CAPABILITY-PRESERVING MOLECULAR ADAPTATION VIA RECIPROCAL PREFERENCE OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Molecular large language models (LLMs) are commonly adapted from general-purpose LLMs through large-scale domain-specific supervised fine-tuning (SFT). While effective at improving targeted molecular tasks, such specialization may overwrite capabilities already present in the base model, leading to degradation in general-purpose abilities and molecular reasoning capabilities outside the fine-tuning distribution. In this work, we investigate whether preference optimization can provide a more capability-preserving alternative for molecular adaptation. Our key hypothesis is that, by learning relative preferences rather than directly maximizing the likelihood of target responses, while remaining anchored to a reference policy, preference optimization can induce a more conservative adaptation regime than conventional SFT. To make preference optimization feasible without human-annotated molecular preference data, we introduce **ReMPA**, a **Re**ciprocal **M**olecular **P**reference **A**daptation framework. ReMPA repurposes existing paired molecule–response supervision into reciprocal preference pairs by pairing template-aligned, structurally similar molecules, such that each response is preferred under its corresponding molecular condition and rejected under the other. Across molecular property prediction, molecular reasoning and comprehension, and evaluations of general capabilities, our experiments demonstrate that, compared with supervised fine-tuning, ReMPA achieves meaningful gains in targeted molecular capabilities while better preserving capabilities already present in the base model. These results suggest that reciprocal preference optimization offers a promising capability-preserving alternative to conventional supervised domain adaptation for molecular LLMs.

1 INTRODUCTION

General-purpose large language models (LLMs) already exhibit non-trivial chemical knowledge and reasoning abilities, even without explicit molecular-domain adaptation (Guo et al., 2023). However, their performance remains limited and uneven on specialized molecular tasks, motivating the development of molecular LLMs that adapt general-purpose language models with domain-specific molecular data and supervision (Edwards et al., 2022; Christofidellis et al., 2023; Liu et al., 2023a). Large-scale instruction resources such as Mol-Instructions and SMolInstruct have demonstrated that supervised instruction tuning can substantially strengthen LLMs across a broad range of molecule-oriented tasks (Fang et al., 2024; Yu et al., 2024). More recent molecular LLMs further combine instruction tuning with molecular representations or structured chemical knowledge to improve molecular understanding and downstream performance (Liu et al., 2023b; 2024b; Kim et al., 2025; Yang et al., 2025). Consequently, domain-specific supervised fine-tuning (SFT) has become a prevailing paradigm for adapting general-purpose LLMs into molecular specialists.

Despite its effectiveness, such specialization is not necessarily free. Domain-specific SFT updates an LLM toward the response distribution emphasized by its fine-tuning data, and prior studies have shown that this process can degrade capabilities acquired before adaptation, including general reasoning and instruction-following abilities (Liu et al., 2024a; Nam et al., 2026). Although such degradation is not inevitable and can be substantially mitigated by milder updates, a trade-off between domain-specific gains and the preservation of general capabilities can still remain (Lin et al., 2026). Recent evidence in molecular LLMs similarly suggests that fine-tuning on molecular instruction data can impair general reasoning capabilities (Yin et al., 2025). More importantly, this

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

concern extends beyond general-purpose abilities: concentrating optimization on a restricted molecular instruction distribution may also interfere with molecular knowledge and reasoning capabilities already present in the base model but not reinforced during fine-tuning. This motivates us to ask a fundamental question for molecular adaptation: *Can we acquire molecular capabilities without unnecessarily overwriting the capabilities already present in the base model?*

This observation motivates us to reconsider the optimization objective used for molecular adaptation. Unlike SFT, which directly maximizes the likelihood of target responses, preference optimization learns from relative comparisons between preferred and dispreferred responses. Direct Preference Optimization (DPO), for example, optimizes such preferences through likelihood ratios defined relative to a fixed reference policy, and can be derived from a KL-regularized preference-learning objective (Rafailov et al., 2023). This distinction suggests a potentially more conservative form of adaptation: rather than requiring the model to reproduce the full target response distribution, preference optimization only needs to make desired responses more favorable than plausible alternatives, while measuring this change relative to the behavior of the reference model (Zhang et al., 2026). We therefore hypothesize that, when applied directly to a general-purpose base model, preference optimization anchored to the reference model can acquire targeted molecular capabilities with less disruption to capabilities already encoded in the model than conventional SFT.

Putting this hypothesis into practice, however, requires suitable preference supervision, which is not naturally available at scale for molecular adaptation. Existing molecular preference-based approaches operate over substantially different preference spaces: some rank generated molecular candidates according to chemical desirability, while others contrast original molecular structures with perturbed alternatives to improve structural utilization (Park et al., 2023; Cavanagh et al., 2026; Lee et al., 2025). Preference optimization over textual outputs has only recently begun to appear in molecular LLMs. For example, ALMol applies preference optimization to molecule–language translation using model-generated translations as dispreferred outputs (Gkoumas, 2024). Yet relatively little work has explored how existing paired molecule–response supervision can itself be repurposed into informative preference pairs. ReMPA addresses this setting by converting paired molecule–response instances into reciprocal preferences, such that each response is preferred under its corresponding molecular condition and rejected under the other. We use structural similarity to select plausible pseudo-negative responses from template-aligned molecules. Crucially, this reciprocal role reversal makes the preference assigned to each response depend on the molecular condition, encouraging the model to learn molecule–response compatibility rather than a fixed ranking over responses.

We systematically evaluate this hypothesis by comparing the unadapted base model, conventional supervised fine-tuning, one-way DPO, and ReMPA across complementary evaluation dimensions, including molecular property prediction, molecular reasoning and comprehension, and general-purpose capabilities. As summarized in Fig. 1, ReMPA improves performance on the targeted molecular adaptation tasks while largely retaining the molecular reasoning and general-purpose capabilities of the base model. In contrast, conventional SFT induces larger degradation on the capabilities to be preserved, particularly under full-scale molecular specialization. We further compare ReMPA with a one-way DPO variant constructed from the same molecular pairs, to examine the effect of reciprocal preference construction. Together, these comparisons examine how both the adaptation objective and reciprocal preference construction affect the adaptation–preservation trade-off during molecular specialization.

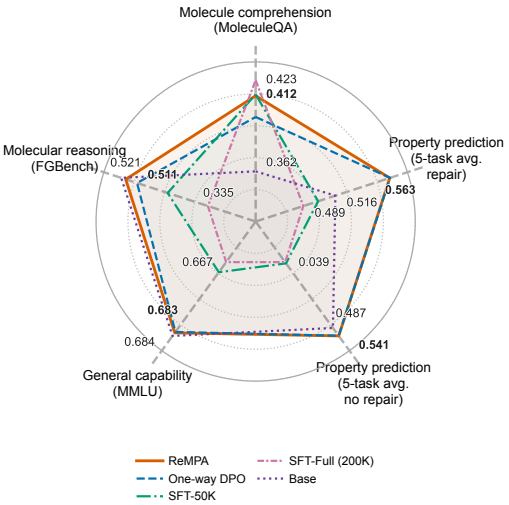

Figure 1: ReMPA achieves a favorable adaptation–preservation trade-off across multiple molecular and general capability dimensions.

Our contributions are summarized as follows:

- We highlight **capability-preserving molecular adaptation** as an important problem, asking how molecular capabilities can be acquired while preserving general-purpose and molecular reasoning capabilities already present in the base model.
- We introduce a **molecular preference-pair construction strategy** that converts paired molecule–response supervision into preferences using template-aligned, structurally similar molecules as pseudo-negatives, without human preference annotation or additional response generation.
- We introduce **ReMPA**, a Reciprocal Molecular Preference Adaptation framework that assigns opposite preference roles to paired responses under their corresponding molecular conditions. ReMPA improves targeted molecular capabilities while better preserving base-model capabilities than conventional SFT, yielding a more favorable adaptation–preservation trade-off.

2 Related Work

Molecular LLM Adaptation. Early molecular language models such as MolT5 and MolXPT used domain-specific pretraining for molecule–text translation and joint molecular–text modeling (Edwards et al., 2022; Liu et al., 2023a). Multimodal foundation models further aligned molecular structures with language through graph–text contrastive learning or additional biomedical knowledge, as in MoMu and MolFM (Su et al., 2022; Luo et al., 2023). More recent work increasingly adapts general-purpose LLMs through molecular instruction tuning, supported by large-scale resources such as Mol-Instructions and SMolInstruct (Fang et al., 2024; Yu et al., 2024). Recent molecular LLMs enrich adaptation with graph-based, multimodal, and chemically informed representations (Liu et al., 2023b; 2024b; Chen et al., 2025; Kim et al., 2025; Livne et al., 2024; Yang et al., 2025), while UniMoT and EDT-Former explore alternative molecule–language alignment through discrete molecular tokenization and dynamic graph tokens (Guo et al., 2025; Jing et al., 2026). Across these variants, supervised likelihood-based training remains central to molecular specialization; we instead study preference-based adaptation when both capability acquisition and preservation matter.

Specialization and Capability Preservation. Domain-specific adaptation can improve specialized performance while degrading capabilities acquired before adaptation, closely related to catastrophic forgetting in LLM fine-tuning (Li et al., 2024; Liu et al., 2024a). Such degradation can be mitigated through milder optimization (Lin et al., 2026), parameter-efficient adaptation such as LoRA (Biderman et al., 2024), or regularized supervised training (Nam et al., 2026). In the molecular domain, MoRA keeps the LLM backbone frozen and performs instance-specific dynamic adaptation to reduce degradation of general reasoning (Yin et al., 2025). Complementary work asks whether the training objective itself can promote preservation: MMPO reports stronger preference alignment while better retaining general language capabilities (Najafi & Fyshe, 2026), and MetaGDPO uses reference-constrained preference optimization to limit parameter drift and alleviate forgetting (Zhang et al., 2026). We instead formulate molecular adaptation explicitly as an adaptation–preservation problem and investigate reference-anchored preference optimization as a means of acquiring targeted molecular capabilities while preserving both broader molecular capabilities and general-purpose competence.

Preference Optimization for Molecular Models. DPO learns directly from pairwise preferences relative to a reference model (Rafailov et al., 2023). For molecular generation, preference optimization has been used to align generated molecules with chemist preferences (Park et al., 2023), construct preference pairs from molecular scores for de novo design (Hou, 2025), and contrast successful and unsuccessful generations to improve adherence to desired properties (Cavanagh et al., 2026). Mol-LLM instead contrasts correct and perturbed molecular graphs to improve structural utilization (Lee et al., 2025), while ALMol applies preference optimization to molecule–language translation outputs (Gkoumas, 2024). Little work has explored whether paired molecule–response supervision itself can provide conditional preference data. ReMPA constructs such preferences reciprocally: each response is preferred under its corresponding molecular condition and rejected under the other, making preference depend on molecule–response compatibility without human preference annotation or additional response generation.

3 REMPA: RECIPROCAL MOLECULAR PREFERENCE ADAPTATION

ReMPA enables molecular adaptation by converting paired molecule–response supervision into reciprocal preferences and optimizing them relative to the base model.

3.1 PROBLEM FORMULATION

Let π_0 denote a general-purpose base policy and let $\mathcal{D}_{\text{mol}} = \{(x_i, m_i, y_i)\}_{i=1}^{N}$ denote a molecular adaptation dataset, where x_i denotes the task instruction, m_i denotes the molecular input, and y_i is the corresponding target response. Conventional molecular adaptation typically obtains an adapted policy π_θ by supervised fine-tuning (SFT), which minimizes the negative log-likelihood of the target responses:

$$\mathcal{L}_{\text{SFT}}(\theta) = -\mathbb{E}_{(x,m,y)\sim\mathcal{D}_{\text{mol}}} \left[\log \pi_\theta(y \mid x, m)\right]. \tag{1}$$

While effective for fitting the target molecular distribution, domain-specific SFT can alter capabilities acquired before adaptation, motivating methods that balance domain-specific gains with the preservation of existing model capabilities (Liu et al., 2024a; Lin et al., 2026).

We therefore formulate *capability-preserving molecular adaptation* as the problem of improving task-relevant molecular capabilities relative to the base policy π_0 while minimizing unnecessary degradation of capabilities already encoded in π_0. Importantly, capability preservation is not introduced as an additional training loss in our framework: ReMPA does not require general-domain replay data or explicit supervision on the capabilities to be preserved. Instead, we investigate whether changing the adaptation objective itself can yield a more favorable adaptation–preservation trade-off.

