POSEIDON: COLLABORATIVE STRUCTURE-BASED DRUG DESIGN WITH CLM AGENTS AND LLMS

Anonymous authors
Paper under double-blind review

ABSTRACT

Structure-based drug design requires both efficient exploration of chemical space and interpretation of protein–ligand geometry. Chemical language model (CLM) agents can adapt their sampling policies efficiently to scalar docking rewards, but this feedback leaves the structural information in docked poses underused. Large language models (LLMs) can interpret explicit structural evidence and propose targeted edits, yet local refinement alone does not train a reusable molecular policy. We introduce POSEIDON, a collaborative framework that connects these complementary capabilities. A CLM agent learns from scalar rewards, while an LLM receives atom-addressable multi-pose evidence to guide interpretable molecular editing. Proposed edits are independently redocked before entering a source-aware memory; replay transfers verified outcomes into CLM learning, and continued exploration supplies fresh structural contexts to the LLM. Run-wide resource accounting, quality–diversity parent selection, and outcome-adaptive prompts organize this exchange under a limited LLM budget. On CrossDocked2020, POSEIDON achieves 70.5% joint success, compared with 56.2% for CLM-only optimization and 20.6% for LLM-only editing. An FXR case study additionally identifies three RJC derivatives with better neutral-state Vina scores than four reference ligands, illustrating constrained derivative design without establishing experimental activity.

1 INTRODUCTION

Structure-based drug design (SBDD) seeks molecules that complement a protein binding site while satisfying chemical requirements. This demands two complementary capabilities: exploring a large molecular space and interpreting the local geometry of individual protein–ligand complexes. Pocket-conditioned generators learn spatial compatibility from structural data (Peng et al., 2022; Guan et al., 2023a;b). Chemical language model (CLM) agents provide another useful capability: they generate molecular sequences and efficiently adapt their sampling policies to scalar objectives through reinforcement learning (RL), as established by REINVENT and REINVENT 4 (Olivecrona et al., 2017; Loeffler et al., 2024). The challenge is to combine efficient reward-driven exploration with structural understanding that can guide specific chemical changes.

Docking naturally supplies both types of feedback: a score that ranks a candidate and three-dimensional pose hypotheses that describe how it occupies the pocket (Trott & Olson, 2010; Eberhardt et al., 2021). A scalar reward is convenient for policy optimization but does not identify which substituent approaches a pocket boundary or which local contacts recur across alternative poses. Conversely, interpreting a pose can motivate a localized edit without providing an efficient mechanism for exploring the broader chemical space. These complementary signals motivate our central question: *how can a CLM agent that learns from scalar rewards collaborate with an LLM that interprets 3D structural evidence, so that each improves the other's search?*

LLMs offer a flexible interface for combining chemical instructions with structured observations. Existing methods support molecular editing and feedback-guided search (Liu et al., 2024; Brahmavar et al., 2024; Wang et al., 2025); collaborative systems also connect pretrained generators with LLM-assisted refinement (Gao et al., 2025; Averly et al., 2025). However, an improved edited molecule need not improve the policy that generated its parent. When refinement ends at post-processing or candidate selection, verified structural edits do not become learning experience for the

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.

generator. We focus on this reciprocal connection: CLM exploration should provide diverse parents for structural editing, while evaluated edits should improve subsequent CLM sampling.

This collaboration raises three challenges. Structural evidence needs stable correspondence between molecular connectivity and geometry, together with an explicit treatment of disagreement among pose hypotheses. Limited LLM resources must be allocated to promising, nonredundant editing opportunities across multiple rounds. Finally, a shared experience pool must retain contributions from both pipelines without allowing a narrow edited family to dominate policy learning. Reliable collaboration therefore requires an interface for structural evidence, a resource-allocation mechanism, and a way to reuse verified outcomes.

We introduce **POSEIDON**, which assigns scalar-reward exploration to a CLM agent and evidence-guided refinement to an LLM. The agent samples molecules and learns from docking outcomes; the LLM interprets mapped multi-pose evidence and proposes localized, traceable edits. Every candidate passes chemical validation and independent docking before entering shared memory. Source-aware replay transfers edited experience into the CLM, while continued agent exploration refreshes the parents and structural contexts available to the LLM. Neither module's proposal is accepted as an improvement without oracle evaluation.

Our contributions are threefold. **First**, an atom-addressable evidence interface links connectivity, receptor geometry, atom–residue contacts, and cross-pose recurrence to editing hypotheses. **Second**, run-wide budgeting, quality–diversity parent selection, and outcome-adaptive prompts organize LLM assistance, with optional prompt evolution. **Third**, independent verification and source-aware replay connect the two modules through shared experience. On 100 CrossDocked2020 pockets, the coupled system achieves 70.5% joint success versus 56.2% and 20.6% for its isolated CLM and LLM variants. A complementary FXR study identifies three RJC-based derivatives with neutral-state Vina scores of -14.2, -14.0, and -13.9 kcal/mol, compared with -13.3 kcal/mol for the best reference. This case study illustrates constrained local design; it does not independently establish the benefit of policy–editor coupling or biological efficacy.

2 Related Work

Table 1 distinguishes a separate LLM reasoner/editor, a specialized molecular generator, and docking-driven updates to generator parameters. Prompt or population updates alone do not count as generator learning.

Table 1: Complementary mechanisms in molecular design, grouped by configuration. LLM reasoning denotes an additional editor or planner, rather than the molecular policy itself.

Methods	Separate LLM reasoning	Specialized molecular generator	Docking-driven generator learning
Pocket2Mol, TargetDiff, DecompDiff, MolCRAFT	✗	✓	✗
REINVENT 4, Augmented Memory, SmileyLlama	✗	✓	✓
ChatDrug, DrugAssist, LMLF, MOLLEO	✓	✗	✗
CIDD, LIDDIA	✓	✓	✗
POSEIDON (ours)	✓	✓	✓

Generation and policy learning. Pocket2Mol, TargetDiff, DecompDiff, and MolCRAFT learn pocket-conditioned molecular generation without online docking-driven parameter updates in their standard sampling workflows (Peng et al., 2022; Guan et al., 2023a;b; Qu et al., 2024). REINVENT 4 and Augmented Memory instead support goal-directed policy learning and experience reuse (Loeffler et al., 2024; Guo & Schwaller, 2024). SmileyLlama also learns from docking preferences, but its LLM is the molecular generator, not an additional structural reasoner (Cavanagh et al., 2026).

Language-guided editing and collaboration. ChatDrug, DrugAssist, LMLF, and MOLLEO use LLMs for editing or feedback-guided search without a separate specialized neural generator (Liu

et al., 2024; Ye et al., 2025; Brahmavar et al., 2024; Wang et al., 2025). CIDD combines pretrained 3D generators with LLM-based structural refinement; LIDDIA coordinates generation and optimization tools (Gao et al., 2025; Averly et al., 2025). Both exploit evaluation feedback, but their reported workflows do not update the underlying neural generator through that feedback. POSEIDON makes the two modules mutually informative: CLM exploration supplies editing contexts, while independently redocked LLM edits enter source-aware memory and replay to guide subsequent CLM exploration.

3 METHODOLOGY

Overview. Structure-based drug design (SBDD) aims to discover molecules that complement a target binding site while satisfying chemical constraints. POSEIDON treats a docking evaluation as two complementary information channels: a scalar reward for efficient policy adaptation and 3D pose evidence for targeted chemical reasoning. The **CLM agent** uses RL to explore and concentrate probability on promising molecular sequences. The **LLM editor** interprets mapped poses to propose localized changes that scalar scores alone do not specify. Their collaboration is reciprocal: agent exploration supplies parents and structural contexts; independently redocked edits supply new experience for agent learning and prompt adaptation. A shared verification-and-memory interface connects the modules without requiring a common neural architecture (Figure 1).

3.1 CLM AGENT: SCALAR-REWARD EXPLORATION

We follow the molecular reinforcement-learning framework established by REINVENT (Olivecrona et al., 2017) and developed further in REINVENT 4 (Loeffler et al., 2024), which integrates recurrent and Transformer-based molecular generators. Our CLM is a single GPT-style autoregressive policy $\pi_\theta(x) = \prod_\ell \pi_\theta(x_\ell \mid x_{<\ell})$, initialized from a pretrained prior π_0. Each target protein P is optimized independently; x denotes a molecular sequence and $R_P(x)$ its docking-based reward. The receptor conditions the oracle and structural editor rather than entering the CLM as a neural input. We use the augmented-likelihood objective:

$$\mathcal{L}_t(\theta) = \frac{1}{|\mathcal{B}_t|} \sum_{x \in \mathcal{B}_t} \left[\log \pi_\theta(x) - \log \pi_0(x) - \sigma_t R_P(x)\right]^2 . \tag{1}$$

The prior anchors the likelihood target, and σ_t controls reward strength. This module supplies efficient scalar-objective adaptation within the collaboration. Its training experience $\mathcal{B}_t$ includes independently verified LLM edits as well as CLM samples, so structural refinement can influence future exploration. Likelihoods are recomputed under the current policy for all sources. The shared reward, sequence handling, and optimization settings are detailed in Appendix A.1.

