DUALEVO: CHEMICAL-SPACE-CONTROLLED ONLINE EVOLUTION FOR DUAL-TARGET MOLECULAR GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Dual-target molecular generation seeks molecules that satisfy activity objectives for two biological targets while maintaining useful exploration of chemical space. Existing generative and feedback-learning methods can enrich high-value candidates, but continual generation also requires deciding how evaluated experience is accumulated and which structures are reused for updating. We introduce DualEvo, a chemical-space-controlled online evolution framework that combines fixed dual-target evaluation, cumulative experience archiving, reward-guided filtering, scaffold-balanced selection, and lightweight LoRA updates. Reward ranks candidate value, whereas scaffold balancing controls the structural composition of the high-value update set. Across six tasks, DualEvo attains the highest six-task mean unique predicted-success yield at the final 500-sample readout; relative to Reward-only, per-task means are higher on five of six tasks and successful-scaffold coverage is broader. Larger fixed-checkpoint readouts show more distinct successful molecules and scaffolds despite a lower successful-event frequency. Saved LoRA states can be reused for independent sampling and as fixed proposal models in downstream search.

1 INTRODUCTION

Dual-target molecular generation aims to design a single molecule that simultaneously satisfies activity objectives for two disease-relevant proteins. Compared with single-target generation, the problem must coordinate two activity objectives within one chemical structure while preserving synthetic accessibility, medicinal-chemistry properties, and useful coverage of chemical space. Prior work has approached this setting through pharmacophore or fragment combination, conditional generation, reward-driven optimization, transfer from target-specific models, and multi-objective search (Huang et al., 2023; Niu et al., 2025; Su et al., 2026).

Reward-driven optimization and replay of evaluated candidates can adapt a generator toward task constraints (Guo & Schwaller, 2024a; Loeffler et al., 2024; Guo et al., 2026). Continual online generation, however, also raises two linked questions: how evaluated experience should accumulate across stages, and which high-value structures should be reused for parameter updates. If update experience concentrates on a few high-scoring molecules or closely related scaffolds, the generator can repeatedly revisit a narrow high-value region, reducing the number of distinct successful structures found under a fixed sampling budget. Molecular-generation benchmarks accordingly measure chemical-space coverage and performance under finite evaluation budgets (Brown et al., 2019; Gao et al., 2022).

DualEvo addresses these questions by organizing generation, fixed dual-target evaluation, cumulative archiving, reward-guided selection, scaffold-balanced sampling of update experience, and LoRA adaptation into a continual loop. Each task maintains a cumulative archive and a lightweight adaptation state. At every stage, the current generator produces complete paired-tree sequences; decoded molecules are evaluated by fixed target-specific predictors; valid unique molecules enter the archive; and a high-value subset is selected for updating. Reward determines task-value priority, while scaffold round-robin controls the structural composition of the update set. The resulting adapter can be saved and reused independently of the archive.

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.

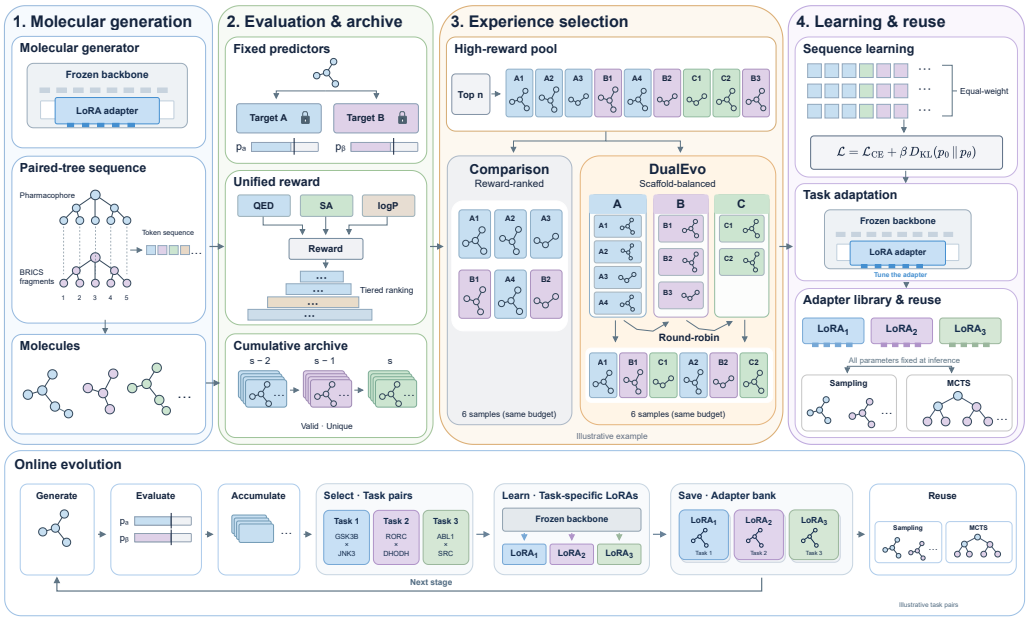

Figure 1: Each stage performs molecular generation, fixed dual-target evaluation, cumulative experience archiving, reward-guided/scaffold-balanced selection, and a LoRA update. The updated generator is used in the next stage, while the saved adapter can later be reused for fixed-generator sampling or as a proposal model in external search.

Our contributions are:

1. **Continual dual-target adaptation.** We introduce a chemical-space-controlled online evolution framework that connects generation, fixed dual-target evaluation, cumulative experience archiving, reward-guided/scaffold-balanced experience selection, and lightweight LoRA updates in a single loop.
2. **Explicit control of update-set structure.** We separate task-value ranking from scaffold composition: reward identifies high-value experience, and scaffold balancing redistributes which high-value structures enter the update while retaining complete paired-tree sequences as reusable training experience.
3. **Controlled mechanism and reuse analyses.** Across six dual-target tasks, we evaluate final independent sampling and structural coverage, and use controlled experiments on experience accumulation, reward direction, scaffold balancing, an alternative learning interface, larger sampling budgets, and downstream search to characterize the learned adaptation state.

2 Related Work

2.1 Dual-Target and Multi-Objective Molecular Generation

Dual-target molecular generation has been approached through structural constraints, joint-objective optimization, and cross-target knowledge reuse. AIxFuse combines pharmacophore fusion with dual-target structural feedback; POLYGON jointly guides exploration with predicted target activity, drug-likeness, and synthesizability; and DualDiff reprograms target-specific diffusion models for dual-target generation (Chen et al., 2024; Munson et al., 2024; Zhou et al., 2024). DiffDTM conditions on protein sequences, FuseDiff jointly models a ligand and two pocket-specific poses, and LaMGen generates protein-informed three-dimensional molecules (Huang et al., 2023; Wu et al., 2026; Su et al., 2026). REUSE searches the input-noise space of a frozen single-target diffusion

generator (Zeng et al., 2026). These approaches establish different routes to satisfying joint molecular constraints. DualEvo instead focuses on continual generation, in which evaluated experience persists across stages and directly shapes subsequent generator states.

2.2 Reward Feedback, Experience Reuse, and Diversity Control

Goal-directed molecular generation can modify the sampling distribution through policy gradients, generative reinforcement learning, Hill-Climb-style fitting of high-scoring samples, or replay of evaluated experience (Williams, 1992; You et al., 2018; Thomas et al., 2022; Guo & Schwaller, 2024a; Loeffler et al., 2024; Guo et al., 2026). REINVENT 4, Augmented Memory, and Saturn reuse historical evaluations, while diversity filters and memory-management mechanisms address repeated concentration on similar structures. FREED++ develops docking-reward reinforcement learning in a fragment-based generator (Telepov et al., 2023). GFlowNet and quality-diversity or multi-objective methods instead emphasize coverage of multiple high-value modes or Pareto regions (Zhu et al., 2023; Niu et al., 2025). LIMO optimizes molecular properties in a learned latent space (Eckmann et al., 2022). DualEvo combines cumulative experience reuse with an explicit separation between reward-based value ranking and scaffold-based control of the update-set composition.

2.3 Chemical-Space Search and Policy Reuse

A separate line of work searches discrete chemical spaces using graph edits, MCMC, MCTS, or genetic operators (Xie et al., 2021; Yang et al., 2017; Jensen, 2019). CombiMOTS performs dual-target multi-objective search with Pareto MCTS in a synthesizable fragment space (Southiratn et al., 2025). RxnFlow generates candidates along reaction-constrained synthetic pathways (Seo et al., 2025). These approaches maintain exploration state in candidate sets, populations, or search trees. DualEvo instead learns a reusable generator state; we test whether that state remains useful when loaded for fixed-checkpoint sampling or substituted as the proposal generator in an otherwise unchanged MCTS procedure.

3 Method

DualEvo maintains, for each task and stage s, a cumulative experience archive A_s and a LoRA adaptation state ϕ_s. The current generator samples complete paired-tree sequences; decoded molecules are evaluated by fixed predictors; valid unique molecules are added to the archive; and update sequences are chosen by reward ranking followed by scaffold-balanced selection. All selected sequences enter the same autoregressive objective, and the updated adapter is used at the next stage.

3.1 Problem Formulation and Paired-Tree Representation

Let the pretrained unconditional generator have frozen base parameters θ_0. For a complete paired-tree sequence $x = (x_1, \ldots, x_L)$,

$$p_{\theta_0}(x) = \prod_{k=1}^{L} p_{\theta_0}(x_k \mid x_{<k}).$$

For a dual-target task $\mathcal{T} = (t_a, t_b)$, stage s samples from the current policy $\pi_{\phi_{s-1}}$ using the fixed decoding settings and reconstructs each sequence as a molecule $m = D(x)$. Target-specific predictors evaluate the molecule, and the joint reward determines which archived sequences are eligible for later updates. The adapted model is

$$p_{\theta_0,\phi}(x) = \prod_{k=1}^{L} p_{\theta_0,\phi}(x_k \mid x_{<k}), \qquad x \sim \pi_{\phi_{s-1}}.$$

Each sequence encodes two trees with a shared ordered topology: a pharmacophore-organization tree and a BRICS fragment tree. Their leaf nodes are aligned, and the fragment tree reconstructs the complete molecule. Molecular properties and rewards are computed on canonical molecules,

whereas updates operate on the complete paired-tree token sequence. During online adaptation, the Transformer backbone remains frozen and rank-2 LoRA adapters are trained on the attention query and value projections (Vaswani et al., 2017; Hu et al., 2022). Additional pretraining and representation details are given in Appendix A.

3.2 Fixed Dual-Target Feedback and Persistent Experience

For target t_j, the activity predictor $f_j(m)$, decision threshold τ_j, and task-level property thresholds are fixed before generation begins. Predictors are trained from ChEMBL human single-protein IC50 records: after standardization, pChEMBL ≥ 6 is labeled active and pChEMBL ≤ 5 inactive; the intermediate interval is excluded. Data are split by Bemis-Murcko scaffold, model selection uses validation PR-AUC, and the operating threshold is selected using MCC (Bento et al., 2014; Mendez et al., 2019). Full predictor and threshold details are reported in Appendix B.

