MERLE: MANY-OBJECTIVE EXTRAPOLATIVE SEARCH VIA REPRESENTATION DISTILLATION FOR BIOLOGICAL LANDSCAPE EXPLAINABILITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Biological landscape search balances many objectives with limited measurements. Restricted initial observations demand extrapolation under substantial predictive uncertainty. We study sequential many-objective search where each query reveals a complete response profile, aiming to expand the observed Pareto frontier under a fixed measurement budget. We introduce MERLE, which distills pretrained mutation preferences into interpretable, objective-specific priors and combines them with an adaptive position-level response model. Together they guide frontier search and reveal which positions matter to the campaign. At each position, the model pools substitutions to estimate assay effects and uncertainty. MERLE scores candidates against unmet frontier requirements and allocates queries across teacher-guided, position-only, and embedding-coverage proposals based on observed frontier gains. Across diverse landscapes and altered campaign settings, MERLE improves frontier discovery over competitive baselines. Ablations identify when the teacher and position field help; site deletion and independent-assay measurements link frontier gains to consequential, reproducible choices of mutation positions across laboratories and distinct antibody and serum panels studied.

1 INTRODUCTION

Biological design rarely has a single criterion of success. A protein variant may need to retain activity across several conditions, trade potency against stability, or exhibit a response profile suited to a particular downstream use (Fram et al., 2024; Luo et al., 2025). The desired output is therefore often a repertoire of nondominated, experimentally validated variants rather than one predicted optimum (Stanton et al., 2022). Finding such a repertoire is difficult when measurements are scarce and the useful region of the landscape differs systematically from the initially observed region. Each experiment must both offer a chance of immediate discovery and inform the experiments that follow, making the task one of sequential many-objective search under extrapolation.

Prior work addresses complementary parts of this task. Black-box optimization methods adaptively propose biological sequences under expensive evaluation (Angermueller et al., 2020), and hypervolume-based Bayesian optimization directly values improvements to a Pareto frontier (Daulton et al., 2020). LaMBO combines this principle with an autoencoder and multi-task Gaussian process for biological sequence design (Stanton et al., 2022); MosPro uses neural property predictors to guide discrete sampling toward multi-property improvements (Luo et al., 2025); and Pareto-conditioned diffusion generates sequences for desired objective trade-offs from offline data (Shrestha et al., 2026). These approaches show the benefit of modeling trade-offs explicitly across both adaptive and offline design.

Recent work suggests that interpretability can instead participate in search. BioBO combines pretrained gene representations with pathway-enrichment priors, providing pathway-level explanations while guiding perturbation selection (Li et al., 2026); IDEAS identifies important sequence motifs with an explainable model and uses them to steer evolutionary mutations (Pandey et al., 2026). Yet neither addresses many-condition sequence landscapes where prior knowledge may have a different, even opposite, meaning for each response. Protein language models make this issue concrete. Their likelihoods encode useful evolutionary constraints (Meier et al., 2021; Lin et al., 2023), but

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.

zero-shot fitness performance varies substantially across assays (Notin et al., 2023). Evolutionary plausibility is not an assay-independent fitness function: a mutation disfavored in nature may reduce one phenotype while improving escape in another. Thus, a useful prior must be adapted and inspected separately for each objective as measurements arrive.

Existing multi-objective methods commonly derive their flexibility from learned latent spaces, generators, or neural property predictors. These are effective search engines, but a predicted value or acquisition score alone does not identify which biological regularity supports a decision, whether it transfers beyond the observed region, or why it favors one response condition over another. We therefore introduce MERLE, a framework for *many-objective extrapolative search via representation distillation*. MERLE compresses a pretrained model's mutation preferences into a small, readable representation and learns how each objective interprets it. This assay-adaptive reading reveals when and in which direction pretrained knowledge transfers; uncertainty over the unexplored landscape then guides experiments toward expansion of the measured functional frontier. Rather than explaining an arbitrary black-box prediction after selection, MERLE exposes a transferable signal that participates directly in selection and can be revised as measurements accumulate.

Contributions. First, we distill masked-token preferences from a pretrained protein model into a compact and readable prior over candidate mutations, avoiding repeated teacher inference during search while preserving a biologically meaningful ordering of substitutions. Second, we learn an objective-specific interpretation of this prior, exposing both the direction and empirical support of its transfer to every assay response. Third, we integrate the adaptive prior with uncertainty-aware, frontier-oriented acquisition in a scalable sequential policy designed for limited experimental budgets. Finally, we evaluate predictive generalization, uncertainty calibration, frontier expansion, and the learned interpretations across diverse biological landscapes, separating better landscape modeling from better discovery.

2 RELATED WORK

Sequential model-guided sequence design. PEX models low-order mutational effects to favor local search (Ren et al., 2022); GFlowNet-based active learning uses predictive uncertainty to propose diverse, informative batches (Jain et al., 2022); and LaMBO brings multi-objective Bayesian optimization to protein sequences (Stanton et al., 2022). These methods learn proposals or surrogates to search sequence space. In contrast, MERLE selects from a finite variant library and receives the complete response profile for each query, making one measurement informative across several trade-offs. With a fixed assay panel, one chooses the next available variant, not a new sequence.

Multi-objective protein design. Evolutionary multi-objective methods preserve diverse trade-offs instead of reducing them to a scalar objective, with NSGA-II a widely used example (Deb et al., 2002). Hong & Kortemme (2024) combine sequence, structure, and folding models in an NSGA-II protein-design pipeline; MP2D uses diffusion and constrained tree search with Pareto rewards to generate multi-property sequences (Kong et al., 2026). Offline multi-objective optimization instead assumes a fixed dataset and does not acquire new oracle labels during search (Xue et al., 2024). The closest active-learning line, A-GPS, couples iterative queries of discrete black-box objectives with a preference-conditioned generator, including on protein-design benchmarks (Steinberg et al., 2025). MERLE instead studies campaign continuation over a fixed library initialized by prior measurements, with each assay revealing a complete multi-condition response profile.

Biological priors and interpretability. Simon & Zou (2025) extract interpretable features from ESM-2 representations; Hunklinger & Ferruz (2026) survey explainability across protein language models. KnowRLM combines an amino-acid knowledge graph with a PLM policy and active fitness feedback for directed evolution (Wang et al., 2024). IDEAS makes motif signals actionable by using them to guide evolutionary sequence design (Pandey et al., 2026). MERLE instead learns an objective-specific reading of mutation preferences from assay evidence and uses that reading to prioritize subsequent experiments. Because pretrained preferences may reverse across conditions, MERLE estimates signed objective-specific readings and shrinks unsupported transfer toward zero. These readings directly shape the next assay, making interpretation part of selection.

3 MANY-OBJECTIVE EXTRAPOLATIVE SEQUENCE SEARCH

Preliminaries. For objective profiles $a, b \in [0,1]^C$, a dominates b, written $a \succ b$, if $a_c \geq b_c$ for every objective and the inequality is strict for at least one. Given a response mapping f and an evaluated set V, the Pareto set and its image, the Pareto front, are

$$\mathcal{P}(V) = \{x \in V : \nexists y \in V,\ f(y) \succ f(x)\}, \qquad \mathcal{F}(V) = \{f(x) : x \in \mathcal{P}(V)\}. \tag{1}$$

The hypervolume (HV) of a front is the volume it dominates relative to a reference point, jointly measuring response quality and trade-off coverage. We call $C \in \{2,3\}$ multi-objective and $C \geq 4$ many-objective, and use the zero vector as the reference point.

Campaign-to-campaign discovery. Let $\mathcal{X}$ be a finite sequence library and $f : \mathcal{X} \to [0,1]^C$ an unknown response function, with all objectives to be maximized. Initially, complete profiles are available for $\mathcal{D}_0 = \{(x_i, f(x_i))\}_{i=1}^N$. This warm start represents a common experimental workflow: $\mathcal{D}_0$ may contain variants measured in one or more preceding campaigns, while a follow-up campaign receives a fresh budget B for selecting previously unmeasured variants. At step t, one query returns the full C-condition profile,

$$x_t = \pi(\mathcal{D}_{t-1}) \in \mathcal{X} \setminus V_{t-1}, \qquad \mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(x_t, f(x_t))\}, \tag{2}$$

where V_t denotes all variants observed through step t. Initial measurements do not consume the new campaign's budget, and responses of unqueried variants remain unavailable.

Relation to existing settings. Existing formulations either stop at a static dataset (Xue et al., 2024) or continue by searching a learned latent or generative sequence space (Stanton et al., 2022; Steinberg et al., 2025). Here, search instead continues within a fixed experimental pool: each query selects an existing variant, jointly reveals all conditions, and must extrapolate beyond the response region covered by the preceding campaign rather than merely refine an already characterized region.

Discovery objective. Writing V_0 for the initially evaluated variants and V_B for those observed after B queries, the goal is to learn a policy that maximizes expected hypervolume gain,

$$\pi^\star \in \arg\max_{\pi \in \Pi} \mathbb{E}_\pi \left[\mathrm{HV}(\mathcal{F}(V_B)) - \mathrm{HV}(\mathcal{F}(V_0)) \mid \mathcal{D}_0 \right]. \tag{3}$$

The expectation covers response uncertainty and policy randomness. A query may be valuable even without immediate frontier improvement when it informs later choices.

4 METHODOLOGY

MERLE builds a structured response model that combines evidence shared across substitutions at a position with calibrated pretrained preferences. Its acquisition A_t ranks proposals by expected frontier gain, with or without teacher-based substitution resolution. Embedding-based coverage supplies complementary measurements, while realized gains allocate queries between these mechanisms. Each new profile updates the response field, unmet frontier requirements, and subsequent source allocation. We develop the response model, acquisition, and allocation in turn.

4.1 TRANSFER FROM PRETRAINED PREFERENCES

Transferring mutation preferences. Here, *representation distillation* means reducing a frozen PLM's contextual mutation preferences to a scalar per variant, then learning how that scalar transfers to each measured condition. This is a compact readout of the teacher, without training a student network or matching hidden states. For a variant v replacing residue a_p^{wt} with $a(v)$ at position $p = p(v)$, one masked ESM-2 call per position (Lin et al., 2023) supplies a mutant-versus-wild-type log-odds score (Meier et al., 2021):

$$\Delta_v = \log \frac{p_{\mathrm{PLM}}(a(v) \mid x_{\setminus p})}{p_{\mathrm{PLM}}(a_p^{\mathrm{wt}} \mid x_{\setminus p})}, \qquad \ell_v = \frac{\Delta_v - \mathbb{E}_u[\Delta_u]}{\mathrm{SD}_u[\Delta_u]}. \tag{4}$$

We standardize and cache this sequence-level score as ℓ_v, then learn its signed transfer to each assay condition. The score's role is determined by measured responses; unscored variants receive zero.

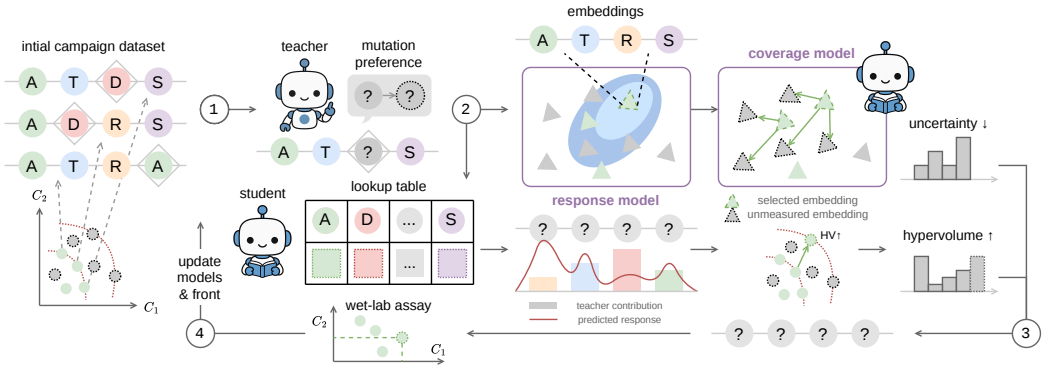


Figure 1: **MERLE in four steps.** (1) An initial campaign supplies measured variants, while a pretrained teacher supplies mutation preferences. (2) A compact preference lookup, response model, and embedding coverage model characterize unmeasured variants. (3) Frontier gain and uncertainty guide the next query. (4) The new measurement updates the models and the cycle repeats.

A mutation preferred by the pretrained model need not improve every experimental response. We therefore learn a signed coefficient for each condition, using the probit response $z_{vc} = \Phi^{-1}(f_c(v))$:

$$z_{vc} = \theta_{p(v),c} + \gamma_c \ell_v + \epsilon_{vc}, \qquad \epsilon_v \sim \mathcal{N}(0, \Sigma_{p(v)}). \tag{5}$$

The position effect $\theta_{p,c}$ shares evidence among substitutions at p, while $\gamma_c \ell_v$ distinguishes them through the pretrained ordering. The covariance Σ_p captures their remaining joint variation.

Estimating transferable signal. To fit Equation 5, we alternate ridge regression for the teacher coefficient with Gaussian position updates under $\theta_{p,c} \sim \mathcal{N}(0, \tau_c^2)$ and working residual variance σ_c^2. To control overfitting, we estimate transfer strength on held-out observations. Using two folds, let $h_{vc} = \widehat{\beta}_c^{(-v)} \ell_v$ and $\widehat{\theta}_{p,c}^{(-v)}$ be predictions fitted on the opposite fold. The retained fraction follows from a constrained regression:

$$s_c = \underset{0 \le s \le 1}{\arg\min} \sum_v \left(z_{vc} - \widehat{\theta}^{(-v)}_{p(v),c} - s h_{vc}\right)^2 = \left[\frac{\sum_v h_{vc}(z_{vc} - \widehat{\theta}^{(-v)}_{p(v),c})}{\sum_v h_{vc}^2}\right]_{[0,1]}. \tag{6}$$

Both sums use $v \in V_{t-1}$. Setting $\gamma_c = s_c \widehat{\beta}_c$, with $\widehat{\beta}_c$ refit on all observations, completes the transfer: the distilled score contributes only as supported by held-out responses. The constraint prevents amplification or a new sign reversal.

4.2 Local Response Geometry

From shared directions to local covariance. The mean alone cannot distinguish positions offering different trade-offs. We estimate a shared residual covariance from campaign queries, shrink it toward $\mathrm{diag}(\sigma^2)$, and write its eigenvectors and eigenvalues as u_k and λ_k. Initial observations are excluded from this correlation estimate because their restricted response range can distort dependence. A local covariance $\Sigma_p = \sum_k \lambda_k w_{pk} u_k u_k^\top$ then requires only one variance ratio per direction. Let a_{pk} be the sum of squared projections $(u_k^\top z_v)^2$ at position p, corrected by subtracting the prior position-effect contribution $n_p \sum_c u_{ck}^2 \tau_c^2$. For $n_p > 0$, the ratio $a_{pk}/(n_p \lambda_k)$ compares local variation with the shared direction's variance. We shrink its departure from unity:

$$w_{pk} = \left[1 + \underbrace{\frac{n_p}{\nu_k + n_p}}_{\text{evidence weight}} \cdot \underbrace{\left(\frac{a_{pk}}{n_p \lambda_k} - 1\right)}_{\text{local departure}}\right]_{[w_{\min}, w_{\max}]}. \tag{7}$$

The evidence weight increases with the number of observations, allowing well-measured positions to depart further from the shared covariance. Positive departures expand a direction; negative departures contract it. For an unmeasured position, we set $w_{pk} = 1$. We select ν_k by two-fold held-out Gaussian likelihood, while the bounds keep variances positive and finite. Because these moments use raw response profiles, they can also include teacher-aligned variation.

As queries extend the initial response range, we also update the working variance split. After removing the teacher contribution, within-position scatter estimates σ_c^2, bounded inside $(0, 1)$, and $\tau_c^2 = 1 - \sigma_c^2$ supplies the complementary position variance. These diagonal variances govern position updates; Σ_p supplies joint variation for acquisition.

4.3 Marginal Frontier Expansion

Sampling available opportunities. Let $\mu_{p,c}$ and $q_{p,c}$ be the posterior mean and standard deviation of $\theta_{p,c}$. We temper each draw by its remaining substitution count $m_{p,t}$ relative to the average $\bar{m}_t$ among considered positions. With independent $\xi_{p,c} \sim \mathcal{N}(0, 1)$,

$$\widetilde{\theta}_{p,c} = \mu_{p,c} + \kappa q_{p,c} \left[1 + \eta \left(\sqrt{\frac{m_{p,t}}{\bar{m}_t}} - 1\right)\right]_+ \xi_{p,c}. \tag{8}$$

At average inventory, the draw has standard deviation $\kappa q_{p,c}$; η increases it where more substitutions remain. The rule does not depend on the remaining budget. Substitution into Equation 5 gives $\widetilde{z}_v = \widetilde{\theta}_{p(v)} + \gamma \ell_v + \epsilon_v$, with $\epsilon_v \sim \mathcal{N}(0, \Sigma_{p(v)})$.

Valuing frontier expansion. For $g \sim \mathcal{N}(0, I_C)$, $\Phi(g)$ is uniform on $[0, 1]^C$, so dominated volume is the probability of exceeding this random requirement. Let $d_t(z)$ be the hypervolume gain from adding $\Phi(z)$ to the current front. Excluding requirements already met gives

$$d_t(z) = \mathrm{HV}\big(\mathcal{F}(V_{t-1}) \cup \{\Phi(z)\}\big) - \mathrm{HV}\big(\mathcal{F}(V_{t-1})\big) = \Pr_g(g \le z,\ \nexists u \in V_{t-1} : g \le z_u)\,. \tag{9}$$

We use a fixed cloud of J requirements, with $\mathcal{U}_t$ indexing those still unmet. Conditional on sampled position effects, acquisition becomes their average satisfaction probability:

$$A_t(v) = \mathbb{E}_{\epsilon_v}[d_t(\widetilde{z}_v)] \ \approx\ \frac{1}{J} \sum_{j \in \mathcal{U}_t} \Pr_{\epsilon_v}\Big(\epsilon_v \ge g_j - \widetilde{\theta}_{p(v)} - \gamma \ell_v\Big)\,. \tag{10}$$

We estimate the tails with M shared Gaussian draws under Σ_p (Equation 7). Unmet requirements encode frontier gaps; covariance governs plausible responses. The resulting A_t prices each available variant against the same unmet-requirement cloud.

To reduce ranking cost, we screen by lone-point volume $\mathbb{E}_\epsilon[\prod_c \Phi(\widetilde{z}_{vc})]$ and rerank a shortlist using Equation 10. Teacher-guided proposals use the lowest, middle, and highest remaining score ranks at each position. The two modes share A_t: teacher-guided proposals retain $\gamma \ell_v$ and query the highest-scoring shortlisted variant; position-only proposals omit it, select the highest-scoring position, and sample a remaining substitution there uniformly. The requirement cloud stays fixed across rounds, so the acquisition responds to new measurements rather than changes in its Monte Carlo reference set. Appendix C.4 details candidate sampling and shortlist reranking.

4.4 Adaptive Query Allocation

Supporting the frontier model. An auxiliary coverage source reduces uncertainty across the remaining library. It uses mean-pooled ESM-2 embeddings, reduced by PCA and augmented with an intercept to give $\phi(v) \in \mathbb{R}^d$. These features retain information beyond ℓ_v. A Bayesian linear model links measured responses through them. With posterior weight covariance $S_t \in \mathbb{R}^{d \times d}$, observing v reduces latent variance by

$$C_t(v) = \sum_{u \in \mathcal{X} \setminus V_{t-1}} \phi(u)^\top \big(S_t - S_{t|v}\big)\phi(u) = \sum_{u \in \mathcal{X} \setminus V_{t-1}} \frac{\big(\phi(u)^\top S_t \phi(v)\big)^2}{\sigma^2_{\mathrm{obs}} + \phi(v)^\top S_t \phi(v)}. \tag{11}$$

Here $S_{t|v}$ is the covariance after a hypothetical query and σ^2_{obs} is observation noise variance. Coverage queries the sampled candidate maximizing $C_t(v)$ and updates the response model with its measurement. This pool-wide variance reduction follows the integrated-variance principle of active learning (Cohn et al., 1996). Within MERLE, coverage measurements update the position-aware response field and unmet requirements; their realized frontier gains determine future source allocation.

Table 1: **Full-budget performance** ($B = 192$)**.** Mean over 20 datasets $\times$ 5 seeds, $\pm$ SE across datasets; higher is better, best policy in bold; reference bounds are unranked. Last columns: lexicon gain for the largest third by candidate-pool size (7/20 datasets), and its slope against $\log_{10}|\mathcal{X} \setminus V_0|$ over all 20; brackets are bootstrap 95% CIs and p uses the same draws.