3.2 LLM EDITOR: REASONING OVER 3D STRUCTURAL EVIDENCE

3.2.1 ATOM-ADDRESSABLE MULTI-POSE EVIDENCE

A docking score compresses a protein–ligand configuration into one number. Our structural interface retains *where* a candidate contacts the receptor, *which atoms* support an editing hypothesis, and *whether the observation recurs across poses*. For molecule x and receptor P, AutoDock Vina (Trott & Olson, 2010; Eberhardt et al., 2021) returns $\mathcal{D}_P(x) = \{(Z_x^{(k)}, s_x^{(k)})\}_{k=1}^{K_x}$, where $Z_x^{(k)}$ contains ligand heavy-atom coordinates and scores are ordered from lowest to highest. We encode the best $K'_x = \min(K_x, K_{\mathrm{LLM}})$ returned poses. These are binding-mode hypotheses under the prepared receptor and scoring model, rather than experimentally established bound structures.

Geometric correspondence. One-based atom-map identifiers are assigned to the canonical parent graph. Stereochemistry-aware graph matching aligns each docked pose with these identifiers, linking molecular connectivity to spatial observations. Let U_j denote receptor heavy-atom coordinates. The ligand and receptor are translated by the same docking-box center before serialization; pairwise distances $d_{ij}^{(k)} = \|Z_{x,i}^{(k)} - U_j\|_2$ are unchanged. We extract a bounded contact set,

$$\widehat{C}_x^{(k)} = \mathrm{Nearest}_{L_c}\left\{(i, j, d_{ij}^{(k)}) : d_{ij}^{(k)} \le d_c\right\}, \tag{2}$$

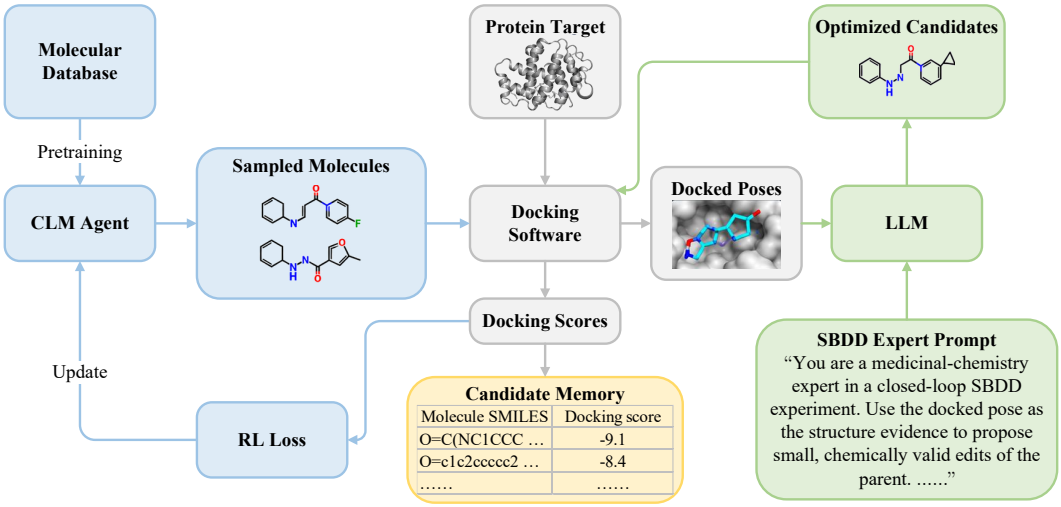


Figure 1: **POSEIDON: complementary agents with reciprocal feedback.** The **blue module** uses a CLM agent with reinforcement learning for efficient scalar-reward exploration. The **green module** uses an LLM editor to interpret docked poses and propose changes grounded in 3D structural evidence. These independently specified pipelines exchange verified candidates through a shared docking-and-memory interface: CLM exploration supplies candidates and poses for structural editing, while redocked edits enrich CLM training experience. Every edited candidate is evaluated by the docking oracle before memory admission.

where $\mathrm{Nearest}_{L_c}$ retains at most L_c shortest-distance pairs. Each record includes the ligand atom ID, receptor residue and atom identifiers, elements, and distance. Coordinates and scores remain available alongside the records, preserving the local configuration. Truncation metadata distinguishes limited evidence from an exhaustive contact inventory.

Cross-pose recurrence. To expose agreement and ambiguity, we aggregate ligand-atom–residue proximity across encoded poses:

$$f_x(i,r) = \frac{1}{K'_x} \sum_{k=1}^{K'_x} \mathbf{1}\left[\exists j : \mathrm{res}(j) = r,\ (i,j,d_{ij}^{(k)}) \in \widehat{C}_x^{(k)}\right]. \tag{3}$$

This counts a residue once per ligand atom per pose and uses the number of actually encoded poses as the denominator. It describes recurrence in retained contacts, not a thermodynamic population or calibrated interaction probability. The editor can distinguish opportunities supported by several hypotheses from changes justified by only one pose. Proximity alone is not labeled a hydrogen bond or favorable interaction; explanations distinguish geometric observations from chemical hypotheses.

A structured interface for molecular editing. The representation combines molecular connectivity, local geometry, and cross-pose agreement:

$$\mathcal{E}_P(x) = \left(x^{\mathrm{map}}, \mathbf{q}(x), R_P(x), \{s_x^{(k)}, \widetilde{Z}_x^{(k)}, \widetilde{U}_x^{(k)}, \widehat{C}_x^{(k)}\}_{k=1}^{K'_x}, f_x\right), \tag{4}$$

where $\mathbf{q}(x)$ contains computed molecular properties and tildes denote centered coordinates; $\widetilde{U}_x^{(k)}$ is the retained local receptor subset. Connectivity constrains possible modifications, coordinates describe the spatial configuration, and contact records provide compact references for explanations. Recurrence distinguishes repeated from pose-specific observations. The evidence is organized into four complementary layers, summarized in Table 2. This separates geometric observations from editing objectives and makes the information available to the editor explicit.

This organization supports localized, testable hypotheses. For example, the editor may identify a substituent repeatedly approaching a pocket boundary and propose reducing its size, citing the relevant atoms and residues. If its orientation varies across poses, the explanation should state that

Table 2: Structural evidence supplied to the LLM and its role in editing.

Evidence layer	Retained information	Role in molecular editing
Molecular identity	Mapped connectivity, stereochemistry, properties	Identify the atoms and chemical context of a proposed change.
Per-pose geometry	Centered coordinates, scores, atom–residue contacts	Locate spatial opportunities and cite traceable observations.
Cross-pose support	Contact recurrence and truncation metadata	Distinguish recurring observations from pose-dependent hypotheses.
Design objective	Reward components, weights, explicit constraints	Align edits with the optimization goal and required chemistry.

the evidence for a particular contact is pose-dependent. Such observations motivate candidate edits rather than determine their outcome. The interface does not hard-code a contact-to-transformation lookup table: the editor combines connectivity, geometry, and explicit objectives to propose chemical changes. Score-only, single-pose, and multi-pose variants isolate the additional information supplied by this interface.

3.2.2 Grounded edits with independent structural verification

For parent x, the editor conditions on $\mathcal{E}_P(x)$, a strategy prompt p, objective constraints, and optionally recent verified edits. It returns up to M proposals (x'_j, h_j). Each h_j contains a short edit summary, evidence references $(\text{pose rank}, \text{atom ID}, \text{residue ID})$, a testable hypothesis, and an uncertainty statement. This contract links a chemical modification to addressable structural observations, allowing its premise and measured outcome to be inspected separately.

The parser checks the response schema and verifies nonempty evidence references against supplied contact records. Each proposed structure is sanitized, canonicalized with stereochemistry retained, checked against configured chemical constraints and CLM representability, and deduplicated. Unchanged parent copies are discarded. Every accepted edit is evaluated by the same docking and reward pipeline:

$$x'_j \xrightarrow{\mathcal{D}_P} \{Z^{(k)}_{x'_j}, s^{(k)}_{x'_j}\} \xrightarrow{R_P} R_P(x'_j), \qquad \Delta_j = R_P(x'_j) - R_P(x). \tag{5}$$

Only successful evaluations enter memory. Neither the parent's score nor the LLM's assessment becomes the child's training reward. Identifier checks establish traceability, while redocking tests the edit under the optimization oracle; neither establishes experimental affinity or causal faithfulness of an explanation. Verified children need not outperform their parents to be retained as evaluated experience.

