BEYOND THE INITIAL RANKING: HOW FEEDBACK SHAPES ACTIVE SEARCH

Anonymous authors
Paper under double-blind review

ABSTRACT

Active search aims to discover as many targets as possible within a limited measurement budget. Initial ranking quality captures only part of this task: observations change future choices, and planning acts through those changes. We separate initial predictions, feedback updates, and acquisition rules to evaluate their distinct roles in discovery. In a coherent probability model, we characterize when observing a fixed query has strictly positive value for a subsequent batch selection: no single continuation set is optimal across all possible outcomes. An exact correlated-group model expresses this value through dependence strength, selection margins, and the remaining capacity to exchange candidates. We then investigate these distinctions in retrospective protein-variant searches on Domainome and ProteinGym, with hits defined by within-pool top-5% fitness. In nested comparisons holding initial predictions and greedy acquisition fixed, adding sequence-position feedback improves discovery by approximately 0.4–1.0 hits at budgets of 12 and 24 measurements. A separate common-prior evaluation extends positive feedback gains to ridge and Gaussian-process updates on both corpora. Under its matched selected configurations, however, neither position-planning gains nor update-by-acquisition interactions pass multiplicity correction. Terminal-set comparisons distinguish changing query order, exchanging candidates, and improving yield. Together, these comparisons ask which relations carry useful feedback, whether candidate exchanges improve yield, and when planning adds value. Ranking quality, feedback utility, and planning value are distinct evaluation targets: a useful update can improve discovery without guaranteeing an additional benefit from planning.

1 INTRODUCTION

A protein-variant screen may contain thousands of candidates but permit only a few measurements. In our retrospective deep mutational scanning (DMS) setting, variants in the top 5% of measured fitness are hits, and an algorithm may reveal only $K = 12$ or 24 labels. Each observation can change the next query; the objective is the number of hits discovered within the budget. Active search therefore values a measurement both for its immediate reward and for what its label reveals about unmeasured candidates (Garnett et al., 2012; Jiang et al., 2017). But what makes this feedback useful?

Two greedy searches with identical initial scores can diverge after a label if one additionally updates substitutions at the same sequence position. Conversely, lookahead can change the next query yet ultimately measure the same candidates. Changing an action, changing terminal membership, and improving hit count are distinct events. Figure 1 shows the discovery comparison and its variation across inference units.

Our empirical intervention uses sequence-position membership, a known relation between single-substitution variants, rather than three-dimensional structure; the baseline already shares feedback through nearest neighbors in free-feature space. We add a position-level label update centered at the global hit-rate prior, zero before any observation, so every initial probability and the first greedy choice are preserved. Holding greedy acquisition fixed separates the starting ranking from the update, though position membership supplies extra prior information. The intervention evaluates the composite update; Section 3.2 specifies its sharing and calibration components.

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.

In a coherent probability model, a fixed query has strictly positive information value for a batch continuation exactly when no continuation set is optimal across all outcomes. An exact correlated-group model connects this value to dependence, selection margins and remaining budget. This conceptual model motivates the exchange analysis; its coherent-posterior assumptions and connection to the empirical scorer are explicit in Section 4.

On Domainome and ProteinGym, nested position-shared greedy updates add 0.40–0.97 hits over the feature-neighborhood baseline. A common-prior comparison extends feedback gains to ridge and Gaussian-process updates, while matched planning remains inconclusive. Terminal-set controls separate productive exchanges from order changes and forced coverage (Sections 5–7).

Our contribution is an evaluation framework that separates initial ranking, feedback updating, and acquisition, together with empirical evidence showing why these distinctions matter for discovery. The matched design makes feedback utility and planning value separately testable. We organize the evidence around three questions. *Relation*: which candidates should share feedback? *Exchange*: do changes in the measured candidates improve their yield? *Planning*: does lookahead add value under the supplied update?

Forced coverage changes membership without improving yield; position feedback improves discovery across update families; and historical planning interactions are not confirmed by the new matched tests. Component interventions locate conditional support for contingency on Domainome, leaving replication and its link to adaptive gains unresolved (Section 7.1). A frozen activation diagnostic also fails to improve on always using the known relation (Appendix H). These contrasts make ranking quality, feedback utility and planning value separately testable.

2 Related work

Bayesian active search maximizes discoveries under a budget (Garnett et al., 2012). ENS approximates nonmyopic search through a batch continuation (Jiang et al., 2017), while batch active search studies delayed feedback (Jiang et al., 2018). Knowledge-gradient policies value observations through changed decisions (Frazier et al., 2009). Multifidelity active search extends planning to queries with different information quality and costs (Hernández-García et al., 2024); imitation learning amortizes the computation of nonmyopic policies (Nguyen et al., 2024), and diffusion-based search samples action sequences while accounting for costs (Banerjee & Schneider, 2026). We hold ENS fixed to test how planning value depends on its feedback update.

Cluster-based and hierarchical batch search use related candidates to guide sampling (Renders, 2018; Myers & Greenside, 2020), and position-level sharing is established in DMS estimation (Rao et al., 2024). Protein fitness models combine assay measurements with evolutionary information and structured kernels (Hsu et al., 2022; Groth et al., 2024). Gaussian-process Bayesian optimization provides another route to sequential protein design (Benjamins et al., 2024), while coevolution-informed optimization incorporates evolutionary dependencies (Kulathunga et al., 2026). Our known sequence-position partition enables matched feedback comparisons, and our site-kernel GP probes estimator dependence. Our objective counts top-tail discoveries rather than maximizing the best fitness found.

In low-budget protein engineering, CIDER combines conformal top-tail calibration, information-directed acquisition, and diversity-aware batches, with a primary objective of finding at least one top-1% candidate (Bhattacharya, 2026). Our objective counts the targets recovered. Diversity-aware active search can also encode diminishing returns in the reward (Nguyen & Garnett, 2023); our quota control retains the candidate-count reward to test coverage as an explanation for gains. Experimental active learning for far-red fluorescent-protein evolution illustrates prospective laboratory feedback (Brown et al., 2026), complementing retrospective replay. Our controlled decomposition isolates feedback utility at matched initial rankings, links terminal-set exchanges to discovery gains, and evaluates planning conditional on the feedback model.

3 Separating ranking, feedback, and acquisition

3.1 Finite-pool active search

A pool $\mathcal{X} = \{1, \dots, n\}$ contains unknown binary labels $y_i \in \{0, 1\}$. At step t, the history is $D_t = \{(i_\tau, y_{i_\tau})\}_{\tau=1}^{t}$ and the unmeasured set is $R_t = \mathcal{X} \setminus \{i_1, \dots, i_t\}$. Each query reveals one label

and costs one unit. The objective is the terminal reward

$$H_K = \sum_{t=1}^{K} y_{i_t}. \tag{1}$$

The pool also provides free features x_i and a known partition $g(i)$. Here $g(i)$ is a sequence position, so several substitutions share a group. Group membership is available before search, whereas its hit rate is not. The objective in Equation 1 counts candidates, not distinct positions.

We represent a search procedure by initial scores $p_i \in (0,1)$, a feedback update $U : D_t \mapsto (r_i(D_t))_{i \in R_t}$, and an acquisition rule acting on the resulting scores. A static policy ranks p once. A greedy policy selects the largest current r_i. A lookahead policy uses hypothetical labels and the same update U to value future choices. Holding the initial ranking and greedy rule fixed isolates changes in the update.

3.2 Feature-neighborhood and position-shared updates

The baseline follows the local probability model used in active search (Garnett et al., 2012; Jiang et al., 2017), with a candidate-specific prior and an explicit prior weight. Let L_t be the measured indices, and $N(i)$ the fixed nearest neighbors of candidate i in free-feature space. The baseline score is

$$q_i(D_t) = \frac{\gamma p_i + \sum_{j \in L_t \cap N(i)} y_j}{\gamma + |L_t \cap N(i)|}, \qquad s_i(D_t) = \operatorname{logit} q_i(D_t). \tag{2}$$

Thus the baseline already shares feedback: observing j updates every candidate whose neighbor list contains j. We use $|N(i)| = 50$ and $\gamma = 25$ in the headline comparison.

For each position h, let $n_h(t)$ be the number of measured substitutions and $z_h(t)$ their hit count. With base rate π_0 and prior strength κ, define

$$\theta_h(D_t) = \frac{\kappa\pi_0 + z_h(t)}{\kappa + n_h(t)}, \qquad s_i^+(D_t) = s_i(D_t) + \lambda \left[\operatorname{logit} \theta_{g(i)}(D_t) - \operatorname{logit} \pi_0\right]. \tag{3}$$

Acquisition uses $r_i^+ = \sigma(s_i^+)$, with logistic σ. Hits and misses raise or lower a position's scores relative to π_0; $\lambda = 0$ recovers the baseline.

A common starting point. At empty history, $n_h = z_h = 0$ and $\theta_h = \pi_0$, so $q_i(\varnothing) = p_i$ and $s_i^+(\varnothing) = \operatorname{logit} p_i$. With common ties, both greedy arms have the same first action. The position arm adds a known relation through a composite update combining sharing topology, shrinkage, log-odds scaling and later calibration; these components are evaluated together. Neighborhood lists may already connect same-position substitutions, allowing overlap between the two sharing channels. This composite scoring update need not define a coherent joint Bayesian posterior.

3.3 Changing acquisition while retaining the update

We compare greedy with an efficient nonmyopic search (ENS) approximation (Jiang et al., 2017). For $b = K - t$ queries remaining, let $T_m(v; S)$ denote the sum of the m largest scores in S, with $T_0 = 0$. The lookahead score of candidate i is

$$a_b(i; D_t) = r_i(D_t) + \sum_{y \in \{0,1\}} w_y(i, D_t)\, T_{b-1}(r(D_t \cup \{(i,y)\}); R_t \setminus \{i\}), \tag{4}$$

where $w_1 = r_i(D_t)$ and $w_0 = 1 - r_i(D_t)$. The same update handles hypothetical and realized labels. A one-observation batch continuation approximates the full decision tree; here the weights are plug-in scores, not demonstrated calibrated label probabilities, and supply no joint-posterior optimality guarantee. Candidate actions are evaluated among the current top 100 scores while the continuation sum uses all remaining candidates (Appendix D). Unlike the greedy comparison, the two lookahead arms need not choose the same first candidate despite identical initial scores. A quota arm, which keeps the baseline scores but selects the highest-scoring candidate among the least-measured positions, separates sharing from forced coverage; it is one specific diversity control, and all arms keep the same candidate-count objective.

4 WHEN FEEDBACK CHANGES THE CONTINUATION SET

Which responses to a label can improve the subsequent discoveries? We make this question precise in a coherent probability model. The analysis concerns one fixed query followed by a batch of m selections, the continuation used by ENS (Jiang et al., 2017); it is not a guarantee for the heuristic update in Equation 3 or for a fully adaptive policy.

Condition on the current history, fix query i, and write U for the candidates remaining after that query. Let p_U contain their conditional hit probabilities and $p_U^{(i,y)}$ their probabilities after observing $Y_i = y$. The value of observing the label before committing the continuation batch is

$$J_i = \mathbb{E}\Big[T_m\Big(p_U^{(i,Y_i)}; U\Big)\Big] - T_m(p_U; U). \tag{5}$$

The query's immediate reward cancels because both alternatives make the same query. By the posterior martingale property, $J_i \geq 0$, with equality exactly when some m-candidate set is optimal under every observation outcome of positive probability. Different returned sets under ties need not imply positive value. This is a standard value-of-information argument; Appendix F gives the proof.

An exact group model. Suppose the queried candidate and its r remaining group members have marginal hit probability $p \in (0, 1)$ and exchangeable pairwise correlation $\rho \in [0, 1]$. Their common posterior probabilities after a hit and miss are

$$q_1 = p + \rho(1 - p), \qquad q_0 = p(1 - \rho). \tag{6}$$

A coherent construction with these marginals exists for every such ρ (Appendix F). The n outside candidates retain probabilities $c_1 \geq \cdots \geq c_n$ after either outcome. For $0 \leq m \leq r + n$, the feasible number of selected group members ranges from $\ell = \max(0, m - n)$ to $u = \min(r, m)$.

Proposition 1 (Value at replaceable selection boundaries)**.** *In this model,*

$$J_i = \sum_{k=\ell}^{u-1} h(c_{m-k}), \qquad h(c) = \min\{p(q_1 - c)_+,\, (1 - p)(c - q_0)_+\}\,, \tag{7}$$

where $(a)_+ = \max(a, 0)$ *and an empty sum is zero. Thus* $J_i > 0$ *exactly when at least one replaceable outside probability lies strictly between* q_0 *and* q_1*.*

Each summand is a possible exchange of an outside candidate for a group member, so an update creates value only if it reaches one of these boundaries; changes confined within the selected or the unselected candidates contribute nothing.

Misspecified sharing. If a scorer assumes the wrong dependence, its assumed correlation determines which boundaries it crosses, while the true dependence determines the reward of those crossings (Proposition 2, Appendix F.5). Crossing a score boundary alone therefore does not certify a discovery gain.

From the model to the experiments. The model motivates comparing identical initial scores, intervening on the sharing relation, and counting acquired sets and hits. These test whether score changes alter membership and yield. The coherent model supplies a conceptual account of productive exchanges; the empirical interventions evaluate score-based updates, with latent correlation and calibrated information value left unestimated.

5 RELATION: WHICH FEEDBACK IMPROVES DISCOVERY?

5.1 EVALUATION PROTOCOL

We replay searches on Domainome 1.0 (Beltran et al., 2025) and ProteinGym v1.3 (Notin et al., 2023), admitting 245 pools in 67 sequence clusters and 161 pools in 97 studies respectively. A hit is a variant in the within-pool top 5% of measured fitness, the headline budgets are $K \in \{12, 24\}$, and each algorithm sees labels only when it queries their candidates. These are retrospective evaluations, not prospective laboratory experiments.

Table 1: Nested greedy comparison with identical initial predictions. Calibration excludes outer and inner validation units. Means weight each cluster or study equally; SE uses paired unit differences. Holm p adjusts the two budgets within each corpus. Historical planner comparisons use a separate protocol.

	Domainome		ProteinGym	
Method	$K = 12$	$K = 24$	$K = 12$	$K = 24$
Greedy (kNN)	3.3296	6.1019	1.4822	2.8589
Greedy (position)	3.7510	7.0677	1.8787	3.7516
Position update gain	$+0.4214$	$+0.9658$	$+0.3965$	$+0.8927$
Paired SE	0.1562	0.2369	0.1088	0.2060
Holm p	0.00758	0.00016	0.00016	0.00008
Inference units	67	67	97	97
Pools	245	245	161	161

The matched arms share precomputed ProteinMPNN log-likelihood-ratio scores (Dauparas et al., 2022) and a structure-confidence feature, fixed before each search; "free" means outside the label budget, not computationally costless. The initial prior is a calibrated transform of the sequence score and both channels define feature neighborhoods (Appendix A).

Assays are searched separately, then averaged within 67 Domainome clusters or 97 ProteinGym studies with equal unit weights. Paired sign-flip tests use these units and named Holm families. The primary greedy comparison uses exact nested refitting, excluding both outer and inner validation units from calibration. This closes the historical indirect outer-label path and changes the four gains by at most 0.0174 hits. Historical planners were not nested-refitted; fixed-grid interactions and the new common-prior comparisons are separate protocols (Appendix A.2). Benchmark reuse and residual biological dependence remain relevant to all comparisons.

Evidence map. The comparisons answer complementary questions; their estimates should not be pooled.