For task $\mathcal{T}$, let $q_{\mathcal{T}} = \max(q_a, q_b)$ and define the engineering gate

$$E_{\mathcal{T}}(m) = \mathbf{1}[\mathrm{QED}(m) \geq q_{\mathcal{T}} \wedge \mathrm{SA}(m) \leq s_{\mathcal{T}} \wedge -2 \leq \log P(m) \leq 6].$$

A molecule is a predicted dual-target success only when it passes the engineering gate and both fixed activity thresholds:

$$S_{\mathcal{T}}(m) = E_{\mathcal{T}}(m)\, \mathbf{1}[f_a(m) \geq \tau_a]\, \mathbf{1}[f_b(m) \geq \tau_b].$$

Invalid or undecodable samples have $S_{\mathcal{T}}(m) = 0$. For ranking, $C(m) = \mathbf{1}[f_a(m) \geq \tau_a] + \mathbf{1}[f_b(m) \geq \tau_b]$, and the reward tier is $h(m) = 2 + C(m)$ when $E_{\mathcal{T}}(m) = 1$ and $h(m) = 1$ otherwise. Within a tier, signed normalized activity margins and QED/SA/logP margins define

$$\Delta(m) = \mathrm{clip}\left(0.9\, \frac{\delta_a(m) + \delta_b(m)}{2} + 0.1\, \delta_{\mathrm{prop}}(m), -1, 1\right),$$

and the scalar ranking reward is

$$r(m) = \frac{h(m)}{4} + 0.1\Delta(m).$$

The reward is used for ranking, not as a calibrated probability. Exact margin definitions and the deterministic tie-breaking key are given in Appendix B. At stage s, valid molecules are canonicalized, merged with historical experience, and deduplicated to form A_s; each retained record stores the complete paired-tree sequence, canonical molecule, reward, and Bemis-Murcko scaffold.

3.3 Reward-Guided Scaffold-Balanced Experience Selection

At stage s, the deduplicated archive is sorted by the fixed reward key. The target update-set size is

$$N_s = \min(|A_s|, \lceil 0.25 \times 200s \rceil) = \min(|A_s|, 50s), \qquad s = 1, \ldots, 10.$$

Thus the update budget is tied to cumulative generation attempts rather than to 25% of the current archive. Let

$$M_s = \max(N_s, \lceil 0.5|A_s| \rceil), \qquad H_s = \mathrm{Top}_{M_s}(A_s; r).$$

Reward-only selects the top N_s records from H_s. DualEvo instead groups H_s by Bemis-Murcko scaffold, preserves reward order within each group, orders groups by their best record, and then selects one record from each nonempty group per round until N_s records are chosen:

$$B_s^{\mathrm{reward}} = \mathrm{Top}_{N_s}(H_s; r), \qquad B_s^{\mathrm{bal}} = \mathrm{RoundRobinScaffold}(H_s, N_s).$$

Reward-only and DualEvo use the same joint reward, high-value-pool construction rule, and update procedure; they differ only in the experience-selection rule, with DualEvo applying scaffold-balanced selection within the high-value pool at each stage. Strictly equal scaffold quotas are not enforced when groups are exhausted or the final round is incomplete.

3.4 Autoregressive Online Adaptation

All sequences in B_s contribute equally to next-token prediction. With $N = |B_s|$,

$$\mathcal{L}_{\mathrm{AR}}(\theta_s) = \frac{1}{N} \sum_i \mathrm{CE}(x_i; \theta_s),$$

$$\mathcal{L}_{\mathrm{KL}}(\theta_s) = \frac{1}{N} \sum_i \frac{1}{T_i} \sum_{t=1}^{T_i} \mathrm{KL}[p_{\theta_0}(\cdot \mid \mathrm{prefix}_{it}) \parallel p_{\theta_s}(\cdot \mid \mathrm{prefix}_{it})],$$

$$\mathcal{L}(\theta_s) = \mathcal{L}_{\mathrm{AR}}(\theta_s) + \beta \mathcal{L}_{\mathrm{KL}}(\theta_s), \qquad \beta = 0.05.$$

The adapter ϕ_s is initialized from ϕ_{s-1}. We use LoRA rank 2, alpha 4, five epochs per selected update set, batch size 4, and learning rate 2×10^{-4}. The main procedure has 10 stages with 200 generated candidate events per stage (2,000 candidate events total). The updated generator is used immediately at the next stage.

For a learning-interface control, REINFORCE-LoRA uses the same fixed evaluator, LoRA rank/alpha, and 2,000-candidate online budget but forms a policy gradient from scalar trajectory returns rather than fitting selected complete sequences. Full settings are given in Appendix F.

3.5 Saveable Generator State and External Proposal Interface

After each update, the frozen backbone plus the current LoRA adapter defines a saveable generator state. The adapter can later be loaded for independent sampling without the historical archive, or used as a fixed proposal generator in MCTS. Complete molecules proposed by external procedures such as genetic-algorithm crossover or mutation can also enter the same evaluation, archive, selection, and update pipeline.

4 Experimental Setup

4.1 Tasks, Predictors, and Comparison Methods

We evaluate six target pairs: GSK3B-JNK3, RORC-DHODH, ABL1-SRC, FGFR1-KDR, EGFR-ERBB2, and ESR1-ESR2. All conditions start from the same base-generator checkpoint and use the same frozen target-specific classifiers and predefined success thresholds. Each condition is run with three independent random seeds.

The main comparison includes the unadapted frozen generator (Frozen), one-shot supervised adaptation (One-shot), reward-ranked online adaptation (Reward-only), and DualEvo. One-shot uses the same frozen-source budget of 2,000 candidate events, selects the top 25% high-reward valid unique experiences once, and matches the total training exposure used by the online conditions. Reward-only and DualEvo generate from the current model stage by stage and use the same reward, high-value-pool construction rule, and update procedure; DualEvo differs only in applying scaffold-balanced selection within each stage's high-value pool.

4.2 Training and Evaluation Protocol

Online conditions run for 10 stages of 200 generated candidate events each. Direct autoregressive sampling uses temperature 0.8 and top-$k = 10$. After training, parameters are frozen and 500 independent samples are drawn for the primary final-checkpoint readout; a separate 5,000-sample

Table 1: Final fixed-generator readout across six tasks. Entries are mean ± sample SD of unique predicted-success yield (%) over three seeds from 500 independent samples.

Target pair	Frozen	One-shot	Reward-only	DualEvo
GSK3B–JNK3	19.2 ± 0.5	48.3 ± 1.3	65.5 ± 3.7	67.7 ± 1.6
RORC–DHODH	48.2 ± 2.6	67.9 ± 2.8	77.4 ± 1.7	78.5 ± 4.1
ABL1–SRC	41.4 ± 1.6	74.9 ± 4.8	72.1 ± 3.7	77.6 ± 5.5
FGFR1–KDR	40.3 ± 0.7	73.8 ± 3.5	73.5 ± 6.6	76.3 ± 2.0
EGFR–ERBB2	44.4 ± 3.1	76.0 ± 1.9	73.6 ± 5.2	80.2 ± 3.2
ESR1–ESR2	11.5 ± 2.1	35.5 ± 1.8	57.1 ± 0.8	54.1 ± 7.7

readout probes larger sampling budgets. Readout samples are never written back to the archive and never trigger updates. Results are mean ± sample standard deviation over three seeds.

For a fixed-checkpoint budget B, let $\mathcal{U}_B$ be the set of unique canonical predicted-success molecules and $g(m)$ the Bemis-Murcko scaffold. We report: (i) predicted-success sampling rate $H_B = B^{-1} \sum_i S(m_i)$, which retains repeated successful events; (ii) unique predicted-success yield $Y_B = |\mathcal{U}_B|/B$; (iii) successful-scaffold count $K_B = |\{g(m) : m \in \mathcal{U}_B\}|$; and (iv) Top-10 successful-scaffold concentration, the fraction of $\mathcal{U}_B$ covered by the ten most frequent scaffolds (or fewer when fewer than ten occur).

4.3 Mechanism Ablations and Reuse Tests

On GSK3B-JNK3 and RORC-DHODH, we compare one-shot, stage-local, and cumulative experience updates; high-reward, random, low-reward, and shuffled reward-to-experience selection; Reward-only versus scaffold-balanced selection; and autoregressive sequence fitting versus REINFORCE-LoRA. Saved checkpoints are additionally evaluated with 5,000 fixed-generator samples and as fixed proposal generators under the same MCTS configuration (500 batch calls, four proposals per call, no generator updates during search). Genetic-algorithm proposal experiments are reported in the appendix.

5 Results

5.1 Higher Mean Final Independent-Sampling Yield Across Six Tasks

With 500 independent samples from the final fixed generator, the six-task equally weighted unique predicted-success yield is 72.4% for DualEvo, compared with 34.2% for Frozen, 62.7% for One-shot, and 69.9% for Reward-only (Table 1). DualEvo therefore has the highest six-task mean in this controlled comparison. Relative to Reward-only, the per-task mean is higher on five of six tasks; ESR1-ESR2 is the exception (54.1% versus 57.1%). These final readouts show that the task adaptation acquired online persists after training, while the Reward-only/DualEvo comparison provides a controlled test of scaffold-composition control in the online update loop. Paired seed-level differences are provided in Appendix Figure S7.

5.2 Controlled Analysis of Online-Evolution Components

Cumulative experience. With the base checkpoint, training-exposure schedule, and online budget held fixed, cumulative reward-based experience reaches unique predicted-success yields of 65.5% and 77.4% on GSK3B-JNK3 and RORC-DHODH, compared with 48.3%/67.9% for One-shot and 38.9%/41.7% for stage-local updating (Table 2; Table S4). The comparison is consistent with later updates benefiting from access to selectable historical feedback rather than only the current stage.

Reward direction. Shuffling the reward-to-experience correspondence reduces unique predicted-success yield to 26.5% and 48.6%, versus 65.5% and 77.4% under correct high-reward ranking. Random and low-reward selection are also lower (Table 2; Table S4). Thus, repeating the same sequence-update procedure without task-aligned experience selection does not reproduce the high-reward condition.

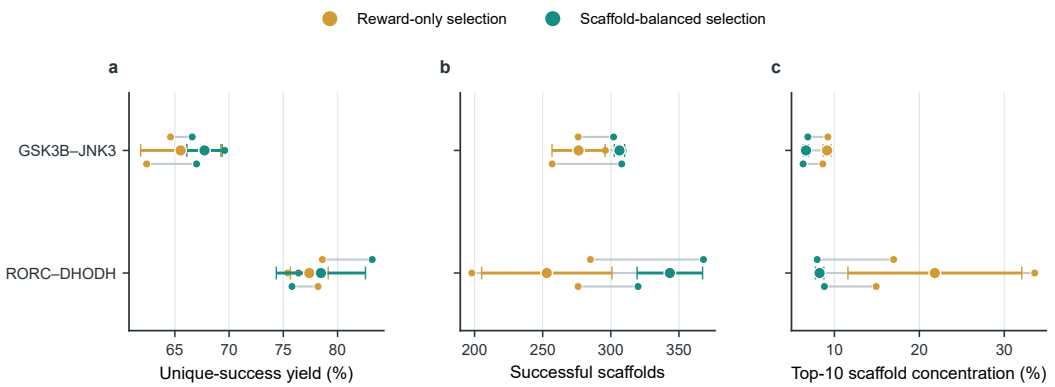

Figure 2: For GSK3B-JNK3 and RORC-DHODH, the figure reports unique predicted-success yield, successful-scaffold count, and Top-10 successful-scaffold concentration. Points are individual seeds; diamonds and error bars show mean ± sample standard deviation ($n = 3$).