Policy	Venue	dHV ↑	dHV-max ↑	IGD+ ↑	eps ↑	R2 ↑	ΔdHV$_{\text{lex}}$	slope
MESMO	NeurIPS 2019	0.314 $\pm$0.024	0.278 $\pm$0.025	0.156 $\pm$0.015	0.123 $\pm$0.011	0.024 $\pm$0.002	−0.002 [−0.029, +0.022]	+0.015 p 0.46
qEHVI	NeurIPS 2020	0.291 $\pm$0.024	0.256 $\pm$0.024	0.144 $\pm$0.014	0.112 $\pm$0.011	0.022 $\pm$0.002	+0.018 [−0.010, +0.046]	+0.000 p 0.99
BOtied	ICML 2024	0.300 $\pm$0.024	0.265 $\pm$0.025	0.152 $\pm$0.014	0.120 $\pm$0.011	0.023 $\pm$0.002	—	—
MOBO-OSD	NeurIPS 2025	0.283 $\pm$0.023	0.248 $\pm$0.023	0.138 $\pm$0.014	0.104 $\pm$0.010	0.021 $\pm$0.002	—	—
STAGE-BO	ICML 2026	0.310 $\pm$0.022	0.277 $\pm$0.022	0.154 $\pm$0.014	0.121 $\pm$0.011	0.023 $\pm$0.002	+0.023 [+0.007, +0.041]	+0.011 p 0.38
ALC	JAIR 1996	0.303 $\pm$0.022	0.262 $\pm$0.022	0.149 $\pm$0.013	0.108 $\pm$0.011	0.022 $\pm$0.002	+0.018 [+0.000, +0.036]	+0.018 p 0.11
BAIT	NeurIPS 2021	0.289 $\pm$0.019	0.244 $\pm$0.018	0.140 $\pm$0.012	0.100 $\pm$0.011	0.021 $\pm$0.002	—	—
PDBAL	Nat. Commun. 2025	0.276 $\pm$0.020	0.227 $\pm$0.019	0.132 $\pm$0.013	0.090 $\pm$0.012	0.020 $\pm$0.002	—	—
MERLE	—	**0.376** $\pm$**0.017**	**0.341** $\pm$**0.017**	**0.188** $\pm$**0.012**	**0.163** $\pm$**0.011**	**0.030** $\pm$**0.002**	+0.047 [+0.017, +0.075]	+0.054 p 0.01
Random	—	0.288 $\pm$0.015	0.237 $\pm$0.013	0.140 $\pm$0.011	0.091 $\pm$0.009	0.021 $\pm$0.001	—	—
White-box	—	0.490 $\pm$0.008	0.478 $\pm$0.010	0.241 $\pm$0.011	0.247 $\pm$0.008	0.041 $\pm$0.001	—	—

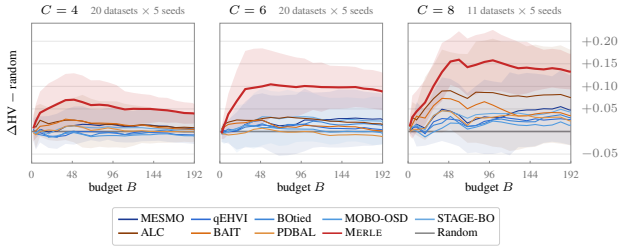


Figure 2: **Gain over random screening as the number of objectives grows.** Paired dHV gain over random, averaged across datasets. Bands show one standard deviation for MERLE and the next two policies in each panel.

Table 2: **Four $\mathcal{D}_0$ rules.** Full-budget dHV (mean $\pm$ standard error over datasets and five seeds). Compare within columns; underlining marks the strongest baseline in each setting.

Policy	uncensored (19)	band (18)	sites (19)	single (19)
MESMO	0.065 $\pm$0.007	0.302 $\pm$0.026	0.326 $\pm$0.025	0.495 $\pm$0.023
qEHVI	0.068 $\pm$0.007	0.296 $\pm$0.026	0.299 $\pm$0.025	0.490 $\pm$0.021
BOtied	0.064 $\pm$0.006	0.303 $\pm$0.025	0.295 $\pm$0.024	0.492 $\pm$0.022
MOBO-OSD	0.066 $\pm$0.006	0.302 $\pm$0.025	0.295 $\pm$0.025	0.477 $\pm$0.023
STAGE-BO	0.068 $\pm$0.007	0.304 $\pm$0.024	0.323 $\pm$0.025	0.489 $\pm$0.024
ALC	0.052 $\pm$0.006	0.306 $\pm$0.023	0.291 $\pm$0.024	0.480 $\pm$0.023
BAIT	0.049 $\pm$0.006	0.266 $\pm$0.020	0.279 $\pm$0.020	0.465 $\pm$0.019
PDBAL	0.041 $\pm$0.005	0.256 $\pm$0.020	0.273 $\pm$0.020	0.454 $\pm$0.020
MERLE	**0.105** $\pm$0.009	**0.347** $\pm$0.021	**0.356** $\pm$0.018	**0.556** $\pm$0.015
Random	0.050 $\pm$0.005	0.280 $\pm$0.020	0.280 $\pm$0.016	0.460 $\pm$0.015
White-box	0.163 $\pm$0.015	0.481 $\pm$0.009	0.488 $\pm$0.010	0.666 $\pm$0.006

Allocating queries by frontier gain. The allocator chooses between the two model-based modes and coverage, which propose queries using A_t or C_t. After an initial trial of each source, it samples a source and rewards the fraction of unmet requirements its query resolves:

$$\widetilde{r}_{a,t} \sim \mathcal{N}\left(\bar{r}_{a,t}, \frac{\widetilde{s}^2_{a,t}}{N_{a,t}}\right), \quad a_t = \arg\max_a \widetilde{r}_{a,t}, \quad r_t = 1 - \frac{|\mathcal{U}_{t+1}|}{|\mathcal{U}_t|}. \tag{12}$$

Here $\bar{r}_{a,t}$, $\widetilde{s}_{a,t}$, and $N_{a,t}$ are the source's empirical mean reward, pooled standard deviation, and query count. Only the selected source receives credit, and $r_t = 0$ when no requirements remain. Teacher guidance uses position-only search until Equation 6 yields a usable coefficient. The response model is refit periodically; Appendix C.5 details the source updates.

5 EXPERIMENTS

We compare MERLE with five multi-objective Bayesian optimization methods: MESMO (Belakaria et al., 2019), qEHVI (Daulton et al., 2020), BOtied (Park et al., 2024), MOBO-OSD (Ngo et al., 2025), and STAGE-BO (Yang et al., 2026); and three active-learning designs: ALC (Cohn et al., 1996), BAIT (Ash et al., 2021), and PDBAL (Tosh et al., 2025). We also report random selection and a white-box oracle as reference bounds. Detailed evaluation settings are given in Appendix D. It covers datasets, campaign design, baseline adaptations, metrics, and uncertainty.

5.1 FRONTIER DISCOVERY ACROSS CAMPAIGN SETTINGS

Full-budget gains persist as the objective count grows. Table 1 shows an advantage across frontier indicators (Appendix D.4), beyond hypervolume alone. Relative to random screening, MERLE recovers 44% of the headroom to the white-box bound. The distilled lexicon contributes most in large candidate pools, where a fixed budget covers fewer variants. This pool-size stratum was defined independently of the gain, which also increases with pool size across the panel. Figure 2 shows gains over random screening at $C = 4$, 6, and 8; MERLE leads each panel at full budget. The separation develops over the campaign rather than appearing only at its end.

Table 3: **Held-out chemistry.** Full-budget dHV with each family excluded from $\mathcal{D}_0$; *crossed* is query enrichment and *across* is the dHV share earned there, averaged over the three held-out families and 19 datasets per arm.

Policy	charged (19)	aromatic (19)	aliphatic (19)	crossed	across
MESMO	0.304 ± 0.023	0.304 ± 0.026	0.299 ± 0.021	0.62×	0.35
qEHVI	0.277 ± 0.023	0.294 ± 0.024	0.287 ± 0.020	0.05×	0.05
BOtied	0.288 ± 0.023	0.290 ± 0.022	0.296 ± 0.021	0.05×	0.05
MOBO-OSD	0.285 ± 0.022	0.291 ± 0.022	0.295 ± 0.022	0.05×	0.05
STAGE-BO	0.299 ± 0.025	0.303 ± 0.025	0.307 ± 0.021	0.45×	0.27
ALC	0.304 ± 0.021	0.292 ± 0.020	0.289 ± 0.021	1.34×	0.52
BAIT	0.283 ± 0.020	0.278 ± 0.018	0.286 ± 0.018	2.05×	0.67
PDBAL	0.259 ± 0.021	0.259 ± 0.021	0.266 ± 0.022	1.89×	0.62
MERLE	**0.354** ± 0.017	**0.361** ± 0.018	**0.362** ± 0.017	1.40×	0.61
Random	0.267 ± 0.016	0.267 ± 0.019	0.278 ± 0.015	1.00×	0.41
White-box	0.481 ± 0.008	0.483 ± 0.009	0.481 ± 0.008	0.99×	0.77

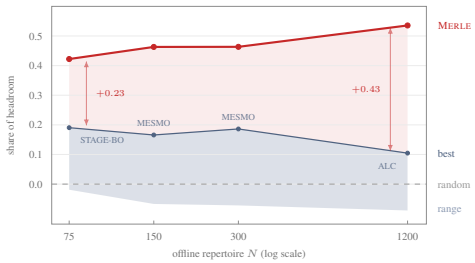


Figure 3: **More offline data favors MERLE.** Share of headroom as $|\mathcal{D}_0|$ increases sixteen-fold, with other settings fixed. "best" is the strongest baseline at each size; "range" spans all eight baseline policies in the panel.

Table 4: **Altered feedback and objectives.** Full-budget dHV (signed headroom share in parentheses); *cost* is half minus complete profile. Noise is scored on clean responses. The new-condition and correlation arms use 12 and 11 datasets; underlining marks the best baseline in each column.

	Half profile		Measurement noise σ			New condition	Objective correlation		
Policy	dHV (share)	cost	0.10	0.25	0.50	+1 objective	Conflicting	Uniform	Aligned
MESMO	0.302 (+0.09)	−0.013	0.324 (+0.22)	0.311 (+0.16)	0.295 (+0.08)	0.352 (+0.23)	0.276 (+0.13)	0.335 (+0.25)	0.387 (+0.12)
qEHVI	0.281 (−0.01)	−0.011	0.289 (+0.02)	0.290 (+0.05)	0.284 (+0.00)	0.335 (+0.14)	0.271 (+0.09)	0.319 (+0.14)	0.369 (−0.16)
BOtied	0.289 (+0.04)	−0.012	0.295 (+0.05)	0.292 (+0.05)	0.297 (+0.07)	0.338 (+0.14)	0.262 (+0.06)	0.325 (+0.17)	0.364 (−0.21)
MOBO-OSD	0.282 (−0.01)	−0.002	0.288 (+0.03)	0.295 (+0.07)	0.297 (+0.07)	0.314 (+0.04)	0.259 (+0.04)	0.304 (+0.04)	0.357 (−0.38)
STAGE-BO	0.305 (+0.12)	−0.005	0.312 (+0.16)	0.306 (+0.14)	0.307 (+0.15)	0.347 (+0.19)	0.277 (+0.13)	0.333 (+0.21)	0.378 (−0.02)
ALC	0.300 (+0.09)	−0.003	0.309 (+0.15)	0.307 (+0.13)	0.309 (+0.14)	0.328 (+0.11)	0.273 (+0.11)	0.316 (+0.12)	0.378 (+0.14)
BAIT	0.293 (+0.04)	+0.004	0.291 (+0.03)	0.290 (+0.03)	0.286 (+0.01)	0.323 (+0.06)	0.259 (+0.03)	0.313 (+0.08)	0.386 (+0.15)
PDBAL	0.279 (−0.05)	+0.003	0.276 (−0.08)	0.280 (−0.05)	0.277 (−0.08)	0.306 (−0.04)	0.230 (−0.10)	0.297 (−0.02)	0.364 (−0.11)
MERLE	**0.362** (+0.40)	−0.014	**0.380** (+0.48)	**0.372** (+0.44)	**0.359** (+0.39)	**0.412** (+0.47)	**0.325** (+0.36)	**0.388** (+0.49)	**0.430** (+0.60)
Random	0.288 (+0.00)	—	0.288 (+0.00)	0.288 (+0.00)	0.288 (+0.00)	0.313 (+0.00)	0.253 (+0.00)	0.303 (+0.00)	0.368 (+0.00)
White-box	0.490 (+1.00)	—	0.485 (+1.00)	0.483 (+1.00)	0.479 (+1.00)	0.549 (+1.00)	0.478 (+1.00)	0.497 (+1.00)	0.498 (+1.00)

The initial repertoire changes which baseline is strongest. Table 2 varies $\mathcal{D}_0$: *uncensored* samples the whole library; *band* restricts the offline responses to their middle range; *sites* samples selected positions; and *single* censors only the first objective. The strongest baseline changes across these rules. Under *sites*, the margin over it is least decisive (95% CI $[-0.001, +0.062]$). Under *uncensored*, MERLE still captures 48% of the random-to-oracle headroom.

The advantage persists as the repertoire changes. Table 3 excludes one chemistry family from $\mathcal{D}_0$ while keeping the full candidate pool. MERLE beats the best baseline by 0.050–0.057 dHV across the three arms, crossing into held-out classes less often than BAIT or PDBAL but earning more dHV there. Figure 3 grows $|\mathcal{D}_0|$ from 75 to 1200: MERLE's headroom share rises from 0.42 to 0.53 while the best baseline's falls from 0.19 to 0.10; at 1200 rows, some baselines are near random screening. Absolute dHV falls as the repertoire grows, but the opposing shares suggest more observations sharpen its site estimates. Together, the two settings indicate that the site field learns from substitutions already seen at a position and ranks new variants there more effectively.

5.2 SEARCH WITH ALTERED FEEDBACK AND OBJECTIVES

The lead persists across four changes to the search problem. Table 4 shows (i) *Partial profiles:* receiving only three of six responses per query lowers MERLE's dHV from 0.376 to 0.362 and its headroom share from 0.46 to 0.40. Its block covariance is approximate, so this does not isolate within-site imputation. (ii) *Noisy measurements:* at $\sigma = 0.50$, MERLE earns 0.359 dHV (0.39 of headroom), above every baseline's clean dHV. Flat baseline rows had little clean margin to lose; this is not evidence of noise immunity. (iii) *Unseen condition:* when a seventh condition is added only at scoring, its share is 0.47 versus 0.46 on the same tables' original six. The campaign never saw it: similar shares indicate transfer, whereas raw dHV levels differ in scale. (iv) *Objective correlation:* MERLE captures 0.36 of headroom on the most conflicting sextet and 0.60 on the most aligned, leading both arms. Five baselines fall below random on the aligned arm.

Table 5: **Targeted demand.** dHV and target headroom share for requirements on three conditions, with a floor on the other three. Upper rows re-score stored campaigns; only rows below the rule were told the demand. Shares use target-specific random/oracle bounds.

Policy	ΔHV	share
MESMO	0.428 ±0.032	+0.04
qEHVI	0.384 ±0.033	−0.14
BOtied	0.400 ±0.032	−0.09
MOBO-OSD	0.389 ±0.035	−0.13
STAGE-BO	0.432 ±0.029	+0.04
ALC	0.410 ±0.031	−0.03
BAIT	0.406 ±0.025	−0.08
PDBAL	0.377 ±0.027	−0.20
MERLE	**0.511** ±0.023	**+0.34**
Random	0.428 ±0.017	+0.00
White-box	0.679 ±0.009	+1.01
arms that were told the demand		
MERLE ▷ demand	**0.527** ±0.020	**+0.40**
White-box ▷ demand	0.679 ±0.010	+1.00

Table 6: **Prior transfer.** New stage-2 conditions; paired dHV gain from carrying the stage-1 prior (95% CI).

Policy	fresh prior	carried	paired difference	improved
MERLE	**0.372**	**0.377**	+0.0052 [−0.0047,+0.0149]	6/9
MESMO	0.325	0.337	+0.0117 [−0.0097,+0.0377]	5/9
STAGE-BO	0.316	0.317	+0.0010 [−0.0200,+0.0192]	5/9
ALC	0.324	0.326	+0.0027 [−0.0072,+0.0119]	6/9
Mean, four policies	0.334	0.339	+0.0051 [−0.0004,+0.0099]	7/9
Random	0.299	—	—	—
White-box	0.476	—	—	—

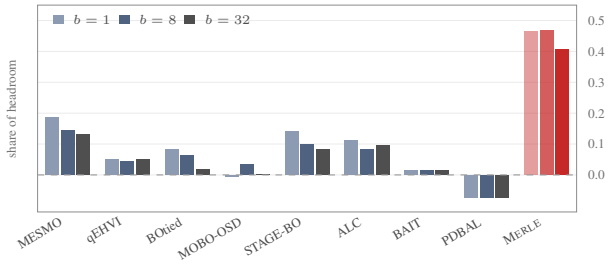


Figure 4: **A plate at a time.** Headroom share after refitting every b queries; $b = 1$ is sequential.

Demand, batching, and transfer. Table 5 re-scores campaigns for a demand concentrated on three conditions. Without seeing it, MERLE captures 0.34 of target headroom; configuring it for that demand raises the share to 0.40, although the paired +0.016 dHV gain has a confidence interval containing zero. Other panel policies have no demand input, so their campaigns are only re-scored. Table 6 starts a second campaign on new conditions with a fresh repertoire. All four policies with a surrogate gain from carrying the prior, but every paired interval contains zero. The gain occurs on only five or six of nine datasets per policy; MERLE's mean gain is +0.0052 dHV. Figure 4 changes when feedback arrives: at batch sizes $b = 1, 8, 32$, MERLE captures 0.46, 0.47, and 0.41 of headroom, staying above every baseline's sequential score with only six model refits.

5.3 ANATOMY AND COST OF THE ADVANTAGE

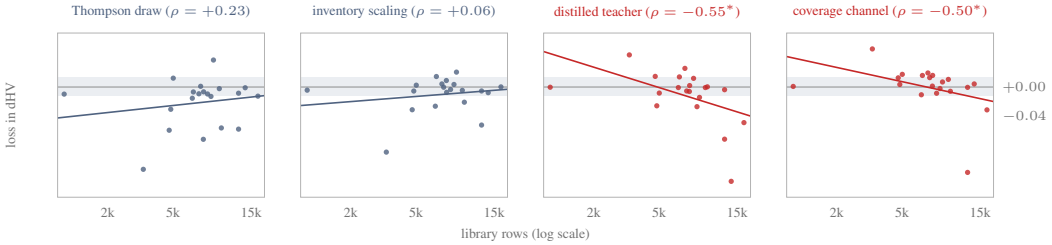


Figure 5: **Ablation loss by library size.** Each point is one library; lines show fitted trends and the grey band marks trajectory noise. Asterisks mark rank correlations with $p < 0.05$.

Table 7: **Ablations by size.**

Part removed	all 20	largest 7	other 13	ρ
Thompson draw	−0.0217	−0.0145	−0.0255	+0.23
Inventory scaling	−0.0102	−0.0100	−0.0104	+0.06
Distilled teacher	−0.0115	−0.0383	+0.0029	−0.55*
Coverage channel	−0.0002	−0.0191	+0.0100	−0.50*
Marginal criterion	−0.0041	−0.0162	+0.0024	−0.27
Within-site axis	−0.0020	−0.0078	+0.0011	−0.19
Variance refresh	−0.0032	−0.0040	−0.0027	+0.02
Thompson allocation	+0.0017	−0.0046	+0.0051	−0.19

Table 8: **Model × acquisition.**

	MESMO acq.	Demand acq., eq. (3)
linear model, eq. (2)	0.314 (+0.19)	0.292 (+0.05)
field model, eq. (4)	0.363 (+0.38)	0.371 (+0.43)
MERLE, as published	**0.376** (+0.46)	

Exploration is broad; reading depends on size. Table 7 re-reads paired component ablations across 20 libraries. Removing the site-effect Thompson draw or inventory scaling lowers dHV across the panel, with little size trend. The teacher instead costs 0.038 dHV when removed from the seven largest libraries (95% CI $[0.009, 0.075]$), but nothing resolved on the other 13; its loss grows with size ($\rho = -0.55$, $p = 0.011$). The coverage channel follows a similar size trend in Figure 5. Table 8 isolates the model's contribution: holding MESMO's acquisition fixed, the position field raises dHV from 0.314 to 0.363. With that field, the demand rule reaches 0.371; the full method scores 0.376.