Rather than directly maximizing the likelihood of a single target response, we consider molecular adaptation through response-level preference supervision. Specifically, we transform the original molecular supervision into a preference dataset

$$\mathcal{D}_{\text{pref}} = \{(x_k, m_k, y_{w,k}, y_{l,k})\}_{k=1}^{M}, \tag{2}$$

where M denotes the number of directional preference examples, $y_{w,k}$ denotes the response preferred under the molecular condition (x_k, m_k), and $y_{l,k}$ denotes a plausible but less compatible alternative. This formulation allows adaptation to be driven by relative response compatibility rather than direct likelihood fitting alone.

Standard molecule response datasets, however, provide only paired supervision (x_i, m_i, y_i) and do not contain an explicit dispreferred response $y_{l,i}$. The central challenge is therefore how to construct informative molecular preference pairs without human preference annotations or additional response generation. We address this challenge by constructing reciprocal preferences directly from paired molecule–response supervision, as described next in Sec. 3.2.

3.2 RECIPROCAL PREFERENCE CONSTRUCTION

Figure 2 illustrates how ReMPA converts existing paired molecule–response supervision into reciprocal preference examples. Given two molecule–response instances sharing the same instruction, we use structural similarity to select plausible alternative responses. The selected pair is then used reciprocally: each response is preferred under its corresponding molecular condition and rejected under the other.

Template-Aligned Molecular Pairing. We construct preference supervision from the molecular instruction data released with Mol-LLaMA (Kim et al., 2025). For each instance (x_i, m_i, y_i), candidate partners (x_j, m_j, y_j) are restricted to examples with the same instruction, $x_j = x_i$, but a different molecule, $m_j \neq m_i$. This template alignment ensures that the two responses address the same task, while differences between them can be associated with the molecular input. Within this candidate set, we use structural similarity as a practical criterion for selecting plausible response alternatives. Specifically, each molecule is represented by a Morgan fingerprint $\phi(m)$ (Rogers & Hahn, 2010), and molecular similarity is measured using the Dice coefficient:

$$s_{ij} = \frac{2|\phi(m_i) \cap \phi(m_j)|}{|\phi(m_i)| + |\phi(m_j)|}. \tag{3}$$

We retain mutual top-K neighbors satisfying $m_j \in \mathcal{N}_K(i)$ and $m_i \in \mathcal{N}_K(j)$, and select disjoint pairs within each instruction template, so that each molecule participates in at most one pair.

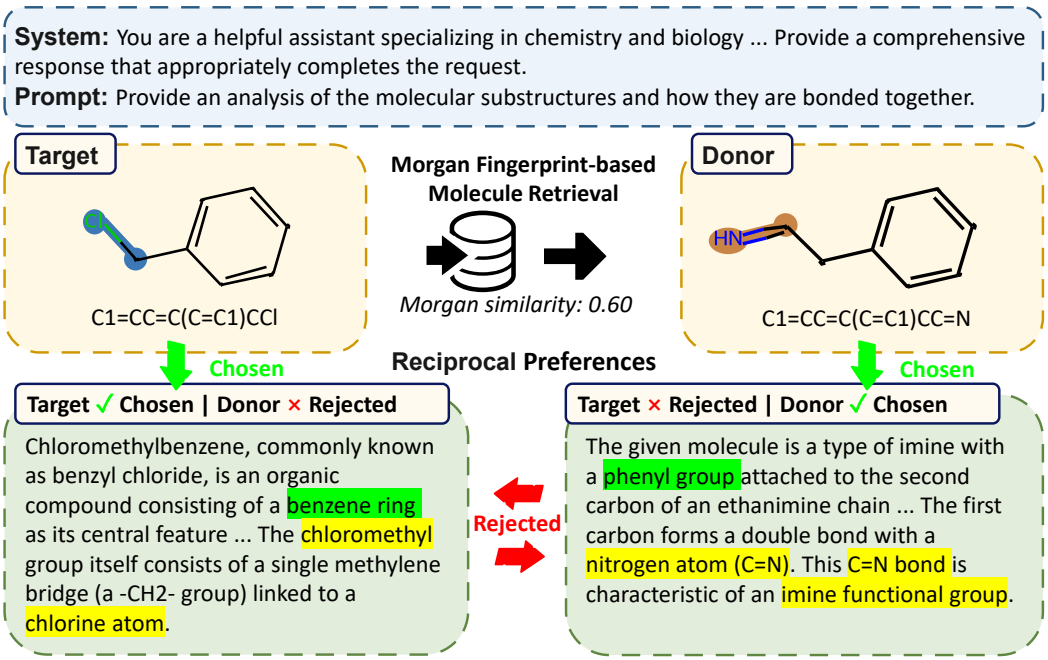

Figure 2: Illustration of reciprocal preference construction in ReMPA. Two template-aligned molecule–response instances are first selected using molecular similarity. Each response is then preferred under its original molecular condition and rejected under the paired molecule, yielding two reciprocal preference examples. The transferred response serves as a plausible pseudo-negative rather than a verified factual negative. Green highlights indicate shared structural or semantic content, while yellow highlights emphasize molecule-specific distinctions.

Reciprocal Preference Construction. For each selected pair (m_i, m_j) sharing instruction x, the response originally paired with each molecule is preferred under its own molecular condition and rejected under the other. This yields two directional preference examples: $(x, m_i, y_w = y_i, y_l = y_j)$, $(x, m_j, y_w = y_j, y_l = y_i)$. Equivalently, the pair defines the two conditional preferences

$$y_i \succ_{(x,m_i)} y_j, \qquad y_j \succ_{(x,m_j)} y_i. \tag{4}$$

Thus, neither response is globally designated as preferred or rejected; its preference role reverses with the molecular condition. This reciprocal construction encourages preference learning to depend on molecule–response compatibility rather than a fixed ranking over responses, while requiring neither human preference annotation nor additional response generation.

3.3 Reciprocal Preference Optimization

Base-Referenced Preference Learning. ReMPA directly adapts the general-purpose base policy π_0 without an intermediate molecular SFT stage. The trainable policy is initialized from π_0, while a frozen copy of the same model serves as the reference policy:

$$\pi_{\theta^{(0)}} = \pi_0, \qquad \pi_{\text{ref}} = \pi_0. \tag{5}$$

Following the reference-relative formulation of DPO (Rafailov et al., 2023), for a directional preference example (x, m, y_w, y_l), we define the reference-relative preference margin

$$\Delta_\theta(x, m; y_w, y_l) = \log \frac{\pi_\theta(y_w \mid x, m)}{\pi_{\text{ref}}(y_w \mid x, m)} - \log \frac{\pi_\theta(y_l \mid x, m)}{\pi_{\text{ref}}(y_l \mid x, m)}. \tag{6}$$

The corresponding directional molecular preference loss is

$$\ell_\theta(x, m; y_w, y_l) = -\log \sigma(\beta \Delta_\theta(x, m; y_w, y_l)), \tag{7}$$

where β scales the reference-relative preference margin.

Reciprocal Optimization. Let $\mathcal{P}$ denote the set of selected molecule–response pairs (x, m_i, y_i, m_j, y_j). As defined in Eq. 4, each pair induces two reciprocal conditional preferences under the corresponding molecular conditions. ReMPA optimizes both directions symmetrically:

$$\mathcal{L}_{\mathrm{ReMPA}}(\theta) = \mathbb{E}_{(x,m_i,y_i,m_j,y_j)\sim\mathcal{P}}\left[\frac{1}{2}\left(\ell_\theta(x, m_i; y_i, y_j) + \ell_\theta(x, m_j; y_j, y_i)\right)\right]. \tag{8}$$

Thus, neither response has a fixed preference role: the preferred response changes with the molecular condition. Optimizing both directions encourages the learned preference to reflect molecule–response compatibility rather than a global ranking between the two responses. Appendix A formalizes this effect, showing that reciprocal optimization limits molecule-independent response reweighting as a standalone adaptation mechanism and requires positive molecule–response interaction gain to improve beyond the reference-policy baseline.

Together, reciprocal preference construction and reference-anchored optimization allow ReMPA to adapt the base model through conditional relative preferences, without directly fitting each target response in isolation or introducing explicit supervision over the capabilities to be preserved.

4 Experiments

Figure 1 provides a high-level overview of the adaptation–preservation trade-off examined in our experiments. We evaluate ReMPA on MoleculeQA (Lu et al., 2024), FGBench (Liu et al., 2025), MMLU (Hendrycks et al., 2021), and molecular property prediction tasks, addressing the following questions: **Q1)** Can ReMPA acquire targeted molecular capabilities directly through preference-based adaptation? **Q2)** Compared with conventional molecular SFT, does ReMPA better preserve molecular reasoning capabilities beyond the adaptation distribution? **Q3)** Does ReMPA better preserve the general-purpose capabilities of the base model after molecular adaptation? **Q4)** When constructed from the same molecular pairs, how does reciprocal preference construction compare with one-way DPO? Detailed benchmark descriptions, evaluation protocols, implementation details, and additional results are provided in the corresponding sections and the Appendix.

4.1 Experimental Setup

Base Model and Adaptation Data. We use Llama-3.1-8B-Instruct as the shared backbone (Grattafiori et al., 2024). To isolate the effects of the adaptation strategy from architectural changes, molecular inputs are represented directly as SMILES strings in the text prompt, and only LoRA parameters are optimized while the pretrained backbone remains frozen. Molecular adaptation data are derived from the template-aligned single-turn supervision in Mol-LLaMA-Instruct (Kim et al., 2025); its multi-turn Comprehensive Conversations are excluded. The same adaptation pool is used for the preference and supervised variants after removing molecules overlapping downstream evaluations using the procedure in Appendix D.2.

Model Variants. We compare the five variants summarized in Tab. 1. SFT-50K and SFT-Full perform supervised adaptation on 50K and approximately 200K examples, respectively, with SFT-Full capped at the SFT-50K optimizer update budget. One-way DPO and ReMPA are constructed from the same 25K molecular pairs; One-way DPO retains one direction per pair, whereas ReMPA uses both reciprocal directions. One-way DPO is trained for twice as many epochs to match directional preference-example exposure. Main comparisons focus on Base, ReMPA, and SFT-Full, while SFT-50K and One-way DPO are used for the controlled analyses in Sec. 4.4. Full optimization and checkpoint-selection details are provided in Appendix C.

4.2 Acquiring Molecular Capabilities

We first evaluate targeted molecular capability acquisition on MoleculeQA (Lu et al., 2024), a multiple-choice benchmark for factual molecular comprehension. We evaluate on its complete 5,786-question test split using the aspect-specific 10-shot setting and report accuracy over four aspects—Structure, Source, Property, and Application—together with overall accuracy. For models evaluated under our protocol, results are averaged across three support-set seeds, except LlaSMol-Llama2-7B, for which outputs are available for two seeds. In addition to Base, ReMPA, and SFT-Full, we include representative instruction-tuned molecular LLMs (Fang et al., 2024; Yu et al.,

Table 1: Model variants. One-way DPO and ReMPA use the same 25K molecular pairs with one versus two preference directions per pair.

Model	Scale	Objective	Directions / Pair
Base	–	–	–
SFT-50K	50K examples	SFT	–
SFT-Full	∼200K examples	SFT	–
One-way DPO	25K directional	DPO	1
ReMPA	50K directional	DPO	2

Table 2: 10-shot accuracy (%) on the MoleculeQA test set (Lu et al., 2024). For models evaluated under our protocol, results are averaged across three support-set seeds, except LlaSMol-Llama2-7B (*), which is averaged over the two available seeds (0 and 42). Results marked with † are reported directly from the original MoleculeQA paper. LlaSMol-Llama2-7B uses one additional format-repair attempt for invalid outputs; all other results under our protocol are computed from the initial predictions.

