site. For example, adding a methyl group might increase steric hindrance, or adding a hydroxyl group could introduce a new hydrogen bonding site. Another thought is to modify the piperazine ring. Changing the substituents on the nitrogen atoms or altering the ring's size could affect the molecule's conformation and its interaction with DNA-PK. I should also consider the molecule's solubility and pharmacokinetics, but since the focus is on binding, I'll prioritize structural changes that might enhance or alter binding affinity.

Let me think about specific modifications. If I replace the fluorine with a chlorine, that would change the substituent from F to Cl. Chlorine is also an electron-withdrawing group but has a different size and polarizability. Alternatively, I could replace the fluorine with a hydrogen, making that position less electronegative. Another option is to add a methoxy group (-OCH<sub>3</sub>) where the fluorine is. Methoxy is an electron-donating group, which could alter the electronic nature of the aromatic ring, potentially affecting binding. I should also consider the possibility of adding a substituent that could act as a hydrogen bond donor or acceptor. For instance, adding an -NH<sub>2</sub> group might provide a new site for hydrogen bonding with DNA-PK. However, I need to ensure that the modification doesn't disrupt the overall structure necessary for binding. It's a balance between altering the molecule enough to change its binding properties and keeping it structurally similar to maintain some level of activity. Perhaps another approach is to change the substitution pattern on the aromatic ring. If the fluorine is in a certain position, maybe moving it or adding another substituent in a different position could create a better fit in the binding site. I also need to think about the molecule's flexibility. Changing substituents can affect how the molecule folds or moves, which in turn affects how it interacts with the target enzyme. In summary, my plan is to identify a substituent on the original molecule that can be altered to create a new molecule with potential DNA-PK binding activity. I'll choose a modification that changes the electronic environment or introduces a new functional group that could interact with DNA-PK's active site.

<answer>Nc1ncn2c1c(-c1ccc(Cl)c(O)c1)nn2CC1CCNCC1</answer>

#### Step-by-Step Explanation:

- Original Structure Analysis:** The original molecule contains a fluorine substituent on an aromatic ring. Fluorine is highly electronegative and can influence the molecule's electronic properties and binding interactions.
- Modification Strategy:** To alter the molecule's binding potential with DNA-PK, I decided to replace the fluorine atom with chlorine. Chlorine is also electronegative but has a larger atomic size and different polarizability compared to fluorine.
- Impact of Change:** This substitution aims to modify the electronic environment of the aromatic ring, potentially affecting how the molecule interacts with the ATP-binding pocket of DNA-PK. The change could influence hydrogen bonding or other interactions crucial for binding.
- Resulting SMILES:** The new molecule's SMILES string reflects the substitution of F with Cl, yielding 'Nc1ncn2c1c(-c1ccc(Cl)c(O)c1)nn2CC1CCNCC1'.

This modification creates a structurally similar yet distinct molecule that may exhibit different binding affinities towards DNA-PK.

Figure 6: Qualitative example on molecular editing.

Stanley, M.; Bronskill, J. F.; Maziarz, K.; Misztela, H.; Lanini, J.; Segler, M.; Schneider, N.; and Brockschmidt, M. 2021. FS-Mol: A few-shot learning dataset of molecules. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Su, B.; Du, D.; Yang, Z.; Zhou, Y.; Li, J.; Rao, A.; Sun, H.; Lu, Z.; and Wen, J.-R. 2022. A molecular multimodal foundation model associating molecule graphs with natural language. *arXiv preprint arXiv:2209.05481*.

Team, Q. 2024. Qwen2 technical report. *arXiv preprint arXiv:2412.15115*.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Wu, C.; Zhang, X.; Zhang, Y.; Wang, Y.; and Xie, W. 2023. PMC-LLaMA: Further finetuning llama on medical papers. *arXiv preprint arXiv:2304.14454*.

Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2018.

How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.

Yang, Y.; He, X.; Pan, H.; Jiang, X.; Deng, Y.; Yang, X.; Lu, H.; Yin, D.; Rao, F.; Zhu, M.; et al. 2025. R1-OneVision: Advancing generalized multimodal reasoning through cross-modal formalization. *arXiv preprint arXiv:2503.10615*.

Yao, S.; Yu, D.; Zhao, J.; Shafran, I.; Griffiths, T.; Cao, Y.; and Narasimhan, K. 2023. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36: 11809–11822.

Zuo, Y.; Qu, S.; Li, Y.; Chen, Z.; Zhu, X.; Hua, E.; Zhang, K.; Ding, N.; and Zhou, B. 2025. MedXpertQA: Benchmarking Expert-Level Medical Reasoning and Understanding. *arXiv preprint arXiv:2501.18362*.