MULTI-OBJECTIVE GENERATIVE OPTIMIZATION VIA EXPECTED HYPERVOLUME FINE-TUNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Scientific discovery requires plausible designs spanning useful trade-offs among competing properties: a molecule with strong binding affinity, for example, may be difficult to synthesize. Pretrained generative models, e.g., flow and diffusion generators, provide a foundation for proposing such designs, but steering them toward high-quality, complementary trade-offs remains challenging. Inference-time search incurs high per-sample costs, while fine-tuning typically relies on fixed per-candidate rewards that do not capture how designs complement one another. We introduce Hypervolume Fine-Tuning (HVFT), which adapts a single pretrained generator to maximize the expected hypervolume of its samples while controlling departure from the pretrained distribution. Expected hypervolume jointly rewards candidate quality and collective trade-off coverage. Nonetheless, we show that it cannot be expressed as the expectation of a fixed per-candidate reward and therefore cannot be optimized directly with standard techniques. A key observation of this work is that its first variation corresponds to the expected marginal hypervolume contribution of a candidate, which yields a tractable adaptive reward via an unbiased estimator based on model samples. HVFT uses this reward to decompose the objective into a sequence of KL-proximal scalar-reward subproblems. We establish concavity, relative smoothness, and convergence guarantees for the resulting distributional procedure. Across therapeutic-peptide and molecular design tasks spanning discrete diffusion and continuous flow-matching models, HVFT achieves state-of-the-art expected hypervolume and trade-off coverage.

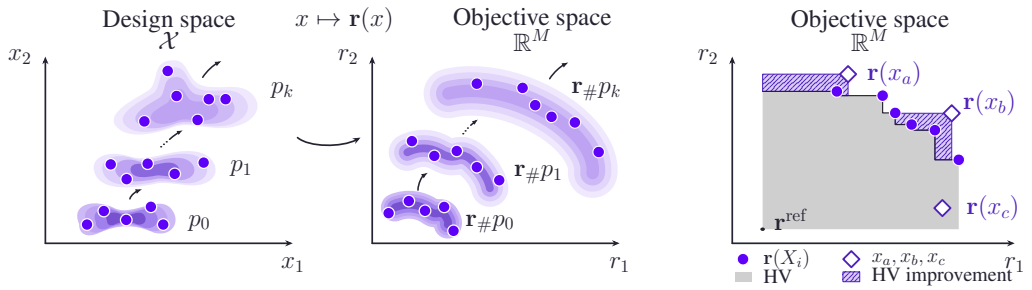
1 INTRODUCTION

Scientific discovery often requires balancing competing objectives. A peptide with high binding affinity to a protein, for example, may have low cell permeability, and vice versa (Tang et al., 2025a). The most useful compromise may remain unknown until experimental evidence or downstream preferences become available. A useful generator should therefore produce candidate sets that collectively represent high-quality, complementary trade-offs.

Pretrained generative models provide a prior over plausible designs, but their sampling distributions are not optimized for the properties of interest. Generative optimization adapts these distributions toward improved objective values (Domingo-Enrich et al., 2025; De Santi et al., 2025). In the multi-objective setting, however, improving candidates individually is not enough: samples should collectively cover complementary trade-offs. The central challenge is therefore to optimize this collective behavior, rather than a fixed score assigned to each candidate in isolation.

Existing methods provide several forms of multi-objective steering. Inference-time search and guidance can seek favorable trade-offs, but they require additional optimization whenever candidates are generated (Tang et al., 2025a; Yao et al., 2024; Annadani et al., 2025) and, without updating the model, may have limited ability to depart from the pretrained prior (Klarner et al., 2024). Fine-tuning amortizes this cost, yet its training signals introduce different limitations. Fixed scalarizations score candidates independently (Tang et al., 2025b; Mahood et al., 2026; Patel et al., 2026; Chen et al., 2025a; Lin et al., 2026); preference-conditioned and composed models require choosing a distribution over desired trade-offs (Golan et al., 2026; Cheng et al., 2025); and dominance-based signals neither distinguish mutually nondominated candidates nor quantify their contributions to collective coverage (Schneuing et al., 2025). None directly trains the sampling distribution for both candidate quality and collective trade-off coverage.

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.



(a) Sequential fine-tuning. (b) Marginal hypervolume reward.

Figure 1: **HVFT for multi-objective generative fine-tuning.** (a) HVFT iteratively constructs a scalar reward from the current distribution p_k and fine-tunes the generator, shifting the pushforward distribution $\mathbf{r}_{\#}p_k$ toward high-quality, complementary trade-offs. (b) The gray region is the hypervolume dominated relative to $\mathbf{r}^{\text{ref}}$. Hatching shows the marginal contributions of x_a and x_b, while the dominated candidate x_c contributes zero. HVFT estimates these contributions against background sets from p_k to define the next fine-tuning reward.

Motivated by this challenge, we address the following question:

> *How can we fine-tune a generative model to directly produce samples that jointly achieve high objective quality and broad trade-off coverage?*

As a response, we introduce Hypervolume Fine-Tuning (HVFT), which trains a single generative model to maximize the expected hypervolume of its generated candidate sets while controlling departure from the pretrained distribution. Hypervolume measures the objective-space region dominated by a candidate set relative to a fixed reference point (Zitzler & Thiele, 1999; Guerreiro et al., 2021). It therefore rewards both strong individual candidates and complementary trade-off coverage, without requiring a distribution over preferences.

Unlike standard fine-tuning objectives, expected hypervolume cannot be expressed as the expectation of a fixed per-candidate reward. A candidate's value depends on the objective-space regions covered by the other samples, so its reward must adapt to the current sampling distribution (Domingo-Enrich et al., 2025; Zheng et al., 2025; Zhu et al., 2025). HVFT addresses this limitation using Flow Density Control (De Santi et al., 2025), which optimizes a distributional objective through a sequence of KL-proximal subproblems driven by its first variation (see Figure 1). A key observation is that, for expected hypervolume, this first variation is the expected marginal hypervolume contribution of a candidate relative to samples from the current model. This identity gives the reward a clear interpretation and makes it tractable to estimate directly from model samples. At each outer iteration, HVFT estimates this adaptive reward under the current distribution, uses it with a model-specific scalar-reward solver, and recomputes it after the update. We further establish concavity and relative smoothness of the objective, yielding monotone improvement and convergence guarantees for this procedure.

Across therapeutic-peptide (Tang et al., 2025a;b) and molecular design (Schneuing et al., 2025) tasks spanning discrete diffusion and continuous flow-matching models, HVFT achieves state-of-the-art expected hypervolume and trade-off coverage without additional inference-time search.

Our main contributions are as follows:

- We introduce expected hypervolume as a distributional training criterion for multi-objective generative fine-tuning, jointly rewarding individual quality and collective trade-off coverage.
- We express the first variation of expected hypervolume in terms of marginal hypervolume contributions and derive an unbiased sample-based estimator. This yields HVFT, a sequence of scalar-reward fine-tuning subproblems for which we establish convergence guarantees.
- We demonstrate state-of-the-art performance on therapeutic-peptide and molecular design tasks spanning discrete diffusion and continuous flow-matching models, improving the expected hypervolume of sampled sets without additional inference-time search.

2 RELATED WORK

We first discuss existing methods for multi-objective steering of pretrained generators and then review the methodological foundations of HVFT. We conclude by briefly discussing adjacent multi-objective optimization settings that target different problems.

Multi-objective steering of pretrained generators. Methods for steering pretrained generators toward multiple objectives differ primarily in the information carried by their steering signal.

Ordinal Pareto signals. Pareto dominance guides inference-time search in PepTune (Tang et al., 2025a), replay-buffer construction and off-policy fine-tuning in TR2D2 and its extensions (Tang et al., 2025b; Mahood et al., 2026; Patel et al., 2026), and the selection of preference pairs for direct preference optimization (Schneuing et al., 2025). Dominance records which candidates are preferable but not by how much: it is invariant to strictly increasing transformations of each objective, cannot order mutually nondominated candidates, and does not quantify their contributions to a candidate set.

Scalarized per-candidate rewards. A second family maps each objective vector to a scalar utility and applies standard reward-based fine-tuning, sometimes incorporating thresholds or predictive uncertainty (Lin et al., 2026; Chen et al., 2025a). TR2-D2-based methods, for example, sum the objectives of candidates in their nondominated replay buffer (Tang et al., 2025b; Mahood et al., 2026; Patel et al., 2026). Because a fixed scalarization scores candidates independently and generally favors a particular trade-off, repeated samples need not cover the front.

Preference-conditioned generation. A third family makes the trade-off an input by conditioning on a preference vector during training (Golan et al., 2026) or inference (Chen et al., 2025b), or by composing separately aligned models at sampling time (Cheng et al., 2025). These methods provide precise control when the desired trade-off is known. Producing a set of alternatives instead requires choosing a distribution over preferences, such as a grid or a Das–Dennis construction. This choice becomes harder as the number of objectives grows and does not directly optimize the coverage of the resulting samples. In Section 4, we evaluate DiffusionBlend by sampling preference vectors.

Positioning of HVFT. The preceding approaches rely on ordinal comparisons, fixed per-candidate utilities, or externally specified preferences. HVFT instead introduces a cardinal set-level training signal: each candidate is valued by the hypervolume it contributes relative to samples from the current model. This signal distinguishes mutually nondominated candidates, accounts for the trade-offs already covered by the generator, and requires no specified preference distribution.

Methodological foundations. We discuss generators and hypervolume as a set-level objective. We further describe the distributional optimization framework that connects them.

Reward-based generative fine-tuning. Pretrained diffusion, flow-matching, and discrete-diffusion models provide priors over plausible molecular structures and biological sequences (Ho et al., 2020; Lipman et al., 2023; Lou et al., 2024; Tang et al., 2025a; Dunn & Koes, 2026). Reward-based methods adapt these models toward target properties (Domingo-Enrich et al., 2025; Liu et al., 2025; Zheng et al., 2025). HVFT likewise adapts a pretrained model using a compatible reward-based method as the inner solver within a sequential fine-tuning procedure, with the marginal hypervolume reward recomputed as the sampling distribution changes.

Hypervolume as an optimization signal. Hypervolume measures the region of objective space dominated by a candidate set relative to a reference point (Zitzler & Thiele, 1999; Guerreiro et al., 2021); its diminishing-returns structure relates it to submodular coverage (Nemhauser et al., 1978; Krause & Golovin, 2014). Beyond evaluation, hypervolume has been used in Bayesian optimization, evolutionary selection, Pareto-set learning, and reinforcement-learning (Daulton et al., 2020; Beume et al., 2007; Bader & Zitzler, 2011; Zhang et al., 2023; Lu et al., 2025; Steinberg et al., 2025). These methods optimize the hypervolume of a finite set selected or maintained by the algorithm. HVFT instead uses the expected hypervolume of samples from the generator as its steering signal.

Distributional optimization of non-additive objectives. Set-level objectives make each candidate's value depend on the others. Related work captures these interactions through marginal-gain estimators and successive additive surrogates (Prajapat et al., 2024; De Santi et al., 2024), or optimizes nonlinear functionals of induced distributions using adaptive rewards (Hazan et al., 2019; Zahavy et al., 2021). Flow Density Control provides a general framework for optimizing nonlinear functionals of generative

distributions: at each outer iteration, it evaluates the first variation at the current sampling distribution and uses it as a scalar reward in a KL-proximal fine-tuning subproblem (De Santi et al., 2025). HVFT uses this framework to optimize expected hypervolume through sequential scalar-reward fine-tuning. To make this possible, we derive the marginal hypervolume first variation, construct an unbiased sample-based estimator, and establish a relative-smoothness bound that controls the step size.

Adjacent multi-objective optimization settings. Several related formulations differ in their evaluation access, prior information, and outputs.

