

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☒ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☒ ☐ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☒ ☐ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	SimpRetro (CPU-based version) was used for retrosynthetic analysis. AutoDock Vina (v1.2.3) was used for molecular docking. PLIP 2.3.1 was used for interaction profiling. The LLM panel comprised results from Gemini-2.5-Pro (Google API), GPT-5.2 (OpenAI API), and DeepSeek-V4-Pro / DeepSeek-V4-Flash (open-weight, herein accessed via DeepSeek API). Source molecules were collected from the GenBench3D repository.
Data analysis	RDKit (v2023.09.6) was used for all chemoinformatics calculations, molecule manipulation, and fingerprint generation (Morgan, ECFP4). The iterative-MCS graph-matching edit-path extraction and the JSON-driven edit-application/RDKit-sanitisation pipeline were implemented in Python and are released as part of the SynCraft-Core codebase. All analysis code is available on GitHub(https://github.com/catalystforyou/SynCraft-Core).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All datasets supporting this study are publicly available in the GitHub repository and have been deposited to figshare. The repository includes: (i) the Synthesis Cliff reference corpus; (ii) the unsynthesizable test sets for Pocket2Mol, ResGen, and RIPK1; (iii) the 47 TargetDiff M^{pro} candidate set used for the agent-skill demonstration; (iv) per-LLM inference outputs across the four-backend panel; and (v) systematic PLIP-retention tables and Vina-paired outputs for RIPK1 and M^{pro}.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

N/A

Reporting on race, ethnicity, or other socially relevant groupings

N/A

Population characteristics

N/A

Recruitment

N/A

Ethics oversight

N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences

☐ Behavioural & social sciences

☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The 'Synthesis Cliff' dataset comprises 3,332 pairs, selected from an initial pool of approximately 20,000 generated molecules (4,000 from each of 5 generative models) to ensure structural diversity. The test sets consisted of 1,025 and 1,715 molecules from ResGen and Pocket2Mol respectively, which provides sufficient statistical power to distinguish model performance. Additional cohorts include: 42 high-scoring RIPK1 candidates (191 source–edit pose pairs for PLIP retention); 47 TargetDiff M^{pro} candidates (260 source–edit pose pairs).

Data exclusions

Molecules that are already synthesizable are excluded as this research focuses on the unsynthesizable molecules.

Replication

The computational results are reproducible using the provided code and seeds. LLM inference (Gemini-2.5-Pro, GPT-5.2, DeepSeek-V4-Pro, DeepSeek-V4-Flash) used each provider's default temperature settings; the chemical edit-application pipeline is deterministic.

Randomization

Randomization was used when (i) sampling the initial pool of unsynthesizable molecules to construct the training/reference dataset; (ii) drawing exemplars for the random-RAG control ablation (matched to similarity-RAG in count and exemplar length), and (iii) sampling 100 cliff pairs for the LLM-direct edit-path audit.

Blinding

Blinding was not applicable as this is a computational benchmark study involving automated metrics calculation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.