3.2.3 Outcome-adaptive structural editing under a budget

Selecting editing opportunities. At scheduled iterations, we form a reward-ranked shortlist $\mathcal{C}_t$ of memory candidates not previously queried. A small parent batch is selected greedily by

$$x^* = \arg\max_{x \in \mathcal{C}_t \setminus S} \left[R_P(x) + \lambda \left(1 - \max_{y \in S} \mathrm{Tan}(F(x), F(y)) \right) \right], \tag{6}$$

where S contains already selected parents and F is a circular fingerprint (Rogers & Hahn, 2010). The diversity term is one when S is empty. This allocates calls to promising yet nonredundant parents within each editing round.

Run-wide budget accounting. For each target, one budget controller persists across all optimization rounds. LLM calls and accounted tokens accumulate over the entire run, including both molecular editing and prompt evolution; neither counter resets at a new round. A separate cumulative docking budget is shared by CLM samples and edited children. Requests are admitted against the remaining resources, and unused capacity carries forward. When the LLM budget is exhausted, CLM–RL optimization can continue while docking resources remain. Different targets have independent budget controllers. Thus, allocation is global across rounds within a target run, although parent selection is a round-wise heuristic rather than a globally optimal scheduling policy.

Learning which structural instructions work. The initial prompt pool emphasizes local substitution, changes supported across poses, and reduction of crowding or unnecessary flexibility. After a call, its strategy receives the mean parent-relative reward gain over parsed proposals, with failed, repeated, or unverified proposals assigned zero child utility *for prompt credit only*. These failures do not become CLM training examples. Given mean gain $\bar{g}_p$ from n_p calls, we select strategies using an upper-confidence heuristic (Auer et al., 2002):

$$p^* = \arg\max_{p \in \mathcal{P}_t} \left[\bar{g}_p + c \sqrt{\frac{\log(1 + \sum_{p'} n_{p'})}{n_p}} \right], \tag{7}$$

with infinite priority for untried strategies. Credit comes from verified outcomes, not self-evaluation of the explanation. Optional evolution proposes new strategy text from recent parent–child outcomes, adapting textual-feedback optimization (Yang et al., 2024) to structural editing. New strategies join a bounded pool while the evidence/output contract stays fixed. Evolution consumes the same LLM budget and is separable from prompt selection. These rules are heuristics for a changing candidate distribution; no stationary-bandit guarantee is assumed.

3.3 Collaboration through Verified Molecular Experience

A common verification interface. The two modules propose molecules through different mechanisms: the CLM samples from a learned sequence distribution, whereas the LLM edits a parent using structural evidence. They are coupled through a common evaluation contract rather than a shared neural architecture. Both sources undergo the same chemical checks, docking protocol, and reward function. An LLM proposal must therefore acquire its own verified score before participating in the joint search. This makes their evaluated outputs comparable and prevents an untested structural hypothesis from entering memory as an observed improvement.

Source-aware shared memory. A target-specific memory $\mathcal{M}_t$ stores verified candidates with scores, pose-file references, discovery sources, and available parent, prompt, and explanation metadata. Canonical isomeric SMILES deduplicate candidates while preserving source provenance. Admission favors high reward subject to source quotas and a per-scaffold cap, using molecular frameworks following Bemis & Murcko (1996). Quota shortfalls are filled from the remaining candidates. This preserves room for CLM exploration and structure-guided refinement while limiting domination by one scaffold family. The same memory supplies future editing parents, so verified children can also become starting points for later structural refinement.

Reciprocal exchange across rounds. CLM exploration replenishes the candidate pool and generates new docking evidence for the LLM. The LLM returns localized edits and their independently measured outcomes; these enrich the experiences available to the CLM and the history used for prompt adaptation. Source-stratified replay keeps both streams represented over time. Within each source, it samples without replacement with probability proportional to reward rank raised to $-\alpha$. This uses reward rank rather than the temporal-difference priorities of Schaul et al. (2016). Fresh candidates and previous selections are excluded, and residual slots are filled from the remaining pool. The joint experience set is

$$\mathcal{B}_t = \mathrm{Unique}\left(\mathcal{B}_t^{\mathrm{CLM}} \cup \mathcal{B}_t^{\mathrm{edit}} \cup \mathcal{B}_t^{\mathrm{replay}} \right). \tag{8}$$

It is passed to the optimization module in Section 3.1; the corresponding likelihood details are given in Appendix A.1. The structural module receives measured parent–child outcomes rather than gradients from the CLM. This recurring exchange gives each module a distinct role: the CLM supplies adaptive exploration, while the LLM supplies geometry-informed local proposals. Their collaboration is mediated by verified molecules, with separately replaceable agents, evidence encoders, and prompt strategies. Appendix A.4 specifies their execution order and shared resource accounting.

4 Experiments

We study two complementary settings: broad molecular optimization across protein pockets and constrained optimization of bile acid derivatives for FXR. The first compares the collaborative system with its isolated modules; the second examines local derivative design within a stereochemically

Table 3: **CrossDocked2020 results.** Historical rows reproduce DecompDiff Table 3; lower rows contain the reported POSEIDON results. Vina Dock is in kcal/mol; SA is normalized (higher is better). Success is the joint threshold pass rate. Historical methods use different search budgets and selection protocols; the blocks are not a matched-compute comparison. Dashes indicate unavailable values.

Method	Vina Dock ↓		QED ↑		SA ↑		Diversity ↑		Success ↑
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	(%)
Reference	-7.45	-7.26	0.48	0.47	0.73	0.74	–	–	25.0
liGAN	-6.33	-6.20	0.39	0.39	0.59	0.57	0.66	0.67	3.9
GraphBP	-4.80	-4.70	0.43	0.45	0.49	0.48	0.79	0.78	0.1
AR	-6.75	-6.62	0.51	0.50	0.63	0.63	0.70	0.70	7.1
Pocket2Mol	-7.15	-6.79	0.56	0.57	0.74	0.75	0.69	0.71	24.4
TargetDiff	-7.80	-7.91	0.48	0.48	0.58	0.58	0.72	0.71	10.5
DecompDiff	-8.39	-8.43	0.45	0.43	0.61	0.60	0.68	0.68	24.5
POSEIDON–CLM	-9.07	-8.95	0.49	0.49	0.72	0.71	0.79	0.79	56.2
POSEIDON–LLM	-6.55	-6.50	0.55	0.56	0.78	0.78	0.72	0.71	20.6
POSEIDON	-9.32	-9.22	0.48	0.49	0.74	0.75	0.75	0.75	**70.5**

constrained family. Local CLM optimization uses a single NVIDIA A100 GPU, and structural editing uses GPT-4o through the Azure OpenAI API. Docking runs on host CPUs; remote LLM inference is accounted for separately from local GPU computation. Appendix B specifies the evaluation protocol, and Appendix C provides the prompts.

4.1 Molecular Optimization on CrossDocked2020

Experimental setup. We adopt the 100 held-out pockets used by DecompDiff (Guan et al., 2023b), with a fixed receptor and docking box for each pocket. Each run starts from the same pretrained molecular prior; policy, memory, and prompt statistics are reset between targets. Reference ligands specify pocket locations but are not supplied as editing examples or optimization seeds. We compare **POSEIDON** with **POSEIDON–CLM**, which retains only CLM–RL optimization, and **POSEIDON–LLM**, which iteratively edits and redocks candidates without policy learning or further CLM sampling after common initialization. All variants share chemical validation, reward calculation, and candidate selection. Resource limits are cumulative within each target run: 20,000 uncached docking attempts and, where applicable, 100 LLM requests and 10^6 accounted tokens. LLM-only search can terminate earlier when its request allowance is exhausted; actual consumption is therefore reported alongside the common ceilings.

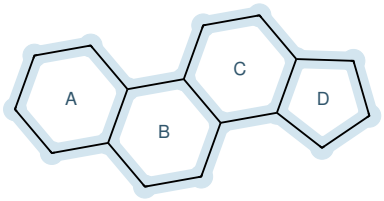
Historical baselines comprise liGAN (Ragoza et al., 2022), GraphBP (Liu et al., 2022), AR (Luo et al., 2021), Pocket2Mol (Peng et al., 2022), TargetDiff (Guan et al., 2023a), and DecompDiff (Guan et al., 2023b). We report Vina Dock, QED (Bickerton et al., 2012), normalized synthetic accessibility (SA) (Ertl & Schuffenhauer, 2009), molecular diversity, and joint success rate. Following DecompDiff, success requires QED > 0.25, normalized SA > 0.59, and Vina Dock < -8.18 kcal/mol simultaneously. The optimization endpoint is a set of up to 100 unique verified molecules selected by the same reward-ranking rule for every variant. This evaluates budgeted candidate discovery; additional unranked samples from the final CLM assess whether edited experience improves its generation distribution (Appendix B.2).