Multi-objective Bayesian optimization methods treat objective evaluations as scarce and repeatedly select batches that improve the hypervolume of an observed archive (Daulton et al., 2020; Stanton et al., 2022; Gruver et al., 2023; Jain et al., 2023; Steinberg et al., 2025). HVFT instead assumes objectives can be queried at training scale and returns a sampler that can subsequently generate candidates without re-solving the optimization problem. *Pareto-set approximation* methods return an explicit approximation of the front or a map from preferences to solutions, using, for example, hypervolume-based survival selection (Beume et al., 2007; Bader & Zitzler, 2011) or learned Pareto-set representations (Lin et al., 2022; Zhang et al., 2023). They do not use a pretrained distribution as a prior over plausible designs. *Offline multi-objective optimization* methods work from a fixed dataset without further objective queries and steer generation using learned predictors or dominance classifiers (Yuan et al., 2025; Annadani et al., 2025). HVFT instead queries the objectives during training and uses only their evaluated values.

3 Hypervolume Fine-Tuning

Let $\mathcal{X}$ denote a design space and let $\mathbf{r}(x) = (r_1(x), \dots, r_M(x)) \in \mathbb{R}^M$ denote a vector of competing objectives, all taken to be maximized. We assume that these objectives can be evaluated on generated candidates, but they may be black-box and nondifferentiable. Let p_0 denote a pretrained generative distribution over $\mathcal{X}$. We seek to adapt this distribution so that its samples collectively achieve high objective quality and broad trade-off coverage while preserving useful behavior learned during pretraining.

Achieving this goal requires a distributional objective that values candidate sets and a way to optimize that objective using generative fine-tuning. The remainder of this section develops these two ingredients. We first define expected hypervolume as a distributional objective (Section 3.1) and then derive its first variation, which is an expected marginal hypervolume contribution (Section 3.2). This local reward leads to the sequential fine-tuning procedure in Section 3.3.

3.1 Expected hypervolume as a distributional objective

Evaluating samples collectively requires accounting for both objective quality and complementary coverage. Hypervolume provides such a measure: relative to a reference point $\mathbf{r}^{\mathrm{ref}} \in \mathbb{R}^M$, it quantifies the volume of objective space jointly dominated by a collection of candidates (Zitzler & Thiele, 1999; Guerreiro et al., 2021). Unlike a fixed scalarization, hypervolume does not reduce the objectives to a prescribed per-candidate score; unlike Pareto dominance alone, it quantifies how much each candidate contributes to collective coverage. The reference point and objective scaling determine the region and relative volumes being measured, and we report these choices for each task in Appendix C.2.

Writing $u \succeq v$ when $u_m \geq v_m$ for every objective m, we define the hypervolume of a set $A \subseteq \mathcal{X}$ as

$$\mathrm{HV}(A; \mathbf{r}^{\mathrm{ref}}) := \mathrm{Leb}_M \left(\bigcup_{x \in A} \left\{ u \in \mathbb{R}^M : \mathbf{r}^{\mathrm{ref}} \preceq u \preceq \mathbf{r}(x) \right\} \right), \tag{1}$$

where Leb_M denotes M-dimensional Lebesgue measure and $\mathrm{HV}(\varnothing; \mathbf{r}^{\mathrm{ref}}) = 0$. A candidate contributes only the portion of its dominated region not already covered by the others. Hypervolume therefore rewards stronger objective values, which expand the dominated region, and complementary trade-offs, which extend that region in previously uncovered directions. See Figure 1b for an illustration of hypervolume.

We lift hypervolume to a distributional training objective by taking its expectation over n independent samples: [1]

$$F_n(p) := \mathbb{E}_{X_{1:n}\sim p^{\otimes n}}\left[\mathrm{HV}\big(\{X_1,\dots,X_n\};\mathbf{r}^{\mathrm{ref}}\big)\right]. \tag{2}$$

The parameter n determines how many samples interact within the training objective rather than how many samples may subsequently be generated. Empirically, we find that training with modest values of n improves hypervolume even when evaluating larger numbers of samples (Section 4).

To limit departure from the pretrained model, we penalize deviations from the pretrained distribution. Let $\mathcal{P}(\mathcal{X})$ denote the set of probability distributions on $\mathcal{X}$. Our optimization target is

$$p^\star \in \arg\max_{p\in\mathcal{P}(\mathcal{X})}\left\{F_n(p) - \alpha D_{\mathrm{KL}}(p\|p_0)\right\}, \tag{3}$$

where $\alpha \geq 0$ balances hypervolume performance against fidelity to the pretrained distribution.

The optimization in Equation (3) is not a standard fixed-reward fine-tuning problem (Domingo-Enrich et al., 2025; Liu et al., 2025). When $n > 1$, the value of a candidate depends on the other $n-1$ samples and hence on the sampling distribution itself. Thus, F_n is generally nonlinear in p and cannot be written as the expectation of a fixed per-candidate reward.

HVFT handles this nonadditivity locally. At iteration k, we linearize the objective around the current distribution p_k. Its first variation assigns each candidate a scalar value equal to its first-order effect on the objective. We optimize this local reward within a KL-controlled neighborhood of p_k to obtain p_{k+1}, recompute the reward at the new distribution, and repeat. The next subsection derives this reward; Section 3.3 then uses it to define the sequence of fine-tuning subproblems.

3.2 First variation and the marginal hypervolume reward

We first derive the scalar reward that locally represents changes in F_n. Following the distributional optimization framework of De Santi et al. (2025), a first variation of F_n at p is a function $f_{n,p}$ such that, for every admissible distribution q,

$$F_n((1-\epsilon)p+\epsilon q) = F_n(p) + \epsilon\left(\mathbb{E}_q[f_{n,p}(X)] - \mathbb{E}_p[f_{n,p}(X)]\right) + o(\epsilon), \qquad \epsilon \downarrow 0. \tag{4}$$

At the current iterate $p = p_k$, both $F_n(p_k)$ and $\mathbb{E}_{p_k}[f_{n,p_k}(X)]$ are constant with respect to q. Maximizing the linear approximation over q therefore reduces to maximizing $\mathbb{E}_q[f_{n,p_k}(X)]$, which has the additive form required by a scalar-reward fine-tuning solver.

For expected hypervolume, this local reward has a direct interpretation. Define the marginal hypervolume contribution of a candidate $x \in \mathcal{X}$ relative to a background set $A \subseteq \mathcal{X}$ as

$$\Delta_{\mathrm{HV}}(x;A) := \mathrm{HV}(A\cup\{x\};\mathbf{r}^{\mathrm{ref}}) - \mathrm{HV}(A;\mathbf{r}^{\mathrm{ref}}). \tag{5}$$

Proposition 1 (Marginal hypervolume reward)**.** *A first variation of F_n at p is given by*

$$f_{n,p}(x) = n\,\mathbb{E}_{A\sim p^{\otimes(n-1)}}\left[\Delta_{\mathrm{HV}}(x;A)\right], \tag{6}$$

where A is formed from $n-1$ independent draws from p.

Equation 6 rewards a candidate for the objective space it covers beyond a background set drawn from the current model (see Figure 1b). The reward is large when the candidate has strong objective values and adds a trade-off that current samples are unlikely to cover. It is not a fixed scalarization of $\mathbf{r}(x)$: the background distribution changes with p, so the reward changes as the model improves. Candidate interactions are therefore retained in the outer loop even though the inner solver receives a scalar reward. The proof of Proposition 1 is given in Appendix A.

3.3 Sequential fine-tuning via marginal hypervolume rewards

Flow Density Control (De Santi et al., 2025) turns the local reward into a distributional update. Starting from the pretrained distribution p_0, we linearize both terms of the penalized objective at p_k and constrain the update with a KL-proximal penalty. The k-th subproblem is

$$p_{k+1} \in \arg\max_p\left\{\mathbb{E}_{X\sim p}\left[f_{n,p_k}(X) - \alpha\log\frac{p_k(X)}{p_0(X)}\right] - \frac{1}{\eta_k}\mathrm{KL}(p\|p_k)\right\}, \tag{7}$$

[1] Throughout the analysis, we assume that $(r_m(x) - r_m^{\mathrm{ref}})_+ \leq C$ for some finite $C > 0$ and all x, m. This bounds the hypervolume by C^M and ensures that the expectation is finite.

Algorithm 1 Hypervolume Fine-Tuning (HVFT)

Require: Pretrained distribution p_0, objective map $\mathbf{r}$, reference point $\mathbf{r}^{\text{ref}}$, set size $n \geq 1$, background count $B \geq 1$, penalty coefficient $\alpha \geq 0$, maximum iterations K, step sizes $\eta_k > 0$

1: **for** $k = 0, \ldots, K-1$ **do**
2: Draw independent background sets $A_b \sim p_k^{\otimes(n-1)}$, $b = 1, \ldots, B$
3: Evaluate and cache $\mathbf{r}(a)$ for each background design $a \in A_b$
4: Construct $\widehat{f}_{n,p_k}$ using Equation (8)
5: Approximately solve Equation (7) with f_{n,p_k} replaced by $\widehat{f}_{n,p_k}$
6: **end for**
7: **return** p_K

The first variation of $p \mapsto D_{\text{KL}}(p\|p_0)$ at p_k is $\log(p_k(x)/p_0(x)) + 1$. The additive constant does not affect the optimizer and is omitted in Equation (7). Therefore, f_{n,p_k} is the expected-hypervolume component of the linearized reward, and the log-density ratio is the pretrained-model penalty component. The final term, $\eta_k^{-1} D_{\text{KL}}(p\|p_k)$, is not part of the original objective but it is included to limit how far the solution of the local subproblem can move from p_k.

The linearized reward remains fixed during the k-th inner solve. Its solution defines p_{k+1}, at which point we recompute both reward components to form the next subproblem. Thus, the outer loop updates the local reward, while the inner loop solves a standard scalar-reward fine-tuning problem.

Theoretical guarantee. The key question is whether progress on successive local subproblems also improves the original penalized objective. As shown in Appendix A.2, expected hypervolume is concave in the sampling distribution, and the pretrained-model KL penalty makes the objective strictly concave when $\alpha > 0$ (Appendix A.3). More importantly, relative smoothness bounds the error of the linear approximation by a multiple of $D_{\text{KL}}(p_{k+1}\|p_k)$ (Appendix A.4). A sufficiently strong KL-proximal penalty dominates this error, linking progress on each local subproblem to progress on the original objective. The following theorem formalizes this guarantee.

Theorem 1 (Convergence of HVFT)**.** *Under the mild regularity conditions, the penalized objective in Equation (3) admits a unique maximizer $p^\star$ for $\alpha > 0$. For any sufficiently small constant step size η, the iterates in Equation (7) are well defined, monotonically improve the penalized objective, and converge geometrically to $p^\star$. Precise conditions and rates are given in Appendix A.5.*

Practical implementation. The expectation in Equation (6) is approximated by Monte Carlo. At iteration k, we draw independent background sets $A_b \sim p_k^{\otimes(n-1)}$, for $b = 1, \ldots, B$, and define

$$\widehat{f}_{n,p_k}(x) := \frac{n}{B} \sum_{b=1}^{B} \Delta_{\text{HV}}(x; A_b). \tag{8}$$

For every fixed x, this estimator is unbiased: $\mathbb{E}[\widehat{f}_{n,p_k}(x)] = f_{n,p_k}(x)$.

During the k-th inner solve, we hold the background sets and their objective values fixed, making $\widehat{f}_{n,p_k}$ a fixed empirical reward. A model-specific scalar-reward method approximately solves the subproblem over the generator parameters. After the outer iteration, we draw new background sets from p_{k+1} and construct the next reward. Our experiments in Section 4 instantiate the inner solver with methods for continuous flow-matching and discrete diffusion models (Zheng et al., 2025; Tang et al., 2025b). Appendix B specifies the corresponding training losses and treatment of local regularization. Algorithm 1 summarizes the procedure.

4 Experiments

We evaluate HVFT on therapeutic-peptide and small-molecule design tasks using discrete diffusion and continuous flow-matching models. Our experiments test whether HVFT improves objective-space quality and trade-off coverage while preserving validity and diversity in the design space.

Evaluation protocol. At each evaluation, we generate candidates until obtaining 1,000 valid designs and record the valid fraction over all generation attempts. From this pool, we repeatedly draw random subsets of 10 and 100 observations without replacement and report their mean hypervolume, while computing diversity and design-space coverage over the full pool. We evaluate 100-hypervolume at regular intervals and use it to select the best-performing epoch, applying the same rule to baselines wherever possible. Across all tasks, HVFT uses $n = 10$; we study the effect of n in Appendix D. We repeat each experiment with three random seeds, corresponding to independent fine-tuning runs for training-based methods, and report means and standard errors across runs. This across-run aggregation is distinct from the repeated subset sampling used to estimate each run's hypervolume. Complete metric definitions and evaluation details are provided in Appendix C.