5.4 What MERLE learns and delivers

Table 9: **Balanced demand.** Share of unmet demand served across quartiles from lopsided (Q1) to balanced (Q4); *tilt* is Q1 minus Q4. The final column is the best balanced point.

Policy	dHV	Q1	Q2	Q3	Q4	tilt	best bal.
MESMO	0.316	0.655	0.644	0.618	0.573	+0.082	0.833
qEHVI	0.293	0.607	0.597	0.571	0.525	+0.082	0.821
BOtied	0.302	0.624	0.614	0.588	0.545	+0.079	0.830
MOBO-OSD	0.285	0.596	0.586	0.558	0.507	+0.088	0.813
STAGE-BO	0.311	0.649	0.635	0.608	0.561	+0.088	0.832
ALC	0.305	0.643	0.628	0.597	0.538	+0.105	0.817
BAIT	0.291	0.622	0.602	0.564	0.498	+0.124	0.809
PDBAL	0.278	0.592	0.574	0.538	0.474	+0.118	0.800
MERLE	**0.377**	**0.765**	**0.757**	**0.738**	**0.705**	+0.060	**0.875**
Random	0.290	0.620	0.601	0.565	0.499	+0.121	0.803
White-box	0.490	0.970	0.968	0.961	0.944	+0.027	0.937

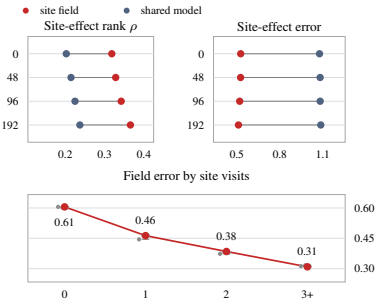


Figure 6: **Site effects.** Top: models at four budgets. Below: error by visits; red is $B = 192$, grey earlier budgets.

Balanced gains and better site estimates. Table 9 shows that MERLE's advantage extends to the most balanced demand quartile while keeping the smallest tilt among the learned policies. Its best balanced point also leads the other deployable policies, suggesting that the overall gain does not come from concentrating service on already favored demand. In Figure 6, the site field ranks true effects more faithfully and has about half the error of the shared surrogate, despite using the same observations. Across the campaign, the number of measured positions grows from 219 to 271. Across budgets, sites visited at least three times have roughly half the error of unvisited sites, with little variation within visit-count groups as budget grows, consistent with gains from repeat visits.

Table 10: **Removing credited positions.** Mean paired dHV after deleting k top, bottom, or candidate-matched random positions from the same campaign.

k	top	bottom	random
5	−0.0494	−0.0096	−0.0010
10	−0.0498	−0.0047	−0.0042
20	−0.0435	−0.0012	−0.0041
40	−0.0384	+0.0001	−0.0053

The report names consequential positions. Removing its top five excludes just 82 of roughly 5,000 candidates and lowers dHV by 0.049, compared with 0.001 for matched random positions. Matching on excluded candidates controls for the amount of library removed. The paired difference is −0.048 (95% CI [−0.073, −0.027]), with intervals excluding zero at all four removal sizes. The effect saturates quickly, concentrating sensitivity in a few positions the report identifies. Removing bottom-ranked positions has little effect, as the attribution predicts.

5.5 Biological relevance of the site readout

The named sites recur across biological measurements. Four matched assay pairs include LASV glycoprotein measured in two laboratories and spike under distinct antibody or serum panels. Separate MERLE campaigns revisit 33.1% of the smaller visited-site set, versus 13.3% expected by chance (Table 11). The site ranking persists across new policy seeds on one instance, while different tables agree at chance (Figure 7). Its weak same-assay correlation with the sequence-only teacher indicates that the repeated site choices are not simply copied from that prior. The modest cross-assay rank correlation confines the biological claim to recurring sites, rather than precise weights. Assigning molecular roles still requires functional annotation.

Appendix E gives a qualitative reading of the site report and the supporting experiments across datasets. It further tests the report's faithfulness, robustness, biological consistency, and usefulness for guiding a subsequent campaign. The deletion tests in Section E.3 and cross-assay controls in Section E.4 most directly check this reading.

Table 11: **Cross-assay site overlap.**

Measure	Mean	95% CI
Pairs	4	5 seeds each
Shared sites	0.331	[0.287, 0.376]
Chance	0.133	[0.108, 0.159]
Lift	3.02×	[2.30, 3.84]
Assay ρ	+0.061	[+0.032, +0.090]
Teacher ρ	+0.042	[+0.030, +0.052]

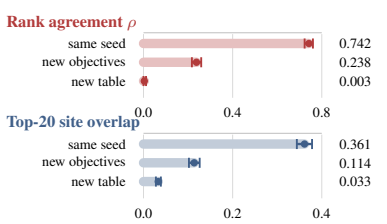


Figure 7: **Site stability.** Rank agreement and top-20 overlap across task redraws; whiskers mark 95% CIs.

AI USE STATEMENT

We acknowledge the use of large language models (LLMs), including ChatGPT and Claude, for limited research assistance: (i) refining academic writing based on author-provided drafts, ideas, and instructions; (ii) assisting with the literature survey; and (iii) supporting the design of TikZ-based visualizations using results from our actual experiments. We did not use autonomous research agents or fully automated research workflows; all substantive content was conceived, directed, and verified by the authors.

The problem formulation, methodological contributions, and experimental design are entirely the work of the authors. All authors reviewed and verified any LLM-assisted output and take full responsibility for the content of this paper. We provide this statement in the interest of research transparency and to support responsible AI use within the research community.

REPRODUCIBILITY STATEMENT

We provide the code required to run the main results reported in this paper, together with the experiment configurations and implementation details needed to reproduce them. Upon acceptance, we will release the complete codebase and all supporting scripts required to reproduce every figure and table. We do not make the full implementation publicly available at this stage because its substantial size would make distribution unwieldy and could disclose ideas underlying our method before publication. We nevertheless commit to ensuring that all reported results are fully reproducible from the released materials.

REFERENCES

Etowah Adams, Liam Bai, Minji Lee, Yiyang Yu, and Mohammed Alquraishi. From mechanistic interpretability to mechanistic biology: Training, evaluating, and interpreting sparse autoencoders on protein language models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 460–476. PMLR, 2025. URL https://proceedings.mlr.press/v267/adams25a.html.

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 *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 324–334. PMLR, 2020. URL https://proceedings.mlr.press/v119/angermueller20a.html.

Jordan T. Ash, Surbhi Goel, Akshay Krishnamurthy, and Sham M. Kakade. Gone fishing: Neural active learning with Fisher embeddings. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8927–8939. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/4afe044911ed2c247005912512ace23b-Abstract.html.

Syrine Belakaria, Aryan Deshwal, and Janardhan Rao Doppa. Max-value entropy search for multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 32, pp. 7793–7803. Curran Associates, Inc., 2019. ISBN 978-1-7138-0793-3. URL https://papers.neurips.cc/paper_files/paper/2019/hash/82edc5c9e21035674d481640448049f3-Abstract.html.

Surojit Biswas, Grigory Khimulya, Ethan C. Alley, Kevin M. Esvelt, and George M. Church. Low-N protein engineering with data-efficient deep learning. *Nature Methods*, 18(4):389–396, 2021. doi: 10.1038/s41592-021-01100-y. URL https://doi.org/10.1038/s41592-021-01100-y.

Can Chen, Yingxue Zhang, Xue Liu, and Mark Coates. Bidirectional learning for offline model-based biological sequence design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 5351–5366. PMLR, 2023. URL https://proceedings.mlr.press/v202/chen23ao.html.

David A. Cohn, Zoubin Ghahramani, and Michael I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996. doi: 10.1613/jair.295. URL https://doi.org/10.1613/jair.295.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective Bayesian optimization. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9851–9864. Curran Associates, Inc., 2020. URL https://papers.neurips.cc/paper_files/paper/2020/hash/6fec24eac8f18ed793f5eaad3dd7977c-Abstract.html.

Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Parallel Bayesian optimization of multiple noisy objectives with expected hypervolume improvement. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2187–2200. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/11704817e347269b7254e744b5e22dac-Abstract.html.

Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002. doi: 10.1109/4235.996017. URL https://doi.org/10.1109/4235.996017.

Kerr Ding, Michael Chin, Yunlong Zhao, Wei Huang, Binh Khanh Mai, Huanan Wang, Peng Liu, Yang Yang, and Yunan Luo. Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. *Nature Communications*, 15 (1):6392, 2024. doi: 10.1038/s41467-024-50698-y. URL https://doi.org/10.1038/s41467-024-50698-y.

Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3): 586–596, 1981. doi: 10.1214/aos/1176345462. URL https://doi.org/10.1214/aos/1176345462.

Benjamin Fram, Yang Su, Ian Truebridge, Adam J. Riesselman, John B. Ingraham, Alessandro Passera, Eve Napier, Nicole N. Thadani, Samuel Lim, Kristen Roberts, Gurleen Kaur, Michael A. Stiffler, Debora S. Marks, Christopher D. Bahl, Amir R. Khan, Chris Sander, and Nicholas P. Gauthier. Simultaneous enhancement of multiple functional properties using evolution-informed protein design. *Nature Communications*, 15(1):5141, 2024. doi: 10.1038/s41467-024-49119-x. URL https://doi.org/10.1038/s41467-024-49119-x.

Sam Gelman, Bryce Johnson, Chase R. Freschlin, Arnav Sharma, Sameer D'Costa, John Peters, Anthony Gitter, and Philip A. Romero. Biophysics-based protein language models for protein engineering. *Nature Methods*, 22(9):1868–1879, 2025. doi: 10.1038/s41592-025-02776-2. URL https://doi.org/10.1038/s41592-025-02776-2.

Michael P. Hansen and Andrzej Jaszkiewicz. Evaluating the quality of approximations to the non-dominated set. Technical Report IMM-REP-1998-7, Institute of Mathematical Modelling, Technical University of Denmark, 1998. URL https://www.imm.dtu.dk/documents/ftp/tr98/tr07_98.abstract.html.

Brian L. Hie, Varun R. Shanker, Duo Xu, Theodora U. J. Bruun, Payton A. Weidenbacher, Shaogeng Tang, Wesley Wu, John E. Pak, and Peter S. Kim. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 42(2):275–283, 2024. doi: 10.1038/s41587-023-01763-2. URL https://doi.org/10.1038/s41587-023-01763-2.

Lu Hong and Tanja Kortemme. An integrative approach to protein sequence design through multiobjective optimization. *PLOS Computational Biology*, 20(7):e1011953, 2024. doi: 10.1371/journal.pcbi.1011953. URL https://doi.org/10.1371/journal.pcbi.1011953.

Chao Hou, Di Liu, and Yufeng Shen. Unsupervised discovery of functional sequence patterns from protein language model with MotifAE. *Nature Communications*, 2026. doi: 10.1038/s41467-026-77333-2. URL https://doi.org/10.1038/s41467-026-77333-2. Advance online publication.

Andrea Hunklinger and Noelia Ferruz. Towards the explainability of protein language models. *Nature Machine Intelligence*, 8(5):649–662, 2026. doi: 10.1038/s42256-026-01232-w. URL https://doi.org/10.1038/s42256-026-01232-w.

Hisao Ishibuchi, Hiroyuki Masuda, Yuki Tanigaki, and Yusuke Nojima. Modified distance calculation in generational distance and inverted generational distance. In *Evolutionary Multi-Criterion Optimization*, volume 9019 of *Lecture Notes in Computer Science*, pp. 110–125. Springer, 2015. doi: 10.1007/978-3-319-15892-1_8. URL https://doi.org/10.1007/978-3-319-15892-1_8.

Moksh Jain, Emmanuel Bengio, Alex Hernandez-Garcia, Jarrid Rector-Brooks, Bonaventure F. P. Dossou, Chanakya Ajit Ekbote, Jie Fu, Tianyu Zhang, Michael Kilgour, Dinghuai Zhang, Lena Simine, Payel Das, and Yoshua Bengio. Biological sequence design with GFlowNets. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9786–9801. PMLR, 2022. URL https://proceedings.mlr.press/v162/jain22a.html.

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

Kaiyi Jiang, Zhaoqing Yan, Matteo Di Bernardo, Samantha R. Sgrizzi, Lukas Villiger, Alisan Kayabolen, B. J. Kim, Josephine K. Carscadden, Masahiro Hiraizumi, Hiroshi Nishimasu, Jonathan S. Gootenberg, and Omar O. Abudayyeh. Rapid in silico directed evolution by a protein language model with EVOLVEpro. *Science*, 387(6732):eadr6006, 2025. doi: 10.1126/science.adr6006. URL https://doi.org/10.1126/science.adr6006.

Zitai Kong, Yifan Dong, Yixuan Wu, Zhaokang Liang, Jian Wu, and Hongxia Xu. MP2D: Constrained Monte Carlo tree-guided diffusion for multi-objective protein sequence design. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence*, pp. 5540–5548, 2026. doi: 10.24963/ijcai.2026/617. URL https://www.ijcai.org/proceedings/2026/617.

Yanke Li, Tianyu Cui, Tommaso Mansi, Mangal Prakash, and Rui Liao. BioBO: Biology-informed Bayesian optimization for perturbation design. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=CF3kJrAwmV.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://doi.org/10.1126/science.ade2574.

Jiaqi Luo, Kerr Ding, and Yunan Luo. Pareto-optimal sampling for multi-objective protein sequence design. *iScience*, 28(3):112119, 2025. doi: 10.1016/j.isci.2025.112119. URL https://doi.org/10.1016/j.isci.2025.112119.

Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alex Rives. Language models enable zero-shot prediction of the effects of mutations on protein function. In *Advances in Neural Information Processing Systems*, volume 34, pp. 29287–29303. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/hash/f51338d736f95dd42427296047067694-Abstract.html.

Lam Ngo, Huong Ha, Jeffrey Chan, and Hongyu Zhang. MOBO-OSD: Batch multi-objective Bayesian optimization via orthogonal search directions. In *Advances in Neural Information Processing Systems*, volume 38, pp. 17259–17294. Curran Associates, Inc., 2025. doi: 10.52202/085713-0581. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/18e780ccbd60697ea4a0184f1e680660-Abstract-Conference.html.

Zhongren Ni and Benjamin Kedem. Normal orthant probabilities in the equicorrelation case. *Journal of Mathematical Analysis and Applications*, 246(1):280–295, 2000. doi: 10.1006/jmaa.2000.6803. URL https://doi.org/10.1006/jmaa.2000.6803.

Pascal Notin, Mafalda Dias, Jonathan Frazer, Javier Marchena-Hurtado, Aidan N. Gomez, Debora Marks, and Yarin Gal. Tranception: Protein fitness prediction with autoregressive transformers and inference-time retrieval. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 16990–17017. PMLR, 2022. URL https://proceedings.mlr.press/v162/notin22a.html.

Pascal Notin, Aaron Kollasch, Daniel Ritter, Lood van Niekerk, Steffanie Paul, Han Spinner, Nathan Rollins, Ada Shaw, Rose Orenbuch, Ruben Weitzman, Jonathan Frazer, Mafalda Dias, Dinko Franceschi, Yarin Gal, and Debora Marks. ProteinGym: Large-scale benchmarks for protein fitness prediction and design. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 64331–64379. Curran Associates, Inc., 2023. doi: 10.52202/075280-2810. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/cac723e5ff29f65e3fcbb0739ae91bee-Abstract-Datasets_and_Benchmarks.html. NeurIPS Datasets and Benchmarks Track.

Akash Pandey, Wei Chen, and Sinan Keten. Interpretability driven evolutionary approach for the design of biological sequences. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=gaq60U4jvU.

Ji Won Park, Natasa Tagasovska, Michael Maser, Stephen Ra, and Kyunghyun Cho. BOtied: Multi-objective Bayesian optimization with tied multivariate ranks. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 39813–39833. PMLR, 2024. URL https://proceedings.mlr.press/v235/park24k.html.

Zhizhou Ren, Jiahan Li, Fan Ding, Yuan Zhou, Jianzhu Ma, and Jian Peng. Proximal exploration for model-guided protein sequence design. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18520–18536. PMLR, 2022. URL https://proceedings.mlr.press/v162/ren22a.html.

Jatan Shrestha, Santeri Heiskanen, Kari Hepola, Severi Rissanen, Pekka Jääskeläinen, and Joni Pajarinen. Pareto-conditioned diffusion models for offline multi-objective optimization. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=S2Q00li155.

Elana Simon and James Zou. InterPLM: Discovering interpretable features in protein language models via sparse autoencoders. *Nature Methods*, 22(10):2107–2117, 2025. doi: 10.1038/s41592-025-02836-7. URL https://doi.org/10.1038/s41592-025-02836-7.

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

Daniel M. Steinberg, Asiri Wijesinghe, Rafael Oliveira, Piotr Koniusz, Cheng Soon Ong, and Edwin V. Bonilla. Amortized active generation of Pareto sets. In *Advances in Neural Information Processing Systems*, volume 38, pp. 85808–85833. Curran Associates, Inc., 2025. doi: 10.52202/085713-2874. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/7c180af017258d239bac6248d1eb26ac-Abstract-Conference.html.

Haoran Sun, Liang He, Pan Deng, Guoqing Liu, Zhiyu Zhao, Yuliang Jiang, Chuan Cao, Fusong Ju, Lijun Wu, Haiguang Liu, Tao Qin, and Tie-Yan Liu. Accelerating protein engineering with fitness landscape modelling and reinforcement learning. *Nature Machine Intelligence*, 7(9):1446–1460, 2025. doi: 10.1038/s42256-025-01103-w. URL https://doi.org/10.1038/s42256-025-01103-w.

Christopher Tosh, Mauricio Tec, Jessica B. White, Jeffrey F. Quinn, Glorymar Ibanez Sanchez, Paul Calder, Andrew L. Kung, Filemon S. Dela Cruz, and Wesley Tansey. A Bayesian active learning platform for scalable combination drug screens. *Nature Communications*, 16 (1):156, 2025. doi: 10.1038/s41467-024-55287-7. URL https://doi.org/10.1038/s41467-024-55287-7.

Yuhao Wang, Qiang Zhang, Ming Qin, Xiang Zhuang, Xiaotong Li, Zhichen Gong, Zeyuan Wang, Yu Zhao, Jianhua Yao, Keyan Ding, and Huajun Chen. Knowledge-aware reinforced language models for protein directed evolution. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 52260–52273. PMLR, 2024. URL https://proceedings.mlr.press/v235/wang24cq.html.

Junhao Xiong, Ishan Gaur, Maria Lukarska, Hunter Nisonoff, Luke M. Oltrogge, David F. Savage, and Jennifer Listgarten. Property guidance for protein sequence generative models with ProteinGuide. *Nature Biotechnology*, 2026. doi: 10.1038/s41587-026-03207-z. URL https://doi.org/10.1038/s41587-026-03207-z. Advance online publication.

Ke Xue, Rongxi Tan, Xiaobin Huang, and Chao Qian. Offline multi-objective optimization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 55595–55624. PMLR, 2024. URL https://proceedings.mlr.press/v235/xue24b.html.

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. doi: 10.1038/s41467-025-55987-8. URL https://doi.org/10.1038/s41467-025-55987-8.

Yaohong Yang, Sammie Katt, and Samuel Kaski. Multi-objective Bayesian optimization via adaptive ε-constraint decomposition. In *Proceedings of the 43rd International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=tu7eFWXBiN.

Ziyi Zhou, Liang Zhang, Yuanxi Yu, Banghao Wu, Mingchen Li, Liang Hong, and Pan Tan. Enhancing efficiency of protein language models with minimal wet-lab data through few-shot learning. *Nature Communications*, 15(1):5566, 2024. doi: 10.1038/s41467-024-49798-6. URL https://doi.org/10.1038/s41467-024-49798-6.

Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and Viviane Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003. doi: 10.1109/TEVC.2003.810758. URL https://doi.org/10.1109/TEVC.2003.810758.

