CRAFT: A CANDIDATE-RETENTION FRAMEWORK FOR OFFLINE-TO-ONLINE SEQUENCE DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Designing biological sequences under a limited experimental budget requires selecting a small number of candidates for evaluation. Even when candidate generation produces high-performing sequences, underestimation of their performance may prevent their selection for experimental evaluation. We propose CRAFT, an offline-to-online framework that separates candidate generation from experimental selection and uses accumulated measurements to guide both. Starting from an offline dataset, it constructs candidate pools via measurement-guided search, including local mutations of high-scoring observed sequences. A separate selector applies the same scoring rule to candidates from all sources, using an upper quantile of the measured property values of their nearest neighbors, and forms evaluation batches favoring high scores and sequence diversity. After each round, new measurements update candidate generation and scoring, while recent performance determines how future evaluations are allocated among candidate sources. Theoretically, for fixed candidate pools under bounded scoring error, we derive an upper bound on the gap in top-K mean objective value between the candidate pool and the evaluated batch, and relate this gap to cumulative improvement across rounds. Across six benchmarks spanning RNA, DNA, and protein design, CRAFT achieves the highest mean top-128 score on five tasks and ranks second on GFP under matched online query budgets. Our code is available at `https://anonymous.4open.science/r/craft-code-private-3C88/`.

1 INTRODUCTION

Designing biological sequences with desired properties is important for drug discovery and industrial biotechnology. The combinatorial size of sequence spaces and the cost of experimental evaluation make exhaustive screening impractical (Jain et al., 2022). Machine learning-assisted directed evolution therefore uses measured sequence–property pairs to guide which variants are tested (Wu et al., 2019). Related offline methods include CbAS, which adapts a conditional design distribution, and COMs, which addresses unreliable extrapolation through conservative surrogate modeling (Brookes et al., 2019; Trabucco et al., 2021). We study the setting in which an initial offline dataset is available and a limited number of additional evaluations can be performed in successive batches. The goal is to discover high-scoring, previously unmeasured sequences within this evaluation budget.

Model-based sequence design requires deciding both where to search and which proposed sequences to evaluate. AdaLead and PEX perform model-guided evolutionary search, while GFN-AL combines diverse generation with uncertainty-aware acquisition (Sinai et al., 2020; Ren et al., 2022; Jain et al., 2022). For example, δ-Conservative Search adjusts perturbations of high-scoring measured sequences according to model uncertainty (Kim et al., 2024), while ProSpero guides a pretrained generator with a surrogate updated from experimental feedback (Kmicikiewicz et al., 2026). ProSpero also uses this surrogate to rank its final candidates. Its prediction errors can therefore affect both candidate generation and experimental selection: even when search produces a high-performing sequence, an underestimated score can keep it out of the evaluation batch. The importance of candidate ordering is also reflected in RaM, which learns rankings rather than relying on pointwise regression (Tan et al., 2025). Improving generation alone does not remove this problem. This motivates a design in which the selection rule can be changed and assessed separately from the procedure that generates candidates.

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.

We propose CRAFT, an offline-to-online framework that separates candidate generation, experimental selection, and feedback updates. Like optimizer portfolios such as P3BO, CRAFT shares measurements across search procedures and adapts their allocations of the query budget (Angermueller et al., 2020). CRAFT additionally places a common candidate-selection stage between these procedures and batch construction. Each procedure, or source, supplies a pool with a fixed size limit. The selector scores the offered candidates using the same measured reference, prioritizes candidates within each source, and merges the resulting lists according to the current query allocations. Generation can therefore use its own search model without requiring selection to reuse that model. This separation also allows us to analyze the quality available in the candidate pools separately from the quality retained in the evaluated batch. A broader discussion of related work, including offline and online sequence optimization and optimizer portfolios, is provided in Appendix A.

We instantiate this framework as a sequential experimental-design loop (Box & Wilson, 1992). Candidate generators use accumulated measurements to construct pools of unevaluated sequences, while allowing each generator to employ its own search strategy or predictive model. A shared selector evaluates these candidates against a common measured reference, using an upper quantile of the observed property values among their nearest neighbors without fitting an additional global predictor. Within each candidate pool, score-based screening is combined with farthest-first selection to balance predicted quality and sequence diversity (Johnson et al., 1990). The resulting lists are merged according to generator-specific query quotas, excluding duplicate and previously measured sequences. After each evaluation round, new measurements update both candidate generation and the reference used for scoring. Discounted returns from evaluated candidates adjust future query allocations, giving greater weight to recent outcomes. This process allows offline measurements and online feedback to jointly guide candidate generation and experimental selection.

Our main contributions are as follows:

1. We introduce CRAFT, a framework that distinguishes candidate-pool quality from the quality of the evaluated batch. For fixed candidate pools under bounded scoring error, we bound the gap in top-K mean objective value in terms of batch constraints, scoring error, and approximate selection, and relate this gap to cumulative improvement across online rounds.
2. We instantiate the framework as a sequence-design algorithm combining measurement-guided candidate construction, shared nearest-neighbor quantile scoring, diversity-aware batch selection, and adaptive source allocation. New measurements update both generation and selection without requiring them to use the same predictive model.
3. We evaluate CRAFT on six benchmarks covering RNA, DNA, and protein design under matched online query budgets. After ten rounds of 128 evaluations, CRAFT achieves the highest mean top-128 score on five tasks and ranks second on GFP, counting only newly evaluated sequences. Controlled studies report the largest observed mean gains from history updates, while the effects of selection and source allocation vary across tasks.

2 Problem Formulation

Sequence space and offline data. We consider offline-to-online black-box sequence optimization under a batched evaluation protocol (Jain et al., 2022). Let Σ be a finite alphabet of allowed symbols and L the sequence length. The design space $\mathcal{X} \subseteq \Sigma^L$ contains the sequences admissible for the task, where Σ^L denotes all length-L sequences over Σ. An unknown objective $f : \mathcal{X} \to \mathbb{R}$ assigns each sequence a score to maximize. The learner obtains objective values through costly oracle evaluations: submitting a sequence x returns its score $f(x)$. We assume a deterministic objective, so each sequence has a fixed score. Before online interaction, the learner receives an offline dataset $D_0 = \{(x_i, y_i)\}_{i=1}^{n_0}$ consisting of n_0 sequence–score pairs, where $y_i = f(x_i)$.

Batched online evaluation. The online phase consists of T rounds, each permitting at most B new oracle evaluations, for a total budget of at most TB. At the beginning of round t, the available measurement history is

$$H_{t-1} = D_0 \cup \bigcup_{s=1}^{t-1} \{(x, f(x)) : x \in Q_s\}. \tag{1}$$

Here, Q_s is the set of sequences queried in round s. Thus, H_{t-1} contains the initial offline data and all sequence–score pairs returned in the preceding $t-1$ online rounds, with $H_0 = D_0$.

The learner selects a query batch $Q_t \subseteq \mathcal{X}$ with $|Q_t| \leq B$. Its sequences must be distinct and previously unmeasured: none appears in the offline dataset or any earlier query batch. All decisions use only information available before the current evaluation and internal randomness. The entire batch is fixed before any of its outcomes are observed, so feedback from part of the batch cannot be used to revise the remaining queries within that batch. After evaluation, the history is updated as

$$H_t = H_{t-1} \cup \{(x, f(x)) : x \in Q_t\}. \tag{2}$$

The first batch is therefore guided by offline measurements, while subsequent batches may additionally use feedback from completed evaluations.

Optimization objective. Let $S_t = \bigcup_{s=1}^{t} Q_s$ denote all sequences newly evaluated online through round t, excluding the initial offline designs. For a finite set S and an integer $1 \leq K \leq |S|$, define

$$V_K(S; f) = \max_{\substack{A \subseteq S \\ |A|=K}} \frac{1}{K} \sum_{x \in A} f(x), \qquad J_K = V_K(S_T; f). \tag{3}$$

The maximization identifies the K highest-scoring sequences in S, and $V_K(S; f)$ is their mean objective value. Accordingly, J_K aggregates the queries from all online rounds before computing the top-K mean, rather than averaging round-level scores. Here, B limits the number of queries per round, whereas K specifies the number of high-scoring sequences used for final evaluation; the two need not be equal.

The goal is to maximize J_K subject to $|S_T| = \sum_{t=1}^{T} |Q_t| \leq TB$ and $|S_T| \geq K$. The offline dataset can guide search throughout the online phase, but neither consumes the online query budget nor contributes designs to the final score. The objective therefore measures the quality of newly discovered sequences under a limited evaluation budget, rather than that of designs already present in the offline dataset.

3 CRAFT: Reach, Selection, and Adaptation

CRAFT separates candidate generation from experimental selection. Each source uses its own proposal and filtering rules, while a common selector decides which offered sequences to evaluate. Starting from $H_0 = D_0$, CRAFT constructs candidate pools, selects a query batch, and uses returned measurements to update generation, scoring, and source allocation. This follows the sequential experimental-design principle of using completed measurements to plan subsequent experiments (Box & Wilson, 1992). Figure 1 shows the workflow, and Algorithm 1 in the appendix gives the complete procedure.

Let $\mathcal{A}$ be a finite nonempty set of candidate sources and $\mathcal{X}_{t-1}^{\mathrm{obs}} = \{x : (x, y) \in H_{t-1}\}$ the previously measured sequences. Initial quotas $m_{a,t}$ are set at the start of round t as described in Section 3.3. Generation and selection use only H_{t-1} and internal randomness, without access to current-round outcomes.

3.1 Reach: Candidate Generation

Local-edit sources generate bounded mutations around high-scoring measured sequences, using observed performance to choose where to search rather than relying on an explicit interaction model. For RNA, CRAFT also fits a rank-weighted ridge model. Single-site and pairwise features represent individual and joint contributions of sequence positions, rank weighting emphasizes high-scoring observations, and regularization discourages large coefficients (Hoerl & Kennard, 1970):

$$\begin{aligned} (c_t, \beta_t) &= \arg \min_{c \in \mathbb{R},\, \beta \in \mathbb{R}^d} \sum_{(x_i, y_i) \in H_{t-1}} w_{i,t} \big(y_i - c - \beta^\top \phi(x_i)\big)^2 + \lambda \|\beta\|_2^2, \\ p_t(x) &= c_t + \beta_t^\top \phi(x). \end{aligned} \tag{4}$$

Here $\phi : \mathcal{X} \to \mathbb{R}^d$ contains single-site indicators and pairwise symbol co-occurrences, c_t is an unpenalized intercept, and $\lambda > 0$ controls regularization. Positive weights $w_{i,t}$ follow a prescribed

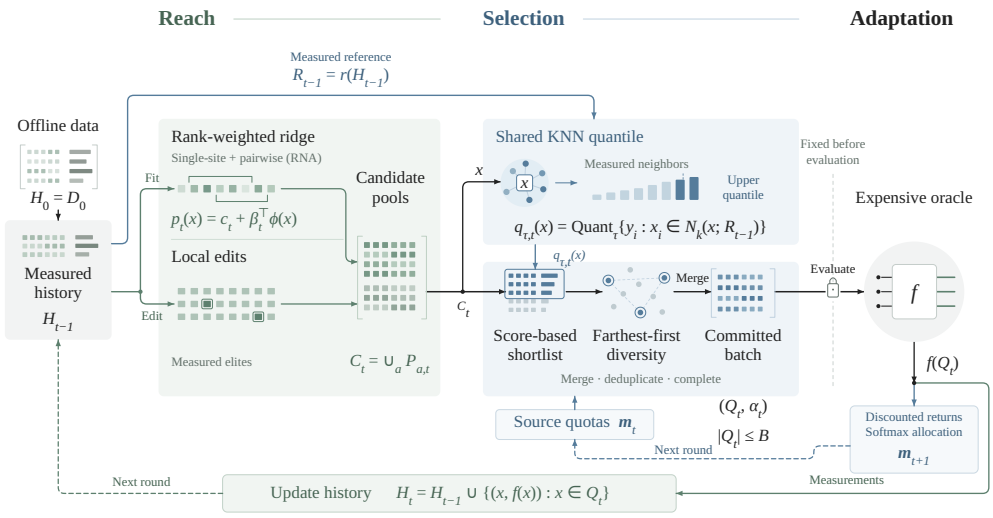

Figure 1: **The CRAFT offline-to-online workflow.** Measured history guides generation and shared scoring. Selection constructs the batch Q_t and source attribution α_t, both fixed before evaluation. Returned measurements update the history and source allocations; dashed arrows indicate next-round feedback.

nonincreasing function of descending measured-score rank, with fixed tie handling. Thus, higher-scoring observations receive no smaller weights. Both the weights and model are recomputed from H_{t-1}.

For RNA, p_t ranks starting sequences from measured elites (high-scoring sequences in the history), empirical symbol-frequency sampling, and uniform sampling. Retained starts undergo bounded local mutations, whose proposals are ranked and truncated by p_t. The complementary local-edit source mutates measured elites without this interaction model, so generation does not depend entirely on its fitted predictions. AAV, GFP, and TFBind use only local-edit sources, differing in enumeration, deduplication, or elite selection. Predictions are never stored as measured labels.

For source a, let g_a include generation, admissibility filtering, and internal truncation, with randomness $\xi_{a,t}$. The offered pools and their union are

$$P_{a,t} = g_a(H_{t-1}; \xi_{a,t}), \qquad |P_{a,t}| \le M_a, \qquad C_t = \bigcup_{a \in \mathcal{A}} P_{a,t}. \tag{5}$$

Each $P_{a,t} \subseteq \mathcal{X}$ contains distinct sequences. The integer cap $M_a \ge 1$ limits offered candidates, not paid queries or all internally considered proposals. The union removes duplicates while retaining source memberships. Generation uses no new oracle calls, and every candidate used for batch completion must already belong to C_t.

3.2 Selection: Batch Construction

Shared scoring. Source-internal scores need not be comparable. CRAFT therefore scores all candidates against the same measured reference rather than combining the sources' internal score scales. To bound neighbor-search size, $R_{t-1} = r(H_{t-1})$ retains the full history up to an integer capacity $M_{\text{ref}} \ge 1$, and its highest-measured-score M_{ref} records otherwise, with fixed tie handling. This restriction does not delete observations from H_{t-1}. For sequences of length L, define

$$\begin{aligned} d_H(x, z) &= \frac{1}{L} \sum_{j=1}^{L} \mathbf{1}\{x_j \ne z_j\}, \\ q_{\tau,t}(x) &= \text{Quant}_\tau\{y_i : x_i \in N_k(x; R_{t-1})\}. \end{aligned} \tag{6}$$

Here $\mathbf{1}$ is the indicator function, $N_k(x; R_{t-1})$ contains the k nearest measured sequences under normalized Hamming distance d_H, and $\tau \in [0, 1]$ is the quantile level. We require $1 \le k \le$

$\min\{n_0, M_{\text{ref}}\}$, where $n_0 = |D_0| \geq 1$; distance ties follow a fixed reference order. For sorted neighbor labels $v_{[0]} \leq \cdots \leq v_{[k-1]}$, including repeats, we use linear interpolation (Hyndman & Fan, 1996):

$$\begin{aligned} j &= \lfloor (k-1)\tau \rfloor, \qquad \theta = (k-1)\tau - j, \\ q_{\tau,t}(x) &= (1-\theta)v_{[j]} + \theta v_{[\min\{j+1,\, k-1\}]}. \end{aligned} \tag{7}$$

A measured neighborhood can contain both high- and low-scoring variants. We use an upper quantile to favor its higher observed values rather than its average, directing selection toward candidates near high-scoring measurements. This is a selection preference, not a calibrated confidence bound or a guarantee of more accurate ranking. We use $k = 16$, $M_{\text{ref}} = 12{,}000$, and $\tau = 0.95$, except on AAV, where $\tau = 0.85$.

Diversity. Selecting only by score can concentrate queries on similar sequences, whereas selecting only by distance need not favor high-scoring regions. We therefore prioritize candidates by score and then use farthest-first selection to spread evaluations within promising sets, following the maximin-distance principle (Johnson et al., 1990). For a finite set U, score function s, and integer $m \geq 0$, $\text{Top}_m(U; s)$ retains the $\min\{m, |U|\}$ highest-scoring distinct elements with prescribed tie handling. Choose an integer shortlist size satisfying $\min\{m_{a,t}, |P_{a,t}|\} \leq \ell_{a,t} \leq |P_{a,t}|$ and set

$$E_{a,t} = \text{Top}_{\ell_{a,t}}(P_{a,t}; q_{\tau,t}). \tag{8}$$

Let W be the working candidate set and $\varnothing \neq U \subseteq W$ the already selected candidates. Starting from a prescribed candidate, farthest-first traversal adds

$$x^{\text{next}} \in \arg\max_{x \in W \setminus U} \min_{z \in U} d_H(x, z) \tag{9}$$

until the required number is selected or W is exhausted. Depending on source settings, W is the high-score shortlist or the entire cutoff-tie group. In the latter case, diversity resolves the tie before $E_{a,t}$ is finalized, while candidates strictly above the cutoff retain priority. This greedy procedure does not guarantee a globally optimal maximin batch.