4.1 THERAPEUTIC PEPTIDES

Peptides are molecules composed of amino-acid residues linked by peptide bonds and are an established class of therapeutic agents (Wang et al., 2022). We fine-tune the pretrained peptide masked diffusion model of Tang et al. (2025a) for five objectives: binding affinity, solubility, non-hemolysis, non-fouling, and membrane permeability. All objectives are evaluated using the classifiers and regressors from the same work and are taken to be maximized.

We consider three binding targets: the transferrin receptor (TfR), a common target for organ-specific drug delivery (Han et al., 2025); the glucagon-like peptide-1 receptor (GLP-1R), which is relevant to type 2 diabetes and obesity (Alfaris et al., 2024); and the glutamate-aspartate transporter (GLAST), which is abundant on astrocytes and implicated in neurological disorders (Pajarillo et al., 2019). Only peptides that pass the `SMILES2PEPTIDE` verifier (Tang et al., 2025a) are scored.

Baselines. Our primary fine-tuning baseline is TR2D2 (Tang et al., 2025b), a state-of-the-art method for multi-objective fine-tuning of discrete diffusion models. We also include TR2D2 ++, an enhanced variant in which rewards are normalized to $[0, 1]$ to avoid scale bias and the best epoch is selected by hypervolume. For additional context, we evaluate the pretrained model and PEPTUNE (Tang et al., 2025a), an inference-time method that guides generation using Monte Carlo tree search.

HVFT configuration. We use the single-objective variant of TR2D2 as HVFT's inner solver. We leave the marginal hypervolume rewards unscaled because scaling gave only minor improvements and requires a problem-dependent factor. See Appendix C.2 for details.

Results. As shown in Table 1, HVFT achieves the highest 10- and 100-hypervolume on all three targets. It also yields the highest design-space diversity and coverage. Figure 2 illustrates these gains on GLAST, where HVFT produces a Pareto set with broader objective-space coverage and stronger individual candidates than TR2D2.

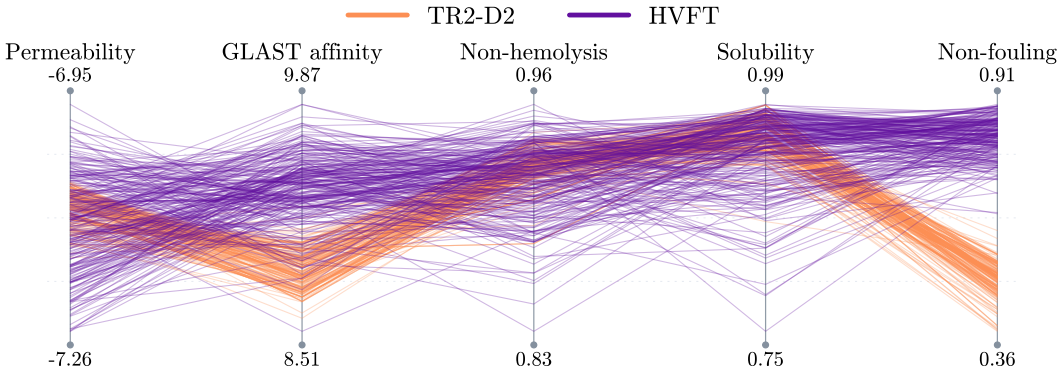


Figure 2: Pareto sets induced by 1,000 samples from TR2D2 (Tang et al., 2025b), a state-of-the-art discrete diffusion fine-tuning algorithm, and HVFT, on the GLAST task. HVFT achieves broader objective-space coverage and higher individual sample quality than TR2D2.

Table 1: Objective-space and design-space metrics for three therapeutic peptide design tasks. Bold and underlined values indicate the highest and second-highest means, within each task. HVFT achieves the highest hypervolume across the three tasks, as well as the highest design-space diversity and coverage.

		Objective-space metrics		Design-space metrics		
Task	Method	100-Hypervolume	10-Hypervolume	Diversity	Coverage	Valid fraction
GLAST	Base	2.78 ± 0.04	1.61 ± 0.01	$\underline{60.35 \pm 0.14}$	$\underline{401.50 \pm 0.28}$	0.321 ± 0.007
	PepTune	2.79 ± 0.07	1.59 ± 0.03	39.78 ± 1.05	134.95 ± 5.87	$\mathbf{0.603 \pm 0.028}$
	TR2D2	7.78 ± 1.06	6.79 ± 1.41	9.08 ± 6.24	136.64 ± 74.24	0.249 ± 0.031
	TR2D2 ++	$\underline{8.56 \pm 0.86}$	$\underline{7.68 \pm 0.77}$	4.13 ± 1.28	89.43 ± 28.69	0.233 ± 0.010
	HVFT	$\mathbf{11.67 \pm 0.27}$	$\mathbf{10.56 \pm 0.25}$	$\mathbf{64.78 \pm 7.34}$	$\mathbf{465.45 \pm 15.71}$	$\underline{0.336 \pm 0.024}$
GLP-1R	Base	3.01 ± 0.04	1.77 ± 0.01	$\underline{60.35 \pm 0.14}$	$\underline{401.50 \pm 0.28}$	0.321 ± 0.007
	PepTune	3.02 ± 0.11	1.77 ± 0.06	40.46 ± 1.06	137.70 ± 6.58	$\mathbf{0.614 \pm 0.012}$
	TR2D2	7.07 ± 0.19	5.64 ± 0.12	13.24 ± 8.00	182.56 ± 77.61	0.167 ± 0.031
	TR2D2 ++	$\underline{8.93 \pm 0.98}$	$\underline{8.12 \pm 0.83}$	3.59 ± 0.82	75.66 ± 21.14	$\underline{0.349 \pm 0.060}$
	HVFT	$\mathbf{11.23 \pm 0.08}$	$\mathbf{10.34 \pm 0.05}$	$\mathbf{93.41 \pm 2.14}$	$\mathbf{521.27 \pm 5.74}$	0.243 ± 0.071
TfR	Base	2.93 ± 0.04	1.69 ± 0.01	$\underline{60.35 \pm 0.14}$	$\underline{401.50 \pm 0.28}$	0.321 ± 0.007
	PepTune	2.82 ± 0.18	1.65 ± 0.09	39.97 ± 0.78	136.73 ± 4.67	$\mathbf{0.551 \pm 0.059}$
	TR2D2	4.41 ± 0.91	2.78 ± 1.14	6.46 ± 1.63	126.79 ± 34.38	0.200 ± 0.015
	TR2D2 ++	$\underline{7.63 \pm 1.87}$	$\underline{6.75 \pm 1.90}$	4.91 ± 1.45	99.92 ± 25.29	$\underline{0.331 \pm 0.068}$
	HVFT	$\mathbf{11.57 \pm 0.16}$	$\mathbf{10.15 \pm 0.18}$	$\mathbf{75.89 \pm 9.42}$	$\mathbf{485.80 \pm 17.27}$	0.250 ± 0.021

4.2 Small molecules

The vast space of drug-like molecules makes exhaustive screening infeasible, motivating generative models that explore chemical space and propose compounds with desirable properties (Gómez-Bombarelli et al., 2016). We fine-tune FlowMol (Dunn & Koes, 2026), a continuous flow-matching model trained on GEOM-Drugs (Axelrod & Gómez-Bombarelli, 2020), and evaluate two nested tasks: *Topology*, which uses graph-level objectives and validity checks, and *Geometry*, which adds conformation-level objectives and 3D validity checks.

For Topology, we maximize Quantitative Estimate of Drug-likeness (QED) (Bickerton et al., 2012) and synthetic accessibility (SA) (Ertl & Schuffenhauer, 2009), as implemented in RDKit (Landrum et al., 2026). We reverse and rescale the original SA score to $[0, 1]$ so that higher values are better. Only molecules that pass the chemical validity checks of Buttenschoen et al. (2025) are scored.

Geometry includes the Topology objectives and additionally maximizes the dipole moment computed at the GFN2-xTB level of theory (Bannwarth et al., 2019). Scored molecules must also pass the PoseBusters checks for bond lengths, bond angles, planar aromatic rings, planar double bonds, and internal steric clashes (Buttenschoen et al., 2024).

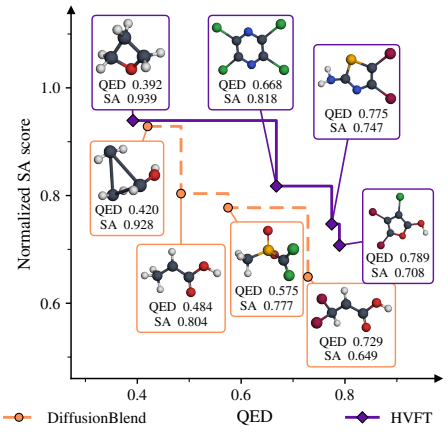


Figure 3: Best four-point subset by hypervolume, selected from 100 samples of DiffusionBlend and HVFT on the *Topology* task. HVFT's samples achieve better coverage of high-quality trade-offs in objective-space.

Baselines. We compare against DiffusionNFT (Zheng et al., 2025), a sample-efficient method that we apply to a sum scalarization of the objectives, and DiffusionBlend (Cheng et al., 2025), a state-of-the-art multi-objective method that combines one separately aligned model per objective during sampling. Although its original implementation uses low-rank adapters, we fully fine-tune each model with DiffusionNFT to avoid restricting model capacity. FlowMol is reported for additional context.

HVFT configuration. We use DiffusionNFT as HVFT's scalar inner solver and keep its hyperparameters fixed across HVFT and all fine-tuning baselines. Details are provided in Appendix C.2.

Table 2: Objective- and design-space metrics for the small-molecule design tasks. Bold and underlined values indicate the highest and second-highest means, respectively, within each task. HVFT achieves the highest hypervolume among state-of-the-art methods while preserving diversity and coverage.

		Objective-space metrics		Design-space metrics		
Task	Method	100-Hypervolume	10-Hypervolume	Diversity	Coverage	Valid fraction
Topology	FlowMol	3.116 ± 0.014	2.665 ± 0.007	$\mathbf{266.1 \pm 4.9}$	$\mathbf{507.1 \pm 6.4}$	0.705 ± 0.008
	DIFFUSIONNFT	3.140 ± 0.004	$\underline{2.754 \pm 0.005}$	128.8 ± 2.7	312.8 ± 8.7	$\mathbf{0.952 \pm 0.009}$
	DIFFUSIONBLEND	$\underline{3.337 \pm 0.024}$	2.750 ± 0.031	150.3 ± 18.8	345.8 ± 30.6	0.674 ± 0.132
	HVFT	$\mathbf{3.416 \pm 0.005}$	$\mathbf{3.247 \pm 0.004}$	$\underline{154.9 \pm 9.0}$	$\underline{351.6 \pm 11.1}$	$\underline{0.734 \pm 0.007}$
Geometry	FlowMol	4.279 ± 0.024	3.655 ± 0.010	$\mathbf{91.4 \pm 2.3}$	$\mathbf{331.6 \pm 7.0}$	0.483 ± 0.009
	DIFFUSIONNFT	$\underline{4.335 \pm 0.029}$	3.669 ± 0.018	70.9 ± 1.6	$\underline{269.4 \pm 6.8}$	$\underline{0.663 \pm 0.007}$
	DIFFUSIONBLEND	4.332 ± 0.066	$\underline{3.670 \pm 0.034}$	51.6 ± 1.9	187.4 ± 2.2	$\mathbf{0.707 \pm 0.043}$
	HVFT	$\mathbf{4.988 \pm 0.049}$	$\mathbf{4.151 \pm 0.024}$	$\underline{87.5 \pm 2.2}$	249.8 ± 1.9	0.518 ± 0.017

Results. Table 2 shows that HVFT achieves the highest 10- and 100-hypervolume on both tasks. It also preserves substantially more design-space diversity than the other fine-tuning methods. Figure 3 compares representative four-candidate subsets on Topology.