Table 2: Controlled comparisons on GSK3B–JNK3 and RORC–DHODH. Yield differences are in percentage points (pp); complete values appear in Tables S4–S5.

Mechanism	Right versus left condition	GSK3B–JNK3	RORC–DHODH
Cumulative experience	Cumulative reward vs. one-shot	+17.2	+9.5
Reward direction	Correct reward vs. shuffled	+39.0	+28.8
Scaffold balancing	DualEvo vs. Reward-only	+2.2	+1.1
Learning interface	DualEvo vs. REINFORCE-LoRA	+15.4	+5.1

Structural coverage for DualEvo versus Reward-only: successful scaffolds increase by 30.0 and 90.3; Top-10 scaffold concentration decreases by 3.08 and 13.78 pp, respectively.

Scaffold-composition control. Holding the reward definition, high-value-pool construction rule, and update procedure fixed, scaffold balancing changes GSK3B-JNK3/RORC-DHODH unique yield from 65.5%/77.4% to 67.7%/78.5%, successful-scaffold count from 276.3/253.0 to 306.3/343.3, and Top-10 successful-scaffold concentration from 10.96%/23.07% to 7.88%/9.29% (Figure 2; Table 2). The larger changes in scaffold count and concentration than in unique yield support the interpretation that scaffold balancing primarily redistributes structural coverage within the high-value region.

Learning interface. DualEvo reaches 67.7% and 78.5% unique predicted-success yield, compared with 52.3% and 73.3% for REINFORCE-LoRA; the paired difference is positive for all six seed/task pairs (Table 2; Table S5). Under the same fixed evaluator and 2,000-candidate budget, the result shows that the update interface materially affects how evaluated experience is converted into the generator state in this implementation.

5.3 Distinct Discovery at Larger Fixed Sampling Budgets

At 5,000 fixed-checkpoint samples, Reward-only has the higher six-task mean predicted-success sampling rate (85.1% versus 80.7%), whereas DualEvo has the higher mean unique predicted-success yield (55.2% versus 48.4%). DualEvo also produces more successful scaffolds on average (1969.7 versus 1533.3) and a lower Top-10 successful-scaffold concentration (4.4% versus 8.9%; Table S3). Thus, more frequent successful events do not necessarily translate into more distinct successful structures.

Figure 3 tracks cumulative discovery from 100 to 5,000 samples on the two mechanism tasks. Under the same fixed sampling budgets, the DualEvo checkpoints continue to accumulate more distinct

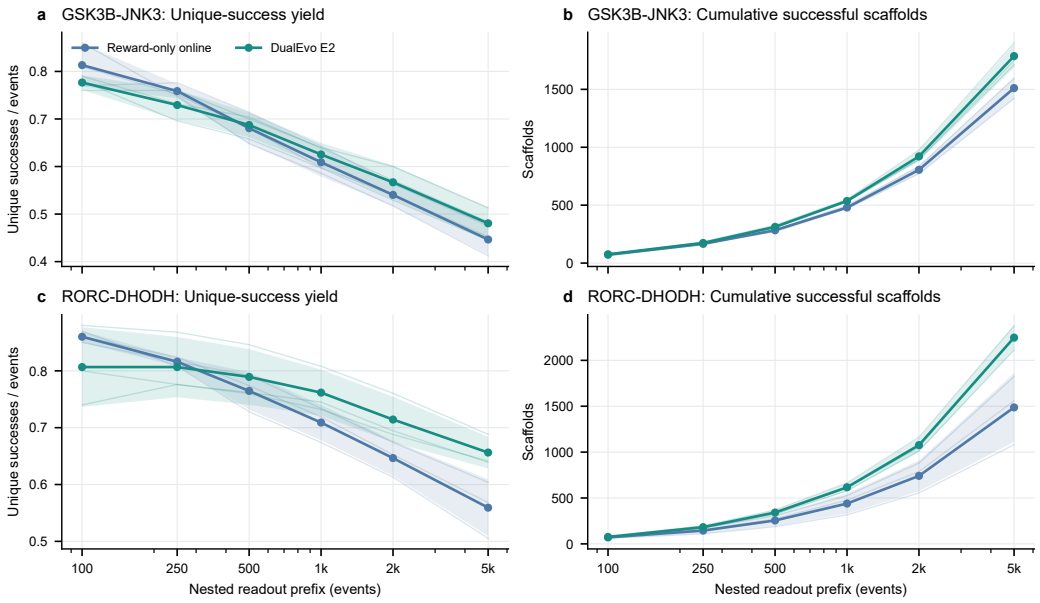

Figure 3: For GSK3B-JNK3 and RORC-DHODH, unique predicted-success molecules and successful scaffolds are counted over prefixes of the same ordered 5,000-sample streams. Curves show the mean across three seeds; error bars show sample standard deviation. Reward-only and DualEvo are the primary comparison, with Frozen and One-shot shown as context.

successful molecules and scaffolds, while Reward-only more often returns repeated successes from already favored regions. This separation between hit frequency and distinct structural discovery is the main evidence for the claimed chemical-space control.

5.4 Saved Adaptation States Support Subsequent Sampling and Search

Loading a saved LoRA adapter onto the frozen backbone enables direct task-matched sampling without revisiting the experience archive. In 5,000-sample readouts, the GSK3B-JNK3 and RORC-DHODH adapters reach unique predicted-success yields of 48.1% and 65.6% and successful-scaffold counts of 1787.3 and 2248.0, respectively (Figure S4; Table S9). Cross-task readouts are reported in the same appendix section.

When the saved DualEvo generator replaces the frozen generator as the fixed proposal model under an otherwise unchanged MCTS procedure, the mean unique-success endpoint rises from 193.7 to 632.7 on GSK3B-JNK3 and from 672.3 to 760.7 on RORC-DHODH; successful-scaffold means rise from 144.3 to 491.0 and from 518.7 to 542.0 (Figure S1; Table S6). Variability is substantial for the RORC-DHODH adapted-generator endpoint, so these results are best interpreted as evidence that the learned generator state can be reused as a task-adapted proposal model, rather than as a standalone claim about MCTS performance. Genetic-algorithm proposal experiments are reported in Appendix H.

6 Discussion

DualEvo separates two functions that are often coupled in replay-based optimization: reward ranking determines which archived molecules are high-value for the task, whereas scaffold balancing determines how structurally concentrated the selected update set is. Across the controlled mechanism experiments, cumulative high-reward experience is associated with higher final unique predicted-success yield, while scaffold balancing produces its clearest effect on successful-scaffold count and

concentration. The 5,000-sample readout further shows that these quantities need not move with successful-event frequency: Reward-only hits the success region more often, but DualEvo discovers more distinct successful molecules and scaffolds under the same sample count.

The saved LoRA state also retains task adaptation outside the online loop. It can be loaded for fixed-generator sampling without the archive and can replace the proposal generator in an external MCTS procedure without changing the search configuration. These results support viewing the adapter as a compact, reusable record of task-directed generator adaptation rather than only as an intermediate state of the training loop.

6.1 Limitations

All primary endpoints use activity predictors and medicinal-chemistry thresholds fixed before generation, so they characterize behavior within a computational evaluation environment rather than experimentally validated dual-target activity. Representative docking poses are therefore kept in Appendix Figure S8 and are presented only as computational structural views; binding and functional activity require experimental validation. Mechanism ablations and downstream reuse tests are concentrated on GSK3B-JNK3 and RORC-DHODH, and structural control is defined through Bemis-Murcko scaffolds. Broader target pairs, structural representations, and search settings remain to be tested.

7 Conclusion

DualEvo combines fixed dual-target evaluation, cumulative experience archiving, reward-guided filtering, scaffold-balanced selection, and lightweight LoRA updates in a continual generation loop. Across six tasks it achieves the highest mean final unique predicted-success yield among the controlled baselines, while scaffold balancing broadens successful-structure coverage and reduces concentration in the larger fixed-checkpoint readout. The mechanism controls and reuse experiments further show how task-aligned historical experience, update-set structure, and the learning interface shape a reusable adapted generator state.

Reproducibility Statement

The main experiments use three independent random seeds. Predictor construction and fixed thresholds are described in Section 3.2 and Appendix B; reward and selection rules in Sections 3.2-3.3; update hyperparameters in Section 3.4 and Appendix A; evaluation metrics and budgets in Section 4; and complete seed-level or endpoint results for the mechanism and reuse experiments in the appendix.

AI Use Statement

The authors used generative AI tools for English translation and language editing of an author-written Chinese manuscript, retrieval and discovery of related work, manuscript organization and reference formatting, and redesign and plotting of figures from recorded experimental results. All AI-assisted text and figures were reviewed against the source manuscript and experimental records. The authors take responsibility for the final text, claims, figures, data presentation, and citations.

References

A. P. Bento et al. The ChEMBL bioactivity database: an update. *Nucleic Acids Research*, 42: D1083–D1090, 2014. doi: 10.1093/nar/gkt1031.

Nathan Brown, Marco Fiscato, Marwin H. S. Segler, and Alain C. Vaucher. GuacaMol: Benchmarking models for de novo molecular design. *Journal of Chemical Information and Modeling*, 59(3): 1096–1108, 2019. doi: 10.1021/acs.jcim.8b00839.

S. Chen et al. Structure-aware dual-target drug design through collaborative learning of pharmacophore combination and molecular simulation. *Chemical Science*, 15:10366–10380, 2024. doi: 10.1039/D4SC00094C.

Peter Eckmann, Kunyang Sun, Bo Zhao, Mudong Feng, Michael Gilson, and Rose Yu. LIMO: Latent inceptionism for targeted molecule generation. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 5777–5792, 2022.

Wenhao Gao, Tianfan Fu, Jimeng Sun, and Connor W. Coley. Sample efficiency matters: A benchmark for practical molecular optimization. In *Advances in Neural Information Processing Systems 35, Datasets and Benchmarks Track*, 2022.

J. Guo and P. Schwaller. Augmented memory: Sample-efficient generative molecular design with reinforcement learning. *JACS Au*, 4(6):2160–2172, 2024a. doi: 10.1021/jacsau.4c00066.

J. Guo and P. Schwaller. Saturn: Sample-efficient generative molecular design using memory manipulation. arXiv:2405.17066, 2024b.

J. Guo, J. Chen, A. GX-Chen, and P. Schwaller. Sample-efficient generative molecular design using memory manipulation. *Nature Machine Intelligence*, 8:449–460, 2026. doi: 10.1038/s42256-026-01200-4.

E. J. Hu et al. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Lei Huang et al. DiffDTM: A conditional structure-free framework for bioactive molecules generation targeted for dual proteins. arXiv:2306.13957, 2023. URL `https://arxiv.org/abs/2306.13957`.

S. Ishida et al. ChemTSv2: Functional molecular design using de novo molecule generator. *WIREs Computational Molecular Science*, 13:e1680, 2023. doi: 10.1002/wcms.1680.

J. H. Jensen. A graph-based genetic algorithm and generative model/monte carlo tree search for the exploration of chemical space. *Chemical Science*, 10:3567–3572, 2019. doi: 10.1039/C8SC05372C.

H. H. Loeffler et al. REINVENT 4: Modern AI-driven generative molecule design. *Journal of Cheminformatics*, 16:20, 2024. doi: 10.1186/s13321-024-00812-5.

D. Mendez et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Research*, 47:D930–D940, 2019. doi: 10.1093/nar/gky1075.