Model	Structure	Source	Property	Application	Overall
Large-scale Universal Models					
Base (Llama3.1-8B-Instruct)	38.93	38.07	26.54	29.99	36.24
Mixtral-8×7B-Instruct-v0.1†	23.32	31.87	32.89	29.96	27.79
GPT-3.5-1106-turbo†	25.60	37.60	28.04	32.22	29.29
GPT-4-1106-preview†	60.94	50.19	35.57	43.91	53.47
Llama-Based Molecular LLMs					
LlaSMol-Llama2-7B*	24.11	25.32	22.85	25.29	24.35
MolInst-Llama3-8B	29.43	27.45	25.72	28.55	28.41
SFT-Full (200K)	48.73	41.60	25.76	30.27	42.26
ReMPA	46.71	41.42	25.81	31.11	41.23

2024) and general-purpose baselines reported in the original MoleculeQA study (Lu et al., 2024) for broader context.

As shown in Table 2, ReMPA improves overall accuracy over Base by 4.99 percentage points. The gain is most pronounced on structure-related questions, with additional improvements on Source and Application, while Property performance remains close to the Base level. ReMPA also achieves overall performance close to SFT-Full, despite adapting directly from the general-purpose base model through preference optimization. These results show that ReMPA can acquire substantial molecular comprehension capabilities without requiring an intermediate molecular SFT stage.

4.3 Preserving Existing Capabilities

Preserving Molecular Reasoning. We first evaluate whether molecular adaptation preserves functional-group-level reasoning using FGBench (Liu et al., 2025), which evaluates reasoning over single functional groups, functional-group interactions, and molecular comparisons. As shown in Table 3, ReMPA remains close to Base across the Boolean reasoning tasks, whereas SFT-Full exhibits substantially larger degradation across all three categories. ReMPA also retains much of the Base model's valid-output behavior on the value-based tasks. Although SFT-Full achieves lower RMSE on these tasks, its substantially lower valid-output rates mean that RMSE is computed over a considerably smaller subset of predictions. RMSE should therefore be interpreted together with output validity. Overall, ReMPA preserves functional-group-level reasoning substantially more effectively than full-scale supervised molecular adaptation.

Preserving General Capabilities. We further evaluate general capability retention on MMLU (Hendrycks et al., 2021). For broader context, we also include representative chemistry-specialized language models (Dai et al., 2025; Yu et al., 2024; Zhang et al., 2024; Zhao et al., 2025). As shown in Table 4, ReMPA remains nearly unchanged from Base in both overall performance and across the four subject groups, whereas SFT-Full shows a broader decline. Together with the FGBench results,

Table 3: Performance on FGBench (Liu et al., 2025). ACC and Valid are reported as proportions; higher is better for both, while lower RMSE is better. Results marked with † are taken from the original FGBench paper (Liu et al., 2025); all other results are obtained under our evaluation protocol.

	Single				Interaction				Comparison			
	Boolean		Value		Boolean		Value		Boolean		Value	
Model	ACC↑	Valid↑	RMSE↓	Valid↑	ACC↑	Valid↑	RMSE↓	Valid↑	ACC↑	Valid↑	RMSE↓	Valid↑
Large-scale Universal Models												
Base	0.592	0.994	174.443	0.851	0.514	0.994	119.899	0.803	0.456	0.989	83.048	0.768
GPT-4o†	0.667	0.999	77.990	0.813	0.488	0.998	43.577	0.891	0.614	0.992	68.857	0.708
o3-mini†	0.687	0.999	101.886	0.960	0.693	1.000	39.943	0.977	0.703	1.000	64.579	0.975
Llama-3.1-8B†	0.548	0.993	162.351	0.840	0.547	0.982	421.325	0.780	0.474	0.991	80.566	0.761
Llama-3.1-70B†	0.683	1.000	84.119	0.973	0.530	1.000	38.646	0.977	0.456	1.000	64.887	0.943
Qwen2.5-7B†	0.590	0.999	63.511	0.576	0.396	0.999	36.307	0.683	0.664	1.000	65.471	0.223
Other Molecular LLMs												
ChemLLM-7B†	0.233	0.997	209.584	0.629	0.235	0.997	162.742	0.566	0.250	1.000	65.428	0.514
nach0-base†	0.606	0.798	104.534	0.539	0.543	0.756	172.929	0.683	0.041	0.149	12221.946	0.879
LlaSMol-Mistral-7B†	0.387	0.922	266.720	0.923	0.298	0.968	262.550	0.983	0.239	1.000	245.298	0.924
Llama-Based Molecular LLMs												
Llama-3-8B-MolInst†	0.107	0.203	328.935	0.496	0.059	0.149	188.376	0.486	0.469	0.873	138.654	0.837
SFT-Full (200K)	0.408	0.834	54.935	0.440	0.312	0.680	43.423	0.574	0.285	0.689	33.996	0.434
ReMPA	0.578	0.973	156.346	0.781	0.532	0.974	55.792	0.817	0.423	0.890	103.008	0.712

Table 4: Performance on MMLU (Hendrycks et al., 2021). Micro and macro accuracy (%) are reported together with performance over four subject groups. Results marked with † are taken from Dai et al. (2025); all other results are obtained using our reproduction of the official MMLU evaluation protocol.

Model	Micro Acc.	Macro Acc.	STEM	Humanities	Social Sci.	Other
Large-scale Universal Models						
Base	68.36	69.12	59.94	64.29	77.67	73.26
Other Molecular LLMs						
LlaSMol-Llama2-7B	37.17	37.44	28.96	36.60	40.85	42.17
ChemDFM†	58.11	–	–	–	–	–
ChemLLM-7B-Chat†	61.79	–	–	–	–	–
ChemLLM-7B-Chat-1.5-SFT†	63.56	–	–	–	–	–
Llama-Based Molecular LLMs						
Llama3-KALE-LM-Chem-1.5-8B†	68.06	–	–	–	–	–
MolInst-Llama3-8B	62.58	64.13	53.51	57.11	73.77	68.32
SFT-Full (200K)	66.68	67.62	57.52	62.53	76.34	72.05
ReMPA	68.29	69.01	59.74	64.34	77.48	73.26

this indicates that reference-anchored preference adaptation can acquire molecular capabilities while largely preserving both molecular reasoning and general-purpose capabilities already present in the base model.

4.4 Controlled Comparisons and Ablations

Molecular Property Prediction. We evaluate five binary property-prediction tasks: BBBP, DILI, hERG, and HIA from Therapeutics Data Commons (TDC) (Huang et al., 2021), and BACE from MoleculeNet (Wu et al., 2018). We report macro-averaged results here, with per-task results in Appendix D.5. As shown in Table 5, the preference-based variants improve balanced accuracy while retaining Base-like initial parse rates. The SFT variants have very low initial parse rates; one repair attempt largely recovers format validity, but classification remains weaker with markedly lower negative-class recall.

Adaptation Objective and Capability Preservation. Table 6 summarizes the primary metrics across all four evaluations. Compared with both SFT variants, ReMPA retains substantially more of the Base model's performance on molecular reasoning and general-purpose benchmarks while still

Table 5: Macro-averaged performance (%) across five molecular property-prediction tasks. Initial metrics are computed from the original model outputs, while Final metrics are computed after one format-repair attempt. BA denotes balanced accuracy and Parse denotes the valid-output rate.

	Initial			Final			Recall	
Model	BA	Macro-F1	Parse	BA	Macro-F1	Parse	Pos.	Neg.
Base	48.69	46.54	95.21	51.63	48.28	100.00	66.41	36.85
SFT-50K	4.76	8.37	9.56	50.16	44.14	97.32	90.10	10.21
SFT-Full (200K)	3.92	7.16	5.96	48.90	42.39	96.59	89.93	7.87
One-way DPO	53.99	52.99	94.76	56.30	53.82	99.62	77.50	35.11
ReMPA	54.09	50.84	94.87	56.27	52.03	99.77	78.69	33.85

Table 6: Cross-benchmark summary (%). FGBench reports mean Boolean accuracy over its three subsets, and Property reports five-task macro BA before/after repair. Best and second-best adapted models are **bold** and underlined; Base is excluded from the ranking.

	MoleculeQA	FGBench	MMLU	Property	
Model	Acc.↑	Bool Avg.↑	Micro Acc.↑	Initial BA↑	Repair BA↑
Base	36.24	52.07	68.36	48.69	51.63
SFT-50K	<u>41.32</u>	42.13	66.91	4.76	50.16
SFT-Full (200K)	**42.26**	33.50	66.68	3.92	48.90
One-way DPO	39.83	<u>48.67</u>	<u>68.27</u>	<u>53.99</u>	**56.30**
ReMPA	41.23	**51.10**	**68.29**	**54.09**	<u>56.27</u>

achieving strong gains in targeted molecular capabilities. In particular, ReMPA approaches SFT-Full on MoleculeQA while remaining much closer to Base on FGBench and MMLU and achieving stronger property-prediction performance. A similar preservation pattern is observed with One-way DPO relative to SFT-50K. Moreover, the learning-rate-matched SFT control in Appendix C.3 attains comparable MoleculeQA performance but remains lower on FGBench, MMLU, and property prediction, indicating that the observed difference is not explained by the initial learning rate alone.

Ablation of Reciprocal Preference Construction. We compare ReMPA with the One-way DPO variant constructed from exactly the same 25K molecular pairs. The two variants share the same underlying molecular pairs and preference-optimization settings, with their reported checkpoints selected independently according to their respective validation trajectories. This ablation therefore measures the effect of using both reciprocal directions rather than a single direction from each molecular pair. ReMPA improves MoleculeQA overall accuracy by 1.40 percentage points and FG-Bench Boolean accuracy by 2.43 points on average, while preserving essentially the same MMLU performance. Aggregate property-prediction performance is also similar between the two variants. Thus, relative to the one-way baseline, ReMPA achieves additional gains in molecular comprehension and functional-group reasoning while maintaining essentially the same general-capability preservation.

5 Conclusion

We introduce ReMPA, which converts paired molecule–response supervision into reciprocal molecule-conditioned preferences for capability-preserving molecular adaptation. Across evaluations of molecular comprehension, property prediction, molecular reasoning, and general-purpose capability, ReMPA acquires targeted capabilities while preserving substantially more base-model capability than conventional SFT. Comparison with the one-way DPO baseline further shows that ReMPA achieves stronger molecular comprehension and functional-group reasoning while maintaining similar general-capability preservation. These results support reference-anchored preference optimization as an alternative to likelihood-based molecular specialization.

AI USE STATEMENT

Generative AI tools were used to assist with manuscript preparation and polishing, literature retrieval and discovery, and as a coding aid. All AI-assisted text and code were reviewed, verified, and, where applicable, tested by the authors. Generative AI was not involved in research ideation, research methodology, or experimental design. The authors take full responsibility for the final content of this work.

REPRODUCIBILITY STATEMENT

We provide detailed information supporting reproducibility throughout the paper and Appendix. Appendix B documents the construction and statistics of the molecular preference data, including the reciprocal and one-way variants. Appendix C specifies the model implementation, training hyperparameters, and additional optimization controls. Appendix D describes the evaluation benchmarks, data-contamination control, metrics, prompts, output-processing procedures, and additional results. Appendix A states the assumptions and derivations underlying our theoretical analysis. All experiments are based on publicly available base models and datasets.

REFERENCES

Dan Biderman, Jacob Portes, Jose Javier Gonzalez Ortiz, Mansheej Paul, Philip Greengard, Connor Jennings, Daniel King, Sam Havens, Vitaliy Chiley, Jonathan Frankle, Cody Blakeney, and John Patrick Cunningham. LoRA learns less and forgets less. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Joseph M. Cavanagh, Kunyang Sun, Andrew Gritsevskiy, Dorian Bagni, Yingze Wang, Thomas D. Bannister, and Teresa Head-Gordon. SmileyLlama: Modifying large language models for directed chemical space exploration. *Nature Computational Science*, 6(8):856–869, 2026. doi: 10.1038/s43588-026-00986-y.

Yongqiang Chen, Quanming Yao, Juzheng Zhang, James Cheng, and Yatao Bian. Hierarchical graph tokenization for molecule-language alignment. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 9664–9690. PMLR, 13–19 Jul 2025.

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

Djork-Arné Clevert, Tuan Le, Robin Winter, and Floriane Montanari. Img2mol – accurate smiles recognition from molecular graphical depictions. *Chemical Science*, 12(42):14174–14181, 11 2021. ISSN 2041-6520. doi: 10.1039/d1sc01839f.

Weichen Dai, Yezeng Chen, Zijie Dai, Yubo Liu, Zhijie Huang, Yixuan Pan, Baiyang Song, Chengli Zhong, Xinhe Li, Zeyu Wang, et al. Kale-lm-chem: Vision and practice toward an ai brain for chemistry. In *Pacific Rim International Conference on Artificial Intelligence*, pp. 623–631. Springer, 2025.