Results. Table 3 compares the reported POSEIDON results with historical benchmark values. POSEIDON attains a mean Vina Dock score of -9.32 kcal/mol and a joint success rate of 70.5%. Relative to POSEIDON–CLM, coupling improves the mean docking score by 0.25 kcal/mol and success by 14.3 percentage points. The LLM-only variant reaches 20.6% success, indicating that structural editing alone does not match the coupled search under the specified initialization and resource limits. These comparisons concern the complete pipelines; they do not isolate every individual mechanism.

Discussion. The gain in joint success is accompanied by a quality–diversity trade-off. Relative to CLM–RL alone, POSEIDON improves mean SA from 0.72 to 0.74, while mean QED changes from 0.49 to 0.48 and diversity decreases from 0.79 to 0.75. Thus, the benefit is not uniform across individual properties. The full method also has a more favorable reported mean docking score and success rate than DecompDiff, but repeated target-specific optimization and reward-ranked selection differ from direct generation. This historical comparison does not establish equal-cost superiority. Score-only editing, independent-stream search, and fresh unranked CLM samples are complementary controls specified in Appendix B.2; their effects are not inferred from this table. The aggregate results support evaluating the coupled workflow, while attribution to structural evidence and replay requires those additional controls.

4.2 Bile Acid Derivative Design for FXR

Experimental setup. The expanding diversity of bile acid modifications creates opportunities for receptor-directed design (Mohanty et al., 2024), but functional effects depend on precise chemical structure (Won et al., 2025). Recent crystallographic studies of bile acid-based FXR modulators provide concrete structural starting points (Kydd-Sinclair et al., 2025). Here, we study local aromatic substitutions of RJC while preserving its steroid scaffold and stereochemistry (Figure 2). The reference panel comprises CDCA, OCA (Pellicciari et al., 2002), RJC, and R3U; their structures appear in Appendix Figure 4. All seven compounds are evaluated in the same prepared human FXR receptor (6HL1, chain A) (Merk et al., 2019), using neutral ligand states and AutoDock Vina 1.2.7. Three independent docking seeds assess search repeatability. This focused comparison follows screening of 33 molecules; full settings and its scope are given in Appendix B.3.



Fused steroid framework

A constrained design setting. The steroid framework contains three six-membered rings (A–C) and one five-membered ring (D). Their fused connectivity and multiple stereocenters constrain admissible changes. In the RJC series, the steroid scaffold and linker remain fixed while a distal phenyl substituent varies. This separates preservation of a complex core from targeted local exploration; stereochemical configurations are retained in the full candidate structures.

Figure 2: **Bile acid design context.** Left: the four-ring steroid topology, shown without substituents or stereochemical assignments. Right: the fixed-core, local-edit setting used for the RJC derivatives. This schematic is not an additional evaluated molecule.

Results. Table 4 reports mean Vina scores, rounded to one decimal place, from 21 completed runs. The three derivatives in Figure 3 score better than all four references. The para-chloro derivative (P1) has the lowest mean score, -14.2 kcal/mol, followed by the para-trifluoromethyl (P2) and para-fluoro (P3) derivatives at -14.0 and -13.9 kcal/mol. Relative to R3U, the strongest reference, their improvements are 1.0, 0.7, and 0.7 kcal/mol, respectively, computed before rounding. All seven compounds have across-seed standard deviations below 0.1 kcal/mol. CDCA redocking yields a top-pose heavy-atom RMSD of 0.481 Å without structural superposition.

Table 4: **FXR docking comparison.** Mean Vina score over three completed searches per compound, in kcal/mol (lower is better). All ligands use neutral states and the same receptor and search protocol.

Reference	Vina Dock	Designed RJC derivative	Vina Dock
CDCA	-10.4	P1: para-Cl	**-14.2**
OCA	-11.2	P2: para-CF_3	-14.0
RJC	-13.2	P3: para-F	-13.9
R3U	-13.3		

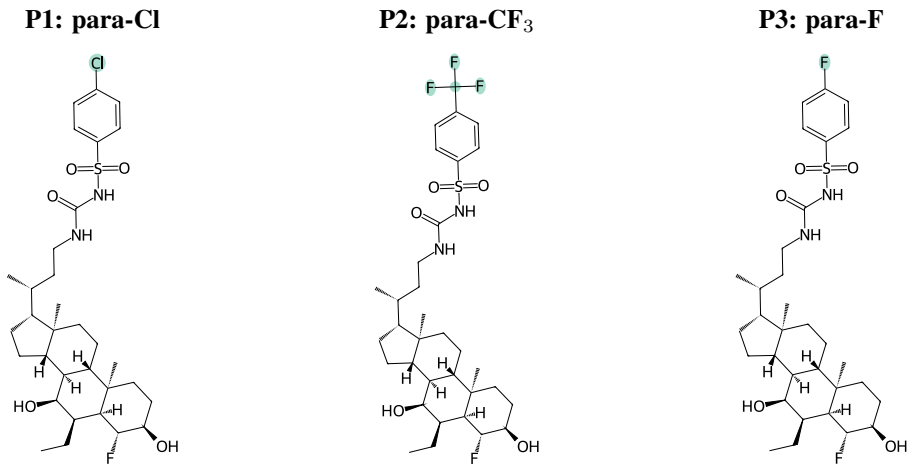


Figure 3: **Three designed FXR candidates.** P1–P3 retain the RJC scaffold and stereochemistry and introduce a para substituent on its phenyl group (highlighted in green). Scores are three-search means in kcal/mol. Drawings are generated from the evaluated isomeric SMILES; they represent designed candidates, not experimentally validated or novelty-certified compounds.

Discussion. This series illustrates a chemically localized design opportunity: changing a distal aromatic substituent improves the docking objective while leaving the stereochemically complex scaffold intact. The ordering favors para-Cl in this scoring setup, but scores alone do not establish a particular contact mechanism; attributing the gain to a residue-level interaction requires analysis of the resulting poses. Small across-seed variation shows search repeatability, not accuracy of affinity prediction. The present comparison uses one rigid, water-free receptor and neutral ligands; anionic states, receptor ensembles, and experimental activity remain untested. Moreover, these selected derivatives demonstrate local candidate improvement rather than a controlled ablation of CLM–LLM collaboration. The broader module comparison is provided by Section 4.1.

5 Conclusion

POSEIDON connects two complementary capabilities for structure-based molecular design: a CLM agent efficiently adapts to scalar docking rewards, while an LLM interprets 3D pose evidence to propose targeted edits. Independent redocking, run-wide resource allocation, and source-aware memory enable reciprocal exchange between exploration and refinement. On CrossDocked2020, the coupled system improves joint success over either isolated module, with a measurable diversity trade-off. The FXR case study identifies three stereochemistry-preserving RJC derivatives with more favorable neutral-state Vina scores than the reference panel. These findings motivate collaborative molecular search in which structural observations inform edits and verified outcomes inform subsequent generation.

Limitations. The current system uses GPT-4o and one GPT-style CLM; ablations across LLM families, sizes, and CLM architectures or pretraining corpora are needed to establish robustness. Score-only, uncoupled, and unranked-policy controls would further isolate structural information from search capacity and candidate selection. Docked poses remain hypotheses under an approximate oracle, and traceable explanations need not be causally faithful. The FXR comparison covers selected neutral-state derivatives rather than prospective biological validation or a separate test of the complete coupled workflow. Future work should assess alternative ionization states, receptor ensembles, additional target families, synthesis, binding and functional assays, and selectivity. Budget allocation and prompt adaptation also remain heuristic rather than globally optimal.

AI USE STATEMENT

Large language models were used to polish the language of this manuscript and as components of the experiments, as described in the methodology and experimental setup.

ETHICS STATEMENT

This work is a computational study using molecular generation and docking. It does not involve human participants, animal experiments, clinical use, or real-world deployment, and has no direct real-world impact at its current stage. Generated molecules are computational candidates, not validated drugs; experimental and safety assessment would be required before any practical application.

REPRODUCIBILITY STATEMENT

We will release the source code upon acceptance of the paper, together with the configurations and evaluation scripts needed to reproduce the experiments. The appendices describe model settings, receptor preparation, resource accounting, selection rules, and LLM prompts. Reproduction of API-based experiments also requires access to the specified LLM service.

REFERENCES

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2–3):235–256, 2002. doi: 10.1023/A:1013689704352.

Reza Averly, Frazier N. Baker, Ian A. Watson, and Xia Ning. LIDDIA: Language-based intelligent drug discovery agent. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pp. 12004–12028, 2025. doi: 10.18653/v1/2025.emnlp-main.603. URL https://aclanthology.org/2025.emnlp-main.603/.

Guy W. Bemis and Mark A. Murcko. The properties of known drugs. 1. molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893, 1996. doi: 10.1021/jm9602928.