B. P. Munson et al. De novo generation of multi-target compounds using deep generative chemistry. *Nature Communications*, 15:3636, 2024. doi: 10.1038/s41467-024-47120-y.

Yifan Niu, Ziqi Gao, Tingyang Xu, Yang Liu, Yatao Bian, Yu Rong, Junzhou Huang, and Jia Li. InversionGNN: A dual path network for multi-property molecular optimization. In *International Conference on Learning Representations*, 2025.

Seonghwan Seo, Minsu Kim, Tony Shen, Martin Ester, Jinkyoo Park, Sungsoo Ahn, and Woo Youn Kim. Generative flows on synthetic pathway for drug design. In *International Conference on Learning Representations*, 2025.

T. Southiratn, B. Koo, Y. Lu, and S. Kim. CombiMOTS: Combinatorial multi-objective tree search for dual-target molecule generation. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *PMLR*, pp. 56650–56691, 2025.

Qun Su, Qiaolin Gou, Hui Zhang, et al. LaMGen: Llm-based 3d molecular generation for multi-target drug design. *Nature Communications*, 17:5091, 2026. doi: 10.1038/s41467-026-71737-w.

Alexander Telepov, Artem Tsypin, Kuzma Khrabrov, et al. FREED++: Improving rl agents for fragment-based molecule generation by thorough reproduction. *Transactions on Machine Learning Research*, 2023. URL `https://openreview.net/forum?id=YVPb6tyRJu`.

M. Thomas, N. M. O'Boyle, A. Bender, and C. de Graaf. Augmented hill-climb increases reinforcement learning efficiency for language-based de novo molecule generation. *Journal of Cheminformatics*, 14:68, 2022. doi: 10.1186/s13321-022-00646-z.

A. Vaswani et al. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, 2017.

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992. doi: 10.1007/BF00992696.

Jianliang Wu, Anjie Qiao, Zhen Wang, Zhewei Wei, and Sheng Chen. FuseDiff: Symmetry-preserving joint diffusion for dual-target structure-based drug design. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026. doi: 10.1145/3770855.3819050.

Y. Xie et al. MARS: Markov molecular sampling for multi-objective drug discovery. In *International Conference on Learning Representations*, 2021.

X. Yang et al. ChemTS: An efficient python library for de novo molecular generation. *Science and Technology of Advanced Materials*, 18:972–976, 2017. doi: 10.1080/14686996.2017.1401424.

J. You et al. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems 31*, 2018.

Qingyuan Zeng, Pengxiang Cai, Zixin Guan, Ziyang Chen, Anglin Liu, Xinyao Lai, and Jintai Chen. Don't retrain, just reuse: Recovering dual-target molecules from single-target diffusion models. arXiv:2605.25681, 2026. URL `https://arxiv.org/abs/2605.25681`.

X. Zhou et al. Reprogramming pretrained target-specific diffusion models for dual-target drug design. In *Advances in Neural Information Processing Systems 37*, 2024.

Y. Zhu et al. Sample-efficient multi-objective molecular optimization with GFlowNets. In *Advances in Neural Information Processing Systems 36*, 2023.

A Generator and Paired-Tree Representation

The base model is a GPT-like autoregressive Transformer molecular generator. The checkpoint used for adaptation was pretrained on ChEMBL molecular data and then distilled on fragment-tree sequences generated by the model; the filtered distillation set contains 756,568 tree sequences. The working vocabulary contains 100 tokens and the maximum context length is 256. Direct generation uses temperature=0.8 and top-k=10. During adaptation, the Transformer backbone is frozen and LoRA with rank=2 and alpha=4 is applied only to the attention query and value projection layers. The model is updated after every 200 candidate samples, with five training epochs per update, batch size 4, learning rate 2×10^{-4}, and KL coefficient 0.05.

The paired-tree representation jointly encodes pharmacophore relationships and BRICS fragment connectivity using a shared ordered topology. Sequence leaf nodes correspond across the two trees, and the fragment tree can be reconstructed into a molecule. Rewards are computed from canonical molecules, while sequence updates operate on paired-tree tokens.

B Activity-Predictor Training and Fixed Thresholds

Target classifiers use ChEMBL human single-protein IC50 records. After parent standardization and aggregation by canonical molecule, pChEMBL$\geq$6 and pChEMBL$\leq$5 define the two classes, and data are split by Bemis-Murcko scaffold. At evaluation time, target-specific activity thresholds and task-specific QED and SA thresholds are fixed in advance. For an activity probability p, the margin above the threshold is (p-τ)/(1-τ), and the margin below the threshold is (p-τ)/τ; both are clipped to [-1,1]. The two target-specific QED margins are (QED-q$_j$)/(1-q$_j$) for $j \in \{a, b\}$, the SA margin is (s-SA)/s, and the logP margin is min(logP+2,6-logP)/8. The logP term is scaled by the full width of the [-2,6] interval, so its maximum at the center of the interval is 0.5. The four property

margins are clipped individually and then averaged. The final joint reward and ranking key are given in Section 3.2.

Table S1a | Final classifiers, fixed activity thresholds, and QED lower bounds for targets appearing in the six tasks. For the four V2 extension tasks, the two QED slots use the same pair-level threshold for the corresponding task. The number of labeled molecules is the total number of active and inactive molecules in the final training dataset.

Target	Final classifier	Activity threshold / QED lower bound	Labeled molecules
ABL1	RBF-SVM (ECFP)	0.7000 / 0.4120	1783
DHODH	random_forest	0.6783 / 0.5138	1501
EGFR	Random forest (ECFP)	0.6500 / 0.3775	8739
ERBB2	RBF-SVM (ECFP)	0.6050 / 0.3775	1977
ESR1	RBF-SVM (ECFP)	0.6850 / 0.4737	2772
ESR2	Random forest (ECFP)	0.4950 / 0.4737	1300
FGFR1	Random forest (ECFP)	0.5600 / 0.4011	3142
GSK3B	Logistic regression (ECFP)	0.4560 / 0.5649	2455
JNK3	Logistic regression (ECFP)	0.4467 / 0.4737	1193
KDR	RBF-SVM (ECFP)	0.8400 / 0.4011	7939
RORC	Logistic regression (ECFP)	0.5747 / 0.4068	7813
SRC	RBF-SVM (ECFP)	0.5200 / 0.4120	2867

Table S1b | Predefined SA and logP constraints for the six dual-target tasks.

Dual-target pair	SA upper bound	logP range
GSK3B–JNK3	2.7178	[-2, 6]
RORC–DHODH	3.5699	[-2, 6]
ABL1–SRC	2.9176	[-2, 6]
FGFR1–KDR	2.8408	[-2, 6]
EGFR–ERBB2	3.0761	[-2, 6]
ESR1–ESR2	3.2304	[-2, 6]

C Complete Metrics from 500 Samples

Table S2 summarizes the main endpoints and successful-structure coverage for four generator states across the six tasks under 500 independent samples. Unique predicted-success yield uses the total number of samples as the denominator and deduplicates by canonical molecule; predicted-success sampling rate retains repeated successful events. Successful-scaffold count and Top-10 successful-scaffold concentration are both computed from unique canonical predicted-success molecules.

Table S2 | 500-event fixed-generator readout across six tasks. Values are mean ± sample standard deviation, n=3. Top-10 successful-scaffold concentration is computed from the unique canonical predicted-success molecules for each seed.

Dual-target pair	Generator state	Unique success yield	Predicted-success sampling rate	Successful scaffolds	Top-10 successful-scaffold concentration
GSK3B–JNK3	Frozen generator	19.2 ± 0.5%	19.3 ± 0.5%	92.3 ± 2.1	14.20 ± 3.09%
GSK3B–JNK3	One-shot supervised adaptation	48.3 ± 1.3%	51.1 ± 1.2%	218.7 ± 7.6	12.68 ± 1.21%
GSK3B–JNK3	Reward-ranked online adaptation	65.5 ± 3.7%	81.7 ± 0.5%	276.3 ± 19.5	10.96 ± 0.71%
GSK3B–JNK3	DualEvo	67.7 ± 1.6%	78.9 ± 4.8%	306.3 ± 3.8	7.88 ± 0.48%
RORC–DHODH	Frozen generator	48.2 ± 2.6%	48.3 ± 2.7%	232.3 ± 12.7	7.75 ± 0.48%
RORC–DHODH	One-shot supervised adaptation	67.9 ± 2.8%	69.9 ± 3.4%	299.3 ± 6.5	11.97 ± 1.77%
RORC–DHODH	Reward-ranked online adaptation	77.4 ± 1.7%	89.9 ± 2.0%	253.0 ± 47.8	23.07 ± 10.85%
RORC–DHODH	DualEvo	78.5 ± 4.1%	82.7 ± 3.1%	343.3 ± 24.0	9.29 ± 0.76%
ABL1–SRC	Frozen generator	41.4 ± 1.6%	41.5 ± 1.7%	199.0 ± 9.5	8.72 ± 1.63%
ABL1–SRC	One-shot supervised adaptation	74.9 ± 4.8%	79.4 ± 3.3%	324.7 ± 22.4	12.02 ± 2.25%
ABL1–SRC	Reward-ranked online adaptation	72.1 ± 3.7%	89.5 ± 2.7%	267.7 ± 46.5	15.84 ± 5.98%
ABL1–SRC	DualEvo	77.6 ± 5.5%	86.5 ± 4.1%	331.3 ± 22.7	10.40 ± 0.40%
FGFR1–KDR	Frozen generator	40.3 ± 0.7%	40.4 ± 0.8%	193.3 ± 5.7	8.95 ± 1.44%
FGFR1–KDR	One-shot supervised adaptation	73.8 ± 3.5%	75.3 ± 3.8%	341.3 ± 16.2	7.55 ± 1.51%
FGFR1–KDR	Reward-ranked online adaptation	73.5 ± 6.6%	85.2 ± 5.1%	298.0 ± 49.8	12.26 ± 3.53%
FGFR1–KDR	DualEvo	76.3 ± 2.0%	83.9 ± 2.5%	338.3 ± 19.4	7.54 ± 1.89%
EGFR–ERBB2	Frozen generator	44.4 ± 3.1%	44.5 ± 3.2%	214.7 ± 18.0	7.88 ± 2.29%
EGFR–ERBB2	One-shot supervised adaptation	76.0 ± 1.9%	80.9 ± 1.6%	327.3 ± 13.8	12.92 ± 1.69%
EGFR–ERBB2	Reward-ranked online adaptation	73.6 ± 5.2%	91.8 ± 0.9%	279.0 ± 64.4	17.59 ± 9.96%
EGFR–ERBB2	DualEvo	80.2 ± 3.2%	91.5 ± 1.6%	356.7 ± 21.4	7.49 ± 0.93%
ESR1–ESR2	Frozen generator	11.5 ± 2.1%	11.5 ± 2.1%	56.7 ± 9.3	18.77 ± 1.52%

Dual-target pair	Generator state	Unique success yield	Predicted-success sampling rate	Successful scaffolds	Top-10 successful-scaffold concentration
ESR1–ESR2	One-shot supervised adaptation	35.5 ± 1.8%	36.9 ± 1.7%	160.7 ± 0.6	14.17 ± 2.51%
ESR1–ESR2	Reward-ranked online adaptation	57.1 ± 0.8%	71.3 ± 4.6%	202.0 ± 16.8	20.50 ± 3.76%
ESR1–ESR2	DualEvo	54.1 ± 7.7%	58.4 ± 8.4%	236.3 ± 35.7	11.07 ± 1.74%

D 5,000-Sample Fixed-Checkpoint Evaluation

The fixed-sampling evaluation uses the saved checkpoint from each condition, draws 5,000 samples, and reevaluates them with the same predicted reward. Table S3 reports predicted-success sampling rate, unique success yield, successful scaffolds, and scaffolds among all valid molecules, capturing both repeated hits and structural coverage.