Tri Dao. Flashattention-2: Faster attention with better parallelism and work partitioning. In *The Twelfth International Conference on Learning Representations*, 2024.

Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. Qlora: Efficient finetuning of quantized llms. In *NeurIPS*, 2023.

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

Yin Fang, Xiaozhuan Liang, Ningyu Zhang, Kangwei Liu, Rui Huang, Zhuo Chen, Xiaohui Fan, and Huajun Chen. Mol-instructions: A large-scale biomolecular instruction dataset for large language models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, 2024.

Dimitris Gkoumas. Almol: Aligned language-molecule translation llms through offline preference contrastive optimisation. In *Proceedings of the 1st Workshop on Language+ Molecules (L+ M 2024)*, pp. 21–27, 2024.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.

Shuhan Guo, Yatao Bian, Ruibing Wang, Nan Yin, Zhen Wang, and Quanming Yao. Unified molecule-text language model with discrete token representation. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 9205–9213. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/1023. AI4Tech: AI Enabling Technologies.

Taicheng Guo, Kehan Guo, Bozhao Nan, Zhenwen Liang, Zhichun Guo, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. What can large language models do in chemistry? a comprehensive benchmark on eight tasks. *Advances in Neural Information Processing Systems*, 36, 2023.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

Junyu Hou. De novo molecular design enabled by direct preference optimization and curriculum learning. In Rita P. Ribeiro, Bernhard Pfahringer, Nathalie Japkowicz, Pedro Larrañaga, Alípio M. Jorge, Carlos Soares, Pedro H. Abreu, and João Gama (eds.), *Machine Learning and Knowledge Discovery in Databases. Research Track - European Conference, ECML PKDD 2025, Porto, Portugal, September 15-19, 2025, Proceedings, Part VI*, volume 16018 of *Lecture Notes in Computer Science*, pp. 95–111. Springer, 2025. doi: 10.1007/978-3-032-06106-5_6.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *ICLR*, 2022.

Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. Therapeutics data commons: Machine learning datasets and tasks for drug discovery and development. In J. Vanschoren and S. Yeung (eds.), *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, 2021.

Zihao Jing, Qiuhao Zeng, Ruiyi Fang, Yan Sun, Boyu Wang, and Pingzhao Hu. Entropy-guided dynamic tokens for graph-llm alignment in molecular understanding. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, 2026.

Dongki Kim, Wonbin Lee, and Sung Ju Hwang. Mol-llama: Towards general understanding of molecules in large molecular language model. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 26921–26960. Curran Associates, Inc., 2025. doi: 10.52202/085713-0908.

Chanhui Lee, Hanbum Ko, Yuheon Song, Yongjun Jeong, Rodrigo Hormazabal, Sehui Han, Kyunghoon Bae, Sungbin Lim, and Sungwoong Kim. Mol-LLM: Multimodal generalist molecular LLM with improved graph utilization. In *NeurIPS 2025 AI for Science Workshop*, 2025.

Hongyu Li, Liang Ding, Meng Fang, and Dacheng Tao. Revisiting catastrophic forgetting in large language model tuning. In *Findings of the association for computational linguistics: EMNLP 2024*, pp. 4297–4308, 2024.

Jiacheng Lin, Zhongruo Wang, Kun Qian, Tian Wang, Arvind Srinivasan, Hansi Zeng, Ruochen Jiao, Xie Zhou, Jiri Gesi, Dakuo Wang, Yufan Guo, Kai Zhong, Weiqi Zhang, sujay sanghavi, Changyou Chen, Hyokun Yun, and Lihong Li. SFT doesn't always hurt general capabilities: Revisiting domain-specific fine-tuning in LLMs. In *The Fourteenth International Conference on Learning Representations*, 2026.

Chengyuan Liu, Yangyang Kang, Shihang Wang, Lizhi Qing, Fubang Zhao, Chao Wu, Changlong Sun, Kun Kuang, and Fei Wu. More than catastrophic forgetting: Integrating general capabilities for domain-specific llms. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 7531–7548, 2024a.

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

Xuan Liu, Siru Ouyang, Xianrui Zhong, Jiawei Han, and Huimin Zhao. Fgbench: A dataset and benchmark for molecular property reasoning at functional group-level in large language models. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference. Curran Associates, Inc., 2025. doi: 10.52202/085713-4331.

Zequn Liu, Wei Zhang, Yingce Xia, Lijun Wu, Shufang Xie, Tao Qin, Ming Zhang, and Tie-Yan Liu. Molxpt: Wrapping molecules with text for generative pre-training. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 1606–1616, 2023a.

Zhiyuan Liu, Sihang Li, Yanchen Luo, Hao Fei, Yixin Cao, Kenji Kawaguchi, Xiang Wang, and Tat-Seng Chua. Molca: Molecular graph-language modeling with cross-modal projector and uni-modal adapter. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 15623–15638, 2023b.

Micha Livne, Zulfat Miftahutdinov, Elena Tutubalina, Maksim Kuznetsov, Daniil Polykovskiy, Annika Brundyn, Aastha Jhunjhunwala, Anthony Costa, Alex Aliper, Alán Aspuru-Guzik, et al. nach0: multimodal natural and chemical languages foundation model. *Chemical Science*, 15(22): 8380–8389, 2024.

Xingyu Lu, He Cao, Zijing Liu, Shengyuan Bai, Leqing Chen, Yuan Yao, Hai-Tao Zheng, and Yu Li. Moleculeqa: A dataset to evaluate factual accuracy in molecular comprehension. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 3769–3789, 2024.

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

Saeed Najafi and Alona Fyshe. Offline preference optimization via maximum marginal likelihood estimation. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6751–6764. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.eacl-long.318.

Yunhun Nam, Jaehyung Kim, and Jongheon Jeong. Learning from the undesirable: Robust adaptation of language models without forgetting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 32537–32545, 2026.

Arka Pal, Deep Karkhanis, Samuel Dooley, Manley Roberts, Siddartha Naidu, and Colin White. Smaug: Fixing failure modes of preference optimisation with dpo-positive. *CoRR*, abs/2402.13228, 2024. doi: 10.48550/ARXIV.2402.13228.

Ryan Park, Ryan Theisen, Rayees Rahman, Anna Cichońska, Marcel Patek, and Navriti Sahni. Preference optimization for molecular language models. In *NeurIPS 2023 Generative AI and Biology (GenBio) Workshop*, 2023.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural*

Information Processing Systems, volume 36, pp. 53728–53741. Curran Associates, Inc., 2023. doi: 10.52202/075280-2338.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.

Michael A. Stravs, Kai Dührkop, Sebastian Böcker, and Nicola Zamboni. Msnovelist: de novo structure generation from mass spectra. *Nature Methods*, 19:865–870, 2022. doi: 10.1038/s41592-022-01486-3.

Bing Su, Dazhao Du, Zhao Yang, Yujie Zhou, Jiangmeng Li, Anyi Rao, Hao Sun, Zhiwu Lu, and Ji-Rong Wen. A molecular multimodal foundation model associating molecule graphs with natural language. *CoRR*, abs/2209.05481, 2022. doi: 10.48550/ARXIV.2209.05481.

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

Zaifei Yang, Hong Chang, RuiBing Hou, Shiguang Shan, and Xilin Chen. Knowmol: Advancing molecular large language models with multi-level chemical knowledge. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference. Curran Associates, Inc., 2025. doi: 10.52202/085713-1328.

Tao Yin, Xiaohong Zhang, Jiacheng Zhang, Li Huang, Zhibin Zhang, Yuansong Zeng, Jin Xie, and Meng Yan. Mora: On-the-fly molecule-aware low-rank adaptation framework for llm-based multi-modal molecular assistant. *arXiv preprint arXiv:2510.12245*, 2025.

Botao Yu, Frazier N. Baker, Ziqi Chen, Xia Ning, and Huan Sun. LlaSMol: Advancing large language models for chemistry with a large-scale, comprehensive, high-quality instruction tuning dataset. In *First Conference on Language Modeling*, 2024.

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

Lanxue Zhang, Yuqiang Xie, Fang Fang, Fanglong Dong, Rui Liu, and Yanan Cao. Metagdpo: Alleviating catastrophic forgetting with metacognitive knowledge through group direct preference optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 34755–34763, 2026.

Zihan Zhao, Da Ma, Lu Chen, Liangtai Sun, Zihao Li, Yi Xia, Bo Chen, Hongshen Xu, Zichen Zhu, Su Zhu, Shuai Fan, Guodong Shen, Kai Yu, and Xin Chen. Developing chemdfm as a large language foundation model for chemistry. *Cell Reports Physical Science*, 6(4):102523, 2025. ISSN 2666-3864. doi: https://doi.org/10.1016/j.xcrp.2025.102523.

APPENDIX CONTENTS

Appendix A: Theoretical Analysis of Reciprocal Preference Construction. Appendix A analyzes how reciprocal preference construction constrains molecule-independent response reweighting and relates optimization beyond the reference-policy baseline to molecule–response interaction gain.

Appendix B: Preference Dataset Construction and Analysis. Appendix B.1 describes the template-aligned molecular pairing procedure and the construction of reciprocal and one-way preference data; Appendix B.2 reports the composition and statistics of the resulting preference datasets.

Appendix C: Molecular Adaptation Implementation Details. Appendix C.1 describes the backbone, adapter architecture, and optimization settings for the compared adaptation variants; Appendix C.2 reports training dynamics and compute budgets; and Appendix C.3 examines the effect of supervised fine-tuning learning rate on the adaptation–preservation trade-off.

Appendix D: Evaluation Details and Additional Results. Appendix D.1 summarizes the downstream benchmarks and evaluation setup; Appendix D.2 describes the data-contamination control procedure; Appendix D.3 defines the evaluation metrics; Appendix D.4 provides the MoleculeQA evaluation protocol, prompt, repair procedure, and results across support-set seeds; and Appendix D.5 provides the molecular property-prediction prompts, repair procedure, and per-task results.

A THEORETICAL ANALYSIS OF RECIPROCAL PREFERENCE CONSTRUCTION

We analyze how reciprocal preference construction constrains molecule-independent response reweighting. With equally weighted directions, such reweighting produces opposite reference-relative margins and cannot, by itself, reduce the average DPO loss below the reference-policy baseline. We further show that improvement beyond this baseline requires a positive reference-relative interaction gain between molecular conditions and responses.

Setup and reference-relative margins. Consider two molecule–response instances (m_i, y_i) and (m_j, y_j) sharing the same instruction x. Their reciprocal preferences are

$$y_i \succ_{(x,m_i)} y_j, \qquad y_j \succ_{(x,m_j)} y_i.$$

Fix a reference policy π_{ref}. For any policy π under analysis, assume that π and π_{ref} assign positive probabilities to the four evaluated molecule–response combinations. Define the reference-relative log-likelihood score

$$r_\pi(x, m, y) = \log \frac{\pi(y \mid x, m)}{\pi_{\mathrm{ref}}(y \mid x, m)}. \tag{9}$$

For the selected pair, abbreviate the two directional margins as

$$\Delta_i(\pi) = r_\pi(x, m_i, y_i) - r_\pi(x, m_i, y_j), \tag{10}$$

$$\Delta_j(\pi) = r_\pi(x, m_j, y_j) - r_\pi(x, m_j, y_i). \tag{11}$$

When $\pi = \pi_\theta$, these are the two instances of the reference-relative margin in Eq. 6.

Following the DPO objective (Rafailov et al., 2023), let

$$h_\beta(u) = -\log \sigma(\beta u) = \log\left(1 + \exp(-\beta u)\right), \qquad \beta > 0. \tag{12}$$

With the same β and equal weights for both directions, the reciprocal-pair loss is

$$L_{\mathrm{rec}}(\pi) = \frac{1}{2}\left[h_\beta\big(\Delta_i(\pi)\big) + h_\beta\big(\Delta_j(\pi)\big)\right]. \tag{13}$$

At the reference policy, both margins vanish. Hence the reference-policy baseline is

$$L_0 := L_{\mathrm{rec}}(\pi_{\mathrm{ref}}) = h_\beta(0) = \log 2. \tag{14}$$

Since ReMPA initializes $\pi_{\theta^{(0)}} = \pi_{\mathrm{ref}} = \pi_0$, this is also its zero-update baseline.