APPENDIX CONTENTS

A The Landscape of Biological Sequence Design **16**
A.1 A Comparative Map of Settings and Open Gaps 16
A.2 Multi-Objective and Many-Objective Optimization 16
A.3 Sequential Model-Guided Search 17
A.4 Bayesian Optimization and Active Learning for Sequence Discovery 17
A.5 Pretrained Models as Informative Priors 17
A.6 Interpretable Machine Learning Methods 17

B Principles of Extrapolative Pareto Search **18**
B.1 When Distilled Preferences Improve Prediction 18
B.2 Why Local Covariance Matters for Frontier Expansion 19
B.3 Coverage and Many-Objective Acquisition Uncertainty 21
B.4 Decision Relevance of the Guarantees 22

C Algorithmic Realization **24**
C.1 End-to-End Query Cycle 24
C.2 Sequence Feature Preparation 25
C.3 Response Model Fitting 26
C.4 Candidate Scoring and Screening 26
C.5 Coverage Queries and Adaptive Allocation 27
C.6 Computational Complexity 28

D Empirical Evaluation Protocol **29**
D.1 Dataset Panel 29
D.2 Reference Campaign Design 29
D.3 Baseline Implementations and Adaptations 29
D.4 Metrics and Scoring 30
D.5 Aggregation and Uncertainty 31

E Further Empirical Evidence **32**
E.1 Qualitative analysis of the site readout 32
E.2 Performance across datasets 33
E.3 Faithfulness and robustness of the readout 35
E.4 Biological consistency across assays 38
E.5 Decision utility of position explanations 39

F Discussion: Scope, Limitations, and Broader Implications **40**
F.1 Reproducibility Statement 40
F.2 Scope, Limitations, and Future Directions 40
F.3 Ethics, Societal Impact, and Responsible Use 40

A THE LANDSCAPE OF BIOLOGICAL SEQUENCE DESIGN

A.1 A COMPARATIVE MAP OF SETTINGS AND OPEN GAPS

Figure 8 compares MERLE with representative work along two questions: *what is selected* and *how it is selected*. Separating the task from the method exposes the joint contribution.

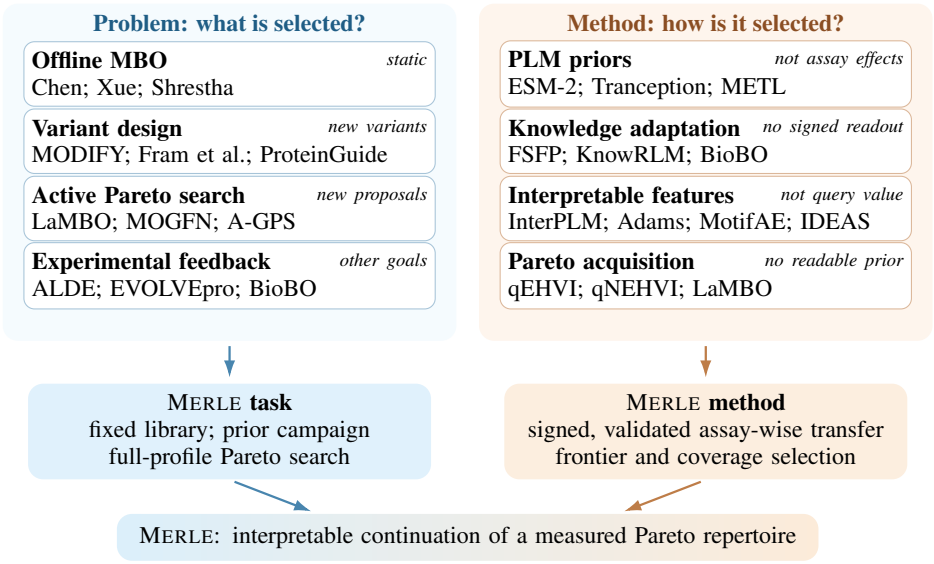


Figure 8: Representative work along the task and modeling axes. The lower boxes identify MERLE's joint contribution; arrows denote conceptual links.

On the *task* axis, LaMBO and A-GPS generate sequences online (Stanton et al., 2022; Steinberg et al., 2025); ALDE iterates assays in a combinatorial mutation space (Yang et al., 2025), whereas offline multi-objective methods take no new measurements (Xue et al., 2024; Shrestha et al., 2026). MERLE instead continues a prior campaign over a fixed pool, using complete response profiles to expand the measured frontier.

On the *method* axis, BioBO uses biological knowledge in perturbation BO (Li et al., 2026), while IDEAS steers mutations using interpretable motifs (Pandey et al., 2026). InterPLM and related analyses expose PLM concepts without valuing the next full-profile query (Simon & Zou, 2025; Adams et al., 2025). MERLE estimates signed, empirically checked transfer for each assay and uses it in the acquisition policy.

A.2 MULTI-OBJECTIVE AND MANY-OBJECTIVE OPTIMIZATION

Multi-objective GFlowNets sample diverse Pareto candidates, including a version embedded in an active-learning loop (Jain et al., 2023). ProteinGuide conditions a generative model on desired properties and can address two properties in tension (Xiong et al., 2026); evolution- informed design has experimentally improved several properties at once (Fram et al., 2024). These studies establish the value of retaining trade-offs, rather than choosing a single weighted score in advance.

The optimization target still matters. MODIFY balances predicted fitness against sequence diversity to build an *initial* enzyme library (Ding et al., 2024), whereas preference-conditioned generation searches across designs as feedback arrives (Steinberg et al., 2025). In MERLE, the library is already available and the budget purchases measurements, not new candidate construction. Because the output is an experimentally observed frontier across many conditions, predicted diversity is not a substitute for measured trade-off coverage.

A.3 Sequential Model-Guided Search

Low-N protein engineering demonstrates that useful models can be built from few measurements (Biswas et al., 2021). EVOLVEpro combines PLM embeddings with iterative few-shot feedback (Jiang et al., 2025), and active learning-assisted directed evolution studies how encodings, surrogates, and uncertainty affect successive laboratory rounds (Yang et al., 2025). These works motivate carrying experimental evidence forward rather than beginning each search from scratch.

Their decision spaces nevertheless differ from the continuation task here. Model-guided search may synthesize new sequences or navigate a combinatorial mutation space; μProtein, for example, uses learned fitness and reinforcement learning to find multi-mutant variants (Sun et al., 2025). Offline biological optimization instead proposes designs from a static labeled collection (Chen et al., 2023). MERLE starts with measurements from a previous campaign and asks which *remaining library member* to assay so that all condition responses inform the next decision.

A.4 Bayesian Optimization and Active Learning for Sequence Discovery

Expected hypervolume improvement values a candidate by the dominated volume it may add, directly reflecting Pareto discovery (Daulton et al., 2020). Its noisy extension integrates uncertainty about the current front (Daulton et al., 2021), while LaMBO combines multi-objective BO with a denoising autoencoder for sequence design (Stanton et al., 2022). These results motivate MERLE's use of frontier gain, but they do not dictate its fixed-library acquisition: the method prices each available query against response requirements not yet met by the measured repertoire.

Biology-informed BO shows that pretrained information can also influence perturbation selection (Li et al., 2026); active learning-assisted evolution emphasizes the practical role of calibrated uncertainty (Yang et al., 2025). MERLE joins these concerns by learning how a mutation preference transfers to *each* assay before using it to select a full-profile measurement. The distinction is between valuing uncertain outcomes and assuming that an evolutionary prior has the same direction or reliability for every objective.

A.5 Pretrained Models as Informative Priors

General PLMs can prioritize useful antibody mutations without target-specific training (Hie et al., 2024), and Tranception improves variant-effect prediction by combining autoregressive modeling with homolog retrieval (Notin et al., 2022). METL brings biophysical simulation into pretraining and improves some low-data extrapolation tasks (Gelman et al., 2025). These results support transfer, but different pretraining signals need not align with a particular experimental response.

Assay evidence can substantially alter that picture. Few-shot fine-tuning improves PLM fitness prediction with limited labels (Zhou et al., 2024), while ProteinGym documents substantial variation in zero-shot performance across assays (Notin et al., 2023). Rather than fine-tuning the entire ESM-2 model (Lin et al., 2023), MERLE distills its mutation preferences into a compact covariate and learns a signed, held-out-validated contribution for each condition. This makes the source and direction of transfer visible during search.

A.6 Interpretable Machine Learning Methods

Sparse autoencoders recover biologically meaningful PLM features (Adams et al., 2025; Simon & Zou, 2025); MotifAE encourages coherent feature activations to identify functional sequence patterns (Hou et al., 2026). These works show that a pretrained representation can be made legible. They principally explain or probe the model, however, rather than calibrating the value of a particular experimental query across conflicting response conditions.

IDEAS takes a step toward actionable interpretation by using motifs to guide evolutionary mutations (Pandey et al., 2026), while KnowRLM injects amino-acid knowledge into an actively updated generation policy (Wang et al., 2024). MERLE makes a different link between interpretation and action: a readable, mutation-level PLM preference is assigned an empirically supported effect for each assay, then participates directly in frontier-oriented selection. Interpretability here is not only a post hoc description of chosen sequences; it is a constrained component of the acquisition policy.

B PRINCIPLES OF EXTRAPOLATIVE PARETO SEARCH

The three results follow a measurement decision in MERLE: determine whether a pretrained preference improves assay prediction, value the resulting joint response for frontier expansion, and identify assays that inform the remaining library. Each result isolates one source of information within this full-profile query setting.

B.1 WHEN DISTILLED PREFERENCES IMPROVE PREDICTION

Fix a condition c and the two held-out folds used in Equation 6. Index their union by $i = 1, \dots, n$ and write $h_i = \widehat{\beta}_c^{(-i)} \ell_i$ and $y_i = z_{ic} - \widehat{\theta}_{p(i),c}^{(-i)}$. Let I_1, I_2 be the folds and $S_j = \sum_{i \in I_j} h_i^2$, $S = S_1 + S_2 > 0$. Suppose that, conditional on the data used to fit the opposite fold, each h_i and the conditional mean m_i of y_i are fixed, while the errors $e_i = y_i - m_i$ within the held-out fold are independent, centered, and σ-sub-Gaussian. The two folds need not be independent of each other. Define the teacher's *conditional alignment* with the residual response by

$$\alpha_c = \frac{\sum_{i=1}^n h_i m_i}{S}. \tag{13}$$

For any $s \in [0, 1]$, let $R_c(s)$ be the mean squared prediction error of sh_i on independent replicate responses with these same conditional means m_i. Thus $R_c(0)$ is the risk of the position-only prediction on this evaluation design.

Proposition B.1 (Benefit of condition-specific distillation) *Assume $0 \le \alpha_c \le 1$, and let s_c be the clipped cross-fitted slope in Equation 6. For every $\delta \in (0, 1)$, with probability at least $1 - \delta$ over the held-out errors,*

$$R_c(0) - R_c(s_c) \;\ge\; \frac{S\alpha_c^2}{n} - \frac{4\sigma^2}{n} \log \frac{4}{\delta}. \tag{14}$$

In particular, a positive gain is certified when $S\alpha_c^2 > 4\sigma^2 \log(4/\delta)$. The condition and the certificate are separate for each assay.

Proof. Equation 6 is the projection onto $[0, 1]$ of the unconstrained least-squares slope. Set $T_j = \sum_{i \in I_j} h_i e_i$. Since $y_i = m_i + e_i$, its numerator separates into signal and noise:

$$\begin{aligned} s_c &= \Pi_{[0,1]}\left(\frac{\sum_i h_i (m_i + e_i)}{S} \right) \\ &= \Pi_{[0,1]}\left(\alpha_c + \frac{T_1 + T_2}{S} \right). \end{aligned} \tag{15}$$

Conditional on the training data outside fold I_j, the weighted sum T_j has moment-generating function bounded by $\mathbb{E}[\exp(\lambda T_j) \mid I_j^{\text{train}}] \le \exp(\lambda^2 \sigma^2 S_j / 2)$. Applying the resulting sub-Gaussian tail bound to both signs of T_j gives

$$\begin{aligned} \Pr\left(|T_j| > \sigma \sqrt{2 S_j \log(4/\delta)} \mid I_j^{\text{train}} \right) &\le 2 \exp[-\log(4/\delta)] \\ &= \delta/2. \end{aligned} \tag{16}$$

If $S_j = 0$, then $T_j = 0$ and the same bound is immediate. Averaging over the training fold and taking a union bound over $j = 1, 2$ needs no independence between folds. On the resulting event, Cauchy–Schwarz gives

$$\begin{aligned} |T_1 + T_2| &\le \sigma \sqrt{2 \log(4/\delta)} (\sqrt{S_1} + \sqrt{S_2}) \\ &\le 2\sigma \sqrt{S \log(4/\delta)}, \end{aligned} \qquad \text{with probability at least } 1 - \delta. \tag{17}$$

Projection onto an interval cannot increase distance to $\alpha_c \in [0, 1]$. Hence Equations 15 and 17 imply $|s_c - \alpha_c|^2 \le |T_1 + T_2|^2 / S^2 \le 4\sigma^2 \log(4/\delta)/S$. Finally, replicate noise contributes the same additive constant to every $R_c(s)$. Expanding the squares and using $\sum_i h_i m_i = S\alpha_c$ makes the gain

explicit:

$$\begin{aligned} R_c(0) - R_c(s_c) &= \frac{1}{n}\sum_i \big[m_i^2 - (m_i - s_c h_i)^2\big] \\ &= \frac{2s_c}{n}\sum_i h_i m_i - \frac{s_c^2}{n}\sum_i h_i^2 \\ &= \frac{S}{n}\big[\alpha_c^2 - (s_c - \alpha_c)^2\big]. \end{aligned} \tag{18}$$

Substitution of the concentration bound proves Equation 14. □

Interpretation for MERLE. The quantity S/n is the mean squared size of the teacher prediction *after* subtracting the position effect; α_c measures how much of that prediction remains aligned with the true assay residual. Thus $S\alpha_c^2/n$ is the largest risk reduction available along this one teacher direction. The other term in Equation 14 is the price of estimating its strength from noisy held-out assays. The two-fold construction matters because each h_i is fitted without using its own response, even though the folds are reused in opposite roles. This explains why Equation 6 learns a separate retained fraction for each condition rather than treating PLM preference as a universal fitness score. Indeed, using the teacher at full strength gives $R_c(0) - R_c(1) = S(2\alpha_c - 1)/n$: a weakly aligned teacher can increase error. The bound concerns the cross-fitted evaluation design; the full-data coefficient requires the stability analysis below.

Effect of the final refit. The implemented coefficient uses $\gamma_c = s_c\widehat{\beta}_c$, with $\widehat{\beta}_c$ refitted on all observations. The preceding bound applies directly to the cross-fitted predictions h_i. Its effect on the final refit can also be made explicit. Let $R_c^{\mathrm{refit}}(s)$ denote the replicate risk of $s\widehat{\beta}_c\ell_i$ on the same evaluation design. Set $\Delta_i = \widehat{\beta}_c\ell_i - h_i$, $D = (n^{-1}\sum_i \Delta_i^2)^{1/2}$, and $L = (n^{-1}\sum_i (m_i - s_c h_i)^2)^{1/2}$. The replicate noise cancels, so expansion followed by Cauchy–Schwarz gives

$$\begin{aligned} R_c^{\mathrm{refit}}(s_c) - R_c(s_c) &= -\frac{2s_c}{n}\sum_i (m_i - s_c h_i)\Delta_i + \frac{s_c^2}{n}\sum_i \Delta_i^2 \\ &\le 2s_c\left(\frac{1}{n}\sum_i (m_i - s_c h_i)^2\right)^{1/2}\left(\frac{1}{n}\sum_i \Delta_i^2\right)^{1/2} + s_c^2 D^2 \\ &= 2s_c L D + s_c^2 D^2. \end{aligned} \tag{19}$$

Combining this inequality with Proposition B.1 yields

$$R_c(0) - R_c^{\mathrm{refit}}(s_c) \ge \frac{S\alpha_c^2 - 4\sigma^2\log(4/\delta)}{n} - 2s_c L D - s_c^2 D^2. \tag{20}$$

Thus the certificate carries to the final coefficient when its full-data refit remains close to the cross-fitted predictions. The term D is a separate stability requirement, not a consequence of cross-fitting alone.

Takeaway. MERLE's assay-specific shrinkage can improve marginal predictions when teacher–assay alignment outweighs estimation noise. Joint responses then determine whether those predictions support frontier expansion.

B.2 WHY LOCAL COVARIANCE MATTERS FOR FRONTIER EXPANSION

Section B.1 bounds assay-wise prediction, but frontier expansion requires favorable responses *together*. The joint Gaussian tail in Equation 10 uses the local covariance from Equation 7. Conditional on a sampled position effect, let a candidate have mean μ_c and residual standard deviation $\sigma_c > 0$ in each condition. Compare the equicorrelated residual law $R_\rho = (1-\rho)I_C + \rho\mathbf{1}\mathbf{1}^\top$, $0 < \rho < 1$, with its diagonal approximation; both have the same marginal predictions. For requirement g_j, put $h_{jc} = (g_{jc} - \mu_c)/\sigma_c$ and let $Q_\rho(h) = \Pr(X_c \ge h_c\ \forall c)$ for $X \sim \mathcal{N}(0, R_\rho)$. The corresponding scores on the same unmet cloud are

$$A_t^{(\rho)}(v) = \frac{1}{J}\sum_{j\in\mathcal{U}_t} Q_\rho(h_j), \qquad A_t^{(0)}(v) = \frac{1}{J}\sum_{j\in\mathcal{U}_t}\prod_{c=1}^{C}\bar{\Phi}(h_{jc}), \tag{21}$$

where $\bar{\Phi} = 1 - \Phi$. Gaussian orthant probabilities have been studied independently of sequence design (Ni & Kedem, 2000); the following comparison specifies their consequence for MERLE's acquisition.

Proposition B.2 (Covariance-aware demand separation) *For $C \geq 2$ and any finite unmet thresholds,*

$$A_t^{(\rho)}(v) - A_t^{(0)}(v) = \frac{1}{J} \sum_{j \in \mathcal{U}_t} \left[Q_\rho(h_j) - \prod_{c=1}^{C} \bar{\Phi}(h_{jc}) \right] \geq 0, \tag{22}$$

strictly when $\mathcal{U}_t \neq \varnothing$. For a balanced requirement $h_j = 0$, write $q_C(\rho) = Q_\rho(0)$. Then

$$q_C(0) = 2^{-C}, \qquad q_C\left(\tfrac{1}{2}\right) = \frac{1}{C+1}, \qquad \lim_{C \to \infty} \frac{-\log q_C(\rho)}{\log C} = \frac{1-\rho}{\rho}. \tag{23}$$

Thus the diagonal model can assign exponentially less probability to simultaneous success despite matching every marginal prediction.

Proof. Take independent standard normals $Z_0, Z_1, \ldots, Z_C$ and represent $X_c = \sqrt{\rho}\, Z_0 + \sqrt{1-\rho}\, Z_c$. Conditional on $Z_0 = z$, the coordinates are independent, and the probability of satisfying condition c is

$$p_c(z) = \Phi\left(\frac{\sqrt{\rho}\, z - h_c}{\sqrt{1-\rho}} \right), \qquad Q_\rho(h) = \mathbb{E}\left[\prod_{c=1}^{C} p_c(Z_0) \right], \quad \mathbb{E}[p_c(Z_0)] = \bar{\Phi}(h_c). \tag{24}$$

Each p_c is positive and strictly increasing. For any two increasing functions F, G and an independent copy Z_0' of Z_0,

$$2\,\mathrm{Cov}(F(Z_0), G(Z_0)) = \mathbb{E}[(F(Z_0) - F(Z_0'))(G(Z_0) - G(Z_0'))] \geq 0. \tag{25}$$

Since products of positive increasing functions remain increasing, repeated application of Equation 25 gives

$$\begin{aligned} Q_\rho(h) &= \mathbb{E}\left[\textstyle\prod_{c=1}^{C} p_c(Z_0) \right] \\ &\geq \mathbb{E}\left[\textstyle\prod_{c=1}^{C-1} p_c(Z_0) \right] \mathbb{E}[p_C(Z_0)] \\ &\geq \textstyle\prod_{c=1}^{C} \mathbb{E}[p_c(Z_0)] = \prod_{c=1}^{C} \bar{\Phi}(h_c). \end{aligned} \tag{26}$$

The first inequality is strict for finite thresholds and $\rho > 0$. Summing over $\mathcal{U}_t$ proves Equation 22; the result also holds after averaging over sampled position effects.

For balanced thresholds, set $a = \sqrt{\rho/(1-\rho)}$ in Equation 24. Then

$$q_C(\rho) = \mathbb{E}[\Phi(aZ_0)^C]. \tag{27}$$

At $\rho = 0$, independence gives 2^{-C}. At $\rho = 1/2$, $a = 1$ and $\Phi(Z_0)$ is uniform on $[0,1]$, giving $q_C(1/2) = \int_0^1 u^C\, du = 1/(C+1)$.