Protocol	Question and inference units	Location
Nested primary	Matched greedy gain; 67 clusters / 97 studies	Table 1; Fig. 1a–c
Unified	Cross-model feedback and matched planning; 67 / 83 connected units	Table 2; Fig. 3a,b
Historical	Exchanges and fixed-grid interaction; 67 / 97 units	Figs. 2, 3c
Mechanistic	Component and allocation interventions; 67 / 97 units	Section 7.1
Descriptive	SKEMPI affinity extension; four related contexts	Appendix G.5

5.2 Position feedback adds value beyond a static ranking

Holding greedy acquisition fixed, adding the position update increases hits in all four headline cells (Table 1). The paired increases are 0.4214 and 0.9658 on Domainome, and 0.3965 and 0.8927 on ProteinGym. Table 1 gives paired standard errors and adjusted tests for this nested comparison; each corpus retains its two-budget correction family. The arms share initial scores, so the separation arises after labels arrive rather than from a different ranking.

Static-ranking comparisons. Historical position greedy increases $K = 24$ yield over matched static ranking by $16.6\%/30.1\%$ on Domainome/ProteinGym (Table 9); this measures hits per query, not turnaround savings. On the 161 ProteinGym pools, it also exceeds the preselected five-checkpoint ESM-1v static ensemble (Meier et al., 2021) at both budgets (Figure 1d; Table 12). Unlike the matched comparisons, this compares complete strategies with different initial predictors; no ESM-1v result is inferred for Domainome.

Who benefits? The nested $K = 24$ comparison retains 15 losses, 7 ties and 45 gains among Domainome clusters, and 18 losses, 33 ties and 46 gains among ProteinGym studies (Figure 1c).



Figure 1: **Discovery gains beyond the initial ranking, with their heterogeneity retained.** (a,b) Nested $K = 12, 24$ outcomes for every pool, with smoothed half-violin densities. Small points show 245 Domainome pools and 161 ProteinGym assays per arm and budget; large markers and 95% intervals retain equal-cluster/study means and whole-unit bootstrap uncertainty (67/97 units), not pool-independent inference. (c) Every unit's nested position-minus-neighborhood gain at $K = 24$, including losses and ties. (d) Historical ProteinGym discovery strategies, including the preselected ESM-1v static ranker; these use different initial predictors and are separated from the matched nested panels. Pools are averaged within each cluster or study, then units receive equal weight.

The largest loss is three hits in either corpus. These realized differences do not supply a validated activation rule; negative-unit details and posthoc dependence sensitivity are in Appendix A.1.

Transfer and sensitivity. ESM-1v preserves update gains, but not a corrected acquisition interaction (Appendix I); development-exposed MaveDB also gives positive transfer (Appendix G.4). A frozen SKEMPI affinity task yields descriptive feedback gains of $0.50/1.50$ hits at $K = 12$ under two priors, without added lookahead hits. Its four contexts share one ovomucoid study. Appendix G.5 retains all SKEMPI outcomes; prior development exposure precludes untouched validation.

5.3 Which candidates should share a label?

We replace position membership by two alternatives while keeping the update formula: random groups with the same size multiset, and contiguous groups formed by sorting the free sequence score. Both use the same extended grid, including $\lambda = 0$ so an unhelpful term can switch off, and ask whether arbitrary or score-based grouping reproduces the position effect.

Table 6 shows that neither control reproduces the position gains, and the position-versus-score-group contrast is positive in all four cells; the weakest, Domainome at $K = 12$, passes the within-corpus correction but not the wider cross-corpus family. This supports these specific relational comparisons, not superiority over every feature-based predictor: the score-group control matches one-dimensional score continuity, not the full feature geometry. These controls test the composite update; the next comparison changes the estimator itself.

Table 2: **Feedback gains across update models.** Common-prior $K = 24$ evaluation; means weight the 67 Domainome or 83 ProteinGym components equally. On/off denote availability of the position term. Intervals are pointwise bootstrap 95% intervals; Holm p covers all six contrasts. This protocol differs from Table 1.

Corpus	Update	Off	On	Gain [95% interval]	Holm p
Domainome	Composite	6.519	6.978	$+0.460$ $[-0.003, 0.960]$	0.06535
	Ridge	6.091	7.061	$+0.971$ $[0.351, 1.606]$	0.00585
	GP	6.218	6.972	$+0.754$ $[0.128, 1.416]$	0.04800
ProteinGym	Composite	2.848	3.852	$+1.004$ $[0.642, 1.412]$	0.00030
	Ridge	2.705	4.110	$+1.405$ $[0.959, 1.881]$	0.00030
	GP	2.398	3.444	$+1.046$ $[0.591, 1.503]$	0.00040

5.4 Does feedback utility extend across update models?

We evaluate all 406 pools at $K = 24$ with one fixed initial prior and greedy acquisition, turning the position term on or off within the composite, ridge and Gaussian-process (GP) updates. Every arm starts with the same probabilities. The prior uses fixed slope 1 and mean 0.05, without fitting labels; tuning excludes the full held-out component and weights training components equally. Domainome retains 67 clusters; linking ProteinGym studies that share a protein gives 83 connected components rather than 97 studies. This is a new common-prior comparison on the developed corpora, not a rerun of Table 1's learned-calibration protocol.

Position-enabled ridge and GP gain 0.75–1.40 hits, with all four comparisons passing the prespecified six-test Holm family (Table 2). Domainome GP is close to the threshold ($p = 0.0480$), and the composite update on Domainome does not pass ($p = 0.0653$). The gains therefore extend beyond one update formula without establishing equal strength across every setting. The on/off arms select their hyperparameters separately: this tests specified relation-enabled tuning procedures, not matched capacity or calibration. Appendix K retains grids, matched-nuisance sensitivities and selection limits.

The control comparisons and their inference boundaries are collected in Table 8.

6 Exchange: do changed candidates yield more hits?

An update benefit alone does not distinguish query-order changes from productive terminal-set exchanges.

For equal-budget trajectories, shared candidates cancel: the hit difference is the number of incoming hits minus outgoing hits. This retrospective diagnostic separates membership changes from their quality (Equation 10). Neighborhood ENS often changes its next action but replaces only 0.74%–5.95% of the greedy terminal set, with gains near zero. Position feedback replaces 20.44%–30.95% while increasing hits. Quota replaces 18.38%–52.21%, yet loses 0.6924 and 1.5371 hits on Domainome and gains nothing clear on ProteinGym (Figure 2). Thus changing order, changing membership and changing reward are distinct. The terminal-set identity locates each realized gain in the candidates exchanged. Replacement-prediction and spatial controls are in Appendices G.1–G.3.

7 Planning: when does lookahead add value?

Productive greedy feedback does not imply additional lookahead value. We test this distinction using the composite update under the unified protocol of Section 5.4. For each held-out component, greedy training reward selects the update configuration; ENS and greedy then share that configuration. We measure position ENS minus position greedy and the difference of planning increments,

$$I(u) = \big[\mathrm{ENS}_{\mathrm{pos}}(u) - \mathrm{greedy}_{\mathrm{pos}}(u)\big] - \big[\mathrm{ENS}_{\mathrm{knn}}(u) - \mathrm{greedy}_{\mathrm{knn}}(u)\big]. \tag{8}$$

Position-planning gains are $+0.238$ and $+0.591$ hits on Domainome and ProteinGym; interactions are $+0.335$ and $+0.556$. None passes the four-test Holm correction ($p = 0.2260, 0.0898$ for gains;

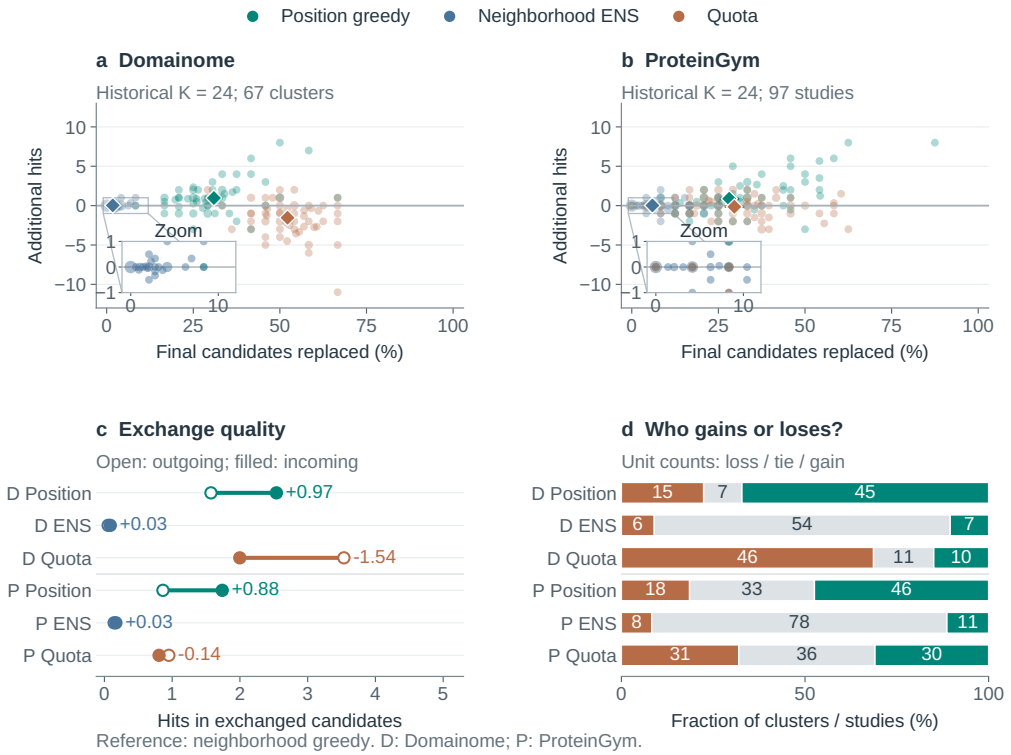

Figure 2: **Productive exchanges, not membership changes alone.** Historical $K = 24$ comparisons against neighborhood greedy. (a,b) Replacement fractions and hit gains for every cluster/study; diamonds mark strategy means and insets magnify the marked near-origin region. Coincident points are sized by unit count. (c) Mean hits entering and leaving the terminal set; their difference equals the net gain. (d) Fractions of units losing, tying or gaining hits. Position greedy, neighborhood ENS and quota represent different ways of changing search behavior under a common reward. All summaries weight units equally; bars are descriptive pointwise 95% whole-unit bootstrap intervals. Shared candidates cancel from the exchange identity.

$0.2260, 0.0955$ for interactions). Positive estimates leave useful effects plausible, but do not establish them under this protocol. Ridge and GP remain greedy-only because their regression means are not binary predictive probabilities (Appendix K).

Historically, neighborhood ENS gains were near zero (-0.0074 to $+0.0319$ hits). Averaging Equation 8 uniformly across the declared (κ, λ) grid yields positive interactions in all four corpus–budget cells and $137/144$ grid cells; three aggregate tests pass correction (Figure 3c). This development-informed grid average differs from the unified selected-configuration comparison; the new tests do not confirm a general planning advantage. Reweighting checks and neighborhood sweeps remain in Appendices C.1 and D.

7.1 What explains planning value?

For query i, let U_i exclude that query, m be the continuation budget and $\bar{r}_i = r_i r_{U_i}^{(i,1)} + (1-r_i) r_{U_i}^{(i,0)}$. The full score is

$$B_i = r_i + T_m(r_{U_i}), \quad C_i = r_i T_m(r_{U_i}^{(i,1)}) + (1-r_i) T_m(r_{U_i}^{(i,0)}) - T_m(\bar{r}_i),$$
$$D_i = T_m(\bar{r}_i) - T_m(r_{U_i}), \qquad a_i = B_i + C_i + D_i. \tag{9}$$

B is baseline opportunity, $C \geq 0$ outcome-contingent selection and D mean-score drift. Coherence implies $D = 0$, but the converse fails; heuristic C need not be calibrated information value (Appendix F.6). Removing C while retaining the feedback update costs more under position feedback

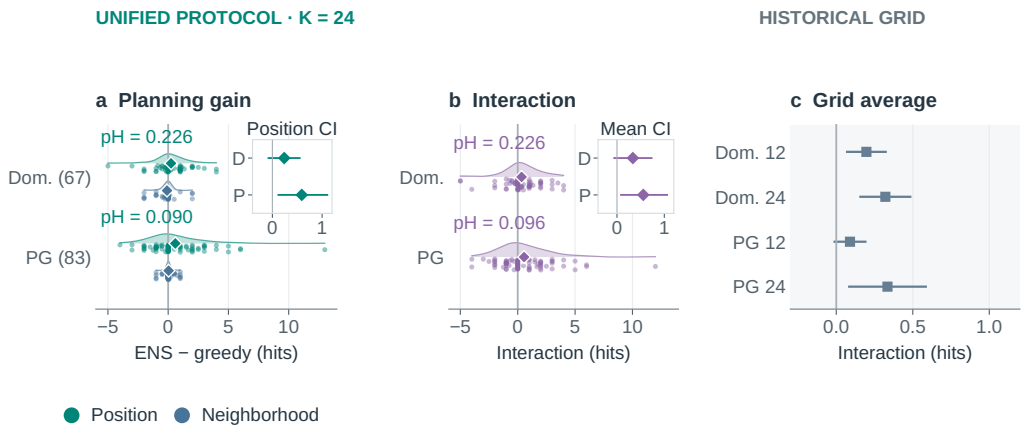

Figure 3: **Planning value requires evidence separate from feedback utility.** (a) Unified $K = 24$ ENS-minus-greedy gains under each update. (b) Unified update-by-acquisition interactions. Small points and densities show every cluster/component $(67/83)$; large markers retain the original means. Insets enlarge position-gain and interaction means with Domainome above ProteinGym. Bars are pointwise whole-unit bootstrap 95% intervals; all four primary tests (position gains and interactions) miss their joint Holm threshold. Neighborhood gains are descriptive references. (c) Historical grid-average interactions at two budgets, a different estimand with approximate pointwise normal intervals. Its three corrected detections do not constitute selected-policy validation under the new protocol. Full grids remain in Figure 7.

on Domainome at $K = 24$: interaction $+0.558$ hits, 95% interval $[0.182, 0.971]$, Holm $p = 0.0218$. ProteinGym's $+0.265$ interval spans zero; neither drift-removal interaction nor any $K = 12$ component interaction passes correction (Table 14). Support for contingency is therefore conditional. Two further interventions compare branch attenuation against a scalar-matched bonus, and an outcome-dependent fixed-query batch against its mean-branch batch. All four primary tests miss joint Holm correction: adaptive effects are $-0.051/+0.099$ hits (Domainome/ProteinGym; both $p = 0.6194$), and fixed-query gains $+0.357/+0.163$ (both $p = 0.0572$). Thus the link from contingency allocation to adaptive discovery remains unresolved. Appendix J and Figure 8 retain protocols, manipulation checks and all outcomes.

8 Discussion and Conclusion

Ranking quality, feedback utility, and planning value are distinct achievements. Our matched evaluation makes them separately testable. The nested comparison establishes 0.40–0.97 additional hits from the composite position update. Terminal-set analysis separates changed order from productive exchanges. A common-prior evaluation extends feedback gains to ridge and GP updates, while matched planning tests remain inconclusive. These results show why a good initial ranking, useful feedback and added lookahead value should be evaluated as distinct properties of a search system.

Component interventions probe the planning interaction. Contingency removal supports an effect on Domainome, while replication on ProteinGym and the link from contingency allocation to adaptive gain remain open.

Primary evidence covers retrospective single-substitution DMS at constructed 5% prevalence. The SKEMPI affinity extension is descriptive, with one shared scaffold/study and inconsistent planning increments (Appendix G.5). Unknown prevalence, measurement noise, delayed or batched feedback, and laboratory turnaround remain untested. Residual dependence and benchmark reuse limit inference; MaveDB is development-exposed. No validated rule identifies which pools should activate position feedback. ESM-1v's acquisition interaction and rarer-hit gains remain uncertain (Appendix D).