Molecule-independent response reweighting. Let $b(x, y)$ be a real-valued response score that depends on the instruction and response, but not on the molecular condition. Define the response-reweighted policy

$$\pi_b(y \mid x, m) = \frac{\pi_{\text{ref}}(y \mid x, m) \exp b(x, y)}{Z_b(x, m)}, \tag{15}$$

where

$$Z_b(x, m) = \sum_{y'} \pi_{\text{ref}}(y' \mid x, m) \exp b(x, y'), \qquad 0 < Z_b(x, m) < \infty \tag{16}$$

for all inputs under analysis. This defines a class of molecule-independent adaptation mechanisms: the same response log-weight is applied across molecular conditions. The resulting reference-relative score is

$$r_{\pi_b}(x, m, y) = b(x, y) - \log Z_b(x, m). \tag{17}$$

The normalizer may depend on the molecule, but cancels when comparing responses under the same molecular condition.

One-way preference fitting. A directional DPO loss depends only on the difference between the chosen and rejected reference-relative scores. Consequently, it can decrease even when the chosen likelihood decreases, provided the rejected likelihood decreases more strongly (Pal et al., 2024).

This behavior can be realized by a normalized response-only reweighting. For a single directional preference with $y_i \neq y_j$, write

$$q_i = \pi_{\text{ref}}(y_i \mid x, m_i), \qquad q_j = \pi_{\text{ref}}(y_j \mid x, m_i),$$

and suppose that the remaining responses have total reference probability $q_o = 1 - q_i - q_j > 0$. For $t \geq 0$, choose

$$b_t(x, y) = \begin{cases} -t, & y = y_i, \\ -2t, & y = y_j, \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

At (x, m_i), the normalizer is

$$Z_t := Z_{b_t}(x, m_i) = q_o + q_i e^{-t} + q_j e^{-2t}.$$

Using Eq. 17,

$$\Delta_i(\pi_{b_t}) = (-t - \log Z_t) - (-2t - \log Z_t) = t. \tag{19}$$

Thus,

$$h_\beta\big(\Delta_i(\pi_{b_t})\big) = h_\beta(t) \longrightarrow 0 \qquad \text{as } t \to \infty.$$

Meanwhile, $Z_t \to q_o > 0$ implies

$$\frac{\pi_{b_t}(y_i \mid x, m_i)}{\pi_{\text{ref}}(y_i \mid x, m_i)} = \frac{e^{-t}}{Z_t} \longrightarrow 0, \qquad \frac{\pi_{b_t}(y_j \mid x, m_i)}{\pi_{\text{ref}}(y_j \mid x, m_i)} = \frac{e^{-2t}}{Z_t} \longrightarrow 0. \tag{20}$$

Therefore, arbitrarily small one-way preference loss is compatible with an arbitrarily large relative reduction in the chosen likelihood, achieved here through molecule-independent response reweighting.

Proposition 1 (Response-only adaptation cannot improve upon the reciprocal baseline). For any policy π_b in Eq. 15, let

$$a = b(x, y_i) - b(x, y_j).$$

Then

$$\Delta_i(\pi_b) = a, \qquad \Delta_j(\pi_b) = -a, \tag{21}$$

and

$$L_{\text{rec}}(\pi_b) \geq L_0, \tag{22}$$

with equality if and only if $a = 0$.

Proof. Starting from the definition of the forward margin and substituting Eq. 17, we obtain

$$\begin{aligned}\Delta_i(\pi_b) &= r_{\pi_b}(x, m_i, y_i) - r_{\pi_b}(x, m_i, y_j) \\ &= [b(x, y_i) - \log Z_b(x, m_i)] - [b(x, y_j) - \log Z_b(x, m_i)] \\ &= b(x, y_i) - b(x, y_j) = a. \end{aligned} \tag{23}$$

The two responses are evaluated under the same molecular condition m_i, so the molecule-dependent normalizer cancels from their reference-relative difference. For the reciprocal direction,

$$\begin{aligned}\Delta_j(\pi_b) &= r_{\pi_b}(x, m_j, y_j) - r_{\pi_b}(x, m_j, y_i) \\ &= [b(x, y_j) - \log Z_b(x, m_j)] - [b(x, y_i) - \log Z_b(x, m_j)] \\ &= b(x, y_j) - b(x, y_i) = -a. \end{aligned} \tag{24}$$

Thus, the same molecule-independent difference $b(x, y_i) - b(x, y_j)$ enters the two reciprocal directions with opposite signs. The scalar loss satisfies

$$h'_\beta(u) = -\frac{\beta}{1 + \exp(\beta u)} < 0, \qquad h''_\beta(u) = \frac{\beta^2 \exp(\beta u)}{(1 + \exp(\beta u))^2} > 0. \tag{25}$$

Hence increasing a decreases the forward loss and increases the reverse loss. By strict convexity,

$$\begin{aligned}L_{\mathrm{rec}}(\pi_b) &= \frac{1}{2}\left[h_\beta(a) + h_\beta(-a)\right] \\ &\geq h_\beta\left(\frac{a + (-a)}{2}\right) = h_\beta(0) = L_0. \end{aligned} \tag{26}$$

Equality holds precisely when $a = -a$, or equivalently $a = 0$. □

In particular, suppressing y_j through a molecule-independent response bias improves its margin as the rejected response under m_i, but worsens its margin as the chosen response under m_j. Reciprocal construction therefore makes the same response-level change act in opposite directions.

Reference-relative interaction gain. For a general policy π, define

$$\begin{aligned}\Gamma(\pi) &:= \Delta_i(\pi) + \Delta_j(\pi) \\ &= r_\pi(x, m_i, y_i) + r_\pi(x, m_j, y_j) \\ &\quad - r_\pi(x, m_i, y_j) - r_\pi(x, m_j, y_i). \end{aligned} \tag{27}$$

This contrast compares the matched combinations (m_i, y_i) and (m_j, y_j) with the exchanged combinations (m_i, y_j) and (m_j, y_i).

To express this comparison directly in terms of probabilities, let

$$C(\pi) = \frac{\pi(y_i \mid x, m_i)\pi(y_j \mid x, m_j)}{\pi(y_j \mid x, m_i)\pi(y_i \mid x, m_j)}. \tag{28}$$

Expanding Eq. 27 yields

$$\Gamma(\pi) = \log \frac{C(\pi)}{C(\pi_{\mathrm{ref}})}. \tag{29}$$

Thus, $\Gamma(\pi) > 0$ means that the policy has increased the matched-to-exchanged probability ratio relative to the reference policy. The alternating sum in Eq. 27 cancels any additive molecule-only or response-only score term.

Response reweighting redistributes directional margins. The invariance of the interaction gain extends to response reweighting applied to any current policy. Given π, define

$$\pi^b(y \mid x, m) = \frac{\pi(y \mid x, m) \exp b(x, y)}{Z_{\pi,b}(x, m)}, \qquad Z_{\pi,b}(x, m) = \sum_{y'} \pi(y' \mid x, m) \exp b(x, y'), \tag{30}$$

where $0 < Z_{\pi,b}(x, m) < \infty$. With the reference policy held fixed,

$$r_{\pi^b}(x, m, y) = r_\pi(x, m, y) + b(x, y) - \log Z_{\pi,b}(x, m). \tag{31}$$

Writing $a = b(x, y_i) - b(x, y_j)$ and again canceling the normalizers gives

$$\Delta_i(\pi^b) = \Delta_i(\pi) + a, \qquad \Delta_j(\pi^b) = \Delta_j(\pi) - a. \tag{32}$$

Consequently,

$$\Gamma(\pi^b) = \Gamma(\pi). \tag{33}$$

Molecule-independent response reweighting therefore transfers margin between the two directions without increasing their total interaction gain.

Proposition 2 (Reciprocal improvement requires positive interaction gain). For any policy π satisfying the positivity conditions above,

$$L_{\mathrm{rec}}(\pi) \geq h_\beta\left(\frac{\Gamma(\pi)}{2}\right), \tag{34}$$

with equality if and only if $\Delta_i(\pi) = \Delta_j(\pi)$. In particular,

$$L_{\mathrm{rec}}(\pi) < L_0 \quad \Longrightarrow \quad \Gamma(\pi) > 0. \tag{35}$$

Proof. By the strict convexity established in Eq. 25,

$$\begin{aligned} L_{\mathrm{rec}}(\pi) &= \frac{1}{2}\left[h_\beta\big(\Delta_i(\pi)\big) + h_\beta\big(\Delta_j(\pi)\big)\right] \\ &\geq h_\beta\left(\frac{\Delta_i(\pi) + \Delta_j(\pi)}{2}\right) \\ &= h_\beta\left(\frac{\Gamma(\pi)}{2}\right). \end{aligned} \tag{36}$$

Equality holds precisely when the two margins are equal. If $L_{\mathrm{rec}}(\pi) < L_0$, then

$$h_\beta\left(\frac{\Gamma(\pi)}{2}\right) < h_\beta(0).$$

Since h_β is strictly decreasing, this implies $\Gamma(\pi) > 0$. □

At a fixed interaction gain, the lower bound in Eq. 34 is attained when the two directional margins are equal. Response reweighting can therefore balance existing margins, while Eq. 33 shows that it cannot increase the interaction gain determining this lower bound. Starting from the reference policy, this gain is zero, which recovers the response-only baseline restriction in Proposition 1.

Averaging over reciprocal pairs. Let $\mathcal{P}$ be the finite set of selected pairs in Eq. 8, and let $\mathbb{E}_{P\sim\mathcal{P}}$ denote its empirical average. Restoring the pair argument, write $L_{\mathrm{rec}}(\pi; P)$ and $\Gamma(\pi; P)$ for $P = (x, m_i, y_i, m_j, y_j)$. Applying Proposition 2 to each pair and then applying convexity across pairs gives

$$\begin{aligned} \mathcal{L}_{\mathrm{ReMPA}}(\theta) &= \mathbb{E}_{P\sim\mathcal{P}}\left[L_{\mathrm{rec}}(\pi_\theta; P)\right] \\ &\geq \mathbb{E}_{P\sim\mathcal{P}}\left[h_\beta\left(\frac{\Gamma(\pi_\theta; P)}{2}\right)\right] \\ &\geq h_\beta\left(\frac{1}{2}\mathbb{E}_{P\sim\mathcal{P}}\left[\Gamma(\pi_\theta; P)\right]\right). \end{aligned} \tag{37}$$

It follows that

$$\mathcal{L}_{\mathrm{ReMPA}}(\theta) < L_0 \quad \Longrightarrow \quad \mathbb{E}_{P\sim\mathcal{P}}\left[\Gamma(\pi_\theta; P)\right] > 0. \tag{38}$$

Moreover, Proposition 1 holds for every selected pair, so any molecule-independent response reweighting of the reference policy satisfies

$$\mathbb{E}_{P\sim\mathcal{P}}\left[L_{\mathrm{rec}}(\pi_b; P)\right] \geq L_0. \tag{39}$$

Thus, reducing the training objective below its zero-update baseline requires a positive average reference-relative interaction gain.

First-order training signal at reference initialization. Suppose the evaluated log-likelihoods are differentiable with respect to θ at $\theta^{(0)}$. Since $\pi_{\theta^{(0)}} = \pi_{\mathrm{ref}}$, both directional margins are zero at initialization. Using $h'_\beta(0) = -\beta/2$, the chain rule gives

$$\begin{aligned}\nabla_\theta L_{\mathrm{rec}}(\pi_\theta)|_{\theta=\theta^{(0)}} &= -\frac{\beta}{4}\left[\nabla_\theta \Delta_i(\pi_\theta) + \nabla_\theta \Delta_j(\pi_\theta)\right]|_{\theta=\theta^{(0)}} \\ &= -\frac{\beta}{4}\,\nabla_\theta \Gamma(\pi_\theta)|_{\theta=\theta^{(0)}}\,. \end{aligned} \tag{40}$$

Averaging this identity over the finite pair set yields

$$\nabla_\theta \mathcal{L}_{\mathrm{ReMPA}}(\theta)|_{\theta=\theta^{(0)}} = -\frac{\beta}{4}\,\nabla_\theta \mathbb{E}_{P\sim\mathcal{P}}\left[\Gamma(\pi_\theta; P)\right]|_{\theta=\theta^{(0)}}\,. \tag{41}$$

The initial first-order optimization signal is therefore the average interaction gain. Along a differentiable response-only reweighting path through the reference policy, this gain remains zero, and the reciprocal objective has zero first-order change.

