ALMol: Aligned Language-Molecule Translation LLMs through Offline Preference Contrastive Optimisation

Dimitris Gkoumas
Queen Mary University of London, London, UK
d.gkoumas@qmul.ac.uk

Abstract

The field of chemistry and Artificial Intelligence (AI) intersection is an area of active research that aims to accelerate scientific discovery. The integration of large language models (LLMs) with scientific modalities has shown significant promise in this endeavour. However, challenges persist in effectively addressing training efficacy and the out-of-distribution problem, particularly as existing approaches rely on larger models and datasets. In this context, we focus on machine language-molecule translation and deploy a novel training approach called contrastive preference optimisation, which avoids generating translations that are merely adequate but not perfect. To ensure generalisability and mitigate memorisation effects, we conduct experiments using only 10% of the data. Our results demonstrate that our models achieve up to a 32% improvement compared to counterpart models. Finally, we introduce a fine-grained, domain-agnostic evaluation method to assess hallucination in LLMs and promote responsible use.

1 Introduction

The world is facing unprecedented complexity in the form of global challenges such as climate change, healthcare, and pandemics. Innovative scientific solutions are urgently needed to address these challenges. Chemistry has been at the forefront of developing such solutions, pioneering new drugs (Ferguson and Gray, 2018), creating advanced materials (Kippelen and Brédas, 2009), or enhancing chemical processes (Zhong et al., 2023). However, these frontiers are vast and require the involvement of Artificial Intelligence (AI) technology to navigate them effectively.

Large language models (LLMs) have shown promising potential for accelerating scientific discovery across various domains, including chemistry, biology, and materials science (Zhang et al., 2023; AI4Science and Quantum, 2023). Existing work has applied successful paradigms from natural language processing (NLP) and multi-modal representation learning to the chemistry domain. One common approach involves converting the inherent three-dimensional structures of molecules into SMILES, which provide a mapping to symbolic character-level representations. Subsequently, researchers have explored learning language-molecule representations either in separate yet coordinated spaces (Edwards et al., 2022, 2021; Liu et al., 2023a), in a joint space (Liu et al., 2023b), or through hybrid approaches (Luo et al., 2023; Christofidellis et al., 2023). In light of the recent significant advancements in the field, none of the above approaches effectively tackle the inherent challenges in training such models. Instead, they rely on sparse or noisy synthetic data, often necessitating exponentially more data than is typically used in NLP tasks (Edwards et al., 2024).

However, training on larger models and datasets does not necessarily guarantee higher performance. A successful paradigm that augments the capabilities of LLMs across multiple NLP tasks is Reinforcement Learning with Human Feedback (RLHF) (Ouyang et al., 2022). Although initially challenged by issues of slowness and instability, recent research has addressed many of these challenges by shifting towards closed-form losses that operate directly on offline preference data (Rafailov et al., 2024). RLHF has demonstrated superior performance compared to standard minimising cross-entropy optimisation approaches.

In this context, we address challenges related to effectively training robust language models when integrated with scientific modalities. We deploy a novel way of training LLMs for language-molecule translation that avoids generating translations that are only adequate but not perfect, called contrastive preference optimisation (CTO) (Xu et al., 2024). CTO is based on offline preferences instead of supervised fine-tuning, mimicking reference translations. To ensure that our models can effectively generalise instead of memorising patterns, we conduct experiments using only 10% of the L+M-24 dataset (Edwards et al., 2024). Our contributions have as follows:

- Our models achieve significant performance improvements across various evaluation metrics compared to models trained on extensive in-distribution and out-of-distribution data (§ 4.4).
- We showcase their robustness through experiments comparing pivot and minor cross-modals. Our empirical results demonstrate that our models consistently outperform the leading baseline, Meditron, which is trained on the entire dataset, even in agnostic cross-modal scenarios (§ 4.4).
- We propose a fine-grained evaluation method that is domain-independent, assessing factual consistency in generated captions using a question-answering evaluation metric and measuring overlaps of unigrams in generated molecules against references (§ 3.3). Our analysis shows that our models achieve improved factual consistency and character-level unigram overlaps for caption and molecule generation (§ 4.5).

