TOPU-LBVS: A REALISTIC MULTI TARGET BENCHMARK FOR LIGAND BASED VIRTUAL SCREENING

Anonymous authors
Paper under double-blind review

ABSTRACT

Ligand-based virtual screening (LBVS) is a practical first-pass tool in early-stage drug discovery, but existing benchmarks can overestimate performance through random negatives, easy decoys, limited target coverage, and non-standardized evaluation protocols. We introduce **TopU-LBVS**, a multi-target benchmark for LBVS under hard-negative screening conditions. Starting from curated ChEMBL 35 bioactivity data, TopU-LBVS covers 93 protein targets across 7 protein classes and constructs target-specific screening libraries with property-matched, structurally similar decoys at a fixed 1:40 active-to-decoy ratio. Libraries contain roughly 400 to 10,000 compounds and are designed to reduce simple physicochemical and nearest-neighbor fingerprint shortcuts. TopU-LBVS provides three fixed protocols. **TopU-LBVS-full** evaluates ChEMBL* $\rightarrow$ TopU generalization across all 93 targets. **TopU-LBVS-low** evaluates low-data TopU $\rightarrow$ TopU learning within the hard-negative distribution. **TopU-LBVS-mini** provides a compact seven-target protocol with a paired random-decoy control that changes only the test decoys, enabling low-cost development and direct measurement of the gap between random ChEMBL* and TopU decoys. Across ten reference baselines spanning fingerprint methods, molecular GNNs, fingerprint hybrids, and modern molecular models, performance under random-decoy evaluation degrades sharply under hard-negative screening. We release data, fixed splits, evaluation code, and baseline implementations for reproducible comparison of future LBVS and molecular representation learning methods. Code and data are available at [1].

1 INTRODUCTION

Ligand-based virtual screening (LBVS) ranks candidate compounds using known active ligands and remains a practical first-pass tool in early-stage drug discovery, especially when structural information is limited but historical structure–activity relationship (SAR) data are available Lagarde et al. (2015); Walters & Wang (2020). Despite this importance, widely used LBVS benchmarks can substantially simplify the screening problem Sieg et al. (2019); Lagarde et al. (2015). Many rely on random negatives or easy decoys, have limited or uneven target coverage, and lack fixed learning protocols for supervised model comparison Chen et al. (2019); Kramer et al. (2025). Random or scaffold-agnostic splits can also place closely related analogues across train and test sets, leading to overly optimistic estimates of performance Hu et al. (2020); Yang et al. (2019). Consequently, it remains unclear which methods continue to work when candidate libraries contain property-matched, structurally similar hard negatives rather than random background compounds.

We introduce **TopU-LBVS**, a multi-target benchmark for ligand-based virtual screening under hard-negative conditions. Starting from ChEMBL 35, we curate a cleaned bioactivity resource, ChEMBL*, covering 93 protein targets across 7 protein classes. For each target, we construct a TopU screening library by pairing curated actives with hard decoys at a fixed 1:40 active-to-decoy ratio. TopU decoys are selected to match actives in physicochemical properties while remaining structurally similar, reducing simple property-based and trivial nearest-neighbour shortcuts. This design targets a central weakness of standard LBVS evaluation: models that perform well against randomly sampled negatives may fail on focused hard-negative screening libraries Chen et al. (2019); Sieg et al. (2019).

[1] Code: https://anonymous.4open.science/r/topu-lbvs-2D5A/
Data: https://huggingface.co/datasets/topu-benchmark/topu-lbvs

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.

TopU-LBVS is organized as a benchmark suite with three complementary protocols. **TopU-LBVS-full** evaluates historical-to-hard-library generalization: models are trained on target-specific ChEMBL* SAR data and tested on the corresponding TopU library across all 93 targets. **TopU-LBVS-low** evaluates low-data learning within the TopU hard-negative distribution, again across all 93 targets, where training, validation, and test compounds come from the same focused candidate library. **TopU-LBVS-mini** provides a compact seven-target version, one target per protein class, for rapid development, debugging, and low-cost comparison. Its TopU arm follows the same curation, training protocol, test ratio, evaluation metric, and evaluation code as TopU-LBVS-full.

A key component of TopU-LBVS-mini is a paired random-decoy control for hardness validation. For each mini target, we keep the same training data and validation protocol, the same test actives, and the same 1:40 test ratio, but replace TopU hard decoys with randomly sampled ChEMBL* decoys. The paired comparison isolates the effect of test-decoy construction and directly measures how much random-decoy evaluation can overestimate performance relative to hard-negative screening.

We evaluate ten baselines spanning classical fingerprints, molecular GNNs Hu et al. (2020); Xu et al. (2019); Veličković et al. (2018), fingerprint hybrids, and modern molecular models Yang et al. (2019); Ross et al. (2022). Recent studies likewise show that ECFP-based or similarly simple molecular representations can remain highly competitive with substantially more complex learned models (Praski et al., 2025; Boulougouri et al., 2024; van Tilborg et al., 2026), motivating their inclusion as strong reference baselines rather than weak classical controls. Across TopU-LBVS-full, Morgan-RF is the strongest overall baseline, while learned models become more competitive in the low-data protocol. All models use the same predefined splits and identical target-specific protocols. We use EF@1% for TopU-LBVS-full and TopU-LBVS-mini, and EF@10% for TopU-LBVS-low, where smaller enrichment budgets are unstable for the smallest test libraries. PR-AUC is used for validation and reported as a secondary metric; ROC-AUC, BEDROC, and additional enrichment metrics are reported in the appendix.

Contributions. We summarize the benchmark contributions as follows:

- **A hard-negative LBVS benchmark.** TopU-LBVS provides 93 target-specific tasks across 7 protein classes, using property-matched, structurally similar decoys at a 1:40 active-to-decoy ratio.
- **A curated ChEMBL* resource.** We release a standardized ChEMBL 35 subset with assay filtering, molecular standardization, activity binarization, artifact removal, and duplicate resolution.
- **Three fixed evaluation protocols.** TopU-LBVS-full tests ChEMBL* $\rightarrow$ TopU generalization; TopU-LBVS-low tests low-data TopU $\rightarrow$ TopU learning; and TopU-LBVS-mini provides a compact seven-target entry point for rapid development.
- **A paired random-decoy hardness control.** The mini protocol includes matched random ChEMBL* control splits that differ only in test-decoy construction, quantifying the gap between random-decoy and TopU hard-negative evaluation.
- **Reference baselines and open infrastructure.** We provide ten baseline results and release the data, fixed splits, implementations, and evaluation code for reproducible LBVS comparison.

2 Background and Existing LBVS Benchmarks

Desiderata for learning-based LBVS evaluation. A useful LBVS benchmark should combine four properties: *challenging negatives* that avoid simple physicochemical or structural shortcuts; *target diversity* across ligand chemistries and protein families; *screening-relevant imbalance*; and a *fixed reproducible protocol* with released splits, seeds, and evaluation code.

Classical decoy-based benchmarks. DUD-E (Mysinger et al., 2012) remains widely used for ligand- and structure-based virtual screening. Its decoys are physicochemically matched while avoiding close 2D similarity, but can often be separated by descriptors such as molecular weight or logP (Chen et al., 2019), and no canonical supervised-learning protocol or fixed splits are provided. DEKOIS 2.0 (Bauer et al., 2013) improves decoy construction but was likewise developed primarily for virtual-screening evaluation rather than standardized supervised comparison.

MUV (Rohrer & Baumann, 2009) is an important precursor to hard-negative evaluation. It uses experimentally screened PubChem compounds and spatial-statistical optimization to reduce analogue

Table 1: **Comparison with representative LBVS benchmarks.** *Hard-negative design* denotes explicit construction intended to reduce easy active–inactive separation or artificial enrichment. *Learning protocol* denotes a canonical training-based setup, and *fixed splits* denotes released benchmark-level train/validation/test splits. TopU-LBVS additionally provides a 93-target low-data protocol, a seven-target mini benchmark, and a paired random-decoy control.

Benchmark	Targets	Classes	Hard-negative design	Act:Dec ratio	Learning protocol	Fixed splits
DUD-E (Mysinger et al., 2012)	102	8	✗	1:50	✗	✗
MUV (Rohrer & Baumann, 2009)	17	4	✓	1:500	✗	✗
DEKOIS 2.0 (Bauer et al., 2013)	81	5	✗	1:30	✗	✗
LIT-PCBA (Tran-Nguyen et al., 2020)	15	3	†	varied	✗	✗
FS-Mol (Stanley et al., 2021)	varied	varied	✗	∼balanced	✓	✓
TopU-LBVS (ours)	**93**	**7**	✓	**1:40**	✓	✓

† LIT-PCBA uses experimentally confirmed inactives but does not explicitly construct structurally challenging negatives around known actives. MUV uses a distinct spatial-statistical design to reduce analogue bias and artificial enrichment rather than TopU-style physicochemical and structural matching.

bias and artificial enrichment, yielding challenging screening sets at a 1:500 ratio. However, MUV has only 17 targets and was not designed as a fixed supervised-learning benchmark with historical-SAR training, low-data learning, and standardized train/validation/test protocols. LIT-PCBA (Tran-Nguyen et al., 2020) similarly uses experimentally confirmed PubChem inactives, reducing false-negative concerns, but covers only 15 targets and lacks a canonical training-based protocol.

Learning-based evaluation beyond classical benchmarks. Riniker and Landrum (Riniker & Landrum, 2013) provide fixed compound lists and evaluation scripts across 88 targets, but study fingerprint retrieval rather than supervised target-specific learning and use random inactives. FS-Mol (Stanley et al., 2021) provides standardized few-shot ChEMBL tasks and fixed splits, but uses approximately balanced ratios and random inactives rather than hard-negative screening libraries. CACHE (Ackloo et al., 2022) benchmarks prospective computational hit-finding with experimental validation, but targets a structure-based prospective setting rather than retrospective LBVS. More broadly, ChEMBL-based studies (Yang et al., 2019; Ross et al., 2022) vary in target selection, activity thresholds, negative sampling, and splitting rules, complicating cross-paper comparison.