Table S3 | 5,000-event fixed-checkpoint readout across six tasks. Values are mean ± sample standard deviation, n=3. Predicted-success sampling rate retains repeated events; the remaining successful-structure metrics are based on unique canonical predicted-success molecules.

Dual-target pair	Generator state	Unique success yield	Predicted-success sampling rate	Successful scaffolds	Scaffolds among all valid molecules	Top-10 successful-scaffold concentration
GSK3B–JNK3	Reward-ranked online adaptation	44.7 ± 3.2%	83.1 ± 0.1%	1511.0 ± 88.6	2090.7 ± 87.2	6.28 ± 0.34%
GSK3B–JNK3	DualEvo	48.1 ± 3.4%	79.8 ± 2.3%	1787.3 ± 100.0	2498.3 ± 144.9	3.84 ± 0.41%
RORC–DHODH	Reward-ranked online adaptation	55.9 ± 5.0%	88.8 ± 2.4%	1487.3 ± 375.7	1760.7 ± 424.9	15.19 ± 7.09%
RORC–DHODH	DualEvo	65.6 ± 2.8%	82.8 ± 2.9%	2248.0 ± 133.2	2741.3 ± 191.0	5.48 ± 0.90%
ABL1–SRC	Reward-ranked online adaptation	46.6 ± 8.3%	89.6 ± 2.2%	1536.0 ± 356.0	1885.7 ± 448.9	6.77 ± 1.70%
ABL1–SRC	DualEvo	56.2 ± 5.2%	87.0 ± 2.3%	1977.7 ± 128.3	2398.3 ± 115.0	4.03 ± 0.17%
FGFR1–KDR	Reward-ranked online adaptation	54.4 ± 6.7%	86.3 ± 4.1%	1791.0 ± 309.0	2140.0 ± 371.6	6.43 ± 2.15%
FGFR1–KDR	DualEvo	57.9 ± 1.3%	84.2 ± 2.2%	2091.7 ± 90.7	2555.7 ± 89.8	4.01 ± 0.26%

Dual-target pair	Generator state	Unique success yield	Predicted-success sampling rate	Successful scaffolds	Scaffolds among all valid molecules	Top-10 successful-scaffold concentration
EGFR–ERBB2	Reward-ranked online adaptation	48.1 ± 8.7%	91.1 ± 0.8%	1631.7 ± 466.0	1939.7 ± 502.4	8.81 ± 4.61%
EGFR–ERBB2	DualEvo	60.2 ± 1.9%	91.1 ± 0.7%	2201.0 ± 87.8	2523.3 ± 65.6	3.31 ± 0.61%
ESR1–ESR2	Reward-ranked online adaptation	40.9 ± 2.1%	71.7 ± 3.8%	1242.7 ± 88.5	2026.0 ± 207.0	9.77 ± 1.40%
ESR1–ESR2	DualEvo	43.0 ± 3.9%	59.2 ± 7.2%	1512.3 ± 119.9	2821.3 ± 122.7	5.50 ± 0.92%

E Ablations of Online Updating and Experience Selection

Table S4 compares one-shot supervised adaptation, stage-local experience updates, cumulative experience updates across stages, and experience-selection variants based on reward ranking, random selection, low reward, shuffled reward, and scaffold balancing. Cumulative experience updating and one-shot supervised adaptation are compared under the same training-exposure schedule. Ranking ablations keep the model and training procedure fixed and change only the source of experience entering the sequence update.

Table S4 | Online-update and experience-selection results on the two mechanism tasks. Values are mean ± sample standard deviation, n=3. Top-10 successful-scaffold concentration is computed from unique canonical predicted-success molecules.

Dual-target pair	Experience update / selection	Unique success yield	Successful scaffolds	Top-10 successful-scaffold concentration
GSK3B–JNK3	One-shot supervised adaptation	48.3 ± 1.3%	218.7 ± 7.6	12.68 ± 1.21%
GSK3B–JNK3	Stage-local experience adaptation	38.9 ± 7.0%	123.3 ± 30.5	30.86 ± 9.64%
GSK3B–JNK3	Cumulative reward-experience adaptation	65.5 ± 3.7%	276.3 ± 19.5	10.96 ± 0.71%
GSK3B–JNK3	Random experience selection	26.5 ± 6.3%	119.3 ± 22.2	17.18 ± 2.65%
GSK3B–JNK3	Low-reward experience selection	2.0 ± 0.2%	9.7 ± 0.6	100.00 ± 0.00%
GSK3B–JNK3	Shuffled reward ranking	26.5 ± 5.1%	120.0 ± 28.1	17.79 ± 5.77%
GSK3B–JNK3	Scaffold-balanced selection	67.7 ± 1.6%	306.3 ± 3.8	7.88 ± 0.48%

Dual-target pair	Experience update / selection	Unique success yield	Successful scaffolds	Top-10 successful-scaffold concentration
RORC–DHODH	One-shot supervised adaptation	67.9 ± 2.8%	299.3 ± 6.5	11.97 ± 1.77%
RORC–DHODH	Stage-local experience adaptation	41.7 ± 5.2%	119.7 ± 29.7	34.72 ± 8.65%
RORC–DHODH	Cumulative reward-experience adaptation	77.4 ± 1.7%	253.0 ± 47.8	23.07 ± 10.85%
RORC–DHODH	Random experience selection	47.2 ± 3.0%	218.7 ± 13.5	10.02 ± 0.78%
RORC–DHODH	Low-reward experience selection	25.5 ± 3.9%	122.7 ± 23.2	12.04 ± 5.53%
RORC–DHODH	Shuffled reward ranking	48.6 ± 5.9%	222.3 ± 30.0	10.81 ± 1.46%
RORC–DHODH	Scaffold-balanced selection	78.5 ± 4.1%	343.3 ± 24.0	9.29 ± 0.76%

F REINFORCE-LoRA Control with the Same Reward

REINFORCE-LoRA and DualEvo use the same fixed dual-target evaluation, LoRA rank=2/alpha=4, 2,000-event online budget, and 500-event independent readout. REINFORCE uses the same oracle return for valid molecules; invalid sequences receive return 0 and still participate in the policy update. The policy is updated once per batch of four complete trajectories. The advantage is the scalar reward minus an EMA baseline. For the first batch, the baseline is the batch mean; thereafter it is updated with weights 0.9/0.1. Advantages are not standardized by return variance. The objective is the mean negative advantage-weighted trajectory log probability minus 0.001 times mean entropy, with gradient norm clipped to 1.0. This condition does not include the KL term used in the DualEvo autoregressive objective. Table S5 reports paired-seed results.

Table S5 | Paired-seed results for DualEvo and REINFORCE-LoRA. Differences are computed as DualEvo minus REINFORCE-LoRA.

Dual-target pair	Random seed	DualEvo	REINFORCE-LoRA	Paired difference
GSK3B–JNK3	0	67.0%	54.8%	+12.2 pp
GSK3B–JNK3	1	69.6%	51.8%	+17.8 pp
GSK3B–JNK3	2	66.6%	50.4%	+16.2 pp
RORC–DHODH	0	75.8%	74.4%	+1.4 pp
RORC–DHODH	1	76.4%	69.2%	+7.2 pp
RORC–DHODH	2	83.2%	76.4%	+6.8 pp

G Fixed Checkpoints and MCTS Search

The frozen generator and the DualEvo-adapted generator use the same MCTS parameters and random seeds. For each of the two mechanism tasks, we run 500 batch calls, with four candidates proposed per call. The generator state remains fixed throughout search. Figure S1 shows cumulative unique predicted-success discoveries, and Table S6 summarizes the endpoints.

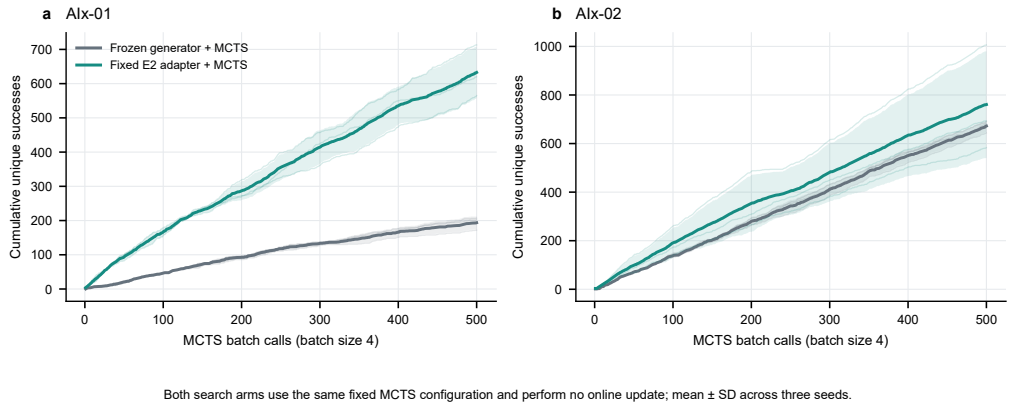

Figure S1: Cumulative unique-success discovery after plugging the frozen generator or the DualEvo-adapted generator into the same MCTS procedure. Curves show the mean over three random seeds; shading indicates ± sample standard deviation.

Table S6 | Sampling endpoints for the fixed MCTS searcher. Values are mean ± sample standard deviation, n=3.

Dual-target pair	**Generator and searcher**	**Unique successful molecules**	**Unique success yield**	**Successful scaffolds**
GSK3B–JNK3	Frozen generator + MCTS	193.7 ± 18.9	9.7 ± 0.9%	144.3 ± 17.8
GSK3B–JNK3	DualEvo-adapted generator + MCTS	632.7 ± 74.8	31.6 ± 3.7%	491.0 ± 110.8
RORC–DHODH	Frozen generator + MCTS	672.3 ± 24.8	33.6 ± 1.2%	518.7 ± 22.2
RORC–DHODH	DualEvo-adapted generator + MCTS	760.7 ± 219.5	38.0 ± 11.0%	542.0 ± 96.7

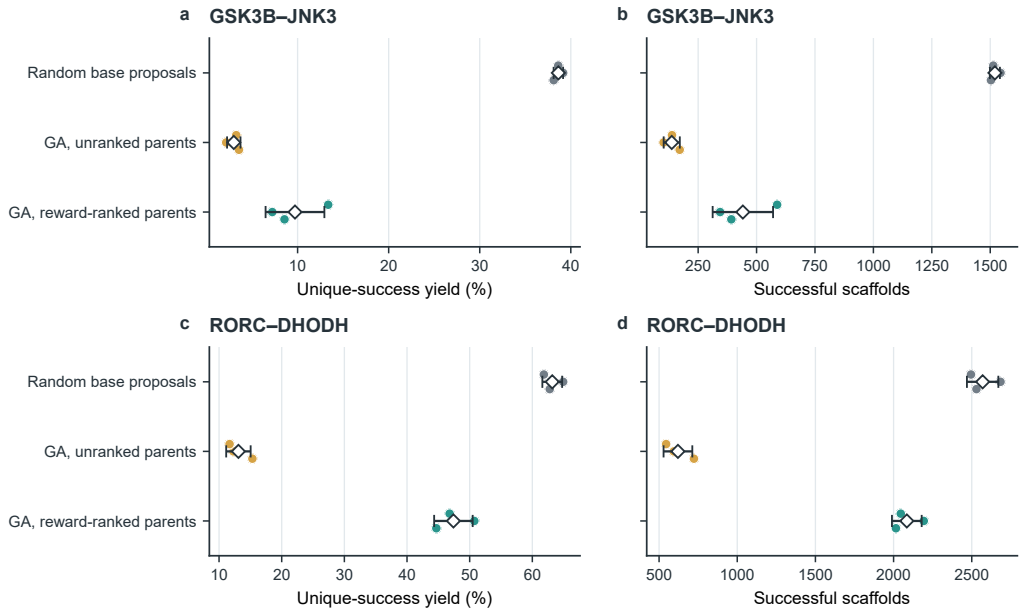

Figure S2: Independent-sampling results after adaptation with different sources of genetic-algorithm proposal experience. Each point is one seed; diamonds and error bars show mean ± sample standard deviation (n=3).