2 Background

Reinforcement Learning with Human Feedback (RLHF) optimisation (Ouyang et al., 2022) operates with a triple dataset $\mathcal{D} = \{x, y_w, y_l\}$, where y_w and y_l represent preferred and dis-preferred outputs, corresponding to input x, such that $y_w \succ y_l$ for x. The probability of y_w over y_l in pairwise comparisons is typically computed using the Bradley-Terry model (Bradley and Terry, 1952):

$$p^*(y_w \succ y_l|x) = \sigma(r^*(x, y_w) - r^*(x, y_l)) \quad (1)$$

where σ is the logistic function, and r^* denotes the reward function that underlies the preferences.

As obtaining the reward directly from a human would be prohibitively expensive, a reward model r_ϕ is trained to act as a surrogate by minimising the negative log-likelihood of the preference data:

$$\mathcal{L}(r_\phi) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}[\log \sigma(r_\phi(x, y_w) - r_\phi(x, y_l))] \quad (2)$$

Additionally, the Kullback-Leibler (KL) divergence between the outputs generated by π_{ref} and the parameterised π_θ models serves as an additional reward signal, ensuring that the generated responses closely align with the reference model. Consequently, an optimal model π_θ is one that maximises:

$$\mathbb{E}_{(x\in\mathcal{D},y\in\pi_\theta)}[r_\phi(x, y)] - \beta\mathcal{D}_{\text{KL}}(\pi_\theta(y|x)||\pi_{\text{ref}}(y|x)) \quad (3)$$

where β is the temperature parameter typically $\in [0.1, 0.5]$.

RLHF can present challenges due to its inherent slowness and instability, especially in distributed settings (Zheng et al., 2024). Recent work has shifted towards closed-form losses to align LLMs with human preferences. Here, we experiment with contrastive preference optimisation that adopts a closed-form loss for RLHF.

3 Methodology

3.1 Task Formulation

Let (x, y) be a pair of source and target sequences mapped to X and Y spaces, respectively. We cast the problem of language-molecule translation as a cross-modal translation task that operates on offline preference data $\mathcal{D} = \{x^{(i)}, y_w^{(i)}, y_l^{(i)}\}_{i=1}^N$, where x is an input, y_w are preferred (e.g. human gold standard) and y_l dis-preferred outputs (typically synthetic, obtained from an appropriate translation model), and N is the total number of pairs. The goal is to learn an optimal function $f : X \leftrightarrow Y$ through a model π_θ parameterised by θ. We coordinate the two spaces through instructional modelling to regulate the translation process in both directions. Specifically, for LMolT, we use instructions for language-to-molecule and molecule-to-language translation (see Appx. A).

3.2 Contrastive Preference Optimisation

Contrastive preference optimisation (CTO) (Xu et al., 2024) addresses challenges stemming from the inherent limitation in RLHF, as discussed in § 2, and from the necessity of high-quality data. CTO is a general approximation of Eq. 3 using a uniform reference model, which assumes equal likelihood for all possible generated outputs:

$$\mathcal{L}(\pi_\theta; U) = -\mathbb{E}_{(x,y_w,y_l)\sim D} \left[\log \sigma\Big(\beta \log \pi_\theta(y_w|x) - \beta \log \pi_\theta(y_l|x)\Big)\right] \quad (4)$$

where π_θ is parameterised model by θ and β hyperparameter (please refer § 2). Eq. 4 implies that the loss is calculated based on how well the generated translations match this uniform distribution of possible translations, rather than being biased towards any particular translation. To maintain π_θ close to the preferred data distribution, a behaviour cloning (BC) (Hejna et al., 2023) regulariser is introduced:

$$\min_{\theta} \mathcal{L}(\pi_\theta, U) \quad \text{s.t.}$$
$$\mathbb{E}_{(x,y_w)\sim D}\Big[\mathbb{KL}(\pi_w(y_w|x)||\pi_\theta(y_w|x))\Big] < \epsilon, \quad (5)$$