It remains to establish the general exponent rather than merely a lower bound. Let $M_C = \max_{1 \leq c \leq C}(-Z_c)$, independent of Z_0. The event $X_c \geq 0$ for every c is $Z_0 \geq M_C/a$. The event $\{M_C \leq t, Z_0 \geq t/a\}$ implies success; conversely, success with $M_C > t$ requires $Z_0 > t/a$. Since $\Pr(M_C \leq t) = \Phi(t)^C$ and M_C is independent of Z_0, these inclusions give, for any $t > 0$,

$$\Phi(t)^C \bar{\Phi}(t/a) \leq q_C(\rho) \leq \Phi(t)^C + \bar{\Phi}(t/a). \tag{28}$$

Fix $\varepsilon \in (0,1)$ and use $t_\pm = (1 \pm \varepsilon)\sqrt{2 \log C}$. To make the tail step explicit, integration of the standard normal density φ gives $\varphi(x + 1/x)/x \leq \bar{\Phi}(x) \leq \varphi(x)/x$ for $x > 0$. Consequently $\bar{\Phi}(b\sqrt{2\log C}) = C^{-b^2 + o(1)}$ for each fixed $b > 0$. Thus $\Phi(t_+)^C \to 1$, whereas $\Phi(t_-)^C \leq \exp\{-C^{1-(1-\varepsilon)^2 + o(1)}\}$. Using t_+ for the lower bound and t_- for the upper bound in Equation 28, then taking negative logarithms, gives the two steps

$$\begin{aligned} C^{-(1+\varepsilon)^2/a^2 + o(1)} \leq q_C(\rho) \leq C^{-(1-\varepsilon)^2/a^2 + o(1)}, \\ \frac{(1-\varepsilon)^2}{a^2} \leq \liminf_{C \to \infty} \frac{-\log q_C(\rho)}{\log C} \leq \limsup_{C \to \infty} \frac{-\log q_C(\rho)}{\log C} \leq \frac{(1+\varepsilon)^2}{a^2}. \end{aligned} \tag{29}$$

Letting $\varepsilon \downarrow 0$ and using $a^{-2} = (1-\rho)/\rho$ proves Equation 23. □

Corollary B.3 (Covariance changes MERLE's candidate ranking) *In* MERLE*, suppose a single cloud requirement remains unmet and two shortlisted variants have the same conditional mean and marginal residual variances, with threshold $h = 0$. If their residual correlations are $R_{1/2}$ and I_C, respectively, Equation 10 assigns them scores $1/[J(C+1)]$ and $1/(J2^C)$. For $C \geq 2$, the covariance-aware acquisition ranks the former higher, whereas a product of marginal survival probabilities ties them.*

Proof. The scores follow by substituting the two balanced orthant probabilities from Equation 23 into Equation 21. Their inequality is strict for $C \geq 2$; equal marginals give the diagonal tie. □

Role of local covariance. The corollary isolates a ranking change from the joint response geometry, not from a change in any marginal prediction. The local loadings w_{pk} in Equation 7 permit a position to expand the co-improvement direction, whose equicorrelated eigenvalue is $1 + (C-1)\rho$; Equation 10 then values its joint tail against unmet requirements. Other dependence patterns, including trade-offs, yield rankings determined by their own thresholds and local covariances. Observed frontier gains in Equation 12 subsequently determine the budget assigned to this proposal mode.

Takeaway. Accurate assay-wise means do not determine joint success: cross-condition dependence can change MERLE's candidate rankings. Coverage then asks whether a proposed measurement can inform other unmeasured variants across the remaining library.

B.3 COVERAGE AND MANY-OBJECTIVE ACQUISITION UNCERTAINTY

Sections B.1–B.2 concern the mean and joint tail of model-based frontier proposals. Coverage addresses a different question: which measurement informs other variants in the library? We analyze this under the auxiliary Bayesian linear model in Equation 11, holding the current front fixed. For each condition c, let $\widetilde{f}_c(u) = \phi(u)^\top b_c$; the b_c are conditionally independent with posterior covariance S_t. A full-profile assay has noise variance σ^2_{obs} per condition. Write $\mathcal{P}_t = \mathcal{X} \setminus V_{t-1}$.

Proposition B.4 (Coverage and acquisition uncertainty) *Let R_t be the integrated posterior squared-error risk for predicting the auxiliary response profiles over $\mathcal{P}_t$. After a hypothetical query at v, evaluate risk on the* same *pool $\mathcal{P}_t$. Then*

$$R_t = \sum_{u \in \mathcal{P}_t} \sum_{c=1}^{C} \mathrm{Var}_t\big(\widetilde{f}_c(u)\big), \qquad R_t - R_{t|v} = C\, C_t(v). \tag{30}$$

Let $h_t(y)$ be the hypervolume gain from adding $y \in [0,1]^C$ to the current front, and define $H_u = h_t\big(\Pi_{[0,1]^C} \widetilde{f}(u)\big)$. If $\widehat{H}_u = \mathbb{E}_t[H_u]$ and $\widehat{u} \in \arg\max_{u \in \mathcal{P}_t} \widehat{H}_u$, then

$$\sum_{u \in \mathcal{P}_t} \mathbb{E}_t\Big[(H_u - \widehat{H}_u)^2\Big] \leq R_t, \qquad \mathbb{E}_t\Big[\max_{u \in \mathcal{P}_t} H_u - H_{\widehat{u}}\Big] \leq 2\sqrt{R_t}. \tag{31}$$

Thus maximizing $C_t(v)$ maximizes the one-step reduction in both profile risk and the upper certificate $2\sqrt{R_t}$ for auxiliary acquisition-ranking regret.

Proof. Gaussian conditioning supplies a rank-one covariance update. Inserting it into the integrated profile risk yields the full reduction:

$$\begin{aligned} S_t - S_{t|v} &= \frac{S_t \phi(v) \phi(v)^\top S_t}{\sigma^2_{\text{obs}} + \phi(v)^\top S_t \phi(v)}, \\ R_t - R_{t|v} &= C \sum_{u \in \mathcal{P}_t} \phi(u)^\top (S_t - S_{t|v}) \phi(u) \\ &= C \sum_{u \in \mathcal{P}_t} \frac{\big(\phi(u)^\top S_t \phi(v)\big)^2}{\sigma^2_{\text{obs}} + \phi(v)^\top S_t \phi(v)} = C\, C_t(v). \end{aligned} \tag{32}$$

The covariance update does not depend on the observed response, so Equation 32 is a preposterior identity.

For the second claim, changing only coordinate c of a response profile changes its newly dominated region within a slab of volume at most $|y_c - y'_c|$. Clipping preserves this Lipschitz bound. Let $H_u^{(c)}$ replace condition c of $\widetilde{f}(u)$ by an independent posterior copy; then $|H_u - H_u^{(c)}| \le |\widetilde{f}_c(u) - \widetilde{f}'_c(u)|$. The objectives are conditionally independent, so the Efron–Stein inequality (Efron & Stein, 1981), with $s_t^2(u) = \phi(u)^\top S_t \phi(u)$, gives

$$\mathrm{Var}_t(H_u) \le \frac{1}{2}\sum_{c=1}^{C} \mathbb{E}_t[(H_u - H_u^{(c)})^2] \le \frac{1}{2}\sum_{c=1}^{C} 2s_t^2(u) = C s_t^2(u). \tag{33}$$

Independent copies differ with variance $2s_t^2(u)$, proving the last inequality. Since $\widehat{H}_u = \mathbb{E}_t[H_u]$, summation over u proves the first part of Equation 31. Let u^* maximize H_u. The choice of $\widehat{u}$ ensures $\widehat{H}_{u^*} - \widehat{H}_{\widehat{u}} \le 0$, so decomposing $H_{u^*} - H_{\widehat{u}}$ around these two posterior means bounds it by $2\max_{u\in\mathcal{P}_t} |H_u - \widehat{H}_u|$. The maximum is bounded by the Euclidean norm of the deviations. Taking expectations and applying Jensen's inequality now gives

$$\begin{aligned}
\mathbb{E}_t[H_{u^*} - H_{\widehat{u}}] &\le 2\,\mathbb{E}_t\left[\max_{u\in\mathcal{P}_t} |H_u - \widehat{H}_u|\right] \\
&\le 2\,\mathbb{E}_t\left[\Big(\sum_{u\in\mathcal{P}_t} (H_u - \widehat{H}_u)^2\Big)^{1/2}\right] \\
&\le 2\left(\sum_{u\in\mathcal{P}_t} \mathbb{E}_t[(H_u - \widehat{H}_u)^2]\right)^{1/2} \le 2\sqrt{R_t}.
\end{aligned} \tag{34}$$

The argument also holds after querying v with $S_{t|v}$ and the front fixed. When $R_t > 0$, Equation 30 makes the decrease in the ranking certificate explicit:

$$2\sqrt{R_t} - 2\sqrt{R_{t|v}} = \frac{2C\,C_t(v)}{\sqrt{R_t} + \sqrt{R_t - C\,C_t(v)}}. \tag{35}$$

This decrease is increasing in $C_t(v)$; if $R_t = 0$, all candidate coverage scores vanish. □

Corollary B.5 (Library-sensitive coverage in MERLE) *Consider two groups of sizes m_1, m_2 with intercept-augmented features $(1,1)^\top, (-1,1)^\top$ and $S_t = s^2 I_2$. A query v_i has cross-covariance $2s^2$ within group i and zero outside. Thus Equations 11 and 30 give*

$$C_t(v_i) = \frac{4m_i s^4}{\sigma_{\mathrm{obs}}^2 + 2s^2}, \qquad R_t - R_{t|v_i} = \frac{4Cm_i s^4}{\sigma_{\mathrm{obs}}^2 + 2s^2}. \tag{36}$$

If both groups occur in MERLE*'s sampled candidate set and $m_1 > m_2$, the coverage source selects from the first group, while pointwise uncertainty assigns all candidates the same variance $2s^2$.*

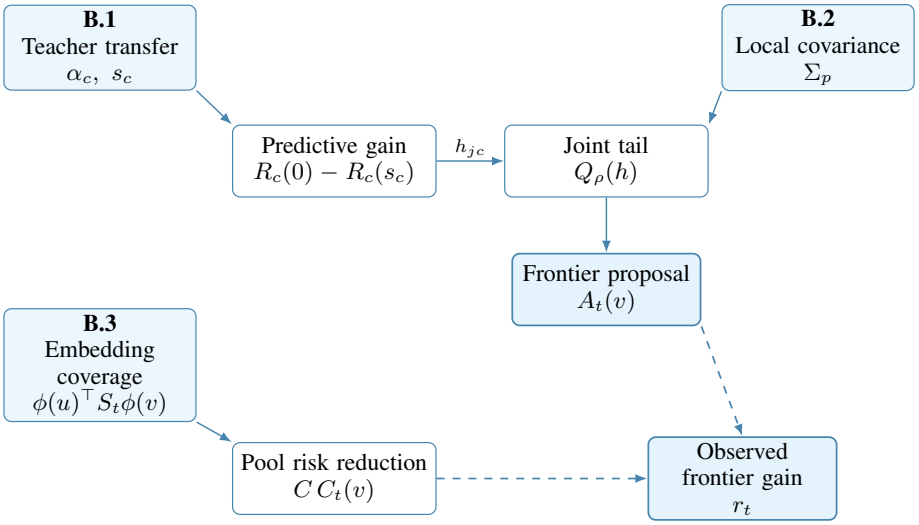
Takeaway. Even a well-valued frontier proposal need not inform other variants. Coverage favors assays that do so across all C conditions, while realized frontier gains determine this source's share of the remaining measurement budget.

B.4 Decision Relevance of the Guarantees

The guarantees form a dependency structure rather than three competing scores. The teacher in Proposition B.1 affects the accuracy of assay-wise means and therefore the thresholds at which frontier requirements are evaluated. The local covariance in Proposition B.2 determines whether responses clear those thresholds together. Both shape the frontier proposal score A_t. Coverage takes a separate path: Proposition B.4 quantifies what a query teaches about unmeasured variants, irrespective of its immediate frontier value. The graph below distinguishes these logical links from the empirical step needed to assess a query.

Where the many-objective effects enter. At balanced thresholds and $\rho = 1/2$, Equation 23 implies a joint-to-diagonal probability ratio of $2^C/(C+1)$. Thus diagonalization can increasingly underprice co-improvement as C grows, even with accurate marginal means. In contrast, the exact $C\,C_t(v)$ reduction in Equation 30 measures information shared across variants, not simultaneous success of one variant. The different scalings belong to different branches of the graph; adding their raw scores would have no probabilistic interpretation.

Closing the loop with measurements. The solid arrows below are conditional theoretical links. Neither prediction-risk reduction nor posterior variance reduction alone proves an observed Pareto gain; the dashed arrows therefore end at the measured reward r_t in Equation 12. This also suggests where to look when search stalls: weak held-out teacher gain can distort the thresholds, diagonalized covariance can misrank joint opportunities, and narrow feature overlap can limit the value of coverage queries. Their effect on search is assessed on the same frontier-coverage scale only after the assay is observed.



A quantitative interaction. The upper branches meet inside the joint tail, not by adding two scores. In the conditional Gaussian comparison of Section B.2, write $\mu_c = \widetilde{\theta}_{p,c} + \gamma_c \ell_v$, so $h_{jc} = (g_{jc} - \mu_c)/\sigma_c$ for fixed $\sigma_c > 0$. Differentiating the survival integral with respect to its lower limit and then applying the chain rule gives

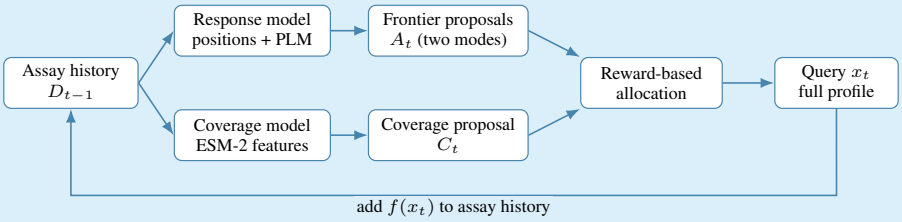
$$\begin{aligned} \frac{\partial A_t^{(\rho)}(v)}{\partial \gamma_c} &= -\frac{\ell_v}{J\sigma_c} \sum_{j \in \mathcal{U}_t} \frac{\partial Q_\rho(h_j)}{\partial h_{jc}} \\ &= \frac{\ell_v}{J\sigma_c} \sum_{j \in \mathcal{U}_t} \varphi(h_{jc}) \Pr(X_{-c} \geq h_{j,-c} \mid X_c = h_{jc})\,. \end{aligned} \tag{37}$$

Here φ is the standard normal density and X_{-c} denotes the other conditions. The density emphasizes requirements near the candidate's predicted boundary, while the conditional probability asks whether the other conditions can succeed there. Thus transfer may improve a marginal prediction yet have little frontier effect when joint success remains unlikely; this is the coupling represented by the graph's upper path.

C ALGORITHMIC REALIZATION

C.1 END-TO-END QUERY CYCLE

Takeaway. Measured frontier gains select the next proposal source.



The campaign starts from the previously measured repertoire V_0 and a fixed, unmeasured library. Sequence-side quantities are prepared once, before any additional assay is purchased (Section C.2). The response field, unmet-requirement cloud, and allocation statistics then evolve with the measurements. Algorithm 1 makes explicit that a proposal source determines *how* candidates are generated, whereas every source is credited using the same realized frontier gain.

Algorithm 1 Sequential query cycle of MERLE

Given: Library $\mathcal{X}$, prior assays V_0, budget B, prepared features ℓ_v and $\phi(v)$
Return: Ordered queries $(x_1, \ldots, x_B)$ and observed frontier

Initialize: measured repertoire and proposal sources

Initialize the position field from V_0; tune the coverage model on V_0 if embeddings are available.
Draw the fixed requirement cloud; let $\mathcal{U}_1$ contain requirements not met by V_0.
Index remaining substitutions by position; activate the available proposal sources.

Repeat: select, assay, and learn

for $t = 1, \ldots, B$ **do**
  if $t = 1$ or the response-model refresh interval has elapsed **then**
    Recalibrate the variance split if $t > 1$; refresh transfer, position effects, and covariance.
  end if
  Draw source a_t using the reward-based allocator.
  if $a_t =$ coverage **then**
    Fit the embedding model on V_{t-1}; maximize Eq. 11 over fresh sampled candidates.
  else
    Fall back to position mode if teacher transfer is unusable.
    Select x_t by Algorithm 4.
  end if
  Measure the complete profile $f(x_t)$ and form z_{x_t}.
  $\mathcal{U}_{t+1} \leftarrow \mathcal{U}_t \setminus \{j : g_j \le z_{x_t}\}$; compute r_t by Eq. 12.
  Credit r_t to the source used; add x_t to the response field and remove it from the available library.
end for
return $(x_1, \ldots, x_B)$ and the frontier of V_B.

The field is refreshed at the first round and periodically thereafter; between refreshes, each complete assay updates the count and second moment at its position without rescanning earlier profiles:

$$n_{p,t} = n_{p,t-1} + 1, \qquad Q_{p,t} = Q_{p,t-1} + z_{x_t} z_{x_t}^\top, \quad p = p(x_t). \tag{38}$$

The fixed requirement cloud likewise needs only its unmet subset updated. Until transfer is validated, a teacher proposal falls back to the position source and is credited there.

Each available source receives an initial trial. Subsequent source choices sample reward estimates from its observed, normalized cloud gains, with the variance pooled across sources. Only the source used for a query receives credit. The planned budget stops the loop but does not enter the acquisition or allocation scores: given the same initial state and random stream, the first k queries are unchanged if the campaign is allowed to continue.

C.2 Sequence Feature Preparation

Two frozen-PLM summaries play different roles. The teacher score ℓ_v retains a single, interpretable ordering among substitutions at a position; the embedding map $\phi(v)$ supports the separate coverage model. Preparing both before the campaign prevents a newly measured assay from changing the representation itself. Only the downstream response models are updated.

Algorithm 2 Preparing the two sequence-side channels

Given: Fixed variant identifiers and wild-type sequence; frozen ESM-2 outputs available for the library
Return: Teacher scores ℓ_v and, when available, features $\phi(v)$

Teacher: positionwise mutation preference
Parse each single substitution into position $p(v)$ and mutant residue $a(v)$.
for each distinct scored position p **do**
  Read the cached masked-token probabilities at p.
  for each scored substitution v at p **do**
    Compute Δ_v as the mutant-to-wild-type log ratio in Eq. 4.
  end for
end for
Standardize Δ_v across scored library variants to obtain ℓ_v; set unscored entries to zero.
Coverage: sequence-level embedding geometry
if row-aligned mean-pooled ESM-2 embeddings are available **then**
  Center the full library's embedding matrix and retain its leading principal components.
  Pin each component's sign, scale it to unit empirical SD, and append an intercept to form $\phi(v)$.
else
  Disable the embedding-coverage source.
end if
return ℓ_v and the available $\phi(v)$.

The masked score uses one teacher evaluation per position and is cached in variant order. Its standardization excludes missing scores, which are set to zero only afterward. For mean-pooled ESM-2 embedding e_v, let U hold the retained PCA directions of the unlabeled library and s their empirical SDs. The fixed coverage features are

$$\phi(v) = \big(1,\ \mathrm{diag}(s)^{-1}U^{\top}(e_v - \bar{e})\big)^{\top}. \tag{39}$$

The intercept and standardized coordinates are reused throughout the campaign; absent usable embeddings, only the coverage source is disabled.

Table 12: Principal defaults in the MERLE implementation.

Setting	Value	Role
PCA rank	64	coverage representation
Requirement cloud J	4096	frontier approximation
Local draws M	32	Monte Carlo acquisition
Shortlist size	64	marginal reranking
Refit interval	8	queries between full field updates
Draw scales (κ, η)	$(0.5, 2.0)$	position and inventory tempering

Previously measured and unmeasured positions are sampled separately for screening, while coverage candidates are redrawn each turn. The requirement cloud and Gaussian residual base draws remain fixed, making gains comparable across proposal modes. The lexicon and PCA transform likewise stay fixed, so new assays share feature coordinates. Sampling changes candidate exposure and computation, not the assay budget.

Teacher scores are joined by variant identifier; embedding rows must match the assay table's order. This prevents either sequence summary from being assigned to the wrong substitution. PCA signs are pinned so fixed-seed weight-space draws are reproducible, although their orientation leaves linear predictions unchanged. No preprocessing uses unqueried assay labels.

C.3 Response Model Fitting

Signed transfer and position effects. At each refresh, ridge teacher updates alternate with Gaussian position updates on $z_{vc} = \Phi^{-1}(f_c(v))$. The two-fold calculation in Equation 6 keeps only transfer that generalizes to held-out assays. For a position with n_p observations, the posterior used in Equation 8 has mean and variance

$$q_{p,c}^2 = \left(\frac{n_p}{\sigma_c^2} + \frac{1}{\tau_c^2}\right)^{-1}, \qquad \mu_{p,c} = \frac{q_{p,c}^2}{\sigma_c^2} \sum_{v \in V_{t-1}:\, p(v)=p} (z_{vc} - \gamma_c \ell_v). \tag{40}$$

An unmeasured position retains its prior mean and variance. Repeated measurements tighten the posterior; the signed teacher readout enters only to the extent supported by held-out assays.