Batch construction. Source lists prioritize retained candidates and keep remaining pool entries for completion. We merge them in the configured source order toward initial quotas, skipping measured and duplicate sequences. Unused slots are filled from eligible candidates in C_t until B queries are selected or none remain. Let $\alpha_t : Q_t \to \mathcal{A}$ assign each query to its supplying source. Then

$$Q_t \subseteq C_t \setminus \mathcal{X}_{t-1}^{\text{obs}}, \qquad |Q_t| \leq B, \qquad x \in P_{\alpha_t(x),t} \quad (x \in Q_t). \tag{10}$$

Each sequence is evaluated and attributed once. Completion may change final source counts relative to initial quotas; the procedure is not global top-B selection over C_t. Both Q_t and α_t are fixed before any current-batch outcome is observed.

3.3 Adaptation: History and Allocation Updates

After all outcomes return, set $H_t = H_{t-1} \cup \{(x, f(x)) : x \in Q_t\}$. The expanded history is used to refit proposal models, refresh measured elites, and rebuild the shared scoring reference. New measurements can therefore change which candidates are generated and how they are ranked, while offline observations remain available.

Because source performance can change as search progresses, CRAFT adjusts future allocations using recent evaluated outcomes. For $Q_t^a = \{x \in Q_t : \alpha_t(x) = a\}$, define

$$C_{a,t} = \begin{cases} |Q_t^a|^{-1} \displaystyle\sum_{x \in Q_t^a} h_t(f(x)), & |Q_t^a| > 0, \\ 0, & |Q_t^a| = 0. \end{cases} \tag{11}$$

Averaging makes this a per-query return rather than a total that depends directly on the number of attributed queries. The prescribed h_t is the identity or $h_t(y) = (y - \mu_t)/\max\{\sigma_t, \varepsilon\}$, where μ_t and σ_t are the completed batch's mean and population standard deviation, and $\varepsilon > 0$ is fixed. An

empty batch gives zero returns without computing these statistics. The transformation affects only allocation, not stored oracle scores; returns summarize evaluated queries, not entire pools.

Initialize return sums $S_{a,1} = s_0$ and discounted round counts $N_{a,1} = 1$ identically across sources, with fixed $s_0 \in \mathbb{R}$. These scalar statistics differ from the cumulative query set S_t. After each round, update

$$S_{a,t+1} = \rho S_{a,t} + C_{a,t}, \qquad N_{a,t+1} = \rho N_{a,t} + 1, \qquad \rho \in (0, 1]. \tag{12}$$

For $\rho < 1$, discounting gives recent returns greater influence. Every source count advances each round, including sources with no attributed queries. At the start of round t, compute allocation weights and quotas:

$$\pi_{a,t} = \frac{\exp(S_{a,t}/(\eta N_{a,t}))}{\sum_{b\in\mathcal{A}} \exp(S_{b,t}/(\eta N_{b,t}))}, \qquad \boldsymbol{m}_t = \mathrm{Quota}_B(\boldsymbol{\pi}_t). \tag{13}$$

Here $\boldsymbol{\pi}_t = (\pi_{a,t})_{a\in\mathcal{A}}$, $\boldsymbol{m}_t = (m_{a,t})_{a\in\mathcal{A}}$, and the temperature $\eta > 0$ controls how strongly mean-return differences affect allocation weights. The prescribed rounding map Quota_B enforces $\sum_a m_{a,t} = B$ and $m_{a,t} \geq m_{\min}$ for a fixed integer $m_{\min} \geq 1$, requiring $|\mathcal{A}|m_{\min} \leq B$. The minimum reserves an initial allocation opportunity, not a guaranteed final count after merging and completion. Weighting, proposal, diversity, tie, completion, and rounding rules are fixed; task settings are reported in the appendix.

Pool caps M_a limit offered candidates, while B limits new oracle evaluations per round. The objective $J_K = V_K(S_T; f)$ uses only newly evaluated sequences, excluding initial offline designs.

4 Theoretical Analysis

We analyze how selection retains candidate quality and how retained improvements accumulate across rounds. Additional results on candidate screening, reference coverage, and diversity appear in Appendix E.

Let $\mathcal{F}_t^{(K)}$ contain all previously unmeasured batches from C_t with sizes between K and B. For $Q_t \in \mathcal{F}_t^{(K)}$, define

$$\begin{aligned} L_{K,t} &= V_K(C_t; f) - V_K(Q_t; f), \\ \Delta_{\mathcal{F}_t} &= V_K(C_t; f) - \max_{Q\in\mathcal{F}_t^{(K)}} V_K(Q; f), \\ r_{q,t} &= \max_{Q\in\mathcal{F}_t^{(K)}} V_K(Q; q_{\tau,t}) - V_K(Q_t; q_{\tau,t}). \end{aligned}$$

These quantities measure selection loss, eligibility loss, and the executed batch's gap under the shared score, respectively.

Theorem 1 (Candidate quality retention)**.** *If* $\max_{x\in C_t} |q_{\tau,t}(x) - f(x)| \leq \epsilon_t$*, then*

$$0 \leq L_{K,t} \leq \Delta_{\mathcal{F}_t} + 2\epsilon_t + r_{q,t}. \tag{14}$$

Proof. Uniform scoring error at most ϵ_t changes the top-K mean of any feasible batch by at most ϵ_t. Hence,

$$\begin{aligned} V_K(Q_t; f) &\geq V_K(Q_t; q_{\tau,t}) - \epsilon_t \\ &\geq \max_{Q\in\mathcal{F}_t^{(K)}} V_K(Q; f) - r_{q,t} - 2\epsilon_t \\ &= V_K(C_t; f) - \Delta_{\mathcal{F}_t} - r_{q,t} - 2\epsilon_t. \end{aligned}$$

Rearrangement gives the upper bound. Nonnegativity follows from $Q_t \subseteq C_t$. □

Quota execution, screening, diversity ordering, and completion contribute to $r_{q,t}$. The bound therefore separates scoring error from losses introduced by batch construction.

For $t \geq 2$, define the available and unretained improvements:

$$\begin{aligned} O_{K,t} &= V_K(S_{t-1} \cup C_t; f) - V_K(S_{t-1}; f), \\ U_{K,t} &= V_K(S_{t-1} \cup C_t; f) - V_K(S_t; f). \end{aligned}$$

Table 1: **Online optimization.** J_{128} after 1,280 new queries (mean ± SEM, eight seeds 112–119; higher is better).

Method	RNA1	RNA2	RNA3	AAV	GFP	TFBind
GFN-AL	0.5111±0.0039	0.4957±0.0028	0.5084±0.0035	0.0000±0.0000	1.6320±0.0029	0.9245±0.0046
AdaLead	0.8963±0.0116	0.8921±0.0126	0.9206±0.0238	0.6495±0.0045	3.5706±0.0012	0.9665±0.0010
PEX	0.7720±0.0085	0.6375±0.0012	0.7387±0.0085	0.5835±0.0021	3.5922±0.0005	0.9505±0.0009
CbAS	0.5523±0.0056	0.5594±0.0054	0.5668±0.0076	0.2047±0.0052	3.3766±0.0146	0.8912±0.0060
DbAS	0.5645±0.0038	0.5491±0.0081	0.5547±0.0052	0.2047±0.0052	3.3761±0.0149	0.8840±0.0054
BO	0.5507±0.0102	0.5499±0.0066	0.5372±0.0057	0.0104±0.0018	2.1397±0.1350	0.8644±0.0034
GFN-AL+δ-CS	<u>0.9211</u>±0.0129	<u>0.9313</u>±0.0041	<u>0.9468</u>±0.0236	0.6438±0.0054	3.5770±0.0017	<u>0.9713</u>±0.0015
MLDE (adapted)	0.8518±0.0035	0.8567±0.0161	0.8345±0.0179	0.5892±0.0124	3.5894±0.0024	0.9420±0.0011
ProSpero	0.8186±0.0193	0.8370±0.0157	0.8298±0.0243	<u>0.6731</u>±0.0058	**3.6549**±0.0042	0.9386±0.0021
CRAFT	**0.9397**±0.0033	**0.9385**±0.0021	**0.9878**±0.0039	**0.6768**±0.0070	<u>3.6127</u>±0.0005	**0.9716**±0.0008

Bold/<u>underline</u>: highest/second-highest eligible mean, not statistical significance. Scores retain task-specific oracle scales. GFN-AL on AAV is a repository extension; the separate MLDE TFBind extension is excluded from rank markings.

Theorem 2 (Cumulative discovery)**.** *Suppose* $|Q_1| \geq K$. *Then*

$$J_K = V_K(Q_1; f) + \sum_{t=2}^{T} \big(O_{K,t} - U_{K,t}\big). \tag{15}$$

Whenever $|Q_t| \geq K$,

$$0 \leq U_{K,t} \leq \min\{O_{K,t}, L_{K,t}\}.$$

If the scoring assumption of Theorem 1 also holds, then

$$\begin{aligned} &V_K(S_t; f) - V_K(S_{t-1}; f) \\ &\quad \geq \big[O_{K,t} - \Delta_{\mathcal{F}_t} - 2\epsilon_t - r_{q,t}\big]_+, \end{aligned} \tag{16}$$

where $[u]_+ = \max\{u, 0\}$.

Proof. Subtracting the definitions gives the round-wise improvement $O_{K,t} - U_{K,t}$. Summing over rounds proves Eq. equation 15. Set inclusion gives $0 \leq U_{K,t} \leq O_{K,t}$. The diminishing-return property of the top-K mean gives $U_{K,t} \leq L_{K,t}$ when $|Q_t| \geq K$ (Appendix E.3). Combining these inequalities with Theorem 1 and the nonnegativity of cumulative improvement proves Eq. equation 16. □

Positive improvement is guaranteed when the available opportunity exceeds the retention-loss bound. History and allocation updates affect both quantities, so their benefits must be assessed along their own feedback trajectories. Unqueried objective values are used only for post-hoc diagnostics and never enter the search policy.

5 Experiments

We evaluate CRAFT on six sequence-design landscapes—RNA1–RNA3, AAV, GFP, and TFBind—covering RNA, protein, and DNA design (Sinai et al., 2020; Jain et al., 2022; Trabucco et al., 2022). Each method starts from offline measurements and receives ten rounds of 128 new oracle queries. We report J_{128} as mean ± SEM across eight algorithm seeds, excluding initial designs. Oracles and online query budgets are matched; candidate-generation computation follows each implementation.

Comparators include BO and AdaLead (Sinai et al., 2020), CbAS/DbAS (Brookes et al., 2019; Brookes & Listgarten, 2018), PEX (Ren et al., 2022), GFN-AL (Jain et al., 2022), GFN-AL+δ-CS (Kim et al., 2024), ProSpero (Kmicikiewicz et al., 2026), and the documented online MLDE-Population adaptation (Wu et al., 2019; Wittmann et al., 2021). Datasets, implementations, baseline adaptations, and evaluation protocols are provided in Appendix B.

5.1 Main Results

CRAFT achieves the highest mean on five of six tasks and ranks second to ProSpero on GFP (Table 1). On RNA3, CRAFT exceeds the strongest competing mean by 0.0410, while the margins on AAV and TFBind are small. These results demonstrate competitive discovery under matched online query budgets, with task-dependent gains.

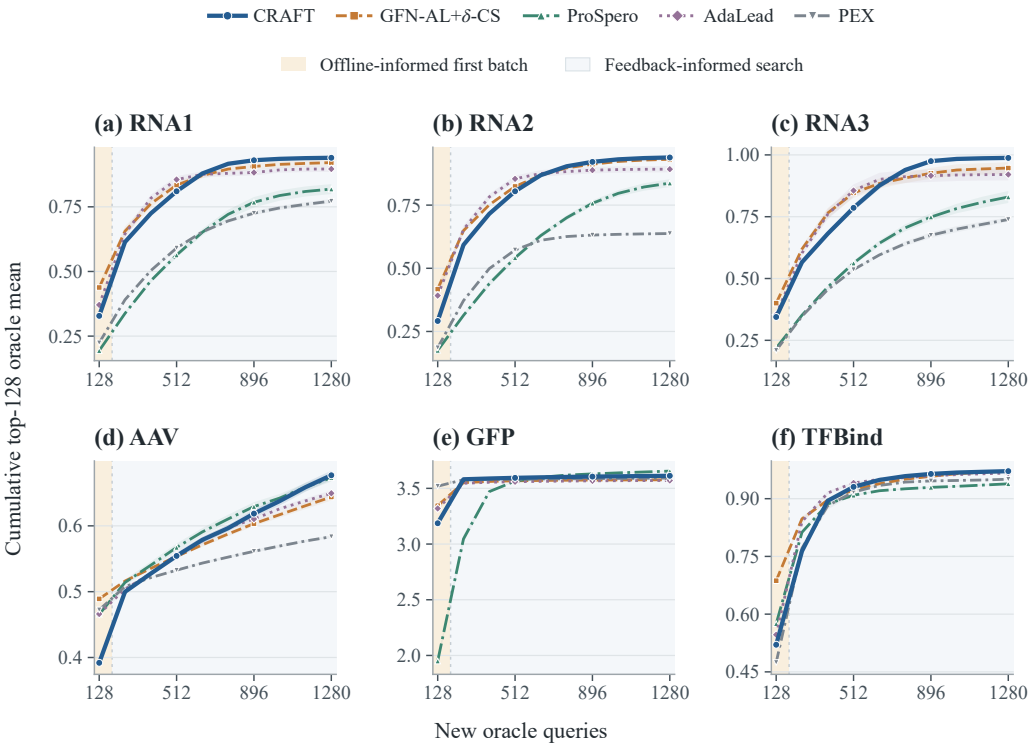

Figure 2: **Discovery over the online budget.** Cumulative $V_{128}(S_t; f)$ versus completed new queries $(128t)$. Lines and bands show the mean and one SEM across eight seeds. The first batch uses offline measurements; later batches can incorporate returned outcomes. Only methods with complete recorded trajectories are shown.

GFN-AL+δ-CS has higher first-batch quality on all six tasks, yet CRAFT overtakes it at 256–768 queries, depending on the task, and finishes ahead on each task (Figure 2). This distinguishes offline-informed initialization from subsequent retained improvements, as in Eq. equation 15; the matched interventions below examine the role of feedback.

5.2 Ablations and Mechanisms

On RNA3, AAV, and GFP, we independently vary history updates and source allocation using eight label-blind 50% subsets of the offline reservoir, with two repeats per subset. The four branches share their initial pools, first batch, and source assignments, then diverge from round two. Table 2 also includes direct querying, which bypasses common selection and diversity while retaining history and allocation updates, and a development-selected single source with matched offered-candidate capacity. These reduced-support runs are separate from the main comparison.

Feedback changes subsequent candidate quality. History updates produce the larger observed mean gains, while allocation adds task-dependent improvements. With fixed allocation, updating history raises the RNA3 mean from 0.5655 to 0.9206; adaptive allocation further increases it to 0.9658 (Table 2). Updated histories also produce stronger later candidate pools, reflecting the joint

Table 2: **Feedback and component ablations.** Final J_{128} under 50% offline support (mean ± SEM across eight archive-level means, after averaging two repeats per archive).

History	Allocation / control	RNA3	AAV	GFP
Frozen	Fixed	0.5655 ±0.0030	0.5009 ±0.0008	3.5715 ±0.0016
Updated	Fixed	0.9206 ±0.0170	0.6248 ±0.0046	3.6098 ±0.0008
Frozen	Adaptive	0.6058 ±0.0030	0.5044 ±0.0009	3.5727 ±0.0015
Updated	Adaptive (CRAFT)	<u>0.9658</u> ±0.0072	**0.6604** ±0.0053	**3.6121** ±0.0004
Updated	Direct query	**0.9676** ±0.0061	<u>0.6556</u> ±0.0035	<u>3.6113</u> ±0.0008
Updated	Best single source	0.8692 ±0.0231	0.6554 ±0.0038	3.6108 ±0.0006

Shading identifies the full loop. **Bold**/<u>underline</u> indicate the highest/second-highest means, not statistical significance. Uncertainty is conditional on the fixed offline reservoir.

response of generation and scoring to new measurements (Appendix C.2). The capacity-matched single-source control remains weaker on RNA3, with smaller differences on AAV and GFP. The nine primary feedback contrasts do not reach family-wise significance at 0.05 after Holm correction (Holm, 1979); complete protocols, paired effects, and interaction analyses appear in Appendix C.2.

Shared selection along online trajectories. We evaluate selection rules along their own complete feedback trajectories under matched query budgets. On AAV with full offline support, the quantile selector improves both first-batch quality and final discovery relative to native ordering with the same diversity procedure, with consistent directional evidence in an independent ten-seed replication. On GFP, improvements are concentrated in the first batch, while RNA3 shows no consistent endpoint gain. The benefits of shared selection therefore depend on the task and evaluation budget, and alternative scorers can perform better. Full online comparisons and the independent AAV replication appear in Appendices C.4 and C.5; fixed-pool diagnostics and computational costs appear in Appendix C.3. We additionally examine whether high-scoring discoveries are diverse within the selected set and novel relative to the initial offline measurements.