4.3 DISCUSSION

HVFT achieves the highest 10- and 100-hypervolume on every task, showing consistent gains in candidate quality and collective trade-off coverage across generative modalities and model architectures. Crucially, HVFT builds on established methods as inner solvers: single-objective TR2D2 for peptides and DIFFUSIONNFT for molecules. Nevertheless, it outperforms TR2D2 and TR2D2 ++ on the peptide tasks and DiffusionNFT on the molecular tasks, supporting the benefit of the sequential marginal-hypervolume rewards. HVFT also outperforms DiffusionBlend and achieves these gains without manually specified scalarization weights or preference allocation across objectives.

Beyond objective-space performance, HVFT also performs strongly in the design space. Among steering methods, it attains the highest diversity on every task and the highest coverage on all but Geometry. In the peptide experiments, HVFT exceeds even the pretrained model in both metrics while maintaining a comparable valid fraction. One possible explanation is that distinct objective trade-offs may correspond to widely separated regions of the design space, so seeking objective-space coverage can encourage broader design-space exploration. Finally, the PEPTUNE comparison places these fine-tuning gains in the context of inference-time steering. The inference-time baseline achieves substantially lower hypervolume than the training-based methods on the peptide tasks. This pattern is consistent with evidence that inference-time steering can have limited ability to move generation beyond regions favored by the pretrained model (Klarner et al., 2024; Steinberg et al., 2025; Santi et al., 2026).

Despite strong performance, HVFT has limitations: it relies on a bi-level optimization structure and introduces hyperparameters (regularization α, Monte-Carlo sample size B, set-size n; see Appendices B and D). Additionally, hypervolume is expensive to compute in many-objective settings. Monte-Carlo estimation can help reduce this cost, and extending HVFT to address many-objective scalability remains a direction for future work.

5 CONCLUSION

Multi-objective generative design requires samples that are individually valuable and collectively span useful trade-offs, yet standard fine-tuning relies on fixed per-candidate rewards that do not capture this collective behavior. We introduced Hypervolume Fine-Tuning (HVFT), which adapts a pretrained generator to maximize expected hypervolume while regularizing toward the pretrained distribution. Its key observation is that the first variation of expected hypervolume is given by the expected marginal contribution of a candidate and can be estimated directly from model samples, yielding scalar-reward subproblems for which we establish convergence guarantees.

Across therapeutic-peptide and small-molecule design tasks, HVFT achieves the highest 10- and 100-hypervolume with both discrete diffusion and continuous flow-matching models. It outperforms existing multi-objective methods, including standalone versions of its inner solvers, while preserving strong design-space diversity and coverage. Together, these results establish HVFT as a principled and tractable approach to fine-tuning generative models for high-quality, complementary designs without predefined scalarization weights or additional inference-time optimization.

Reproducibility statement

The objective, first variation, estimator, and proofs are given in Section 3 and Appendix A. Section 4 and Appendix C specify the evaluation rules and implementation choices. Code will be released upon acceptance of this work.

AI use statement

In this work, we used generative AI tools to identify related work, polish the writing, implement our algorithms in code, and assist with mathematical proofs. All AI-assisted code was carefully checked and experimentally validated. All AI-assisted proofs were independently verified. All literature retrieved through AI was checked against the original sources. We take full responsibility for the final content of this work, including all text, claims, and artifacts.

References

Nasreen Alfaris, Stephanie Waldrop, Veronica Johnson, Brunna Boaventura, Karla Kendrick, and Fatima Cody Stanford. GLP-1 single, dual, and triple receptor agonists for treating type 2 diabetes and obesity: a narrative review. *eClinicalMedicine*, 75:102782, 2024. doi: 10.1016/j.eclinm.2024.102782.

Yashas Annadani, Syrine Belakaria, Stefano Ermon, Stefan Bauer, and Barbara E. Engelhardt. Preference-guided diffusion for multi-objective offline optimization. In *Advances in Neural Information Processing Systems*, 2025.

Simon Axelrod and Rafael Gómez-Bombarelli. Geom: Energy-annotated molecular conformations for property prediction and molecular generation. *ArXiv*, abs/2006.05531, 2020. URL https://api.semanticscholar.org/CorpusID:219558923.

Johannes Bader and Eckart Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, 2011. doi: 10.1162/EVCO_a_00009. URL https://doi.org/10.1162/EVCO_a_00009.

Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of Chemical Theory and Computation*, 15 (3):1652–1671, 2019. doi: 10.1021/acs.jctc.8b01176.

Nicola Beume, Boris Naujoks, and Michael Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653–1669, 2007. doi: 10.1016/j.ejor.2006.08.008. URL https://doi.org/10.1016/j.ejor.2006.08.008.

G. Richard Bickerton, Gaia V. Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L. Hopkins. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2):90–98, 2012. doi: 10.1038/nchem.1243.

Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. Posebusters: Ai-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15(9):3130–3139, 03 2024. ISSN 2041-6520. doi: 10.1039/d3sc04185a. URL https://doi.org/10.1039/d3sc04185a.

Martin Buttenschoen, Yael Ziv, Garrett M. Morris, and Charlotte M. Deane. An evaluation of unconditional 3d molecular generation methods. 2025. URL https://api.semanticscholar.org/CorpusID:278237274.

Lianghong Chen, Dongkyu Kim, Mike Domaratzki, and Ping-Zhao Hu. Uncertainty-aware multi-objective reinforcement learning-guided diffusion models for 3d de novo molecular design. *ArXiv*, abs/2510.21153, 2025a. URL https://api.semanticscholar.org/CorpusID:282384440.

Tong Chen, Yinuo Zhang, Sophia Tang, and Pranam Chatterjee. Multi-objective-guided discrete flow matching for controllable biological sequence design. *ArXiv*, abs/2505.07086, 2025b. URL https://api.semanticscholar.org/CorpusID:278502057.

Mingrui Cheng, Fatemeh Doudi, Dileep M. Kalathil, Mohammad Ghavamzadeh, and Panganamala Ramana Kumar. Diffusion blend: Inference-time multi-preference alignment for diffusion models. *ArXiv*, abs/2505.18547, 2025. URL https://api.semanticscholar.org/CorpusID:278905652.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Riccardo De Santi, Manish Prajapat, and Andreas Krause. Global reinforcement learning: Beyond linear and convex rewards via submodular semi-gradient methods. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 10235–10266. PMLR, 2024. URL https://proceedings.mlr.press/v235/de-santi24b.html.

Riccardo De Santi, Marin Vlastelica, Ya-Ping Hsieh, Zebang Shen, Niao He, and Andreas Krause. Flow density control: Generative optimization beyond entropy-regularized fine-tuning. In *Advances in Neural Information Processing Systems*, 2025. Spotlight.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *International Conference on Learning Representations*, 2025.

Ian Dunn and David R. Koes. Flowmol3: flow matching for 3d de novo small-molecule generation. *Digital Discovery*, 5(5):2052–2066, 05 2026. ISSN 2635-098X. doi: 10.1039/d5dd00363f. URL https://doi.org/10.1039/d5dd00363f.

P Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1:8, 06 2009. doi: 10.1186/1758-2946-1-8.

Dan Friedman and Adji Bousso Dieng. The vendi score: A diversity evaluation metric for machine learning. *Trans. Mach. Learn. Res.*, 2023, 2022. URL https://api.semanticscholar.org/CorpusID:252715476.

Shelly Golan, Michael Finkelson, Ariel Bereslavsky, Yotam Nitzan, and Or Patashnik. ParetoSlider: Diffusion models post-training for continuous reward control. *arXiv preprint arXiv:2604.20816*, 2026. doi: 10.48550/arXiv.2604.20816. URL https://arxiv.org/abs/2604.20816.

Rafael Gómez-Bombarelli, David Kristjanson Duvenaud, José Miguel Hernández-Lobato, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4:268 – 276, 2016. URL https://api.semanticscholar.org/CorpusID:3347345.

Nate Gruver, Samuel C. Stanton, Nathan C Frey, Tim G. J. Rudner, Isidro Hotzel, Julien Lafrance-Vanasse, Arvind Rajpal, Kyunghyun Cho, and Andrew Gordon Wilson. Protein design with guided discrete diffusion. *ArXiv*, abs/2305.20009, 2023. URL https://api.semanticscholar.org/CorpusID:258987335.

Andreia P. Guerreiro, Carlos M. Fonseca, and Luís Paquete. The hypervolume indicator: Problems and algorithms. *ACM Computing Surveys*, 54(6), 2021. doi: 10.1145/3453474. URL https://doi.org/10.1145/3453474.

Emily L. Han, Sophia Tang, Dongyoon Kim, Amanda M. Murray, Kelsey L. Swingle, Alex G. Hamilton, Kaitlin Mrksich, Marshall S. Padilla, Rohan Palanki, Jacqueline J. Li, and Michael J. Mitchell. Peptide-functionalized lipid nanoparticles for targeted systemic mrna delivery to the brain. *Nano Letters*, 25(2):800–810, 01 2025. ISSN 1530-6984. doi: 10.1021/acs.nanolett.4c05186. URL https://doi.org/10.1021/acs.nanolett.4c05186.

Elad Hazan, Sham Kakade, Karan Singh, and Abby Van Soest. Provably efficient maximum entropy exploration. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2681–2691. PMLR, 2019. URL https://proceedings.mlr.press/v97/hazan19a.html.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

Moksh Jain, Sharath Chandra Raparthy, Alex Hernández-García, Jarrid Rector-Brooks, Yoshua Bengio, Santiago Miret, and Emmanuel Bengio. Multi-objective GFlowNets. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 14631–14653. PMLR, 2023. URL https://proceedings.mlr.press/v202/jain23a.html.

Leo Klarner, Tim G. J. Rudner, Garrett M. Morris, Charlotte M. Deane, and Yee Whye Teh. Context-guided diffusion for out-of-distribution molecular and protein design. *ArXiv*, abs/2407.11942, 2024. URL https://api.semanticscholar.org/CorpusID:271218010.

Andreas Krause and Daniel Golovin. Submodular function maximization. In Lucas Bordeaux, Youssef Hamadi, and Pushmeet Kohli (eds.), *Tractability: Practical Approaches to Hard Problems*, chapter 3, pp. 71–104. Cambridge University Press, 2014. doi: 10.1017/CBO9781139177801.004.

Greg Landrum, Paolo Tosco, Ricardo Rodriguez, Brian Kelley, David Cosgrove, Riccardo Vianello, sriniker, Peter Gedeck, Gareth Jones, Dan Nealschneider, Eisuke Kawashima, NadineSchneider, tadhurst cdd, Andrew Dalke, Niels Maeder, Matt Swain, Yakov Pechersky, Brian Cole, Kevin Boyd, Samo Turk, Aleksandr Savelev, Rachel Walker, Alain Vaucher, Maciej Wójcikowski, Hussein Faara, Ichiru Take, Vincent F. Scalfani, Steven Kearnes, Kazuya Ujihara, and Daniel Probst. rdkit/rdkit: 2026_03_6 (q1 2026) release, August 2026. URL https://doi.org/10.5281/zenodo.22140358.

Guang Lin, Shi-Kui Tu, and Lei Xu. Fine-tuning diffusion models for molecular generation via reinforcement learning and fast sampling. *ArXiv*, abs/2606.01220, 2026. URL https://api.semanticscholar.org/CorpusID:288860919.

Xi Lin, Zhiyuan Yang, and Qingfu Zhang. Pareto set learning for neural multi-objective combinatorial optimization. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=QuObT9BTWo.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Peng-Fei Wan, Di Zhang, and Wan-Li Ouyang. Flow-grpo: Training flow matching models via online rl. *ArXiv*, abs/2505.05470, 2025. URL https://api.semanticscholar.org/CorpusID:278394068.

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32819–32848. PMLR, 2024. URL https://proceedings.mlr.press/v235/lou24a.html.

Yining Lu, Zilong Wang, Shiyang Li, Xin Liu, Changlong Yu, Qingyu Yin, Zhan Shi, Zixuan Zhang, and Meng Jiang. Learning to optimize multi-objective alignment through dynamic reward weighting. *arXiv preprint arXiv:2509.11452*, 2025. URL https://arxiv.org/abs/2509.11452.