Implication for reciprocal molecular adaptation. Equal-weight reciprocal preferences assign opposite roles to the same responses under different molecular conditions. Molecule-independent response reweighting consequently produces opposing directional margins: it can redistribute these margins, but cannot increase the matched-to-exchanged interaction gain or, by itself, improve upon the reference-policy baseline. Improvement beyond this baseline requires a positive reference-relative interaction gain, and the initialization gradient directly reflects this quantity.

This dependence on the molecular condition is aligned with the supervision that ReMPA aims to extract from paired molecular data. For two examples sharing the same instruction, the preference between y_i and y_j reverses when the molecular input changes from m_i to m_j. The informative distinction is therefore not a global property of either response alone, but how each response matches the molecule under which it is evaluated. Reciprocal construction makes this condition-dependent distinction relevant to both preference directions, limiting pure response bias as a standalone adaptation mechanism and supplying training pressure for molecule-dependent changes in relative response scores.

B Preference Dataset Construction and Analysis

B.1 Template-Aligned Molecular Pairing

Data source and preprocessing. We construct the preference data from the template-aligned, single-turn molecular instruction data released with Mol-LLaMA-Instruct (Kim et al., 2025). Specifically, we use the *Detailed Structural Descriptions*, *Structure-to-Chemical Features Relationships*, and *Structure-to-Biological Features Relationships* subsets, which contain 7, 9, and 9 instruction templates, respectively, for a total of 25 templates. We exclude *Comprehensive Conversations*, since it contains diverse multi-turn interactions rather than template-aligned single-turn supervision. Before constructing molecular pairs, we remove molecules that overlap with the downstream molecular evaluation benchmarks using the benchmark-contamination control procedure described in Appendix D.2.

To identify the instruction template of each example, we remove the standalone molecule-placeholder line from the user message and group examples according to the remaining normalized instruction text. Molecules are canonicalized from their SMILES representations, and the resulting unique canonical molecules are split into training and development sets at a 9:1 ratio using random seed 42. All records associated with the same canonical molecule are assigned to the same split.

Template-aligned pair selection. Molecular pairing is performed separately within each source subset, instruction template, and data split. For each molecule–response instance (x_i, m_i, y_i), candidate partners (x_j, m_j, y_j) are restricted to examples sharing the same normalized instruction template, $x_j = x_i$, while containing a different molecule, $m_j \neq m_i$. This restriction ensures that the paired responses address the same molecular task.

Within each template-aligned candidate set, structural similarity is used to select plausible response alternatives. Molecules are represented using Morgan fingerprints with radius 2, 2048 bits, and

Table 7: Statistics of the template-aligned molecular pairs and the preference datasets derived from them. ReMPA includes both reciprocal directions of each selected pair, while One-way DPO retains one direction per pair.

Statistic	Value
Training molecular pairs	25,000
Development molecular pairs	2,500
ReMPA training directional examples	50,000
ReMPA development directional examples	5,000
One-way DPO training directional examples	25,000
One-way DPO development directional examples	2,500
Instruction templates	25 (7 / 9 / 9)
Structural / chemical / biological training pairs	7K / 9K / 9K
Mean Dice similarity	0.738

chirality enabled (Rogers & Hahn, 2010). Candidate molecules are ranked by Dice similarity, and we retain up to the top 10 neighbors with a minimum Dice similarity of 0.50. We further require the two molecules to occur in each other's top-10 neighbor sets and select disjoint molecular pairs, so that each molecule participates in at most one selected pair within a template. Competing candidate edges are resolved greedily in increasing mutual-neighbor rank, traversing molecules in a fixed seed-42 order and accepting the first eligible edge whose two endpoints have not yet been selected. Pairs whose original responses are identical are discarded.

Reciprocal and one-way preference construction. For each selected molecular pair, ReMPA constructs the reciprocal preferences defined in Section 3, such that each response is preferred under its corresponding molecular condition and dispreferred under the other. Consequently, each selected pair contributes two directional preference examples.

For the one-way DPO control, we deterministically order the two records in each selected molecular pair by record identifier and retain only the first-to-second preference direction. The molecular pair identities, responses, preprocessing procedure, and pair-selection criteria are otherwise unchanged. Consequently, ReMPA and One-way DPO differ in whether both conditional preference directions are included, while being derived from the same underlying set of template-aligned molecular pairs.

B.2 Preference Dataset Statistics

Dataset composition. The final pair set contains 25,000 training molecular pairs and 2,500 development pairs across 25 instruction templates. The training set comprises 7,000 structural, 9,000 chemical, and 9,000 biological pairs, reflecting the 7, 9, and 9 templates in the three source subsets, respectively.

Each selected pair contributes two directional examples to the reciprocal ReMPA dataset and one directional example to the One-way DPO dataset. Accordingly, ReMPA contains 50,000 training and 5,000 development preference examples, whereas One-way DPO contains 25,000 training and 2,500 development examples. Because both variants are derived from the same molecular pairs, they share the same pair-level composition and structural-similarity statistics. The selected molecular pairs have a mean Dice similarity of 0.738. Table 7 summarizes the resulting datasets.

C Molecular Adaptation Implementation Details

C.1 Model and Adapter Implementation

Backbone and parameter-efficient adaptation. All models in our controlled comparisons use Llama-3.1-8B-Instruct as the common backbone (Grattafiori et al., 2024). The Base model denotes the unadapted backbone without molecular-domain training. For all adapted variants, we keep the pretrained backbone frozen and optimize low-rank adapters using a QLoRA-style implementation (Hu et al., 2022; Dettmers et al., 2023). The backbone weights are loaded with 4-bit NF4 quantization and double quantization, with bfloat16 computation. We use LoRA rank $r = 8$ and scaling

Table 8: Implementation and optimization settings for the molecular adaptation variants. ReMPA and One-way DPO are constructed from the same 25K molecular pairs; ReMPA includes both reciprocal directions, whereas One-way DPO retains one direction per pair. Training One-way DPO for 12 epochs and ReMPA for 6 epochs matches their total directional preference-example exposure. SFT-Full is capped at 2,346 optimizer updates, matching the update budget of the 50K SFT controls.

Setting	**SFT-50K**	**SFT-Full**	**One-way DPO**	**ReMPA**
Training scale	50K	∼200K	25K dir.	25K pairs / 50K dir.
Objective	SFT	SFT	DPO	DPO
Preference directions per pair	–	–	1	2
Training duration	6 epochs	2,346 steps	12 epochs	6 epochs
Initial LR	1×10^{-4}	1×10^{-4}	5×10^{-6}	5×10^{-6}
Minimum LR	5×10^{-6}	5×10^{-6}	5×10^{-7}	5×10^{-7}
Warmup	200 steps	200 steps	20 steps	20 steps
Weight decay	0.05	0.05	0.05	0.05
Global batch size	128	128	128	128
Max. sequence length	1024	1024	1024	1024
Quantization	NF4 4-bit	NF4 4-bit	NF4 4-bit	NF4 4-bit
LoRA r/α	8 / 32	8 / 32	8 / 32	8 / 32
DPO β	–	–	0.1	0.1

factor $\alpha = 32$, with adapters inserted into all attention projections and MLP projections. Gradient checkpointing and FlashAttention-2 (Dao, 2024) are enabled for all adapted models.

All training examples follow the Llama-3 chat format with a maximum sequence length of 1024 tokens. For SFT, the language-modeling loss is applied only to assistant-response tokens, while system and user tokens are masked from the objective. For preference optimization, sequence log-likelihoods are likewise computed only over the assistant completions. The LoRA configuration uses a dropout rate of 0.1 for supervised fine-tuning. For both ReMPA and One-way DPO, stochastic dropout is disabled during preference optimization so that the chosen and rejected branches are scored deterministically. This removes stochastic variation from their likelihood difference and yields a directly comparable preference margin against the fixed cached reference scores.

Optimization settings. All adapted models are optimized with AdamW using weight decay 0.05. Gradient accumulation is used to maintain an effective global batch size of 128 across all adaptation methods.

SFT-50K is trained on 50K examples for six epochs with a peak learning rate of 1×10^{-4} and cosine decay to 5×10^{-6}, following the learning-rate range used for Mol-LLaMA instruction tuning (Kim et al., 2025), with 200 linear warmup steps. SFT-FULL uses the same optimization schedule on the full approximately 200K-example molecular adaptation set, but is capped at 2,346 optimizer updates, matching the update budget of SFT-50K and corresponding to approximately 1.5 passes over the full training set.

Both preference-based variants directly initialize from the same Llama-3.1-8B-Instruct base model and optimize the standard DPO objective with $\beta = 0.1$. They use an initial learning rate of 5×10^{-6}, 20 linear warmup steps, and cosine decay to 5×10^{-7}. ReMPA is trained for six epochs on the reciprocal preference dataset, in which each selected molecular pair contributes two directional preference examples. One-way DPO is constructed from the same molecular pairs but retains only one direction per pair and is therefore trained for twelve epochs. This doubles the number of passes over the one-way dataset so that ReMPA and One-way DPO receive the same total number of directional preference-example exposures.

For model selection, the SFT variants use the checkpoint with the lowest validation loss. For the preference-based variants, we select the earliest checkpoint entering the validation-loss plateau. The pre-training evaluation at optimizer step 0 corresponds to the base-model initialization and is not counted as a training epoch. The selected checkpoints are at optimizer step 775 for One-way DPO (epoch 3) and optimizer step 1,939 for ReMPA (epoch 4) under our zero-based epoch indexing.

The main implementation and optimization settings are summarized in Table 8.

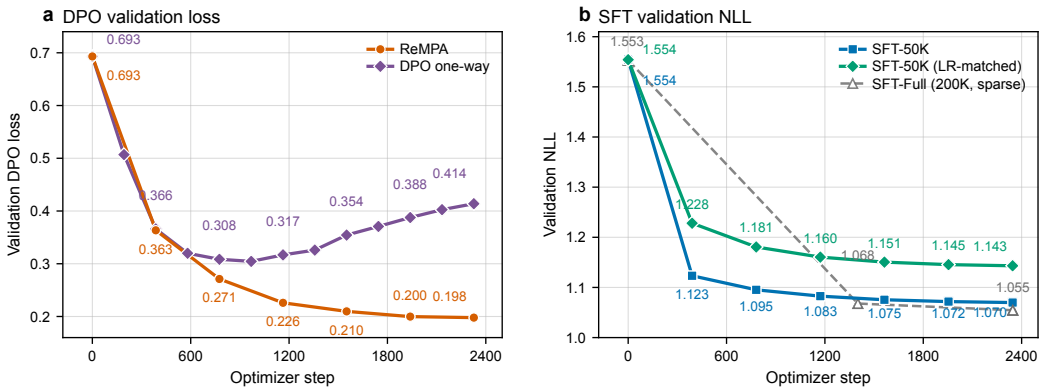

Figure 3: Validation-objective trajectories over optimizer steps. **(a)** Validation DPO loss for ReMPA and One-way DPO. The selected checkpoints are optimizer step 775 (epoch 3) for One-way DPO and optimizer step 1,939 (epoch 4) for ReMPA under zero-based epoch indexing, corresponding to the fifth and sixth plotted validation points, respectively, when the step-0 base-model initialization is counted as the first point. **(b)** Validation negative log-likelihood (NLL) for SFT-50K, the learning-rate-matched SFT control, and SFT-Full; only sparse validation measurements are available for SFT-Full. Loss values are comparable within, but not across, the two objective families.

C.2 Training Dynamics and Compute Budget

Optimization budget. The supervised and preference-based variants are trained under closely comparable optimizer-update budgets. SFT-50K is trained for 2,346 optimizer updates, while SFT-FULL is explicitly capped at the same number of updates. At their maximum scheduled training horizons, ReMPA and One-way DPO each receive approximately 2.3K optimizer updates. Because the directional dataset of One-way DPO is half the size of that of ReMPA, One-way DPO is trained for twice as many epochs, yielding approximately the same total number of directional preference-example exposures.

Preference optimization is more expensive per training example than SFT because each preference example contains both a preferred and a rejected completion, requiring the policy to score two response branches rather than one. With cached reference-model log-likelihoods, the frozen reference model does not require an additional forward pass during optimization. In our implementation, this makes the training time per preference example approximately twice that of an SFT example.