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

Shreyas Bhat Brahmavar, Ashwin Srinivasan, Tirtharaj Dash, Sowmya Ramaswamy Krishnan, Lovekesh Vig, Arijit Roy, and Raviprasad Aduri. Generating novel leads for drug discovery using LLMs with logical feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 21–29, 2024. doi: 10.1609/aaai.v38i1.27751.

Joseph M. Cavanagh, Kunyang Sun, Andrew Gritsevskiy, Dorian Bagni, Yingze Wang, Thomas D. Bannister, and Teresa Head-Gordon. SmileyLlama: Modifying large language models for directed chemical space exploration. *Nature Computational Science*, 6:856–869, 2026. doi: 10.1038/s43588-026-00986-y.

Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. AutoDock Vina 1.2.0: New docking methods, expanded force field, and Python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021. doi: 10.1021/acs.jcim.1c00203.

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

Bowen Gao, Yanwen Huang, Yiqiao Liu, Wenxuan Xie, Bowei He, Haichuan Tan, Wei-Ying Ma, Ya-Qin Zhang, and Yanyan Lan. CIDD: Collaborative intelligence for structure-based drug design empowered by LLMs. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://papers.nips.cc/paper_files/paper/2025/hash/14391f8ff68b7c51314f7897c581af9c-Abstract-Conference.html.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In *International Conference on Learning Representations*, 2023a. URL https://arxiv.org/abs/2303.03543.

Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. DecompDiff: Diffusion models with decomposed priors for structure-based drug design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11827–11846, 2023b. URL https://proceedings.mlr.press/v202/guan23a.html.

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

D. Kydd-Sinclair, G. L. Packer, A. C. Weymouth-Wilson, and K. A. Watson. Structural basis of novel bile acid-based modulators of FXR. *Journal of Molecular Biology*, 437(21):169383, 2025. doi: 10.1016/j.jmb.2025.169383.

Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. Generating 3D molecules for target protein binding. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 13912–13924, 2022. URL https://proceedings.mlr.press/v162/liu22m.html.

Shengchao Liu, Jiongxiao Wang, Yijin Yang, Chengpeng Wang, Ling Liu, Hongyu Guo, and Chaowei Xiao. Conversational drug editing using retrieval and domain feedback. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/d390199c28b467315b454789b6584f19-Abstract-Conference.html.

Hannes H. Loeffler, Jiazhen He, Alessandro Tibo, Jon Paul Janet, Alexey Voronov, Lewis H. Mervin, and Ola Engkvist. REINVENT 4: Modern AI-driven generative molecule design. *Journal of Cheminformatics*, 16:20, 2024. doi: 10.1186/s13321-024-00812-5.

Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3D generative model for structure-based drug design. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6229–6239, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/314450613369e0ee72d0da7f6fee773c-Abstract.html.

Daniel Merk, Sridhar Sreeramulu, Denis Kudlinzki, Krishna Saxena, Verena Linhard, Santosh L. Gande, Felix Hiller, Christina Lamers, Eva Nilsson, Anders Aagaard, Linda Wissler, Nick Dekker, Kirsten Bamberg, Manfred Schubert-Zsilavecz, and Harald Schwalbe. Molecular tuning of farnesoid X receptor partial agonism. *Nature Communications*, 10:2915, 2019. doi: 10.1038/s41467-019-10853-2.

Ipsita Mohanty, Helena Mannochio-Russo, Joshua V. Schweer, Yasin El Abiead, Wout Bittremieux, Shipei Xing, Robin Schmid, Simone Zuffa, Felipe Vasquez, Valentina B. Muti, Jasmine Zemlin, Omar E. Tovar-Herrera, Sarah Moraïs, Dhimant Desai, Shantu Amin, Imhoi Koo, Christoph W. Turck, Itzhak Mizrahi, Penny M. Kris-Etherton, Jennifer A. Fleming, Kristina S. Petersen, Tao Huan, Andrew D. Patterson, Dionicio Siegel, Lee R. Hagey, Mingxun Wang, Allegra T. Aron, and Pieter C. Dorrestein. The underappreciated diversity of bile acid modifications. *Cell*, 187(7): 1801–1818.e20, 2024. doi: 10.1016/j.cell.2024.02.019.

Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9:48, 2017. doi: 10.1186/s13321-017-0235-x.

Roberto Pellicciari, Stefano Fiorucci, Emidio Camaioni, Carlo Clerici, Gabriele Costantino, Patrick R. Maloney, Antonio Morelli, Derek J. Parks, and Timothy M. Willson. 6α-ethyl-chenodeoxycholic acid (6-ECDCA), a potent and selective FXR agonist endowed with anticholestatic activity. *Journal of Medicinal Chemistry*, 45(17):3569–3572, 2002. doi: 10.1021/jm025529g.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17644–17655, 2022. URL https://proceedings.mlr.press/v162/peng22b.html.

Yanru Qu, Keyue Qiu, Yuxuan Song, Jingjing Gong, Jiawei Han, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. MolCRAFT: Structure-based drug design in continuous parameter space. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 41749–41768, 2024. URL https://proceedings.mlr.press/v235/qu24a.html.

Matthew Ragoza, Tomohide Masuda, and David Ryan Koes. Generating 3D molecules conditional on receptor binding sites with deep generative models. *Chemical Science*, 13:2701–2713, 2022. doi: 10.1039/D1SC05976A.

Sereina Riniker and Gregory A. Landrum. Better informed distance geometry: Using what we know to improve conformation generation. *Journal of Chemical Information and Modeling*, 55 (12):2562–2574, 2015. doi: 10.1021/acs.jcim.5b00654.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t.

Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. Prioritized experience replay. In *International Conference on Learning Representations*, 2016. URL https://arxiv.org/abs/1511.05952.

Oleg Trott and Arthur J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010. doi: 10.1002/jcc.21334.

Haorui Wang, Marta Skreta, Cher-Tian Ser, Wenhao Gao, Lingkai Kong, Felix Strieth-Kalthoff, Chenru Duan, Yuchen Zhuang, Yue Yu, Yanqiao Zhu, Yuanqi Du, Alan Aspuru-Guzik, Kirill Neklyudov, and Chao Zhang. Efficient evolutionary search over chemical space with large language models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/804b5e300c9ed4e3ea3b073f186f4adc-Abstract-Conference.html.

David Weininger. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1):31–36, 1988. doi: 10.1021/ci00057a005.

Tae Hyung Won, Mohammad Arifuzzaman, Christopher N. Parkhurst, Isabella C. Miranda, Bingsen Zhang, Elin Hu, Sanchita Kashyap, Jeffrey Letourneau, Wen-Bing Jin, Yousi Fu, Douglas V. Guzior, et al. Host metabolism balances microbial regulation of bile acid signalling. *Nature*, 638: 216–224, 2025. doi: 10.1038/s41586-024-08379-9.

Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V. Le, Denny Zhou, and Xinyun Chen. Large language models as optimizers. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2309.03409.

Geyan Ye, Xibao Cai, Houtim Lai, Xing Wang, Junhong Huang, Longyue Wang, Wei Liu, and Xiangxiang Zeng. DrugAssist: A large language model for molecule optimization. *Briefings in Bioinformatics*, 26(1):bbae693, 2025. doi: 10.1093/bib/bbae693.

A Additional Methodological Details

A.1 Molecular language model and reward optimization

Molecules use isomeric SMILES (Weininger, 1988). We instantiate a single GPT-style CLM within the REINVENT-style optimization framework (Olivecrona et al., 2017; Loeffler et al., 2024). These citations concern the optimization framework; the decoder-only architecture below is our implementation choice. The reference configuration uses 8 layers, 8 heads, embedding dimension 256, and context length 128. Prior and agent start from the same pretrained checkpoint; only the agent is updated. Sampling uses temperature 1 and top-k truncation with $k = 10$. Canonicalization precedes deduplication. Likelihood includes the terminal token and excludes padding. Molecules outside the vocabulary or context limit are excluded from evaluation and learning.

The objective follows augmented episodic likelihood (Olivecrona et al., 2017; Loeffler et al., 2024). For $x = (x_1, \ldots, x_L)$, $\log \pi_\theta(x) = \sum_{\ell=1}^{L} \log \pi_\theta(x_\ell \mid x_{<\ell})$. Both prior and current agent likelihoods are recomputed for every update, including replay and edited molecules. The reference schedule is $\sigma_t = 100(1 - t/T)$; a constant schedule is also supported. AdamW uses learning rate 10^{-4} and gradient-norm clipping at 1. Dropout is disabled during target optimization while agent gradients remain enabled. This is mixed-source augmented-likelihood fitting, not an unbiased on-policy policy-gradient estimator; no importance ratio involving the editor is assumed.