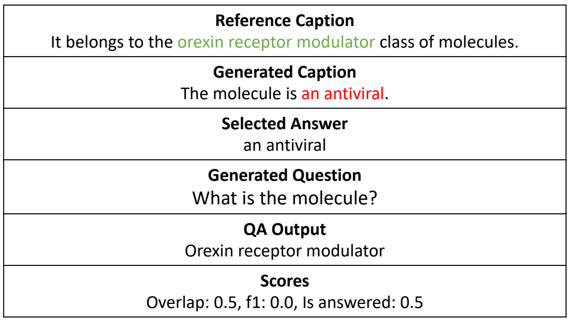
Here, ϵ denotes a small positive constant, and $\mathbb{KL}$ signifies the Kullback-Leibler divergence. The regulariser is enhanced with an additional SFT term on the preferred data, bolstering the CPO loss as:

$$\mathcal{L}_{\text{CPO}} = \min_{\theta} \underbrace{\mathcal{L}(\pi_\theta, U)}_{\mathcal{L}_{\text{prefer}}} - \underbrace{\mathbb{E}_{(x,y_w)\sim D}[\log \pi_\theta(y_w|x)]}_{\mathcal{L}_{\text{NLL}}} \quad (6)$$

3.3 Proposed Evaluation Methodology

Prior studies have utilised embedding representations, for assessing the semantics in chemical-domain models (Jaeger et al., 2018; Edwards et al., 2021; Christofidellis et al., 2023). However, these approaches require domain adaptation for out-of-distribution data (Edwards et al., 2024) and might lead to opaque and arbitrary outcomes (Steck et al., 2024). We address these limitations by introducing a scalable fine-grained evaluation methodology for assessing the presence of hallucinations[1] in generated outputs.

Language Evaluation: For molecule-to-language translation, we deploy the QAFactEval (Fabbri et al., 2022) metric to evaluate the factual consistency of generated captions. QAFactEval first selects noun phrases and named entities (NER) from the generated outputs. A question generation (QG) model then formulates associated questions, which a question answering (QA) model addresses based on the reference text. QAFactEval measures the semantic overlap between the QA model's responses and the selected answers to produce the final metric score. An example is illustrated in Fig. 1. Here, we report the semantic *overlap*, the f_1 *accuracy* between the QA model and the selected answer, and *answerability*, which is the probability of the question being answered by the reference caption.



Reference Caption
It belongs to the orexin receptor modulator class of molecules.

Generated Caption
The molecule is an antiviral.

Selected Answer
an antiviral

Generated Question
What is the molecule?

QA Output
Orexin receptor modulator

Scores
Overlap: 0.5, f1: 0.0, Is answered: 0.5

Figure 1: A toy example illustrating a factual inconsistency between a generated and a reference caption. The QAFactEval metric selects a noun-phrase answer from the generated caption. A QG model then generates an associated question that a QA model answers based on the reference caption. The scores measure the semantic overlap between the QA model's answer and the selected answer from the generated caption

Molecule Evaluation: For language-to-molecule translation, we employ the Chr-F metric, an F-score statistic, to evaluate character n-gram matches between prediction-reference pairs (Popović, 2015). This metric assesses the matches in generated molecules against their references by averaging the scores of unigram, bigram, and trigram matches. A higher Chr-F score indicates better performance.

Bias Evaluation: We also calculate the character and token length bias in generated-reference pairs of molecules and captions, respectively, to investigate potential length bias in the evaluated LLMs.

4 Experiments

4.1 Data

We conduct experiments on the *L+M-24* benchmark dataset, which encompasses both molecule and linguistic modalities (Edwards et al., 2024). It is divided into four categories, each with significant applications in small-molecule domain; biomedical; light and electricity; human interaction and organoleptics; and agriculture and industry. The training and validation subsets consist of approximately 127k and 34k language-molecule pairs, respectively. Here, we utilise 10% of these subsets for training and validation. To operationalise CTO, we recreate a triples dataset consisting of preferred and dis-preferred outputs (see § 2), where the former are the golden references and the latter are generated from MolT5 (Edwards et al., 2022). For evaluation, we randomly selected 3k unseen pairs

[1] Hallucination in LLMs refers to a phenomenon where the generated outputs are inaccurate, nonsensical, or contradictory to the provided factual information.

from a distinct dataset provided by the research group of L+M-24.[2]

4.2 Bechmark Models

We compare our results with established language-molecule models as captured in the literature:

- TxtChem-T5 (Christofidellis et al., 2023): A T5 model trained on both linguistic and molecule modalities with a multi-task objective across various datasets, including the CheBI-20 dataset (Edwards et al., 2022), akin to *L+M-24*.
- Chem-LLM (Zhang et al., 2024): An InternLM2-Base-7B model, trained on an extensive chemical domain knowledge dataset, with the direct preference optimisation objective (Rafailov et al., 2024), achieves results comparable to GPT-4.
- Meditron (Chen et al., 2023): A Meditron-7B model fine-tuned on the entire *L+M-24* for unidirectional language-molecule translation.
- SFT-Meditron: We fine-tune Meditron-7B on a 10% subset of *L+M-24* for bi-directional machine language-molecule translation.