Supplementary analyses of discovery diversity and novelty, including metric definitions and the scope of archived-record comparisons, appear in Appendix D.

6 Discussion

CRAFT separates candidate generation from experimental selection, using offline measurements and online feedback to guide candidate construction, shared scoring, and source allocation. By making candidate pools explicit, it distinguishes the quality available from search from the quality retained in evaluated batches. Our theoretical analysis connects retained candidate quality to cumulative discovery under bounded scoring error. Experiments on six benchmarks demonstrate competitive performance under matched online query budgets. The framework also supports controlled comparisons of selection rules along their own feedback trajectories, allowing their effects on immediate batch quality and subsequent discovery to be assessed separately.

7 Limitations.

CRAFT depends on measurement coverage, candidate representations, and the suitability of the shared score; the quantile rule is not universally optimal. Our analysis applies to finite candidate pools and provides no global search guarantee. The evaluation assumes fixed-length sequences, synchronous batches, and fixed query-count budgets. Robustness to laboratory noise and effectiveness in distant sequence regions remain to be established.

AI USE STATEMENT

We used generative AI tools to improve the phrasing and readability of the manuscript and to correct grammatical errors. The authors reviewed and revised the AI-assisted text to preserve the intended scientific meaning and ensure consistency with the methods, results, and claims presented in the paper. We take full responsibility for the final content of this work, including all AI-assisted text and any associated claims or artifacts.

ETHICS STATEMENT

This work studies design optimization using existing benchmark datasets and computational oracles. The reported evaluation is computational and does not involve recruiting human participants or conducting new wet-laboratory experiments. Benchmark performance should not be interpreted as evidence of biological safety, experimental feasibility, or clinical efficacy. Although the framework is intended to support scientific discovery, applications to biological sequence design may raise dual-use concerns. Any deployment involving biological synthesis or real-world testing should therefore include application-specific risk assessment, appropriate biosafety oversight, and independent experimental validation.

REPRODUCIBILITY STATEMENT

The main text and supplementary material describe the proposed method, experimental settings, and evaluation procedures. The accompanying code provides the CRAFT implementation and instructions for its use. Additional experimental details and statistical reporting are provided in the appendix, together with the assumptions and proofs supporting the theoretical results. These materials document the scope of the evaluation and support reproduction and further investigation of the method.

REFERENCES

Christof Angermueller, David Belanger, Andreea Gane, Zelda Mariet, David Dohan, Kevin Murphy, Lucy Colwell, and D Sculley. Population-based black-box optimization for biological sequence design. In *International conference on machine learning*, pp. 324–334. PMLR, 2020.

Philip J Ball, Laura Smith, Ilya Kostrikov, and Sergey Levine. Efficient online reinforcement learning with offline data. In *International Conference on Machine Learning*, pp. 1577–1594. PMLR, 2023.

George EP Box and Kenneth B Wilson. On the experimental attainment of optimum conditions. In *Breakthroughs in statistics: methodology and distribution*, pp. 270–310. Springer, 1992.

David Brookes, Hahnbeom Park, and Jennifer Listgarten. Conditioning by adaptive sampling for robust design. In *International conference on machine learning*, pp. 773–782. PMLR, 2019.

David H Brookes and Jennifer Listgarten. Design by adaptive sampling. *arXiv preprint arXiv:1810.03714*, 2018.

Can Chen, Yingxueff Zhang, Jie Fu, Xue Steve Liu, and Mark Coates. Bidirectional learning for offline infinite-width model-based optimization. *Advances in Neural Information Processing Systems*, 35:29454–29467, 2022a.

Can Chen, Yingxueff Zhang, Jie Fu, Xue Steve Liu, and Mark Coates. Bidirectional learning for offline infinite-width model-based optimization. *Advances in Neural Information Processing Systems*, 35:29454–29467, 2022b.

Cuong Dao, Phi Le Nguyen, Truong Thao Nguyen, Nghia Hoang, et al. Root: Rethinking offline optimization as distributional translation via probabilistic bridge. *Advances in Neural Information Processing Systems*, 38:98783–98817, 2026.

Thomas Desautels, Andreas Krause, and Joel W Burdick. Parallelizing exploration-exploitation tradeoffs in gaussian process bandit optimization. *The Journal of Machine Learning Research*, 15 (1):3873–3923, 2014.

David Eriksson, Michael Pearce, Jacob Gardner, Ryan D Turner, and Matthias Poloczek. Scalable global optimization via local bayesian optimization. *Advances in neural information processing systems*, 32, 2019.

Clara Fannjiang and Jennifer Listgarten. Autofocused oracles for model-based design. *Advances in Neural Information Processing Systems*, 33:12945–12956, 2020.

Adam Foster, Desi R Ivanova, Ilyas Malik, and Tom Rainforth. Deep adaptive design: Amortizing sequential bayesian experimental design. In *International conference on machine learning*, pp. 3384–3395. PMLR, 2021.

Javier González, Zhenwen Dai, Philipp Hennig, and Neil Lawrence. Batch bayesian optimization via local penalization. In *Artificial intelligence and statistics*, pp. 648–657. PMLR, 2016.

Nate Gruver, Samuel Stanton, Nathan Frey, Tim GJ Rudner, Isidro Hotzel, Julien Lafrance-Vanasse, Arvind Rajpal, Kyunghyun Cho, and Andrew G Wilson. Protein design with guided discrete diffusion. *Advances in neural information processing systems*, 36:12489–12517, 2023.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

Matthew Hoffman, Eric Brochu, Nando De Freitas, et al. Portfolio allocation for bayesian optimization. In *UAI*, volume 11, pp. 327–336, 2011.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pp. 65–70, 1979.

Rob J Hyndman and Yanan Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365, 1996.

Moksh Jain, Emmanuel Bengio, Alex Hernandez-Garcia, Jarrid Rector-Brooks, Bonaventure FP Dossou, Chanakya Ajit Ekbote, Jie Fu, Tianyu Zhang, Michael Kilgour, Dinghuai Zhang, et al. Biological sequence design with gflownets. In *International conference on machine learning*, pp. 9786–9801. PMLR, 2022.

Mark E Johnson, Leslie M Moore, and Donald Ylvisaker. Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2):131–148, 1990.

Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

Hyeonah Kim, Minsu Kim, Taeyoung Yun, Sanghyeok Choi, Emmanuel Bengio, Alex Hernández-García, and Jinkyoo Park. Improved off-policy reinforcement learning in biological sequence design. *arXiv preprint arXiv:2410.04461*, 2024.

Michal Kmicikiewicz, Vincent Fortuin, and Ewa Szczurek. Prospero: Active learning for robust protein design beyond wild-type neighborhoods. *Advances in Neural Information Processing Systems*, 38:131015–131048, 2026.

Siddarth Krishnamoorthy, Satvik Mehul Mashkaria, and Aditya Grover. Diffusion models for black-box optimization. In *International Conference on Machine Learning*, pp. 17842–17857. PMLR, 2023.

Seunghyun Lee, Younggyo Seo, Kimin Lee, Pieter Abbeel, and Jinwoo Shin. Offline-to-online reinforcement learning via balanced replay and pessimistic q-ensemble. In *Conference on Robot Learning*, pp. 1702–1712. PMLR, 2022.

Satvik Mehul Mashkaria, Siddarth Krishnamoorthy, and Aditya Grover. Generative pretraining for black-box optimization. In *International Conference on Machine Learning*, pp. 24173–24197. PMLR, 2023.

Mitsuhiko Nakamoto, Simon Zhai, Anikait Singh, Max Sobol Mark, Yi Ma, Chelsea Finn, Aviral Kumar, and Sergey Levine. Cal-ql: Calibrated offline rl pre-training for efficient online fine-tuning. *Advances in Neural Information Processing Systems*, 36:62244–62269, 2023.

Zhizhou Ren, Jiahan Li, Fan Ding, Yuan Zhou, Jianzhu Ma, and Jian Peng. Proximal exploration for model-guided protein sequence design. In *International Conference on Machine Learning*, pp. 18520–18536. PMLR, 2022.

Sam Sinai, Richard Wang, Alexander Whatley, Stewart Slocum, Elina Locane, and Eric D Kelsic. Adalead: A simple and robust adaptive greedy search algorithm for sequence design. *arXiv preprint arXiv:2010.02141*, 2020.

Jasper Snoek, Hugo Larochelle, and Ryan Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

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 *International conference on machine learning*, pp. 20459–20478. PMLR, 2022.

Rong-Xi Tan, Ke Xue, Shen-Huan Lyu, Haopu Shang, Yao Wang, Yaoyuan Wang, Fu Sheng, and Chao Qian. Offline model-based optimization by learning to rank. In *International Conference on Learning Representations*, volume 2025, pp. 47623–47650, 2025.

Brandon Trabucco, Aviral Kumar, Xinyang Geng, and Sergey Levine. Conservative objective models for effective offline model-based optimization. In *International Conference on Machine Learning*, pp. 10358–10368. PMLR, 2021.

Brandon Trabucco, Xinyang Geng, Aviral Kumar, and Sergey Levine. Design-bench: Benchmarks for data-driven offline model-based optimization. In *International Conference on Machine Learning*, pp. 21658–21676. PMLR, 2022.

Bruce J Wittmann, Yisong Yue, and Frances H Arnold. Informed training set design enables efficient machine learning-assisted directed protein evolution. *Cell systems*, 12(11):1026–1045, 2021.

Zachary Wu, SB Jennifer Kan, Russell D Lewis, Bruce J Wittmann, and Frances H Arnold. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proceedings of the National Academy of Sciences*, 116(18):8852–8858, 2019.

Jason Yang, Ravi G Lal, James C Bowden, Raul Astudillo, Mikhail A Hameedi, Sukhvinder Kaur, Matthew Hill, Yisong Yue, and Frances H Arnold. Active learning-assisted directed evolution. *Nature Communications*, 16(1):714, 2025.

Michael S Yao, Yimeng Zeng, Hamsa Bastani, Jacob Gardner, James C Gee, and Osbert Bastani. Generative adversarial model-based optimization via source critic regularization. *Advances in neural information processing systems*, 37:44009–44039, 2024.

Sihyun Yu, Sungsoo Ahn, Le Song, and Jinwoo Shin. Roma: Robust model adaptation for offline model-based optimization. *Advances in Neural Information Processing Systems*, 34:4619–4631, 2021.

Zishun Yu and Xinhua Zhang. Actor-critic alignment for offline-to-online reinforcement learning. In *International Conference on Machine Learning*, pp. 40452–40474. PMLR, 2023.

Haichao Zhang, We Xu, and Haonan Yu. Policy expansion for bridging offline-to-online reinforcement learning. *arXiv preprint arXiv:2302.00935*, 2023.

A Related Work

Offline model-based optimization. Offline model-based optimization seeks high-scoring designs from fixed measurements (Trabucco et al., 2022). CbAS adapts a conditional design distribution, while autofocused oracles reweight surrogate training to follow the evolving search distribution (Brookes et al., 2019; Fannjiang & Listgarten, 2020). COMs, RoMA, and GABO address unreliable extrapolation through conservative modeling, robust model adaptation, and source critic regularization, respectively (Trabucco et al., 2021; Yu et al., 2021; Yao et al., 2024) Bidirectional learning transfers information from the offline dataset through forward and backward mappings (Chen et al., 2022a;b), whereas RaM learns candidate rankings rather than relying on pointwise regression (Tan et al., 2025). Generative approaches include conditional diffusion in DDOM and distributional translation in ROOT (Krishnamoorthy et al., 2023; Dao et al., 2026); BONET further learns optimization trajectories from offline data for search with a limited online evaluation budget (Mashkaria et al., 2023). CRAFT focuses on how offline-informed proposals become online experiments: offline measurements support both candidate construction and a shared reference for selection, while returned measurements update both in subsequent rounds.

Online sequence optimization and active learning. Machine learning-assisted directed evolution uses measured variants to guide screening, with informed training-set construction and iterative active learning improving how experimental data are collected and reused (Wu et al., 2019; Wittmann et al., 2021; Yang et al., 2025). AdaLead and PEX perform model-guided evolutionary search (Sinai et al., 2020; Ren et al., 2022), while LaMBO and LaMBO-2 combine Bayesian optimization with denoising autoencoders and guided discrete diffusion, respectively (Stanton et al., 2022; Gruver et al., 2023). GFN-AL combines diverse generation with uncertainty-aware acquisition and explicitly uses existing labeled sequences to train its generator (Jain et al., 2022). More recently, δ-Conservative Search perturbs high-scoring offline sequences according to proxy uncertainty (Kim et al., 2024), and ProSpero combines a pretrained generator with a surrogate updated from experimental feedback (Kmicikiewicz et al., 2026). CRAFT's distinction lies in separating source-specific generation from common selection: candidates from bounded source pools are compared using the same measurement-based score before quota-constrained batch construction. This supports separate evaluation of candidate quality and queried quality under the same measured history.

Design of experiments and batch selection. Sequential design of experiments (DOE) uses completed measurements to plan subsequent experiments, as exemplified by response-surface optimization (Box & Wilson, 1992). Bayesian optimization selects evaluations through acquisition functions based on a probabilistic surrogate (Jones et al., 1998; Snoek et al., 2012), while Deep Adaptive Design learns history-conditioned experimental policies using information-based objectives (Foster et al., 2021). For parallel evaluations, GP-BUCB accounts for pending feedback, and local penalization discourages redundant queries within a batch (Desautels et al., 2014; González et al., 2016). CRAFT adopts the sequential DOE perspective to connect offline measurements with budgeted online experiments, but focuses on how much candidate quality is retained in the evaluated batch. For fixed pools under bounded scoring error, its top-K analysis separates the effects of batch constraints, scoring error, and approximate selection, rather than deriving a new Bayesian acquisition criterion.

Optimizer portfolios and adaptive allocation. Portfolio methods distribute evaluations across multiple search strategies. GP-Hedge selects among acquisition functions using an online bandit strategy (Hoffman et al., 2011), while TuRBO allocates evaluations across local Bayesian optimization models (Eriksson et al., 2019). P3BO constructs experimental batches by sampling proposals from an optimizer portfolio, shares returned measurements among its members, and updates sampling weights using discounted performance (Angermueller et al., 2020). CRAFT exposes bounded proposal pools to an explicit candidate-selection stage. A shared score derived from the measured history prioritizes candidates within source quotas, followed by diversity handling and deduplication. This interface supports separate evaluation of the quality available in the proposal pools and the quality retained in the queried batch, including controlled substitutions of the selection rule. Any source-specific scores guide generation and internal truncation, whereas shared scoring determines selection within source quotas. This separates the quality supplied by a source from the quality of its evaluated proposals, enabling controlled comparisons of generation, selection, and subsequent allocation.

Offline-to-online learning. Offline-to-online reinforcement learning studies how prior experience can support further interaction. Balanced replay and RLPD reuse offline samples alongside newly collected data (Lee et al., 2022; Ball et al., 2023); Policy Expansion preserves an offline policy while introducing new trainable policies (Zhang et al., 2023). Actor–critic alignment and Cal-QL address value-function mismatch during the transition to online learning (Yu & Zhang, 2023; Nakamoto et al., 2023). CRAFT instead operates on measured sequence–score pairs, using offline data as both an input to candidate generation and a reference for common selection. Online outcomes augment this shared history and revise future source allocations. Its offline-to-online contribution therefore concerns the use of prior measurements in budgeted design decisions and candidate-to-query quality retention, rather than policy preservation or transferred-critic calibration.

B Experimental Setup and Supplementary Results

Algorithm 1 summarizes the procedure in Section 3; task-specific settings are reported below.

Algorithm 1 CRAFT: Offline-to-Online Sequence Design

Require: Offline data D_0, oracle f, sources $\mathcal{A}$, rounds T, batch budget B, and fixed source, scoring, and allocation settings.
Ensure: Online discoveries S_T and measured history H_T.