Recent work also underscores the continued strength of classical molecular representations. Praski et al. (Praski et al., 2025) compare 25 pretrained molecular embedding models across 25 datasets and find that nearly all provide negligible or no improvement over an ECFP baseline. Boulougouri et al. (Boulougouri et al., 2024) similarly show that simple set representations using atom invariants can match or outperform established graph models on widely used molecular benchmarks. Van Tilborg et al. (van Tilborg et al., 2026) further report an ECFP-based predictor matching or exceeding their learned molecular model in the in-distribution setting. Together, these findings motivate treating fingerprint methods as strong reference baselines rather than weak classical controls.

Benchmark gap and TopU-LBVS. The resources above address important aspects of LBVS evaluation, but do not jointly provide a large multi-class target panel, explicit hard-negative libraries, screening-relevant imbalance, fixed supervised-learning protocols, and low-cost evaluation. MUV is the closest classical precursor in avoiding easy screening sets; TopU-LBVS extends this philosophy to a larger, learning-oriented benchmark for contemporary molecular ML. This is particularly important as increasingly complex molecular models are evaluated on benchmarks where strong classical representations may already capture much of the available signal.

TopU-LBVS covers 93 targets across 7 protein classes at a fixed 1:40 ratio, with released data, splits, and evaluation code. **TopU-LBVS-full** evaluates generalization from historical ChEMBL* SAR data to separate hard TopU libraries. **TopU-LBVS-low** evaluates learning from limited labeled data within the TopU hard-negative distribution. **TopU-LBVS-mini** provides a seven-target entry point and paired random-decoy control that holds training, validation, test actives, ratio, and metric fixed while replacing only TopU test decoys with random ChEMBL* decoys. Together, these protocols test whether LBVS methods remain effective under focused hard-negative screening.

3 The TopU–LBVS Benchmark

3.1 Target Panel and Protein Classes

TopU-LBVS contains **93 protein targets** from ChEMBL 35 spanning **7 protein classes**: cytochrome P450 enzymes, GPCRs, ion channels, kinases, nuclear receptors, proteases, and other enzymes. Targets were selected for biological diversity, ligand-discovery relevance, and sufficient assay coverage.

The same panel supports the 93-target full and low-data protocols and the seven-target mini benchmark. In TopU-LBVS-full, when too few target-specific inactives remain after removing TopU compounds, we augment *training negatives only* with compounds from the same protein class. Compounds with documented activity against the recipient target are excluded, and augmentation never affects TopU test libraries. Full target counts and the augmentation audit are given in Appendix A.5.

3.2 ChEMBL* Curation

We curate ChEMBL 35 into a cleaned source dataset, denoted ChEMBL*. For each target, we retain reliable target and assay records, standardize molecular structures, canonicalize compounds, resolve duplicates, binarize activity, restrict molecules to common organic elements, and apply drug-like physicochemical filters.

Because hard-negative construction can amplify label uncertainty, we audit the inactive pool before decoy selection. Compounds with conflicting active/inactive annotations are excluded from the decoy pool, and no final TopU decoy is recorded as active for its own target under the benchmark criterion. Detailed curation, label-consistency statistics, and leakage audits are provided in Appendix A.1.

ChEMBL* supplies the historical SAR pool for TopU-LBVS-full after all target-specific TopU compounds are removed. It also supplies the random-decoy control in TopU-LBVS-mini, where TopU test decoys are replaced by randomly sampled ChEMBL* decoys while preserving test actives, the active-to-decoy ratio, and the training/validation protocol.

3.3 TopU Construction and Statistics

For each target, we construct a TopU library by pairing curated actives with hard decoys from the corresponding ChEMBL* inactive pool. Candidate decoy sets are optimized to reduce active–decoy distinguishability: a dynamic search minimizes mutual information between activity labels and six physicochemical descriptors plus MFP3 features while enforcing library size and class ratio. A Random Forest is then used only as a *post-construction hardness audit*; easily separable candidate libraries are rejected. Thus, the RF criterion works against simple Morgan/RF separability rather than favoring the downstream Morgan-RF baseline, which is trained independently from scratch on disjoint ChEMBL* data.

Each TopU library uses a fixed **1:40** active-to-decoy ratio. We treat this as a focused hard-negative screening operating point rather than an estimate of raw HTS prevalence. It is comparable to established benchmarks such as DUD-E ($\sim$1:50) and DEKOIS 2.0 (1:30) (Mysinger et al., 2012; Bauer et al., 2013), while allowing stringent hard-negative constraints to be applied consistently across all 93 targets. Increasing the nominal ratio substantially would often require adding easier background negatives, changing the benchmark question.

Library sizes range from several hundred to more than ten thousand compounds. The same TopU libraries serve as test sets in TopU-LBVS-full, train/validation/test pools in TopU-LBVS-low, and the seven-target subset in TopU-LBVS-mini. Detailed construction parameters, per-target statistics, and post-construction robustness checks using alternative fingerprints and molecular descriptors are provided in Appendix A.2.

4 Tasks and Evaluation Settings

We formulate LBVS as target-specific ranking. For each target t and candidate compound c, a model outputs a score $s_t(c)$, which induces a ranking over the candidate library. Metrics are computed per target and then aggregated across targets.

TopU-LBVS contains 93 target-specific tasks across 7 protein classes. Each target has a TopU library constructed at a fixed 1:40 active-to-decoy ratio using property-matched, structurally challenging decoys (Section 3); library sizes range from roughly 400 to more than 10,000 compounds. We define three protocols: **TopU-LBVS-full** for historical ChEMBL* $\rightarrow$ TopU generalization, **TopU-LBVS-low** for low-data learning within TopU hard-negative libraries, and **TopU-LBVS-mini** for low-cost evaluation with a paired random-decoy control. Figure 1 illustrates the benchmark pipeline, and Table 2 summarizes the protocol details.

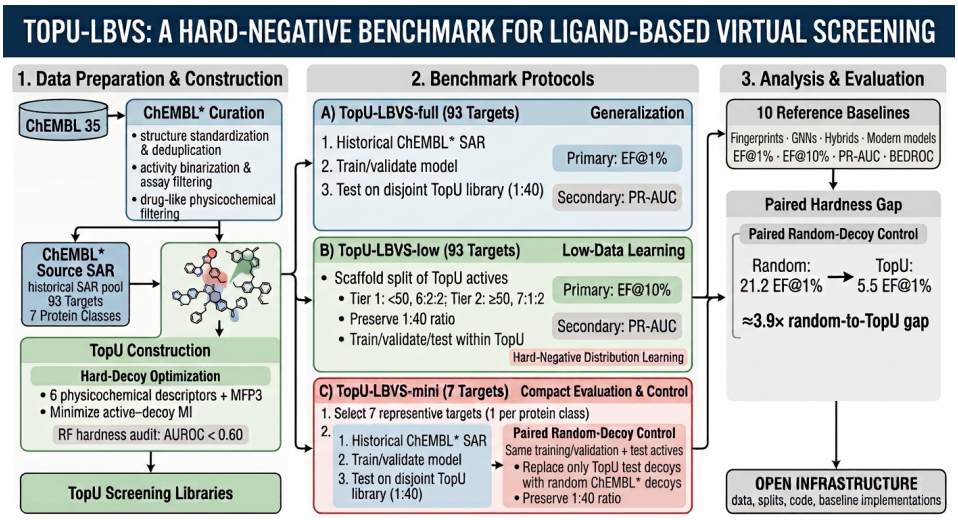

Figure 1: **TopU-LBVS framework.** ChEMBL 35 is curated into ChEMBL*, which supplies historical SAR data and the inactive pool for constructing TopU hard-negative libraries. TopU-LBVS-full evaluates generalization from historical ChEMBL* SAR to disjoint TopU screening libraries; TopU-LBVS-low evaluates learning from limited labeled data within the TopU hard-negative distribution; and TopU-LBVS-mini provides a seven-target protocol with a paired random-decoy control that changes only the test decoys.

4.1 TopU-LBVS-full: ChEMBL* → TopU Evaluation

TopU-LBVS-full tests whether models trained on historical SAR data generalize to hard-negative screening libraries. For each target, all TopU compounds are removed from ChEMBL* to prevent train-test overlap. The remaining compounds form a fixed 1:10 active-to-inactive training set. If target-specific inactives are insufficient, we add same-class inactives using seed 2026; documented recipient-target actives are excluded, and selections are released.

We reserve 15% of this pool as a stratified random validation set, preserving the 1:10 ratio, and use the remaining 85% for training. The test set is the full TopU library for the target, with a 1:40 active-to-decoy ratio. The primary metric is EF@1%; PR-AUC is used for validation and reported as a secondary metric.

4.2 TopU-LBVS-low: Low-Data TopU → TopU Evaluation

TopU-LBVS-low evaluates learning from limited labeled data when train, validation, and test compounds all come from the same hard TopU distribution. It is not intended to model a target-naive discovery campaign; rather, it tests whether models can learn from limited labels within a focused hard-negative candidate distribution. For each target, splits are drawn from its TopU library while preserving the 1:40 active-to-decoy ratio.

Because active counts vary widely, targets with fewer than 50 TopU actives (Tier 1) use a 6:2:2 active split, while targets with at least 50 actives (Tier 2) use a 7:1:2 split. Active compounds are split by Bemis–Murcko scaffold, while inactives are assigned to preserve the 1:40 ratio in each split. Splits use the predefined seeds 2026, 2027, and 2028.

The primary metric is EF@10%, which retains a direct screening-budget interpretation while avoiding the instability of EF@1% and EF@5% for the smallest low-data test libraries. PR-AUC is used for validation and reported as a secondary metric; BEDROC provides a complementary continuous early-recognition measure and is reported in the appendix. Full split rules and per-target active counts are given in Appendix E.

4.3 TopU-LBVS-mini: Compact Evaluation and Random-Decoy Control

TopU-LBVS-mini is a seven-target subset for rapid development, debugging, and low-cost comparison. For each protein class, we select the target with the most TopU actives, yielding one representative target per class. Selected targets and counts are listed in Appendix E.

Table 2: **Overview of TopU-LBVS evaluation protocols.** TopU-LBVS-full and TopU-LBVS-low contain 93 target-specific tasks. TopU-LBVS-mini follows the full protocol on seven representative targets, one from each protein class. Ratios and predefined splits are shared across all baselines.