4.3 Experimental Settings

Here, we train Meditron with CTO on a 10% subset of *L+M-24*. We experiment with both language and molecule weight initialisation obtained from Meditron trained on the entire data (Edwards et al., 2024). We refer to them as CTO-Meditron$_{\overrightarrow{Lan.}}$ and CTO-Meditron$_{\overrightarrow{Mol.}}$, respectively. We train the models with QLoRA (Dettmers et al., 2024). For evaluation, we adopt established metrics in (Edwards et al., 2022).

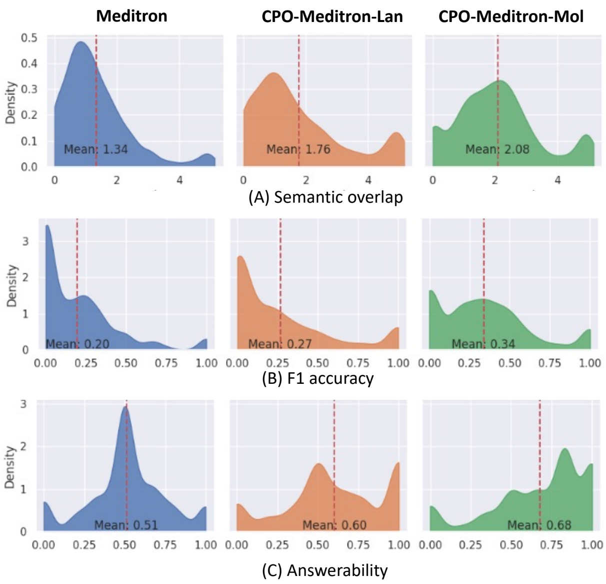
4.4 Experiment Results

Table 2 presents a summary of the molecule-to-language results. We observed a significant decrease in performance for benchmark models trained on extensive data with SFT when tested on out-of-distribution data. Among the baseline models, Meditron demonstrated the highest performance, likely due to its training on the entire *L+M-24* dataset utilised in our experiments. Training Meditron with SFT for bi-directional language-molecule translation has demonstrated neither effectiveness (see Table 1) nor efficiency (refer to Appx. B). This suggests that the performance in our experiments is not dependent on memorised patterns from Meditron trained on the entire dataset. In contrast, our models trained with the CTO objective on only 10% of *L+M-24* achieved a remarkable improvement in performance across diverse evaluation metrics, up to 32% compared to Meditron trained on the entire dataset. This improvement is consistent, as our model consistently enhances performance when initialised from agnostic cross-modals, i.e., CTO-Meditron$_{\overrightarrow{Lan}}$ in Table 1.

We observed similar performance patterns for language-to-molecule translation as reported in Table 2. However, even though our model achieved better performance compared to Meditron when initialised from agnostic cross-modals, it struggled to learn molecular patterns (see CTO-Meditron$_{\overrightarrow{Mol.}}$ in Table 2). This suggests that language plays a pivotal role in the molecule modality. In the future, we aim to explore more advanced initialised methods to address this challenge.

4.5 Evaluation Results

Fig. 2 illustrates the evaluation results on the factual consistency of generated captions against references for the molecule-to-language task. CTO-Meditron$_{\overrightarrow{Mol.}}$, trained on 10% of the available data, exhibited superior factual consistency, achieving a semantic overlap of 2.08, $f1$ accuracy of 0.34, and answerability of 0.68, compared to 1.34, 0.20, and 0.51, respectively, for Meditron trained on the entire dataset. CTO-Meditron$_{\overrightarrow{Lan.}}$ also outperformed Meditron but showed lower performance than CTO-Meditron$_{\overrightarrow{Mol.}}$. We attribute this to the model being initialised by agnostic cross-modals.