REPRODUCIBILITY STATEMENT

Appendix A specifies the replay protocol, parameter grids, statistical units, and calibration limits. Appendix B gives the episode algorithm, and Appendix E summarizes artifact provenance; the source README maps comparisons to their producing experiments. Appendix H reports the frozen fixed-prefix replay, all diagnostic selection results, and the distinction between assay-level search and study-level inference. Appendix I specifies the preselected external static-ranking comparison. Appendix F.6 specifies the complete-cohort component interventions and independent finite-world formula checks, including their frozen settings and multiplicity families. Appendix G.5 adds frozen SKEMPI affinity-task replays with source-bound structures, native probability arrays, complete paid-query traces and independently checked outcomes. The source includes recorded inputs and asset provenance. For Table 1, a separate replay bundle contains all frozen candidate pools and features, pinned code and dependencies, calibration references and resumable per-pool outputs. It reruns calibration, selection and greedy episodes and reproduces all archived aggregate statistics (Appendix A.1); figure regeneration alone does not do so. Upstream feature generation and historical planner episodes are not reproduced by this primary replay bundle.

AI USE STATEMENT

Generative AI assisted with experiment implementation, analysis, literature checking, manuscript drafting, figure and table code, and internal critical review under human direction. The manuscript reanalyzes existing experimental evidence, includes a frozen replay on the same pools, evaluates a preselected static ranker, and includes exact calculations in a specified probabilistic model; its reported protein measurements come from recorded runs, not language-model predictions. The authors retain responsibility for the scientific claims, citations, evaluation limitations, and final submission.

REFERENCES

Arundhati Banerjee and Jeff Schneider. Cost-aware diffusion active search, 2026.

Antoni Beltran, Xiang'er Jiang, Yue Shen, and Ben Lehner. Site-saturation mutagenesis of 500 human protein domains. *Nature*, 637(8047):885–894, 2025.

Carolin Benjamins, Shikha Surana, Oliver Bent, Marius Lindauer, and Paul Duckworth. Bayesian optimization for protein sequence design: Back to simplicity with Gaussian processes. In *AI for Accelerated Materials Design - NeurIPS Workshop 2024*, 2024.

Rishabh Bhattacharya. CIDER: Conformal information-directed agents for low-budget protein engineering. In *ICML Workshop on Generative and Agentic AI for Biology (GenBio)*, 2026.

Daniel V. Brown, Ryan S. Cross, Shiying Zhu, Thomas Hill, Chhon Ling Sok, Misty R. Jenkins, Marija Dramicanin, and Rory Bowden. Active learning enables efficient directed evolution of a far-red fluorescent protein with minimal experimental data. *bioRxiv*, 2026.

J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022.

Peter Frazier, Warren Powell, and Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4):599–613, 2009.

Roman Garnett, Yamuna Krishnamurthy, Xuehan Xiong, Jeff G. Schneider, and Richard P. Mann. Bayesian optimal active search and surveying. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*, 2012.

Peter Mørch Groth, Mads Herbert Kerrn, Lars Olsen, Jesper Salomon, and Wouter Boomsma. Kermut: Composite kernel regression for protein variant effects. In *Advances in Neural Information Processing Systems*, volume 37, pp. 29514–29565, 2024.

Alex Hernández-García, Nikita Saxena, Moksh Jain, Cheng-Hao Liu, and Yoshua Bengio. Multi-fidelity active learning with gflownets. *Trans. Mach. Learn. Res.*, 2024.

Chloe Hsu, Hunter Nisonoff, Clara Fannjiang, and Jennifer Listgarten. Learning protein fitness models from evolutionary and assay-labeled data. *Nature Biotechnology*, 40(7):1114–1122, 2022.

Justina Jankauskaitė, Brian Jiménez-García, Justas Dapkūnas, Juan Fernández-Recio, and Iain H. Moal. SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, 35(3):462–469, 2019.

Shali Jiang, Gustavo Malkomes, Geoff Converse, Alyssa Shofner, Benjamin Moseley, and Roman Garnett. Efficient nonmyopic active search. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1714–1723, 2017.

Shali Jiang, Gustavo Malkomes, Matthew Abbott, Benjamin Moseley, and Roman Garnett. Efficient nonmyopic batch active search. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Dulitha P. Kulathunga, Divyanshu Shukla, and Davit A. Potoyan. Coevolution-informed Bayesian optimization for sample-efficient protein design. *bioRxiv*, 2026.

Stephen M. Lu, Wuyuan Lu, M. A. Qasim, Stephen Anderson, Izydor Apostol, Wojciech Ardelt, Theresa Bigler, Yi Wen Chiang, James Cook, Michael N. G. James, Ikunoshin Kato, Clyde Kelly, William Kohr, Tomoko Komiyama, Tiao-Yin Lin, Michio Ogawa, Jacek Otlewski, Soon-Jae Park, Sabiha Qasim, Michael Ranjbar, Misao Tashiro, Nicholas Warne, Harry Whatley, Anna Wieczorek, Maciej Wieczorek, Tadeusz Wilusz, Richard Wynn, Wenlei Zhang, and Michael Laskowski, Jr. Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases. *Proceedings of the National Academy of Sciences*, 98 (4):1410–1415, 2001.

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, 2021.

Vivek Myers and Peyton Greenside. A hierarchical approach to scaling batch active search over structured data. *arXiv preprint arXiv:2007.10263*, 2020.

Quan Nguyen and Roman Garnett. Nonmyopic multiclass active search with diminishing returns for diverse discovery. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *PMLR*, pp. 5231–5249, 2023.

Quan Nguyen, Anindya Sarkar, and Roman Garnett. Amortized nonmyopic active search via deep imitation learning, 2024.

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 *Advances in Neural Information Processing Systems*, volume 36, pp. 64331–64379, 2023.

Jingyou Rao, Ruiqi Xin, Christian Macdonald, Matthew K. Howard, Gabriella O. Estevam, Sook Wah Yee, Mingsen Wang, James S. Fraser, Willow Coyote-Maestas, and Harold Pimentel. Rosace: a robust deep mutational scanning analysis framework employing position and mean-variance shrinkage. *Genome Biology*, 25, 2024.

Jean-Michel Renders. Active search for high recall: a non-stationary extension of Thompson sampling. In *Advances in Information Retrieval*, pp. 722–728, 2018.

A Replay protocol and evaluation boundaries

The main evaluation protocol map separates nested primary, common-prior, historical, mechanistic and descriptive evidence. The following details specify the original replay.

Candidates and labels. Upstream corpus admission requires at least 960 scored single substitutions, at least 48 hits at the nominal 5% threshold, and a realized hit fraction no greater than twice nominal. The subsequent feature join retains finite rows, removes pools with fewer than 200 candidates, and recomputes hits as the top $\mathrm{round}(0.05n)$ remaining scores. Position keys, at least two positions, and the precomputed feature channels must be available. No admission rule uses a policy's performance. Labels derive from measured fitness already present in the dataset; each episode masks all labels except those queried by that policy. The hit threshold is a benchmark convention, and its known prevalence is extra information that need not be available in deployment. A separately prepared corpus with unresolved assay direction was excluded from all scientific conclusions, regardless of effect sign.

Free features and initial probabilities. The neighborhood model uses within-pool standardized ProteinMPNN log-likelihood ratios and a precomputed pLDDT confidence channel. They are held fixed across arms. For a standardized sequence score z_i, the calibrated prior is $p_i = \sigma(a + bz_i)$, with the intercept a chosen numerically so the pool-average prior equals 0.05. The implementation clips extreme probabilities for numerical stability. In the historical experiment, the slope for held-out unit u is eight times the mean sequence-score/label correlation across the other pools. The position prior π_0 is the mean realized hit fraction in those other pools. These are the historical experiment definitions, not logistic maximum-likelihood estimates. The archived features determine the reported comparison; the exact structure-predictor checkpoint cannot be established from a model-family citation alone. Consequently, it does not claim to reproduce upstream feature generation from the original measurements.

Parameters and validation. The position grid used in both historical and nested greedy selection is

$$\kappa \in \{2, 4, 6, 8, 12, 16, 24, 32, 48\}, \qquad \lambda \in \{0.25, 0.375, 0.5, 1\}.$$

Each outer unit is evaluated using the pair selected on all other units; the selection objective averages their pools rather than reweighting their units. The baseline fixes $\lambda = 0$, making κ irrelevant. For Domainome the modal greedy configurations are $(6, 0.375)$ at $K = 12$ and $(2, 0.375)$ at $K = 24$; for ProteinGym they are $(2, 0.375)$ and $(2, 0.5)$ under the nested refit (the historical $K = 24$ mode was $(4, 0.5)$). Some choices lie on the grid boundary. The extended control grid adds $\kappa \in \{0.5, 1, 96\}$ and $\lambda \in \{0, 1.5, 2, 3\}$. Extended-grid controls are not silently substituted into the headline table.

Independence and benchmark reuse. Domainome clusters join shared Pfam families, UniProt accessions or loader sequence-identity links; ProteinGym units are studies. Related units and corpus sequence segments may overlap. Both corpora were development-exposed; most sensitivity analyses are post hoc. Earlier-file exposure counts use a larger cohort and are not quoted for this denominator. Adjusted tests assume the declared units and sign symmetry; they do not correct development history or establish untouched, independent or prospective validation.

A.1 Exact nested heterogeneity and replay artifacts

Complete unit distributions. A faithful rerun retains the original calibration, ordered parameter grid, deterministic ties, pool-weighted selection, unit-weighted reporting and two-budget Holm family within each corpus. All 96 aggregate fields across historical and nested modes reproduce the archived means, effects, MDEs and raw/adjusted tests to absolute tolerance 10^{-10}; elapsed runtime is excluded from this comparison. The rerun additionally saves each pool's final and training results, selected cells, calibration exclusions and unit vectors. Figure 1c uses these exact nested outputs.

Domainome at $K = 12$ has $12/25/30$ losses/ties/gains; Domainome at $K = 24$ has $15/7/45$ losses/ties/gains; ProteinGym at $K = 12$ has $9/61/27$ losses/ties/gains; ProteinGym at $K = 24$ has $18/33/46$ losses/ties/gains. These are complete descriptive sign counts, not additional significance tests.

Largest negative units. Table 3 selects the three lowest $K = 24$ unit gains in each corpus, breaking equal-gain ties by unit identifier. All six happen to be single-pool units, so these losses are not created by averaging a large multi-assay study. Both arms find hits in every listed unit, but the position update finds fewer. Their selected (κ, λ) values are $(2, 0.375)$ for the three Domainome units,

Table 3: Three lowest-gain units per corpus under the exact nested $K = 24$ protocol of Table 1. Units are sorted by gain, then lexical identifier to resolve ties. Hits are means across constituent pools; gain is position minus neighborhood. Selection is descriptive and does not define a predictive rule.

Corpus	Unit	Pools	Neighbor	Position	Gain
Domainome	P61024_PF01111_1	1	8.00	5.00	-3.00
Domainome	Q9UBF6_PF12678_48	1	7.00	4.00	-3.00
Domainome	P41002_PF00646_28	1	7.00	5.00	-2.00
ProteinGym	Firnberg_2014	1	7.00	4.00	-3.00
ProteinGym	Gersing_2022	1	6.00	4.00	-2.00
ProteinGym	Loggerenberg_2023	1	4.00	2.00	-2.00

and $(2, 0.5)$, $(4, 0.5)$ and $(2, 0.5)$ for the listed ProteinGym units. These outcome-selected examples characterize observed failures, not a biological cause or a rule for recognizing failures before search. The frozen diagnostic in Appendix H does not validate an activation policy.

Dependence sensitivity of the primary comparison. Joining ProteinGym studies transitively through verified shared UniProt identifiers merges 97 studies into 83 components. We first average the original nested study effects within each component, then weight components equally. The posthoc means are $+0.400$ hits at $K = 12$ (normal 95% interval $[0.183, 0.617]$) and $+0.894$ at $K = 24$ ($[0.491, 1.296]$). Intervals use the standard error across component means and are descriptive, not new primary tests. Regrouping changes the weighting and retains the original fitted predictions; it does not refit calibration or selection with entire components excluded. Shared-protein components are not protein families and need not be independent. Domainome already merges domains through Pfam, accession and sequence-identity links; we introduce no new claim of independence for its 67 clusters. Neither check repairs benchmark-development dependence.

Reproduction from fixed features. The replay bundle includes all 406 frozen input pools, their ordered variant keys, fitness and labels, standardized features, an independent calibration reference, source dependency closure, pinned numerical packages and a command to rerun both greedy arms. Per-pool content-bound caches preserve completed work and refuse changed inputs, protocols or environments. Unit vectors reconstruct Table 1; a separate comparison script checks all archived aggregate fields, including corrected tests. This reproduces calibration, selection and scientific episodes from fixed inputs. It does not regenerate upstream protein-model features or establish an unrecorded checkpoint identity.

A.2 Calibration sensitivity

In the historical protocol, the leave-one-unit-out slope is computed once for every unit. When a candidate configuration is evaluated on the remaining units during model selection, their precomputed slopes can contain labels from the outer unit. This is an indirect information path through selection despite the target unit's own slope excluding it. Fixed-slope checks remove that particular slope-fitting path; they do not retroactively change the headline run. Their constants were chosen after examining fitted slopes, and other development choices remain historical. Table 4 reports every retained corpus and slope.

The exact nested refit. The path can be closed rather than estimated. For every ordered pair of units (u, w), refit the slope over the pools in neither, so the training mean that selects unit u's configuration reads no label from u through any route. The refit reported here covers the two greedy arms only; the historical planner comparisons were not recomputed under this nested selection protocol.

Two controls run before any refitted number is read, and the procedure refuses rather than reports if either fails. First, the reimplemented single-exclusion slope must reproduce the shipped one on every unit to 10^{-12}. Second, the harness in legacy mode must reproduce the published rounded effect sizes to 1.5×10^{-4}; until it does, a refitted-minus-legacy difference would be an artifact of reimplementation rather than the path under study. Both passed.

Table 4: Position–Greedy minus kNN–Greedy with fixed calibration slopes. This is a sensitivity analysis, not a new confirmatory evaluation; slope choices followed inspection of historical fitted ranges. Holm correction covers six tests per corpus.

Corpus	Fixed slope	K	Δ hits	p_{Holm}
Domainome	0.85	12	$+0.2219$	0.41442
Domainome	0.85	24	$+0.3425$	0.41442
Domainome	1.00	12	$+0.0421$	0.79404
Domainome	1.00	24	$+0.5618$	0.05368
Domainome	1.60	12	$+0.4332$	0.02200
Domainome	1.60	24	$+0.9454$	0.00048
ProteinGym	0.85	12	$+0.3362$	0.00388
ProteinGym	0.85	24	$+0.7002$	0.00114
ProteinGym	1.00	12	$+0.3850$	0.00088
ProteinGym	1.00	24	$+1.0690$	0.00012
ProteinGym	1.60	12	$+0.3207$	0.00388
ProteinGym	1.60	24	$+0.9726$	0.00012

Algorithm 1 Greedy search with optional position feedback

Require: Initial probabilities p, neighbor lists $N(i)$, position map g, budget K, constants $\gamma, \kappa, \lambda, \pi_0$

$D \leftarrow \varnothing$; initialize $n_h = z_h = 0$ for each position h
for $t = 0, \ldots, \min(K, n) - 1$ **do**
  Compute q_i from Equation 2 for each unmeasured i
  Compute s_i^+ from Equation 3
  Choose the unmeasured $i^\star$ maximizing s_i^+, breaking ties by lowest index
  Reveal $y_{i^\star}$; add $(i^\star, y_{i^\star})$ to D
  $h \leftarrow g(i^\star)$; $n_h \leftarrow n_h + 1$; $z_h \leftarrow z_h + y_{i^\star}$
end for
return Measured set and its hit count

Corpus	K	Historical	Nested primary	Change	Units whose cell moved
Domainome	12	$+0.4332$	$+0.4214$	-0.0118	$3/67$
Domainome	24	$+0.9658$	$+0.9658$	$+0.0000$	$0/67$
ProteinGym	12	$+0.3943$	$+0.3965$	$+0.0022$	$1/97$
ProteinGym	24	$+0.8753$	$+0.8927$	$+0.0174$	$66/97$

At ProteinGym $K = 24$, many selections move but the effect changes by only 0.0174 hits: the selection surface is flat. This bounds the active calibration path, not benchmark reuse.