Figure 3 shows the validation-loss trajectories of the adaptation variants. The supervised models generally approach their validation-loss plateau earlier than the preference-based variants. This trend is also observed for the additional SFT-50K (LR-matched) control, which uses the same initial learning rate as ReMPA and is analyzed further in Appendix C.3. Among the preference-based models, ReMPA decreases steadily across training, whereas One-way DPO reaches its minimum earlier and its validation loss subsequently increases.

C.3 SFT Learning-Rate Control

To control for the effect of optimization strength, we train an additional supervised baseline, denoted SFT-50K (LR-matched), on the same 50K-example supervised dataset using the same initial learning rate as ReMPA, 5×10^{-6}. This control examines whether the preservation advantage of ReMPA can be explained primarily by its lower learning rate rather than the adaptation objective.

Using the lower learning rate alleviates the degradation associated with supervised molecular adaptation in general-purpose capability, output-format adherence, and molecular reasoning performance, consistent with previous findings that milder SFT updates can better preserve existing capabilities (Lin et al., 2026). Nevertheless, even with the same initial learning rate as ReMPA, the LR-matched SFT control remains weaker on the preservation-oriented evaluations and molecular property prediction, while achieving comparable MoleculeQA performance.

Table 9: Comparison between learning-rate-matched SFT and ReMPA. FGBench reports mean Boolean accuracy over its three subsets. Property reports five-task macro-averaged balanced accuracy before and after one format-repair attempt, together with the initial valid-output rate.

	MoleculeQA	FGBench	MMLU	Property		
Model	**Acc.**	**Bool Avg.**	**Micro Acc.**	**Initial BA**	**Initial Parse**	**Repair BA**
SFT-50K (LR-matched)	41.88	49.60	67.85	43.43	82.30	52.04
ReMPA	41.23	51.10	68.29	54.09	94.87	56.27

These results suggest that reducing the SFT learning rate improves the adaptation–preservation trade-off, but does not fully recover the broader preservation profile of ReMPA. The observed difference between supervised and preference-based adaptation therefore cannot be attributed to the initial learning rate alone.

D Evaluation Details and Additional Results

D.1 Benchmarks and Evaluation Setup

We evaluate the models across four complementary settings covering targeted molecular capability acquisition, functional-group-level molecular reasoning, general-purpose capability preservation, and molecular property prediction. Within each benchmark, all of our model variants are evaluated on the same examples under an identical prompting, decoding, and scoring protocol.

MoleculeQA. MoleculeQA is a multiple-choice benchmark for evaluating factual molecular comprehension from molecular representations (Lu et al., 2024). It organizes questions into four aspects: *Structure*, *Source*, *Property*, and *Application*. We evaluate on the complete 5,786-question test split using an aspect-specific 10-shot protocol aligned with the original MoleculeQA evaluation. For each aspect, ten demonstrations are drawn from the official training split and kept fixed across all compared model variants for a given support-set seed. Each test instance is presented with its molecular SMILES, question, and four answer choices, and the model is instructed to return one of A/B/C/D. Generation is deterministic, with sampling disabled. The detailed prompt construction, output parsing, format-repair procedure, and results across support-set seeds are provided in Appendix D.4.

FGBench. FGBench evaluates molecular property reasoning at the functional-group level (Liu et al., 2025). The benchmark contains three reasoning settings, *Single*, *Interaction*, and *Comparison*, each with Boolean and value-based questions. We evaluate on the complete 7,146-example FGBench test set using the benchmark's Llama-3.1 prompting format. Models are evaluated zero-shot with deterministic greedy decoding and a maximum generation budget of 1,024 new tokens. We preserve the benchmark questions without truncation and use the official FGBench answer-extraction and evaluation procedure to compute the reported Boolean accuracy, numerical RMSE, and valid-output rates.

MMLU. We use Massive Multitask Language Understanding (MMLU) (Hendrycks et al., 2021) to measure preservation of the general-purpose capabilities of the base model. MMLU consists of multiple-choice questions spanning 57 subjects across STEM, humanities, social sciences, and other professional and general-knowledge domains. We reproduce the original 5-shot evaluation protocol using the raw MMLU CSV test sets, with the first five development examples of each subject serving as in-context demonstrations. Rather than generating a free-form response, we score the conditional log-likelihood of the four answer continuations and select the option with the highest score. The maximum input length is 2,048 tokens. We report overall micro and subject-level macro accuracy together with results over the four broad MMLU subject groups.

Molecular Property Prediction. We additionally evaluate molecular property prediction on five binary classification tasks. BBBP, DILI, hERG, and HIA use the `BBB_Martins`, `DILI`, `hERG`, and `HIA_Hou` datasets, respectively, from Therapeutics Data Commons (TDC) (Huang et al., 2021), while BACE uses the BACE classification dataset from MoleculeNet (Wu et al., 2018). These tasks

cover both molecular activity and ADME/toxicity-related properties, providing a complementary view of how molecular adaptation changes downstream prediction behavior. We use a SMILES-only, text-based evaluation with one Direct prompt per task, following the text-only Direct evaluation format used in EDT-Former (Jing et al., 2026). For each task, all model variants are evaluated on the same frozen test set using an identical task-specific prompt. Predictions are parsed only from the generated completion. We record the initial prediction and, when the initial output cannot be parsed into one of the two task labels, allow exactly one format-repair generation. Detailed task prompts, parsing and repair rules, and per-task results are provided in Appendix D.5.

D.2 Data Contamination Control

To prevent molecular identity overlap between the training data and downstream evaluations, we perform molecule-level decontamination before constructing the final training sets. We consider all molecular evaluations used in this work, including MoleculeQA (Lu et al., 2024), FGBench (Liu et al., 2025), and the five property-prediction tasks drawn from MoleculeNet (Wu et al., 2018) and Therapeutics Data Commons (Huang et al., 2021). MMLU contains no molecular inputs and therefore does not require molecule-level decontamination.

Connectivity-level molecular identity. We determine molecular identity by connectivity rather than by literal SMILES matching. Connectivity-level, stereochemistry-agnostic identities are a common choice for molecular data curation and deduplication. For example, MSNovelist uses a two-dimensional InChIKey representation based on atom connectivity without stereochemistry for molecular identity, while Img2Mol removes stereochemical information before duplicate filtering and canonical-SMILES construction (Stravs et al., 2022; Clevert et al., 2021). Following the same conservative principle, we parse each SMILES string with RDKit and convert it to a canonical non-isomeric SMILES representation, with canonicalization enabled and stereochemical information omitted. Molecules yielding the same representation are therefore treated as identical for decontamination, even when their original SMILES strings or stereochemical annotations differ.

Filtering procedure. Before constructing the final training sets, we remove molecular training records whose connectivity overlaps with the downstream molecular evaluation sets. For MoleculeQA, the complete test split is considered. For FGBench, both the target and reference molecules appearing in each test example are included in the overlap check. We similarly exclude molecules occurring in the test sets of the five property-prediction tasks. Filtering is performed before the molecule-level train–development split and before template-aligned molecular pairing, preventing an evaluation molecule from appearing as either member of a selected molecular pair.

The supervised controls follow the same contamination-control criterion. SFT-50K and SFT-Full are constructed from the cleaned molecular training pool. ReMPA and One-way DPO are derived from the same decontaminated set of template-aligned molecular pairs, with One-way DPO retaining only one directional preference from each pair.

Overlap audit. The original molecular training pool contains non-negligible overlap with the downstream evaluations, motivating explicit molecule-level decontamination. As summarized in Table 10, the union of the molecular evaluation sets contains 4,298 unique connectivity identities, of which 1,692 also occur in the original training pool and correspond to 3,463 molecule–response records. The per-benchmark statistics further show that the extent of overlap varies substantially across datasets. All such overlaps are removed before construction of the training sets.

D.3 Evaluation Metrics

We use benchmark-specific metrics following the corresponding evaluation protocols. Accuracy-based metrics, balanced accuracy, Macro-F1, recall, and valid-output rates are better when higher, whereas RMSE is better when lower.

Accuracy. For classification and multiple-choice evaluations, accuracy is defined as

$$\mathrm{Acc} = \frac{N_{\mathrm{correct}}}{N}. \tag{42}$$

Table 10: Connectivity-level overlap between the original molecular training pool and the downstream molecular evaluation sets before decontamination. Source records denotes the number of molecule–response records in the original training pool associated with the overlapping connectivities. Per-benchmark rows are not additive because the same molecular connectivity may occur in multiple evaluation sets.

Evaluation set	Unique eval. connectivities	Overlapping connectivities	Source records
MoleculeQA test	2,703	1,333	2,783
FGBench test	834	182	340
BBBP test	391	126	242
BACE test	152	1	1
DILI test	96	55	114
hERG test	121	18	31
HIA test	112	42	83
Union	4,298	1,692	3,463

We use accuracy for MoleculeQA (Lu et al., 2024), the Boolean subsets of FGBench (Liu et al., 2025), and MMLU (Hendrycks et al., 2021). MoleculeQA reports accuracy separately for *Structure*, *Source*, *Property*, and *Application*, together with overall accuracy over the complete test set. For FGBench Boolean tasks, invalid outputs are counted as incorrect.

For MMLU, we report both micro and subject-macro accuracy. Micro accuracy is computed over all test questions jointly, whereas subject-macro accuracy first computes accuracy independently for each of the 57 subjects and then averages the resulting subject-level scores:

$$\mathrm{Acc}_{\mathrm{macro}} = \frac{1}{S}\sum_{s=1}^{S}\mathrm{Acc}_s, \qquad S = 57. \tag{43}$$

We additionally report accuracy over the four broad subject groups: STEM, Humanities, Social Sciences, and Other.

Recall and Balanced Accuracy. For the binary molecular property-prediction tasks, we report recall for both the positive and negative classes:

$$R_+ = \frac{TP}{TP+FN}, \qquad R_- = \frac{TN}{TN+FP}. \tag{44}$$

Balanced accuracy (BA) gives the two classes equal weight and is defined as

$$\mathrm{BA} = \frac{R_+ + R_-}{2}. \tag{45}$$

We use BA as the primary classification metric for the five property tasks, since it remains informative when the class distributions are imbalanced.

Macro-F1. For each class $c \in \{0,1\}$, we compute the standard class-specific F1 score $F1_c$. Macro-F1 is then the unweighted mean over the two classes:

$$\text{Macro-F1} = \frac{F1_0 + F1_1}{2}. \tag{46}$$

Together with positive- and negative-class recall, Macro-F1 provides a class-balanced view of the prediction behavior beyond BA alone.

Root Mean Squared Error. For the value-based FGBench tasks (Liu et al., 2025), numerical predictions are evaluated using root mean squared error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}}(\hat{y}_i - y_i)^2}, \tag{47}$$

where $\mathcal{V}$ denotes the subset of predictions considered valid by the official FGBench evaluator. Because invalid outputs are excluded from RMSE, we interpret RMSE together with the corresponding valid-output rate.

Output Validity. We explicitly track output validity for evaluations requiring structured answers. FGBench reports *Valid*, the fraction of outputs from which the official evaluator can extract a valid Boolean or numerical answer. For molecular property prediction, we analogously report *Parse*, the fraction of outputs that can be parsed into one of the two task labels.

For property prediction, unparsable outputs are not discarded when computing BA, recall, or Macro-F1. Instead, an unparsable output is treated as an incorrect prediction for its gold class and therefore contributes a false negative to that class. Metrics are computed separately for the initial output and after allowing one format-repair attempt.

Finally, the aggregate property-prediction results are macro-averaged across the five tasks:

$$\mathrm{Metric}_{\mathrm{avg}} = \frac{1}{5}\sum_{t=1}^{5}\mathrm{Metric}_t, \tag{48}$$

so that each task contributes equally irrespective of its test-set size.

D.4 MoleculeQA 10-Shot Evaluation

Evaluation Prompt. Following the aspect-specific 10-shot setting of MoleculeQA (Lu et al., 2024), for each seed, we construct a separate 10-shot support set for each of the four aspects and keep these demonstrations fixed across all compared model variants. All demonstrations are drawn from the official training split and are shared across the compared model variants. Since the released training examples do not provide explicit aspect labels, we assign them to aspects by matching their normalized question stems to those in the officially annotated test split and transferring the corresponding aspect labels. Within each aspect, we select ten examples with distinct normalized question–choice patterns, providing a reproducible proxy for the topic diversity used in the original MoleculeQA evaluation.