(A) Semantic overlap

(B) F1 accuracy

(C) Answerability

Figure 2: Factual consistency in generated captions against references, assessed through (A) semantic overlap, (B) F1 accuracy, and (C) answerability using QAFactEval (§ 3.3) across various LLMs.

[2] Sampling is conducted from a distinct subset.

Model	Blue-2 ↑	Blue-4 ↑	Rouge-1 ↑	Rouge-2 ↑	Rouge-L ↑	METEOR ↑
TxtChem-T5	0.08	0.09	0.19	0.06	0.17	0.16
Chem-LLM	0.03	0.00	0.11	0.02	0.09	0.14
Meditron	0.42	0.30	0.63	0.47	0.49	0.54
SFT-Meditron	0.37	0.26	0.54	0.39	0.38	0.60
CTO-Meditron$_{\overrightarrow{Lan}}$	0.62 (+0.20)	0.45 (+0.15)	0.67 (+0.03)	0.50 (+0.03)	0.48 (-0.01)	0.62 (+0.08)
CTO-Meditron$_{\overrightarrow{Mol}}$	0.74 (+0.32)	0.53 (+0.23)	0.76 (+0.10)	0.56 (+0.09)	0.53 (+0.04)	0.71(+0.17)

Table 1: Molecule-to-language translation results. Arrows next to metrics indicate the higher value the better performance. Numbers in parentheses show deviations from Meditron trained on the entire dataset.

Model	BLEU ↑	Exact ↑	Levenshtein ↓	MACCS FTS ↑	RDK FTS ↑	Morgan FTS ↑	FCD ↓	Validity ↑
TxtChem-T5	0.18	0.00	133.29	0.21	0.10	0.03	37.67	0.58
Chem-LLM	0.04	0.00	732.74	0.00	0.00	0.00	59.44	0.19
Meditron	0.43	0.00	66.16	0.35	0.29	0.19	13.64	0.57
SFT-Meditron	0.30	0.00	186.99	0.70	0.62	0.41	11.14	0.98
CTO-Meditron$_{\overrightarrow{Lan.}}$	0.71 (+0.28)	0.00	42.65 (-23.51)	0.78 (+0.43)	0.70 (+0.41)	0.48 (+0.29)	4.19 (-9.45)	1.00 (+0.43)
CTO-Meditron$_{\overrightarrow{Mol.}}$	0.52 (+0.09)	0.00	76.95 (+10.43)	0.52 (+0.17)	0.49 (+0.20)	0.37 (+0.18)	27.39 (+13.75)	0.58 (+0.01)

Table 2: Language-to-molecule translation results. Arrows next to metrics indicate whether higher or lower values denote better performance. Numbers in parentheses show deviations from Meditron trained on the entire dataset.

For the language-to-molecule task, we observed that both Meditron$_{\overrightarrow{Lan.}}$ and Meditron$_{\overrightarrow{Mol.}}$ achieved similar performance in terms of uni-, bi-, and tri-gram overlaps between generated and reference pairs, outperforming Meditron (see Fig. 3). However, when the model was initialized with known cross-modal weights, i.e., Meditron$_{\overrightarrow{Lan.}}$, it achieved a slightly increased performance

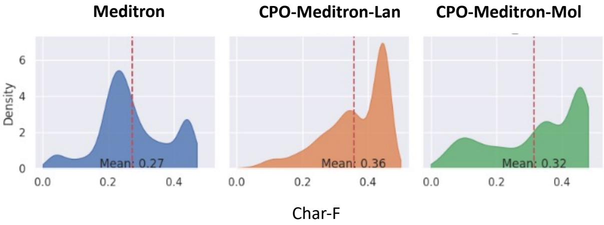
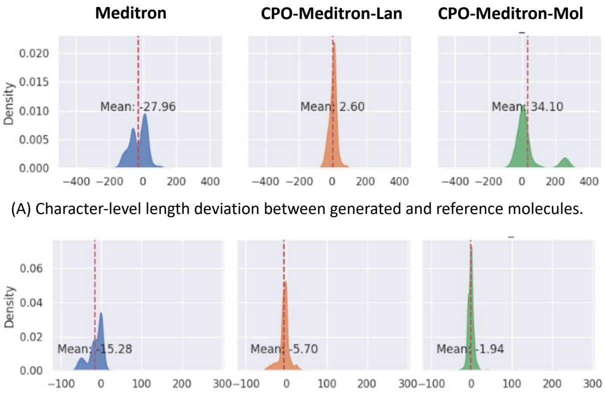


Figure 3: Overlaps of n-gram matches between generated and reference molecules as captured by the char-F (§ 3.3) score across various LLMs.