Parameter	**TopU-LBVS-full** ChEMBL* → TopU	**TopU-LBVS-low, Tier 1** TopU → TopU (< 50 actives)	**TopU-LBVS-low, Tier 2** TopU → TopU (≥ 50 actives)	**TopU-LBVS-mini** ChEMBL* → TopU
Training source	ChEMBL*	TopU	TopU	ChEMBL*
Test source	TopU	TopU	TopU	TopU
Number of targets	93	46	47	7
Train/val/test split	85% / 15% / TopU test	6:2:2	7:1:2	85% / 15% / paired test
Train ratio (act:dec)	1:10	1:40	1:40	1:10
Validation ratio (act:dec)	1:10	1:40	1:40	1:10
Test ratio (act:dec)	1:40	1:40	1:40	1:40
Split seeds	{2026, 2027, 2028}	{2026, 2027, 2028}	{2026, 2027, 2028}	{2026, 2027, 2028}
Primary test metric	EF@1%	EF@10%	EF@10%	EF@1%
Secondary test metric	PR-AUC	PR-AUC	PR-AUC	PR-AUC
Validation metric	PR-AUC (AP)	PR-AUC (AP)	PR-AUC (AP)	PR-AUC (AP)

In the random-control variant, TopU-LBVS-mini keeps the same training and validation protocol but replaces TopU test decoys with randomly sampled ChEMBL* decoys, preserving test active count and 1:40 ratio.

The TopU evaluation follows the same protocol as TopU-LBVS-full: models are trained on ChEMBL* after removing all target-specific TopU compounds, validated on the same fixed 15% stratified split, and tested on the full TopU library. EF@1% is the primary metric, and PR-AUC is used for validation.

TopU-LBVS-mini also includes a paired random-decoy control. For each mini target, we keep the training data, validation protocol, test actives, and 1:40 test ratio fixed, but replace the TopU test decoys with randomly sampled ChEMBL* decoys. The resulting gap isolates the effect of test-decoy difficulty and quantifies how much random-decoy evaluation can overestimate performance relative to hard-negative screening. Final benchmark claims should be reported on TopU-LBVS-full; TopU-LBVS-mini is intended as an accessible development and hardness-analysis protocol.

Across all protocols, enrichment factors are the primary test metrics: EF@1% for TopU-LBVS-full and TopU-LBVS-mini, and EF@10% for TopU-LBVS-low. PR-AUC is used for validation and reported as a secondary metric. BEDROC, ROC-AUC, and additional enrichment factors are defined and reported in Appendix D.

5 BASELINES

We evaluate ten baselines spanning three families: (i) classical fingerprint methods, (ii) molecular GNNs and fingerprint hybrids, and (iii) modern molecular models. Each produces a target-specific ranking score. Models are selected by validation PR-AUC, and test performance uses EF@1% for TopU-LBVS-full and TopU-LBVS-mini and EF@10% for TopU-LBVS-low. Architectures, hyperparameter searches, optimization, and training details are provided in Appendix B.

Classical fingerprint baselines. We include Morgan-RF and Tanimoto-NN, two standard LBVS references based on 2048-bit ECFP4-style Morgan fingerprints computed with RDKit (Rogers & Hahn, 2010; Landrum, 2006). Morgan-RF trains a Random Forest classifier, while Tanimoto-NN ranks compounds by their maximum Tanimoto similarity to training actives (Willett, 2006). These methods provide strong, interpretable reference points for practical ligand-based screening.

GNN and hybrid baselines. We evaluate three atom-bond molecular graph architectures: GIN (Xu et al., 2019), GAT (Veličković et al., 2018), and GPS (Rampášek et al., 2022). Each is paired with a fingerprint-augmented variant, GIN+FP, GAT+FP, and GPS+FP, which combines the learned graph embedding with a learned projection of the Morgan fingerprint before classification. These hybrids test whether learned molecular representations add signal beyond classical substructure fingerprints under TopU hard-negative evaluation.

Modern molecular baselines. We include D-MPNN/Chemprop (Yang et al., 2019; Heid et al., 2023), a widely used directed message-passing model for molecular property prediction, and MolFormer (Ross et al., 2022), a pretrained SMILES Transformer. For MolFormer, we fine-tune the publicly released 10separately for each target.

All learned models are trained on the corresponding TopU-LBVS splits using class-weighted binary cross-entropy and model-specific optimization, with selection by validation PR-AUC. Hyperparameters are chosen on the training/validation splits, and results are averaged over the three predefined runs corresponding to seeds 2026, 2027, and 2028. on fixed data splits. A separate model is trained for each target and protocol; implementation details are given in Appendix B.

6 EXPERIMENTAL RESULTS

We evaluate ten baselines across the three TopU-LBVS protocols. Metrics are computed per target from the induced ranking and then averaged by protein class and overall. We report the primary metric for each setting in the main text: EF@1% for TopU-LBVS-full and TopU-LBVS-mini, and EF@10% for TopU-LBVS-low.

Full per-target results and additional metrics for TopU-LBVS-full and TopU-LBVS-low, including EF@5%, PR-AUC, ROC-AUC, BEDROC, and LogAUC, are reported in the Appendix. Table 13 provides a roadmap of all appendix tables.

6.1 TOPU-LBVS-FULL

Table 3 reports class-averaged EF@1% under the ChEMBL*→TopU protocol across all 93 targets.

Table 3: **TopU-LBVS-full: class-averaged EF@1%.** Mean EF@1% under the ChEMBL*→TopU protocol, averaged over targets within each protein class. The **Overall** column is the mean over all 93 targets. Higher is better.

Group	Model	Cytochrome	GPCR	Ion Channel	Kinase	Nuc. Receptor	Misc.	Protease	Overall
Classical	Morgan-RF	14.18	5.21	12.41	9.15	11.12	10.01	6.66	9.18
	Tanimoto-NN	9.84	5.06	9.15	8.84	9.13	7.50	4.40	7.58
GNN & Hybrid	GIN	6.96	3.15	3.87	5.62	5.63	3.92	2.58	4.47
	GIN+FP	14.70	4.88	7.00	7.38	6.29	6.42	4.53	6.64
	GAT	6.03	2.82	3.68	5.13	3.98	3.42	1.10	3.76
	GAT+FP	13.87	4.78	6.41	7.89	7.33	6.05	3.04	6.57
	GPS	5.22	2.02	3.64	4.58	4.23	3.11	0.69	3.36
	GPS+FP	14.95	4.84	7.06	7.84	7.10	5.75	3.64	6.60
Modern	D-MPNN	13.54	4.81	5.04	8.62	8.80	7.52	4.34	7.38
	MolFormer	11.19	4.19	7.51	7.84	7.87	6.08	4.48	6.65

Classical fingerprint methods are strongest in the full protocol. Morgan-RF achieves the best overall EF@1% (9.18), outperforming all GNN, fingerprint-augmented hybrid, and modern learned baselines across the 93-target panel. Tanimoto-NN (7.58) also exceeds all pure GNN architectures. The gap is substantial: the best pure GNN, GIN, reaches 4.47, while the strongest modern learned baseline, D-MPNN, reaches 7.38. Thus, strong performance on easier negative distributions does not necessarily transfer to the TopU hard-negative regime. Although TopU construction reduces trivial property and nearest-neighbour shortcuts, it does not eliminate all substructure-level signal, and Morgan-RF remains a strong reference baseline. This is consistent with recent studies showing that ECFP-based or similarly simple molecular representations can match or outperform substantially more complex learned models across diverse molecular prediction benchmarks (Praski et al., 2025; Boulougouri et al., 2024; van Tilborg et al., 2026). Our results extend this observation to target-specific LBVS under focused hard-negative screening, where the evaluated learned representations do not consistently improve upon fixed fingerprints.

Fingerprint augmentation recovers part of the GNN deficit. Adding Morgan fingerprints to learned graph embeddings consistently improves GNN performance: GIN+FP, GAT+FP, and GPS+FP improve over their pure counterparts by 2.17, 2.81, and 3.24 EF@1% points, respectively. The three hybrids cluster tightly around 6.6 overall, indicating that fingerprint information contributes substantially across GNN backbones in this setting. However, the hybrids remain about 2.5 EF@1% points below Morgan-RF, suggesting that the learned graph representations do not yet add reliable signal beyond the fingerprint features under hard-negative evaluation. Similarly, D-MPNN (7.38) and MolFormer (6.65), despite greater model complexity and, for MolFormer, large-scale pretraining, do not surpass the strongest fingerprint baseline. These results establish TopU-LBVS-full as a stringent benchmark: under the evaluated training protocols, the learned molecular models do not provide a consistent advantage over well-tuned fixed fingerprints.

6.2 TOPU-LBVS-LOW

Table 4 reports class-averaged EF@10% under the low-data TopU→TopU protocol. EF@10% is the primary metric for this setting because EF@1% and EF@5% are unstable for the smallest Tier 1 libraries, where only a few actives appear within the top-ranked budget.

Table 4: **TopU-LBVS-low: class-averaged EF@10%.** Mean EF@10% under the TopU→TopU few-shot protocol, averaged over targets within each protein class. Tier 1 contains targets with fewer than 50 TopU actives; Tier 2 contains targets with at least 50 TopU actives. Higher is better.