B THE EPISODE ALGORITHM

Algorithm 1 makes the observation boundary explicit. The position counters use only realized labels after a query. Setting $\lambda = 0$ removes the position term while keeping the remaining bookkeeping identical. The description is a direct implementation of the scoring equations, not a new optimality result.

Lookahead replaces the argmax with Equation 4. For each candidate action, both hypothetical labels update its reverse neighbors and its position counts before the continuation is ranked. Hypothetical changes are discarded before evaluating the next action; only the realized label changes the persistent state. The action shortlist has at most 100 candidates. The final-step score reduces to the immediate probability, because the continuation budget is zero.

Why common initial scores matter. Zero observations give identical priors, excluding a stronger starting ranking as the matched greedy explanation. Joint beliefs, subsequent predictive risk and first lookahead actions need not match.

Terminal-set accounting. For two complete, equal-budget trajectories on the same pool, let A and B be their terminal sets and $v = |A \setminus B| = |B \setminus A|$. The common-label sum cancels, leaving the difference between exclusive sums. For $v > 0$,

$$H(A) - H(B) = v\left(\overline{y}_{A\setminus B} - \overline{y}_{B\setminus A}\right). \tag{10}$$

For $v = 0$ the difference is zero. This retrospective identity separates replacement quantity from quality; exchanged labels are read after replay. Same-state action disagreement and between-trajectory replacement are distinct measurements. For unit u with pools $\mathcal{P}_u$, the reported average gain is

$$\frac{1}{C}\sum_{u=1}^{C}\frac{1}{|\mathcal{P}_u|}\sum_{p\in\mathcal{P}_u} v_p\left(\overline{y}_{A_p\setminus B_p} - \overline{y}_{B_p\setminus A_p}\right).$$

This expression does not factor into a product of two unweighted averages. In particular, historical pooled hit rates among exchanged candidates use different weights from the headline unit means; they are not multiplied in our figures.

C Paired inference and supplementary comparisons

The primary nested greedy comparison's standard errors and two-test-per-corpus Holm values are in Table 1. The supplementary tables below retain the historical protocol and its own uncertainty estimates; they must not be substituted for nested inference. Tests randomly flip the signs of paired unit differences, using exact enumeration where feasible and 10^5 draws otherwise. Each table retains the correction family of its source experiment. The original lookahead family contains four contrasts at two budgets within each corpus; the quota experiment has its own eight-comparison family. A small adjusted p in one such family is not adjustment for the full history of model development. The weakest headline position contrast changes its significance under broader historical families, so we report effects and families rather than a universal four-cell significance claim. Failure to reject a null is not evidence of equivalence, and power calculations are not used as upper confidence bounds on an absent effect.

C.1 Sensitivity to grid weighting

These post-review descriptive summaries reuse every archived grid-cell effect. They exclude the largest sharing strength, reweight cells by inverse sharing strength, or omit each strength in turn; none selects a favorable configuration for a test unit. All four corpus–budget estimates are reported for each rule.

Corpus	K	Uniform	Exclude $\lambda = 1$	Weight $1/\lambda$	Leave-one range
Domainome	12	$+0.1969$	$+0.1366$	$+0.1477$	$[+0.1366, +0.2396]$
Domainome	24	$+0.3207$	$+0.2601$	$+0.2688$	$[+0.2601, +0.3701]$
ProteinGym	12	$+0.0896$	$+0.0753$	$+0.0748$	$[+0.0753, +0.1041]$
ProteinGym	24	$+0.3347$	$+0.2787$	$+0.2773$	$[+0.2787, +0.3938]$

The direction remains positive under these weighting changes, but the size changes. This does not establish corrected significance of any reweighted endpoint: the cell summaries do not provide their joint unit-level covariance, so no reweighted standard error or new test is inferred from individual cell errors. It is sensitivity of point estimates on development-exposed data, not independent replication.

The score-contiguous partition preserves group sizes by sorting the sequence feature and assigning consecutive candidates to groups of those sizes. Random groups preserve the same size multiset; the reported random control averages three permutations. The feature geometry is not otherwise matched, and averaging permutations is not identical to evaluating one fixed relation. In the position-versus-score comparison, the within-corpus correction has twenty tests; the wider sixty-test family includes the historical additional corpus and is reported only to illustrate sensitivity to multiplicity. No outcomes from that excluded corpus support our scientific conclusions.

Table 5: Paired effects across inference units. Approximate SEs divide rounded MDEs by 2.80 (ENS) or 2.8016 (quota). Holm correction retains each corpus/experiment's eight-test family; the interaction has its own four-test family. Position–Lookahead and Position–Greedy are separately tuned. The interaction instead averages fixed cells without test-unit configuration selection; $137/144$ cells are positive.

Contrast	Corpus	K	Δ hits	SE	p_{Holm}
Position–Greedy vs. kNN–Greedy					
	Domainome	12	$+0.4332$	0.1563	0.02352
	Domainome	24	$+0.9658$	0.2369	0.00056
	ProteinGym	12	$+0.3943$	0.1088	0.00080
	ProteinGym	24	$+0.8753$	0.2044	0.00024
kNN–Lookahead vs. kNN–Greedy					
	Domainome	12	-0.0074	0.0064	0.46876
	Domainome	24	$+0.0319$	0.0255	0.46876
	ProteinGym	12	$+0.0098$	0.0194	1.00000
	ProteinGym	24	$+0.0263$	0.0483	1.00000
Position–Lookahead vs. Position–Greedy					
	Domainome	12	$+0.1589$	0.0952	0.29796
	Domainome	24	$+0.4731$	0.1550	0.01476
	ProteinGym	12	$+0.2811$	0.1231	0.07296
	ProteinGym	24	$+0.7568$	0.2514	0.00624
Quota vs. kNN–Greedy					
	Domainome	12	-0.6924	0.1578	0.00030
	Domainome	24	-1.5371	0.2681	0.00016
	ProteinGym	12	-0.0569	0.0679	1.00000
	ProteinGym	24	-0.1445	0.1266	1.00000
Interaction, Equation 8: grid-averaged, own four-test family					
	Domainome	12	$+0.1969$	0.0680	0.01056
	Domainome	24	$+0.3207$	0.0869	0.00096
	ProteinGym	12	$+0.0896$	0.0550	0.10666
	ProteinGym	24	$+0.3347$	0.1314	0.02092

Table 6: Extended-grid partition controls. Random denotes the average of three permutations. Score bins preserve group sizes and contiguity in one MPNN score, not complete feature geometry. SEs are approximate, recovered from rounded MDEs divided by 2.80. Reported Holm values use 20 tests per corpus; under the original 60-test cross-corpus family, the Domainome $K = 12$ score-bin contrast has $p = 0.10556$. This grid differs from Table 9.

Partition contrast	Corpus	K	Δ hits	SE	p_{Holm}
True position − random					
	Domainome	12	$+0.4305$	0.1563	0.07224
	Domainome	24	$+0.8931$	0.2372	0.00272
	ProteinGym	12	$+0.4194$	0.1106	0.00120
	ProteinGym	24	$+1.2452$	0.2490	0.00040
True position − score bins					
	Domainome	12	$+0.4634$	0.1590	0.04732
	Domainome	24	$+1.0010$	0.2624	0.00272
	ProteinGym	12	$+0.5238$	0.1348	0.00120
	ProteinGym	24	$+1.1830$	0.2573	0.00040

The fixed-parameter planning comparison chooses the globally modal grid cell among the cross-validation selections and runs both policies at that cell. It eliminates between-arm parameter differences at evaluation time, but uses a data-informed cell and is not a new outer-fold-nested comparison. The table displays both the greedy-modal and lookahead-modal versions where the archived run provides them; absent rows are not inferred.

Table 7: Position–Lookahead minus Position–Greedy. The modal-cell diagnostics hold both arms at one global modal parameter setting, selected from the historical cross-validation choices; they are not a new nested evaluation. Each matched contrast is substituted into its original eight-test family. ProteinGym $K = 12$ has no matched-modal result in this artifact.

Corpus	Parameter choice	K	Δ hits	p_{Holm}
Domainome	Separately tuned	12	$+0.1589$	0.29796
Domainome	Greedy modal cell	12	$+0.1248$	0.70314
Domainome	Lookahead modal cell	12	$+0.3309$	0.01848
Domainome	Separately tuned	24	$+0.4731$	0.01476
Domainome	Greedy modal cell	24	$+0.4808$	0.03160
Domainome	Lookahead modal cell	24	$+0.4602$	0.10384
ProteinGym	Separately tuned	12	$+0.2811$	0.07296
ProteinGym	Separately tuned	24	$+0.7568$	0.00624
ProteinGym	Greedy modal cell	24	$+0.5525$	0.07496
ProteinGym	Lookahead modal cell	24	$+0.5695$	0.08912

D Robustness, alternative estimators, and unresolved explanations

Table 8: **Alternative explanations, controls and inference boundaries.** Protocols differ; these are complementary controls, not independent replications.

Explanation	Controlled comparison	What the evidence supports and leaves open
Stronger starting ranking	Identical initial probabilities and greedy acquisition (Table 1)	Gains arise after feedback; ESM-1v compares different initial predictors.
Any grouping helps	Size-matched random groups and score-contiguous groups (Table 6)	Controls do not reproduce the gains; other feature geometries remain open and weak contrasts depend on correction.
Only the Beta offset works	Position-weighted ridge and site-kernel GP (Table 2)	Ridge/GP gains extend across formulas; Domainome GP is marginal. Capacity and calibration are not matched.
Broader coverage suffices	Quota acquisition and terminal-set comparison (Section 6)	Membership changes need not improve yield; one coverage rule is tested under candidate-count reward.

Static and diversity controls. Static queries the initial top K without feedback and loses to position greedy at both budgets in both corpora. Round-robin quota imposes coverage on neighborhood predictions; a diminishing-return score discounts sampled sites, often selecting zero discount. These controls do not exhaust diversity objectives.

Planning sensitivity. The neighborhood lookahead checks vary $\gamma \in \{2, 4, 8, 25\}$, neighbor count $k \in \{12, 25, 50, 100, 200\}$, and action cap in $\{100, 300, 1000\}$. No corrected positive planning increment is found in those tested sweeps; some increments are negative. The cap runs have equal reported summary statistics, which does not prove identical trajectories or equivalence to an uncapped planner. These experiments address the specific planning null, not the maximum achievable performance of every feature-only model.

Budget and hit definition. The post-hoc budget sweep extends from $K = 6$ to 48. Position-update point estimates remain positive, with weaker evidence at the smallest Domainome budget. At $K = 100$ and 200, the position-greedy contrast remains positive, but the additional planning effect differs across corpora. Changing the hit quantile from 5% to 10% or 1% also changes the pattern: corrected evidence is least consistent at the rarest threshold. The constructed prevalence and the composition of each pool are therefore part of the result's scope, not incidental benchmark details.

Table 9: Mean hits found (higher is better), averaging pools within units and units equally. Position adds same-position label sharing; Lookahead is ENS. Position update gain compares Position–Greedy with kNN–Greedy; ± denotes approximate paired SE. Position–Lookahead is separately tuned. ESM-1v changes the initial predictor; dashes mean not evaluated. Full inference: Table 5.

	Domainome		ProteinGym	
Method	$K = 12$	$K = 24$	$K = 12$	$K = 24$
ESM-1v static	—	—	1.420	2.531
Matched static	3.173	6.063	1.538	2.871
kNN–Greedy	3.330	6.102	1.482	2.859
kNN–Lookahead	3.322	6.134	1.492	2.885
Position–Greedy	3.763	7.068	1.877	3.734
Position–Lookahead	3.922	7.541	2.158	4.491
Position update gain	$+0.433$	$+0.966$	$+0.394$	$+0.875$
Paired SE	± 0.156	± 0.237	± 0.109	± 0.204

Other estimators. The historical estimator comparison replaces the update while retaining greedy acquisition and the common initial prior. The ridge model gives the position-indicator block its own weight; the site-kernel Gaussian process adds a same-position kernel to an RBF kernel over the standardized sequence and structure-confidence channels. It is a related model, not a full reproduction of Kermut (Groth et al., 2024). The following paired hit gains are against neighborhood greedy; parentheses give Holm p values from the original 16-test family within each corpus, not the nested primary family.

Corpus / budget	Position-weighted ridge	Site-kernel GP
Domainome / 12	0.2236 (1.00000)	0.2690 (1.00000)
Domainome / 24	0.6904 (0.08008)	0.6872 (0.23606)
ProteinGym / 12	0.4158 (0.00448)	0.4147 (0.00414)
ProteinGym / 24	1.2761 (0.00032)	1.1123 (0.00032)

These results support a useful position relation beyond the additive Beta formula on ProteinGym, not superiority of that formula or a fully identified shrinkage mechanism. The ridge search at Domainome $K = 12$ selected its upper grid boundaries in $59/67$ units; its nondetection is therefore not evidence against a better-tuned ridge model. These historical selected-model comparisons retain the calibration and development limitations; the archived summary alone does not constitute a complete executable reproduction.

What remains unexplained. A development ranking-stability certificate never triggered; a proposed gain law missed magnitudes. Neither explains the empirical effects. Synthetic association changes also changed prevalence, so they do not isolate association or establish a law for useful sharing.

E Evidence map and artifact reproduction

The source package contains figure/table generators, recorded result summaries and content hashes; its README maps comparisons to producing experiments. Regenerating these assets checks presentation provenance. The separate primary replay bundle reproduces Table 1 from frozen candidate pools and features (Appendix A.1); upstream feature generation and historical planner episodes remain outside that bundle.

E.1 Historical tuned multiarm comparisons

The following table retains the historical selection protocol, including planner arms not recomputed with the nested calibration repair. Its values are distinct from the nested greedy-only primary comparison in Table 1.

F EXACT ANALYSIS OF CONTINGENT BATCH SELECTION

All probabilities below condition on the currently observed history. The queried candidate is fixed, and m subsequent candidates are selected simultaneously after its label is revealed. The analysis does not compare alternative query choices or solve the full adaptive decision tree. In particular, coherence is an assumption of this section, not an established property of Equation 3.

F.1 BINARY CONDITIONAL UPDATES

Let Y_i, Y_j be binary labels with conditional means p_i, p_j, and assume $0 < p_i < 1$. Writing $d_{ij} = \mathrm{Cov}(Y_i, Y_j)$ gives

$$\mathbb{E}[Y_j \mid Y_i = y] - p_j = (y - p_i)\frac{d_{ij}}{p_i(1-p_i)}. \tag{11}$$

Indeed, $\mathbb{E}[Y_iY_j] = p_ip_j + d_{ij}$, so the posterior after a hit is $p_j + d_{ij}/p_i$. Averaging hit and miss posteriors must recover p_j, so the miss posterior is $p_j - d_{ij}/(1-p_i)$. This proves the identity. When $p_i \in \{0, 1\}$ the observation is deterministic and the supported posterior equals the prior; the displayed division is then inapplicable. Equation 11 is an elementary identity: equal initial marginal probabilities do not specify how labels update other candidates.