$H_0 \leftarrow D_0$; $S_0 \leftarrow \varnothing$; $(S_{a,1}, N_{a,1}) \leftarrow (s_0, 1)$ for every $a \in \mathcal{A}$.
for $t = 1, \dots, T$ **do**
  Compute allocation weights $\boldsymbol{\pi}_t$ and quotas $\boldsymbol{m}_t$ using Eq. (13).
  Reach: construct offered pools.
  Refit explicit priors where used and refresh measured elites from H_{t-1} (Section 3.1).
  Construct $P_{a,t} = g_a(H_{t-1}; \xi_{a,t})$ with $|P_{a,t}| \le M_a$ for every source, without new oracle calls.
  $C_t \leftarrow \bigcup_{a \in \mathcal{A}} P_{a,t}$; retain each candidate's source memberships.
  Selection: compare and commit queries.
  $R_{t-1} \leftarrow r(H_{t-1})$; $\mathcal{X}^{\mathrm{obs}}_{t-1} \leftarrow \{x : (x, y) \in H_{t-1}\}$.
  Compute $q_{\tau,t}(x)$ for all $x \in C_t$ from measured neighbors using Eqs. (6)–(7).
  for all $a \in \mathcal{A}$ **do**
    Choose $\ell_{a,t}$ with $\min\{m_{a,t}, |P_{a,t}|\} \le \ell_{a,t} \le |P_{a,t}|$.
    Construct $E_{a,t}$ and source priorities by score screening and farthest-first selection, with the prescribed cutoff-tie handling (Eqs. (8)–(9)).
    Keep remaining pool entries available for completion.
  end for
  Merge source lists toward $\boldsymbol{m}_t$ in the configured order, skipping measured and duplicate designs.
  Complete Q_t from remaining eligible entries in C_t until $|Q_t| = B$ or none remain; attribute each query once to its supplying source via α_t.
  Commit (Q_t, α_t) before observing any current outcome.
  Adaptation: incorporate completed measurements.
  Evaluate each $x \in Q_t$ once and wait for all outcomes $f(x)$.
  $H_t \leftarrow H_{t-1} \cup \{(x, f(x)) : x \in Q_t\}$; $S_t \leftarrow S_{t-1} \cup Q_t$.
  for all $a \in \mathcal{A}$ **do**
    Compute the attributed return $C_{a,t}$ using Eq. (11) and the configured whole-batch transformation; use zero for an empty attributed set.
    $S_{a,t+1} \leftarrow \rho S_{a,t} + C_{a,t}$; $N_{a,t+1} \leftarrow \rho N_{a,t} + 1$.
  end for
end for
return (S_T, H_T).

B.1 Benchmarks and Evaluation Protocol

We evaluate six computational sequence-design landscapes (Table B.1). All objectives are maximized on their reported task scales. Each run starts from D_0 and receives $T = 10$ rounds of $B = 128$ distinct, previously unmeasured queries. The final metric is $J_{128} = V_{128}(S_T; f)$, computed from all online discoveries; initial designs are excluded. Curves report $V_{128}(S_t; f)$, whereas batch and pool diagnostics report $V_{128}(Q_t; f)$ and $V_{128}(C_t; f)$, respectively.

The entire batch and its source attribution are fixed before current outcomes are observed. Online oracle budgets are matched; proposal and surrogate computation follow each implementation.

Table B.1: Online benchmarks. Every run uses $T = 10$ and $B = K = 128$.

Task	L	$\lvert\Sigma\rvert$	$\lvert D_0\rvert$	Oracle
RNA1	14	4	5,000	FLEXS RNA1
RNA2	14	4	5,000	FLEXS RNA2
RNA3	14	4	5,000	FLEXS RNA3
AAV	90	20	15,307	FLEXS AAV
GFP	237	20	10,200	GFP-Transformer
TFBind	8	4	32,898	Exact lookup

Candidate-oracle evaluations are separate post-hoc diagnostics and never update search histories, scores, or allocation states.

B.2 Experimental Cohorts and Statistics

Main results use seeds 112–119. RNA ridge selection uses 168–175; one-factor sweeps and streaming configuration selection use 168–171. The previously inspected 208–215 cohort is a descriptive component audit; independent stage comparisons use 216–223. Feedback studies use eight offline subsets with two repeats each, and same-pool replay uses rounds 1, 5, and 10 from those runs. These cohorts are not pooled or substituted for main-comparison results.

Unless stated otherwise, results are means with $\mathrm{SEM}(z) = \sqrt{\sum_{i=1}^{n}(z_i - \bar{z})^2/[n(n-1)]}$. Feedback and replay first average repeats within each archive and then use eight archive-level units; Main-seed variability is conditional on a fixed archive, and incomplete baseline initialization-pairing records preclude treating the entire main table as a paired experiment.

B.3 Implementation and Baseline Settings

Reach. RNA uses 1,512 single-site and pairwise indicator features, with ridge penalty $\lambda = 10$ and rank-weight parameter 0.01. It scores 512 measured elites, 1,024 empirical-frequency starts, and 1,024 uniform starts; retains 512 expansion anchors; and proposes at most 16,384 variants within mutation radius three. The prior retains 128 candidates, while a separate elite-local source supplies radius-one/two proposals. The selected configuration uses zero coordinate sweeps. For AAV/GFP/TFBind, three elite-centered sources use single substitutions without the explicit pairwise prior. Two differ in whether deduplication occurs during prefix filling or after bounded enumeration; the third uses a broader offline/online elite mixture.

Table B.2: CRAFT settings. Raw caps limit per-round proposal computation, not the deduplicated union. All tasks use reference capacity 12,000, $\rho = 0.75$, and minimum source quota one.

Task	Sources	Raw cap	k	τ	η	Return
RNA1–3	2	24,576	16	0.95	0.08	Batch-standardized
AAV	3	4,096	16	0.85	0.02	Raw scores
GFP	3	8,192	16	0.95	0.08	Raw scores
TFBind	3	8,192	16	0.95	0.08	Raw scores

Selection and adaptation. Neighbors use normalized Hamming distance. The reference contains the full measured history up to 12,000 records and its highest-scoring prefix thereafter. Quantiles use linear interpolation (Hyndman & Fan, 1996). RNA retains cutoff ties before diversity; AAV/GFP/TFBind use farthest-first selection within score prefixes, with shortlist multipliers four/two/two. Source-priority merging removes measured and duplicate designs before bounded completion. Source returns are computed after the entire batch returns; all sources have identical initial states and synchronous discounted-count updates. Neighbor, tie, rounding, and completion rules remain fixed within matched comparisons.

Baselines. BO, CbAS/DbAS, and AdaLead use FLEXS implementations with the CNN-UCB proxy; BO proposes 2,000 model-ranked candidates per round, and CbAS/DbAS retain released VAE defaults (Brookes & Listgarten, 2018; Brookes et al., 2019; Sinai et al., 2020). PEX uses MuFacNet, ensemble-mean prediction, two random mutations, and frontier-neighbor size five (Ren et al., 2022). Required starting sequences are the highest-scoring measured designs.

GFN-AL and GFN-AL+δ-CS use the recorded generator/proxy family (Jain et al., 2022; Kim et al., 2024), with $(r_{\max}, c_\sigma)$ set to $(0.5, 5)$ on RNA/TFBind, $(0.05, 1)$ on GFP, and $(0.05, 0.1)$ on AAV. GFN-AL on RNA/AAV is a repository extension; its zero AAV result is retained as an adaptation outcome. ProSpero uses its released AAV runner and task-interface adaptations elsewhere (Kmicikiewicz et al., 2026).

MLDE-Population is an online adaptation (Wu et al., 2019; Wittmann et al., 2021): each round constructs up to 16,384 candidates from measured elites and token-frequency sampling, then queries 128 designs. Three-fold cross-validation selects linear regression, Bayesian ridge, or random forest on full-sequence one-hot features. Its separate TFBind extension is excluded from main-table rank markings.P3BO is discussed in Appendix A as an optimizer-portfolio reference and is not included in the empirical comparisons. DyNA-PPO is excluded because its adapter accesses the oracle during proposal construction.

B.4 Supplementary Online Results

The complete method comparison is reported in the main text. CRAFT has the highest mean on five tasks and the second-highest on GFP; rank markings do not denote statistical significance. Table B.3 supplies its seed-level endpoints. Trajectories are available for CRAFT, GFN-AL+δ-CS, ProSpero, AdaLead, and PEX; endpoint-only methods are not interpolated.

Table B.3: CRAFT seed-level J_{128}, retaining the supplied precision.

Seed	RNA1	RNA2	RNA3	AAV	GFP	TFBind
112	.9309	.9341	.9802	.6347	3.6122	.9729
113	.9390	.9378	.9999	.6706	3.6128	.9706
114	.9370	.9284	.9664	.6826	3.6146	.9691
115	.9324	.9477	.9810	.6749	3.6139	.9680
116	.9522	.9367	.9904	.6877	3.6103	.9738
117	.9482	.9410	.9956	.6712	3.6128	.9718
118	.9283	.9376	.9956	.6953	3.6140	.9711
119	.9496	.9449	.9936	.6977	3.6108	.9751

C Ablation Studies and Mechanistic Analyses

C.1 Candidate Quality and Controlled Stage Comparisons

Using seeds 216–223, Table C.1 compares the best eligible source, the deduplicated union, and the executed first batch under the same initial history. The gaps $\Gamma_{128,1}$ and $L_{128,1}$ follow Section E.1; all candidate labels are obtained after decisions. On every task, pool-to-query loss exceeds union gain. This loss includes batch constraints and execution, not scoring error alone; the larger union does not establish source complementarity at matched capacity.

The five stage contrasts in Table C.2 separate current-batch effects from trajectory effects. RNA3 removes the measurement-informed source and reruns the loop. AAV compares ordering on identical first-round candidates and quotas, and allocation after an identical first batch. GFP ordering is evaluated end to end. Intervals use 10,000 paired-bootstrap resamples; exact two-sided sign-flip tests are Holm-adjusted across these five contrasts (Holm, 1979), separately from the feedback family.

The RNA3 prior supplies fewer but higher-quality initial candidates; removing it also changes later measurements and computation. GFP's negative ordering effect and near-zero allocation effect preclude universal component gains. Earlier 208–215 audits included RNA variants matching

Table C.1: First-round candidate quality. The first three numeric columns are $V_{128}(\cdot; f)$. Means and gaps are rounded separately; GFP's union gain rounds to zero.

Task	Best source	Union	Queried batch	$\Gamma_{128,1}$	$L_{128,1}$
RNA1	.476	.483	.337	+.008	.146
RNA2	.424	.434	.292	+.010	.142
RNA3	.441	.474	.343	+.034	.132
AAV	.505	.512	.391	+.007	.121
GFP	3.580	3.580	3.194	+.000	.386
TFBind	.820	.854	.537	+.034	.317

Table C.2: Five paired stage contrasts. Positive effects favor the full, adaptive, or shared variant. First batch: $V_{128}(Q_1; f)$; final: J_{128}.

Task	Contrast	Outcome	Δ	95% CI	Wins	p_{Holm}
RNA3	Full − no primary	Final	+.376	$[+.353, +.398]$	8/8	.039
AAV	Adaptive − uniform	Final	+.029	$[+.017, +.043]$	8/8	.039
AAV	Shared − native	First batch	+.052	$[+.043, +.063]$	8/8	.039
GFP	Shared − native	Final	−.002	$[-.003, -.001]$	2/8	.109
GFP	Adaptive − uniform	Final	+.000	$[-.002, +.002]$	4/8	.875

or exceeding the full method. Frozen-pool transfer probes on GFN-AL+δ-CS/RNA2 and ProSpero/AAV/GFP also reduced selected quality when common scoring replaced native ordering; these are not end-to-end generator replacements.

C.2 Feedback Interventions and Candidate Opportunities

Protocol. RNA3, AAV, and GFP use eight label-blind 50% subsets of a fixed offline reservoir, containing 2,500, 7,653, and 5,100 labels, respectively, with two repeats per subset. The four history/allocation branches share their initial pools, first batch, and source assignments, then diverge from round two. Frozen history retains initial labels for generation and scoring but continues excluding queried designs; the entire reservoir, including unexposed designs, is ineligible for new queries. Fixed allocation preserves initial symmetric proportions and the integer-quota rule.

The endpoint results are in Table 2. Direct querying keeps history and allocation updates but bypasses common selection and diversity. The single source is chosen by development-stage first-pool quality and receives matched offered-candidate capacity and query budget, not necessarily matched raw computation. The $3 \times 8 \times 2 \times 6 = 288$ runs are separate from the main study.

Paired effects. For each archive, let $\bar{J}^{uv}$ average the two repeats, with u indicating history updates and v adaptive allocation. We use

$$\Delta_{\text{full}} = \bar{J}^{11} - \bar{J}^{00}, \qquad \Delta_{\text{history}} = \bar{J}^{10} - \bar{J}^{00}, \tag{C.1}$$

$$\Delta_{\text{allocation|updated}} = \bar{J}^{11} - \bar{J}^{10},$$

$$\Delta_{\text{interaction}} = (\bar{J}^{11} - \bar{J}^{10}) - (\bar{J}^{01} - \bar{J}^{00}). \tag{C.2}$$

Eight archive-level differences are analyzed with 10,000 paired-bootstrap resamples and exact two-sided sign-flip tests, assuming sign exchangeability under the paired null. Holm correction covers the nine primary contrasts jointly. All adjusted p-values are 0.0703; pointwise intervals therefore do not establish family-wise significance at 0.05. Secondary intervals are descriptive. Subsets may overlap, and uncertainty is conditional on the fixed reservoir.

History updates yield the larger mean gains and jointly affect generation and scoring. AAV's positive interaction is consistent with allocation becoming more useful after history updates; RNA3 and GFP interaction intervals span zero. The full-minus-direct intervals span zero on all three tasks; they establish neither selection superiority nor equivalence.

Table C.3: Archive-paired ΔJ_{128} and pointwise 95% bootstrap intervals. Effects retain source precision rather than being recomputed from rounded endpoints.

Contrast	RNA3	AAV	GFP
Primary contrasts: nine tests			
Full feedback	$+.4004$ $[+.3848, +.4151]$	$+.1595$ $[+.1506, +.1698]$	$+.0406$ $[+.0381, +.0441]$
History update	$+.3551$ $[+.3246, +.3892]$	$+.1239$ $[+.1161, +.1319]$	$+.0383$ $[+.0360, +.0411]$
Allocation: updated history	$+.0452$ $[+.0207, +.0721]$	$+.0356$ $[+.0217, +.0489]$	$+.0023$ $[+.0006, +.0037]$
Secondary controls and interaction			
Allocation: frozen history	$+.0403$ $[+.0362, +.0445]$	$+.0035$ $[+.0027, +.0042]$	$+.0011$ $[+.0008, +.0014]$
Full − direct query	$-.0018$ $[-.0116, +.0060]$	$+.0049$ $[-.0037, +.0152]$	$+.0008$ $[-.0009, +.0022]$
Full − best source	$+.0966$ $[+.0636, +.1338]$	$+.0050$ $[-.0052, +.0159]$	$+.0014$ $[+.0008, +.0020]$
Interaction	$+.0050$ $[-.0214, +.0337]$	$+.0321$ $[+.0179, +.0459]$	$+.0012$ $[-.0006, +.0026]$

Candidate opportunities. Figure C.1 tracks pool quality and the opportunity decomposition in Eq. equation E.9. Updated histories produce stronger later pools; frozen histories remain near their initial level. A small retention gap can merely reflect weak available opportunities, so available and realized gains must be read jointly. Candidate labels are post-hoc and outside the online budget; source-tagged entries can repeat designs and are not exact diagnostic oracle-call counts.

C.3 Same-Pool Selection and Computational Cost

Six scoring rules are replayed on full-feedback pools from rounds 1, 5, and 10, fixing history, candidates, quotas, source priority, deduplication, and tie/diversity procedures. Neighbor statistics use identical neighbors. The shared linear model fits one-hot ridge regression to available labels with penalty 10 and LSQR tolerance 10^{-6}. Original-quantile replay reproduces the executed batch. Table C.4 averages rounds and repeats within each archive before computing mean and SEM across eight archives.

Table C.4: Same-pool current-batch quality $V_{128}(Q_t; f)$ and end-to-end optimization time. Values are mean ± SEM across eight archive means; timing is measured separately for full and direct-query policies, not for every replayed selector.