The system message first specifies the target aspect and its definition, followed by ten solved demonstrations. Each demonstration contains the molecular SMILES, the question, four answer choices, and the corresponding A/B/C/D answer. The user message then presents a single test instance using the same SMILES–question–choice format. Figure 4 illustrates the resulting prompt template.

The prompt follows the 10-shot construction used in the original MoleculeQA evaluation (Lu et al., 2024), while adapting the output instruction to our single-instance inference setting. The same prompt structure and output format are applied to all models evaluated under our protocol.

Output Parsing and Repair. We parse each generated response into one of the four answer choices A/B/C/D. If the initial response cannot be parsed into a valid choice, we allow exactly one format-repair attempt. Specifically, the original response is appended to the prompt together with an additional `Answer:` marker, and the model is asked to continue the response once more under the same deterministic decoding setting. The repaired continuation is then parsed using the same answer-extraction rule. If no valid choice can be extracted after this attempt, the prediction is treated as incorrect.

For completeness, we compute both the initial accuracy and the repair-once accuracy for every evaluated model. Unless otherwise noted, the results reported in the main text use the initial predictions before repair. The only exception is LlaSMol-Llama2-7B, for which we report the repair-once accuracy to better accommodate its limited instruction-following capability and frequent deviations from the required answer format.

Results across Support-Set Seeds. To assess sensitivity to the particular in-context demonstrations, we repeat the 10-shot evaluation using support sets constructed with seeds 0, 42, and 123. Table 11 reports the resulting per-seed accuracies. LlaSMol-Llama2-7B does not have outputs for seed 123, so its average reported in Table 2 is computed over the two available seeds; all other models are averaged over all three seeds.

The overall results are stable across support-set choices. In each of the three runs, ReMPA improves over Base and the One-way DPO variant, while remaining close to SFT-Full. The aspect-level results show the same general pattern as the averaged results in the main text: ReMPA's largest gain over

System: You are a chemistry research assistant, and I'd like to test your professional ability on molecule understanding. Please complete the following task.
You are provided with the SMILES representation of a molecule and asked a multiple-choice question about the molecule's **Structure** knowledge. Here, **Structure** refers to *details about the architecture, composition, and interaction of atoms within a molecule*. Three of the four options do not describe the given molecule, and you must select the correct option.

Other aspect definitions used in our evaluation are:
Source: the natural or synthetic origin, as well as the production context related to a molecule.
Property: physical, biological, or chemical properties in various environments or reactions.
Application: the utilization of a molecular compound in various applications and scenarios.

10-shot demonstrations:
Example 1:
Molecular SMILES: CCO
Question: Which statement best describes the structure of this molecule?
Choices:
A: It is a three-carbon ketone containing a carbonyl group.
B: It contains a two-carbon chain attached to a hydroxyl group.
C: It is an aromatic ring bearing a chlorine substituent.
D: It is an imine containing a C=N bond.
Answer: B
###

. . . Examples 2–10 follow the same format and are omitted for brevity. . . .

User: Please give me your choice for this instance in the above examples' style. No other information is required.
Instance ID: test-0421
Molecular SMILES: CC(=O)O
Question: Which statement best describes the structure of this molecule?
Choices:
A: It contains a carboxyl group attached to a methyl group.
B: It is a six-membered aromatic hydrocarbon.
C: It is a primary amine with two carbon atoms.
D: It is an ether with two ethyl substituents.

Assistant: Answer: A

Figure 4: Illustration of the aspect-specific 10-shot prompt used for MoleculeQA evaluation. This figure shows the *Structure* version; the prompts for the other three aspects differ only in the aspect definition in the system instruction, while keeping the same overall format.

Base occurs on Structure, with additional improvements on Source and Application, while Property performance remains close to the Base level. Thus, the molecular-comprehension gains reported in Sec. 4.2 are not driven by a particular choice of 10-shot demonstrations.

D.5 Molecular Property Prediction

Evaluation Prompts We evaluate each property-prediction task using a single SMILES-only Direct prompt, following the text-based Direct evaluation format of EDT-Former (Jing et al., 2026). The system message specifies the target molecular property and the two permitted output labels, while the user message provides the task instruction together with the raw molecular SMILES. The resulting messages are rendered using the native chat template of the evaluated model. The same task-specific prompt is used for all compared model variants.

For BBBP, we use the task-specific Direct prompt released in EDT-Former (Jing et al., 2026), with `Penetrant` and `Non-penetrant` as the two output labels. For BACE, DILI, hERG, and HIA, we retain the same Direct prompt structure and instantiate it with the corresponding task semantics

Table 11: Per-seed 10-shot accuracy (%) on MoleculeQA. All models except LlaSMol-Llama2-7B are evaluated with support-set seeds 0, 42, and 123. LlaSMol-Llama2-7B has outputs for seeds 0 and 42 only, and its results use the single format-repair attempt described above. All other entries are computed from the initial predictions.

Model	Structure	Source	Property	Application	Overall
Support-set seed = 0					
Base	38.55	39.24	25.44	31.05	36.28
LlaSMol-Llama2-7B	24.35	24.94	22.30	26.38	24.44
MolInst-Llama3-8B	28.85	26.36	24.90	27.71	27.65
SFT-50K	48.06	39.39	24.76	30.72	41.31
SFT-Full (200K)	49.18	42.81	25.44	29.55	42.67
One-way DPO	43.91	41.40	24.49	31.89	39.63
ReMPA	46.00	41.10	24.21	32.22	40.68
Support-set seed = 42					
Base	39.09	36.34	27.50	28.38	35.88
LlaSMol-Llama2-7B	23.87	25.69	23.39	24.21	24.27
MolInst-Llama3-8B	30.90	30.75	26.54	28.38	30.06
SFT-50K	47.57	40.28	27.77	30.22	41.58
SFT-Full (200K)	48.47	40.21	26.40	29.55	41.81
One-way DPO	43.94	39.46	26.95	30.22	39.34
ReMPA	46.96	39.91	27.91	29.38	41.10
Support-set seed = 123					
Base	39.16	38.64	26.68	30.55	36.57
LlaSMol-Llama2-7B	–	–	–	–	–
MolInst-Llama3-8B	28.53	25.24	25.72	29.55	27.51
SFT-50K	46.68	41.18	25.58	30.72	41.08
SFT-Full (200K)	48.54	41.77	25.44	31.72	42.31
One-way DPO	44.43	43.26	25.17	32.89	40.53
ReMPA	47.16	43.26	25.31	31.72	41.89

System: You are a drug discovery assistant tasked with predicting whether a molecule can penetrate the blood-brain barrier (BBBP). Output the final decision strictly using the provided answer format string. Your final answer should be formatted as either: 'Final answer: Penetrant' or 'Final answer: Non-penetrant'.

User: Determine the BBBP class of the given molecule.
Molecule: C1=CC=CC=C1O.

Assistant: Final answer: Penetrant

Task-specific variants:
BBBP: Penetrant / Non-penetrant. **BACE:** Active / Inactive.
DILI: Active / Inactive. **hERG:** Active / Inactive. **HIA:** Active / Inactive.

Figure 5: Illustration of the SMILES-only Direct prompt used for molecular property prediction, shown here using the BBBP prompt.

and output labels; these four tasks use `Active` and `Inactive`. Figure 5 illustrates the resulting prompt using BBBP as an example.

Output Parsing and Repair. Predictions are parsed exclusively from the generated completion. For each task, we locate the last occurrence of the `Final answer:` marker and accept the prediction only if the following text exactly matches one of the two task-specific labels (up to capitalization and minor surrounding punctuation). Outputs that do not satisfy this format are treated as unparseable.

For an unparseable initial output, we allow exactly one format-repair attempt. We append the original generated completion to the prompt, followed by a new `Final answer:` marker, and ask the model to continue generation once under the same decoding setting. The resulting continuation is parsed using the same rule. We report both the metrics computed from the original outputs and those obtained after this single repair attempt.

The two-stage evaluation separates output-format reliability from classification behavior. Across the five tasks, the SFT variants frequently fail to produce a parseable class label in their initial outputs, whereas Base, One-way DPO, and ReMPA generally retain high initial parse rates. A single format-repair attempt recovers valid outputs for most SFT predictions, indicating that a substantial part of their pre-repair degradation is associated with failure to follow the required output format. However, their class-balanced performance remains limited on several tasks after repair, with particularly low negative-class recall. We therefore report both pre- and post-repair metrics together with class-specific recall.

Per-Task Results. Table 12 reports the complete results for the five property-prediction tasks. The preference-based variants maintain initial output validity close to Base across all tasks, in contrast to the substantially lower initial parse rates of the SFT variants. After repair, the relative performance of ReMPA and One-way DPO is task-dependent: One-way DPO performs better on BBBP, BACE, and DILI, whereas ReMPA performs better on hERG and HIA. In particular, ReMPA shows strong gains on the latter two tasks while maintaining near-complete output validity. The SFT variants recover high parse rates after repair but frequently remain strongly biased toward the positive class, as reflected by their low negative-class recall on several datasets. Overall, the per-task results show that preference-based adaptation largely preserves format-following behavior.

Table 12: Per-task molecular property prediction results (%). The five test sets contain 898 examples in total. Initial metrics are computed from the original model outputs, while Final metrics are computed after allowing one format-repair attempt. Parse denotes the valid-output rate. Positive- and negative-class recall are reported after the repair stage.

	Initial			Final			Recall	
Model	BA	Macro-F1	Parse	BA	Macro-F1	Parse	Pos.	Neg.
BBBP ($N_{\text{test}} = 406$)								
Base	53.63	49.33	92.61	55.31	49.50	100.00	55.49	55.13
SFT-50K	6.53	11.39	12.32	52.64	52.47	96.80	71.95	33.33
SFT-Full (200K)	7.29	13.03	11.33	51.82	51.72	97.54	72.87	30.77
One-way DPO	56.68	54.07	89.90	62.51	56.04	99.51	62.20	62.82
ReMPA	53.17	51.07	91.63	56.40	52.40	99.51	62.80	50.00
BACE ($N_{\text{test}} = 152$)								
Base	51.01	49.25	95.39	54.82	51.26	100.00	91.30	18.33
SFT-50K	4.35	7.55	9.21	50.00	37.70	100.00	100.00	0.00
SFT-Full (200K)	3.26	6.06	4.61	50.00	37.70	100.00	100.00	0.00
One-way DPO	52.64	50.62	95.39	55.65	52.67	99.34	91.30	20.00
ReMPA	46.56	41.92	94.08	47.93	42.80	99.34	85.87	10.00
DILI ($N_{\text{test}} = 96$)								
Base	48.22	48.38	93.75	52.39	51.21	100.00	70.00	34.78
SFT-50K	3.00	5.36	6.25	50.00	34.72	97.92	100.00	0.00
SFT-Full (200K)	3.00	5.66	3.13	49.00	33.79	100.00	98.00	0.00
One-way DPO	58.13	56.80	92.71	59.22	56.13	100.00	88.00	30.43
ReMPA	49.26	39.35	96.88	51.35	41.36	100.00	94.00	8.70
hERG ($N_{\text{test}} = 132$)								
Base	46.10	40.85	96.97	48.56	42.44	100.00	37.11	60.00
SFT-50K	6.19	10.62	12.88	46.16	43.88	95.45	86.60	5.71
SFT-Full (200K)	6.07	11.03	9.85	46.55	46.02	91.67	84.54	8.57
One-way DPO	46.52	46.32	98.48	47.04	46.70	99.24	59.79	34.29
ReMPA	56.36	56.36	96.21	58.31	56.95	100.00	68.04	48.57
HIA ($N_{\text{test}} = 112$)								
Base	44.51	44.89	97.32	47.08	46.97	100.00	78.16	16.00
SFT-50K	3.72	6.93	7.14	51.98	51.93	96.43	91.95	12.00
SFT-Full (200K)	0.00	0.00	0.89	47.13	42.71	93.75	94.25	0.00
One-way DPO	55.95	57.15	97.32	57.10	57.58	100.00	86.21	28.00
ReMPA	65.08	65.49	95.54	67.38	66.63	100.00	82.76	52.00