H Genetic-Algorithm Proposal Experience

The genetic algorithm supplies complete-molecule proposals. The corrected procedure starts from an initial pool produced by the frozen generator. In later stages, parents are selected from the cumulative valid archive according to the joint reward, followed by crossover and mutation. The resulting sequences enter the same scaffold-balanced selection and LoRA update pipeline. Each task and random seed uses 2,000 proposal samples, and the updated generator is evaluated with 5,000 independent samples. Figure S2 and Table S7 report results for reward-ranked parents, the original parent selection, and random proposal sources side by side.

Table S7 | Effect of genetic-algorithm proposal experience on the adapted generator. Values are mean ± sample standard deviation, n=3. Top-10 successful-scaffold concentration is computed from unique canonical predicted-success molecules for each seed.

Dual-target pair	Genetic-algorithm proposal experience	Independent-sampling unique success yield	Successful scaffolds	Top-10 successful-scaffold concentration
GSK3B–JNK3	GA proposals with original parent selection	3.0 ± 0.7%	137.0 ± 34.5	15.28 ± 5.97%
RORC–DHODH	GA proposals with original parent selection	13.1 ± 2.0%	620.7 ± 91.9	5.75 ± 1.12%
GSK3B–JNK3	GA proposals with reward-ranked parents	9.7 ± 3.2%	441.3 ± 129.3	7.31 ± 0.51%

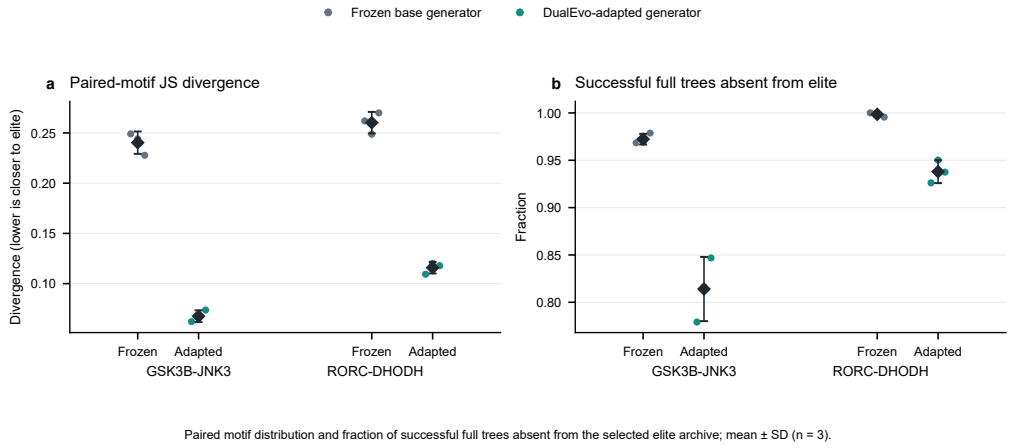

Figure S3: Paired-tree motif distributions and new combinations in successful molecules. Mean ± sample standard deviation, n=3.

Dual-target pair	Genetic-algorithm proposal experience	Independent-sampling unique success yield	Successful scaffolds	Top-10 successful-scaffold concentration
RORC–DHODH	GA proposals with reward-ranked parents	47.4 ± 3.1%	2084.7 ± 95.1	4.23 ± 0.41%
GSK3B–JNK3	Random molecule proposals	38.6 ± 0.5%	1520.0 ± 21.4	5.57 ± 0.31%
RORC–DHODH	Random molecule proposals	63.2 ± 1.6%	2569.7 ± 100.6	4.01 ± 0.55%

I Paired-Tree Distributions and Recombination in Successful Molecules

Figure S3 and Table S8 compare the paired-tree motif Jensen-Shannon distance between the frozen and DualEvo-adapted generators on four extension tasks, together with the fraction of canonical successful molecules that were not present in the reward-elite sequences. The first metric measures how closely the paired-tree motif-frequency distribution matches elite experience, while the second describes recombination at the level of complete successful molecules.

Table S8 | Paired-tree statistics for the four extension tasks.

Dual-target pair	Generator state	Paired-tree motif JS distance	Successful molecules unseen in elite experience
ABL1–SRC	DualEvo-adapted generator	0.078 ± 0.002	84.6 ± 3.3%
ABL1–SRC	Frozen generator	0.253 ± 0.011	99.2 ± 0.3%
FGFR1–KDR	DualEvo-adapted generator	0.076 ± 0.004	88.2 ± 0.4%
FGFR1–KDR	Frozen generator	0.246 ± 0.017	99.0 ± 0.0%

Dual-target pair	Generator state	Paired-tree motif JS distance	Successful molecules unseen in elite experience
EGFR–ERBB2	DualEvo-adapted generator	0.070 ± 0.008	$88.0 \pm 3.3\%$
EGFR–ERBB2	Frozen generator	0.252 ± 0.006	$99.5 \pm 0.0\%$
ESR1–ESR2	DualEvo-adapted generator	0.133 ± 0.013	$88.8 \pm 0.9\%$
ESR1–ESR2	Frozen generator	0.315 ± 0.021	$98.2 \pm 2.1\%$

J Adapter Loading and Cross-Task Evaluation

The saved GSK3B-JNK3 and RORC-DHODH LoRA adapters are each loaded onto the same frozen base generator and evaluated with 5,000 independent samples on both the source task and the other mechanism task. Figure S4 summarizes the 2×2 task-matched and cross-task evaluation matrices, and Table S9 reports unique predicted-success yield and successful-scaffold count. In the task-matched readout, the GSK3B-JNK3 adapter reaches a 48.1% unique predicted-success yield and 1787.3 successful scaffolds on GSK3B-JNK3; the RORC-DHODH adapter reaches 65.6% and 2248.0, respectively, on RORC-DHODH.

Table S9 | 5,000-event independent-sampling evaluation after loading saved adapters. Values are mean $\pm$ sample standard deviation, n=3.

Adapter training task	Evaluation task	Unique success yield	Successful scaffolds
GSK3B–JNK3	GSK3B–JNK3	$48.1 \pm 3.4\%$	1787.3 ± 100.0
GSK3B–JNK3	RORC–DHODH	$31.1 \pm 4.5\%$	1255.0 ± 158.2
RORC–DHODH	GSK3B–JNK3	$17.7 \pm 1.9\%$	643.0 ± 76.0
RORC–DHODH	RORC–DHODH	$65.6 \pm 2.8\%$	2248.0 ± 133.2

K Native-Call Candidate Streams and Fixed-Checkpoint Readouts

Table S10 summarizes candidate streams and fixed-checkpoint readouts for the run protocols used on the two mechanism tasks. Direct-generation methods record 500 native calls. Batch generation produces four candidates per call, so 500 calls correspond to 2,000 candidate events. The DualEvo three-round protocol reports the round-1 online candidate stream and the 500-event fixed-checkpoint readout after three training rounds. Historical MCTS-v2 is truncated at the 500th unique oracle call and retains all proposal events before that cutoff. Rates in each row use the listed candidate events as the denominator. This three-round protocol is reported separately from the main 10-stage $\times$ 200-candidate online procedure and is not combined with the main final-checkpoint readout.

Table S10 | Native-call candidate streams and fixed-checkpoint readouts on the two mechanism tasks. Values are mean $\pm$ sample standard deviation, n=3. Calls denotes the number of native calls, and Candidate events is the actual number of candidate events in that row. Predicted-success sampling rate, unique predicted-success yield, and validity are computed by dividing successful events, unique canonical successful molecules, and valid events, respectively, by the listed Candidate events. The DualEvo round-3 fixed readout follows three rounds totaling 1,500 batch-rollout calls.

Dual-target pair	Generation / search method	Seeds	Calls	Events	Sampling hit rate	Unique success yield	Validity
GSK3B–JNK3	ChemTSv2	3	500	500	$9.5 \pm 2.5\%$	$9.5 \pm 2.5\%$	$100.0 \pm 0.0\%$

Dual-target pair	Generation / search method	Seeds	Calls	Events	Sampling hit rate	Unique success yield	Validity
GSK3B–JNK3	Historical MCTS-v2 (500 unique oracle calls)	3	500	605	28.4 ± 1.1%	27.9 ± 1.2%	83.3 ± 0.8%
GSK3B–JNK3	DualEvo three-round: round-1 online stream	3	500	2000	30.5 ± 1.4%	21.7 ± 2.6%	78.8 ± 3.3%
GSK3B–JNK3	GraphGA	3	500	500	15.3 ± 3.7%	15.3 ± 3.7%	100.0 ± 0.0%
GSK3B–JNK3	MARS	3	500	500	26.8 ± 21.7%	1.5 ± 0.8%	100.0 ± 0.0%
GSK3B–JNK3	REINVENT4	3	500	500	5.7 ± 2.6%	5.7 ± 2.6%	98.7 ± 0.5%
GSK3B–JNK3	Saturn	3	500	500	13.0 ± 4.2%	12.9 ± 4.3%	95.9 ± 1.6%
GSK3B–JNK3	DualEvo three-round: round-3 fixed readout	3	500	500	76.7 ± 3.6%	41.1 ± 3.5%	90.5 ± 4.3%
GSK3B–JNK3	Frozen Student + MCTS (500 batch calls)	3	500	2000	11.4 ± 0.6%	9.7 ± 0.9%	77.6 ± 0.9%
GSK3B–JNK3	Frozen Student: round-0 fixed readout	3	500	500	20.0 ± 2.1%	19.9 ± 2.1%	97.9 ± 0.8%
RORC–DHODH	ChemTSv2	3	500	500	19.4 ± 0.7%	19.3 ± 0.6%	100.0 ± 0.0%
RORC–DHODH	Historical MCTS-v2 (500 unique oracle calls)	3	500	594	52.4 ± 6.2%	51.8 ± 5.7%	84.8 ± 2.2%
RORC–DHODH	DualEvo three-round: round-1 online stream	3	500	2000	47.5 ± 5.8%	39.3 ± 6.1%	74.0 ± 5.8%
RORC–DHODH	GraphGA	3	500	500	39.3 ± 2.1%	39.3 ± 2.1%	100.0 ± 0.0%
RORC–DHODH	MARS	3	500	500	58.6 ± 6.7%	6.4 ± 1.5%	100.0 ± 0.0%
RORC–DHODH	REINVENT4	3	500	500	20.3 ± 1.8%	20.3 ± 1.8%	97.9 ± 0.1%
RORC–DHODH	Saturn	3	500	500	27.9 ± 0.6%	27.9 ± 0.6%	95.8 ± 1.1%
RORC–DHODH	DualEvo three-round: round-3 fixed readout	3	500	500	87.7 ± 3.7%	38.9 ± 4.2%	93.3 ± 1.9%
RORC–DHODH	Frozen Student + MCTS (500 batch calls)	3	500	2000	38.1 ± 1.3%	33.6 ± 1.2%	76.8 ± 0.2%

Dual-target pair	Generation / search method	Seeds	Calls	Events	Sampling hit rate	Unique success yield	Validity
RORC–DHODH	Frozen Student: round-0 fixed readout	3	500	500	49.7 ± 3.5%	49.7 ± 3.5%	98.2 ± 0.4%

L Fixed Dual-Target Scoring of External Candidate Sets

Candidate sets from Saturn, GraphGA, MARS, ChemTSv2, REINVENT4, and Historical MCTS-v2 are scored using the same fixed dual-target evaluator and predicted-success criterion defined in Section 3.2 (Guo & Schwaller, 2024b; Jensen, 2019; Xie et al., 2021; Ishida et al., 2023; Loeffler et al., 2024). Table S11 reports predicted-success sampling rate, unique predicted-success yield, validity, QED, and SA. The Historical MCTS-v2 rows use the 500 unique canonical valid molecules produced by the trace within the 500-unique-oracle-call cutoff.