Rule / policy	RNA3	AAV	GFP
Same-pool batch quality			
Upper quantile	$.6734 \pm .0056$	$.4497 \pm .0048$	$3.4129 \pm .0074$
Neighbor mean	$.6802 \pm .0058$	$.4688 \pm .0037$	$3.4436 \pm .0059$
Neighbor median	$.6767 \pm .0056$	$.4629 \pm .0036$	$3.4347 \pm .0062$
Neighbor maximum	$.6725 \pm .0054$	$.4397 \pm .0035$	$3.4070 \pm .0112$
Native order	$.6732 \pm .0057$	$.4328 \pm .0023$	$3.2355 \pm .0146$
Shared linear model	$.6968 \pm .0056$	$.5845 \pm .0025$	$3.5651 \pm .0015$
Ten-round optimization time (seconds)			
Full CRAFT	$11.7 \pm .2$	$11.8 \pm .2$	$24.1 \pm .2$
Direct query	$11.9 \pm .1$	$2.9 \pm .0$	$15.0 \pm .3$

Neighbor means, medians, and the shared linear model yield higher mean batch quality than the deployed quantile. These fixed-history results do not establish final-discovery improvements: replacement selectors would generate different histories. Native-order replay retains diversity, unlike direct querying. Timing uses two RTX 4080 SUPER GPUs and four concurrent workers, including generation, selection, and online evaluation but excluding loading and post-hoc pool evaluation. Di-

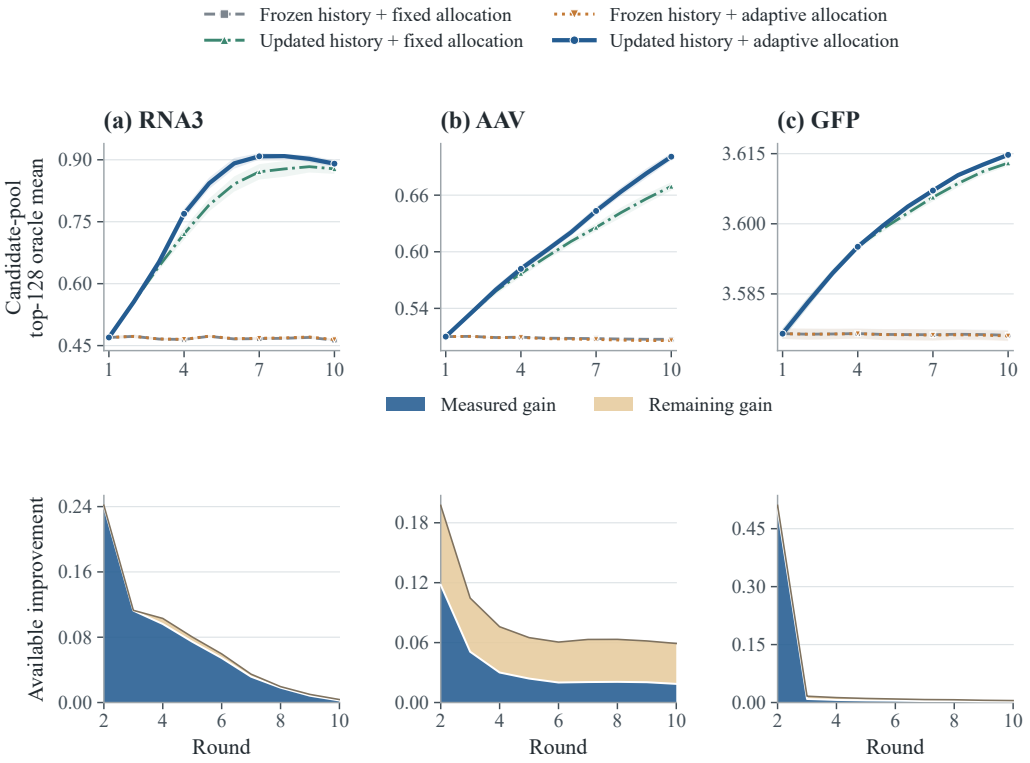

Figure C.1: Feedback and candidate opportunities. Top: $V_{128}(C_t; f)$ under four feedback conditions. Bottom: full-feedback $O_{128,t}$ split into measured improvement and unretained opportunity $U_{128,t}$ from round two. Candidate labels never enter the policy.

rect querying is faster on AAV/GFP. Under the same reduced-support feedback protocol, the paired endpoint intervals for full CRAFT minus Direct Query include zero (Table C.3); no paired timing intervals are inferred from marginal SEMs.

C.4 Full-Trajectory Selector Comparisons

Protocol and selection policies. We compare Quantile, KNN Mean, Shared Ridge, Native + Diversity, and Direct Query on RNA3, AAV, and GFP. Each run uses the complete offline dataset and ten rounds of 128 new queries. Policies with the same task and algorithm seed share their initial measurements and first-round candidate membership, but form their own query batches and subsequently update generation, scoring, and allocation from their own feedback. Thus, later candidate pools may differ across policies. All queries are distinct within a trajectory and exclude the initial offline dataset, as checked by the archived run audits.

KNN Mean replaces quantile aggregation with the mean of the same neighbors. Both use $k = 16$ and at most 12,000 reference sequences; Quantile uses $\tau = 0.95$, 0.85, and 0.95 for RNA3, AAV, and GFP, respectively. Shared Ridge fits the accumulated measurements using one-hot features, an intercept, ridge penalty 10, and LSQR tolerance 10^{-6}. Native + Diversity retains source-native ordering and the deployed task-specific diversity procedure; Direct Query bypasses shared scoring and selection-stage diversity while retaining generation, quota merging, deduplication, and feedback updates. The evaluated RNA3 implementation retains its tie-closed candidate ordering, while AAV/GFP use their deployed farthest-first procedure. Ridge uses all accumulated measurements, whereas KNN uses its reference cap; training-set sizes and computational costs are therefore not matched across scorers.

Cohorts, metrics, and implementation version. The exploration cohort uses seeds 1000–1004: 75 trajectories and 96,000 online oracle calls. Separate holdout protocols use seeds 1005–1009 for AAV (25 trajectories; 32,000 calls) and RNA3/GFP (50 trajectories; 64,000 calls). All 150 trajectories completed, with zero diagnostic oracle calls. Each task retains a fixed offline dataset; seeds vary algorithmic randomness rather than resampling that dataset. Means and SEMs are computed across seeds, and contrasts pair policies by task and seed. Cohorts are reported separately.

The primary endpoint is query-only final discovery, $J_{128} = V_{128}(S_{10}; f)$, excluding the offline observations. We additionally report first-batch quality $V_{128}(Q_1; f)$ and the discrete trajectory mean

$$\mathrm{AUC}_{\mathrm{disc}} = \frac{1}{10} \sum_{t=1}^{10} V_{128}(S_t; f). \tag{C.3}$$

This is an equally weighted mean at budgets 128, 256, …, 1280, not a trapezoidal integral, and includes no zero-budget observation.

All results in this subsection and Appendix C.5 were recomputed using the implementation with fixed-reference-order neighbor tie handling specified in Section 3.2.

Results and inferential scope. Table C.5 reports all five policies; Figure C.2 shows their complete trajectories. On AAV, Quantile exceeds Native + Diversity at the endpoint in both cohorts, with paired mean differences of 0.043571 and 0.058230, respectively (Table C.6). AAV was selected after inspecting the exploration cohort. Before running its holdout, the positive endpoint contrast Quantile minus Native + Diversity was specified as the primary comparison: all five differences are positive, giving a one-sided exact paired sign-test $p = 0.03125$. Shared Ridge has the highest AAV mean in both cohorts, so this result supports a benefit over native ordering rather than optimality of quantile scoring.

On GFP, Quantile improves first-batch quality over Native + Diversity by 0.207157 and 0.182137 in the exploration and holdout cohorts, respectively, with five positive differences in each cohort. The corresponding endpoint differences are -0.002064 and -0.000225: the early advantage does not persist to the endpoint. RNA3 likewise shows no consistent endpoint advantage of Quantile over Native + Diversity. The separately specified primary comparison for the RNA3/GFP holdout was Shared Ridge minus Quantile on the RNA3 endpoint. Its mean difference is -0.011119, with two positive, one tied, and two negative differences; the one-sided exact sign-test $p = 0.6875$ does not support the anticipated positive direction.

Table C.5: Full-trajectory final J_{128} after 1,280 new queries (mean $\pm$ SEM over five seeds). Both cohorts use complete offline support. AAV and RNA3/GFP holdouts followed separate protocols.

Policy	RNA3	AAV	GFP
Exploration: seeds 1000–1004, $n = 5$			
Quantile	0.9545 ± 0.0243	0.6707 ± 0.0070	3.6110 ± 0.0012
KNN Mean	0.9547 ± 0.0250	0.6808 ± 0.0096	3.6110 ± 0.0014
Shared Ridge	0.9689 ± 0.0283	0.7614 ± 0.0009	3.6099 ± 0.0006
Native + Diversity	0.9569 ± 0.0211	0.6271 ± 0.0073	3.6131 ± 0.0010
Direct Query	0.9677 ± 0.0111	0.6278 ± 0.0100	3.6126 ± 0.0008
Holdout: seeds 1005–1009, $n = 5$			
Quantile	0.9715 ± 0.0125	0.7005 ± 0.0063	3.6108 ± 0.0015
KNN Mean	0.9872 ± 0.0029	0.6570 ± 0.0087	3.6116 ± 0.0005
Shared Ridge	0.9604 ± 0.0231	0.7621 ± 0.0006	3.6101 ± 0.0002
Native + Diversity	0.9754 ± 0.0042	0.6423 ± 0.0085	3.6110 ± 0.0011
Direct Query	0.9658 ± 0.0132	0.6470 ± 0.0109	3.6095 ± 0.0012

The original exploration analysis treats its 12 endpoint contrasts as exploratory and uses two-sided exact sign-flip tests with Holm correction; none reaches the 0.05 threshold. The directional holdout tests above were specified separately within their respective protocols. Other effects reported here, including the GFP first-batch contrasts, are descriptive; they are not additional confirmatory tests.

Table C.6: Paired mean differences, Quantile minus the indicated control. Δ_1, Δ_T, and Δ_A denote first-batch, final-discovery, and $\mathrm{AUC}_{\mathrm{disc}}$ differences. The $+/0/-$ column counts positive, tied, and negative endpoint differences across five seeds. These are descriptive effect summaries.

Task	Control	Δ_1	Δ_T	$+/0/-$	Δ_A
Exploration: seeds 1000–1004					
RNA3	Native + Div.	-0.000492	-0.002417	1/1/3	-0.007632
RNA3	Direct Query	-0.000492	-0.013262	2/0/3	-0.007430
AAV	Native + Div.	$+0.057929$	$+0.043571$	5/0/0	$+0.047725$
AAV	Direct Query	$+0.052129$	$+0.042907$	4/0/1	$+0.041795$
GFP	Native + Div.	$+0.207157$	-0.002064	1/0/4	$+0.019385$
GFP	Direct Query	$+0.122991$	-0.001547	1/0/4	$+0.011382$
Holdout: seeds 1005–1009					
RNA3	Native + Div.	$+0.014277$	-0.003872	2/1/2	$+0.009725$
RNA3	Direct Query	$+0.014277$	$+0.005736$	2/1/2	$+0.012420$
AAV	Native + Div.	$+0.060821$	$+0.058230$	5/0/0	$+0.051549$
AAV	Direct Query	$+0.055085$	$+0.053541$	5/0/0	$+0.054534$
GFP	Native + Div.	$+0.182137$	-0.000225	2/0/3	$+0.018030$
GFP	Direct Query	$+0.164180$	$+0.001308$	4/0/1	$+0.016863$

C.5 Independent AAV Replication

Protocol. We further evaluate Quantile, Native + Diversity, and Direct Query on ten new seeds, 1010–1019, using the same frozen implementation, complete offline support, and ten rounds of 128 queries. The Quantile configuration remains $k = 16$ and $\tau = 0.85$. All 30 trajectories completed, totaling 38,400 online oracle calls and zero diagnostic oracle calls. Query uniqueness is enforced within each trajectory; sequences may recur across separate runs. All seeds are retained.

AAV was chosen based on the earlier exploration and five-seed holdout results. Before running the new seeds, the positive Quantile minus Native + Diversity contrast in final J_{128} was specified as the primary comparison, with a one-sided exact paired sign test at $\alpha = 0.05$. Quantile minus Direct Query, first-batch quality, and $\mathrm{AUC}_{\mathrm{disc}}$ were listed as secondary outcomes. This cohort is analyzed separately from the earlier cohorts.

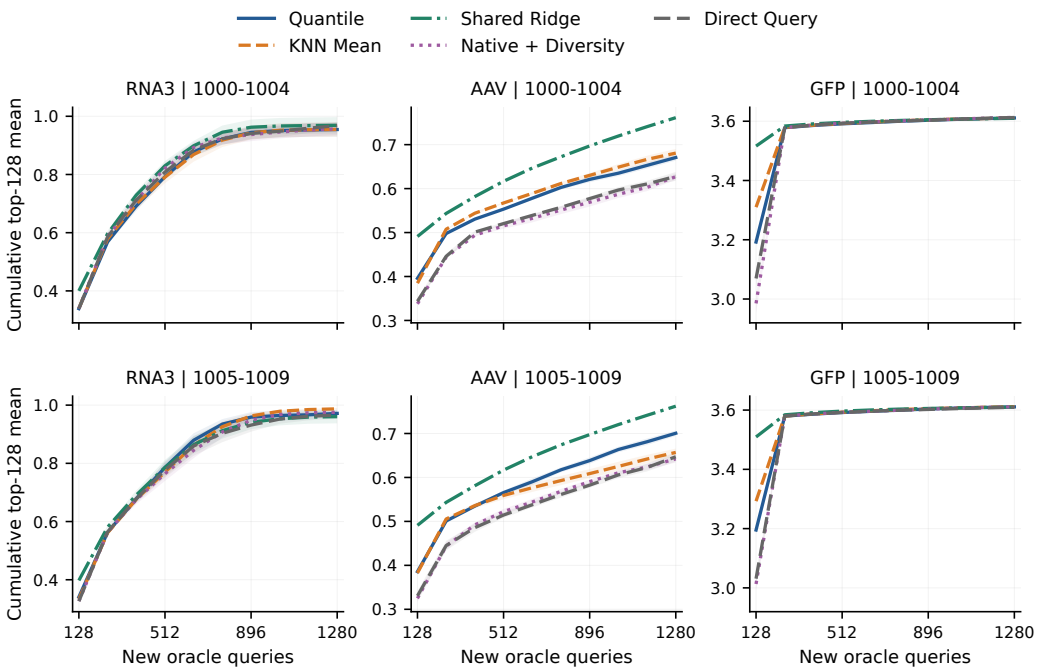

Figure C.2: Complete selector trajectories under full offline support. Top: exploration seeds 1000–1004. Bottom: holdout seeds 1005–1009. Lines and bands show mean cumulative top-128 quality and one SEM over five seeds. Every policy follows its own feedback history; the curves are not fixed-pool replays.

Table C.7: Independent AAV replication: mean ± SEM across ten new algorithm seeds. All metrics exclude the initial offline data.

Policy	First batch	Final J_{128}	$\mathrm{AUC}_{\mathrm{disc}}$
Quantile	0.3900 ± 0.0029	0.6883 ± 0.0060	0.5798 ± 0.0041
Native + Diversity	0.3381 ± 0.0069	0.6429 ± 0.0075	0.5417 ± 0.0059
Direct Query	0.3408 ± 0.0068	0.6457 ± 0.0101	0.5383 ± 0.0046

Primary and secondary results. Quantile exceeds Native + Diversity by a paired mean of 0.045461 in final discovery, with nine positive differences out of ten (Tables C.7 and C.8). The prespecified one-sided sign-test $p = 0.010742$; a two-sided sensitivity analysis gives $p = 0.021484$. The sign test assesses the direction of paired differences, not a null hypothesis about their mean. A descriptive 95% paired percentile-bootstrap interval for the mean difference is $[0.030786, 0.059746]$, obtained by resampling the ten seed-level paired differences 20,000 times.

The mean first-batch improvement over Native + Diversity is 0.051851, and the $\mathrm{AUC}_{\mathrm{disc}}$ improvement is 0.038076; both are positive in all ten seeds. Against Direct Query, the mean endpoint improvement is 0.042599 with nine positive differences and a descriptive bootstrap interval of $[0.021544, 0.062109]$. The corresponding first-batch and trajectory-mean improvements are 0.049216 and 0.041507, both positive in all ten seeds. These secondary outcomes are reported descriptively, without additional confirmatory significance claims.

Table C.8: All seed-level final J_{128} values in the independent AAV replication. Δ_T is Quantile minus Native + Diversity; the negative difference at seed 1012 is retained.

Seed	Quantile	Native + Diversity	Direct Query	Δ_T
1010	0.709184	0.645644	0.634274	+0.063540
1011	0.689611	0.646031	0.665446	+0.043579
1012	0.685286	0.685969	0.669638	−0.000683
1013	0.644908	0.620192	0.621228	+0.024716
1014	0.696854	0.615017	0.622369	+0.081837
1015	0.695909	0.618321	0.651405	+0.077588
1016	0.685908	0.648761	0.713851	+0.037147
1017	0.714117	0.674391	0.643863	+0.039726
1018	0.684709	0.648665	0.599390	+0.036044
1019	0.676754	0.625639	0.635783	+0.051116

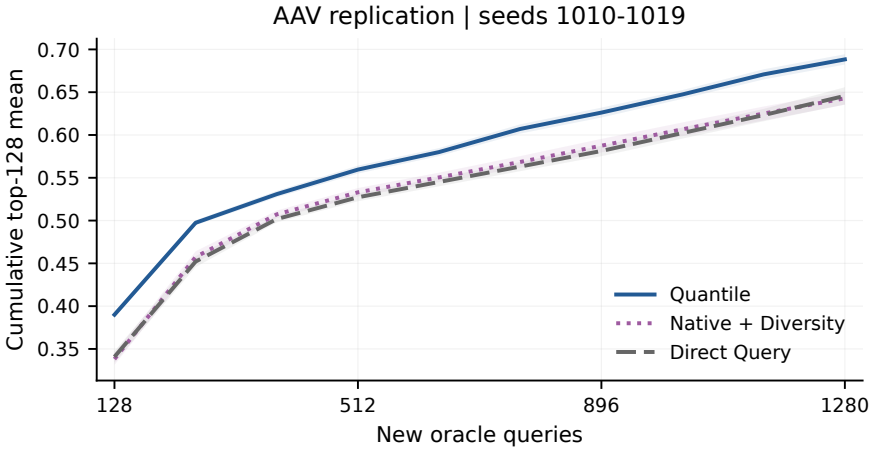

Figure C.3: Independent AAV replication over the full online budget. Lines and bands show mean cumulative top-128 quality and one SEM across seeds 1010–1019.