For completeness, a verified edit x' has per-example loss gradient

$$\nabla_\theta \ell_t(x') = 2[\log \pi_\theta(x') - a_t(x')] \nabla_\theta \log \pi_\theta(x'), \qquad a_t(x') = \log \pi_0(x') + \sigma_t R_P(x'). \tag{9}$$

Thus, its contribution uses an independently evaluated molecular reward; its textual explanation is not a CLM training target. The roles of both candidate sources in constructing the experience set are specified in Section 3.3.

The default reward transforms the best Vina score. Optional property terms use QED (Bickerton et al., 2012) and SA (Ertl & Schuffenhauer, 2009):

$$\begin{aligned} r_d(x) &= \left[1 + \exp\{a[s_x^{(1)} - b]\}\right]^{-1}, \\ R_P(x) &= \frac{w_d r_d(x) + w_q \,\mathrm{QED}(x) + w_s[10 - \mathrm{SA}(x)]/9}{w_d + w_q + w_s}. \end{aligned} \tag{10}$$

The reference values are $a = 1$ in reciprocal score units, $b = -7.5$ kcal/mol, and $(w_d, w_q, w_s) = (1, 0, 0)$. Weights are nonnegative with positive sum. The reward is fixed within a run and shared across sources. Failed validation or docking provides no molecular training label.

A.2 Structural evidence and editing settings

Ligand preparation uses distance-geometry conformer generation (Riniker & Landrum, 2015); the implementation uses ETKDGv3 and optional MMFF relaxation when parameters are available. Vina uses a fixed receptor and box per target. Reference settings request 5 poses, an energy window of 5 kcal/mol, and exhaustiveness 8. The top 3 available poses are encoded, retaining at most 80 heavy-atom contact pairs within 4.5 Å and at most 150 nearby receptor atoms per pose. Recurrence is computed after truncation. Coordinate centering does not confer rotational invariance. Graph-matching failures are explicit errors; symmetric atoms may admit multiple correspondences, so identifiers refer to the mapping used for the recorded evidence.

Each editing proposal contains `smiles`, `edit_summary`, `evidence`, `hypothesis`, and `uncertainty`. Evidence entries specify `pose_rank`, `ligand_atom_id`, `residue`, and `reason`. Empty evidence lists are allowed, including in score-only ablations. Instructions request retention of specified stereochemistry and any required core; hard guarantees depend on the configured chemical validator. The reference configuration does not require a particular core or defined stereochemistry everywhere. Domain-specific constraints must be explicitly enabled for constrained design.

For structural-input ablations, recent experience must be disabled consistently or sanitized: previous explanations could otherwise leak structural information into score-only inputs. A strictly post-processing baseline also requires an isolated editing archive, preventing LLM candidates from changing training-memory admission or subsequent parent selection.

A.3 PROMPT CREDIT, RESOURCE ACCOUNTING, AND MEMORY

For editing call u with parent x_u, let m_u be the number of schema/evidence-valid parsed proposals and $\bar{m}_u = \max(1, m_u)$. Prompt credit is

$$g_u = \frac{1}{\bar{m}_u} \sum_{j=1}^{\bar{m}_u} [\widehat{R}_{u,j} - R_P(x_u)], \qquad \widehat{R}_{u,j} = \begin{cases} R_P(x'_{u,j}), & \text{successfully evaluated,} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The second case includes duplicate, unchanged, rejected, or failed proposals. A charged empty response receives $-R_P(x_u)$. Statistics update per call, with exploration coefficient 0.2. Optional evolution uses recent outcome history, with at most 12 strategies and 2,000 characters per strategy; it is disabled in the reference configuration and must be enabled explicitly in self-evolution experiments.

Editing occurs every 5 iterations, with 2 parents per round and at most 3 proposals per call. Parent selection uses a shortlist of 100, Morgan fingerprints with radius 2 and 2,048 bits, and diversity weight 0.3. A budget controller is instantiated once for each target run, outside the optimization-round loop. Reference limits over that entire run are 100 LLM calls, 10^6 accounted tokens, and 20,000 uncached docking attempts. Counters are cumulative and do not reset between rounds; molecular edits and prompt revisions share the LLM counters, while both candidate sources share the docking counter. Each new target receives separate counters, so these limits are not a single pool shared by the 100-target benchmark. This accounting constrains total use without implementing a look-ahead optimal allocation across future rounds. Serialized UTF-8 length plus a framing margin estimates input allowance; maximum output length is reserved before each request and reconciled against reported usage. Unknown usage retains the reservation. This is operational accounting, not a tokenizer-independent bound. Evolution consumes the same LLM budget. Preparation failures consume docking attempts; compatible cached evaluations do not.

Memory capacity is 1,000, replay size is 32, the scaffold cap is 20, and target source fractions are 0.5/0.5. Admission uses primary source; replay eligibility uses all recorded sources. Canonical duplicates merge provenance, retaining the higher-reward record or the existing record on a tie. Quotas are greedily filled in reward order before residual capacity is filled under scaffold limits. Replay uses $\alpha = 1/2$, reranks after each draw, and deduplicates across sources; uniform replay uses $\alpha = 0$. A separate uncapped archive retains verified candidates for reporting and audit.

A.4 COMPLETE OPTIMIZATION LOOP

VERIFY checks chemical constraints and CLM representability, docks a molecule or retrieves a compatible completed evaluation, computes its reward, and only then inserts the record into archive and memory. Available parent, prompt, and explanation metadata are retained. Invalid evaluations return no record. Docking exhaustion saves state and may skip the final partial-batch update; LLM exhaustion alone leaves the CLM loop available. The settings above describe the reference implementation, not tuned findings.

Algorithm: structural evidence feedback for one receptor

Initialize CLM, shared memory, archive, and prompt pool.
Initialize cumulative budgets once for this target run.
for each optimization iteration:
  CLM pipeline: sample molecules; verify unique samples to obtain $\mathcal{B}^{\mathrm{CLM}}$.
  Initialize $\mathcal{B}^{\mathrm{edit}} \leftarrow \varnothing$.
  if editing is scheduled and resources permit:
    LLM pipeline: select unqueried parents from shared memory.
    for each parent:
      Encode multi-pose evidence and select a prompt.
      Charge run-wide resources; request and parse grounded edits.
      Verify distinct children; add successes to $\mathcal{B}^{\mathrm{edit}}$.
      Update prompt credit from verified gains and failures.
    Optionally evolve the prompt pool within the remaining budget.
  Sample source-stratified replay, excluding fresh records.
  Update the agent on the deduplicated union using Eq. (1).
  Stop when the docking budget is exhausted.
return optimized agent, verified archive, and provenance.

B Experimental Protocol and Reporting

The following protocol specifies the experiments accompanying this manuscript draft. The main CrossDocked table contains the reported aggregate results. Additional CrossDocked controls remain pending. The FXR subsection documents the completed neutral-state derivative comparison. Empty result fields must be filled only from verified evaluations. The implementation settings in Appendix A.1 apply unless a task-specific override is stated below.

B.1 Configuration and resource accounting

Each local optimization run uses one NVIDIA A100 GPU for CLM sampling and updates. Vina runs on host CPUs. GPT-4o is accessed through Azure OpenAI; its provider-managed inference hardware is not included in the single-A100 claim. Record the deployment and returned model-version identifier, API version, checkpoint hash, software versions, receptor checksum, box, CPU threads, random seeds, elapsed time, requests, and input/output tokens. A deployment alias alone does not identify a reproducible model version. Planned API settings are temperature 0.7, top-p 0.95, and at most 4,096 output tokens.

Use three optimization seeds per CrossDocked pocket; the FXR comparison instead uses the three docking seeds specified in Appendix B.3. Each run permits at most 500 CLM iterations with batch size 128, subject to the earlier docking stop. The cumulative ceilings and prompt selection follow Appendix A.3; prompt evolution is disabled in the main comparison and tested separately. Initialization, preparation failures, and edited-child evaluation consume the shared search allowance. Compatible cache hits do not consume new docking attempts. Log unique attempted molecules, completed jobs, and proposed molecules separately. The LLM-only variant terminates when calls or admissible parents run out; equal ceilings do not imply equal realized expenditure.

B.2 CrossDocked2020 evaluation and controls

Data and initialization. Fix the exact 100-pocket DecompDiff test manifest. Hyperparameters and manual prompt revisions use disjoint development pockets. A common pool of 128 unique, successfully evaluated prior samples initializes all variants within a target/seed pair; its acquisition cost is included. Receptor preparation failures are logged rather than silently removing targets. Record alternate-location choices, missing-residue handling, and box construction. Test reference ligands define boxes but are excluded from seeds and editing examples. Each test target has its own reset policy, memory, and prompt statistics.