F.2 NONNEGATIVITY AND THE EXACT EQUALITY CONDITION

For the remaining set U, let $\mathcal{S}_m$ contain its subsets of size m. Choose any prior-optimal $S^* \in \mathcal{S}_m$. The tower property gives

$$J_i = \sum_{y:\,\Pr(Y_i=y)>0} \Pr(Y_i = y) \left[T_m(p_U^{(i,y)}; U) - \sum_{j \in S^*} p_j^{(i,y)} \right]. \tag{12}$$

Every bracket is nonnegative. Consequently $J_i \geq 0$, and equality implies that S^* is optimal for every supported outcome. Conversely, suppose some set S is optimal under every supported outcome. Averaging its posterior dominance over every competing set shows that S is also prior-optimal. Substituting it into Equation 12 gives $J_i = 0$.

Thus zero value is equivalent to the existence of a common optimizer, including ties. A numerical algorithm may return distinct tied optimizers while value is zero. If the query is deterministic, there is only one supported outcome and zero value. If $m = 0$ or $m = |U|$, there is only one feasible set and again zero value. This argument requires coherent conditional probabilities; arbitrary score updates need not satisfy the tower property.

F.3 A FEASIBLE EXCHANGEABLE GROUP

To realize Equation 6, draw a mixture indicator with success probability $\rho \in [0, 1]$. On success, all $r+1$ group labels share one Bernoulli(p) draw; otherwise, draw them independently as Bernoulli(p). Every marginal is p, and for distinct members

$$\mathbb{E}[Y_iY_j] = \rho p + (1-\rho)p^2, \qquad \mathrm{Cov}(Y_i, Y_j) = \rho p(1-p). \tag{13}$$

Equation 11 therefore yields q_1 and q_0. The outside label vector can be independent of this construction, with arbitrary internal dependence and marginal probabilities $c_1, \ldots, c_n$. For $0 < \rho < 1$, a Beta–Bernoulli mixture with concentration $1/\rho - 1$ and mean p gives the same one-observation posteriors. Both constructions hold prevalence p fixed as correlation changes. The restriction $p \in (0, 1)$ avoids undefined correlations at zero variance.

F.4 PROOF OF PROPOSITION 1

For a common sibling probability q, a feasible continuation batch contains k siblings and the highest $m - k$ outsiders, where $\ell \leq k \leq u$. Its optimal value is

$$F_m(q) = \ell q + \sum_{a=1}^{m-\ell} c_a + \sum_{k=\ell}^{u-1} (q - c_{m-k})_+. \tag{14}$$

To prove this expression, begin with ℓ siblings. Increasing their count from k to $k+1$ replaces outsider c_{m-k} with a sibling, for incremental value $q - c_{m-k}$. These increments are nonincreasing with k, because the outsider probabilities are sorted. Taking every positive increment maximizes the batch sum; zero increments can be taken or omitted without changing the value. This yields Equation 14. Every outsider index in a nonempty last sum is between 1 and n. If $n = 0$, then $\ell = u = m$ and that sum is empty.

Substitute Equation 14 into $J_i = pF_m(q_1) + (1-p)F_m(q_0) - F_m(p)$. The constant and affine terms cancel because $pq_1 + (1-p)q_0 = p$. Each remaining term at a boundary c is

$$h(c) = p(q_1 - c)_+ + (1-p)(q_0 - c)_+ - (p - c)_+. \tag{15}$$

Since $q_0 \le p \le q_1$, direct expansion gives

$$h(c) = \begin{cases} 0, & c \le q_0 \text{ or } c \ge q_1, \\ (1-p)(c - q_0), & q_0 \le c \le p, \\ p(q_1 - c), & p \le c \le q_1. \end{cases} \tag{16}$$

The branches agree at their endpoints. They equal $\min\{p(q_1 - c)_+, (1-p)(c - q_0)_+\}$, proving Equation 7. Each term is positive exactly for $q_0 < c < q_1$, since p and $1 - p$ are positive. This proves the strict-positivity condition, including the zero-value case of an empty sum and boundary ties.

For equal outsider probabilities c, there are $s = u - \ell$ identical terms. Substituting $c = p + d$ gives the margin form quoted in Section 4,

$$J_i = s \min\{p[\rho(1-p) - d]_+,\ (1-p)[\rho p + d]_+\}. \tag{17}$$

Positive value requires both $s > 0$ and $-\rho p < d < \rho(1-p)$. At $d = 0$, its value is $s\rho p(1-p)$. If $m = r + n$, then $s = 0$. Thus the number of updated siblings alone does not determine value: there must remain candidates to exchange at a boundary reached by the update.

F.5 Misspecified feedback relations

When the relation is wrong. Proposition 1 uses true conditional probabilities; a deployed scorer acts on an assumed dependence, and that case is the one worth characterizing.

Proposition 2 (Value under a misspecified relation)**.** *Let a policy act on assumed correlation $\hat{\rho}$, with $\hat{q}_y$ from Equation 6 at $\hat{\rho}$, while labels follow ρ. Relative to not sharing, its continuation value changes only at boundaries c with $\hat{q}_0 < c < \hat{q}_1$, and*

$$\Delta = p \sum_{\text{crossed after a hit}} (q_1 - c) \ + \ (1-p) \sum_{\text{crossed after a miss}} (c - q_0). \tag{18}$$

Thus $\hat{\rho}$ enters only through which *boundaries are crossed, and ρ only through what each crossing is worth.*

Call a crossing unjustified when the true posterior lies on the opposite side of c from the believed one. Because $\rho \ge 0$ gives $q_1 \ge p \ge q_0$, its cost is at most $p(c-p)_+ + (1-p)(p-c)_+$ — the boundary's own *margin* — while a justified crossing after a hit gains $p(q_1 - c)$, bounded only by $p(1-c)$. Near a boundary the margin is small by construction, so an unjustified crossing can cost little while a justified crossing can yield a larger gain. The margin is read from current scores before any label; whether it also predicts the *sign* of a realized replacement is empirical and not settled here (Appendix F.5).

Assume uniform outsider probability c, true group marginal p, true correlation $\rho \in [0,1]$, and an assumed correlation $\widehat{\rho} \in [0,1]$. Let k_* be a fixed prior-optimal sibling count: u when $p > c$, ℓ when $p < c$, and any feasible count when $p = c$. The model uses $\widehat{q}_1 = p + \widehat{\rho}(1-p)$ and $\widehat{q}_0 = p(1-\widehat{\rho})$, selecting $k_y = u$ if $\widehat{q}_y > c$ and $k_y = \ell$ if $\widehat{q}_y < c$. On equality it retains k_*. The true expected improvement of this contingent rule over the fixed batch is

$$\Delta_{\text{true}} = p(k_1 - k_*)(q_1 - c) + (1-p)(k_0 - k_*)(q_0 - c). \tag{19}$$

Indeed, conditional on outcome y, a batch with k siblings has true expected reward $mc + k(q_y - c)$. Subtract the fixed k_* batch and average the two outcomes. This argument also covers $p = c$ and the stated tie rule.

If $p < c$ and $\widehat{q}_1 > c$, only a hit changes the sibling count, giving $\Delta_{\text{true}} = ps(q_1 - c)$. Provided $s > 0$, its sign is negative, zero, or positive according as ρ is below, equal to, or above $(c - p)/(1 - p)$. If $p > c$ and $\widehat{q}_0 < c$, only a miss changes the count, giving $\Delta_{\text{true}} = (1 - p)s(c - q_0)$; for $s > 0$ its corresponding threshold is $(p - c)/p$. If $s = 0$, the gain is zero for every true and assumed correlation, regardless of these inequalities.

For $r = n = m = 1$, $p = 0.2$, $c = 0.3$, and $\widehat{\rho} = 0.5$, the assumed posteriors are 0.6 and 0.1. The model predicts a continuation improvement $0.2(0.6) + 0.8(0.3) - 0.3 = 0.06$. Under true independence ($\rho = 0$), the sibling remains at probability 0.2 after either outcome, so the same choices produce $0.2(0.2) + 0.8(0.3) = 0.28$, a loss of 0.02 relative to the fixed outside choice. The starting marginals are correct; only the assumed feedback relation is wrong. If instead $m = 2$, both candidates must be selected and the loss is zero. These calculations are illustrative expectations under specified probability models, not new measurements on either biological corpus.

The margin bounds an unjustified crossing. Proposition 2 is Equation 19 written per boundary: k_y is fixed by where $\widehat{q}_y$ falls among the outsider probabilities, and each term is valued at the true q_y. Call a crossing at boundary c *unjustified* when the true posterior lies on the opposite side of c from the believed one. Its cost is bounded by the boundary's own margin, in two cases.

After a hit with $c > p$, the policy takes the sibling because $\widehat{q}_1 > c$; if unjustified then $q_1 < c$ and the cost is $p(c - q_1) \le p(c - p)$, since $\rho \ge 0$ forces $q_1 \ge p$. After a miss with $c < p$, it abandons the sibling because $\widehat{q}_0 < c$; if unjustified then $q_0 > c$ and the cost is $(1 - p)(q_0 - c) \le (1 - p)(p - c)$, since $\rho \ge 0$ forces $q_0 \le p$. Together these give $p(c - p)_+ + (1 - p)(p - c)_+$, the two-sided form; a one-sided version counting only $c > p$ omits the second channel and does not bound the loss. A justified crossing after a hit is worth $p(q_1 - c)$, bounded only by $p(1 - c)$, so the two sides are not symmetric: the loss from misjudging a boundary is controlled by its margin and the gain from judging it correctly is not.

The same case split appears in $h(c)$ above, which is why this is a consequence of the model already stated rather than an additional assumption. It concerns one continuation under exact posteriors; it is not a guarantee about the heuristic scorer or about any adaptive trajectory.

F.6 Acquisition-score decomposition and its limits

Equation 9 holds algebraically for any finite scores in $[0, 1]$, with every continuation excluding the query and all previously observed candidates. The complete score is $B_i + C_i + D_i$, not $r_i + C_i + D_i$: dropping $T_m(r_{U_i})$ changes comparisons between queries because their exclusion changes the continuation pool. The sum-of-largest-m function is convex, hence $C_i \ge 0$. Under coherent conditional probabilities the tower property gives $\bar{r}_i = r_{U_i}$ and $D_i = 0$. For a heuristic update, C_i is a score-based selection quantity rather than guaranteed expected discovery, and D_i is one scalar projection of incoherence, not its total size. Removing D_i does not construct coherent posteriors.

A candidate-level consistency diagnostic would instead retain every residual

$$\epsilon_{ij}(D_t) = r_j(D_t) - \big[(1 - r_i(D_t))r_j(D_t \cup \{(i, 0)\}) + r_i(D_t)r_j(D_t \cup \{(i, 1)\})\big].$$

Coherent conditional probabilities require these residuals to vanish; vanishing residuals alone do not establish joint coherence or empirical calibration. The scalar D_i can hide candidate-level discrepancies through aggregation and selection. We do not report a representative-history calibration or residual audit, nor a matched coherent hierarchical comparator, so the score interventions cannot certify a value-of-information mechanism.

For the neighborhood update, observing i moves a reverse neighbor j from q_j to $(A_j q_j + y)/(A_j + 1)$, where $A_j = \gamma + |L_t \cap N(j)|$. Averaging using q_i gives mean-minus-current displacement $(q_i - q_j)/(A_j + 1)$ (the negative of ϵ_{ij}); at $\gamma = 25$ its magnitude is at most one twenty-sixth of the score gap. This identity does not determine which acquisition component changes an argmax or improves terminal reward.

An earlier subsample diagnostic omitted the query-dependent baseline continuation term when selecting actions and did not compare updates on a common history. Its trajectory was therefore not the specified full planner. We withdraw its magnitude ratios and mechanism attribution; they are not evidence for the interaction in Equation 8. The corrected comparison retains the full baseline, checks

Table 10: All component-intervention arm means (hits). Parameters are fixed; these are not nested-tuned or grid-averaged comparisons. Pools are averaged within inference units, then units equally. Removing C retains $B + D$; removing D retains $B + C$.

	Domainome		ProteinGym	
Update / acquisition	$K = 12$	$K = 24$	$K = 12$	$K = 24$
kNN / Greedy	3.4254	6.5187	1.5122	2.8267
kNN / Full ($B + C + D$)	3.3895	6.4213	1.5124	2.8792
kNN / Without C ($B + D$)	3.4028	6.5410	1.4915	2.8714
kNN / Without D ($B + C$)	3.3751	6.2701	1.5270	2.9410
Position / Greedy	3.6208	7.0574	1.9023	3.8346
Position / Full ($B + C + D$)	3.9803	7.7016	2.0547	4.3478
Position / Without C ($B + D$)	3.9060	7.2634	1.8272	4.0754
Position / Without D ($B + C$)	3.9916	7.5508	2.0734	4.1772

reference trajectory parity, and distinguishes same-history score interventions from total trajectory effects.

Frozen component intervention. We replay all 245 Domainome pools and 161 ProteinGym pools, with the same candidate, label, feature, and inference-unit identities as the historical replay. Both updates start from calibration slope 1, with $\pi_0 = 0.05$, $\gamma = 25$, and 50 nearest neighbors (or all other candidates in a smaller pool). The position update uses $\kappa = 2$ and $\lambda = 0.375$; the neighborhood update uses $\lambda = 0$. These settings were frozen before the component outcomes were computed, but were informed by earlier development on these corpora. This is a development reanalysis, not an independent confirmatory sample.

The primary budget is $K = 24$; $K = 12$ is a separately executed secondary search, not a prefix of the longer trajectory. For each update we run greedy, full $B + C + D$, without drift $B + C$, and without contingency $B + D$. At each step the eligible lookahead shortlist contains the 100 highest current scores; candidates outside it retain their immediate scores, matching the reference ENS implementation. The full and greedy policies were checked for exact query and label trajectory agreement with that implementation on both real-pool pilot cases and small regression cases, including tied scores. NumPy and the reference sorting operation are fixed to preserve ties. The ablations change scoring only; realized posterior updates and the label-access boundary are unchanged.

Shadow decisions use the neighborhood-greedy trajectory as a common paid history for both updates. The query removed for a continuation-set diagnostic is the same reference query under each update. Thus shadow comparisons condition on the same information, while complete trajectories measure total policy effects after histories diverge. Step-level disagreement counts are descriptive and do not increase the number of independent inference units.

For component $X \in \{C, D\}$, the primary contrast is

$$J_X = \left(H_{\text{full}} - H_{\text{without } X}\right)_{\text{pos}} - \left(H_{\text{full}} - H_{\text{without } X}\right)_{\text{knn}}.$$

Each budget has one Holm family containing its four corpus–component tests. Pool differences are averaged within cluster or study, then inference units receive equal weight. We report 5,000 whole-unit percentile bootstrap draws and 20,000 paired sign flips, with the plus-one correction and deterministic contrast-specific seeds. Other policy contrasts have descriptive paired intervals. The complete available cohort fixes sample size; normal 80% power approximations use a Bonferroni four-test threshold and the observed standard error, and are sensitivity summaries rather than prospective power guarantees. The declared practical reference differences are 0.5 hits at $K = 24$ and 0.25 at $K = 12$; nonsignificance alone does not establish equivalence. Neither a difference in significance nor average component magnitude establishes C-versus-D dominance.

Independent finite-world calculation. We also validate the fixed-query formulas by enumerating all 32 binary worlds for one query, two siblings and two independent outsiders. The frozen grid crosses $p \in \{0.05, 0.2, 0.5\}$, true dependence $\rho \in \{0, 0.25, 0.5, 1\}$, assumed dependence $\hat{\rho} \in \{0, 0.5, 1\}$, $c = \text{clip}(p + \delta, 0, 1)$ with $\delta \in \{-0.1, 0, 0.1\}$, and continuation capacity $m \in \{0, 1, 2, 3, 4\}$: 540 conditions in total. World enumeration and exhaustive continuation-set

selection agree with the analytic values to maximum absolute error below 10^{-14}. This checks the stated model, including misspecified continuation choices; it is not evidence of measured protein-policy improvement or of posterior coherence in the empirical updater.