Elizabeth H. Mahood, William J. Pattie, Ethan M. Jones, Sophia Vincoff, Adi Mashiach, and Pranam Chatterjee. MOOSE: Targeted small molecule generation via multi-objective-guided discrete diffusion. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=0cPAO03FYj.

H. L. Morgan. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5(2):107–113, 05 1965. ISSN 0021-9576. doi: 10.1021/c160017a018. URL https://doi.org/10.1021/c160017a018.

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978. doi: 10.1007/BF01588971.

Edward Pajarillo, Asha Rizor, Jayden Lee, Michael Aschner, and Eunsook Lee. The role of astrocytic glutamate transporters glt-1 and glast in neurological disorders: Potential targets for neurotherapeutics. *Neuropharmacology*, 161:107559, 2019. ISSN 0028-3908. doi: https://doi.org/10.1016/j.neuropharm.2019.03.002. URL https://www.sciencedirect.com/science/article/pii/S0028390819300802. Neurotransmitter Transporters.

Sawan Patel, Sherwood Yao, Sophia Tang, and Pranam Chatterjee. Full-length mRNA design with reward-guided masked diffusion model fine-tuning. In *ICLR 2026 Workshop on Foundation Models for Science*, 2026.

Manish Prajapat, Mojmír Mutný, Melanie N. Zeilinger, and Andreas Krause. Submodular reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=loYSzjSaAK.

Ric De Santi, Bruce D. Lee, Cristian Perez Jensen, Kimon Protopapas, Sophia Tang, Cheng-Hao Liu, Pranam Chatterjee, Yisong Yue, and Andreas Krause. Active flow expansion for out-of-distribution discovery: from theory to molecules. *ArXiv*, abs/2606.08802, 2026. URL https://api.semanticscholar.org/CorpusID:289097154.

Arne Schneuing, Ilia Igashov, Adrian W. Dobbelstein, Thomas Castiglione, Michael Bronstein, and Bruno Correia. Multi-domain Distribution Learning for De Novo Drug Design. *arXiv e-prints*, art. arXiv:2508.17815, August 2025. doi: 10.48550/arXiv.2508.17815.

Adrian M. Schreyer and Tom Blundell. USRCAT: real-time ultrafast shape recognition with pharmacophoric constraints. *Journal of Cheminformatics*, 4(1):27, 2012. ISSN 1758-2946. doi: 10.1186/1758-2946-4-27. URL https://doi.org/10.1186/1758-2946-4-27.

Samuel Stanton, Wesley Maddox, Nate Gruver, Phillip Maffettone, Emily Delaney, Peyton Greenside, and Andrew Gordon Wilson. Accelerating bayesian optimization for biological sequence design with denoising autoencoders. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 20459–20478. PMLR, 2022. URL https://proceedings.mlr.press/v162/stanton22a.html.

Daniel M. Steinberg, Asiri Wijesinghe, Rafael Oliveira, Piotr Koniusz, Cheng Soon Ong, and Edwin V. Bonilla. Amortized active generation of Pareto sets. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://arxiv.org/abs/2510.21052.

Sophia Tang, Yinuo Zhang, and Pranam Chatterjee. PepTune: De novo generation of therapeutic peptides with multi-objective-guided discrete diffusion. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 59017–59065. PMLR, 2025a. URL https://proceedings.mlr.press/v267/tang25n.html.

Sophia Tang, Yuchen Zhu, Molei Tao, and Pranam Chatterjee. TR2-D2: Tree search guided trajectory-aware fine-tuning for discrete diffusion. *arXiv preprint arXiv:2509.25171*, 2025b.

Lei Wang, Nanxi Wang, Wenping Zhang, Xurui Cheng, Zhibin Yan, Gang Shao, Xi Wang, Rui Wang, and Caiyun Fu. Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy*, 7(1):48, 2022. doi: 10.1038/s41392-022-00904-4.

Yinghua Yao, Yuangang Pan, Jing Li, Ivor Tsang, and Xin Yao. PROUD: PaRetO-gUided diffusion model for multi-objective generation. *arXiv preprint arXiv:2407.04493*, 2024.

Ye Yuan, Can Chen, Christopher Pal, and Xue Liu. ParetoFlow: Guided flows in multi-objective optimization. In *International Conference on Learning Representations*, 2025.

Tom Zahavy, Brendan O'Donoghue, Guillaume Desjardins, and Satinder P. Singh. Reward is enough for convex MDPs. In *Advances in Neural Information Processing Systems*, volume 34, pp. 25746–25759, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/d7e4cdde82a894b8f633e6d61a01ef15-Abstract.html.

Richard Zhang and Daniel Golovin. Random hypervolume scalarizations for provable multi-objective black box optimization. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11096–11105, 2020.

Xiaoyuan Zhang, Xi Lin, Bo Xue, Yifan Chen, and Qingfu Zhang. Hypervolume maximization: A geometric view of Pareto set learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://papers.neurips.cc/paper_files/paper/2023/hash/7a7f6cc5dc2a84fb4edf0feb8e5cfd50-Abstract-Conference.html.

Kaiwen Zheng, Huayu Chen, Haotian Ye, Haoxiang Wang, Qinsheng Zhang, Kai Jiang, Hang Su, Stefano Ermon, Jun Zhu, and Mingying Liu. Diffusionnft: Online diffusion reinforcement with forward process. *ArXiv*, abs/2509.16117, 2025. URL https://api.semanticscholar.org/CorpusID:281411162.

Yuchen Zhu, Wei Guo, Jaemoo Choi, Petr Molodyk, Bo Yuan, Molei Tao, and Yongxin Chen. Enhancing reasoning for diffusion llms via distribution matching policy optimization. *ArXiv*, abs/2510.08233, 2025. URL https://api.semanticscholar.org/CorpusID:281950780.

Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, 1999.

A Theory

This section establishes the first variation, concavity, and relative smoothness of the expected-hypervolume objective in Equation (2), and convergence of the regularized HVFT updates. Appendix A.1 proves that the first variation is an expected marginal hypervolume contribution. Appendix A.2 expresses the objective in terms of coverage probabilities, and establishes concavity in the sampling distribution. Appendix A.3 studies the regularized maximizer; Appendices A.4 and A.5 establish curvature bounds and convergence for the update with the linearized KL penalty. We use the notation of Section 3 throughout. In particular, the objective map $\mathbf{r}$ is measurable, the reference point $\mathbf{r}^{\mathrm{ref}}$ is fixed, and the assumed bound $(r_m(x) - r_m^{\mathrm{ref}})_+ \le C$ ensures that $0 \le \mathrm{HV}(A; \mathbf{r}^{\mathrm{ref}}) \le C^M$ for every finite candidate set A.

A.1 First variation

Proposition 1 (Marginal hypervolume reward)**.** *A first variation of F_n at p is given by*

$$f_{n,p}(x) = n\, \mathbb{E}_{A \sim p^{\otimes(n-1)}} \left[\Delta_{\mathrm{HV}}(x; A) \right], \tag{6}$$

where A is formed from $n-1$ independent draws from p.

Proof. Fix an integer $n \ge 1$ and probability distributions p, q on $\mathcal{X}$. Set $h = q - p$ and $p_\epsilon = (1-\epsilon)p + \epsilon q = p + \epsilon h$, where $0 \le \epsilon \le 1$. The signed measure h has total mass zero: $h(\mathcal{X}) = q(\mathcal{X}) - p(\mathcal{X}) = 0$. Writing $x_{1:n} = (x_1, \dots, x_n)$, the definition of expected hypervolume gives

$$F_n(p_\epsilon) = \int_{\mathcal{X}^n} \mathrm{HV}(\{x_1, \dots, x_n\}; \mathbf{r}^{\mathrm{ref}})\, d\big(p_\epsilon^{\otimes n}\big)(x_{1:n}). \tag{9}$$

Expanding the product, its term of order zero is $p^{\otimes n}$, and each term of order one replaces exactly one factor p by h:

$$p_\epsilon^{\otimes n} = p^{\otimes n} + \epsilon \sum_{j=1}^{n} p^{\otimes(j-1)} \otimes h \otimes p^{\otimes(n-j)} + O(\epsilon^2). \tag{10}$$

Substituting Equation (10) into Equation (9) gives

$$F_n(p_\epsilon) = F_n(p) + \epsilon \sum_{j=1}^{n} \int_{\mathcal{X}^n} \mathrm{HV}(\{x_1, \ldots, x_n\}; \mathbf{r}^{\mathrm{ref}})\, d\big(p^{\otimes(j-1)} \otimes h \otimes p^{\otimes(n-j)}\big)(x_{1:n}) + O(\epsilon^2)$$

$$= F_n(p) + \epsilon n \int_{\mathcal{X}} \mathbb{E}_{A \sim p^{\otimes(n-1)}} \left[\mathrm{HV}(A \cup \{x\}; \mathbf{r}^{\mathrm{ref}})\right]\, dh(x) + O(\epsilon^2). \tag{11}$$

For the second equality, hypervolume is unchanged by a permutation of its candidates, so all n integrals are equal. In each one, x denotes the candidate integrated against h, and A is the set formed by the other $n-1$ independent draws from p. The expected background hypervolume is independent of x, and hence

$$\int_{\mathcal{X}} \mathbb{E}_{A \sim p^{\otimes(n-1)}} [\mathrm{HV}(A; \mathbf{r}^{\mathrm{ref}})]\, dh(x) = \mathbb{E}_A[\mathrm{HV}(A; \mathbf{r}^{\mathrm{ref}})]\, h(\mathcal{X}) = 0.$$

We may therefore subtract it inside the linear term in Equation (11). Using the definition of Δ_{HV} in Equation (5), we obtain

$$\begin{aligned} & n \int_{\mathcal{X}} \mathbb{E}_A[\mathrm{HV}(A \cup \{x\}; \mathbf{r}^{\mathrm{ref}})]\, dh(x) \\ & \quad = n \int_{\mathcal{X}} \mathbb{E}_A \left[\mathrm{HV}(A \cup \{x\}; \mathbf{r}^{\mathrm{ref}}) - \mathrm{HV}(A; \mathbf{r}^{\mathrm{ref}})\right]\, dh(x) \\ & \quad = n \int_{\mathcal{X}} \mathbb{E}_A \left[\Delta_{\mathrm{HV}}(x; A)\right]\, dh(x) \\ & \quad = \int_{\mathcal{X}} f_{n,p}(x)\, d(q-p)(x) = \mathbb{E}_q[f_{n,p}(X)] - \mathbb{E}_p[f_{n,p}(X)]. \end{aligned} \tag{12}$$

Consequently,

$$F_n(p_\epsilon) = F_n(p) + \epsilon \left(\mathbb{E}_q[f_{n,p}(X)] - \mathbb{E}_p[f_{n,p}(X)]\right) + O(\epsilon^2),$$

which is the first-variation identity in Equation (4), since $O(\epsilon^2) = o(\epsilon)$. When $n = 1$, the background set is empty and the formula reduces to $f_{1,p}(x) = \mathrm{HV}(\{x\}; \mathbf{r}^{\mathrm{ref}})$, consistently with the linearity of F_1. □

A.2 Per-candidate coverage

For an objective-space point $u \succeq \mathbf{r}^{\mathrm{ref}}$, define the probability that a single candidate drawn from p covers u by

$$S_p(u) := \Pr_{X \sim p}(\mathbf{r}(X) \succeq u) = \int_{\mathcal{X}} \mathbb{1}\{\mathbf{r}(x) \succeq u\}\, p(dx). \tag{13}$$

Coverage of u means that every objective value is at least the corresponding coordinate of u. All coverage integrands below vanish outside the finite box $\prod_{m=1}^{M} [r_m^{\mathrm{ref}}, r_m^{\mathrm{ref}} + C]$, so the integrals are finite under the standing bound.