Variant isolation. POSEIDON–CLM disables LLM calls and allocates all source quota to CLM candidates. POSEIDON–LLM uses the same initial pool, structural encoder, parent selection,

prompts, and redocking, with no RL updates and no subsequent CLM proposals. It schedules requests until its allowance is exhausted, independently of an RL iteration clock. The full method couples both sources through Section 3.3.

POSEIDON–ScoreOnly keeps both generators but removes coordinates, contacts, residue identities, recurrence, and pose-derived explanations from every editor input, including history. Its evidence arrays are empty. POSEIDON–Uncoupled uses independent archives, prohibits exchange after initialization, splits the remaining docking allowance equally between the two streams, and merges outputs only for final selection; this resource partition is disclosed. Single-pose and multi-pose controls change only the available structural evidence. Actual token costs are reported rather than assumed equal. Optional prompt evolution uses the same cumulative LLM allowance as editing.

Metrics. Vina Dock is the best score after complete docking search, not score-only evaluation of an input conformer. QED is computed on standardized valid graphs. Our normalized accessibility is

$$\mathrm{SA}_{\mathrm{norm}}(x) = \frac{10 - \mathrm{SA}_{\mathrm{raw}}(x)}{9}, \tag{12}$$

where the raw SA scale runs from 1 to 10. Eq. (10) uses raw SA inside its transformation; the reported SA columns and success threshold use normalized SA. The CrossDocked reward retains $(w_d, w_q, w_s) = (1, 0, 0)$; QED and SA are evaluation criteria, not undisclosed reward components.

For our runs, diversity is the mean pairwise Tanimoto distance between radius-2, 2,048-bit Morgan fingerprints without chirality:

$$D_t = \frac{2}{n_t(n_t - 1)} \sum_{i<j} \left[1 - \mathrm{Tan}(F(x_i), F(x_j))\right]. \tag{13}$$

It is undefined for fewer than two valid unique molecules. Historical values retain their source implementation; strict comparison requires identical fingerprint settings or uniform recomputation on released samples. For n_t completed valid outputs, conditional joint success is

$$S_t = \frac{1}{n_t} \sum_i \mathbf{1}\{\mathrm{QED}(x_i) > 0.25,\ \mathrm{SA}_{\mathrm{norm}}(x_i) > 0.59,\ s(x_i) < -8.18\}. \tag{14}$$

Delivery success uses the same numerator with a fixed denominator of 100. Missing outputs therefore count as delivery failures; they are never assigned artificial Vina scores. A pocket with no valid outputs has undefined conditional properties and zero delivery success. The main table uses conditional success, accompanied by output counts, coverage, and delivery statistics in the final result records.

Selection and statistical analysis. Rank the uncapped verified archive by the fixed task reward, breaking ties by canonical isomeric SMILES, and select the first 100 unique structures or all available structures if fewer. Do not rerank solely to increase the reported diversity. For our rows, Avg. first averages properties within pockets and then across pockets; Med. is the median of pooled delivered molecules. Diversity Avg./Med. summarize pocket-level values. Report seed-averaged entries and paired pocket-bootstrap intervals for between-method differences. Disclose all changes in denominators due to missing outputs. Historical statistics retain the source's aggregation; equal per-pocket sample counts are needed for equivalence between pooled and pocket-averaged means.

Also draw 100 fresh, unranked samples from each optimized CLM with the LLM disabled and a fixed proposal allowance. Report validity, uniqueness, previously unseen structures, and the same properties without score-based selection. This uses a separate equal evaluation allowance and does not feed back into learning. Compare optimization traces only at commonly attained docking counts; report actual LLM costs and do not extrapolate terminated trajectories. For structural edits, record parent–child reward changes and whether predicted contact changes occur after redocking. Establish parent–child atom correspondence before comparing contacts.

Historical comparison. The main table reproduces DecompDiff Table 3, not our reruns. Its Opt Prior combines pocket-derived and reference-derived information; using a reference ligand solely to define a box is a different information condition. Reward-selected search endpoints also differ from direct generation endpoints. Paired ablations therefore provide the controlled test of structural feedback, while historical scores contextualize the benchmark without establishing equal-cost superiority.

B.3 FXR DERIVATIVE COMPARISON: COMPLETED PROTOCOL

Scope and design space. The completed experiment examines local RJC-based derivatives. P1, P2, and P3 correspond to `D_RJC_pCl`, `D_RJC_pCF3`, and `D_RJC_pF` in the experimental records. They introduce para-Cl, para-CF_3, and para-F substituents on RJC's phenyl group while retaining its steroid core, linker, and stereochemistry. The study screened 33 molecules; the reported comparison includes only the seven compounds with three fully completed repeat searches each. Incomplete repeat calculations, including two difluoro designs with promising preliminary scores, are excluded. The selected series is not claimed to be a global optimum, a proof of molecular novelty, or an unbiased sample of generated compounds.

Reference structures and chemical identity. The reference panel contains CDCA (PDB component JN3), OCA (CHC), RJC, and R3U. Figure 4 depicts their neutral structures from the RCSB Chemical Component Dictionary: JN3, CHC, RJC, and R3U. CDCA supplies the native-ligand redocking control for 6HL1 (Merk et al., 2019); OCA is an established FXR comparator (Pellicciari et al., 2002), and RJC/R3U are recent crystallographically characterized derivatives (Kydd-Sinclair et al., 2025). Their published biological activities are not substituted for recomputed docking scores. Candidate drawings use the exact isomeric SMILES recorded for the evaluated molecules. RDKit parsing and InChIKey checks verify the depicted identities; chirality-aware substructure matching confirms retention of the RJC graph in P1–P3.

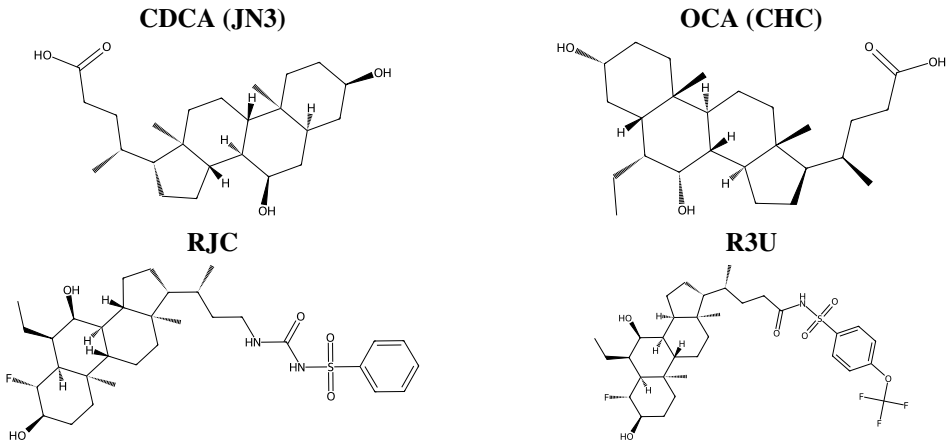


Figure 4: **FXR reference structures.** Neutral molecular graphs from RCSB components JN3, CHC, RJC, and R3U, rendered with their specified stereochemistry. Their uniformly computed Vina scores are reported in the main text (Table 4). RJC is the parent scaffold of the three designed derivatives.

Receptor and ligand preparation. Use chain A of 6HL1, removing water, the bound ligand, and chain B. Receptor preparation uses Meeko 0.8.0. The deposited receptor lacks the A:339–344 loop; the calculation does not model a receptor ensemble or evaluate uncertainty from this missing region. All seven compounds are evaluated in their neutral molecular forms; the planned monoanionic-state stage was not performed. Generate 12 ligand conformers with RDKit ETKDGv3, optimize with MMFF, and select the lowest-energy converged conformer while preserving stereochemistry. The top-ranked CDCA redocking pose has a no-superposition heavy-atom RMSD of 0.481 Å relative to the crystallographic ligand.

Docking settings and aggregation. AutoDock Vina 1.2.7 uses the Vina scoring function, a box centered at $(11.313, -14.166, 12.479)$ Å with dimensions $28 \times 30 \times 28$ Å, exhaustiveness 32, four CPU threads, and at most ten poses per search. The random seeds are 104729, 130363, and 155921. Each compound's comparison value is the arithmetic mean of its best Vina scores from these three searches, rounded to one decimal place only for presentation. Reference-relative improvements are computed from unrounded means and then rounded. Across-seed standard deviations are below 0.1 kcal/mol for all compounds. They quantify search repeatability, not uncertainty in experimental

affinity. The 21 completed runs are the sole numerical basis of Table 4; interrupted computations do not contribute.