G ARCHIVED TRAJECTORIES AND EXACT MODEL CALCULATIONS

Source reconciliation and denominators. Archived picks are unique within trajectories, shared candidate labels agree, and neighborhood-greedy sequences match across bundles. Unit means agree to 10^{-12} and all 32 rounded arm means reproduce. No new protein-policy runs enter this reanalysis. Replacement fractions and incoming/outgoing hit counts use budget K as denominator; pool values average within the original units, then units equally. ECDFs use unit means. $K = 24$ prefixes are histories, not separate $K = 12$ experiments.

Representative selection before visualization. At $K = 24$, select the pool nearest the componentwise median of position-greedy versus neighborhood-greedy $(v/K, \mathrm{gain}/K)$ by squared Euclidean distance, breaking ties by pool index. No effect-sign filter is used. Selection is pool-weighted, not a change to the unit-weighted estimand. Both representatives have zero gain: Domainome exchanges seven candidates and two hits for two; ProteinGym exchanges eight and zero hits for zero. Pool identity, unit, size and complete index/label sequences match pick archives. The position-offset decomposition uses its recorded history, not an alternative rollout.

Descriptive intervals. Figure 6a reconstructs an approximate standard error from the archived minimum detectable effect (MDE), then plots mean $\pm 1.96\,\mathrm{SE}$. The producing budget-sweep code used $\mathrm{MDE} = 2.8016\,\mathrm{SE}$; the partition and estimator summaries used $2.80\,\mathrm{SE}$. Rounded source summaries introduce rounding error. These marginal normal intervals are descriptive and do not replace the paired sign-flip tests or Holm corrections in Appendix C.

Three exact illustrations. Figure 4 uses the coherent shared-versus-independent Bernoulli mixture from Appendix F. All three examples keep the sibling marginal hit probability $p = 0.2$ fixed. Panel (a) has one sibling, one outsider, and one continuation slot; it varies ρ and outsider score c. Panel (b) has four siblings and four outsiders with $c = 0.25$ and $\rho = 0.5$, varying continuation budget from zero to eight. Panel (c) has one sibling, one outsider at $c = 0.3$, and one slot: the policy assumes $\widehat{\rho} = 0.5$ while the true correlation varies. Its value under the assumed model is 0.06; at true independence its expected gain is -0.02. These are exact model expectations, not fitted predictions or measured protein-policy effects. Tests compare the closed form with direct top-m sums and independent finite-world enumeration; the surface also agrees pointwise with the direct two-outcome expectation to numerical precision. Code, calculation inputs, selected replay records, and source hashes are bundled.

Assay identities and experimental categories. Archived 24-query outcomes join uniquely to all 161 official ProteinGym v1.3 `coarse_selection_type` annotations through producing order, checked study labels and candidate counts. Pick NPZs lack assay IDs; reconciled manifests and metadata accompany the per-assay CSV. Sort by gain then name, retaining every assay.

The five assay-weighted mean gains are 1.125 for organismal fitness (64 assays), 1.395 for stability (43), 0.677 for activity (31), 0.643 for expression (14), and 0.889 for binding (9). These summaries retain the historical calibration limits and do not estimate biological differences between assay types. In particular, the 42 Tsuboyama assays contribute separately to the assay-weighted summary but together form one unit in the primary study-weighted analysis. No new significance tests are added.

Endpoint uncertainty. The nested panels of Figure 1 summarize recorded $K = 12$ and $K = 24$ episodes, with pools averaged within units and then units equally. Pointwise whole-unit intervals are descriptive, not simultaneous bands or additional corrected tests. Its external-ranking panel retains the historical protocol. Half-violin shapes in Figures 1 and 3 use Gaussian kernel density estimates with Scott bandwidth, restricted to each observed range. Points retain exact outcomes; jitter affects only categorical positions. Discovery densities weight pools equally for display, whereas the overlaid means and intervals retain equal cluster/study weighting. Planning densities show the inference units.

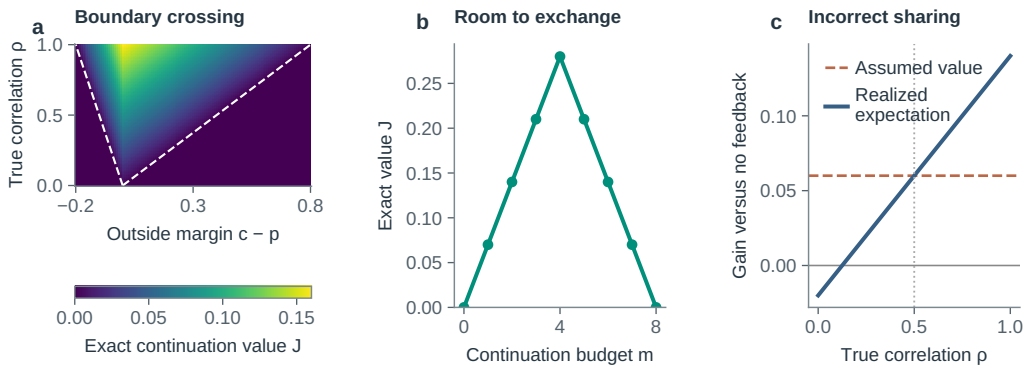

Figure 4: **Correlation is useful only when it changes a feasible choice.** Exact coherent-model calculations at fixed marginal hit probability $p = 0.2$. (a) Feedback value is positive between the two posterior crossing boundaries (dashed). (b) With four siblings and four outsiders, value vanishes when none or all can be selected. (c) Assuming correlation 0.5 can yield negative true gain when the relation is weaker. These model expectations are separate from the protein experiments; parameters and finite-world checks appear in Appendix G.

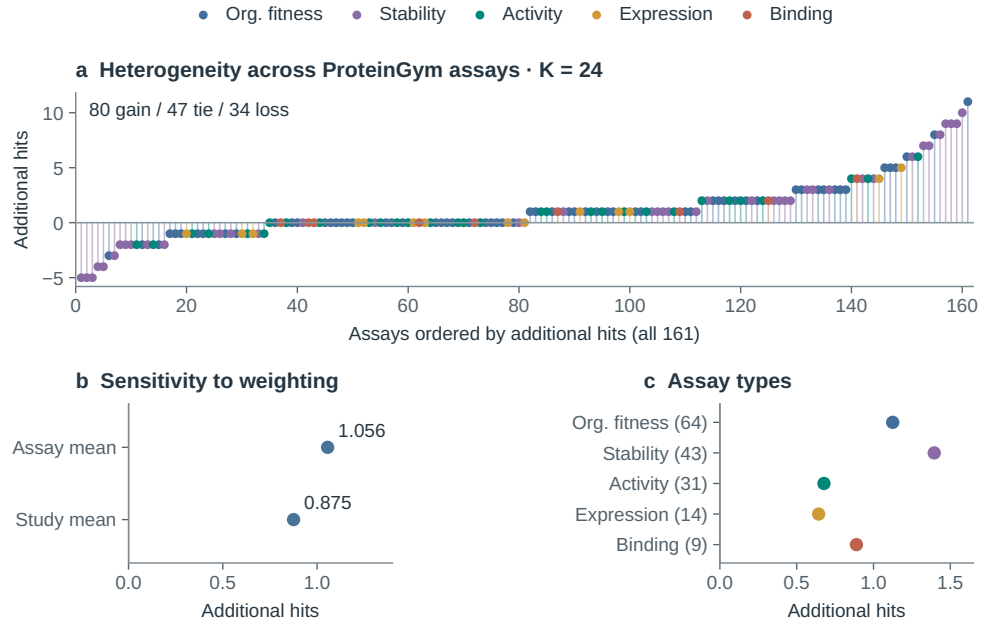

Figure 5: **Historical assay heterogeneity and descriptive strata.** (a) All 161 ProteinGym assays ordered by position-minus-neighborhood greedy hits at $K = 24$, colored by assay type: 80 gain, 47 tie and 34 lose. (b) Equal-assay versus equal-study means. (c) Assay-type means with counts. These historical descriptive summaries are distinct from the nested-primary and new common-prior protocols.

G.1 Predicting a replacement's value at decision time

Every step at which the position arm departs from the neighborhood arm's choice on its own history is a replacement: it takes candidate a where the baseline would take b. Recording those steps over both corpora gives $8{,}680$ replacements on the position arm, each carrying the scores, boundary margins and position evidence visible at that step, and a value $y_a - y_b$ knowable only afterwards.

The exploratory predictor uses decision-time fields only. Five outer folds split inference units, and three inner folds choose among ridge penalties $0.1, 1, 10, 100$ and two histogram boosting regressors (depth 3, 100 iterations or depth 4, 200 iterations; learning rate 0.06), using mean validation Spearman correlation. Outer-training feature means and standard deviations standardize both partitions. Results average three fold seeds.

Within-episode sign-AUC compares each positive-valued replacement to each negative-valued replacement in the same episode, scores a strict predicted ordering as one and a tie as one half, then divides the pooled numerator by the total number of comparable pairs. Zero-valued replacements do not form pairs; episodes without both signs contribute no pairs. Thus episodes are weighted by comparable pairs, not equally or by biological inference unit. This descriptive metric prevents a constant episode-level productivity score from explaining within-episode ordering, but it is not a policy-value estimate.

Twenty within-episode outcome-permutation refits give finite-sample mean metrics of 0.513 on Domainome and 0.510 on ProteinGym, below the observed 0.576 and 0.578; none of those null draws exceeds the observed score. These sample means do not establish a theoretical null expectation different from 0.5. Global and within-unit permutations also give near-chance scores. The margin-only scores are 0.468 and 0.429. This is exploratory evidence of association in the chosen features; the limited permutation draws and absence of biological-unit uncertainty do not establish reliable decision-time prediction or a deployable gating policy.

G.2 Label-informed policy comparisons

We compare two uses of retrospective labels. A label-informed one-step veto restores the neighborhood pick exactly when that candidate is a hit and the position candidate is not. This optimizes the immediate choice between those two candidates, not terminal reward: changing the query also changes the observed history and subsequent choices. We therefore make no upper-bound claim for this veto policy or for the class of online replacement rules.

The per-pool strength comparator takes the maximum terminal reward over the eight predeclared constant-λ policies. This is an upper bound for choosing one of those constants before a pool's search begins, and only for that finite class. It does not bound policies that adapt λ during an episode. Both comparators use a baseline constant selected to maximize reward on the evaluated episodes; that baseline is itself outcome-selected, not deployable without additional validation. The reported quantities are descriptive policy comparisons under this auxiliary protocol, not the nested headline comparison.

Corpus	K	Best constant λ	One-step veto	Per-pool maximum λ
Domainome	12	3.6082	3.8041 (+5.4%)	4.1714 (+15.6%)
	24	6.7020	6.8612 (+2.4%)	7.6286 (+13.8%)
ProteinGym	12	2.7764	2.9689 (+6.9%)	3.2919 (+18.6%)
	24	5.4161	5.5590 (+2.6%)	6.2857 (+16.1%)

The veto increment is below the constant-grid spread (0.44–1.35 hits) in every cell, but the original H3 criterion was stricter: an increment below one quarter of that spread. The observed ratios are 0.440, 0.178, 0.333, and 0.106 in table order. Only two of four cells meet the criterion, and the recorded joint H3 outcome is false. The broader "below the spread" observation must not be substituted for this failed criterion.

The per-pool grid maximum is between 2.7 and 6.1 times the veto increment. This numerical ordering neither proves that strength adaptation is generally superior nor follows from the fixed-query theoretical bound. Selection on realized outcomes can reflect idiosyncratic noise. Counting each

locally avoided miss also overstates the measured trajectory difference because candidates can be queried later; policy replay and stepwise counterfactual counts answer different questions.

G.3 A SECOND TRUE RELATION: SPATIAL PROXIMITY

The partitions compared in Section 5.3 are one true relation and two controls, so nothing there tests whether the account survives a *different* true relation. Of the 161 ProteinGym pools, 114 have a structure on disk and 47 do not; the 47 are excluded for that reason alone, and they are not a random subset of the corpus. Each of the 114 is also required to align — the wild-type amino acid in its mutation keys must match the structure's residue at the matched number for at least 90% of positions, after a search over offsets, because a wrong offset would silently produce a meaningless partition that reads as "spatial sharing does not pay". On this corpus that guard refused none of the 114, so it is a precaution with no positive control here rather than a filter that did work.

The position partition holds one position per group, so a spatial partition matched to it would be that same partition. The spatial relation is therefore a coarsening: groups of g positions that are close in space, with $g \in \{2, 4, 8\}$, and each g carries its own size-matched random control so no comparison crosses group sizes. The clustering does use the structure — mean C_β distance within a spatial group is 8.0 Å against 25.8 Å within its random control.

K	g	Spatial	Size-matched random	Difference
12	2	+0.6228	+0.5000	+0.1228
	4	+0.4737	+0.2895	+0.1842
	8	+0.4298	+0.2982	+0.1316
24	2	+1.2281	+1.1140	+0.1140
	4	+1.0526	+0.8772	+0.1754
	8	+0.9737	+0.8333	+0.1404

Gains are over the same no-sharing arm, taken as the best value on the shared λ grid *on these same episodes*. That selection inflates every arm's absolute gain, the random control's included, so only the spatial-minus-random difference in the last column carries a claim. Sequence position remains the strongest single relation at $K = 24$ (+1.2719), and coarsening costs value monotonically in g.

A single draw of the control is not an interval, and the conjunction over six cells was the whole strength of the claim, so the control construction was redrawn five more times by re-running the same script under different seeds; the published seed reproduces its six differences exactly. The mean advantage is positive in every cell, from +0.048 to +0.206, and its mean exceeds one across-draw standard deviation in five of the six. The exception is $K = 12$, $g = 8$: mean +0.048 against a spread of 0.068, positive in five draws of six and negative in the sixth. This compares random-control constructions on fixed biological data; it is not an inferential interval or a significance test over biological units.

That last pattern refutes a reading we had entertained. Reach — the share of groups touching the selected set, times how many of their members sit near the boundary — rises with g (0.25, 0.59, 1.33 at $K = 24$) while gain falls, so reach alone does not order relations. What the table isolates instead is the other factor: size-matching makes the spatial and random arms nearly equal in reach (0.5873 against 0.5863 at $g = 4$, $K = 24$), while the average spatial-minus-random estimate remains positive. This is consistent with relation identity mattering beyond this opportunity summary; it does not isolate a unique causal factor or establish superiority over all alternative relations.

G.4 TRANSFER TO A THIRD CORPUS

MaveDB transfer fixes $\lambda = 0.375$, the modal development-corpus choice, without retuning. Grid maxima below are selection upper bounds, not the transfer result.

K	No sharing	Frozen λ	Gain	p	MDE
12	1.2378	1.5418	+0.3039	0.0048	0.2845
24	2.5047	3.0821	+0.5774	0.0072	0.6493

Best over the grid would give $+0.3237$ at $\lambda = 1$ and $+0.7756$ at $\lambda = 0.5$; those are upper bounds only for selecting among the declared constant-grid policies. The frozen-λ estimates are positive at both budgets but smaller than the corresponding nested development-corpus estimates. This supports sign transfer under the tested setting, not equivalence of effect sizes.

MaveDB was development-exposed, not untouched. Its 31 clusters offer fewer units, and the $K = 24$ paired-test gain lies below its SD-based MDE; these are different sensitivity criteria. The pre-search axis/structure manifest retains 64 score sets, with 9 hit sets re-derived after the shipped loader selected the wrong tail.