By Equation (1), the hypervolume of a sampled set is the integral of the indicator that at least one of its candidates covers u. Tonelli's theorem and independence therefore give

$$\begin{aligned} F_n(p) &= \int_{u \succeq \mathbf{r}^{\mathrm{ref}}} \Pr(\exists i \le n : \mathbf{r}(X_i) \succeq u)\, du \\ &= \int_{u \succeq \mathbf{r}^{\mathrm{ref}}} \left[1 - \prod_{i=1}^{n} \Pr(\mathbf{r}(X_i) \not\succeq u)\right] du \\ &= \int_{u \succeq \mathbf{r}^{\mathrm{ref}}} \left[1 - (1 - S_p(u))^n\right] du. \end{aligned} \tag{14}$$

Thus, expected hypervolume integrates the probability that n independent draws collectively cover each objective-space point.

The coverage representation in Equation 14 clarifies the preferences induced by F_n. The marginal value of increasing a point's coverage probability is $\phi_n'(s) = n(1-s)^{n-1}$, which decreases in s

for $n > 1$. Thus, the same increase in coverage probability yields a greater benefit at a point that is rarely covered than at one that is already reached reliably. This diminishing-returns structure favors allocating probability mass across complementary trade-offs, without requiring a fixed set of preference directions to be specified in advance. Beyond providing this interpretation, the coverage representation makes the concavity of F_n transparent.

Theorem 2 (Concavity of expected hypervolume)**.** *For every integer $n \geq 1$, F_n is concave on $\mathcal{P}(\mathcal{X})$. In particular, for any probability distributions p, q and $t \in [0, 1]$,*

$$F_n((1-t)p + tq) \geq (1-t)F_n(p) + tF_n(q). \tag{15}$$

Proof. For each fixed u, the quantity $S_p(u)$ is linear in the distribution p, since it is just the expectation $\mathbb{E}_p$ of a fixed indicator function. Meanwhile, ϕ_n is increasing and concave on $[0, 1]$. Thus, $\phi_n(S_p(u))$ is concave in p, and integration preserves concavity. □

For the remainder of this section, we prove convergence of HVFT under the stated measurability and boundedness conditions, in the idealized setting where first variations are evaluated exactly and each distributional maximization subproblem, once shown to have a unique solution, is solved exactly. Informally, sufficiently small steps improve the regularized objective and converge geometrically to its unique maximizer. We restate the main-text guarantee here.

Theorem 1 (Convergence of HVFT)**.** *Under the mild regularity conditions, the penalized objective in Equation (3) admits a unique maximizer $p^\star$ for $\alpha > 0$. For any sufficiently small constant step size η, the iterates in Equation (7) are well defined, monotonically improve the penalized objective, and converge geometrically to $p^\star$. Precise conditions and rates are given in Appendix A.5.*

We first establish existence and uniqueness of $p^\star$ in Appendix A.3, then bound the linear approximation error in Appendix A.4. These results lead to the precise convergence theorem in Appendix A.5.

A.3 KL REGULARIZATION, EXISTENCE, AND UNIQUENESS

For $\alpha > 0$, write the objective in Equation (3) as

$$G_\alpha(p) := F_n(p) - \alpha D_{\mathrm{KL}}(p\|p_0), \qquad p \in \mathcal{D} := \{p \in \mathcal{P}(\mathcal{X}) : D_{\mathrm{KL}}(p\|p_0) < \infty\}. \tag{16}$$

KL divergence is strictly convex in its first argument on the convex domain $\mathcal{D}$, so $-\alpha D_{\mathrm{KL}}(p\|p_0)$ is strictly concave. Together with Theorem 2, this gives the following two corollaries.

Corollary 1 (Strict concavity of the regularized objective)**.** *For every $n \geq 1$ and $\alpha > 0$, G_α is strictly concave on $\mathcal{D}$.*

Proof. By Theorem 2, F_n is concave. Its sum with the strictly concave functional $-\alpha D_{\mathrm{KL}}(\cdot\|p_0)$ is strictly concave. □

Corollary 2 (Uniqueness of the regularized maximizer)**.** *For $\alpha > 0$, if G_α attains its maximum on $\mathcal{D}$, its maximizer $p^\star$ is unique.*

Proof. If distinct $p, q \in \mathcal{D}$ both maximized G_α, strict concavity would give $G_\alpha((p+q)/2) > (G_\alpha(p) + G_\alpha(q))/2$, contradicting maximality. □

Theorem 3 (Existence of a regularized maximizer)**.** *Under the standing measurability and boundedness assumptions on $\mathbf{r}$, for every integer $n \geq 1$ and $\alpha > 0$, G_α attains its maximum on $\mathcal{D}$.*

Proof. Since $p_0 \in \mathcal{D}$ and $0 \leq F_n \leq C^M$, the supremum is finite. Choose a maximizing sequence (p_j) with $G_\alpha(p_j) \geq G_\alpha(p_0) - 1$. Then

$$\alpha D_{\mathrm{KL}}(p_j\|p_0) = F_n(p_j) - G_\alpha(p_j) \leq C^M - G_\alpha(p_0) + 1.$$

Consequently, the densities $v_j = dp_j/dp_0$ are uniformly integrable: their entropy integrals $\int v_j \log v_j \, dp_0$ are bounded and $a \log a$ grows superlinearly. By the Dunford–Pettis theorem, a subsequence converges weakly in $L^1(p_0)$ to a density $v \geq 0$ with $\int v \, dp_0 = 1$. Let $p = v\, p_0$. Testing this convergence against the bounded indicator $\mathbb{1}\{\mathbf{r}(x) \succeq u\}$ gives $S_{p_j}(u) \to S_p(u)$ for each u. The

coverage representation in Equation (14) and dominated convergence on the finite coverage box imply $F_n(p_j) \to F_n(p)$. The entropy integral is weakly lower semicontinuous in $L^1(p_0)$, so $p \in \mathcal{D}$ and

$$G_\alpha(p) \geq \limsup_j G_\alpha(p_j) = \sup_{q \in \mathcal{D}} G_\alpha(q).$$

Thus p attains the supremum. □

Corollary 3 (Existence and uniqueness of the regularized maximizer)**.** *Under the assumptions of Theorem 3, the optimization problem in Equation (3) has exactly one maximizer $p^\star \in \mathcal{D}$.*

Proof. Existence follows from Theorem 3 and uniqueness from Corollary 2. Distributions outside $\mathcal{D}$ have $G_\alpha = -\infty$ and cannot be maximizers. □

Observation: first variation of the regularized objective. As in the main text, write $p(x)$ and $p_0(x)$ for densities with respect to a common dominating measure. When p and p_0 are mutually absolutely continuous and $\log(p/p_0)$ is bounded p_0-almost everywhere, a first variation of G_α is

$$g_{\alpha,p}(x) := f_{n,p}(x) - \alpha \log \frac{p(x)}{p_0(x)}. \tag{17}$$

Indeed, the first variation of $D_{\mathrm{KL}}(p\|p_0)$ is $\log(p(x)/p_0(x)) + 1$. The constant $-\alpha$ can be omitted because it integrates to zero against $q - p$. Thus, the full linearized reward combines the marginal hypervolume reward with the log-ratio penalty used in Equation (7).

A.4 Bounding Linear Approximation Error: Relative Smoothness

We bound the error of the linear approximations used by HVFT in terms of KL divergence. For a concave functional, this means bounding how far the functional can fall below its tangent approximation; equivalently, its negative is smooth relative to entropy.

A.4.1 Relative smoothness of $F_n(p)$

Define

$$L_n := \frac{n(n-1)C^M}{4}, \qquad \|q - p\|_{\mathrm{TV}} := \sup_{E \subseteq \mathcal{X} \text{ measurable}} |q(E) - p(E)|.$$

Proposition 2 (Relative smoothness of expected hypervolume)**.** *For every integer $n \geq 1$ and probability distributions p, q with $D_{\mathrm{KL}}(q\|p) < \infty$,*

$$0 \leq F_n(p) + \int_{\mathcal{X}} f_{n,p}(x)\, d(q-p)(x) - F_n(q) \leq L_n D_{\mathrm{KL}}(q\|p). \tag{18}$$

Proof. For $n \geq 2$, set $p_t = (1-t)p + tq$ and $\psi(t) = F_n(p_t)$. Since $S_{p_t}(u) = (1-t)S_p(u) + tS_q(u)$, differentiating the coverage representation in Equation (14) twice gives

$$\psi''(t) = -n(n-1) \int_{u \succeq \mathbf{r}^{\mathrm{ref}}} (1 - S_{p_t}(u))^{n-2} (S_q(u) - S_p(u))^2 \, du.$$

The finite coverage box and bounded integrands justify differentiation under the integral; endpoint values are understood by continuity. Because $S_q(u) - S_p(u)$ is the difference of the probabilities of the same measurable event, $|S_q(u) - S_p(u)| \leq \|q - p\|_{\mathrm{TV}}$. Thus

$$-n(n-1)C^M \|q-p\|_{\mathrm{TV}}^2 \leq \psi''(t) \leq 0.$$

Taylor's formula with integral remainder, together with $\psi'(0) = \int f_{n,p}(x)\, d(q-p)(x)$, implies

$$\begin{aligned} 0 \leq \psi(0) + \psi'(0) - \psi(1) &= -\int_0^1 (1-t)\psi''(t)\, dt \\ &\leq \frac{n(n-1)C^M}{2} \|q-p\|_{\mathrm{TV}}^2 \leq L_n D_{\mathrm{KL}}(q\|p), \end{aligned}$$

where the last inequality is Pinsker's inequality. For $n = 1$, F_1 is linear and the result holds with $L_1 = 0$. □

A.4.2 Relative smoothness of $G_\alpha(p)$

The KL regularizer also contributes to the linearization error. Using the first variation in Equation (17), we account for the log-ratio term explicitly. Its contribution to the error is exactly $\alpha D_{\mathrm{KL}}(q\|p)$, which adds to the relative smoothness constant and supplies a quantitative strong-concavity bound.

Proposition 3 (Relative smoothness and strong concavity of G_α)**.** *Let $\alpha > 0$, $p, q \in \mathcal{D}$, and suppose that p is equivalent to p_0 and $\log(p/p_0) \in L^\infty(p_0)$. Then*

$$\begin{aligned}\alpha D_{\mathrm{KL}}(q\|p) &\le G_\alpha(p) + \int_{\mathcal{X}} g_{\alpha,p}(x)\, d(q-p)(x) - G_\alpha(q) \\ &\le (L_n + \alpha) D_{\mathrm{KL}}(q\|p). \end{aligned} \tag{19}$$

Thus $-G_\alpha$ is $(L_n+\alpha)$-smooth relative to entropy, and G_α is α-strongly concave relative to entropy for these pairs of distributions.

Proof. The assumptions ensure that all terms in the entropy identity

$$D_{\mathrm{KL}}(q\|p_0) - D_{\mathrm{KL}}(p\|p_0) - \int_{\mathcal{X}} \log \frac{p(x)}{p_0(x)}\, d(q-p)(x) = D_{\mathrm{KL}}(q\|p)$$

are finite. Substituting $G_\alpha(p) = F_n(p) - \alpha D_{\mathrm{KL}}(p\|p_0)$ and $g_{\alpha,p}(x) = f_{n,p}(x) - \alpha \log(p(x)/p_0(x))$ into the linearization error gives

$$\begin{aligned} & G_\alpha(p) + \int_{\mathcal{X}} g_{\alpha,p}(x)\, d(q-p)(x) - G_\alpha(q) \\ &\quad = F_n(p) - \alpha D_{\mathrm{KL}}(p\|p_0) + \int_{\mathcal{X}} \Big[f_{n,p}(x) - \alpha \log \frac{p(x)}{p_0(x)} \Big]\, d(q-p)(x) \\ &\qquad - F_n(q) + \alpha D_{\mathrm{KL}}(q\|p_0) \\ &\quad = \left[F_n(p) + \int_{\mathcal{X}} f_{n,p}(x)\, d(q-p)(x) - F_n(q) \right] \\ &\qquad + \alpha \left[D_{\mathrm{KL}}(q\|p_0) - D_{\mathrm{KL}}(p\|p_0) - \int_{\mathcal{X}} \log \frac{p(x)}{p_0(x)}\, d(q-p)(x) \right] \\ &\quad = \left[F_n(p) + \int_{\mathcal{X}} f_{n,p}(x)\, d(q-p)(x) - F_n(q) \right] + \alpha D_{\mathrm{KL}}(q\|p). \end{aligned}$$

The last equality uses the KL identity above. By Proposition 2, the bracketed linearization error of F_n lies between 0 and $L_n D_{\mathrm{KL}}(q\|p)$. Adding $\alpha D_{\mathrm{KL}}(q\|p)$ to these two bounds gives Equation (19): the lower bound establishes relative strong concavity, and the upper bound establishes relative smoothness. □

A.5 Convergence of the linearized HVFT update

We now analyze the exact distributional update in Equation (7), which uses the full reward g_{α,p_k}, including the linearized log-ratio penalty:

$$p_{k+1} = \arg\max_{p \in \mathcal{D}} \left\{ \mathbb{E}_p[g_{\alpha,p_k}(X)] - \eta^{-1} D_{\mathrm{KL}}(p\|p_k) \right\}. \tag{20}$$

The pretrained-model penalty enters through $-\alpha \log(p_k(x)/p_0(x))$ in the reward. Its curvature is accounted for by the $L_n + \alpha$ step-size bound below.