Table S11 | Scores of external candidate sets under the fixed dual-target evaluator. Values are mean ± sample standard deviation, n=3. Historical MCTS-v2 uses 500 unique canonical valid molecules; the other methods use candidate sets exported from the corresponding benchmark runs.

Dual-target pair	Method	Seeds	Sampling hit rate	Unique success yield	Validity	QED	SA
GSK3B–JNK3	Saturn	3	13.0 ± 4.2%	12.9 ± 4.3%	95.9 ± 1.6%	0.662 ± 0.031	2.574 ± 0.068
RORC–DHODH	Saturn	3	27.9 ± 0.6%	27.9 ± 0.6%	95.8 ± 1.1%	0.659 ± 0.014	2.658 ± 0.052
GSK3B–JNK3	GraphGA	3	15.3 ± 3.7%	15.3 ± 3.7%	100.0 ± 0.0%	0.681 ± 0.006	2.700 ± 0.019
RORC–DHODH	GraphGA	3	39.3 ± 2.1%	39.3 ± 2.1%	100.0 ± 0.0%	0.697 ± 0.007	2.808 ± 0.016
GSK3B–JNK3	ChemTSv2	3	9.5 ± 2.5%	9.5 ± 2.5%	100.0 ± 0.0%	0.715 ± 0.008	3.222 ± 0.133
RORC–DHODH	ChemTSv2	3	19.4 ± 0.7%	19.3 ± 0.6%	100.0 ± 0.0%	0.718 ± 0.007	3.294 ± 0.047
GSK3B–JNK3	REINVENT4	3	5.7 ± 2.6%	5.7 ± 2.6%	98.7 ± 0.5%	0.574 ± 0.009	2.865 ± 0.026
RORC–DHODH	REINVENT4	3	20.3 ± 1.8%	20.3 ± 1.8%	97.9 ± 0.1%	0.580 ± 0.003	2.906 ± 0.041
GSK3B–JNK3	MARS	3	26.8 ± 21.7%	1.5 ± 0.8%	100.0 ± 0.0%	0.507 ± 0.072	2.929 ± 0.273
RORC–DHODH	MARS	3	58.6 ± 6.7%	6.4 ± 1.5%	100.0 ± 0.0%	0.502 ± 0.028	3.173 ± 0.257
GSK3B–JNK3	Historical MCTS-v2 (500 unique valid molecules)	3	33.7 ± 1.2%	33.7 ± 1.2%	100.0 ± 0.0%	0.633 ± 0.006	2.748 ± 0.013
RORC–DHODH	Historical MCTS-v2 (500 unique valid molecules)	3	61.4 ± 5.5%	61.4 ± 5.5%	100.0 ± 0.0%	0.636 ± 0.003	2.870 ± 0.133

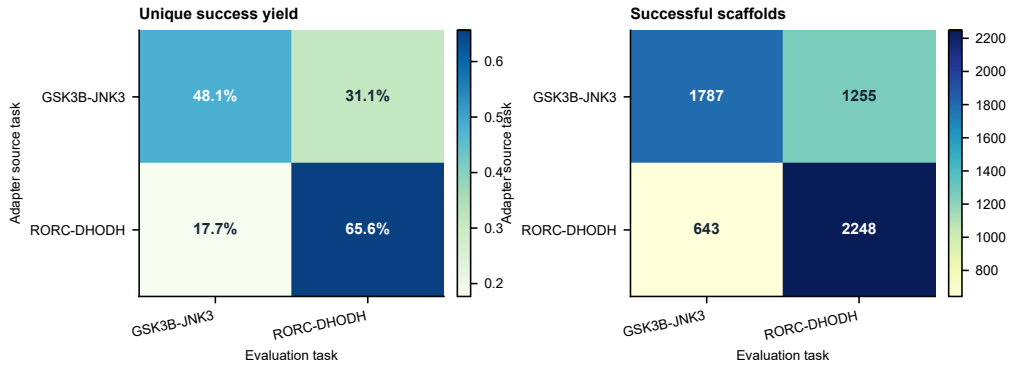

Figure S4: Task-matched and cross-task fixed-generator readout of saved adapters. The heatmaps show unique predicted-success yield and successful-scaffold count. Rows indicate the adapter source task and columns indicate the evaluation task.

M Discovery Curves over Sampling Budget

Table S12 uses prefixes of a separate ordered 5,000-sample stream to show how the cumulative numbers of unique successful molecules and successful scaffolds grow with the sampling budget. All prefixes share the same random stream and are aggregated over three seeds. The 500-sample prefix is from a different sampling stream than the independent 500-sample readout in Table 1. The cumulative curves are shown in Figure 3 and are not repeated here.

Table S12 | Cumulative successful output and scaffold coverage for DualEvo at selected sampling prefixes. Values are mean ± sample standard deviation, n=3. Top-10 successful-scaffold concentration is computed from unique canonical predicted-success molecules within each prefix.

Dual-target pair	Cumulative samples	Unique successful molecules	Successful scaffolds	Top-10 successful-scaffold concentration
GSK3B–JNK3	100	77.7 ± 1.5	75.7 ± 0.6	15.44 ± 0.99%
GSK3B–JNK3	500	343.7 ± 12.7	312.7 ± 13.4	7.36 ± 1.26%
GSK3B–JNK3	1,000	625.3 ± 24.5	536.3 ± 12.1	6.28 ± 0.50%
GSK3B–JNK3	2,000	1133.7 ± 70.9	921.3 ± 46.5	4.84 ± 0.28%
GSK3B–JNK3	5,000	2402.7 ± 171.1	1787.3 ± 100.0	3.84 ± 0.41%
RORC–DHODH	100	80.7 ± 7.0	75.7 ± 6.4	18.62 ± 2.29%
RORC–DHODH	500	394.7 ± 24.5	341.7 ± 24.2	9.27 ± 2.07%
RORC–DHODH	1,000	761.7 ± 40.6	616.7 ± 43.7	7.43 ± 1.06%
RORC–DHODH	2,000	1428.7 ± 80.2	1076.3 ± 77.2	6.64 ± 0.99%
RORC–DHODH	5,000	3281.3 ± 139.3	2248.0 ± 133.2	5.48 ± 0.90%

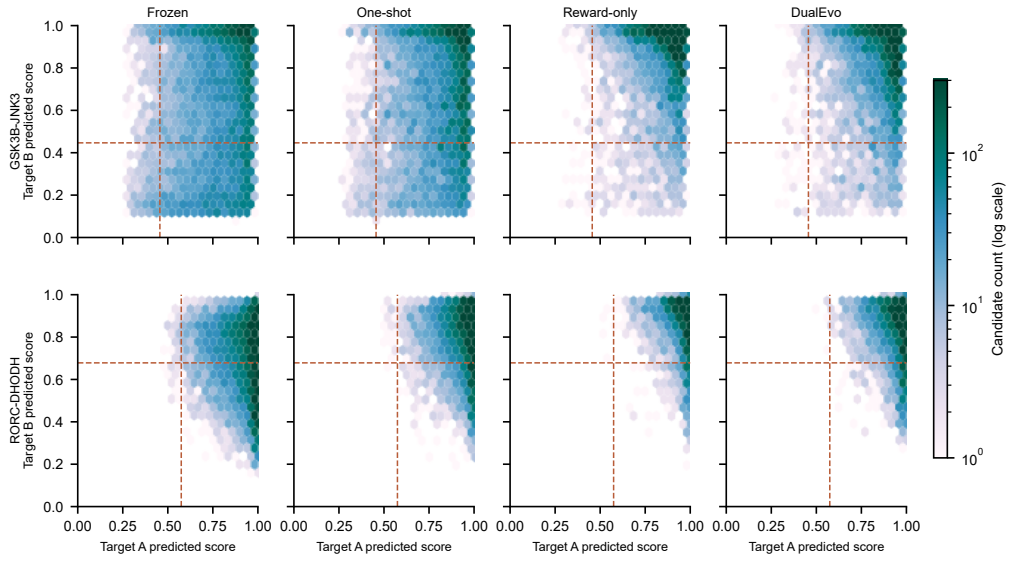

Figure S5: Joint distribution of predicted scores for the two targets. Hexbin density represents the number of candidates, dashed lines mark target-specific thresholds, and both axes show outputs from the fixed activity predictors.

N Dual-Target Score Space

Figure S5 plots the scores from the fixed activity predictors for target A and target B in the same coordinate system and marks the target-specific decision thresholds. Each point corresponds to a valid candidate molecule; the upper-right quadrant contains candidates that meet both fixed prediction thresholds.

Table S13 | Summary of dual-target scores under the fixed predictors. Scores and the number of candidates above both thresholds are reported as mean ± sample standard deviation across seeds.

Dual-target pair	**Generator state**	**Target A threshold**	**Target B threshold**	**Mean target A score**	**Mean target B score**	**Molecules above both thresholds**
GSK3B–JNK3	Frozen generator	0.4560	0.4467	0.799 ± 0.002	0.655 ± 0.004	3378.3 ± 33.0
GSK3B–JNK3	DualEvo-adapted generator	0.4560	0.4467	0.874 ± 0.009	0.853 ± 0.006	4627.0 ± 42.2
RORC–DHODH	Frozen generator	0.5747	0.6783	0.938 ± 0.001	0.692 ± 0.003	2837.3 ± 29.3
RORC–DHODH	DualEvo-adapted generator	0.5747	0.6783	0.965 ± 0.008	0.847 ± 0.018	4406.7 ± 120.7

O Archive Overlap and Nearest-Neighbor Structural Relationships

Figure S6 compares exact canonical overlap between all valid unique query molecules and the time-aligned archive. For query molecules absent from that archive, it also shows maximum nearest-neighbor Tanimoto similarity using Morgan radius-2, 2048-bit fingerprints.