Group	Model	Cytochrome	GPCR	Ion Channel	Kinase	Nuc. Receptor	Misc.	Protease	Overall
Classical	Morgan-RF	1.24	1.02	0.58	0.79	0.57	0.30	0.35	0.63
	Tanimoto-NN	1.13	0.53	0.58	0.67	0.51	0.38	0.05	0.50
GNN & Hybrid	GIN	1.15	1.46	0.75	0.96	0.98	0.90	1.09	1.03
	GIN+FP	0.91	0.85	0.63	1.00	0.83	0.86	0.42	0.84
	GAT	1.14	0.94	0.88	0.95	1.27	0.96	0.74	0.97
	GAT+FP	0.96	0.75	0.56	0.70	0.60	0.89	0.73	0.75
	GPS	1.39	0.95	1.36	0.99	1.18	1.11	0.76	1.05
	GPS+FP	0.66	0.72	1.00	0.84	1.01	0.90	0.74	0.85
Modern	D-MPNN	1.64	1.13	1.21	1.06	1.15	0.52	0.96	0.96
	MolFormer	1.72	1.04	1.06	1.09	1.24	1.22	0.75	1.12
Tier 1 (<50 actives, 46 targets)		–	0.89	0.81	0.76	1.04	0.81	0.65	0.80
Tier 2 (≥50 actives, 47 targets)		1.19	0.95	1.07	0.98	0.84	0.78	0.70	0.94

Low-data hard-negative screening changes the model ranking. The fingerprint baselines that dominate TopU-LBVS-full perform poorly in this regime: Morgan-RF reaches 0.63 overall EF@10%, and Tanimoto-NN reaches 0.50, both below random enrichment on most classes. With only a limited number of labeled TopU actives for training, nearest-neighbour fingerprint retrieval loses much of its advantage because the candidate library contains structurally challenging decoys. Learned models become more competitive: MolFormer obtains the best overall EF@10% (1.12), followed by GPS (1.05) and GIN (1.03). However, these gains remain modest. Since random ranking corresponds to EF=1, the low-data hard-negative regime remains largely unsolved, especially for Tier 1 targets, whose overall EF@10% is only 0.80 across baselines.

Fingerprint augmentation becomes protocol-dependent. Unlike in TopU-LBVS-full, adding fingerprints does not consistently help in the low-data setting. GIN+FP underperforms GIN (0.84 vs. 1.03), GAT+FP underperforms GAT (0.75 vs. 0.97), and GPS+FP underperforms GPS (0.85 vs. 1.05). This suggests that fingerprint similarity provides less reliable additional signal when learning from a small labeled subset of an already hard-negative candidate distribution. The contrast with TopU-LBVS-full shows that robustness across data-rich and low-data hard-negative regimes is non-trivial. The Tier 1/Tier 2 split further highlights the data bottleneck: increasing from fewer than 50 to at least 50 TopU actives improves overall EF@10% from 0.80 to 0.94, but even Tier 2 remains close to random enrichment.

To provide an accessible entry point for low-data molecular and graph learning, we also introduce **TopU-LBVS-low-mini**, a compact 13-target subset covering one Tier 1 and one Tier 2 target per protein class where available; full details are given in Appendix A.4.

6.3 TopU-LBVS-mini: Random-Decoy Control

TopU-LBVS-mini is a compact subset of TopU-LBVS-full, containing one representative target from each protein class. It is designed as a *low-cost molecular benchmark for the broader machine learning community* and also enables a paired random-decoy control. Table 5 compares TopU hard-decoy (T) test sets with randomly sampled ChEMBL* decoy (R) test sets across the seven mini targets. The two settings use the same training data, validation protocol, test actives, active count, and active-to-decoy ratio; only the test decoys differ. Thus, the R−T gap isolates the effect of test-decoy difficulty.

Random-decoy evaluation substantially overestimates performance. Across all ten models and seven targets, mean EF@1% drops from 21.2 on random decoys to 5.5 on TopU decoys, a nearly fourfold reduction caused solely by changing the test negatives. The gap is highly consistent: random-decoy EF@1% exceeds TopU EF@1% in 67 of 70 paired model-target comparisons and appears across all model families. This controlled comparison shows that strong performance against random negatives can largely reflect separation from chemically dissimilar background compounds rather than robust enrichment under hard-negative screening.

The hardness gap is pervasive but target-dependent. The largest average R−T gaps occur on egfr (21.6) and thrb (21.4), whereas cp3a4 shows almost no average gap between the two decoy regimes. On cp3a4, several models achieve similar performance under TopU and random decoys: GAT+FP changes from 18.2 to 14.3, while GPS+FP achieves the best TopU EF@1% (16.4) with a

Table 5: **TopU-LBVS-mini: TopU vs. random decoy test sets (EF@1%).** Each target shows TopU hard decoys (T) and randomly sampled ChEMBL* decoys (R). The gap between R and T measures how much random-decoy evaluation overestimates performance. The last two columns are averages across all seven targets. Higher is better.

	aa2ar		aces		cp3a4		egfr		esr1		kcnh2		thrb		Average	
Model	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R
Morgan-RF	6.2	27.2	10.0	28.8	15.9	16.2	6.4	32.1	6.2	24.4	5.4	19.2	10.1	30.6	**8.6**	**25.5**
Tanimoto-NN	3.6	18.6	8.0	18.7	10.9	8.5	5.6	17.4	8.0	17.5	4.8	12.0	8.3	22.6	**7.0**	**16.5**
GIN	2.8	18.3	2.6	16.2	10.3	9.8	2.8	22.6	2.1	18.1	1.5	10.2	2.2	19.1	**3.5**	**16.3**
GIN+FP	6.2	29.9	5.1	30.2	15.6	16.2	3.9	32.8	3.3	25.0	2.4	20.7	3.5	31.5	**5.7**	**26.6**
GAT	4.1	13.4	2.8	16.5	9.0	9.4	2.0	16.9	2.4	15.4	2.0	10.2	2.2	18.1	**3.5**	**14.3**
GAT+FP	5.2	29.7	4.8	32.2	14.3	18.2	2.0	31.7	4.2	23.5	2.5	20.5	3.5	31.8	**5.2**	**26.8**
GPS	1.0	9.3	3.4	13.1	5.8	6.6	2.6	15.8	2.4	10.1	2.5	9.4	2.5	15.9	**2.9**	**11.5**
GPS+FP	5.7	30.2	6.8	30.5	16.4	16.9	3.9	30.8	3.3	22.3	1.9	21.1	1.9	31.8	**5.7**	**26.2**
D-MPNN	3.1	28.1	6.5	30.5	14.9	15.8	7.4	28.0	5.9	24.7	9.1	20.1	4.5	29.6	**7.3**	**25.3**
MolFormer	2.8	25.0	3.4	27.3	14.5	13.5	3.5	28.4	2.7	19.6	3.8	19.1	5.7	27.7	**5.2**	**22.9**
Avg.	**4.1**	**23.0**	**5.3**	**24.4**	**12.8**	**13.1**	**4.0**	**25.6**	**4.0**	**20.1**	**3.6**	**16.2**	**4.4**	**25.9**	**5.5**	**21.2**

nearly identical random-decoy result (16.9). Thus, hard-negative difficulty varies substantially across targets even under a fixed construction protocol. Overall, no baseline achieves consistently strong enrichment across all seven TopU mini targets, making TopU-LBVS-mini a low-cost diagnostic for the gap between apparent random-decoy performance and hard-negative screening performance.

6.4 Discussion, Limitations, and Release

Taken together, the three protocols reveal a consistent picture: the evaluated learned molecular models do not reliably outperform strong fingerprint baselines under hard-negative screening, while the nearly 4× gap exposed by the paired mini control shows how random-decoy evaluation can substantially overestimate enrichment. In TopU-LBVS-low, learned models become more competitive, but performance remains close to random enrichment on the most data-limited tasks. TopU-LBVS is designed to make these regime-dependent failures visible and measurable.

TopU-LBVS also has important limitations. It is retrospective and ligand-based: no 3D binding-site information, docking scores, or prospective experimental validation are included, and ChEMBL-derived inactive labels may still contain assay noise or unobserved activity despite our consistency and leakage audits. TopU operationalizes hard-negative similarity using physicochemical descriptors and MFP3 features; alternative construction spaces could yield different libraries, although post-construction robustness checks using other molecular representations are reported in the Appendix. The 93 targets are not class-balanced and favor families with richer public assay coverage. Finally, Tier 1 TopU-LBVS-low tasks contain very limited labeled data, so individual-target enrichment estimates can have high variance and class-level or overall results are more reliable.

All data, fixed splits, evaluation code, and baseline implementations are released with the benchmark, including the ChEMBL* curation pipeline, TopU construction scripts, train/validation/test splits, and implementations of all ten baselines. A common evaluation script accepts ranked compound lists and returns all reported metrics. Future benchmark versions will be separately versioned so that results against the current release remain directly comparable.

7 Conclusion

We introduced TopU-LBVS, a multi-target benchmark for ligand-based virtual screening under hard-negative conditions. Built from curated ChEMBL 35 data, TopU-LBVS provides 93 target-specific tasks, fixed 1:40 active-to-decoy libraries, and three protocols: ChEMBL* → TopU evaluation, low-data TopU → TopU evaluation, and a compact mini benchmark with a paired random-decoy control. Across ten reference baselines, the benchmark shows that performance under random-decoy evaluation can degrade sharply when the same models are tested against TopU hard negatives; in the mini control, mean EF@1% is nearly 4× lower with TopU decoys than with random ChEMBL* decoys. The evaluated learned molecular models also do not consistently outperform strong fingerprint baselines under these conditions. By releasing curated data, fixed splits, evaluation code, and reference baselines, TopU-LBVS provides a reproducible testbed for assessing future LBVS and molecular representation learning methods under focused hard-negative screening conditions.

AI USE STATEMENT

Generative AI tools were used to improve the clarity and English of the manuscript, assist with drafting and revising portions of the text, and help design and refine the pipeline figure. All AI-assisted text and visual content were reviewed, edited, and approved by the authors. Generative AI was not used to generate experimental data, conduct experiments, or produce reported numerical results. The authors take full responsibility for the final manuscript and its contents.

ETHICS STATEMENT

This work is entirely computational and uses publicly available molecular bioactivity data from ChEMBL; it involves no human participants, personally identifiable information, animal experiments, or new wet-lab experiments. TopU-LBVS is intended as a research benchmark for evaluating ligand-based virtual screening methods and should not be interpreted as providing experimentally validated binding predictions or clinical or therapeutic guidance. As with other ChEMBL-derived resources, activity labels may contain assay noise or incomplete annotations; we document these limitations and apply consistency and leakage checks during curation. The released data and code are intended to support reproducible methodological research under hard-negative screening conditions.

REPRODUCIBILITY STATEMENT

We release the curated benchmark data, fixed train/validation/test splits, target metadata, TopU construction code, baseline implementations, and evaluation scripts. All experiments use the released splits and specified evaluation protocols, with model selection based on validation PR-AUC. Results are reported over multiple training seeds on fixed data splits, and the Appendix provides implementation details, hyperparameters, per-target results, and additional metrics. The evaluation code accepts ranked compound lists and computes all reported metrics, enabling direct comparison with future methods. Our code and data are available at
Code: https://anonymous.4open.science/r/topu-lbvs-2D5A/
Data: https://huggingface.co/datasets/topu-benchmark/topu-lbvs

REFERENCES

Suzanne Ackloo et al. CACHE (Critical Assessment of Computational Hit-finding Experiments): a public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding. *Nature Reviews Chemistry*, 6(4):287–295, 2022.