The replication supports a contribution of the archived shared selection procedure on AAV under this protocol. Native + Diversity retains the diversity procedure and therefore provides the more targeted comparison of score-based prioritization; Direct Query removes both shared scoring and selection-stage diversity. The stronger Shared Ridge results in the five-policy study and the cross-task results delimit the conclusion: this experiment does not establish that quantile scoring is the best selector or that its endpoint benefit generalizes across tasks.

D DIVERSITY, NOVELTY, AND QUALITY OF ONLINE DISCOVERIES

Evaluation objects. For each run, let $\widehat{S}_K = \mathrm{Top}_K(S_T; f)$ be the $K = 128$ highest-scoring distinct online discoveries, with fixed tie handling, and let $\mathcal{X}_0^{\mathrm{obs}} = \{x : (x, y) \in D_0\}$. The score is $J_K = K^{-1} \sum_{x \in \widehat{S}_K} f(x)$. Using normalized Hamming distance $d_H(x, z) = L^{-1} \sum_{j=1}^{L} \mathbf{1}\{x_j \neq z_j\}$, define

$$\begin{aligned} \mathrm{Diversity}_K &= \frac{1}{K(K-1)} \sum_{\substack{x,z \in \widehat{S}_K \\ x \neq z}} d_H(x, z), \qquad K \geq 2, \\ \mathrm{Novelty}_K &= \frac{1}{K} \sum_{x \in \widehat{S}_K} \min_{z \in \mathcal{X}_0^{\mathrm{obs}}} d_H(x, z). \end{aligned} \tag{D.1}$$

Diversity measures internal dispersion; novelty measures distance from the initial offline data, not the final history. Ordered and unordered pair averaging give identical diversity. These are sequence-level, within-task descriptions rather than measures of functional diversity. All views concern final top-128 sets, not individual-distance distributions or all-query novelty.

Score and distance alignment. MLDE score summaries agree with Table 1. For all six tasks, including AAV, scores, novelty, and diversity are computed from the same eight runs with seeds 112–119. Joint summaries use scores and distance metrics from matched runs.

Records and visualization. The available distance records contain eight runs (seeds 112–119) per method and task, under the recorded ten-round, 128-query protocol. Boxes show run-level quartiles, medians, and minimum/maximum whiskers, with all eight seed markers. Offsets on the categorical method axis only improve visibility; metric values are unchanged. Joint panels pair scores and distances from the same run and contain 56 points, except AAV, which contains 48 points plus the MLDE score-only reference. Neither numerical coordinate is jittered. Filled markers identify current CRAFT/MLDE records; open markers identify archived baselines and the MLDE AAV distance-only records. Other methods retain their saved run-level scores, rather than replacing them with main-table means. Archived reference/metric definitions and the mapping of Delta-CS to the current GFN-AL+δ-CS are not fully verified (the latter is marked *); these comparisons remain descriptive, not controlled rankings or significance claims.

D.1 RNA1–RNA3

Diversity and novelty. Table D.1 and Figure D.1 show higher mean diversity for CRAFT than MLDE, AdaLead, PEX, and ProSpero on all three RNA tasks. Relative to MLDE, mean diversity is higher by 53.5%, 53.6%, and 45.2% on RNA1–RNA3, respectively.

GFN-AL and archived Delta-CS remain more diverse; CRAFT's RNA3 diversity ranges from 0.2132 to 0.4709 across runs. CRAFT's novelty exceeds MLDE by 7.6%, 3.8%, and 5.2%, and is highest among the displayed means on RNA1/RNA2 (Figure D.1). On RNA3 it is nearly identical to AdaLead and below ProSpero. GFN-AL combines higher diversity with lower novelty than CRAFT, illustrating the different properties measured by the two distances.

Joint interpretation. Using the main-table MLDE scores, CRAFT's mean scores of 0.9397, 0.9385, and 0.9878 exceed MLDE's 0.8518, 0.8567, and 0.8345. The corresponding differences are 0.0879, 0.0818, and 0.1533 at the displayed precision. CRAFT has the highest displayed RNA mean scores, whereas GFN-AL's greater diversity accompanies substantially lower scores (Figure D.2). On RNA3, CRAFT and AdaLead have nearly identical novelty (0.4581 and 0.4584) but different scores (0.9878 and 0.9206). High-quality discovery can therefore coexist with a dispersed retained set without maximizing either distance measure.

D.2 AAV, GFP, AND TFBIND

We jointly examine the score, novelty, and diversity of the final retained top-128 on AAV, GFP, and TFBind (Table D.2). All MLDE entries refer to **MLDE (adapted)**, including AAV. The three metrics are reported together, rather than treating the AAV distance records as a separate method.

Table D.1: RNA results: mean score / novelty / diversity of the final top-128 across eight runs. MLDE scores agree with the main-table version. Distances are normalized Hamming fractions; archived comparisons are descriptive.

Method	RNA1	RNA2	RNA3
CRAFT	.9397 / .5264 / .4132	.9385 / .5534 / .4132	.9878 / .4581 / .3746
MLDE (adapted)	.8518 / .4890 / .2691	.8567 / .5333 / .2690	.8345 / .4353 / .2579
AdaLead	.8963 / .5214 / .2641	.8921 / .5391 / .2689	.9206 / .4584 / .2686
PEX	.7720 / .4867 / .3514	.6375 / .3922 / .2310	.7387 / .4122 / .3053
GFN-AL	.5112 / .4081 / .7094	.4989 / .4120 / .7086	.5129 / .3797 / .7224
GFN-AL+δ-CS*	.9211 / .5144 / .4314	.9313 / .5472 / .4565	.9468 / .4547 / .3885
ProSpero	.8186 / .5073 / .2999	.8370 / .5455 / .2791	.8298 / .4655 / .3028

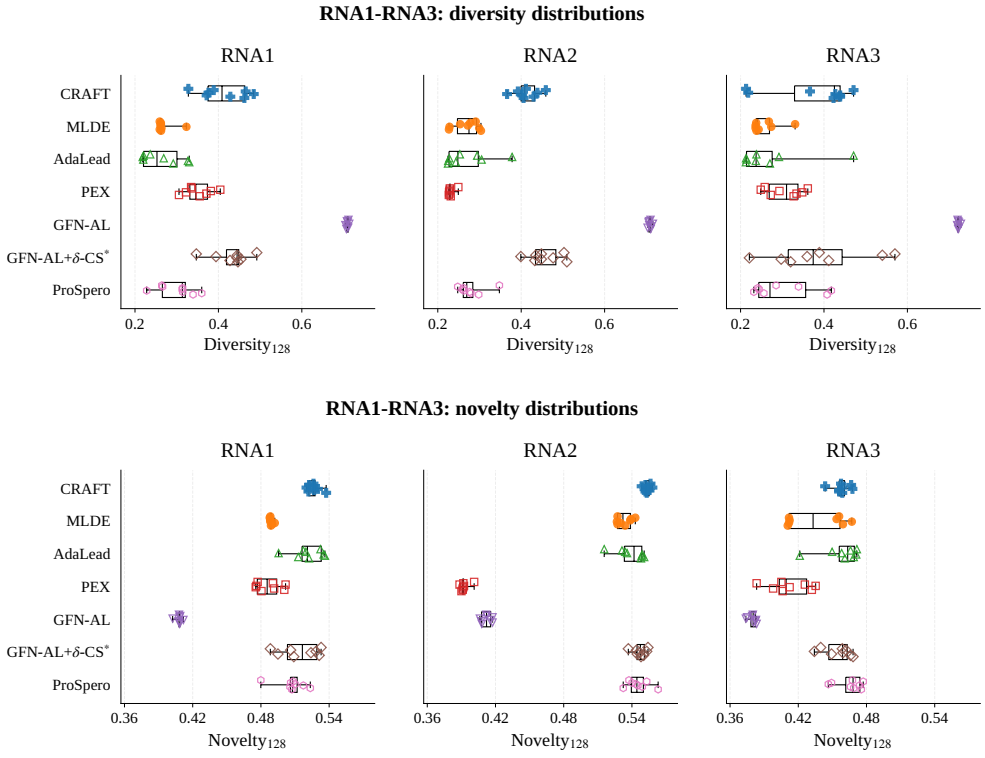

Figure D.1: RNA run-level distributions: diversity (top) and novelty (bottom). Each task retains eight observations per method. Boxes span the interquartile range, central lines are medians, and whiskers show the observed minimum and maximum. Filled/open markers distinguish current and archived records.

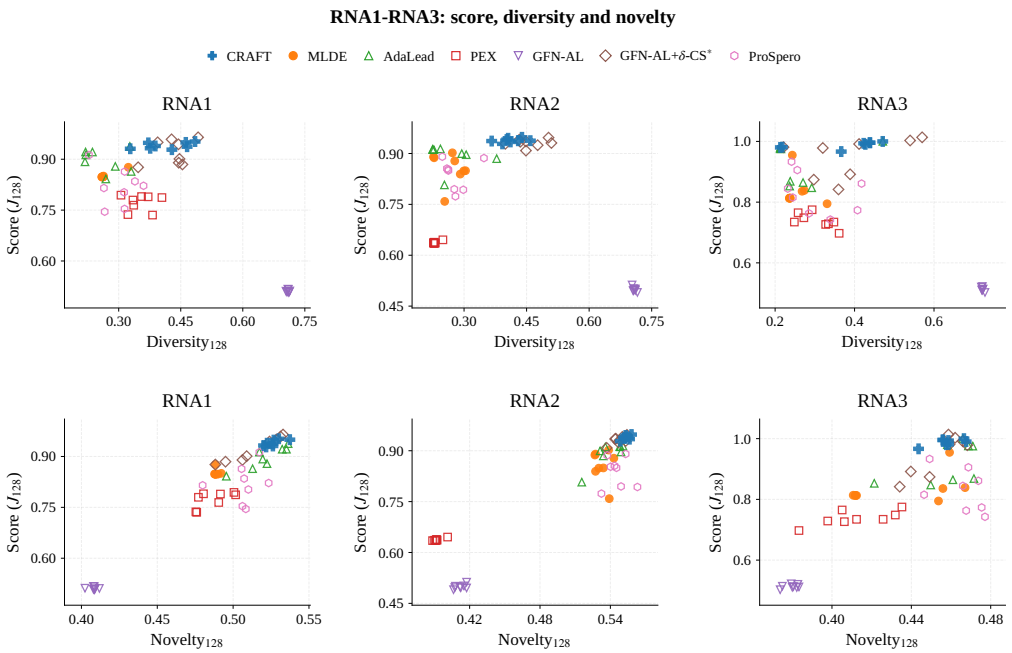

Figure D.2: RNA score–diversity (top) and score–novelty (bottom). Each task panel contains 56 same-run observations. MLDE scores and distance metrics are paired by run. Overlapping coordinates are not displaced; the two rows share a score scale within each task.

AAV. CRAFT attains a mean score of 0.6768, compared with 0.5892 for MLDE (adapted), while retaining lower novelty (0.0887 versus 0.1591) and diversity (0.0349 versus 0.0638). Its mean score is also close to ProSpero's 0.6731, despite lower novelty (0.0887 versus 0.1544) and diversity (0.0349 versus 0.0630). Relative to PEX, however, CRAFT has both a higher score (0.6768 versus 0.5835) and higher novelty (0.0887 versus 0.0635), while remaining less diverse. These comparisons show that high-quality AAV discoveries can form a concentrated set, without a consistent ordering of quality by either distance metric. The recorded GFN-AL result provides a contrasting case: its large novelty and diversity accompany a zero score.

GFP. CRAFT's mean score of 3.6127 exceeds MLDE's 3.5894 by 0.0233, but its novelty and diversity are lower (0.0405/0.0128 versus 0.0549/0.0407). Compared with AdaLead and PEX, CRAFT has higher scores and novelty, but lower diversity. Thus, distance from the initial data and dispersion within the retained set describe distinct properties: a set can be farther from its closest offline references without being more internally dispersed. ProSpero has the highest displayed mean score, 3.6549, so these records do not establish uniform quality superiority for CRAFT. Across its eight runs, CRAFT's diversity remains between 0.0113 and 0.0151, indicating consistently concentrated retained sets under these settings.

TFBind. CRAFT has higher mean score and diversity than MLDE (0.9716/0.6356 versus 0.9420/0.5696), despite lower novelty (0.1451 versus 0.1782). Its score is close to that of the recorded GFN-AL+δ-CS* result (0.9713), while its diversity is higher (0.6356 versus 0.4809) and its novelty is lower (0.1451 versus 0.8574). CRAFT also has higher score and diversity than PEX and ProSpero, whereas AdaLead and GFN-AL retain more diverse sets with lower mean scores. Consequently, a high-quality set can remain close to offline measurements while being internally dispersed; neither greater novelty nor greater diversity is by itself a sufficient indicator of discovery quality.

Visualization and interpretation. Figure D.3 separately shows the eight run-level diversity and novelty values per method; boxes denote quartiles, medians, and observed minimum/maximum whiskers. Figure D.4 pairs scores and distances

For AAV MLDE, one larger circular marker instead combines its corrected mean score with its mean distance; horizontal and vertical bars show the corresponding marginal SEMs.

Taken together, these results motivate evaluating discovery quality alongside both distance metrics, rather than maximizing either metric in isolation. Final-set mean pairwise diversity differs from the batch-level maximin criterion, and these endpoint records do not isolate the causal effects of diversity handling, shared scoring, or history updates.

Table D.2: AAV/GFP/TFBind results: mean reported score / novelty / diversity of the final top-128 across eight runs. MLDE denotes the online adaptation. Its score summaries agree with Table 1, and its score and distance metrics are computed from the same runs. Distances are normalized Hamming fractions; archived comparator results remain descriptive.

Method	AAV	GFP	TFBind
CRAFT	.6768 / .0887 / .0349	3.6127 / .0405 / .0128	.9716 / .1451 / .6356
MLDE (adapted)	.5892 / .1591 / .0638	3.5894 / .0549 / .0407	.9420 / .1782 / .5696
AdaLead	.6495 / .0957 / .0806	3.5706 / .0141 / .1236	.9665 / .1666 / .6474
PEX	.5835 / .0635 / .0915	3.5922 / .0280 / .0395	.9505 / .1528 / .5506
GFN-AL	.0000 / .9068 / .9500	1.6320 / .9159 / .9497	.9245 / .1650 / .6929
GFN-AL+δ-CS*	.6438 / .1041 / .1188	3.5770 / .0461 / .1262	.9713 / .2574 / .4809
ProSpero	.6731 / .1544 / .0630	3.6549 / .9165 / .0600	.9386 / 1.0000 / .5440

E Quality Retention and Online Discovery

This appendix proves the finite-pool and cumulative-discovery results in the main text. The objective $f : \mathcal{X} \to \mathbb{R}$ is fixed, all design sets are finite and contain distinct sequences, and all scores are finite. Results hold for each realized history and internal algorithmic choice; candidate generation need not be independent. The method uses exact oracle labels. Bounded label errors below are conditional extensions, recovered by setting the error terms to zero.

E.1 Notation, Feasibility, and Candidate Opportunities

For $K \geq 1$, $|A| \geq K$, and a score h, write

$$V_K(A; h) = \max_{\substack{D \subseteq A \\ |D|=K}} \frac{1}{K} \sum_{x \in D} h(x). \tag{E.1}$$

This mean is monotone under set inclusion, including for signed scores. Let $[u]_+ = \max\{u, 0\}$ and $\|h\|_{\infty,A} = \max_{x \in A} |h(x)|$, with $\|h\|_{\infty,\varnothing} = 0$.

At round t, let $C_t = \bigcup_{a \in \mathcal{A}} P_{a,t}$ and $\mathcal{X}_{t-1}^{\text{obs}} = \{x : (x, y) \in H_{t-1}\}$. Define

$$\begin{aligned} C_t^{\text{new}} &= C_t \setminus \mathcal{X}_{t-1}^{\text{obs}}, \\ \mathcal{F}_t &= \{Q \subseteq C_t^{\text{new}} : |Q| \leq B\}, \quad \mathcal{F}_t^{(K)} = \{Q \in \mathcal{F}_t : |Q| \geq K\}. \end{aligned} \tag{E.2}$$

The executed batch satisfies $Q_t \in \mathcal{F}_t$, and each query is attributed once to a source containing it. Initial quotas are construction targets, not exact final counts; farthest-first ordering likewise imposes no prescribed minimum distance on the final batch. Quota execution, score screening, diversity ordering, and completion therefore belong to the selection procedure, rather than additional hard constraints in equation E.2. Declared eligibility restrictions or hard constraints may instead specify a subfamily containing Q_t; it must not be reduced retrospectively to the executed batch.