Throughout the next three lemmas, assume the standing measurability and boundedness conditions, fix $n \ge 1$ and $\alpha > 0$, initialize at p_0, and choose $0 < \eta \le (L_n + \alpha)^{-1}$. Write $\rho = 1 - \alpha\eta \in [0, 1)$.

Lemma 1 (Well-defined updates)**.** *Every subproblem in Equation (20) has a unique solution $p_{k+1} \in \mathcal{D}$, given by*

$$p_{k+1}(x) = \frac{p_k(x) \exp(\eta g_{\alpha,p_k}(x))}{Z_k}, \qquad Z_k := \mathbb{E}_{p_k}[\exp(\eta g_{\alpha,p_k}(X))]. \tag{21}$$

Each iterate is mutually absolutely continuous with p_0, and $\log(p_k/p_0)$ is bounded p_0-almost everywhere.

Proof. Initially $\log(p_0/p_0) = 0$. If $\log(p_k/p_0)$ is bounded, then g_{α,p_k} is bounded because $0 \le f_{n,p_k} \le nC^M$. Consequently $0 < Z_k < \infty$, and the exponential update defines a probability distribution with the same null sets as p_k. Moreover,

$$\log \frac{p_{k+1}(x)}{p_0(x)} = \log \frac{p_k(x)}{p_0(x)} + \eta g_{\alpha,p_k}(x) - \log Z_k$$

is bounded. Induction therefore gives $p_{k+1} \in \mathcal{D}$ and the conditions required by Proposition 3 at every iteration.

For any $p \in \mathcal{D}$, substituting the exponential update into $D_{\mathrm{KL}}(p\|p_{k+1})$ and rearranging yields

$$\mathbb{E}_p[g_{\alpha,p_k}(X)] - \eta^{-1} D_{\mathrm{KL}}(p\|p_k) = \eta^{-1} \log Z_k - \eta^{-1} D_{\mathrm{KL}}(p\|p_{k+1}).$$

All terms are finite by the bounded log ratios and $p \in \mathcal{D}$. Since KL divergence is nonnegative and vanishes exactly at $p = p_{k+1}$, this distribution uniquely solves the subproblem. □

Lemma 2 (One-step bound and monotonic improvement)**.** *For every $q \in \mathcal{D}$ and $k \ge 0$, the iterates satisfy*

$$\begin{aligned} G_\alpha(q) - G_\alpha(p_{k+1}) \le (\eta^{-1} - \alpha) D_{\mathrm{KL}}(q\|p_k) - \eta^{-1} D_{\mathrm{KL}}(q\|p_{k+1}) \\ - \big(\eta^{-1} - L_n - \alpha\big) D_{\mathrm{KL}}(p_{k+1}\|p_k). \end{aligned} \tag{22}$$

In particular, $G_\alpha(p_{k+1}) \ge G_\alpha(p_k)$.

Proof. Taking logarithms of the update in Lemma 1 gives

$$\eta g_{\alpha,p_k}(x) = \log \frac{p_{k+1}(x)}{p_k(x)} + \log Z_k.$$

Integrate against $q - p_{k+1}$. The constant $\log Z_k$ cancels because both distributions have mass one. Furthermore,

$$\int_{\mathcal{X}} \log \frac{p_{k+1}(x)}{p_k(x)}\, dq(x) = D_{\mathrm{KL}}(q\|p_k) - D_{\mathrm{KL}}(q\|p_{k+1}),$$

while integration of the same logarithm against p_{k+1} gives $D_{\mathrm{KL}}(p_{k+1}\|p_k)$. Therefore

$$\begin{aligned} \int_{\mathcal{X}} g_{\alpha,p_k}(x)\, d(q - p_{k+1})(x) = \eta^{-1} \big[D_{\mathrm{KL}}(q\|p_k) - D_{\mathrm{KL}}(q\|p_{k+1}) \\ - D_{\mathrm{KL}}(p_{k+1}\|p_k) \big]. \end{aligned} \tag{23}$$

Next apply the two bounds in Equation (19) at the same base point p_k. Relative strong concavity bounds the comparator's objective from above:

$$G_\alpha(q) \le G_\alpha(p_k) + \int_{\mathcal{X}} g_{\alpha,p_k}(x)\, d(q - p_k)(x) - \alpha D_{\mathrm{KL}}(q\|p_k).$$

Relative smoothness bounds the next iterate's objective from below:

$$G_\alpha(p_{k+1}) \ge G_\alpha(p_k) + \int_{\mathcal{X}} g_{\alpha,p_k}(x)\, d(p_{k+1} - p_k)(x) - (L_n + \alpha) D_{\mathrm{KL}}(p_{k+1}\|p_k).$$

Subtracting these inequalities cancels $G_\alpha(p_k)$ and combines the two integrals into an integral against $q - p_{k+1}$:

$$\begin{aligned} G_\alpha(q) - G_\alpha(p_{k+1}) \le \int_{\mathcal{X}} g_{\alpha,p_k}(x)\, d(q - p_{k+1})(x) - \alpha D_{\mathrm{KL}}(q\|p_k) \\ + (L_n + \alpha) D_{\mathrm{KL}}(p_{k+1}\|p_k). \end{aligned}$$

Substituting Equation (23) and collecting the coefficients of each KL divergence proves Equation (22). For $q = p_k$, it gives

$$G_\alpha(p_{k+1}) - G_\alpha(p_k) \ge \eta^{-1} D_{\mathrm{KL}}(p_k\|p_{k+1}) + (\eta^{-1} - L_n - \alpha) D_{\mathrm{KL}}(p_{k+1}\|p_k) \ge 0,$$

since $\eta^{-1} \ge L_n + \alpha$. □

Lemma 3 (Geometric contraction and objective convergence)**.** *Let $p^\star$ be the unique maximizer from Corollary 3, and write $D_k := D_{\mathrm{KL}}(p^\star \| p_k)$ and $\Delta_k := G_\alpha(p^\star) - G_\alpha(p_k)$. Then*

$$D_{k+1} \le \rho D_k, \qquad \Delta_{k+1} \le \eta^{-1}(\rho D_k - D_{k+1}), \qquad k \ge 0.$$

Consequently, $D_k \le \rho^k D_0$ and $\Delta_k \le \eta^{-1}\rho^k D_0$ for every $k \ge 1$.

Proof. The initial divergence D_0 is finite because $p^\star \in \mathcal{D}$. Every subsequent D_k is finite as well, since

$$D_k = D_{\mathrm{KL}}(p^\star \| p_0) - \mathbb{E}_{p^\star}\left[\log \frac{p_k(X)}{p_0(X)}\right]$$

and the logarithm is bounded by Lemma 1. Taking $q = p^\star$ in Equation (22), dropping its final nonpositive term, and using $\eta^{-1} - \alpha = \eta^{-1}\rho$ yields

$$0 \le \Delta_{k+1} \le \eta^{-1}(\rho D_k - D_{k+1}).$$

The left inequality follows from maximality of $p^\star$. It forces $D_{k+1} \le \rho D_k$, and iteration gives $D_k \le \rho^k D_0$. For $k \ge 1$, nonnegativity of D_k also gives

$$\Delta_k \le \eta^{-1}(\rho D_{k-1} - D_k) \le \eta^{-1}\rho D_{k-1} \le \eta^{-1}\rho^k D_0.$$

If $\rho = 0$, the same inequalities imply $D_1 = 0$ and $\Delta_1 = 0$, so $p_1 = p^\star$. Otherwise, both bounds decay geometrically because $0 < \rho < 1$. □

The three lemmas establish the following convergence guarantee.

Theorem 4 (Convergence of exact linearized HVFT)**.** *Under the standing measurability and boundedness assumptions, let $n \ge 1$, $\alpha > 0$, and initialize at the pretrained distribution p_0. Choose a constant step size*

$$0 < \eta \le \frac{1}{L_n + \alpha}, \qquad \rho := 1 - \alpha\eta \in [0, 1).$$

Then every update in Equation (20) has a unique solution, and $G_\alpha(p_{k+1}) \ge G_\alpha(p_k)$. For the unique maximizer $p^\star$ from Corollary 3,

$$D_{\mathrm{KL}}(p^\star \| p_k) \le \rho^k D_{\mathrm{KL}}(p^\star \| p_0), \qquad k \ge 1, \tag{24}$$

$$G_\alpha(p^\star) - G_\alpha(p_k) \le \eta^{-1}\rho^k D_{\mathrm{KL}}(p^\star \| p_0), \qquad k \ge 1. \tag{25}$$

In particular, $p_k \to p^\star$ in total variation. When $\rho = 0$, the first update already equals $p^\star$.

Proof. Lemma 1 establishes existence and uniqueness of each update. Lemma 2 proves monotonic improvement, and Lemma 3 gives Equations (24) and (25), including the case $\rho = 0$. Finally, Pinsker's inequality gives

$$\|p_k - p^\star\|_{\mathrm{TV}} \le \sqrt{\frac{D_{\mathrm{KL}}(p^\star \| p_k)}{2}} \le \sqrt{\frac{\rho^k D_{\mathrm{KL}}(p^\star \| p_0)}{2}},$$

which proves convergence in total variation. □

At the largest admissible step size, $\eta = (L_n + \alpha)^{-1}$, the contraction factor is $\rho = L_n/(L_n + \alpha)$. These guarantees apply to exact first variations and exact optimization over distributions. Monte Carlo rewards, neural parameterization, and approximate inner optimization require additional error analysis.

B Practical implementation aspects

B.1 Dealing with the log-ratio

A practical implementation must account for the log-density ratio $\log p(X) - \log p_{\mathrm{pre}}(X)$ in Equation (7), which involves marginal densities that are often intractable for diffusion models. We consider two approaches. The first, used in our experiments, sets $\alpha = 0$ while retaining the local KL penalty relative to p_k. This avoids evaluating the density ratio and applies to both continuous and discrete

generators. The local penalty constrains each update, while the cumulative departure from p_0 depends on the step sizes η_k and the stopping point. This approach is empirically effective, as shown in Section 4.

The second approach, applicable to continuous diffusion and flow-matching models, retains $\alpha > 0$ and uses a solver driven by reward gradients, such as Adjoint Matching (Domingo-Enrich et al., 2025). The adaptive reward has gradient

$$\nabla_x \left[f_{n,p_k}(x) - \alpha \log \frac{p_k(x)}{p_0(x)} \right] = \nabla_x f_{n,p_k}(x) - \alpha \left[\nabla_x \log p_k(x) - \nabla_x \log p_0(x) \right],$$

wherever the required derivatives exist. This approach requires estimates of the output-distribution scores, which are readily available for diffusion and flow-matching models (Domingo-Enrich et al., 2025). The only remaining obstacle is the first term: a differentiable approximation of the marginal hypervolume reward f_{n,p_k}.