Matthias R Bauer, Tamer M Ibrahim, Simon M Vogel, and Frank M Boeckler. DEKOIS 2.0 – a library of challenging data sets for structure-based virtual screening and target identification. *Journal of Chemical Information and Modeling*, 53(6):1447–1462, 2013.

Maria Boulougouri, Pierre Vandergheynst, and Daniel Probst. Molecular set representation learning. *Nature Machine Intelligence*, 6(7):754–763, 2024.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Li Chen, Anthony Cruz, Steven Ramsey, Callum J Dickson, David S Goodsell, Viktor Hornak, David R Koes, and Soumya Dutta. Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PLOS ONE*, 14(12):e0220113, 2019.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Esther Heid, Kevin P Greenman, Yunsie Chung, Shih-Cheng Li, David E Graff, Florence H Vermeire, Haoyang Wu, William H Green, and Charles J McGill. Chemprop: a machine learning package for chemical property prediction. *Journal of chemical information and modeling*, 64(1):9–17, 2023.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations (ICLR)*, 2020.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Christian Kramer et al. The need for continuing blinded pose- and activity prediction benchmarks. *Journal of Chemical Information and Modeling*, 65(5):2180–2190, 2025.

Nathalie Lagarde, Jean-François Zagury, and Matthieu Montes. Benchmarking data sets for the evaluation of virtual ligand screening methods: Review and perspectives. *Journal of Chemical Information and Modeling*, 55(7):1297–1307, 2015. doi: 10.1021/acs.jcim.5b00090.

Greg Landrum. RDKit: Open-source cheminformatics, 2006. URL https://www.rdkit.org.

Michael M Mysinger, Michael Carchia, John J Irwin, and Brian K Shoichet. Directory of useful decoys, enhanced (dud-e): better ligands and decoys for better benchmarking. *Journal of Medicinal Chemistry*, 55(14):6582–6594, 2012.

Adam Paszke et al. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 2019.

Mateusz Praski, Jakub Adamczyk, and Wojciech Czech. Benchmarking pretrained molecular embedding models for molecular representation learning. *arXiv preprint arXiv:2508.06199*, 2025.

Ladislav Rampášek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Lim, Guillaume Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Sereina Riniker and Gregory A Landrum. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *Journal of Cheminformatics*, 5(1):26, 2013.

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

Sebastian G Rohrer and Knut Baumann. Maximum unbiased validation (MUV) data sets for virtual screening based on PubChem bioactivity data. *Journal of Chemical Information and Modeling*, 49 (2):169–184, 2009.

Jerret Ross, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. Large-scale chemical language representations capture molecular structure and properties. *Nature Machine Intelligence*, 4(12):1256–1264, 2022.

Jochen Sieg, Florian Flachsenberg, and Matthias Rarey. In need of bias control: Evaluating chemical data for machine learning in structure-based virtual screening. *Journal of Chemical Information and Modeling*, 59(3):947–961, 2019. doi: 10.1021/acs.jcim.8b00712.

Megan Stanley, John F Bronskill, Krzysztof Maziarz, Hubert Misztela, Jessica Lanini, Marwin Segler, Nadine Schneider, and Marc Brockschmidt. FS-Mol: A few-shot learning dataset of molecules. In *NeurIPS Datasets and Benchmarks Track*, 2021.

Jonathan M Stokes, Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, et al. A deep learning approach to antibiotic discovery. *Cell*, 180(4):688–702, 2020.

Viet-Khoa Tran-Nguyen, Clémence Jacquemard, and Didier Rognan. LIT-PCBA: an unbiased data set for machine learning and virtual screening. *Journal of Chemical Information and Modeling*, 60 (9):4263–4273, 2020.

Derek van Tilborg, Luke Rossen, and Francesca Grisoni. Molecular deep learning at the edge of chemical space. *Nature Machine Intelligence*, pp. 1–13, 2026.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

W. Patrick Walters and Renxiao Wang. New trends in virtual screening. *Journal of Chemical Information and Modeling*, 60(9):4109–4111, 2020. doi: 10.1021/acs.jcim.0c01009.

Peter Willett. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, 11 (23-24):1046–1053, 2006.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *ICLR*, 2019.

Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, et al. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019.

APPENDIX

A TopU-LBVS Benchmark Details **14**

A.1 ChEMBL* Curation . . . 14

A.2 TopU Construction and Statistics . . . 14

A.3 Label, Leakage, and Representation Robustness Audits . . . 15

A.4 TopU-LBVS-low-mini: A Compact Low-Data Benchmark . . . 16

A.5 Target Panel and Protein Classes . . . 17

B Baseline Models **17**

B.1 Category I: Classical Similarity-Based Baselines . . . 18

B.2 Category II: Graph Neural Networks and Hybrid Models . . . 20

B.3 Category III: Modern Pretrained Models . . . 21

C Hyperparameters and Implementation Details **22**

C.1 Classical Baselines . . . 22

C.2 GNN and Hybrid Models . . . 22

C.3 Modern Pretrained Models . . . 23

C.4 Computational Resources . . . 23

D Metric Definitions **24**

D.1 Enrichment Factor at x percent (EF@x%) . . . 24

D.2 Precision Recall AUC (PR AUC) . . . 24

D.3 Receiver Operating Characteristic AUC (ROC AUC) . . . 25

D.4 BEDROC . . . 25

D.5 Aggregation Across Targets and Seeds . . . 25

E Additional Details for Tasks and Evaluation Settings **25**

E.1 Per Target LBVS Task . . . 26

E.2 Setting 1: ChEMBL* to TopU (Main Benchmark) . . . 26

E.3 Setting 2: TopU Only (Few Shot) . . . 26

E.4 Setting 3: ChEMBL* Only (Ablation) . . . 27

F Limitations and Scope **28**

G Broader impacts **28**

H Additional Results **28**

H.1 Part I: Primary-Metric Per-Target Results . . . 29

H.2 Part II: Full Per-Model Metrics, TopU-LBVS-full . . . 29

H.3 Part III: Full Per-Model Metrics, TopU-LBVS-few . . . 32

A TopU-LBVS Benchmark Details

A.1 ChEMBL* Curation

We construct ChEMBL* from ChEMBL 35 to serve as the foundation for the TopU-LBVS. The goal of this curation pipeline is to remove noisy or assay artifacts, standardize molecular representations, and generate a high-confidence, consistent dataset of actives and inactives for subsequent benchmark construction.

Target-specific record collection. For each target of interest, we retrieve bioactivity data from ChEMBL 35 with reliable target assignments and assay annotations. We retain only measurements corresponding to four canonical potency endpoints (IC_{50}, K_i, K_d, EC_{50}) and exclude records with ambiguous or inconsistent target assignments. Where multiple assay types exist for a single target, we harmonize them into a unified target-level activity table.

Molecular standardization. All compounds are standardized prior to filtering and deduplication. This includes canonicalization of SMILES, removal of salts and solvent fragments when possible, and normalization of charges according to the RDKit-based preprocessing pipeline used in our benchmark code. Molecules that fail sanitization or cannot be converted to a valid canonical representation are discarded.

Activity binarization. For each target, we convert continuous activity measurements into binary labels (active versus inactive). These thresholds are chosen to be consistent within target and to preserve a reasonable separation between active and inactive compounds. To establish a definitive supervision signal for downstream benchmark construction, the retained compounds are strictly binarized based on their bioactivity: molecules with a pChEMBL value > 4.9 are designated as actives, while all remaining entries are categorized as inactives. The threshold of 4.9 is chosen to maximize target coverage while maintaining assay reliability.

Deduplication and conflict resolution. After standardization, molecules are deduplicated by canonical SMILES. If multiple measurements map to the same standardized molecule for a given target, they are merged into a single entry. To account for inherent experimental variability, cases with conflicting activity labels are resolved by designating the molecule as active. This step ensures that the final benchmark does not contain duplicate structures with inconsistent supervision.

Removal of problematic compounds. To prevent the distortion of screening benchmarks and eliminate confounding factors, we remove problematic entries—such as those with unrealistically small or large structures, or other obvious cheminformatics anomalies—by applying rigorous physicochemical filtering rules. Specifically, compounds are retained only if they meet the following criteria: (i) atomic composition restricted to C, N, O, F, S, Cl, Br, I, P and H atoms; (ii) molecular weight between 150 and 800 Da; (iii) lipophilicity (LogP) between -3.0 and 6.0; (iv) fewer than 20 rotatable bonds; (v) fewer than 10 hydrogen bond acceptors and donors; (vi) a total formal charge constrained between -2.0 and +2.0.

Resulting curated dataset. The output of this pipeline is the cleaned target-specific dataset we denote by *ChEMBL**. Throughout the paper, "ChEMBL*" refers exclusively to this curated subset of ChEMBL 35. ChEMBL* serves as the source pool for constructing TopU and for defining the historical SAR training data used in Settings 1 and 3.

Tables 7 and 8 report, for each target, the number of raw ChEMBL compounds, the number of compounds remaining after ChEMBL* curation, and the final TopU counts. The gap between raw and cleaned counts reflects the cumulative effect of assay filtering, removal of problematic compounds, and canonical SMILES deduplication.

A.2 TopU Construction and Statistics

For each target, we construct a TopU library by pairing a selected set of actives with hard decoys drawn from the corresponding ChEMBL* inactive pool. The main design goal of TopU is to remove

the trivial shortcuts that often make public LBVS benchmarks overly easy. In particular, we aim to ensure that decoys cannot be separated from actives by simple physicochemical filters, while also remaining structurally similar enough to known actives that nearest-neighbour fingerprint retrieval is no longer sufficient. We intentionally restrict baselines to 2D and sequence-based methods to isolate the effect of representation learning under hard-negative regimes and avoid confounding performance with target-specific structural availability.