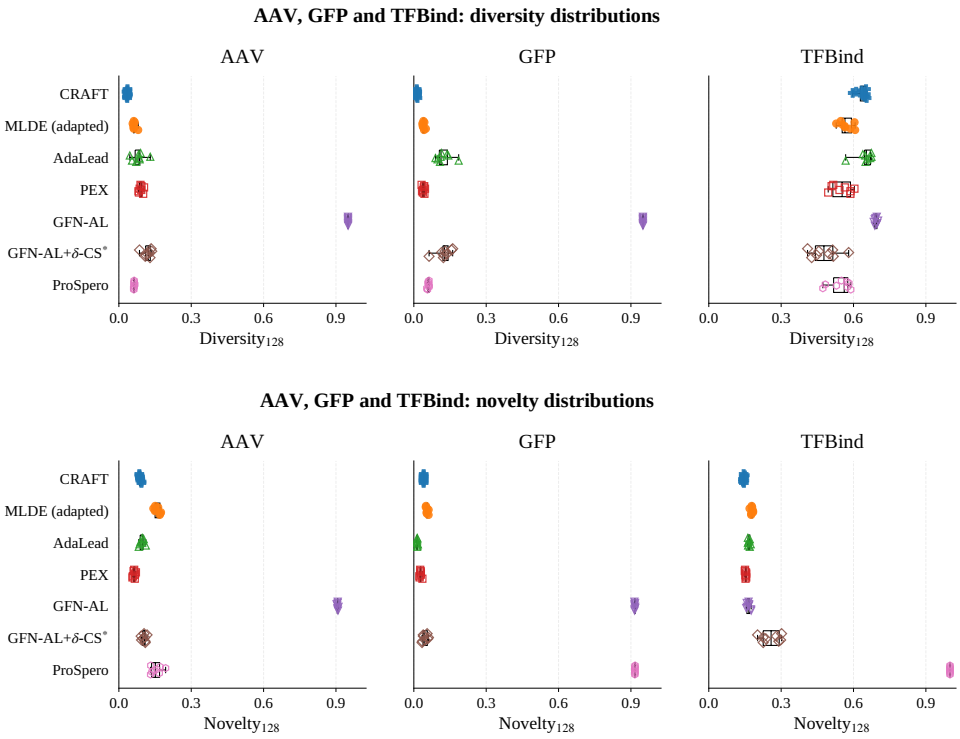

Figure D.3: AAV/GFP/TFBind run-level distributions: diversity (top) and novelty (bottom). All seven methods retain eight distance observations per task, including MLDE (adapted) on AAV. Boxes span the interquartile range, center lines show medians, and whiskers show the observed minimum and maximum. Categorical offsets separate seed markers without changing distance values. Filled markers denote CRAFT and MLDE; open markers denote the archived comparators. No distance value is clipped or changed.

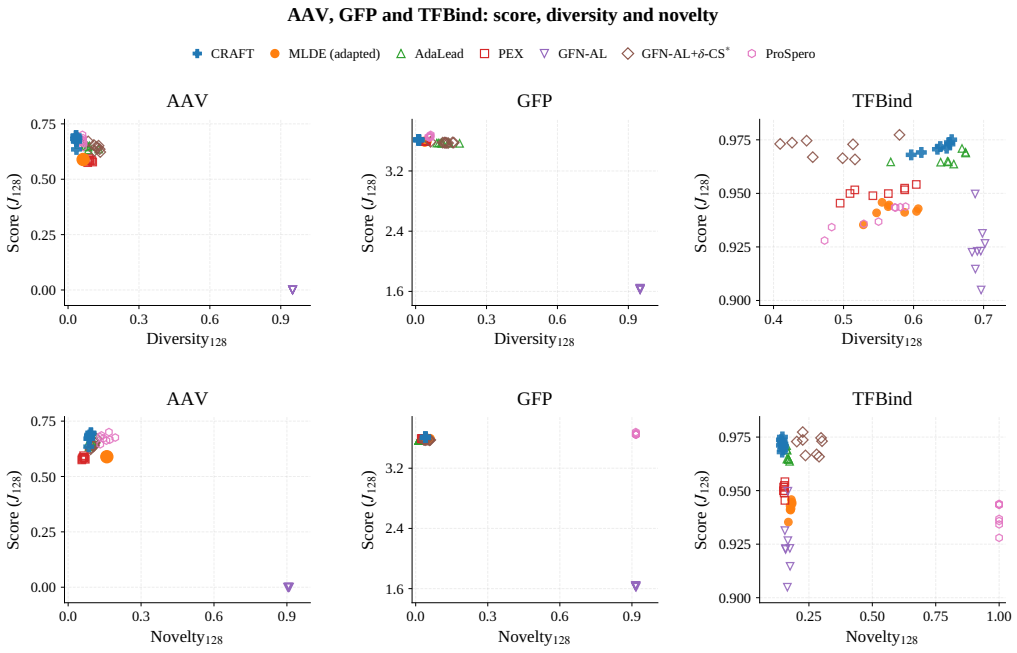

Figure D.4: AAV/GFP/TFBind score–diversity (top) and score–novelty (bottom). GFP and TFBind contain eight same-run points for each of seven methods. Each AAV panel contains 48 same-run points for the other six methods and one larger MLDE (adapted) aggregate marker. Its vertical coordinate is the corrected mean score 0.5892 with SEM 0.0124; its horizontal coordinate is mean diversity 0.0638 or mean novelty 0.1591, with the corresponding distance SEM. The two marginal error bars are not a joint confidence region. No corrected AAV seed scores are inferred, and no numerical coordinate jitter is applied.

When $|Q_t| \geq K$, define the realized, feasibility, and score-execution gaps by

$$\begin{aligned} L_{K,t} &= V_K(C_t; f) - V_K(Q_t; f), \\ \Delta_{\mathcal{F}_t} &= V_K(C_t; f) - \max_{Q \in \mathcal{F}_t^{(K)}} V_K(Q; f), \\ r_{q,t} &= \max_{Q \in \mathcal{F}_t^{(K)}} V_K(Q; q_{\tau,t}) - V_K(Q_t; q_{\tau,t}). \end{aligned} \tag{E.3}$$

All three are nonnegative. For the family in equation E.2, any finite score h satisfies

$$\max_{Q \in \mathcal{F}_t^{(K)}} V_K(Q; h) = V_K(C_t^{\text{new}}; h). \tag{E.4}$$

Indeed, inclusion gives the upper bound, and the top-K eligible designs attain it because $K \leq B$. Thus $\Delta_{\mathcal{F}_t}$ measures exclusion of measured candidates and vanishes when all offered candidates are new. The execution gap is computable from common scores without new oracle evaluations. Additional hard constraints preserve the gap definitions but need not preserve equation E.4.

If some source contains at least K designs, define

$$\begin{aligned} \Gamma_{K,t} &= V_K(C_t; f) - \max_{a:\, |P_{a,t}| \geq K} V_K(P_{a,t}; f), \\ V_K(Q_t; f) &= \max_{a:\, |P_{a,t}| \geq K} V_K(P_{a,t}; f) + \Gamma_{K,t} - L_{K,t}. \end{aligned} \tag{E.5}$$

The union gain is nonnegative, and $\Gamma_{1,t} = 0$. For $K > 1$, its benefit can arise from high-quality designs supplied by different sources. These are comparisons of realized pools under their actual budgets, not capacity-matched counterfactuals; a source with at least K designs need not itself be a feasible query batch.

E.2 Quality Retention by the Executed Batch

[Top-K perturbation] If $|A| \geq K$ and $\|h - \widetilde{h}\|_{\infty,A} \leq \epsilon$, then

$$|V_K(A; h) - V_K(A; \widetilde{h})| \leq \epsilon. \tag{E.6}$$

Proof. For every K-element subset, the two mean scores differ by at most ϵ. Taking maxima over the same finite family yields $V_K(A; h) \leq V_K(A; \widetilde{h}) + \epsilon$; exchanging the scores gives the reverse inequality. Ties and signed scores are allowed. □

[Retention by the executed batch] If $|Q_t| \geq K$ and $\|q_{\tau,t} - f\|_{\infty,C_t} \leq \epsilon_t$, then

$$0 \leq L_{K,t} \leq \Delta_{\mathcal{F}_t} + 2\epsilon_t + r_{q,t}. \tag{E.7}$$

Proof. Write $F_h^\star = \max_{Q \in \mathcal{F}_t^{(K)}} V_K(Q; h)$. Lemma E.2, applied to each feasible batch, gives $|F_{q_{\tau,t}}^\star - F_f^\star| \leq \epsilon_t$. Therefore

$$\begin{aligned} V_K(Q_t; f) &\geq V_K(Q_t; q_{\tau,t}) - \epsilon_t \\ &= F_{q_{\tau,t}}^\star - r_{q,t} - \epsilon_t \\ &\geq F_f^\star - r_{q,t} - 2\epsilon_t \\ &= V_K(C_t; f) - \Delta_{\mathcal{F}_t} - r_{q,t} - 2\epsilon_t. \end{aligned}$$

Rearrangement proves the upper bound; $Q_t \subseteq C_t$ gives nonnegativity. □

Combining equation E.5 and equation E.7 yields

$$V_K(Q_t; f) - \max_{a:\, |P_{a,t}| \geq K} V_K(P_{a,t}; f) \geq \Gamma_{K,t} - \Delta_{\mathcal{F}_t} - 2\epsilon_t - r_{q,t}. \tag{E.8}$$

Thus $\Gamma_{K,t} > \Delta_{\mathcal{F}_t} + 2\epsilon_t + r_{q,t}$ is sufficient to exceed every eligible source-pool quality. The result conditions on scoring accuracy and does not assume that the implemented selector has a small execution gap.

E.3 Cumulative Discovery and Batch-to-Cumulative Transfer

Let $S_0 = \varnothing$, $S_t = S_{t-1} \cup Q_t$, and suppose $|S_T| \geq K$. Define

$$t_0 = \min\{t \in \{1, \dots, T\} : |S_t| \geq K\}, \qquad J_K = V_K(S_T; f).$$

For $t > t_0$, define available and unretained opportunities by

$$\begin{aligned} O_{K,t} &= V_K(S_{t-1} \cup C_t; f) - V_K(S_{t-1}; f), \\ U_{K,t} &= V_K(S_{t-1} \cup C_t; f) - V_K(S_t; f). \end{aligned} \tag{E.9}$$

Inclusion gives $0 \leq U_{K,t} \leq O_{K,t}$. Subtraction and telescoping give the exact identities

$$\begin{aligned} V_K(S_t; f) - V_K(S_{t-1}; f) &= O_{K,t} - U_{K,t}, \\ J_K &= V_K(S_{t_0}; f) + \sum_{t=t_0+1}^{T} (O_{K,t} - U_{K,t}). \end{aligned} \tag{E.10}$$

Empty sums are zero. If $|Q_1| \geq K$, then $t_0 = 1$ and the initial term is the offline-informed first batch $V_K(Q_1; f)$, as in the main text. For $t_0 > 1$, initialization may already use online feedback. Offline designs are excluded from S_t; a metric that includes them instead initializes S_0 to those designs and permits $t_0 = 0$.

Cumulative retention with smaller batches. For $t > t_0$, let stored scores z_{t-1} satisfy $\|z_{t-1} - f\|_{\infty, S_{t-1}} \leq \nu_t$ and suppose $\|q_{\tau,t} - f\|_{\infty, C_t} \leq \epsilon_t$. Set

$$\widetilde{q}_t(x) = \begin{cases} z_{t-1}(x), & x \in S_{t-1}, \\ q_{\tau,t}(x), & x \in C_t \setminus S_{t-1}, \end{cases} \qquad \bar{\epsilon}_t = \max\{\nu_t, \epsilon_t\}.$$

Stored scores take precedence on overlaps, so the hybrid error on $S_{t-1} \cup C_t$ is at most $\bar{\epsilon}_t$. Exact stored labels have $\nu_t = 0$. Using the full family $\mathcal{F}_t$, define

$$\begin{aligned} F^{\mathrm{cum}}_{h,t} &= \max_{Q \in \mathcal{F}_t} V_K(S_{t-1} \cup Q; h), \\ \Delta^{\mathrm{cum}}_{\mathcal{F}_t} &= V_K(S_{t-1} \cup C_t; f) - F^{\mathrm{cum}}_{f,t}, \\ r^{\mathrm{cum}}_{q,t} &= F^{\mathrm{cum}}_{\widetilde{q}_t,t} - V_K(S_t; \widetilde{q}_t). \end{aligned} \tag{E.11}$$

Both gaps are nonnegative. All evaluated sets contain at least K designs, without requiring $|Q| \geq K$. This is a diagnostic objective; the method is not assumed to optimize it.

[Cumulative quality retention] Under the preceding assumptions, for $t > t_0$,

$$\begin{aligned} 0 \leq U_{K,t} &\leq \Delta^{\mathrm{cum}}_{\mathcal{F}_t} + 2\bar{\epsilon}_t + r^{\mathrm{cum}}_{q,t}, \\ V_K(S_t; f) - V_K(S_{t-1}; f) &\geq \left[O_{K,t} - \Delta^{\mathrm{cum}}_{\mathcal{F}_t} - 2\bar{\epsilon}_t - r^{\mathrm{cum}}_{q,t}\right]_+. \end{aligned} \tag{E.12}$$

Proof. Lemma E.2 applied to every $S_{t-1} \cup Q$ gives $|F^{\mathrm{cum}}_{\widetilde{q}_t,t} - F^{\mathrm{cum}}_{f,t}| \leq \bar{\epsilon}_t$. Hence

$$\begin{aligned} V_K(S_t; f) &\geq V_K(S_t; \widetilde{q}_t) - \bar{\epsilon}_t \\ &= F^{\mathrm{cum}}_{\widetilde{q}_t,t} - r^{\mathrm{cum}}_{q,t} - \bar{\epsilon}_t \\ &\geq F^{\mathrm{cum}}_{f,t} - r^{\mathrm{cum}}_{q,t} - 2\bar{\epsilon}_t. \end{aligned}$$

Subtracting from $V_K(S_{t-1} \cup C_t; f)$ proves the first bound. The identity equation E.10 and nonnegative cumulative improvement prove the second. □

[Transfer from batch loss to cumulative loss] For $t > t_0$ and $|Q_t| \geq K$, one has $U_{K,t} \leq L_{K,t}$.

Proof. For $|A| \geq K$, let $v_K(A)$ be its Kth largest objective value. For $x \notin A$,

$$V_K(A \cup \{x\}; f) - V_K(A; f) = \frac{1}{K}[f(x) - v_K(A)]_+.$$

If $A \subseteq B$ and $|A| \geq K$, then $v_K(A) \leq v_K(B)$; adding the same design has no larger gain for B, including zero gain if it already belongs to B. Add $C_t \setminus Q_t$, in the same order, to Q_t and $S_{t-1} \cup Q_t$. Summing gives

$$\begin{aligned} U_{K,t} &= V_K(S_{t-1} \cup C_t; f) - V_K(S_{t-1} \cup Q_t; f) \\ &\leq V_K(C_t; f) - V_K(Q_t; f) = L_{K,t}. \end{aligned}$$

□

Consequently, under the batch-size and scoring assumptions of Theorem E.2,

$$\begin{aligned} 0 \leq U_{K,t} &\leq \min\{O_{K,t}, L_{K,t}\}, \\ V_K(S_t; f) - V_K(S_{t-1}; f) &\geq [O_{K,t} - \Delta_{\mathcal{F}_t} - 2\epsilon_t - r_{q,t}]_+. \end{aligned} \tag{E.13}$$

If these assumptions hold at every $t > t_0$, then

$$J_K \geq V_K(S_{t_0}; f) + \sum_{t=t_0+1}^{T} [O_{K,t} - \Delta_{\mathcal{F}_t} - 2\epsilon_t - r_{q,t}]_+. \tag{E.14}$$

Summing equation E.12 gives the corresponding bound with cumulative gaps and $\bar{\epsilon}_t$, including smaller batches. A positive bracket ensures strict improvement. Batch and cumulative gaps need not coincide. The positive part follows from retaining earlier discoveries, not from nonnegative objective values; comparing different policies requires their own opportunities and losses.

E.4 Local Quantile Error and Stability

Fix the retained reference $R_{t-1} = r(H_{t-1})$ and $1 \leq k \leq |R_{t-1}|$. Neighbors use normalized Hamming distance and the prescribed tie rule. Write

$$r_{k,t}(x) = \max_{x_i \in \mathcal{N}_k(x; R_{t-1})} d_H(x, x_i).$$

For ordered neighbor labels $v_{[0]} \leq \cdots \leq v_{[k-1]}$, the method uses

$$\begin{aligned} j &= \lfloor (k-1)\tau \rfloor, & \theta &= (k-1)\tau - j, \\ q_{\tau,t}(x) &= (1-\theta)v_{[j]} + \theta v_{[\min\{j+1,k-1\}]}, & \tau &\in [0,1]. \end{aligned} \tag{E.15}$$

This includes $k = 1$ and both endpoints of τ.