Shared directions and local scale. At later refreshes, the within-position variance is re-estimated from all observed profiles with the current teacher contribution removed; the complementary position variance preserves the normalized total scale. In particular, with $r_{vc} = z_{vc} - \gamma_c \ell_v$ and $\bar{r}_{p,c}$ the mean residual at an observed position, the update is

$$\sigma_c^2 = \left[\frac{\sum_{v \in V_{t-1}} (r_{vc} - \bar{r}_{p(v),c})^2}{\max\{|V_{t-1}| - |\mathcal{P}_{t-1}^{\mathrm{obs}}|, 1\}}\right]_{[0.05,0.95]}, \qquad \tau_c^2 = 1 - \sigma_c^2. \tag{41}$$

Here $\mathcal{P}_{t-1}^{\mathrm{obs}}$ denotes positions already assayed. For cross-condition dependence, the shared residual covariance uses *campaign* queries rather than the range-restricted prior repertoire. Writing m_t for the number of campaign assays and $\widehat{\Sigma}_{\mathrm{camp},t}$ for their residual covariance, the working estimate shrinks toward the calibrated diagonal:

$$\widetilde{\Sigma}_t = \frac{m_t \widehat{\Sigma}_{\mathrm{camp},t} + 32\,\mathrm{diag}(\sigma^2)}{m_t + 32}, \qquad \Sigma_t = D_t \widetilde{\Sigma}_t D_t, \tag{42}$$

where diagonal D_t restores $\mathrm{diag}(\Sigma_t) = \sigma^2$. Before enough campaign profiles are available, Σ_t remains diagonal. Its eigenvectors define the shared modes. Running positionwise second moments supply the projections needed for Equation 7; its shrinkage strength is selected by two-fold held-out Gaussian likelihood separately for each shared mode. Unobserved positions retain unit loadings and hence the shared covariance.

Algorithm 3 Refreshing the assay-specific response field

Given: Profiles observed through $t-1$, position keys, teacher scores ℓ_v, and current field statistics
Return: γ_c, $(\mu_{p,c}, q_{p,c})$, shared modes, and shrinkage

Transfer: retain assay-supported teacher signal

if this is not the first refresh **then**
  Update σ_c^2 from within-position residuals and set $\tau_c^2 = 1 - \sigma_c^2$.
end if
Alternate ridge teacher updates with the position posterior in Eq. 40.
Cross-fit the teacher predictions and shrink their coefficients by Eq. 6.

Geometry: calibrate response directions

Estimate shared campaign-residual covariance, shrinking toward $\mathrm{diag}(\sigma^2)$; obtain its eigenmodes.
Select modewise shrinkage by held-out likelihood; evaluate w_{pk} by Eq. 7 when scoring positions.
return the updated field for candidate scoring.

C.4 Candidate Scoring and Screening

For frontier proposals, evaluating the unmet cloud for every substitution would be costly. Lone-point volume therefore forms a shortlist, which Equation 10 reranks by marginal gain. The position draw in Equation 8 and covariance in Equation 7 supply common samples for both stages. In Algorithm 4, q is a position or a teacher-ranked variant, and $p(q)$ is its position.

Algorithm 4 Candidate scoring and screening in MERLE

Given: Available variants, active positions, unmet cloud $\mathcal{U}_t$, proposal mode $a \in \{\text{position}, \text{teacher}\}$
Return: Next variant x_t

Screen: plausible single-point gains

Consider a sampled set of active positions $\mathcal{P}_t$.
for each $p \in \mathcal{P}_t$ **do**
  Draw $\widetilde{\theta}_p$ by Eq. 8 and $\epsilon_p^{(1:M)} \sim \mathcal{N}(0, \Sigma_p)$.
  $Q_p \leftarrow \{p\}$ in position mode, or the low, middle, and high remaining ℓ_v ranks in teacher mode.
  for each $q \in Q_p$ **do**
    Set $b_q = \widetilde{\theta}_p + \gamma \ell_q$ in teacher mode, or $b_q = \widetilde{\theta}_p$ in position mode.
    $L(q) \leftarrow M^{-1} \sum_{m=1}^{M} \prod_{c=1}^{C} \Phi(b_{q,c} + \epsilon_{p,c}^{(m)})$.
  end for
end for
Retain the highest-L shortlist $\mathcal{S}_t$.

Rerank: marginal coverage of unmet requirements

for each $q \in \mathcal{S}_t$ **do**
  $\widehat{A}_t(q) \leftarrow (JM)^{-1} \sum_{m=1}^{M} \sum_{j \in \mathcal{U}_t} \mathbf{1}\{g_j \le b_q + \epsilon_{p(q)}^{(m)}\}$.
end for
$q^\star \leftarrow \arg\max_{q \in \mathcal{S}_t} \widehat{A}_t(q)$.
return the ranked variant $q^\star$ in teacher mode; otherwise sample x_t uniformly at position $q^\star$.

The teacher mode resolves a shortlisted position to a specific substitution; the position-only mode leaves that choice uniform among its remaining substitutions. This distinction affects proposal construction, not the criterion used to compare shortlisted opportunities. If the cross-fitted teacher coefficient is not yet usable, the teacher mode reverts to position-only scoring. The shortlist is sorted by its fixed candidate indices before marginal reranking, so ties in Monte Carlo gain resolve consistently across numerical backends.

C.5 Coverage Queries and Adaptive Allocation

Pool-wide coverage. The coverage source fits its Bayesian linear model to every profile observed so far, using the fixed embedding features from Section C.2. Its prior and observation-noise scales are tuned on V_0 once; only the posterior is updated during the campaign. Equation 11 can then be evaluated through the full remaining pool's feature second moment:

$$H_t = \sum_{u \in \mathcal{X} \setminus V_{t-1}} \phi(u)\phi(u)^\top, \qquad C_t(v) = \frac{\phi(v)^\top S_t H_t S_t \phi(v)}{\sigma_{\text{obs}}^2 + \phi(v)^\top S_t \phi(v)}. \tag{43}$$

Only the candidate side is sampled afresh at each coverage turn; H_t still sums over the full unmeasured pool. This separates broad predictive coverage from repeatedly querying near-neighbors of the previous choice. The assay result then updates the same position field and unmet cloud as any frontier-mode query.

Realized-gain allocation. The allocator initially tries each available source. Thereafter it uses the empirical mean $\bar{r}_{a,t}$ and variance $s_{a,t}^2$ of that source's rewards from Equation 12, smoothing the variance toward the pooled reward variance $s_{\text{pool},t}^2$:

$$\widetilde{s}_{a,t}^2 = \frac{N_{a,t} s_{a,t}^2 + 2 s_{\text{pool},t}^2}{N_{a,t} + 2}, \qquad a_t = \arg\max_a \left\{ \bar{r}_{a,t} + \frac{\widetilde{s}_{a,t}}{\sqrt{N_{a,t}}} \xi_a \right\}, \quad \xi_a \sim \mathcal{N}(0,1). \tag{44}$$

Here $N_{a,t}$ counts realized queries credited to source a; a source with no credited query is tried before Equation 44 is applied. Only the selected source receives the new reward. If teacher transfer is not usable, its proposed turn executes position-only search and credits the position source instead. Thus the allocator compares observed frontier effects on a common scale, not the numerical values of the frontier and coverage acquisition functions.

C.6 COMPUTATIONAL COMPLEXITY

Let N be the remaining library size, n the number of measured profiles, and S the number of positions. A frontier turn screens P positions, reads at most T substitutions per position, and reranks K candidates using M Gaussian draws. Write $W = \lceil J/64 \rceil$ for the words in a bitset over the J requirements. A coverage turn samples H candidates with d-dimensional features.

Table 13: Time costs of the main operations in MERLE.

Operation	Time complexity
Assimilate one assay	$O(C^2 + CW)$
Frontier screen and rerank	$O(PC^3 + PC^2M + PTMC + KMCW)$
Fit and score a coverage proposal	$O(ndh + h^3 + ndC + Nd^2 + Hd^2)$
Refresh the response field	$O(nC^2 + SGC + C^3)$, once per refit interval

The key saving is that cloud scoring scales with the shortlist K, not the library N; coverage still uses the full pool through a $d \times d$ feature moment. The field refit occurs only every few queries. In the table, G is the fixed number of shrinkage candidates and $h = \min\{n, d\}$; frozen-PLM inference and feature preparation are excluded.

D EMPIRICAL EVALUATION PROTOCOL

D.1 DATASET PANEL

The benchmark panel contains twenty sequence–condition response tables drawn from diverse biological assays. Each row is a candidate sequence, each condition is one objective, and entries are rank-normalized within condition before evaluation so that objectives share a common scale.

Table 14: **Dataset panel and provenance.** $n \times C$ gives the number of sequences and conditions.

Dataset	Size ($n \times C$)	Description	Source and license
rbd_ab_escape	1083×102	SARS-CoV-2 RBD escape	Bloom Lab; GPL-3.0
flu_perth_sera	6555×30	H3 HA escape, Perth sera	Bloom Lab; GPL-3.0
h5_ha_escape	4736×71	H5 HA escape	dms-vep; MIT License
tem1_lactamase	4845×6	TEM-1 under antibiotics	MaveDB; CC0 1.0 by default
hsp90_env	13780×6	Hsp90 fitness environments	eLife (2020); CC BY 4.0
lasv_gp_escape	8826×6	Lassa GP escape	dms-vep; MIT License
nipah_rbp_escape	9856×6	Nipah RBP escape	dms-vep; MIT License
flu_h3_hk19_sera	3293×31	H3 HA escape, HK/19 sera	dms-vep; MIT License
lasv_gp_scripps	8118×10	Lassa GP, Scripps mAbs	dms-vep; MIT License
hiv_env_escape	6653×7	HIV-1 Env escape	dms-vep; MIT License
rabv_g_escape	7655×8	Rabies G escape	dms-vep; MIT License
hcov229e_spike_sera	16486×8	229E spike, human sera	dms-vep; MIT License
nipah_f_escape	8521×6	Nipah F escape	dms-vep; MIT License
vim2_lactamase	5016×9	VIM-2 under ampicillin	MaveDB; CC0 1.0 by default
zikv_e_escape	9576×6	Zika E escape	Bloom Lab; see upstream terms
xbb15_infant_sera	7187×11	XBB.1.5, infant sera	dms-vep; MIT License
ba2_crowe_mabs	7370×6	BA.2 spike, Crowe mAbs	dms-vep; MIT License
ba2_spike_sera	7682×9	BA.2 spike, human sera	dms-vep; MIT License
hiv_broad_human_sera	12559×15	HIV-1 Env, broad sera	Bloom Lab; see upstream terms
hiv_vacc_rabbit_sera	12559×6	HIV-1 Env, rabbit sera	Bloom Lab; see upstream terms

D.2 REFERENCE CAMPAIGN DESIGN

Campaign size. Unless a result caption states otherwise, each campaign uses six objectives sampled without replacement, an initial repertoire of $N = 300$ complete profiles, and a sequential budget of $B = 192$ queries. Each query reveals the complete profile on the selected objectives and costs one unit. We use five random seeds per dataset.

Initial repertoire and candidate pool. For each seed, the initial repertoire is sampled from the bottom 75% of candidates ranked by their maximum response across the selected objectives; the remaining candidates form the query pool. This split reserves higher performers for discovery while holding the initial repertoire size fixed across datasets. The selected objectives, repertoire, and pool are shared by all policies within a dataset–seed instance, enabling paired comparisons.

Features and information access. The main-table setting uses the 64-dimensional whitened PCA representation of the sequence features. Policies may use observed responses only after a query is made. The white-box reference is evaluated separately with the full response table and is not available to the querying policies.

D.3 BASELINE IMPLEMENTATIONS AND ADAPTATIONS

All model-based policies use the same conjugate Bayesian linear response model and fixed features, isolating differences in sequence selection. The original methods assume continuous design spaces, other surrogate models, or multi-query batches. We preserve each acquisition principle while adapting optimization to the finite sequence library and complete-profile feedback. Batch methods retain their batch construction, but queries are consumed sequentially without within-batch refitting. Table 15 summarizes these benchmark adaptations; the shared surrogate is specific to our comparison.

Table 15: **Baseline rules and benchmark adaptations.** Shared model and feature map; method-specific changes for finite-library search.

Policy	Adaptation used in this benchmark
MESMO	Max-value entropy search; 16 coherent function draws, with maxima over the finite library.
qEHVI	Sequential $q = 1$ expected hypervolume improvement; 32 quasi-Monte Carlo draws and a fresh 512-candidate shortlist per step.
BOtied	Mean-of-CDF rule with 20 predictive draws; empirical copula replaces the vine CDF, using a fresh 512-candidate shortlist per step.
MOBO-OSD	Enumerates the finite pool over 20 orthogonal directions; local exploration uses 100 feature-space neighbors, one query per step.
STAGE-BO	Exact sampled Pareto front replaces continuous NSGA-II; adaptive max-min gap constrains expected improvement on a round-robin objective.
ALC (targeted)	Variance reduction weighted by predicted single-profile hypervolume contribution (eq. (3)); uniform-pool ALC is reported in the appendix.
BAIT	Fisher selection on shared features; forward-selects 64, prunes to 32, then issues the batch without within-batch refitting.
PDBAL	Posterior diameter over the full pool with 16 weight draws; plate design adapted to one query.

D.4 Metrics and Scoring

Let Y_0 be the response profiles in the initial repertoire, Y_B those observed after the full budget, and $Z = \mathcal{F}(\mathcal{X})$ the true Pareto front of the finite library, available only to the evaluator. All five indicators in Table 1 are reported as improvements over Y_0, so larger values are better.

dHV. The primary score is the gain in dominated hypervolume relative to the zero reference point:

$$\mathrm{dHV} = \mathrm{HV}(Y_B) - \mathrm{HV}(Y_0). \tag{45}$$

We estimate hypervolume with 200,000 uniform samples, shared across policies on each instance.

dHV-max. This score recomputes dHV after deleting the measured profile that contributes most to final hypervolume. It asks whether the gain survives without that single exceptional point:

$$\mathrm{dHV\text{-}max} = \min_{y \in Y_B} \mathrm{HV}(Y_B \setminus \{y\}) - \mathrm{HV}(Y_0). \tag{46}$$

IGD$^+$. For each point on the true front, IGD$^+$ takes the distance to the nearest measured profile, counting only coordinates in which that profile falls short (Ishibuchi et al., 2015):

$$\mathrm{IGD}^+(Z, Y) = \frac{1}{|Z|} \sum_{z \in Z} \min_{y \in Y} \|(z - y)_+\|_2 \,. \tag{47}$$

Here $(\cdot)_+$ clips negative coordinates to zero. The reported score is $\mathrm{IGD}^+(Z, Y_0) - \mathrm{IGD}^+(Z, Y_B)$.

eps. The additive epsilon indicator takes the largest shortfall remaining on the true front after each front point is matched to its best measured profile (Zitzler et al., 2003):

$$\varepsilon(Z, Y) = \max_{z \in Z} \min_{y \in Y} \max_{c} (z_c - y_c). \tag{48}$$

The table reports $\varepsilon(Z, Y_0) - \varepsilon(Z, Y_B)$, the reduction in worst-case shortfall.

R_2. This indicator averages the best weighted Chebyshev distance to the ideal point over 128 weight vectors W drawn from a uniform Dirichlet distribution (Hansen & Jaszkiewicz, 1998):

$$R_2(W, Y) = \frac{1}{|W|} \sum_{w \in W} \min_{y \in Y} \max_{c} w_c (1 - y_c). \tag{49}$$

The reported score is $R_2(W, Y_0) - R_2(W, Y_B)$, the reduction in average weighted distance.

D.5 Aggregation and Uncertainty

Seeds capture trajectory variability within an assay, while datasets are the unit of generalization. Table 16 summarizes how we aggregate results, quantify uncertainty, and handle reference policies. Appendix 17 reports the per-dataset seed means and standard deviations.

Table 16: **Aggregation, uncertainty, and comparison rules.** Each dataset contributes one equally weighted unit to panel estimates and paired policy comparisons.

Stage	Procedure
Seed aggregation	Average the five campaign seeds within each dataset for each policy and metric. Retain the seed standard deviation as trajectory variability; do not pool seeds across datasets.
Panel mean and SE	Give the twenty dataset means equal weight; compute SE across those means, not across seed runs. Assay size does not affect weight.
Paired comparison	Subtract policies' seed averages within each shared dataset before summarizing differences across the panel.
Bootstrap interval	Resample datasets and keep both policies paired in each draw, preserving shared objectives, initial repertoire, and candidate pool.
Win count	Compare seed-averaged policy scores per dataset; ties follow the rule stated in the corresponding result table.
Reference bounds	Random is a screening reference; white-box uses the full response table and is not an admissible querying policy.
Changed settings	Captions state the dataset count and departures in objectives, feedback, budget, or eligibility; interpret each within its panel.

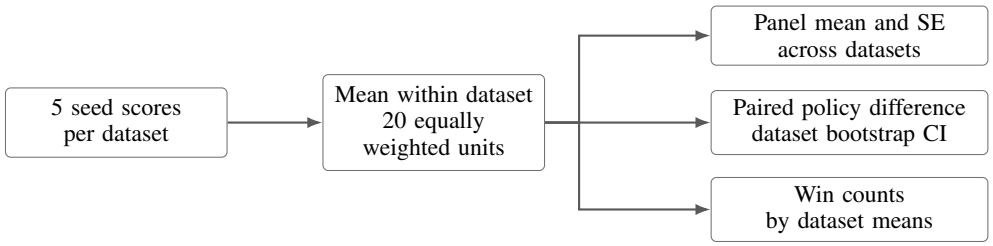


Figure 9: **From repeated campaigns to panel summaries.** Seed runs are first collapsed within each dataset; all downstream estimates use the dataset-level units.

E FURTHER EMPIRICAL EVIDENCE

E.1 QUALITATIVE ANALYSIS OF THE SITE READOUT

The site report maps where a campaign found useful variation under its assayed objectives; it does not assign an epitope or molecular mechanism. Figures 10–12 trace one LASV glycoprotein campaign, from its attained front through the 192 queries to the positions visited in two independent assays of the same protein. The assays need not agree on the effect of each substitution.

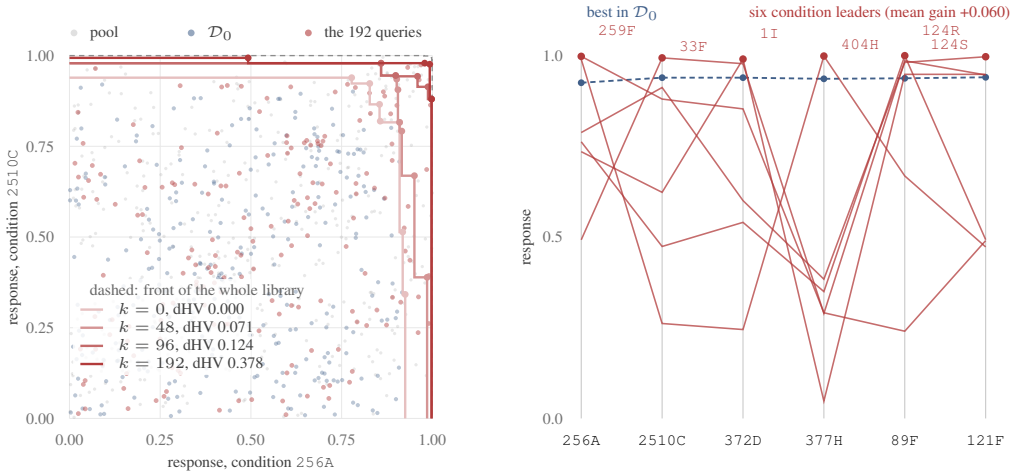


Figure 10: **A campaign's attained front and its leaders.** LASV glycoprotein escape, seed 0. *Left:* rank-normalised front in two of six conditions after k queries; dashed is the full-library front. *Right:* the variants leading the final front across six conditions. No variant leads on every condition.

The plane hides four objectives. The front has no single winner: a leader in one condition may be weak in another. Figure 11 tracks all six across the budget.