To construct the TopU library for each target, we extract carefully curated actives and hard decoys from the corresponding ChEMBL* pool using a modified dynamic genetic algorithm. Unlike standard heuristic matching, this approach dynamically optimizes the size of the maximum matched subset while strictly enforcing a predefined active-to-decoy ratio. To ensure rigorous similarity in both physicochemical and topological spaces, molecules are encoded using a combination of six key physicochemical descriptors (Molecular Weight, LogP, Rotatable Bonds, Hydrogen Bond Acceptors, Hydrogen Bond Donors and Net Charge) and Morgan fingerprints 3 (MFP3). The algorithm's fitness function is designed to minimize the mutual information between these molecular features and bioactivity labels, thereby forcing the distributions of actives and decoys to become virtually indistinguishable. Furthermore, a Random Forest classifier is integrated as an additional validator, dynamically scaling the subset to guarantee that the cross-validation AUROC remains below 0.6. The Random Forest is configured with 100 estimators and a minimum of 10 samples per leaf, and is trained on the concatenated physicochemical and MFP3 features. The AUROC threshold of 0.6 is chosen to ensure that actives and decoys remain nearly indistinguishable while still allowing feasible library construction across diverse targets. The genetic algorithm is run with target-specific parameter tuning: starting active sizes (5–10), maximum generations (100–700), patience (8–18), and step sizes for active increments (5–10). These ranges are chosen to balance search granularity against convergence speed for each target. Targets with highly overlapping active–decoy distributions require smaller starting sizes and step sizes for finer chemical space exploration, while larger values accelerate convergence for easier targets. Higher patience and generation limits are used for targets where the active pool is large, ensuring sufficient iterations to search feasible subsets. Collectively, these results demonstrate that TopU decoys are not random background molecules; rather, they are purposefully curated hard negatives.

Each TopU library is constructed at a fixed **1:40 active-to-decoy ratio**. This ratio is large enough to create a realistic early-enrichment problem, while still keeping library sizes manageable for systematic benchmarking across many targets. By construction, the TopU test libraries used in Setting 1 and Setting 2 therefore represent focused hard-negative screening collections rather than arbitrary random splits of ChEMBL.

The resulting TopU libraries vary considerably in size across targets, reflecting natural variation in public assay coverage and the number of surviving actives after curation. Across the 93 targets, TopU libraries range from a few hundred compounds to well over ten thousand compounds per target. This variation is an intentional feature of the benchmark: some tasks are small and data-limited, while others support larger-scale screening, making the benchmark more representative of real discovery settings.

The statistics in Tables 7 and 8 show the relationship between the original ChEMBL counts, the cleaned ChEMBL* counts, and the final TopU counts for each target. In most cases, the number of TopU actives is substantially smaller than the number of cleaned actives, reflecting the stringency of the TopU selection process. Similarly, although ChEMBL* may contain many thousands of inactives for some targets, only a carefully chosen subset is used in TopU, again to maintain the intended hard-negative regime.

Overall, TopU should be viewed not simply as another decoy set, but as a target-specific benchmark library designed to stress-test LBVS methods under conditions that are much closer to realistic prospective screening than standard random-negative evaluation.

A.3 Label, Leakage, and Representation Robustness Audits

Inactive-label consistency. Because TopU intentionally selects chemically challenging inactives, we performed an additional annotation-consistency audit before decoy selection. Approximately 24.5% of compounds in the initial inactive pool had multiple retained ChEMBL measurements. Among these compounds, only 0.56% showed conflicting active/inactive annotations across mea-

surements; all such conflict-bearing compounds were removed from the inactive pool before TopU construction. After molecular standardization and filtering, none of the 224,440 final TopU decoys is recorded as active against its own target under the benchmark activity criterion. These checks remove known label conflicts, although, as with any retrospective ChEMBL-derived benchmark, unmeasured activity cannot be ruled out completely.

Training-negative augmentation audit. Same-class negative borrowing is used only for TopU-LBVS-full training pools when insufficient target-specific inactives remain after removing TopU compounds; it is never used in the TopU test libraries. Borrowing affects 15 of the 93 targets. After molecular standardization, only 0.039% of borrowed training-negative candidates overlapped compounds with documented activity against the recipient target. These compounds were removed from the released training splits.

Robustness to molecular representation. TopU construction uses physicochemical descriptors and MFP3 fingerprints to operationalize hard-negative similarity. To test whether the resulting difficulty is narrowly specific to this representation, we additionally evaluated active–decoy separability on the fixed TopU libraries using alternative representations, including ECFP4, ECFP6, RDKit fingerprints, MACCS keys, and physicochemical descriptors. The great majority of targets remain difficult under these alternative representations according to the same separability criterion. This is a post-construction robustness audit, rather than independent re-mining of the 93 libraries, and therefore does not claim invariance to every possible hard-negative construction. It nevertheless shows that TopU difficulty is not confined to the MFP3 representation used during library construction.

A.4 TopU-LBVS-low-mini: A Compact Low-Data Benchmark

Motivation. Evaluating new methods across all 93 targets of TopU-LBVS-low can be computationally demanding, particularly during model development, ablation, and hyperparameter analysis. TopU-LBVS-low-mini provides a compact, standardised entry point to the same low-data hard-negative regime, following the design philosophy of TopU-LBVS-mini. It preserves the fixed splits, metrics, active-to-decoy ratio, and evaluation code of the full TopU-LBVS-low protocol while substantially reducing evaluation cost.

Target selection. For each protein class, we select one Tier 1 target (fewer than 50 TopU actives) and one Tier 2 target ($\geq$ 50 TopU actives). Tier 2 targets are chosen to match the TopU-LBVS-mini collection, creating a direct connection between the two compact benchmarks. For Tier 1, we select the target with the largest number of TopU actives below the 50-active threshold, improving the stability of enrichment estimates while retaining the low-data setting. Cytochrome P450 has no Tier 1 target because all four CYP targets contain at least 50 TopU actives; it therefore contributes only its Tier 2 representative. This yields 13 targets across the 7 protein classes. The full target list is given in Table 6.

Table 6: **TopU-LBVS-low-mini target collection.** One Tier 1 and one Tier 2 target per protein class. Tier 2 targets match the TopU-LBVS-mini collection. TopU active counts determine tier assignment ($<$50: Tier 1; $\geq$50: Tier 2). Cytochrome P450 contributes one target only (no Tier 1 CYP exists). Splits, metrics, and evaluation code are identical to TopU-LBVS-low.

	Tier 1 ($<$50 actives, 6:2:2 split)			**Tier 2** ($\geq$50 actives, 7:1:2 split)		
Class	**Target**	**Full name**	**Actives**	**Target**	**Full name**	**Actives**
Cytochrome P450	–	(no Tier 1 target)	–	cp3a4	Cytochrome P450 3A4	154
GPCR	5ht6r	5-hydroxytryptamine receptor 6	47	aa2ar	Adenosine A2a receptor	127
Ion Channel	5ht3a	Serotonin 3a (5-HT3a) receptor	49	kcnh2	hERG	352
Kinase	akt2	RAC-beta serine/threonine kinase	48	egfr	Epidermal growth factor receptor	153
Nuclear Receptor	ppara	Peroxisome proliferator-activated receptor α	40	esr1	Estrogen receptor α	112
Misc. Enzymes	hivint	HIV integrase	48	aces	Acetylcholinesterase	117
Protease	fa10	Coagulation factor Xa	43	thrb	Thrombin	104

Protocol. TopU-LBVS-low-mini is a strict subset of TopU-LBVS-low and uses the same TopU-LBVS-low-mini is a strict subset of TopU-LBVS-low and uses the same fixed TopU libraries and predefined split seeds 2026, 2027, and 2028. Tier 1 targets use a 6:2:2 train/validation/test split and Tier 2 targets use a 7:1:2 split, with the 1:40 active-to-decoy ratio preserved throughout. The primary

metric is EF@10%, with PR-AUC as the secondary metric; BEDROC and the remaining ranking metrics are reported in Appendix H. Because the low-mini targets use the corresponding fixed splits from TopU-LBVS-low, their results are directly comparable to the same targets in the full 93-target benchmark. TopU-LBVS-low-mini is intended for rapid development and ablation; broad benchmark claims should be supported by evaluation on the full TopU-LBVS-low panel.

Relationship to TopU-LBVS-mini. The Tier 2 targets of TopU-LBVS-low-mini coincide with the seven targets in TopU-LBVS-mini and use the same underlying TopU libraries, but the protocols differ in training regime. TopU-LBVS-mini trains on historical ChEMBL* SAR data at a 1:10 training ratio, whereas TopU-LBVS-low trains on a limited labeled subset of the TopU hard-negative distribution at 1:40. The paired target collection therefore provides a controlled way to examine whether a method remains effective when moving from a relatively data-rich historical-SAR setting to a low-data hard-negative setting.

A.5 Target Panel and Protein Classes

TopU-LBVS is built on a panel of **93 protein targets** spanning **7 major protein classes**: Cytochrome P450 enzymes, GPCRs, ion channels, kinases, nuclear receptors, proteases, and other enzymes. Targets labelled as "Miscellaneous" in ChEMBL were excluded because they did not form a sufficiently coherent class for meaningful grouped analysis.

We begin from ChEMBL 35 and retain targets with substantial public bioactivity data, reliable target annotations, and enough active and inactive compounds to support both curation and benchmark construction. From this initial pool, we keep only targets for which a challenging TopU library can be constructed after cleaning and filtering, yielding the final set of 93 benchmark tasks used throughout the paper.

The target panel is intentionally diverse rather than class balanced. Kinases, GPCRs, and other enzymes contribute the largest number of tasks, while ion channels, proteases, nuclear receptors, and cytochrome P450 enzymes add distinct activity landscapes and screening regimes. This breadth is important because benchmark difficulty, decoy quality, and model behaviour can vary substantially across target families. A benchmark concentrated in only one or two families can therefore give a misleading picture of model performance.

The class structure also supports downstream protocol design. In Setting 1, after removing TopU compounds from ChEMBL*, some targets contain too few inactives to support the desired training ratio. In those cases, we allow limited augmentation of the negative pool using inactives drawn from other targets within the same protein class. Restricting this augmentation to the same class preserves a biologically plausible negative background while avoiding unrealistically small training sets for sparse targets.