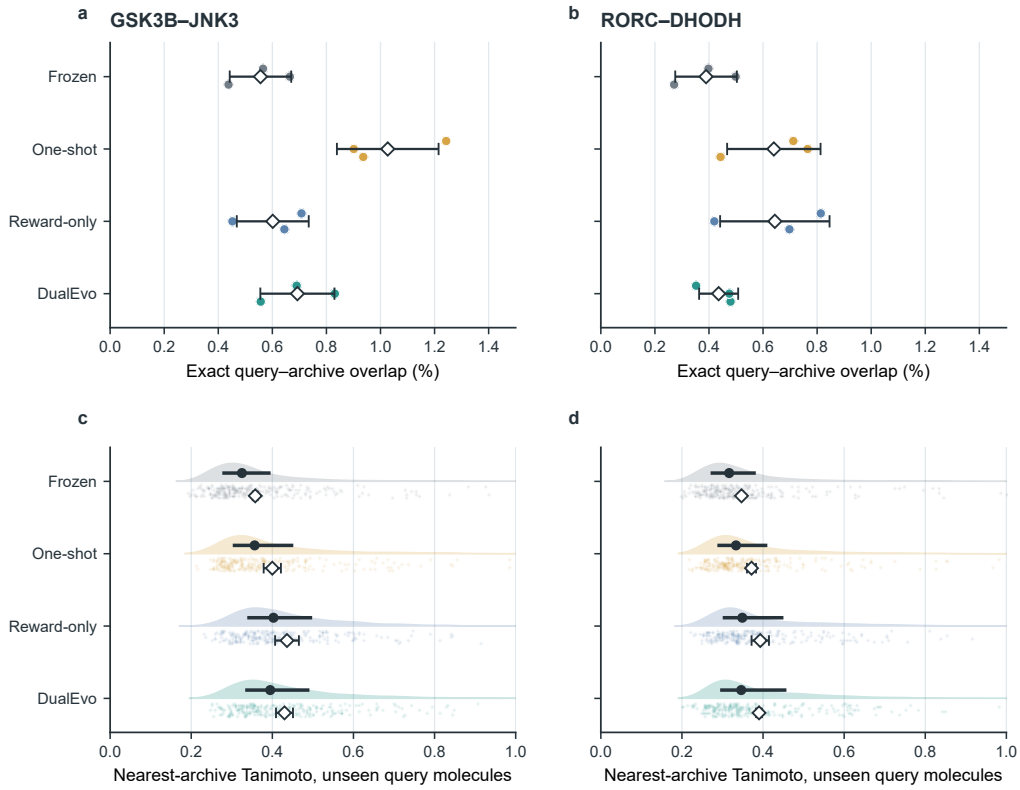

Figure S6: Relationship of generated valid unique molecules to the time-aligned archive at round 3. Top: exact query–archive overlap, with seed values and mean ± SD. Bottom: nearest-archive Tanimoto for all valid unique query molecules absent from the archive. Half violins use all molecule-level values; plotted molecule dots show up to 240 per group; horizontal bars show molecule-level median and IQR; diamonds and whiskers show mean ± SD across three seeds.

Table S14 | Overlap and nearest-neighbor summary for the time-aligned experience archive. This time-aligned analysis uses its corresponding stage-3 experience archive and is not equated with the terminal archive from the main 10-stage procedure. Values are mean ± sample standard deviation, n=3.

Dual-target pair	**Generator state**	**Exact overlap with archive**	**Successful molecules unseen in archive**	**Nearest-neighbor Tanimoto for unseen valid molecules**	**Nearest-neighbor Tanimoto ≥0.70**
GSK3B–JNK3	Frozen generator	0.56 ± 0.11%	98.4 ± 0.5%	0.358 ± 0.006	3.23 ± 0.26%
GSK3B–JNK3	One-shot supervised adaptation	1.03 ± 0.19%	98.3 ± 0.3%	0.400 ± 0.021	5.48 ± 0.30%

Dual-target pair	Generator state	Exact overlap with archive	Successful molecules unseen in archive	Nearest-neighbor Tanimoto for unseen valid molecules	Nearest-neighbor Tanimoto $\geq$0.70
GSK3B–JNK3	Reward-ranked online adaptation	$0.60 \pm 0.13\%$	$99.3 \pm 0.2\%$	0.436 ± 0.029	$5.28 \pm 2.45\%$
GSK3B–JNK3	DualEvo	$0.69 \pm 0.14\%$	$99.2 \pm 0.2\%$	0.430 ± 0.021	$5.17 \pm 0.67\%$
RORC–DHODH	Frozen generator	$0.39 \pm 0.11\%$	$99.4 \pm 0.2\%$	0.347 ± 0.004	$2.60 \pm 0.47\%$
RORC–DHODH	One-shot supervised adaptation	$0.64 \pm 0.17\%$	$99.2 \pm 0.2\%$	0.372 ± 0.011	$3.57 \pm 0.44\%$
RORC–DHODH	Reward-ranked online adaptation	$0.64 \pm 0.20\%$	$99.3 \pm 0.2\%$	0.393 ± 0.021	$4.42 \pm 1.73\%$
RORC–DHODH	DualEvo	$0.44 \pm 0.07\%$	$99.5 \pm 0.1\%$	0.390 ± 0.007	$3.53 \pm 0.70\%$

P Supplementary Figures and Data

The appendix reports the complete fixed-checkpoint, mechanism, and reuse results associated with the main text. Figures S1–S8 present the corresponding visual analyses.

Q Paired Differences in the Primary Readout

Figure S7 reports matched-seed differences in final unique predicted-success yield for DualEvo relative to the frozen generator and Reward-only. Absolute per-task means are given in Table 1.

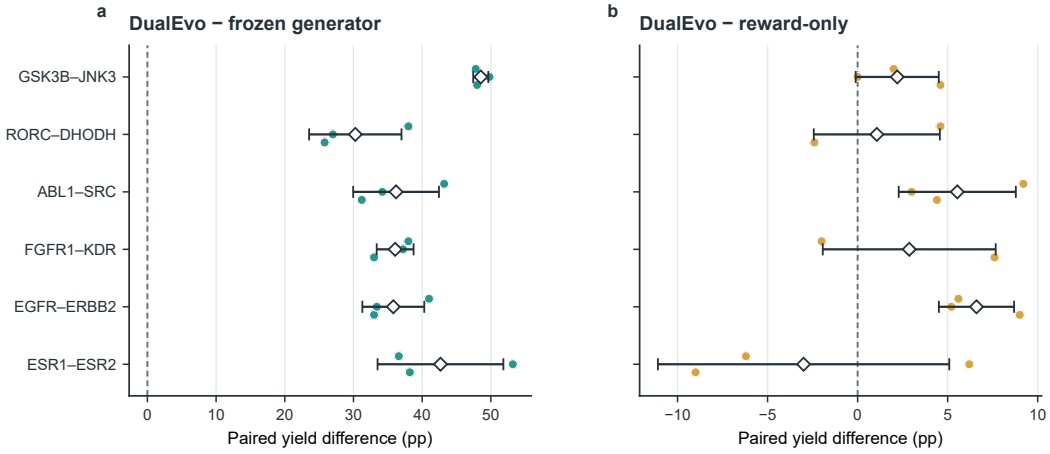

Figure S7: Paired differences in final unique predicted-success yield. Each small point is the difference for the same random seed. Diamonds and error bars show mean ± sample SD across three paired seeds; the dashed line marks zero. Absolute per-task values are in Table 1.

R Representative Dual-Target Candidates and Predicted Docking Poses

Four unique predicted-success Pareto-front molecules from the first 500 independent samples are shown in Figure S8. Two molecules are from RORC–DHODH and two from EGFR–ERBB2. Docking is a computational structural view and does not establish actual binding or functional activity.

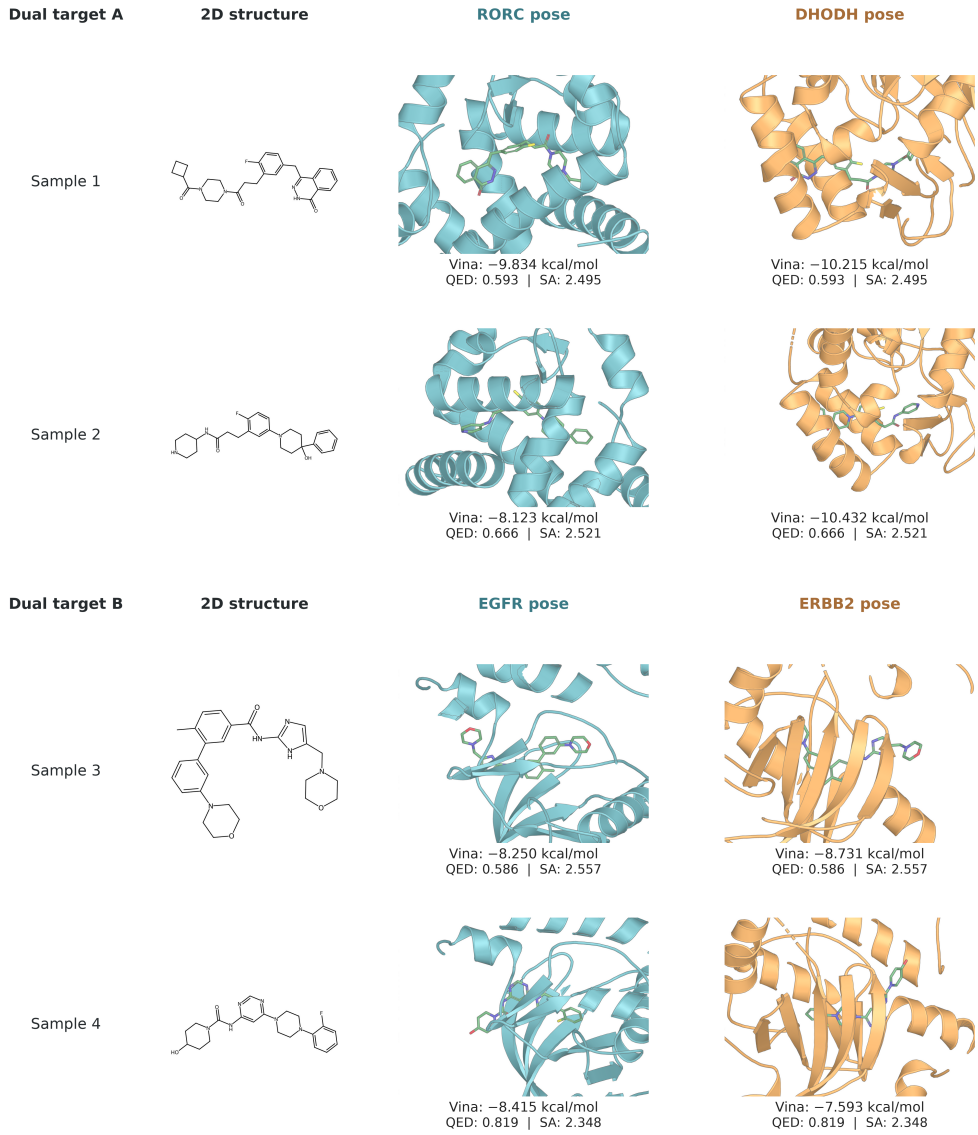

Figure S8: Representative DualEvo dual-target candidates and predicted docking poses. The top two rows show RORC–DHODH and the bottom two rows EGFR–ERBB2. Each row shows one molecule, its 2D structure, predicted poses for both targets, Vina scores, QED, and SA.