[Local quantile error] Suppose every relevant candidate–neighbor pair satisfies $|f(x_i) - f(x)| \leq L_f d_H(x_i, x)$ and $|y_i - f(x_i)| \leq \sigma$, with $L_f, \sigma \geq 0$. Then

$$|q_{\tau,t}(x) - f(x)| \leq L_f r_{k,t}(x) + \sigma \qquad (\tau \in [0,1]). \tag{E.16}$$

Proof. Each neighbor label satisfies $|y_i - f(x)| \leq \sigma + L_f r_{k,t}(x)$. Every order statistic lies in the same interval centered at $f(x)$, and the interpolated quantile is a convex combination of two such values. □

For a nonempty pool satisfying these assumptions, one may therefore take

$$\epsilon_t = L_f \max_{x \in C_t} r_{k,t}(x) + \sigma, \qquad L_{K,t} \leq \Delta_{\mathcal{F}_t} + 2L_f \max_{x \in C_t} r_{k,t}(x) + 2\sigma + r_{q,t}. \tag{E.17}$$

Setting $\sigma = 0$ gives the main-text coverage bound. The radius uses the actual retained reference, not the full history. The same interval argument applies to the neighbor mean; it establishes error control, not superiority of an upper quantile.

[Stability for fixed neighbors] Fix neighbor identities, k, and τ. If corresponding labels obey $|\widetilde{y}_i - y_i| \leq \delta$, then $|\widetilde{q}_{\tau,t}(x) - q_{\tau,t}(x)| \leq \delta$.

Proof. At least $j+1$ original labels are at most $v_{[j]}$; their perturbed values are at most $v_{[j]} + \delta$. Thus $\widetilde{v}_{[j]} \leq v_{[j]} + \delta$. Exchanging the vectors gives $|\widetilde{v}_{[j]} - v_{[j]}| \leq \delta$ for every rank. The fixed interpolation weights then give $|\widetilde{q} - q| \leq (1-\theta)\delta + \theta\delta = \delta$. □

For fixed labels, the quantile is nondecreasing in τ and equals the neighbor maximum at $\tau = 1$; each interpolation segment has nonnegative slope $(k-1)(v_{[j+1]} - v_{[j]})$ when $k > 1$. An upper level expresses a preference for higher measured neighbors, not an outlier-rejection guarantee. Stability concerns bounded label changes with fixed neighbors, not reference replacement or unbounded contamination. Capped references and candidate locations may change between rounds, so coverage error need not decrease monotonically.

E.5 Explicit Prior Estimation and Shortlist Retention

Fix a round with training observations $\{(x_i, y_i)\}_{i=1}^n$, $n \geq 1$, and $y_i = f(x_i) + e_i$. The rank weights $w_i = w_{i,t} > 0$ and $\lambda > 0$ define the same unnormalized weighted-ridge objective as in Method:

$$(c_t, \beta_t) = \arg\min_{c,\beta} \left\{ \sum_{i=1}^n w_i (y_i - c - \beta^\top \phi(x_i))^2 + \lambda \|\beta\|_2^2 \right\}, \qquad p_t(x) = c_t + \beta_t^\top \phi(x). \quad \text{(E.18)}$$

The intercept is unpenalized. Rescaling the data loss requires the same rescaling of λ to preserve the fitted model. No distributional assumption is imposed on e_i.

Let $w_+ = \sum_i w_i$, $\bar{\phi}_w = w_+^{-1} \sum_i w_i \phi(x_i)$, and $\bar{v}_w = w_+^{-1} \sum_i w_i v_i$ for any training vector v. Write $\widetilde{v} = v - \bar{v}_w \mathbf{1}$, $\widetilde{\phi}_t(x) = \phi(x) - \bar{\phi}_w$, and let $\widetilde{\Phi}_t$ have rows $\widetilde{\phi}_t(x_i)^\top$. Set

$$W_t = \mathrm{diag}(w_1, \ldots, w_n), \qquad V_t = \widetilde{\Phi}_t^\top W_t \widetilde{\Phi}_t + \lambda I.$$

Then V_t is positive definite. For arbitrary comparison parameters $(c_\star, \beta_\star)$, define $r(x) = f(x) - c_\star - \beta_\star^\top \phi(x)$ on training and candidate points. This residual differs from the reference operator $r(H)$. Use $\bar{r}_w, \widetilde{r}$ and $\bar{e}_w, \widetilde{e}$ for weighted means and centered training vectors, and $\|v\|_M = (v^\top M v)^{1/2}$ for positive definite M.

[Weighted ridge prediction error] For every analyzed candidate,

$$\begin{aligned} |p_t(x) - f(x)| &\leq |r(x) - \bar{r}_w| + |\bar{e}_w| \\ &\quad + \|\widetilde{\phi}_t(x)\|_{V_t^{-1}} \left(\|W_t^{1/2} \widetilde{r}\|_2 + \|W_t^{1/2} \widetilde{e}\|_2 + \sqrt{\lambda} \|\beta_\star\|_2 \right). \end{aligned} \quad \text{(E.19)}$$

Proof. Eliminating the intercept gives $c_t = \bar{y}_w - \beta_t^\top \bar{\phi}_w$ and $V_t \beta_t = \widetilde{\Phi}_t^\top W_t \widetilde{y}$. Since $\widetilde{y} = \widetilde{\Phi}_t \beta_\star + \widetilde{r} + \widetilde{e}$,

$$\begin{aligned} \beta_t - \beta_\star &= -\lambda V_t^{-1} \beta_\star + V_t^{-1} \widetilde{\Phi}_t^\top W_t (\widetilde{r} + \widetilde{e}), \\ p_t(x) - f(x) &= \widetilde{\phi}_t(x)^\top (\beta_t - \beta_\star) - (r(x) - \bar{r}_w) + \bar{e}_w. \end{aligned} \quad \text{(E.20)}$$

Set $A_t = W_t^{1/2} \widetilde{\Phi}_t$. Since $V_t = A_t^\top A_t + \lambda I$,

$$V_t^{-1/2} A_t^\top A_t V_t^{-1/2} = I - \lambda V_t^{-1} \preceq I,$$

so $\|V_t^{-1/2} A_t^\top\|_{\mathrm{op}} \leq 1$. Also $V_t \succeq \lambda I$ implies $\lambda \|V_t^{-1/2} \beta_\star\|_2 \leq \sqrt{\lambda} \|\beta_\star\|_2$. Cauchy–Schwarz therefore gives

$$\begin{aligned} |\widetilde{\phi}_t(x)^\top (\beta_t - \beta_\star)| &\leq \|\widetilde{\phi}_t(x)\|_{V_t^{-1}} \\ &\quad \times \left(\sqrt{\lambda} \|\beta_\star\|_2 + \|W_t^{1/2} \widetilde{r}\|_2 + \|W_t^{1/2} \widetilde{e}\|_2 \right). \end{aligned}$$

Combining this inequality with equation E.20 proves the claim. □

Exact labels set $e_i = 0$, giving

$$|p_t(x) - f(x)| \leq |r(x) - \bar{r}_w| + \|\widetilde{\phi}_t(x)\|_{V_t^{-1}} \left(\|W_t^{1/2} \widetilde{r}\|_2 + \sqrt{\lambda} \|\beta_\star\|_2 \right). \quad \text{(E.21)}$$

These terms separate representation residuals, measured feature geometry, label error, and regularization bias. Rank weights give high-scoring observations no smaller fitting weight; single-site and pairwise features specify represented effects. Neither property automatically controls candidate error. Conditioning on realized weights permits their dependence on measured ranks. Complementary local-edit sources do not require this explicit prior.

[Retention by a score-ranked prefix] Let $P = \mathrm{Top}_m(A; h)$, with the method's fixed tie rule, and $|P| \geq K$. If $\|h - f\|_{\infty, A} \leq \delta$, then

$$0 \leq V_K(A; f) - V_K(P; f) \leq 2\delta. \quad \text{(E.22)}$$

Proof. The prefix retains an h-maximizing K-element subset even under ties. Consequently,

$$V_K(P; f) \geq V_K(P; h) - \delta = V_K(A; h) - \delta \geq V_K(A; f) - 2\delta.$$

The lower loss bound follows from inclusion. □

For a source using model-ranked truncation, let $G_{a,t}$ contain its distinct admissible proposals immediately before truncation, with $P_{a,t} = \mathrm{Top}_{M_a}(G_{a,t}; p_t)$ and $|P_{a,t}| \geq K$. Then $\|p_t - f\|_{\infty, G_{a,t}} \leq \delta_{a,t}$ implies

$$0 \leq V_K(G_{a,t}; f) - V_K(P_{a,t}; f) \leq 2\delta_{a,t}. \tag{E.23}$$

A valid $\delta_{a,t}$ is the maximum right-hand side of equation E.19 over $G_{a,t}$, or of equation E.21 for exact labels.

[Retention through two score-based prefixes] Let $P = G$ or $P = \mathrm{Top}_M(G; p)$, and $E = \mathrm{Top}_\ell(P; q)$ with $K \leq \ell \leq |P|$. If $\|q - f\|_{\infty,P} \leq \epsilon_q$ and, when $P \neq G$, $\|p - f\|_{\infty,G} \leq \epsilon_p$, then

$$V_K(E; f) \geq V_K(G; f) - 2\delta_p - 2\epsilon_q, \qquad \delta_p = \begin{cases} 0, & P = G, \\ \epsilon_p, & P \neq G. \end{cases} \tag{E.24}$$

Proof. Apply Lemma E.5 to $P \to E$, and also to $G \to P$ when $P \neq G$. Add the two losses; identity truncation has zero loss. □

Thus $V_K(E_{a,t}; f) \geq V_K(G_{a,t}; f) - 2\delta_{a,t} - 2\epsilon_t$ under the corresponding bounds and $\ell_{a,t} \geq K$. Diversity-based cutoff-tie resolution preserves the prefix property if strictly higher-scoring candidates retain priority. These results concern explored proposals, not unproposed regions, and do not require every source or shortlist to contain K designs. Subsequent traversal, merging, deduplication, and completion are assessed on the final Q_t by Theorem E.2.

E.6 Shared Scoring and Source-Score Scales

[Invariance of common selection] Fix history, offered pools, source memberships, quotas, common scores, diversity rules, and internal random choices. Suppose Selection uses internal scalar scores only through within-source priorities, including ties and completion. Replacing those scores while preserving the pools and all such priorities leaves both Q_t and α_t unchanged. In particular, strictly increasing source-specific transformations preserve the result when outputs are fixed and priorities use only within-source ranks.

Proof. The same history assigns identical common scores to the same candidates. Preserved priorities, distances, and random choices give identical screening and diversity outputs. During merging, induction over the unchanged list order gives the same selected set, eligibility checks, and duplicate checks at every step. Completion obeys the same argument, so both the final batch and its attribution agree. □

Pooling native scores does not have this property: scores 0 and 1 from different sources reverse order after adding 2 only to the first source. Common scoring removes this dependence in the stated conditional comparison, without asserting better accuracy. Transformations that change proposals or magnitude-dependent priorities fall outside it.

E.7 Farthest-First Selection Within a Score-Selected Set

Fix a finite working set W, $2 \leq m \leq |W|$, and one prescribed $u_1 \in W$. Starting from $U_1 = \{u_1\}$, let

$$u_j \in \arg\max_{x \in W \setminus U_{j-1}} \min_{z \in U_{j-1}} d_H(x, z), \qquad U_j = U_{j-1} \cup \{u_j\}, \quad 2 \leq j \leq m. \tag{E.25}$$

Define

$$\mathrm{sep}(U) = \min_{\substack{x,z \in U \\ x \neq z}} d_H(x, z), \qquad d_m^\star(W) = \max_{\substack{U \subseteq W \\ |U| = m}} \mathrm{sep}(U).$$

[Within-working-set separation] The single-start traversal satisfies

$$\mathrm{sep}(U_m) \geq \tfrac{1}{2} d_m^\star(W). \tag{E.26}$$

Proof. Let $d_j = \min_{z \in U_{j-1}} d_H(u_j, z)$. Distances to the selected set cannot increase as it grows, so $d_2 \geq \cdots \geq d_m$. Each pair is considered when its later endpoint is inserted, hence $\mathrm{sep}(U_m) = d_m$. The last selection step also gives $\max_{x \in W} \min_{z \in U_{m-1}} d_H(x, z) = d_m$. Thus $m-1$ closed balls of radius d_m, centered at U_{m-1}, cover W. Assign each of the m points of an optimal separated set to a covering center. Two distinct points x, z share some center u, so the triangle inequality gives

$$d_m^\star(W) \leq d_H(x, z) \leq d_H(x, u) + d_H(u, z) \leq 2d_m.$$

Substitute $d_m = \mathrm{sep}(U_m)$. □

The working set is a high-score shortlist or cutoff-tie group: score screening prioritizes quality, while the traversal promotes separation within it. The guarantee is not for arbitrary forced multiple starts or the final batch after source merging and completion; those operations remain in $r_{q,t}$. Minimum separation is not final-set mean pairwise diversity, functional diversity, information gain, or objective improvement.

E.8 Feedback Timing, Discounted Allocation, and Online Trajectories

Query-before-feedback ordering. Let $\mathcal{I}_{t-1}$ contain the initial data, past internal randomness, and completed queries and observations. Let ξ_t collect current internal randomness available before oracle access, and define

$$\mathcal{D}_t = \mathcal{I}_{t-1} \vee \sigma(\xi_t).$$

If controller state is $\mathcal{I}_{t-1}$-measurable and each stage is a measurable function of its inputs and available randomness, composition makes quotas and (Q_t, α_t) $\mathcal{D}_t$-measurable. Quotas without fresh randomness are already $\mathcal{I}_{t-1}$-measurable. After fixing the batch and attribution, observe responses Y_t and set $\mathcal{I}_t = \mathcal{D}_t \vee \sigma(Y_t)$. A measurable state update permits the same construction next round; induction establishes that current outcomes may change later rounds, not the committed batch or its source attribution.

Discounted source returns. For $Q_t^a = \{x \in Q_t : \alpha_t(x) = a\}$, the method computes

$$C_{a,t} = \begin{cases} |Q_t^a|^{-1} \sum_{x \in Q_t^a} h_t(f(x)), & |Q_t^a| > 0, \\ 0, & |Q_t^a| = 0. \end{cases}$$

The prescribed identity or whole-batch standardization affects allocation, not labels stored in H_t. Returns summarize evaluated, attributed queries, not entire source pools. With $S_{a,1} = s_0$, $N_{a,1} = 1$, and $\rho \in (0, 1]$,

$$S_{a,t+1} = \rho S_{a,t} + C_{a,t}, \qquad N_{a,t+1} = \rho N_{a,t} + 1.$$

The scalar $S_{a,t}$ is distinct from the cumulative design set S_t.

[Discounted comparison of realized returns] For every source and round,

$$\begin{aligned} S_{a,t} &= \rho^{t-1} s_0 + \sum_{s=1}^{t-1} \rho^{t-1-s} C_{a,s}, \\ N_{a,t} &= \rho^{t-1} + \sum_{s=1}^{t-1} \rho^{t-1-s} = \sum_{j=0}^{t-1} \rho^j =: N_t. \end{aligned} \tag{E.27}$$

Consequently, softmax weights with temperature $\eta > 0$ satisfy

$$\log \frac{\pi_{a,t}}{\pi_{b,t}} = \frac{1}{\eta N_t} \sum_{s=1}^{t-1} \rho^{t-1-s} (C_{a,s} - C_{b,s}). \tag{E.28}$$

Proof. The expansions hold at $t = 1$ with empty sums. Multiplying by ρ and adding the new return or 1 proves the next round. Every source receives the same count increment, including those with no attributed queries, hence all counts equal N_t. The softmax definition gives $\log(\pi_{a,t}/\pi_{b,t}) = (S_{a,t} - S_{b,t})/(\eta N_t)$; substitution cancels the common initialization. □

For $0 < \rho < 1$, more recent return differences receive larger coefficients; $\rho = 1$ weights past rounds equally. Minimum quotas reserve initial allocation opportunities, not guaranteed final attributed counts when pools overlap or are exhausted. A finite bonus b_t common to every source cancels from softmax:

$$\frac{\exp((u_{a,t} + b_t)/\eta)}{\sum_b \exp((u_{b,t} + b_t)/\eta)} = \frac{\exp(u_{a,t}/\eta)}{\sum_b \exp(u_{b,t}/\eta)}.$$

In particular, a bonus depending only on equal discounted counts and a source-independent round statistic adds no source-specific exploration. These identities characterize discounted-return allocation, not a bandit regret guarantee.

Matched online trajectories. History updates change proposal models, measured elites, and the shared reference; allocation changes how future queries are distributed. For two runs u, v sharing the first batch and attribution, with $|Q_1^u| = |Q_1^v| \geq K$ and equal horizons and budgets, equation E.10 gives

$$J_K^u - J_K^v = \sum_{t=2}^{T} \left[(O_{K,t}^u - O_{K,t}^v) - (U_{K,t}^u - U_{K,t}^v) \right]. \tag{E.29}$$

Each term uses its run's own history and candidate pools. If the first batches differ, their top-K difference must be added. The identity has no predetermined sign: better selection on a fixed pool need not improve discovery along a different feedback history. Quantities requiring unqueried objective values are post-hoc diagnostics and never search inputs.