Tables 7 and 8 report the full list of targets, their PDB identifiers, protein classes, and the numbers of raw ChEMBL, ChEMBL*, and TopU compounds per target. These statistics highlight the wide range of task sizes represented in the benchmark, from targets with only a few dozen TopU actives to targets with several hundred, and from focused libraries of a few hundred compounds to much larger hard-negative collections.

Across all 93 targets, the cleaned ChEMBL* data contain a median of 1 978 actives (min 121, max 8 751) and 2 737 inactives (min 463, max 61 814) per target. The TopU libraries contain a median of 53 actives (min 10, max 352) and a median total size of 2 173 compounds (min 410, max 14 432) per target. These per target statistics are provided to document the benchmark and support downstream analysis; all train, validation, and test splits are defined in the main text.

B Baseline Models

We provide full details of all baseline models, including architectures, training procedures, and hyperparameters. Table 9 summarises all baseline models, their paradigm, molecular representation, and key implementation details.

Table 7: Per target statistics for the TopU LBVS benchmark (part a). For each target we report counts of ChEMBL compounds before and after cleaning, as well as the number of actives and decoys in the TopU library.

			ChEMBL (raw)		ChEMBL* (cleaned)		TopU	
Target	**PDB**	**Class**	**active**	**decoy**	**active**	**decoy**	**active**	**decoy**
cp2c9	1r9o	Cytochrome P450	1893	24225	1882	24052	119	4760
cp1a2	2hi4	Cytochrome P450	1029	22581	1024	22415	136	5440
cp3a4	3nxu	Cytochrome P450	3702	30526	3686	30235	154	6160
cp2d6	4wnt	Cytochrome P450	1891	25216	1884	25028	139	5560
cxcr4	3odu	GPCR	919	855	907	853	15	600
5ht6r	8jlz	GPCR	4168	2212	4131	2189	47	1880
cnr2	5zty	GPCR	6645	3098	6604	3085	55	2200
adrb1	7bu7	GPCR	1408	2964	1384	2920	62	2480
hrh3	7f61	GPCR	4286	3051	4240	3023	65	2600
oprk	9mql	GPCR	5064	3992	5010	3948	84	3360
drd3	3pbl	GPCR	5351	4042	5304	3960	85	3400
oprm	9mqi	GPCR	5546	5403	5489	5329	113	4520
adrb2	3ny8	GPCR	1961	4648	1923	4575	102	4080
oprd	6pt3	GPCR	4321	4824	4265	4771	103	4120
drd2	6cm4	GPCR	7860	5197	7795	5119	113	4520
aa3r	8x16	GPCR	4190	5798	4142	5756	120	4800
aa2ar	3eml	GPCR	5194	6237	5159	6212	127	5080
trpa1	6x2j	Ion Channel	632	463	631	463	10	400
cac1h	9ayh	Ion Channel	399	1421	398	1417	27	1080
scn5a	6lqa	Ion Channel	766	2065	764	2061	43	1720
5ht3a	8bla	Ion Channel	778	2307	772	2289	49	1960
kcnh2	8zyq	Ion Channel	7110	16267	7067	16086	352	14080
wee1	3biz	Kinase	541	627	541	624	12	480
tgfr1	3hmm	Kinase	1129	1030	1127	1028	22	880
kpcb	2i0e	Kinase	620	1252	616	1235	26	1040
mk10	2zdt	Kinase	1262	1653	1258	1648	37	1480
fak1	3bz3	Kinase	1649	2123	1642	2099	44	1760
mp2k1	3eqh	Kinase	889	2043	887	2021	44	1760
rock1	2etr	Kinase	1667	2022	1649	2010	44	1760
csf1r	3krj	Kinase	1884	2059	1867	2048	46	1840
akt2	3d0e	Kinase	1021	2258	1016	2242	48	1920
kit	3g0e	Kinase	1536	2526	1530	2517	53	2120
braf	3d4q	Kinase	4066	2726	4060	2715	58	2320
mapk2	3m2w	Kinase	869	2718	868	2700	59	2360
plk1	2owb	Kinase	1253	25981	1252	25947	66	2640
igf1r	2oj9	Kinase	2428	3289	2427	3273	66	2640
mk01	2ojg	Kinase	3128	18279	3127	18105	68	2720
met	3lq8	Kinase	3784	3622	3777	3597	76	3040
akt1	3cqw	Kinase	3126	3744	3120	3706	77	3080
fgfr1	3c4f	Kinase	2829	3534	2807	3509	77	3080
jak2	3lpb	Kinase	6568	3777	6553	3757	84	3360
mk14	2qd9	Kinase	4362	4056	4356	4032	86	3440
lck	2of2	Kinase	1943	4375	1937	4346	93	3720
cdk2	1h00	Kinase	2133	4403	2125	4358	97	3880
abl1	2hzi	Kinase	2433	4548	2420	4496	97	3880
src	3el8	Kinase	3011	4570	2992	4529	98	3920
vgfr2	2p2i	Kinase	8825	6280	8751	6214	133	5320
egfr	2rgp	Kinase	7575	7645	7509	7582	153	6120

B.1 Category I: Classical Similarity-Based Baselines

These baselines match long-standing practice in LBVS and provide an interpretable reference point: any complex learned model should at least match or exceed their performance.

Table 8: Per target statistics for the TopU LBVS benchmark (part b).

			ChEMBL (raw)		ChEMBL* (cleaned)		TopU	
Target	**PDB**	**Class**	**active**	**decoy**	**active**	**decoy**	**active**	**decoy**
mcr	2aa2	Nuclear Receptor	944	490	943	489	10	400
rxra	1mv9	Nuclear Receptor	610	1066	609	1063	20	800
thb	1q4x	Nuclear Receptor	574	5734	574	5709	20	800
ppard	2znp	Nuclear Receptor	1698	2055	1689	2038	35	1400
ppara	2p54	Nuclear Receptor	2566	2365	2559	2341	40	1600
gcr	3bqd	Nuclear Receptor	2488	2992	2484	2952	56	2240
prgr	3kba	Nuclear Receptor	1877	2743	1874	2737	55	2200
esr2	2fsz	Nuclear Receptor	2135	2936	2108	2796	56	2240
andr	2am9	Nuclear Receptor	2502	3433	2492	3402	68	2720
pparg	2gtk	Nuclear Receptor	3990	3996	3979	3971	73	2920
esr1	1sj0	Nuclear Receptor	3268	5813	3226	5614	112	4480
hxk4	3f9m	Misc. Enzymes	1041	463	1041	463	10	400
pygm	1c8k	Misc. Enzymes	149	687	148	685	10	400
tysy	1syn	Misc. Enzymes	507	505	505	503	10	400
hmdh	3ccw	Misc. Enzymes	269	1064	261	1058	11	440
fnta	3e37	Misc. Enzymes	1745	652	1745	651	12	480
ampc	1l2s	Misc. Enzymes	170	61861	121	61814	12	480
dhi1	3frj	Misc. Enzymes	3295	675	3294	670	14	560
inha	4trj	Misc. Enzymes	267	671	261	671	14	560
pgh2	3ln1	Misc. Enzymes	462	746	462	743	15	600
dyr	3nxo	Misc. Enzymes	978	921	975	918	19	760
glcm	2v3f	Misc. Enzymes	415	12008	410	11966	21	840
parp1	3l3m	Misc. Enzymes	3596	952	3589	947	20	800
nos1	1qw6	Misc. Enzymes	567	1261	544	1246	19	760
pyrd	1d3g	Misc. Enzymes	1093	1084	1089	1084	25	1000
hdac2	3max	Misc. Enzymes	1995	1767	1978	1758	28	1120
hdac8	3f07	Misc. Enzymes	2129	1465	2123	1451	29	1160
pde5a	1udt	Misc. Enzymes	1872	1649	1866	1641	34	1360
hivint	3nf7	Misc. Enzymes	2275	2462	2254	2447	48	1920
cah2	1bcd	Misc. Enzymes	7822	3078	7540	2942	62	2480
pgh1	2oyu	Misc. Enzymes	900	2984	898	2974	63	2520
ptn1	2azr	Misc. Enzymes	2377	3377	2375	3352	64	2560
aofb	7p4h	Misc. Enzymes	3417	3014	3403	2996	68	2720
hivrt	3lan	Misc. Enzymes	4797	3676	4702	3576	75	3000
aces	1e66	Misc. Enzymes	4701	5907	4575	5797	117	4680
ace	3bkl	Protease	581	529	578	529	10	400
mmp13	830c	Protease	2588	762	2585	763	17	680
try1	2ayw	Protease	1223	817	1165	797	16	640
urok	1sqt	Protease	856	868	840	848	17	680
ada17	2oi0	Protease	1728	877	1728	876	18	720
hivpr	1xl2	Protease	4511	1556	4491	1550	25	1000
dpp4	2i78	Protease	4009	1483	3875	1437	31	1240
fa10	3kl6	Protease	5395	2188	5315	2126	43	1720
bace1	3l5d	Protease	7195	2809	7158	2786	54	2160
thrb	1ype	Protease	4661	5339	4434	5256	104	4160

Morgan Fingerprint + Random Forest (Morgan-RF). We compute 2048-bit Morgan (ECFP4) fingerprints (Rogers & Hahn, 2010) with radius $r = 2$ for each compound using RDKit (Landrum, 2006), and train a Random Forest classifier (Breiman, 2001) with balanced class weights to account for the active-to-decoy imbalance. The predicted probability of the active class is used as the ranking score. Morgan-RF is included as a canonical fingerprint-based LBVS baseline, reflecting decades of established practice in the field (Willett, 2006).

Tanimoto Nearest-Neighbour Search (Tanimoto-NN). For each test compound, we compute the maximum Tanimoto coefficient against all training actives using 2048-bit Morgan fingerprints,

Table 9: **Summary of all baseline models.** "Group" indicates the model family (Similarity based, GNN / Hybrid, Modern pretrained). All neural models use BCE loss with positive class weighting and Adam optimisation. Validation metric is PR-AUC for all models.