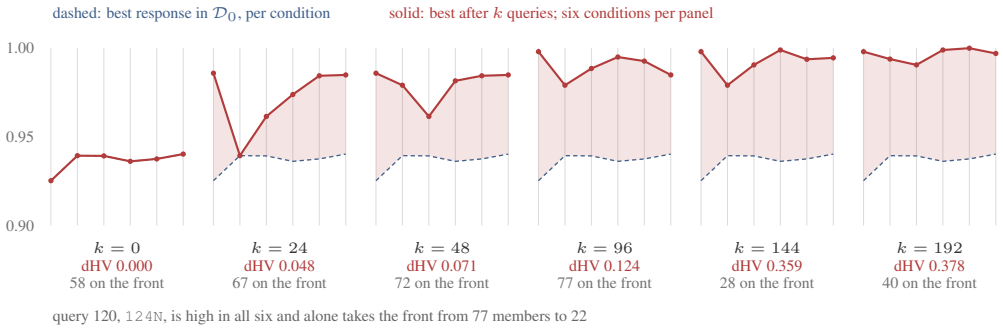


Figure 11: **The same campaign in all six conditions.** Dashed: best starting response; solid: best after k queries; shading: gain. The front count drops at query 120 because one purchase improves all six conditions and dominates many earlier candidates.

One purchase can reshape the front. Condition maxima improve at different rates, with some gains arriving only in the final quarter. At query 120, a variant at a site credited by both assays raises all six ceilings and dominates many older front members. The falling front count thus signals stronger dominance, not a loss of useful responses. Earlier checkpoints already improved several conditions, so this single purchase does not account for the entire campaign's gain. Figure 12 locates the repeatedly queried sites behind this shift. The site readout should be read with the query history: new purchases can change which variants support the front even when its overall shape barely moves.

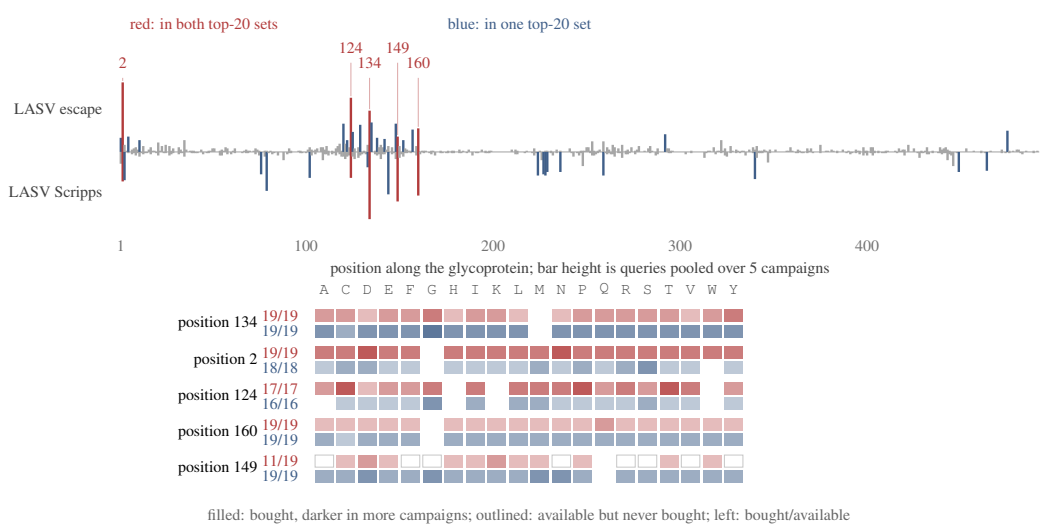


Figure 12: **Where the budget went, on two independent measurements of one protein.** Queries pooled over the five campaigns of each assay, by position; the upper track is one laboratory's measurement of the LASV glycoprotein and the lower track the other's. Both concentrate on a small set of positions and the heavily visited sites largely coincide, while their order does not. The lower panel prints the substitutions bought at the positions both assays rank in their top 20: a credited position is one the campaigns nearly exhaust, not a single variant they happen to like.

What the case illustrates. The position trace in Figures 10–12 connects the attained front to the sites queried in two independent assays. The aggregate deletion, stability, and cross-assay controls are presented with their tables in Sections E.3 and E.4.

E.2 Performance across datasets

Figure 13 tests whether any one protein system drives the panel mean. Table 17 then gives the individual datasets behind that aggregate at the full budget.

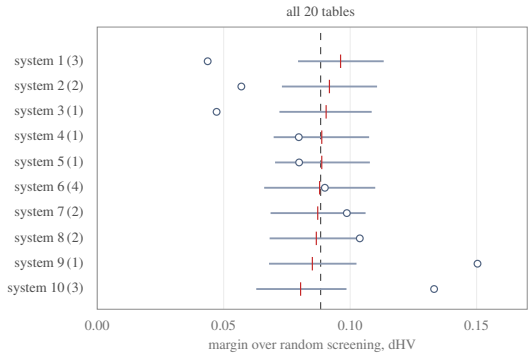


Figure 13: **Leave-one-system-out gain.** Red: MERLE's dHV margin over random on the remaining systems, with bootstrap interval; ring: held-out system; dashed: full-panel mean.

No single protein system carries the panel mean. Several response tables share a protein system, so averaging by table could overweight it. Figure 13 recomputes MERLE's gain after omitting each system in turn: the remaining-system mean stays between 0.080 and 0.096, close to the full-panel 0.088. This guards against one repeated system driving the result, though it does not establish transfer to an unmeasured protein; that still requires a new assay.

Table 17: **Full-budget dHV by dataset ($B = 192$).** Mean ± standard deviation over five seeds; bold marks the best ranked policy per row. Bottom rows give the panel mean and within-dataset mean rank; RANDOM and WHITE-BOX are unranked bounds.

	Multi-objective BO								Active-learning designs						Ours	Bounds	
Dataset	MESMO	PFES	qEHVI	BOtied	MOBO-OSD	STAGE-BO	RHVS-U	RHVS-T	ALC-t	ALC	ALM	BAIT	PDBAL	k-center	MERLE	Random	White-box
SARS-CoV-2 spike: monoclonals and human sera																	
rbd_ab_escape	0.423 ±0.051	0.417 ±0.059	0.439 ±0.049	0.397 ±0.075	0.395 ±0.048	0.428 ±0.055	0.405 ±0.074	0.429 ±0.059	0.444 ±0.052	0.406 ±0.059	0.396 ±0.061	0.411 ±0.056	0.410 ±0.062	0.407 ±0.054	**0.446** ±0.047	0.350 ±0.106	0.458 ±0.041
ba2_spike_sera	0.421 ±0.021	0.357 ±0.087	0.362 ±0.108	0.435 ±0.032	0.315 ±0.105	0.377 ±0.064	0.414 ±0.087	0.400 ±0.135	0.345 ±0.072	0.432 ±0.044	0.400 ±0.070	0.431 ±0.045	0.409 ±0.019	0.387 ±0.106	**0.439** ±0.036	0.379 ±0.026	0.536 ±0.040
ba2_crowe_mabs	0.373 ±0.052	0.370 ±0.081	0.404 ±0.075	0.417 ±0.034	**0.439** ±0.036	0.376 ±0.075	0.382 ±0.045	0.377 ±0.072	0.392 ±0.008	0.344 ±0.046	0.319 ±0.024	0.336 ±0.029	0.338 ±0.014	0.275 ±0.062	0.412 ±0.068	0.311 ±0.090	0.499 ±0.023
xbb15_infant_sera	0.333 ±0.103	0.328 ±0.122	0.313 ±0.117	0.315 ±0.121	0.313 ±0.105	0.298 ±0.132	0.312 ±0.097	0.283 ±0.117	0.285 ±0.108	0.279 ±0.092	0.271 ±0.083	0.306 ±0.069	0.259 ±0.088	0.309 ±0.073	**0.398** ±0.075	0.295 ±0.051	0.511 ±0.047
mean	0.387	0.368	0.379	0.391	0.365	0.370	0.379	0.372	0.367	0.365	0.347	0.371	0.354	0.345	**0.424**	0.334	0.501
Influenza hemagglutinin: H3 sera and H5 escape																	
flu_perth_sera	0.206 ±0.122	0.215 ±0.067	0.250 ±0.050	0.250 ±0.028	0.257 ±0.048	0.286 ±0.179	0.299 ±0.118	0.235 ±0.062	0.246 ±0.066	0.280 ±0.146	0.324 ±0.117	0.250 ±0.170	0.295 ±0.137	0.331 ±0.162	**0.404** ±0.146	0.261 ±0.172	0.573 ±0.039
flu_h3_hk19_sera	0.294 ±0.039	0.278 ±0.040	0.245 ±0.042	0.247 ±0.050	0.252 ±0.044	0.293 ±0.060	0.320 ±0.071	0.294 ±0.048	0.239 ±0.040	0.271 ±0.052	0.289 ±0.061	0.267 ±0.045	0.271 ±0.071	0.336 ±0.125	**0.485** ±0.055	0.337 ±0.089	0.497 ±0.055
h5_ha_escape	0.373 ±0.052	0.329 ±0.039	0.351 ±0.035	0.375 ±0.025	0.263 ±0.115	0.368 ±0.058	0.356 ±0.059	0.344 ±0.072	0.357 ±0.052	0.320 ±0.054	0.317 ±0.073	0.351 ±0.042	0.338 ±0.053	0.357 ±0.063	**0.392** ±0.042	0.284 ±0.094	0.482 ±0.018
mean	0.291	0.274	0.282	0.291	0.258	0.316	0.325	0.291	0.281	0.290	0.310	0.289	0.302	0.341	**0.427**	0.294	0.517
HIV-1 envelope glycoprotein																	
hiv_env_escape	0.384 ±0.046	0.325 ±0.073	0.375 ±0.036	0.342 ±0.032	0.369 ±0.055	0.341 ±0.017	0.340 ±0.022	0.320 ±0.088	0.357 ±0.039	0.364 ±0.037	**0.420** ±0.047	0.354 ±0.063	0.329 ±0.063	0.334 ±0.092	0.398 ±0.051	0.379 ±0.029	0.484 ±0.028
hiv_broad_human_sera	0.118 ±0.053	0.134 ±0.057	0.145 ±0.053	0.131 ±0.060	0.144 ±0.036	0.123 ±0.064	0.123 ±0.048	0.114 ±0.041	0.145 ±0.053	0.152 ±0.044	0.141 ±0.046	0.145 ±0.026	0.155 ±0.042	0.167 ±0.056	**0.192** ±0.083	0.172 ±0.039	0.458 ±0.030
hiv_vacc_rabbit_sera	0.328 ±0.153	0.304 ±0.075	0.161 ±0.061	0.191 ±0.125	0.145 ±0.061	0.301 ±0.146	0.364 ±0.115	0.180 ±0.076	0.339 ±0.056	0.222 ±0.090	0.205 ±0.066	0.208 ±0.069	0.176 ±0.051	0.286 ±0.102	**0.407** ±0.053	0.315 ±0.026	0.469 ±0.022
mean	0.276	0.254	0.227	0.221	0.219	0.255	0.276	0.205	0.280	0.246	0.255	0.236	0.220	0.262	**0.332**	0.289	0.470
Lassa virus glycoprotein																	
lasv_gp_escape	**0.463** ±0.051	0.384 ±0.093	0.379 ±0.107	0.415 ±0.119	0.370 ±0.058	0.433 ±0.070	0.447 ±0.051	0.406 ±0.057	0.393 ±0.094	0.295 ±0.074	0.375 ±0.077	0.345 ±0.120	0.409 ±0.079	0.350 ±0.058	0.461 ±0.057	0.326 ±0.094	0.530 ±0.045
lasv_gp_scripps	0.488 ±0.048	0.482 ±0.048	0.495 ±0.039	0.502 ±0.034	**0.508** ±0.037	0.495 ±0.034	0.488 ±0.034	0.498 ±0.028	0.493 ±0.032	0.430 ±0.061	0.412 ±0.055	0.414 ±0.059	0.393 ±0.047	0.431 ±0.067	0.466 ±0.048	0.404 ±0.068	0.532 ±0.041
mean	**0.475**	0.433	0.437	0.458	0.439	0.464	0.468	0.452	0.443	0.363	0.394	0.380	0.401	0.390	0.463	0.365	0.531
Nipah virus: fusion and receptor-binding proteins																	
nipah_f_escape	0.238 ±0.017	**0.262** ±0.053	0.223 ±0.062	0.242 ±0.062	0.241 ±0.057	0.253 ±0.045	0.228 ±0.033	0.247 ±0.039	0.260 ±0.037	0.205 ±0.026	0.155 ±0.023	0.215 ±0.021	0.186 ±0.016	0.211 ±0.041	0.262 ±0.048	0.223 ±0.054	0.416 ±0.023
nipah_rbp_escape	0.174 ±0.068	0.186 ±0.079	0.165 ±0.068	0.220 ±0.080	0.161 ±0.069	0.181 ±0.074	0.194 ±0.092	0.205 ±0.068	0.172 ±0.046	0.181 ±0.059	0.137 ±0.021	0.210 ±0.062	0.162 ±0.048	0.242 ±0.085	**0.299** ±0.072	0.224 ±0.050	0.521 ±0.025
mean	0.206	0.224	0.194	0.231	0.201	0.217	0.211	0.226	0.216	0.193	0.146	0.213	0.174	0.227	**0.280**	0.223	0.469
Other viruses: seasonal coronavirus, rabies, Zika																	
hcov229e_spike_sera	**0.416** ±0.035	0.393 ±0.018	0.363 ±0.036	0.362 ±0.031	0.346 ±0.063	0.391 ±0.054	0.414 ±0.030	0.366 ±0.052	0.351 ±0.051	0.343 ±0.038	0.338 ±0.027	0.358 ±0.039	0.329 ±0.045	0.385 ±0.031	0.405 ±0.042	0.325 ±0.094	0.469 ±0.025
rabv_g_escape	0.242 ±0.062	0.224 ±0.049	0.198 ±0.052	0.224 ±0.047	0.199 ±0.055	0.233 ±0.072	0.200 ±0.063	0.186 ±0.080	0.226 ±0.064	0.198 ±0.059	0.235 ±0.071	0.204 ±0.068	0.190 ±0.028	**0.269** ±0.110	0.261 ±0.076	0.214 ±0.090	0.476 ±0.029
zikv_e_escape	0.182 ±0.024	0.170 ±0.096	0.182 ±0.057	0.178 ±0.046	0.173 ±0.038	0.159 ±0.043	0.200 ±0.088	0.164 ±0.058	0.149 ±0.074	0.100 ±0.068	0.113 ±0.031	0.135 ±0.035	0.176 ±0.111	0.186 ±0.081	**0.284** ±0.035	0.204 ±0.044	0.464 ±0.027
mean	0.280	0.262	0.248	0.255	0.239	0.261	0.271	0.238	0.242	0.214	0.229	0.232	0.232	0.280	**0.317**	0.248	0.469
Non-viral protein fitness under selection																	
tem1_lactamase	0.238 ±0.040	0.293 ±0.038	0.295 ±0.025	0.291 ±0.043	0.283 ±0.050	0.272 ±0.061	0.240 ±0.048	0.260 ±0.040	0.343 ±0.045	0.291 ±0.038	0.271 ±0.047	0.303 ±0.038	0.239 ±0.071	0.328 ±0.050	**0.371** ±0.044	0.226 ±0.044	0.467 ±0.037
vim2_lactamase	0.370 ±0.035	0.332 ±0.048	0.352 ±0.038	0.338 ±0.034	0.349 ±0.034	0.366 ±0.042	0.364 ±0.015	0.367 ±0.036	0.346 ±0.073	0.321 ±0.042	0.197 ±0.081	0.305 ±0.026	0.220 ±0.076	0.338 ±0.036	**0.377** ±0.024	0.315 ±0.033	0.471 ±0.023
hsp90_env	0.221 ±0.135	0.179 ±0.050	0.131 ±0.031	0.137 ±0.028	0.147 ±0.018	0.221 ±0.135	0.207 ±0.118	0.129 ±0.028	0.178 ±0.019	0.220 ±0.042	0.211 ±0.049	0.235 ±0.043	0.245 ±0.032	0.254 ±0.030	**0.366** ±0.030	0.216 ±0.079	0.486 ±0.029
mean	0.276	0.268	0.259	0.255	0.260	0.287	0.270	0.252	0.289	0.277	0.226	0.281	0.235	0.307	**0.372**	0.252	0.474
Panel mean	0.314	0.298	0.291	0.300	0.283	0.310	0.315	0.290	0.303	0.283	0.276	0.289	0.276	0.309	**0.376**	0.288	0.490
Mean rank	6.20	8.50	8.40	7.15	9.65	7.05	6.80	9.15	7.95	10.05	10.75	9.10	10.60	6.80	**1.85**	—	—

Dataset-level variation. Table 17 expands the panel mean of Table 1 into antigen groups and individual datasets. The within-dataset rank separates consistent performance from a high score driven by one large assay. Results vary even within a protein family: the two LASV glycoprotein measurements have different strongest policies.

Breadth with assay-level exceptions. MERLE ranks first on all three influenza assays and on several SARS-CoV-2, HIV, Nipah-RBP, Zika, and non-viral datasets. Other policies lead on individual rows, including both LASV scans. The row-level ranking therefore shows breadth across assay families while preserving exceptions that a single panel mean would conceal; it does not imply a universal winner.

Repertoires formed by different laboratory choices. The main panel changes the response range represented in $\mathcal{D}_0$; Table 18 instead changes how the initial measurements were selected. Whole positions, one substitution chemistry, a feature-space cover, and the teacher's highest-scored rows leave different gaps for the campaign to search. MERLE has the highest dHV in every arm and gains 0.065–0.094 over the random screen within the same arm. The teacher-top arm gives the smallest margin, so teacher-selected starting rows do not inflate MERLE's measured advantage.

Table 18: **Four ways to assemble the initial repertoire.** Full-budget dHV and paired gain over random screening (parentheses) under each matched selection rule. The columns start from whole sites, one substitution chemistry, a feature-space cover, or teacher-top rows. MERLE leads within each rule, with its smallest margin under teacher-top selection.

Policy	whole sites	one chemistry	a feature cover	the teacher's top rows
MESMO	0.341 (+0.029)	0.335 (+0.057)	0.319 (+0.043)	0.338 (+0.026)
qEHVI	0.351 (+0.038)	0.346 (+0.068)	0.290 (+0.014)	0.316 (+0.005)
BOtied	0.363 (+0.050)	0.351 (+0.072)	0.293 (+0.017)	0.315 (+0.004)
MOBO-OSD	0.356 (+0.044)	0.342 (+0.064)	0.288 (+0.013)	0.313 (+0.001)
STAGE-BO	0.352 (+0.040)	0.343 (+0.065)	0.301 (+0.026)	0.332 (+0.021)
ALC	0.327 (+0.015)	0.319 (+0.041)	0.314 (+0.039)	0.316 (+0.004)
BAIT	0.297 (−0.015)	0.272 (−0.007)	0.301 (+0.026)	0.303 (−0.009)
PDBAL	0.286 (−0.027)	0.269 (−0.009)	0.289 (+0.014)	0.296 (−0.015)
MERLE	**0.389** (+0.077)	**0.362** (+0.083)	**0.369** (+0.094)	**0.377** (+0.065)
Random	0.312 (+0.000)	0.278 (+0.000)	0.275 (+0.000)	0.312 (+0.000)
White-box	0.512 (+0.199)	0.481 (+0.203)	0.480 (+0.204)	0.502 (+0.191)

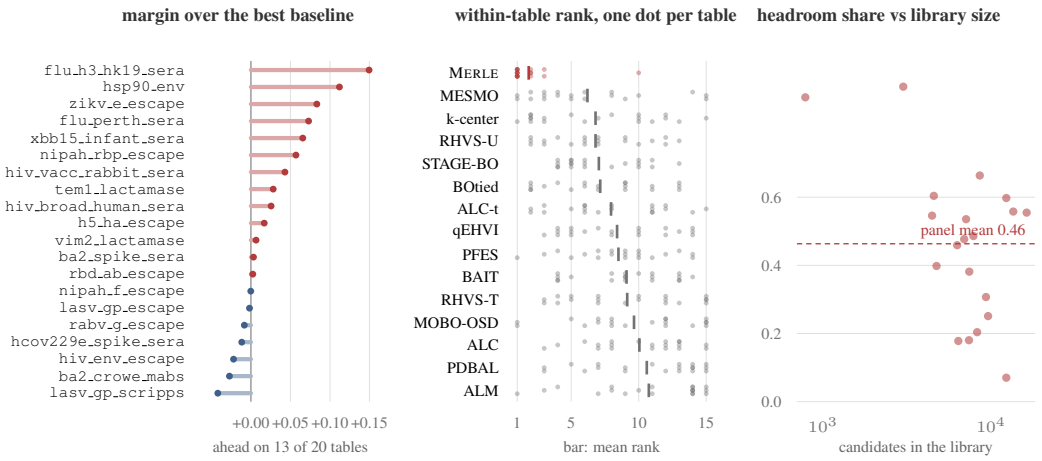


Figure 14: **Where the panel mean comes from.** *Left:* full-budget dHV of MERLE minus the best baseline on the same table, one row per table, means over five seeds. *Middle:* the within-table rank of every policy, one dot per table, ordered by mean rank; dots are offset vertically where they would coincide. *Right:* headroom share, $(\text{dHV} - \text{random})/(\text{oracle} - \text{random})$, against the number of candidates in the library.