For the language-to-molecule task, we observed that Meditron and Meditron$_{\overrightarrow{Mol.}}$ generated significantly shorter and longer outputs, respectively (see Fig. 4). In contrast, Meditron$_{\overrightarrow{Lan.}}$ did not exhibit any length bias, producing outputs similar in length to the actual ones. Conversely, for the molecule-to-language task, our models did not show any significant length bias, while Meditron, trained on the entire dataset, generated significantly shorter answers against references.



(A) Character-level length deviation between generated and reference molecules.

(B) Token-level length deviation between generated and reference captions.

Figure 4: Length-bias across different LLMs.

5 Conclusion

This work address training efficacy and the out-of-distribution problem for automatic language-molecule translation. We train models using only 10% of available data and deploying contrastive preference optimisation which avoids generating translations that are merely adequate but not perfect. We achieve significant improvement in performance when compared with models trained on extensive in and out-of-the-distribution data. Finally, we propose a fine-grained, domain-agnostic evaluation method to assess hallucination in LLMs. Our models show superior factual consistency for caption generation and character-level unigram overlaps for molecule generation.

References

Microsoft Research AI4Science and Microsoft Azure Quantum. 2023. The impact of large language models on scientific discovery: a preliminary study using gpt-4. *arXiv preprint arXiv:2311.07361*.

Ralph Allan Bradley and Milton E Terry. 1952. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Zeming Chen, Alejandro Hernández Cano, Angelika Romanou, Antoine Bonnet, Kyle Matoba, Francesco Salvi, Matteo Pagliardini, Simin Fan, Andreas Köpf, Amirkeivan Mohtashami, et al. 2023. Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*.

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

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2024. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 36.

Carl Edwards, Tuan Lai, Kevin Ros, Garrett Honke, Kyunghyun Cho, and Heng Ji. 2022. Translation between molecules and natural language. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 375–413.

Carl Edwards, Qingyun Wang, Lawrence Zhao, and Heng Ji. 2024. L+m-24: Building a dataset for language+ molecules@ acl 2024. *arXiv preprint arXiv:2403.00791*.

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

Alexander Richard Fabbri, Chien-Sheng Wu, Wenhao Liu, and Caiming Xiong. 2022. Qafacteval: Improved qa-based factual consistency evaluation for summarization. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2587–2601.

Fleur M Ferguson and Nathanael S Gray. 2018. Kinase inhibitors: the road ahead. *Nature reviews Drug discovery*, 17(5):353–377.

Joey Hejna, Rafael Rafailov, Harshit Sikchi, Chelsea Finn, Scott Niekum, W Bradley Knox, and Dorsa Sadigh. 2023. Contrastive prefence learning: Learning from human feedback without rl. *arXiv preprint arXiv:2310.13639*.

Sabrina Jaeger, Simone Fulle, and Samo Turk. 2018. Mol2vec: unsupervised machine learning approach with chemical intuition. *Journal of chemical information and modeling*, 58(1):27–35.

Bernard Kippelen and Jean-Luc Brédas. 2009. Organic photovoltaics. *Energy & Environmental Science*, 2(3):251–261.

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

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

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

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Maja Popović. 2015. chrf: character n-gram f-score for automatic mt evaluation. In *Proceedings of the tenth workshop on statistical machine translation*, pages 392–395.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.

Harald Steck, Chaitanya Ekanadham, and Nathan Kallus. 2024. Is cosine-similarity of embeddings really about similarity? *arXiv preprint arXiv:2403.05440*.

Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. 2024. Contrastive preference optimization: Pushing the boundaries of llm performance in machine translation. *arXiv preprint arXiv:2401.08417*.

Di Zhang, Wei Liu, Qian Tan, Jingdan Chen, Hang Yan, Yuliang Yan, Jiatong Li, Weiran Huang, Xiangyu Yue, Dongzhan Zhou, et al. 2024. Chemllm: A chemical large language model. *arXiv preprint arXiv:2402.06852*.