G.5 SKEMPI AFFINITY EXTENSION

We evaluate a frozen affinity-search extension on SKEMPI. This dataset was previously used during method development and is not untouched validation. Arms S/A/B/C/D are static ranking, neighborhood greedy/lookahead, and position-feedback greedy/lookahead. The fixed settings are slope 1, $\pi_0 = 0.05$, $\kappa = 2$, $\lambda = 0$ or 0.375, 50 neighbors, neighborhood strength 25 and shortlist 100; all episodes start without labels. Canonical position/WT/mutant order resolves ties.

SKEMPI matched task. A phenotype-blind census of SKEMPI2.0 (Jankauskaitė et al., 2019) retained four receptor contexts from one ovomucoid study (Lu et al., 2001): 171 single substitutions at nine sites each. Admission required $N \geq 48$, at least four sites with three alternatives, and at least half the candidates at such sites; pools match declared partners, mutated chain, study, temperature, method and WT affinity reference. All four also meet the $N \geq 96$ secondary-budget rule. Exact positive affinities give $\Delta\Delta G = RT \log(K_{d,\mathrm{mut}}/K_{d,\mathrm{WT}})$; hits include all variants at or below the $\lfloor 0.05N \rfloor$th value (eight hits, no boundary ties here). Unrecorded pH/buffer differences remain unresolved.

The two priors are unbound-partner mutation LLR and complex minus unbound-partner LLR, computed on identical retained coordinates with ProteinMPNN v_48_020 conditional probabilities (seed 37, no backbone noise). Entire biological partners are retained; only the mutated chain is scored (1CHO EFG:I; the other contexts E:I). These dimensionless preferences are not physical free energies. Each prior is standardized within pool and calibrated to mean 0.05; a degenerate prior is uniform. The graph is fixed across priors and budgets, using unbound LLR and BLOSUM62 substitution score. All outcomes follow; $I = (D - C) - (B - A)$ and means weight the four contexts equally. $K = 12$ is primary; $K = 24$ is secondary.

Pool	Prior	K	S	A	B	C	D	C−A	D−C	B−A	I
1CHO	unbound	12	3	2	2	4	4	2	0	0	0
1PPF	unbound	12	1	1	1	1	1	0	0	0	0
1R0R	unbound	12	0	0	0	0	0	0	0	0	0
3SGB	unbound	12	0	0	0	0	0	0	0	0	0
Mean	unbound	12	1.00	0.75	0.75	1.25	1.25	0.50	0.00	0.00	0.00
1CHO	cycle	12	2	4	3	5	5	1	0	-1	1
1PPF	cycle	12	4	3	3	6	6	3	0	0	0
1R0R	cycle	12	0	0	0	2	2	2	0	0	0
3SGB	cycle	12	0	0	0	0	0	0	0	0	0
Mean	cycle	12	1.50	1.75	1.50	3.25	3.25	1.50	0.00	-0.25	0.25
1CHO	unbound	24	4	5	5	6	6	1	0	0	0
1PPF	unbound	24	1	1	1	1	1	0	0	0	0
1R0R	unbound	24	1	0	0	2	3	2	1	0	1
3SGB	unbound	24	0	0	0	3	3	3	0	0	0
Mean	unbound	24	1.50	1.50	1.50	3.00	3.25	1.50	0.25	0.00	0.25
1CHO	cycle	24	4	6	5	6	6	0	0	-1	1
1PPF	cycle	24	6	6	6	6	6	0	0	0	0
1R0R	cycle	24	2	1	2	4	3	3	-1	1	-2
3SGB	cycle	24	0	0	0	2	2	2	0	0	0
Mean	cycle	24	3.00	3.25	3.25	4.50	4.25	1.25	-0.25	0.00	-0.25

At $K = 12$, C−A averages 0.50 and 1.50 hits for unbound and cycle priors; the corresponding $K = 24$ means are 1.50 and 1.25. Nine of the sixteen dependent settings improve and seven tie. All



Figure 6: **Relation controls and separate diagnostic evaluations.** (a) Position, random and score-bin partitions relative to neighborhood greedy at $K = 24$; circles denote Domainome, squares ProteinGym. Approximate pointwise normal intervals use archived MDE$/2.80$, not multiplicity-adjusted intervals. (b) Spatial-minus-own-control gains over six random-control draws: circle, mean; thick teal band, $\pm$SD; thin gray line, range; black tick, original draw. These measure construction variability, not biological uncertainty. (c) Count density of all $8{,}680$ dependent replacement records; logarithmic color denotes observations per bin. There are $1{,}858$ gained, $1{,}061$ lost and $5{,}761$ neutral replacements. (d) Paired unit gains: relation is position minus permuted relation; timing is selected minus always-position. Marker area counts coincident units; ticks mark equal-unit means.

$K = 12$ D$-$C contrasts are zero; its positive cycle interaction reflects a loss under neighborhood lookahead, not an absolute planning gain. At $K = 24$, 1R0R's D$-$C changes from $+1$ to -1 across priors. Thus the composite position update helps descriptively in these dense scans, without consistent planning benefit or isolation of position identity from every alternative relation. One scaffold/study, literature selection and prior exposure preclude inferential tests or scorer-independent generalization. No pool, seed, threshold or scorer was selected using this replay's outcomes; no confidence interval or p-value is reported.

Historical discovery records. Figure 5 retains historical assay gains and Figure 2 shows historical terminal exchanges. Their protocols are distinct from Table 1 and the nested panels of Figure 1.

Relation and replacement diagnostics.

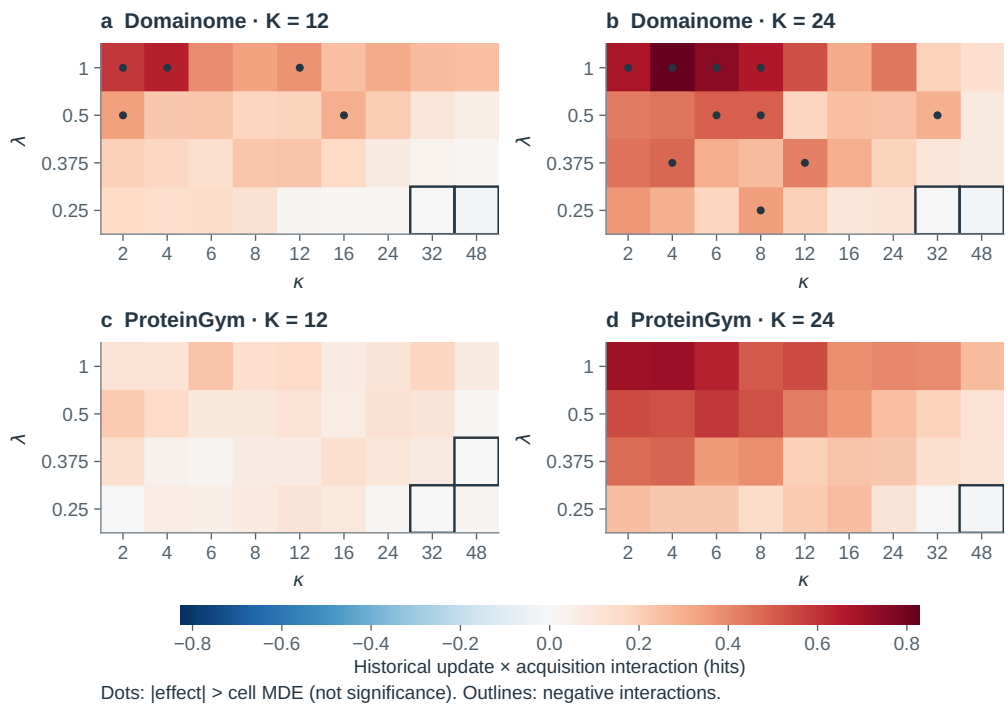

Figure 7: **Historical planning interactions across the complete parameter grid.** Every (κ, λ) cell is shown for both corpora and both budgets on one symmetric color scale. Dots mark absolute effects exceeding the archived cell-specific MDE; outlines identify negative effects. MDE flags are not significance tests. The four aggregate historical effects are displayed separately from the new unified comparison in the main planning figure.

H Frozen evaluation of a pre-continuation diagnostic

This pre-outcome frozen continuation analysis reuses admitted development-exposed corpora; it is not untouched validation and permits no outcome-dependent tuning.

Common paid history and fixed updates. Every policy first measures the six highest initial probabilities, with stable index ties. Those queries and their hits count toward $K = 24$. Each continuation then measures eighteen additional candidates greedily. We fix calibration slope to 1, prior prevalence to 0.05, neighborhood size to 50, prior weight to 25, and position parameters to $(\kappa, \lambda) = (2, 1)$. Calibration uses features only. We compare neighborhood-only updates, true-position updates, and three fixed size-preserving random permutations of position membership. Each partition reconstructs its own group evidence from the identical six labels. Seeds depend only on corpus, unit, pool identity and replicate index. The permutation controls do not match the full feature geometry. Continuations may diverge after the shared prefix; their contrast includes those downstream trajectory differences.

Information available to the diagnostic. The target is position-minus-neighborhood terminal hits. Baseline covariates comprise log pool size, log position count, initial top-24 mean probability, prefix hit fraction, remaining-candidate mean Bernoulli entropy, the neighborhood top-18 boundary margin, and mean group size. All are available before continuation. No initial predictive-accuracy statistic requiring hidden labels is supplied.

Four added diagnostics describe feedback on that same history. Prequential reliability is the mean position-minus-neighborhood Bernoulli log score on the six prefix actions; each action is predicted using only earlier labels. Replacement fraction compares the current top-18 sets of the two scores. Let these sets be S_1 and S_0, respectively, and let r^+ be the sigmoid of the position-adjusted score.

The opportunity diagnostic is

$$B(D_6) = \frac{1}{18}\left[\sum_{j\in S_1} r_j^+(D_6) - \sum_{j\in S_0} r_j^+(D_6)\right]. \tag{20}$$

Finally, sibling exposure is the fraction of remaining candidates sharing a position with any prefix query. These are observable score-based features, not a certified value of information or a new acquisition rule. They are also calculated under each permuted partition before its continuation labels are read.

Relation utility versus selective activation. Mean hits for permuted position, neighborhood, diagnostic-selected and always-position updates are respectively $4.306, 6.463, 6.751, 7.005$ on Domainome and $2.484, 2.795, 3.782, 3.808$ on ProteinGym. Table 11 retains every paired contrast and interval: only true-minus-permuted sharing passes correction. For ProteinGym selected-minus-position, only five of 39 units differ; the absolute-mean sign-flip test gives $p = 1.000$ despite a nonzero interval. Nondetection is not equivalence. The fixed $(2, 1)$ permuted arm cannot switch off; extended-grid selection instead gives Domainome $+0.0000/+0.1110$ at $K = 12/24$, with $65/67$ units selecting $\lambda = 0$ at $K = 12$, none at $K = 24$, and modal $\kappa = 24$ shrinking toward π_0. Appendix G.2's one-step veto does not bound terminal reward; its per-pool grid maximum bounds only constant-grid selection, not every selective-sharing policy.

The seven-covariate versus eleven-feature diagnostic changes held-out MSE from 8.320 to 8.325 and 7.730 to 7.608 (relative improvements $-0.066\%/+1.585\%$). Predicted-versus-realized unit-gain Spearman correlations are $-0.173/+0.090$. Signal strength, model specification and estimation uncertainty remain unresolved.

Frozen development and evaluation. Units are sorted by SHA256 of a fixed protocol string, corpus and unit name; the first $\lfloor 0.6N \rfloor$ develop the models ($40/27$ Domainome clusters and $58/39$ ProteinGym studies, development/evaluation). Separate corpus ridge regressions ($\alpha = 1$, no tuning) use either baseline covariates or the four added diagnostics. Each pool has inverse-development-unit-size weight. Centering, scaling and intercept use development data only; constant features receive scale one. At query six, either frozen predictor chooses position updates iff predicted gain is positive, otherwise neighborhood. The fixed zero threshold uses no evaluation outcomes, and both rules pay for the common prefix rather than selecting between known terminal rewards.

Endpoints and decision rule. Pool squared errors are averaged within units, then units equally; averaged residuals are not squared. Table 11 gives all four prespecified contrasts per corpus. Holm covers all eight two-sided sign-flip tests (20,000 deterministic sign vectors, plus-one correction); percentile intervals use $5{,}000$ whole-unit bootstraps. Inference assumes sign exchangeability and conditions on the fixed development fit and available cohort, not independent biological sampling.

The predictive-and-practical criterion requires both 10% MSE reduction and 0.5 extra hits at $K = 24$, each with Holm $p < 0.05$. Neither corpus meets it. Selected-minus-baseline-gate gains are $-0.046/0.000$ hits. Noncentral-t power under normal paired differences, standardized effect 0.6 and Bonferroni $\alpha = 0.05/8$ is approximately $0.56/0.80$ for $27/39$ units; these are sensitivity calculations, not exact sign-flip power. Diagnostic correlations and permutation repetitions are descriptive. Complete predictions, unit contrasts and source/contract hashes accompany the results.

Pools and statistical units. Assays retain separate searches and budgets. The 161 ProteinGym assays form 97 studies; 42 Tsuboyama assays together receive one study's weight. Evaluation contains 50 assays in 39 studies. Recurring proteins/homologs across studies and splits preclude an unseen-family claim.

I External static-ranking comparison

Both predeclared contrasts favor position greedy (Table 12). At $K = 24$, the paired study-mean difference is 1.204 hits with a percentile interval of $[0.571, 1.869]$. This supplies a concrete comparison with an established zero-shot predictor on the admitted pools; it is not a claim to outperform all available rankers.

Table 11: All prespecified held-out contrasts. MSE reduction is baseline minus augmented squared error; remaining contrasts are hits at $K = 24$. Intervals are marginal 95% whole-unit bootstrap intervals, conditional on the fixed development fits. Two-sided sign-flip p values assume sign exchangeability; Holm adjustment covers these eight tests.

Corpus	Contrast	Mean	95% interval	p	p_{Holm}
Domainome	MSE reduction	-0.006	$[-1.243, 1.393]$	0.99370	1.00000
	Selected − position	-0.254	$[-0.745, 0.230]$	0.33508	1.00000
	Selected − baseline gate	-0.046	$[-0.306, 0.148]$	0.81201	1.00000
	True − permuted	2.699	$[1.637, 3.947]$	0.00010	0.00080
ProteinGym	MSE reduction	0.123	$[-0.333, 0.547]$	0.59637	1.00000
	Selected − position	-0.026	$[-0.205, 0.128]$	1.00000	1.00000
	Selected − baseline gate	0.000	$[-0.179, 0.205]$	1.00000	1.00000
	True − permuted	1.323	$[0.658, 2.117]$	0.00025	0.00175

Table 12: External static-ranking comparison on the same 161 ProteinGym pools, averaged within studies and then equally across 97 studies. Positive differences favor position greedy. Confidence intervals are paired whole-study percentile intervals; Holm correction covers the two predeclared tests.

K	ESM-1v static	Position greedy	Δ hits [95% CI]	Holm p
12	1.420	1.877	+0.457 [0.082, 0.864]	0.02370
24	2.531	3.734	+1.204 [0.571, 1.869]	0.00070

A fixed model choice. We preselected the five-checkpoint ESM-1v ensemble (Meier et al., 2021), using the `ESM1v_ensemble` column of the official ProteinGym v1.3 substitution-score archive. The choice was frozen before inspecting its hit-discovery outcomes; we did not select among language models on these assays. The ensemble is the arithmetic mean of five masked-marginal scores. Larger scores rank first, with exact ties broken by the original candidate index. Ranking is computed once, so none of the measured fitness labels changes the order. Its top-12 and top-24 candidates are evaluated using the same hit labels and candidate subsets as the historical search experiments.