Random hypervolume scalarizations provide a convenient representation of this reward (Zhang & Golovin, 2020). For the objective vector $\mathbf{r}(x) \in \mathbb{R}^M$ and reference point $\mathbf{r}^{\mathrm{ref}}$, define

$$s_\lambda(\mathbf{r}(x) - \mathbf{r}^{\mathrm{ref}}) = \left[\min_{1 \le m \le M} \frac{r_m(x) - r_m^{\mathrm{ref}}}{\lambda_m} \right]_+^M, \qquad \lambda \sim \mathrm{Uniform}(S_+^{M-1}),$$

where $[a]_+ = \max\{a, 0\}$ and

$$S_+^{M-1} = \left\{ \lambda \in \mathbb{R}^M_{>0} : \|\lambda\|_2 = 1 \right\}.$$

The marginal hypervolume reward in Equation (6) can then be written as

$$f_{n,p_k}(x) = n\, c_M\, \mathbb{E}_{\lambda,\, A \sim p_k^{\otimes(n-1)}} \left[\left[s_\lambda(\mathbf{r}(x) - \mathbf{r}^{\mathrm{ref}}) - \max_{x' \in A} s_\lambda(\mathbf{r}(x') - \mathbf{r}^{\mathrm{ref}}) \right]_+ \right], \tag{26}$$

$$c_M = \frac{\pi^{M/2}}{2^M\, \Gamma(M/2 + 1)}.$$

This representation supports Monte Carlo estimation over directions λ and background sets A: for each sample, draw a direction λ and a set A of $n-1$ points from p_k, and evaluate the bracketed expression.

A smooth surrogate of Equation (26) is obtained by replacing the minimum in s_λ and each positive-part operation $[a]_+$ with softmin and softplus respectively. The maximum over the background set term $\max_{x' \in A} s_\lambda(\mathbf{r}(x') - \mathbf{r}^{\mathrm{ref}})$ need not be smoothed, since it is independent of x.

Automatic differentiation of this surrogate then yields an estimate of the reward gradient $\nabla_x f_{n,p_k}(x)$ for use in Adjoint Matching, provided that $\mathbf{r}$ is itself differentiable. Combined with the score terms above, this furnishes the full gradient of the entropy-regularized reward required by the solver.

B.2 Computational Cost

At first glance, Algorithm 1 appears to require a K-fold increase in training computation. In practice, however, the total training budget can be split across the K fine-tuning rounds, so the overall number of gradient steps remains unchanged. HVFT's only additional cost relative to running the inner solving algorithm for the same number of epochs therefore comes from generating and scoring the background sets—a small overhead relative to the full fine-tuning pipeline. Table 3 compares HVFT's runtime to that of its inner solver counterpart in Section 4. For the TR2-D2 comparison, statistics and standard errors are computed over 9 runs spanning all three peptide tasks; for the DiffusionNFT comparison, they are computed over 3 runs of the Geometry task.

C Experimental Details

C.1 Metrics

For each metric, we sample 1000 valid designs from the generator. Since generated designs may not always be valid, we allow repeated sampling until the 1000 quota is fulfilled or a maximum of 10000 trials is exceeded.

Table 3: Average execution time in hours. The first block aggregates results across the three peptide tasks. The second block aggregates runs for the Geometry task. The performance of these runs are reported in Section 4. HVFT's execution time remain comparable to its inner solver counterpart.

Method	Execution time (hours)
TR2-D2	12.384 ± 1.114
HVFT	10.734 ± 0.310
DiffusionNFT	6.399 ± 0.069
HVFT	7.204 ± 0.082

Table 4: Reference points used for hypervolume computation across all peptides tasks.

Property	Binding Affinity	Permeability	Nonfouling	Solubility	Hemolysis
Reference Point	4.5	-10.0	0.0	0.0	0.0

Our main objective-space metric is the mean hypervolume of 100 samples from the generator. We estimate the 100-hypervolume with sub-sampling: we draw $P = 100000$ subsets of size 100 without replacement from the pool and average their hypervolumes. For each index $i \in [1, 1000]$, we average the hypervolume values of the permutations that exclude $\mathbf{r}(X_i)$ to form a leave-one-out estimate $\widehat{\mathrm{HV}}_{-i}$. The final hypervolume is the mean of these leave-one-out estimates, $\widehat{\mathrm{HV}} = \frac{1}{N}\sum_i \widehat{\mathrm{HV}}_{-i}$. 10-hypervolume is computed the same way.

We used the full set of (up to) 1000 samples for the design-space metrics. Validity is computed as the number of valid samples divided by the number of attempts required to reach this quota, where validity is determined by problem-specific criteria. Following Santi et al. (2026), we measure generator diversity via the Vendi score of valid designs (Friedman & Dieng, 2022).

To assess design-space coverage, we adopt a metric similar to the fixed-threshold clustering metric proposed in the same work. Designs are embedded in a metric space and grouped into clusters such that any two designs within a cluster lie within a fixed radius of one another; the resulting number of clusters can be interpreted as the number of hyperspheres needed to cover an equal-sized set of designs from each method, with more clusters indicating broader coverage. Because this metric depends on the choice of radius — which encodes an arbitrary notion of how "similar" two designs must be to count as the same — we avoid committing to a single threshold. Instead, we compute the number of clusters across 100 equally spaced thresholds and integrate numerically, yielding a single coverage score that is independent of any particular similarity cutoff.

C.2 Implementation Details

HVFT introduces four method-specific parameters: the background count B, the set size n, the regularization coefficient α, and the number of outer loop steps K. The background count controls the precision of the Monte Carlo estimator of f_{n,p_k}. We found that $B = 85$ provides sufficient accuracy in practice and use this value throughout our experiments. We also keep $n = 10$ fixed across experiments and examine its effect in Appendix D. Following the discussion in Appendix B, we set the pretrained-model regularization coefficient $\alpha = 0$, while retaining the local KL penalty between successive iterates. The local trust region coefficient η_k is fixed across k so we drop the subscript in the remainder of this section. K is task specific.

C.2.1 Therapeutic Peptides

To compute Vendi score and coverage, we use 1024-bit Morgan fingerprints (radius 2) (Morgan, 1965) and Tanimoto similarity. The reference point for hypervolume computation is written in Table 4 and is shared for all peptide tasks.

Table 5: Hyperparameters for TR2-D2 baselines and HVFT inner-loop finetuning.

Hyperparameter	Name in Tang et al. (2025b)	Value
Num. of Replicates for Weighted Denoising Cross-Entropy	R	16
Buffer Size	B	20
Sequence Length	L_{seq}	200
Num. of Children	M	50
Num. of Decoding Steps	N_{steps}	128
Num. of MCTS Iterations	N_{iter}	20
Resample Every n Steps	N_{resample}	20
Max Token Length	L	200
Learning rate	–	10^{-4}
η^{-1}	α	0.1 (TR2-D2) or 1 (HVFT)

Table 6: Reference points used for hypervolume computation for the Topology and Geometry tasks.

Property	QED	SA	Dipole Moment
Reference Point	-1.0	-1.0	-1.0

TR2-D2, TR2-D2+. Following Tang et al. (2025b), we finetune the base model for 1000 epochs total, evaluating every 100 epochs. Hyperparameters for TR2-D2 and TR2-D2+ are taken directly from Tang et al. (2025b).

HVFT. HVFT decomposes this same 1000-epoch budget into $K = 50$ linear finetuning subproblems, each run for 20 epochs. Its inner solver is the single-objective variant of TR2-D2, and it shares the same hyperparameters as the baselines except for η^{-1}. Table 5 summarizes these settings.

Peptune. To ensure a fair comparison, we depart from the evaluation protocol used for Peptune in Tang et al. (2025a). Specifically, we set $N_{\text{iter}} = 200$ and $M = 50$, so that at most 10000 samples are considered during inference. Every valid candidate encountered is retained in a running buffer until it reaches 1000 samples. Candidates that are unrolled during the tree search but turn out invalid still count toward the validity total.

C.2.2 Molecular Design

In all tasks, we fix the number of atoms to 10. We normalize Synthetic Accessibility via $(10 - SA)/9$ and additionally scale the dipole moment by a factor of 20 to avoid magnitude bias across baselines. Invalid candidates receive a reward of -1, giving the finetuning algorithms a signal when a design is invalid. The reference points for the Topology and Geometry tasks are given in Table 6.

All finetuning methods rely on DiffusionNFT, either as the main solver or as a component of their algorithm. Its hyperparameters are held fixed across methods and are summarized in Table 7.

Topology. For the topology task, we allow 900 finetuning epochs, evaluating every 50. We compute Vendi score and coverage using 1024-bit Morgan fingerprints (radius 2) with Tanimoto similarity. Here, HVFT uses $K = 9$ to get 100 epochs per inner finetuning subproblem.

Geometry. For the geometry task, computational constraints limit us to 600 epochs, evaluating every 50. We compute Vendi score and coverage using USRCAT fingerprints and the USRCAT similarity score (Schreyer & Blundell, 2012), as implemented in RDKit (Landrum et al., 2026). Here, HVFT uses $K = 6$ to get 100 epochs per inner finetuning subproblem.

DiffusionBlend. Cheng et al. (2025) do not describe an automatic weight-selection procedure. We generate evenly spread reward-weighting tradeoffs via Das–Dennis sampling (for Geometry) or uniform sampling in $[0, 1]$ (for Topology).

Table 7: Additional hyperparameters for DiffusionNFT finetuning.

Hyperparameter	**Name in Zheng et al. (2025)**	**Value**
Guidance Strength	β	50
KL Regularization Coefficient	Not in paper, but in official implementation	0.5
Learning Rate	–	1e-5

Table 8: Effect of n on objective- and design-space metrics for the TfR and topology tasks, with the pretrained base model as a reference. Values are means $\pm$ standard errors across three runs. Boldface and underlining indicate the highest and second-highest means within each task, respectively.

		Objective-space metrics		Design-space metrics		
Task	Model / n	100-Hypervolume	10-Hypervolume	Diversity	Coverage	Valid fraction
TfR	Base	2.93 ± 0.04	1.69 ± 0.01	60.35 ± 0.14	401.50 ± 0.28	$\mathbf{0.321 \pm 0.007}$
	10	11.57 ± 0.16	10.15 ± 0.18	75.89 ± 9.42	$\mathbf{485.80 \pm 17.27}$	$\underline{0.250 \pm 0.021}$
	50	$\underline{12.231 \pm 0.153}$	$\underline{10.589 \pm 0.226}$	$\underline{93.2 \pm 3.0}$	437.9 ± 62.9	0.210 ± 0.078
	150	$\mathbf{13.757 \pm 1.721}$	$\mathbf{11.107 \pm 0.891}$	$\mathbf{107.3 \pm 8.6}$	$\underline{454.1 \pm 43.6}$	0.235 ± 0.114
Topology	FlowMol	3.116 ± 0.014	2.665 ± 0.007	$\underline{266.1 \pm 4.9}$	$\underline{507.1 \pm 6.4}$	0.705 ± 0.008
	10	$\mathbf{3.416 \pm 0.005}$	$\mathbf{3.247 \pm 0.004}$	154.9 ± 9.0	351.6 ± 11.1	$\mathbf{0.734 \pm 0.007}$
	100	$\underline{3.405 \pm 0.004}$	$\underline{3.200 \pm 0.006}$	254.5 ± 3.5	473.1 ± 5.0	0.675 ± 0.008
	1000	3.362 ± 0.010	3.082 ± 0.017	$\mathbf{310.6 \pm 6.3}$	$\mathbf{547.4 \pm 6.5}$	$\underline{0.708 \pm 0.012}$

D Ablations

To assess HVFT's sensitivity to n, we conduct additional experiments on representative therapeutic-peptide and small-molecule design tasks, holding all other parameters fixed as specified in Appendix C.2. Table 8 shows substantial hypervolume gains over the pretrained model across the tested values of n, which span up to two orders of magnitude. These results suggest that HVFT is robust to the choice of n.

On the Topology task, which has two objectives, increasing n does not improve hypervolume: $n = 10$ achieves the highest values. This suggests that relatively few samples may suffice to capture the main trade-offs on this task's Pareto front.

Conversely, on the peptide task TfR, increasing n consistently yields better hypervolume values. With five rewards, the Pareto front of this task is likely more complex and not easily covered by a small set of points. The larger variance at $n = 150$ has a simple explanation: Since $B = 85$, each point's marginal contribution is estimated over 85 sets, each containing 149 other scores. Contributing additional hypervolume coverage against a set of that size is hard, so most of these 85 estimates are zero, and only a handful are non-zero. As n grows, this sparsity makes the average marginal contribution noisier, producing higher variance across runs.