Panel heterogeneity. Figure 14 reads the same store three ways. The margin over the strongest baseline on each table is positive on 13 of the 20, so the panel mean is a majority of moderate wins and not a handful of large ones; the rank panel shows the same thing as a distribution rather than the single mean-rank row of Table 17; and the headroom share does not grow with the size of the candidate pool, which rules out the reading that the margin is an artefact of small libraries. This normalized view complements absolute dHV by asking how much of each table's available oracle gap the campaign closes, rather than which table has the largest raw score.

E.3 FAITHFULNESS AND ROBUSTNESS OF THE READOUT

Table 19 expands the intervention summarized in Table 10. The credited positions are removed from the candidate pool and the campaign is rerun. The top, bottom, and random groups are matched by the number of candidate variants excluded, since sites carry different numbers of substitutions. The top-ranked group causes the largest loss at every removal size, and the paired top-minus-random interval excludes zero throughout. The loss is already near its maximum at five positions, indicating that the report concentrates the campaign's sensitivity in a small set. This is faithfulness to the policy's own behavior on these instances; it does not establish the same effect in every assay.

Table 19: **Full position-deletion test.** Paired dHV change from the unmodified campaign, with groups matched on candidate variants removed.

k	removed	top	bottom	random	top − random
5	82	−0.0494 [−0.0771,−0.0240]	−0.0096 [−0.0229,+0.0029]	−0.0010 [−0.0117,+0.0101]	−0.0484 [−0.0732,−0.0273]
10	159	−0.0498 [−0.0838,−0.0190]	−0.0047 [−0.0190,+0.0086]	−0.0042 [−0.0142,+0.0065]	−0.0456 [−0.0748,−0.0198]
20	303	−0.0435 [−0.0777,−0.0163]	−0.0012 [−0.0139,+0.0105]	−0.0041 [−0.0226,+0.0124]	−0.0394 [−0.0753,−0.0069]
40	591	−0.0384 [−0.0757,−0.0053]	+0.0001 [−0.0109,+0.0117]	−0.0053 [−0.0200,+0.0087]	−0.0331 [−0.0665,−0.0031]

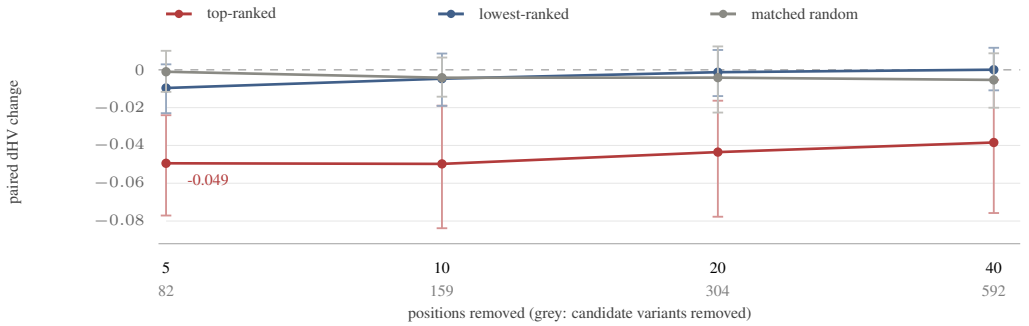


Figure 15: **Deleting the credited positions.** Paired dHV change from the unmodified campaign when the k highest-ranked, lowest-ranked, or candidate-matched random positions leave the candidate pool; whiskers are 95% CIs over the 20 tables, and the grey figures are the candidate variants removed, which the three groups are matched on. Table 19 tabulates the same measurements.

Figure 16 shows the tables behind those four intervals. At five positions the top group costs more than its matched random control on every table in the panel; by forty the two groups have converged on a third of them. The effect is therefore panel-wide and concentrated in the first few credited positions, rather than carried by a few instances.

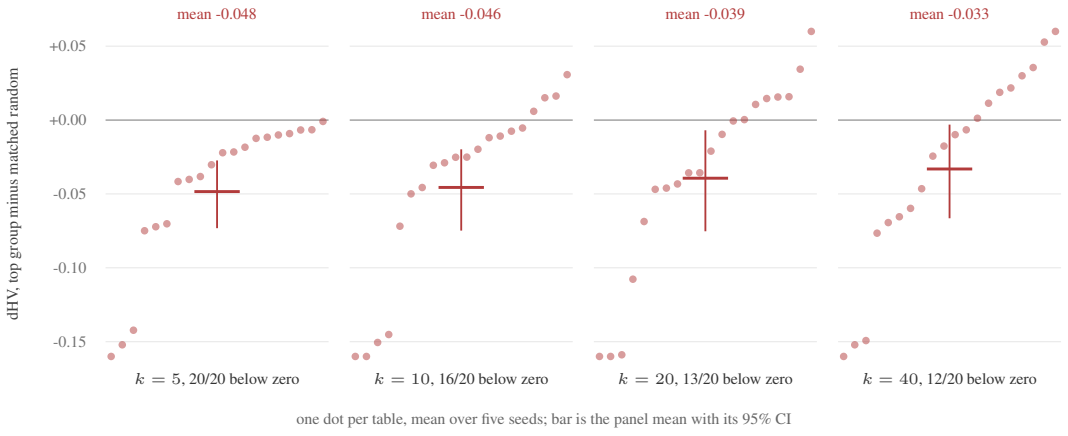


Figure 16: **The deletion test, table by table.** Paired difference between deleting the k top-ranked positions and deleting k candidate-matched random ones, one dot per table (mean over five seeds), sorted within each panel. Negative means the credited positions mattered more. The bar is the panel mean of Table 19 with its 95% CI.

Stability across campaign and task redraws. Table 20 separates a new policy draw on the same instance from redrawn objectives on the same table and from different tables. Rank agreement falls from 0.742 under a new policy draw to 0.238 under new objectives and reaches 0.003 between tables. The leading within-site mode remains similar even between unrelated tables because a common-direction mode is shared by construction; its apparent stability should not be read as biological replication. The second mode and the position ranking are more discriminating checks.

Table 20: **Stability of the position readout.** Rank agreement, top-20 overlap, and within-site modes across three comparison distances, from a redrawn campaign to a different table.

Compared across	n	Spearman	top-20 overlap	|cos| m1	|cos| m2
same instance, redrawn campaign	300	0.742 [+0.723,+0.760]	0.361 [0.344,0.378]	0.952	0.756
same table, redrawn objectives	200	0.238 [+0.217,+0.259]	0.114 [0.102,0.126]	0.829	0.352
different tables	190	0.003 [-0.004,+0.011]	0.033 [0.028,0.039]	0.791	0.327

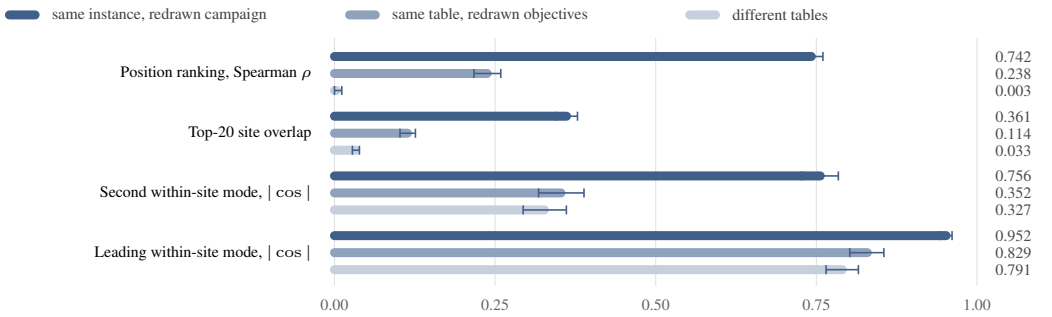


Figure 17: **What survives a redraw.** The four agreement measures of Table 20 on one axis, ordered by how far each falls as the comparison moves from a redrawn campaign to a different table. Whiskers are 95% CIs where the store carries them; the within-site modes are point estimates. A mode shared by construction stays high even between unrelated tables, so only the upper two rows discriminate. Ranking and overlap are the more demanding tests of transferable site priorities.

Robustness to the published constants. Figure 18 changes each of eight constants in turn while holding the other seven fixed. Fourteen of the sixteen intervals cross zero. The two resolved losses occur when the Monte Carlo sample or reranking width is reduced; increasing either does not yield a resolved gain. The margin does not depend on a narrowly tuned setting of one constant, although this one-at-a-time check cannot rule out interactions. The tested upward changes provide no resolved gain; the clearest cost comes from reducing the computation used for Monte Carlo sampling or reranking. These are the two reductions that most directly limit search fidelity.

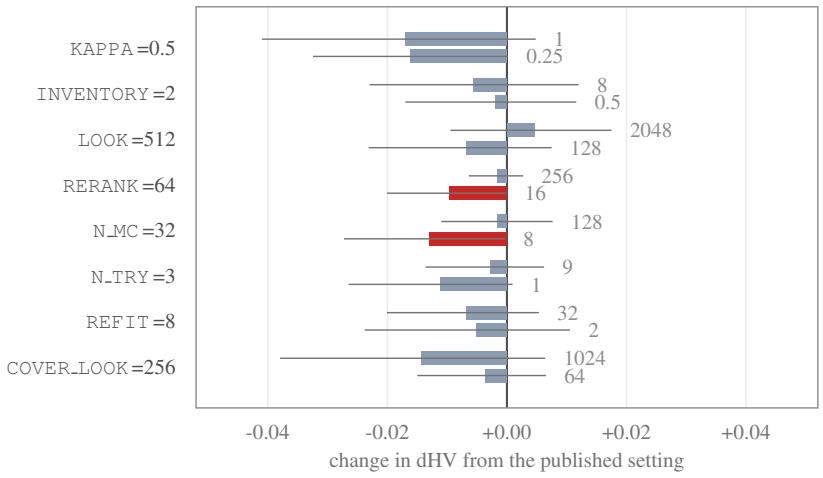


Figure 18: **Sensitivity to eight constants.** Paired dHV change when one constant moves low or high while the other seven retain their published settings. Whiskers are bootstrap intervals over libraries; red marks an interval excluding zero. Labels give the low and high values used in each perturbation.

E.4 BIOLOGICAL CONSISTENCY ACROSS ASSAYS

Table 21 uses four pairs of measurements of a matched protein, including different laboratories and antibody or serum panels. It reports both full position rankings and the overlap of sites that the campaigns actually visited. The sequence-only lexicon is a control: its score is shared across paired assays even when it has learned nothing from their responses. The visited-site overlap is therefore reported against its chance expectation, alongside the readout's same-assay correlation with the lexicon. The all-position correlation is small because many positions were never queried; restricting attention to visited sites gives a 3.02-fold overlap lift. Their ordering still transfers weakly, and the oracle's cross-assay correlation shows that even perfect knowledge of one assay would not imply identical rankings in the other.

Table 21: **Independent-assay replication.** Rank agreement and top-20 overlap across matched protein assays, plus visited-site overlap relative to chance.

Readout	Spearman across the pair	top-20 overlap
MERLE	**0.061** [+0.032,+0.090]	**0.075** [0.050,0.102]
Lexicon (the PLM alone)	0.966 [+0.947,+0.983]	0.900 [0.857,0.938]
Shared surrogate	0.140 [+0.088,+0.195]	0.095 [0.055,0.140]
Oracle (ceiling)	0.449 [+0.358,+0.538]	0.173 [0.123,0.225]
restricted to the positions each campaign actually queried		
shared, as a share of the smaller set	**0.331**	chance 0.133
lift over chance	**3.02**× [2.30,3.84]	Spearman +0.058
MERLE against the teacher on the *same* table: $\rho = +0.042$ [+0.030, +0.052]		

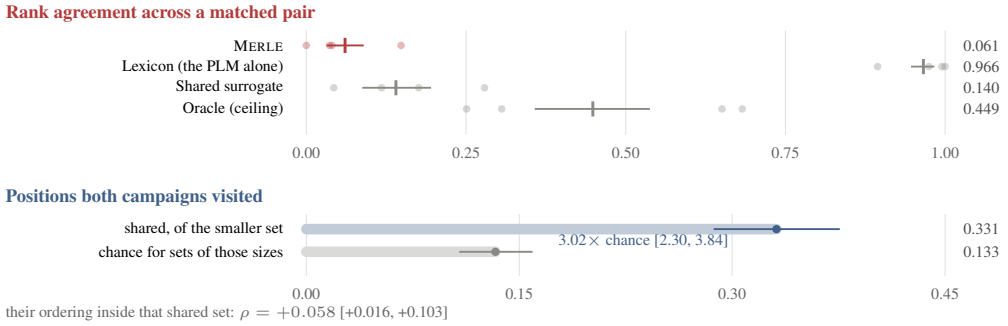


Figure 19: **Replication across an independent assay.** Upper panel: Spearman correlation between the position rankings of two campaigns on a matched pair, one faint dot per pair behind the mean and its 95% CI. Lower panel: the positions both campaigns queried, as a share of the smaller visited set, against the overlap two sets of those sizes would have by chance. The sequence-only lexicon replicates almost perfectly because it answers both assays identically, and the oracle shows that even exact knowledge of one assay does not order the other. Drawn from the measurements in Table 21.

Figure 20 separates the four pairs the table averages. The readout's rank correlation is small on three of them and negative on the fourth, while the visited-site overlap is above its chance level on all four — the pair whose rankings disagree outright still visits the same sites about three times as often as chance. Averaging the four hides that, which is why the set-level claim is the one the section makes.

What transfers between assays. The overlap is evidence that the campaigns repeatedly investigate a restricted set of positions on the same protein, even when the measured response changes. It is not evidence that those positions have the same effect in both assays: the weak, sometimes negative, rank correlations separate a reproducible search region from a reproducible ordering of biological effects. Follow-up tests must establish the assay-specific effect of each position. This distinction matters for transfer: site selection can remain stable while response direction and magnitude vary across laboratories, panels, and assay conditions. The result supports cautious reuse of positional priors, not blind reuse of effect estimates.

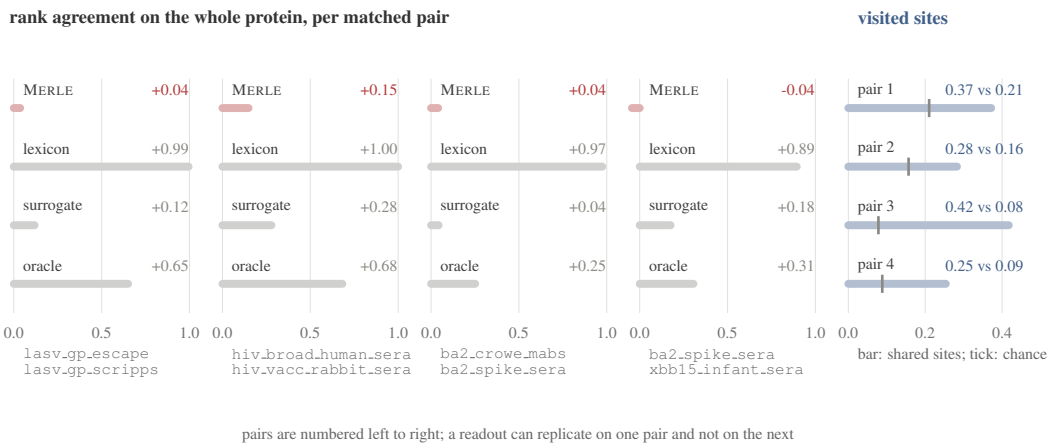


Figure 20: **The four matched pairs, separately.** First four panels: rank agreement across the pair for each readout, from the same store as Table 21. Last panel: the positions both campaigns queried, as a share of the smaller visited set, with the tick marking what two sets of those sizes would share by chance; pooling its five seeds over the four pairs reproduces the 0.331 and 0.133 of that table.

E.5 DECISION UTILITY OF POSITION EXPLANATIONS

A rule from a previous campaign. Table 22 tests actionability as a separate property. A fixed rule receives the prior campaign, chooses at most 40 positions, and commits its queries before seeing new responses. The visit log is the strongest information-matched control in the first test: readout weights do not improve on it. This limits the claim that numerical weights add value beyond knowing which sites were visited.

Comparing different rankings. In Table 23, the MERLE readout yields the strongest non-oracle fixed rule among the rankings evaluated. The independent-assay columns quantify a different property: a sequence-only lexicon replicates almost perfectly because it gives the same answer to both assays, yet its rule performs worse. Thus repeatability and usefulness for query selection must be assessed separately. Neither comparison assigns a molecular mechanism to individual residues.

Table 22: **A fixed rule from a previous campaign.** Explanation-based rules and information-matched controls for the next-seed comparison under a fixed query budget.

Arm	Reads	ΔHV	over the visit log
weight over positions	explanation	0.387 [0.353,0.418]	−0.0031 [−0.0108,+0.0041]
+ chemistry	explanation	0.385 [0.352,0.414]	−0.0053 [−0.0187,+0.0086]
leading axis	explanation	0.359 [0.320,0.394]	**−0.0306 [−0.0450,−0.0182]**
arms that read no explanation			
where it queried (visit log)	—	0.390 [0.359,0.419]	—
what it measured	—	0.367 [0.330,0.402]	**−0.0226 [−0.0377,−0.0074]**
40 random positions	—	0.264 [0.224,0.307]	**−0.1257 [−0.1684,−0.0824]**
MERLE	*no head start*	0.376 [0.342,0.408]	—
Random	*no head start*	0.288 [0.259,0.317]	—
White-box	*no head start*	0.490 [0.475,0.505]	—

Table 23: **Position rankings on two axes.** Independent-assay replication and the dHV of fixed rules that act on each ranking across the matched assay pair under a fixed query budget.

Ranking	replication lift	Spearman across the pair	rule ΔHV
MERLE	**2.79**× [1.98,3.60]	**0.061** [+0.032,+0.090]	**0.388** [0.370,0.406]
Lexicon (the PLM alone)	39.29× [33.90,44.53]	0.966 [+0.947,+0.983]	0.303 [0.284,0.322]
Shared surrogate	3.42× [2.22,4.73]	0.140 [+0.088,+0.195]	0.269 [0.248,0.291]
Repertoire spread	3.32× [2.44,4.29]	0.076 [+0.051,+0.103]	0.199 [0.181,0.218]
Oracle (ceiling)	6.62× [5.27,7.87]	0.449 [+0.358,+0.538]	0.433 [0.419,0.447]
Random (floor)	1.17× [0.61,1.76]	0.000 [-0.019,+0.017]	0.282 [0.260,0.305]

F Discussion: Scope, Limitations, and Broader Implications

F.1 Reproducibility Statement

The accompanying code contains the essential files needed to reproduce the main results, including the method implementation, baseline runners, experiment scripts, and evaluation routines. The repository documents the required inputs and provides runnable entry points so that the reported comparisons can be checked directly. The paper specifies the search setting, initial repertoire rules, query budgets, and metrics. In each replay, a policy receives only profiles revealed by its previous queries; the full table is used by the evaluator to compute reference-front metrics. The white-box oracle serves only as an evaluation bound. Results are aggregated over datasets and five seeds, as indicated in the corresponding captions.

F.2 Scope, Limitations, and Future Directions

Scope. The study concerns sequential selection from a fixed sequence library after an initial campaign has measured a restricted part of its response landscape. Each additional measurement reveals the full profile across the specified objectives. The reported comparisons evaluate how well policies expand the observed Pareto frontier under the same query budget and initial repertoire rules.

Limitations. The empirical findings are grounded in the datasets, initial repertoires, and measurement budgets considered here. As with any benchmark study, the magnitude of the gains may vary with the landscape and experimental context. The results should therefore be interpreted as evidence for the evaluated search setting rather than as a uniform guarantee across all biological landscapes.

Future work. Future studies can examine the method across additional biological landscapes and prospective campaigns, and assess how its representation and query-allocation components adapt as evidence accumulates. Such studies would help characterize where the largest discovery gains arise and how well they transfer across experimental contexts.

F.3 Ethics, Societal Impact, and Responsible Use

Efficient sequence search may reduce the number of measurements needed to identify useful candidates. The same capability may also increase the pace of work on sequences with harmful properties. These retrospective results do not establish the safety or efficacy of any selected sequence. Applications in sensitive biological settings warrant appropriate biosafety review, human oversight, and independent experimental validation. Reporting uncertainty and failure cases alongside frontier gains supports informed assessment, especially when selection moves beyond previously observed sequence or response regions.