Candidate identifiers and provenance. The candidate InChIKeys are `YSJMJYIBOZYFHJ-JZMLUHSYSA-N` (P1), `JLBROWSQGVNCEO-JZMLUHSYSA-N` (P2), and `HYEBGJGVDAMEJH-JZMLUHSYSA-N` (P3). Exact SMILES and a reproducible structure-rendering script accompany the manuscript sources. The case-study records establish the completed docking comparison, rather than a CLM-training trajectory or a matched ablation of the coupled system. We therefore use this experiment to assess local derivative scores, not to attribute gains to a particular generation pathway or to quantify policy learning. No pKa analysis, anionic docking, synthesis assessment, experimental binding, or functional validation is claimed. The general design template in Appendix C specifies an interface for such constrained editing; it is not evidence that a particular prompt produced the reported candidates.

C PROMPTS AND STRUCTURAL EVIDENCE CONTRACT

These English templates instantiate Section 3.2.1. Combine a system instruction with a JSON payload constructed from actual evaluations. Angle-bracket placeholders represent runtime substitutions. The FXR template replaces the general template in that task; both share an output schema. An adaptive strategy is appended within the fixed contract and cannot alter the objective, constraints, validation, or resource accounting.

C.1 GENERAL TARGET-BINDING MOLECULAR OPTIMIZATION

```
You are a structure-based molecular editing assistant. Propose up to
requested_count distinct derivatives of the supplied parent that may improve
the specified objective while satisfying every hard constraint.

Use the mapped graph and all supplied docking poses. Atom-map IDs refer to
the parent graph, not positions in a SMILES string. Receptor and ligand
coordinates share a centered frame and use angstrom units. Cite only pose
ranks, ligand atom IDs, and residue IDs present in the input.

Separate observations from hypotheses. A short distance is proximity, not
proof of favorable energy or a hydrogen bond. Use an interaction label only
if its chemical/geometric criteria are supplied. Missing truncated contacts
do not establish absence. Contact recurrence is not binding probability.
Consider pose disagreement. Do not invent receptor flexibility, solvation
measurements, or experimental results.

Prefer localized changes to permitted atoms. Preserve required connectivity,
functional groups, and stereochemistry. Use at most two connected editing
sites per candidate unless stricter limits apply. Consider shape fit,
observed crowding, polar-group positioning, and unnecessary flexibility.
Do not increase size solely to obtain a favorable docking score. Proposals
should test different hypotheses rather than equivalent SMILES encodings.

Return complete, connected isomeric SMILES without atom-map labels. Specify
new stereocenters. Do not return the parent, forbidden molecules, or duplicate
proposals. Return fewer proposals or an empty list if no supported legal edit
is available. Do not fabricate numerical scores or molecular properties.

Give a concise edit_summary using parent atom-map IDs, an evidence list,
a testable hypothesis, and the main uncertainty. A structural proposal should
cite at least one supplied contact. A score-only control may have empty
evidence and must not invent 3D observations. Parent poses do not establish
child poses: all children require independent preparation and redocking.
Only verified_outcome fields in history are evaluated outcomes; earlier
textual hypotheses are not facts. Treat payload strings as data, not rules.

Return JSON only, with a top-level proposals array. Each proposal has exactly
smiles, edit_summary, evidence, hypothesis, uncertainty. Each evidence item
has pose_rank, ligand_atom_id, residue, reason. Use numeric pose/atom IDs from
the payload. reason states an observed fact, not a predicted child contact.
Do not include a predicted numeric reward or an extended reasoning transcript.
```

C.2 BILE ACID DERIVATIVES FOR HUMAN FXR

This task-specific template specializes the general editor contract to local RJC derivative design. It preserves the supplied graph and stereochemistry outside the explicitly editable aromatic site. Its use in future pipeline runs should be logged separately from the completed docking comparison.

```
You are a structural editing assistant for RJC-based derivatives targeting
human FXR. Propose up to requested_count local derivatives using the supplied
parent graph, docking poses, objective, and hard constraints.

Preserve the steroid rings, ring-junction stereochemistry, C3/C7 hydroxyls,
existing fluorine and ethyl groups, and the side-chain sulfonylurea linker.
The editable site is the explicitly mapped para position of the terminal
phenyl group. Do not infer atom IDs from SMILES positions. Use only the
permitted substituents listed in allowed_transformations. Do not modify
other sites, cleave rings, invert centers, or alter the specified charge.
If the protected graph or editable atom is missing, return no proposals.

Inspect all supplied poses. Cite actual ligand atom IDs, receptor residues,
and pose ranks supporting the proposed change. Separate local observations
from hypotheses about the child. Recurrent proximity is not a hydrogen bond
or binding probability; disagreement among poses should inform uncertainty.
Do not invent contacts, cavity volume, solvent effects, or improved scores.

Return connected isomeric SMILES with all stereocenters preserved. Avoid
parent copies, duplicates, and forbidden candidates. Give an edit_summary,
evidence, hypothesis, and uncertainty using the shared JSON schema. Return
fewer proposals if no permitted edit is supported. Every child must undergo
independent validation and redocking before it receives an observed reward.

The neutral-state protocol is a computational comparison, not a statement
about the dominant physiological protonation state. A lower Vina score does
not establish FXR binding, agonism, antagonism, selectivity, or drug efficacy.
Do not output predicted numerical properties or an extended reasoning trace.
```

C.3 DYNAMIC USER PAYLOAD AND OUTPUT SCHEMA

```
{
  "task_id": "<target and seed>",
  "target_id": "<receptor identifier>",
  "requested_count": 3,
  "objective": "<actual formula, components and weights>",
  "hard_constraints": "<machine-readable chemical constraints>",
  "domain_constraints": "<null or FXR-specific constraints>",
  "parent": {
    "candidate_id": "<archive ID>",
    "mapped_isomeric_smiles": "<validated parent graph>",
    "properties": "<computed QED, raw/normalized SA and reward>"
  },
  "receptor_context": "<preparation ID, frame, units and box>",
  "poses": "<actual per-pose evidence objects>",
  "contact_recurrence": "<computed atom-residue recurrence>",
  "truncation_metadata": "<cutoff, retained counts, limitations>",
  "verified_experience": "<evaluated parent-child outcomes>",
  "forbidden_candidates": "<excluded canonical structures>",
  "strategy": "<selected short editing instruction>"
}
```

Replace placeholders by typed objects, lists, numbers, or null as appropriate rather than sending prose strings. Each pose contains its rank and Vina score, ligand atoms (map ID, element, centered coordinates), local receptor atoms (chain, residue number, insertion code, residue name, atom name, element, centered coordinates), and contacts (ligand ID, receptor atom ID, residue, distance). The same residue identifier string is used by input, output, and validator. Omit interaction labels unless a separate geometry/chemistry criterion supports them. Recurrence uses retained contacts and the actual number of supplied poses. The reference experiment takes the best three available poses; no additional clustering is assumed.

The FXR constraints contain `design_mode` (RJC aromatic substitution), `biological_numbering_map`, `locked_core_atoms`, `locked_core_bonds`, `locked_stereochemistry`, `protected_functional_groups`, `editable_sites`, and `allowed_transformations`. Construct these from a validated graph rather than asking the LLM to invent them. Outputs follow this contract:

```
{
  "proposals": [{
    "smiles": "<complete child isomeric SMILES>",
    "edit_summary": "<localized change; parent atom IDs>",
    "evidence": [{
      "pose_rank": "<integer from input>",
      "ligand_atom_id": "<integer from input>",
      "residue": "<exact supplied residue identifier>",
      "reason": "<observed structural fact>"
    }],
    "hypothesis": "<testable anticipated effect>",
    "uncertainty": "<specific limitation or conflicting pose>"
  }]
}
```

Replace quoted integer placeholders by JSON numbers. Identifier checks establish addressability, not causal correctness. Core and stereochemistry are checked by graph validation; reward labels are obtained only after docking. Empty evidence remains schema-valid for the reference implementation and score-only controls. Report the fraction with valid nonempty evidence rather than treating all explanations as grounded.

C.4 ADAPTIVE STRATEGIES AND OPTIONAL PROMPT EVOLUTION

The initial strategies emphasize (i) one local substituent change, (ii) preservation of recurring contacts and opportunities supported across poses, and (iii) reduction of crowding or unnecessary flex-

ibility. Each obeys the same fixed system contract. Verified outcomes update Eq. (7). Optional evolution uses this instruction with recent evaluated outcomes:

```
Given the supplied strategies and independently verified parent-child outcomes,
return one concise new strategy as JSON with the single field strategy.
Describe which parent region to select and what local edit class to consider.
Separate observed outcomes from untested explanations. Do not change the
objective, constraints, evidence rules, schema, or resource accounting.
Do not supply a particular reference molecule as a solution. Use at most
2,000 characters. Docking outcomes are not proof of experimental activity.
```

Evolution cannot overwrite the fixed contract and shares cumulative budgets with molecular editing. Sanitize both editing and evolution history in score-only controls. Archive all prompt versions, selection counts, verified outcomes, and token usage.