Alignment and provenance. Only the 161 historically admitted ProteinGym pools enter this comparison, even though the official archive covers 217 assays. We require unique mutation keys, exact mutated-sequence agreement with the original assay tables, and finite scores for every admitted candidate. Candidate features, ordered mutation keys and positive indices are checked against the historical pool fingerprints; all archived acquired labels are checked against those same pools. No missing-score candidate is silently dropped. Hash-bound per-assay records preserve selected indices, labels and cumulative hits. This comparison processes precomputed scores and does not rerun a protein model.

Estimation. As in the headline analysis, assay results are averaged within the original author–year studies, then the 97 studies receive equal weight. The two predeclared contrasts are position greedy minus ESM-1v static at $K = 12$ and $K = 24$. Two-sided paired sign-flip tests use $20{,}000$ draws with the plus-one correction, followed by Holm correction across these two contrasts. Percentile intervals use $5{,}000$ whole-study bootstrap resamples and remain conditional on the recorded historical configurations. The full available cohort fixes the sample size; 0.5 additional hits at $K = 24$ is a descriptive practical reference, not a post-hoc power guarantee. Repeated proteins and homologous sequences across studies can still induce dependence.

What the comparison identifies. ESM-1v changes the initial predictor as well as removing adaptation. It therefore tests an external discovery strategy, whereas matched static ranking isolates the value of using feedback from the same initial scores. The computational cost of pretrained models is outside the measurement budget for every method, and greater hit yield is not evidence of faster computation or laboratory turnaround. No external ESM-1v result is inferred for Domainome.

Table 13: Frozen boundary interventions. Equal-weight cluster/study means; pointwise whole-unit bootstrap intervals. Holm correction covers all four rows.

Corpus	Endpoint	Mean hits	95% CI	Holm p
Domainome	Attenuation − scalar	-0.051	[-0.155, 0.040]	0.6194
Domainome	Position fixed-query R	+0.357	[0.096, 0.668]	0.0572
ProteinGym	Attenuation − scalar	+0.099	[-0.066, 0.297]	0.6194
ProteinGym	Position fixed-query R	+0.163	[0.037, 0.294]	0.0572

The same feedback comparison from a different initial predictor. Because the headline holds initial scores fixed, it leaves open whether the gains repair a defect specific to the ProteinMPNN log-likelihood ratio. We therefore substitute the ESM-1v ensemble for the sequence channel and rerun the same arms, holding everything else: the structure-confidence channel, the neighborhood construction, the position update, the parameter grid, the units and the hit definition. The slope is refitted by the same leave-one-unit-out formula on the substituted channel.

The substitution carries its own check. The channel is standardized per pool and the calibration is monotone in it, so the static arm under the substitution is exactly the raw ESM-1v ranking and must reproduce the numbers above; it returns 1.4197 and 2.5306 against 1.420 and 2.531, and the run is written to refuse rather than report if it does not.

This transfer check uses historical calibration and a uniform grid-average endpoint, not the nested selected-configuration endpoint of Table 1. At $K = 12/24$, neighborhood greedy yields $1.4974/2.9261$ study-mean hits and position greedy, averaged over the declared grid, yields $1.8339/3.8711$. The within-ESM-1v update gains are therefore $+0.3365/ + 0.9449$ hits (Holm $p = 0.0027/0.00008$). These are not a matched quantitative comparison of gain sizes between ProteinMPNN and ESM-1v: the historical ProteinMPNN selected-policy gains have a different estimand and are not juxtaposed here.

Neighborhood lookahead changes yield by $-0.0410/ + 0.0048$ hits. The grid-averaged acquisition–update interactions are $+0.1601/ + 0.1390$ (Holm $p = 0.12272/0.35134$). The correction family jointly contains the two update gains and two interactions. Thus the positive update effect is also observed from ESM-1v starting scores; evidence for an acquisition interaction does not transfer under this test. Neither result removes the historical calibration or benchmark-reuse limitations.

J FROZEN BOUNDARY-ALLOCATION INTERVENTION

The complete cohort is unchanged: 245 Domainome pools in 67 clusters and 161 ProteinGym pools in 97 studies. These are development-exposed benchmarks, not a new independent confirmation sample. The $K = 24$ experiment fixes slope 1, $\pi_0 = 0.05$, $\kappa = 2$, position weight 0.375 (zero for the neighborhood update), 50 neighbors, neighborhood strength 25, and a 100-query lookahead shortlist. Outsiders retain immediate scores as in the native implementation. The contraction dose $1/2$, four primary contrasts and multiplicity family were frozen before production. There is no favorable-dose or favorable-pool selection.

Intervention definitions. For each query at a fixed paid history, contract hypothetical continuation scores toward their weighted mean:

$$r^{(i,y)}(\alpha) = \bar{r}_i + \alpha\big(r^{(i,y)} - \bar{r}_i\big), \qquad \alpha \in \{0, \tfrac{1}{2}, 1\}. \tag{21}$$

The mean, hence B_i and D_i, is unchanged at that history. Let $C_i(\alpha)$ be the resulting contingency term. We compare $B + D + C(1/2)$ with $B + D + \gamma C(1)$, where $\gamma = \sum_i C_i(1/2)/\sum_i C_i(1)$ on the same query shortlist (zero when its full mass is zero). This matches total contingency mass while changing its allocation across queries; later histories may diverge. It is a policy intervention, not a claim that the contracted branches form coherent joint posteriors. Native full and without-contingency trajectories are reproduced exactly against the earlier archive.

A complementary fixed-query test freezes the first six neighborhood-greedy queries and its seventh query. For each update, compare its 17-candidate batch selected from the mean branch with the

Table 14: Primary component-removal interactions at $K = 24$: $(\text{full} - \text{without component})_{\text{position}} - (\text{full} - \text{without component})_{\text{kNN}}$. Intervals are pointwise unit-paired bootstrap 95% intervals. Holm p values use their own four-test family across both corpora and components; intervals are not multiplicity adjusted.

Corpus	Removed	Interaction	95% interval	Holm p
Domainome	C	$+0.5578$	$[+0.1819, +0.9707]$	0.0218
Domainome	D	-0.0004	$[-0.3744, +0.3760]$	0.9980
ProteinGym	C	$+0.2646$	$[-0.0413, +0.5808]$	0.2013
ProteinGym	D	$+0.2323$	$[+0.0275, +0.4505]$	0.0901

batch selected after the seventh label, keeping both possible batches fixed before reading that label. Their hit difference R measures productive exchanges at equal total budget $6 + 1 + 17 = 24$. These committed batches are distinct from the fully adaptive trajectories.

Numerical and truth-access checks. At a matched history, contraction preserves the weighted branch mean and probability bounds. The scalar control matches total C on the same shortlist, not its variance or rank ordering. Negative C values within 10^{-10} numerical tolerance are clamped only for control construction; larger violations are rejected. A full mass at most 10^{-12} produces zero scalar bonus. Native endpoint scores retain their original arithmetic and ties. Committed batches break score ties by ascending candidate index. Each updater has its own mean-branch comparator. Both query-outcome batches are frozen before truth is read; realized incoming and outgoing hits are then counted separately. Semantic validators check branch probabilities, means, Top-17 sets, rewards, prefixes and native endpoint trajectories on every pool.

Inference and activity. Pool effects are averaged within the original units and units equally. Two-sided sign-flip tests use 20,000 draws with plus-one correction and require the stated unit-level sign-symmetry assumption. Pointwise percentile intervals use 5,000 whole-unit resamples; both remain conditional on these reused cohorts and settings. Endpoint-specific seeds derive from master seed 20260919 and the endpoint key. Practical reference effects were 0.5 adaptive hits and 0.25 committed-batch hits; these do not imply adequate power for every effect or make a nonsignificant result an equivalence test. All four primary tests remain in the family irrespective of activity. Attenuation and scalar controls change the next argmax in $234/245$ and $158/161$ pools, and exchange on average 1.288 and 1.992 terminal candidates under equal-unit weighting. Different actions therefore need not produce a clearly detectable difference in discovery.

Absolute adaptive yields. In the order full, without contingency, half-branch attenuation and matched-mass scalar, Domainome mean hits and pointwise 95% intervals are 7.702 [6.763, 8.655], 7.263 [6.323, 8.279], 7.600 [6.660, 8.573], 7.651 [6.753, 8.607]. ProteinGym mean hits and pointwise 95% intervals are 4.348 [3.579, 5.177], 4.075 [3.325, 4.872], 4.228 [3.455, 5.035], 4.129 [3.349, 4.938]. These whole-unit bootstrap intervals describe absolute yields, not paired contrasts or additional corrected tests.

Dependence sensitivity. Verified UniProt links transitively merge 97 ProteinGym studies into 83 components; average original effects within components, then equally across components. Descriptive adaptive and fixed-query contrasts are $+0.076$ (95% interval $[-0.067, 0.229]$) and $+0.160$ ($[0.015, 0.304]$). These do not replace frozen units/tests or eliminate homology dependence. Domainome retains its original sequence clusters.

J.1 Component and allocation evidence

Figure 8 separates hypothetical branch responses, component removal, and allocation interventions. Each comparison has its own protocol and inference family; these panels do not establish a causal chain. The complete component-arm yields are in Table 10.

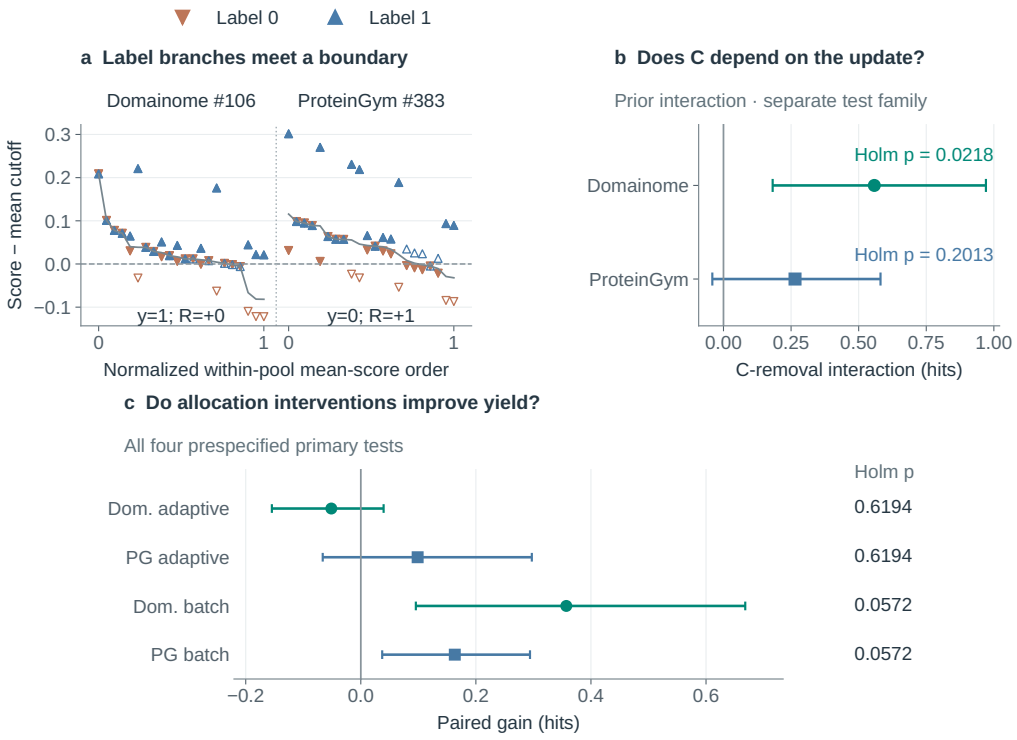

Figure 8: **Branch responses and the limits of the mechanism evidence.** (a) Hypothetical label branches in two representatives selected without outcome labels; filled markers identify selected candidates. (b) Component-removal interactions compare updates under fixed settings. (c) All four primary allocation effects: adaptive attenuation minus matched scalar and fixed-query position batch gain. Pointwise whole-unit bootstrap 95% intervals are distinct from multiplicity-adjusted tests; none of the four allocation tests passes its joint Holm correction. Branch examples, component removal and adaptive allocation answer separate questions.

K Common-prior feedback and matched planning evaluation

All 245 Domainome and 161 ProteinGym pools use $K = 24$ and $p_i = \mathrm{sigmoid}(z_{\ell,i} + b)$: slope one, with b chosen without labels for mean 0.05 (supplied prevalence). Domainome retains 67 clusters; ProteinGym studies sharing proteins link transitively into 83 components. Pools average within components, then components equally in tuning and evaluation. Each test component is fully excluded from selection; on/off arms separately maximize training greedy hits with deterministic ties. Development exposure, overlapping folds and residual dependence remain.

Update definitions and fixed grids. Composite updates use Equation 3 and the original neighborhood graph. Ridge and GP regress paid-label residuals about the shared prior: $r = p + K_{:,L}(K_{L,L} + \eta I)^{-1}(y_L - p_L)$. Ridge uses kernel $K_{ij} = 1 + z_{\ell,i}z_{\ell,j} + z_{p,i}z_{p,j} + \mathbf{1}\{a_i = a_j\} + w\mathbf{1}\{g(i) = g(j)\}$, where a_i is the mutant amino acid; GP uses $K_{ij} = \exp(-\|z_i - z_j\|^2/(2\ell^2)) + w\mathbf{1}\{g(i) = g(j)\}$. The two numerical features are standardized within each pool for these kernels. Scores receive only already purchased labels. With none purchased, every model returns p. These regression means are unbounded and used for greedy ranking only, not as ENS label weights.

Frozen grids (zero denotes the off arm) contain 175 settings per pool:

Update (settings)	Parameter	Values
Composite (37)	κ	$2, 4, 6, 8, 12, 16, 24, 32, 48$
	λ	$0.25, 0.375, 0.5, 1$; one $\lambda = 0$ arm
Ridge (50)	η	$1, 10, 100, 1000, 10000$
	w	$0, 0.25, 1, 4, 16, 64, 256, 1024, 4096, 16384$
GP (88)	w; fixed η	$0, 0.25, 1, 4, 16, 64, 256, 1024$; 0.25
	ℓ	$0.03125, 0.0625, 0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32$

No grid was expanded after results. On/off comparisons test tuning procedures, not identical model capacity.

Inference and sensitivity. The six-test family in Table 2 uses two-sided paired sign flips ($20{,}000$ draws) and pointwise whole-component bootstrap intervals. Domainome GP's adjusted $p = 0.0480$ is marginal. All four ridge/GP estimates exceed the 0.5-hit practical reference, but not all lower bounds do. Imposing the on-selected nuisance parameter on the off arm gives descriptive GP/ridge gains $1.053/1.875$ on Domainome and $0.897/1.815$ on ProteinGym, without additional tests. Domainome GP selects maximum site weight 1024 in $54/67$ components; attainable GP performance is not established.

Matched planning. Training-greedy-selected composite settings are reused for ENS with deterministic index shortlist ties. Neighborhood ENS gains are $-0.097/ + 0.035$ hits (Domainome/ProteinGym). All four primary tests below miss Holm correction; pointwise intervals imply neither zero benefit nor equivalence. These selected-policy estimates differ from historical grid averages.

Corpus	Endpoint	Hits [95% interval]	Holm p
Domainome	Position ENS–greedy	$+0.238$ $[-0.097, 0.567]$	0.22599
	Interaction	$+0.335$ $[-0.077, 0.750]$	0.22599
ProteinGym	Position ENS–greedy	$+0.591$ $[0.102, 1.123]$	0.08980
	Interaction	$+0.556$ $[0.067, 1.079]$	0.09555

The replay records retain $71{,}050$ greedy and 812 ENS episodes with all purchased candidate identities and labels. The summary, source snapshot and input manifest are hash-bound; independent checking reproduced the component exclusions, selected settings, unit means, intervals, correction families and recorded labels. This artifact-level check supports reproducibility of the new episodes, not independent biological validation.