X Zhang, L Wang, J Helwig, Y Luo, C Fu, Y Xie, M Liu, Y Lin, Z Xu, K Yan, et al. 2023. Artificial intelligence for science in quantum, atomistic, and continuum systems. arxiv 2023. *arXiv preprint arXiv:2307.08423*.

Chen Zheng, Ke Sun, Hang Wu, Chenguang Xi, and Xun Zhou. 2024. Balancing enhancement, harmlessness, and general capabilities: Enhancing conversational llms with direct rlhf. *arXiv preprint arXiv:2403.02513*.

Ming Zhong, Siru Ouyang, Yizhu Jiao, Priyanka Kargupta, Leo Luo, Yanzhen Shen, Bobby Zhou, Xianrui Zhong, Xuan Liu, Hongxiang Li, et al. 2023. Reaction miner: An integrated system for chemical reaction extraction from textual data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 389–402.

A Language-molecule Translation Instructions

> Below is an instruction that describes a task, paired with an input that provides further context.
> Write a response that appropriately completes the request.
>
> ### Instruction: You are a researcher. You can come up captions based on your existing knowledge.
> Captions are given against the following input. You should be as detailed as possible.
>
> ### Input: Molecule: {source molecule}
> In that molecule, could you formulate a caption about?
>
> ### Response:{target caption}

Figure 5: Instruction for molecule to language translation, i.e., $M \rightarrow L$

> Below is an instruction that describes a task, paired with an input that provides further context.
> Write a response that appropriately completes the request.
>
> ### Instruction: You are a researcher. You can come up molecule smile strings based on your existing knowledge.
> Molecule smile strings are given against the following input. You should be as detailed as possible.
>
> ### Input: Caption: {source caption}
> In that caption, could you generate a molecule smile string?
>
> ### Response: {target molecule}

Figure 6: Instruction for language to molecule translation, i.e., $L \rightarrow M$

B Training Effectiveness and Efficiency

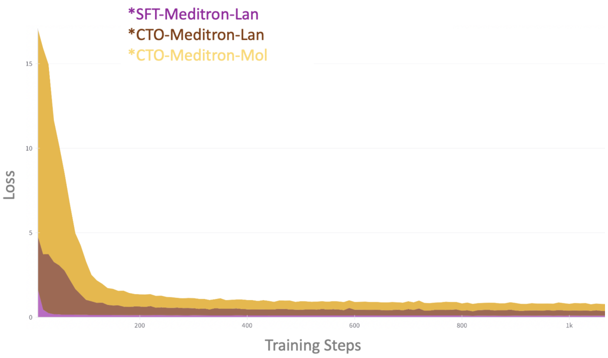


Figure 7: Training convergence

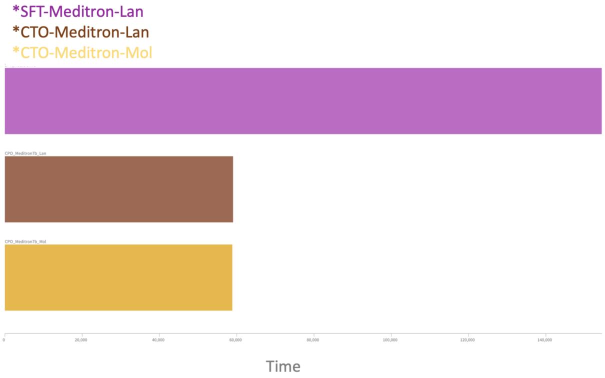


Figure 8: Training efficiency

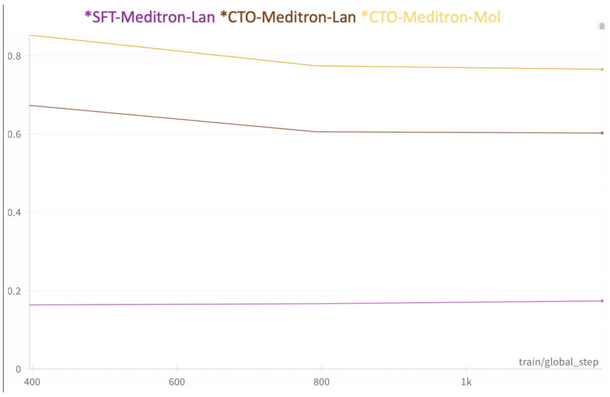


Figure 9: Validation loss