Group	Model	Paradigm	Representation	Parameters
Naive	Morgan-RF (Rogers & Hahn, 2010)	Fingerprint + Ensemble	Morgan FP (2048-bit, $r = 2$)	up to 500 trees
	Tanimoto-NN (Willett, 2006)	Similarity search	Morgan FP (2048-bit, $r = 2$)	–
GNN & Hybrid	GIN (Xu et al., 2019)	Message passing (sum)	Molecular graph	5L, $d = 256$
	GIN+FP	Message passing + FP	Graph ‖ Morgan FP	5L, $d = 256$
	GAT (Veličković et al., 2018)	Attention weighted MP	Molecular graph	5L, $d = 256$, 4 heads
	GAT+FP	Attention weighted MP + FP	Graph ‖ Morgan FP	5L, $d = 256$, 4 heads
	GPS (Rampášek et al., 2022)	Graph Transformer	Molecular graph	5L, $d = 256$, 4 heads
	GPS+FP	Graph Transformer + FP	Graph ‖ Morgan FP	5L, $d = 256$, 4 heads
Modern pre-trained	D-MPNN (Heid et al., 2023)	Directed message passing	Graph + RDKit (200-dim)	3 steps, $d = 300$
	MolFormer (Ross et al., 2022)	Pretrained SMILES Transformer	SMILES (1.1B pretraining)	Fine tuned

and use this maximum similarity as the ranking score. No learning is involved; the method ranks compounds purely by structural proximity to known actives. Tanimoto-NN is one of the most widely deployed LBVS methods in practice (Willett, 2006) and serves as a crucial reference point: a learned model that cannot outperform Tanimoto-NN provides no practical benefit over off-the-shelf similarity search. This baseline is particularly informative under our three-setting design, as its performance gap relative to learned models is expected to change noticeably, and often widen, on the TopU test sets (Settings 1 and 2), where hard negatives are selected by a closely related topological similarity criterion.

B.2 Category II: Graph Neural Networks and Hybrid Models

Graph neural networks (GNNs) operate directly on the molecular graph, learning hierarchical representations of atoms and bonds without relying on fixed hand-crafted features. We evaluate three message-passing architectures and one Graph Transformer, together with their fingerprint-augmented hybrid variants.

Molecular Graph Representation. Each molecule is represented as an undirected graph $\mathcal{G} = (V, E)$, where nodes $v \in V$ correspond to heavy atoms and edges $e \in E$ correspond to covalent bonds. Node features (9 dimensional) encode atomic number, degree, formal charge, number of hydrogens, number of radical electrons, hybridisation state, aromaticity, ring membership, and chirality. Edge features (3 dimensional) encode bond type, conjugation, and ring membership.

Graph Isomorphism Network (GIN). GIN (Xu et al., 2019) uses sum aggregation with a learned ϵ-weighted self-loop, achieving the maximum expressive power of 1-dimensional Weisfeiler-Leman graph isomorphism tests among message-passing GNNs. We use 5 message-passing layers, hidden dimension 256, global sum pooling, and a 2-layer MLP classification head. All hyperparameters follow common practice in molecular GNN benchmarks, with exact values listed in Appendix C.

Graph Attention Network (GAT). GAT (Veličković et al., 2018) replaces uniform neighbourhood aggregation with multi-head attention, allowing the model to assign differential importance to neighbours. We use 4 attention heads, 5 layers, hidden dimension 256, and the same pooling and classification head as GIN (full details in Appendix C).

Graph Transformer (GPS). GPS (Rampášek et al., 2022) combines local message-passing with global Transformer-style self-attention, enabling the model to capture both local chemical environments and long-range atomic dependencies within a single unified architecture. We use 5 GPS layers with hidden dimension 256, 4 attention heads, and sum pooling.

Hybrid Variants (GNN + Morgan FP). For each of the three GNN architectures above, we additionally evaluate a hybrid model combining the learned graph embedding with a 2048-bit Morgan fingerprint via a late-fusion MLP. The GNN encoder produces a graph-level embedding $\mathbf{h}_{\text{graph}}$, while the Morgan fingerprint $\mathbf{fp}_{\text{Morgan}} \in \{0, 1\}^{2048}$ is passed through a separate two-layer MLP with ReLU activation and dropout to obtain a fingerprint embedding $\mathbf{h}_{\text{fp}}$ of the same dimensionality as $\mathbf{h}_{\text{graph}}$.

The two embeddings are concatenated and fed to a two-layer classification head:

$$\mathbf{h}_{\text{hybrid}} = [\mathbf{h}_{\text{graph}} \parallel \mathbf{h}_{\text{fp}}], \qquad \hat{y} = \text{MLP}_{\text{cls}}(\mathbf{h}_{\text{hybrid}}).$$

Each hybrid variant introduces an additional fingerprint-projection MLP and a wider classification head whose input is the concatenated graph and fingerprint embeddings. This late-fusion design allows the model to leverage both the expressive learned graph representations and the well-validated substructural information encoded by Morgan fingerprints, with the list of hyperparameters reported in Appendix C. We denote the hybrid variants GIN+FP, GAT+FP, and GPS+FP.

Training Details (All GNN Models). All GNN models are trained end-to-end using binary cross-entropy loss with positive class weighting: $w_+ = \frac{N_{\text{decoy}}}{N_{\text{active}}}$ to compensate for class imbalance. We use the Adam optimiser (Kingma & Ba, 2014) with learning rate 10^{-3}, weight decay 10^{-4}, batch size 64, and a maximum of 100 epochs with early stopping based on validation PR-AUC (patience $= 20$ epochs). The learning rate is reduced by a factor of 0.5 when validation PR-AUC plateaus for 5 consecutive epochs. A complete list of hyperparameters and training schedules is given in Appendix C.

B.3 Category III: Modern Pretrained Models

Modern pretrained models leverage large-scale self-supervised pretraining on unlabelled molecular data, followed by task-specific fine-tuning on the target dataset. We select two models representing distinct and complementary pretraining paradigms that are among the most widely adopted in contemporary molecular property prediction.

D-MPNN (Chemprop). The Directed Message Passing Neural Network (D-MPNN), implemented in Chemprop (Yang et al., 2019; Heid et al., 2023), propagates messages along directed edges rather than nodes, mitigating oversmoothing and enabling more precise encoding of local chemical environments. The molecular embedding is obtained by aggregating learned directed bond representations, and passed to a feed-forward network (FFN) for classification. Unlike the GNN models in Category II, the D-MPNN additionally accepts an optional vector of global molecular features; we augment the molecular embedding with 200-dimensional RDKit physicochemical descriptors, following the standard recommended configuration (Yang et al., 2019). We use the Chemprop v2 default architecture: 3 message-passing steps, hidden dimension 300, depth-2 FFN, dropout $p = 0.0$, and ReLU activations. Models are trained with binary cross-entropy loss and the Adam optimiser for up to 50 epochs with early stopping on validation PR-AUC (patience $= 10$ epochs). D-MPNN is included as a widely adopted and empirically validated modern LBVS baseline, with successful applications spanning antibiotic discovery (Stokes et al., 2020) and large-scale bioactivity prediction.

MolFormer (Fine-tuned). MolFormer (Ross et al., 2022) is a large-scale chemical language model pretrained on over 1.1 billion SMILES strings from PubChem and ZINC using a masked language modelling objective, with linear attention and rotary position embeddings for computational efficiency. We use the publicly released `ibm/MoLFormer-XL-both-10pct` checkpoint. For each target, the pretrained encoder produces token-level representations, which are pooled into a single molecular embedding by attention-mask-weighted mean pooling over the last hidden state. following the IBM MolFormer fine-tuning reference. Models are trained with binary cross-entropy loss using positive-class weighting (pos_weight $= n_{\text{neg}}/n_{\text{pos}}$) to compensate for class imbalance, and optimized end-to-end with the LAMB optimizer at a single learning rate of 3×10^{-5} (no weight decay). SMILES strings are canonicalized with `isomericSmiles=False` and truncated to 202 tokens (MolFormer's maximum position embedding). We use a batch size of 32, a maximum of 30 fine-tuning epochs, and early stopping on validation PR-AUC (patience $= 5$ epochs); the checkpoint with the highest validation PR-AUC is retained. MolFormer serves as the representative large-scale pretrained transformer baseline, allowing us to test whether chemical foundation model pretraining improves LBVS performance in our hard-negative and data-limited regimes.

C HYPERPARAMETERS AND IMPLEMENTATION DETAILS

This section provides complete hyperparameter configurations and implementation details for all ten baseline models. All models are implemented in PyTorch (Paszke et al., 2019), with GNN models using PyTorch Geometric (Fey & Lenssen, 2019) and fingerprint computation using RDKit (Landrum, 2006). Chemprop v2 is used for D-MPNN, and the pretrained MolFormer weights are loaded from the official HuggingFace repository.[2] All split files, model configurations, training scripts, and random seeds are released alongside the benchmark to ensure full reproducibility.

C.1 CLASSICAL BASELINES

Morgan-RF. Each compound is represented by a 2048-bit Morgan (ECFP4) fingerprint with radius $r = 2$. We train a scikit-learn `RandomForestClassifier` with `max_features="sqrt"` and balanced class weights $w_j = n/(2 \cdot n_j)$ to account for the active-to-decoy imbalance. The random seed (2026, 2027, 2028) controls both bootstrap sampling and feature selection. Trees are grown incrementally with early stopping on validation PR-AUC: starting from 10 trees, we add 10 trees per step and stop after 5 consecutive steps without PR-AUC improvement, with a hard cap of 500 trees. The optimal tree count $n^\star$ is then refit from scratch to obtain the final model. The predicted probability of the active class is used as the ranking score.

Tanimoto-NN. Each compound is represented by a 2048-bit Morgan fingerprint identical to the Morgan-RF representation. The model stores fingerprints of training actives only; the validation set and decoys are not used. At inference, we compute the full $(n_{\text{test}} \times n_{\text{train-actives}})$ Tanimoto similarity matrix and use the row-wise maximum as the ranking score. Tanimoto similarity is defined as $T(a, b) = (a \cdot b)/(\|a\|_1 + \|b\|_1 - a \cdot b)$. No model selection or hyperparameter tuning is performed. Tanimoto-NN has no random state, but each seed produces a different training split, so the three-seed variance reported in the main tables reflects split-induced variance only.

C.2 GNN AND HYBRID MODELS

All GNN models share the training infrastructure described below unless otherwise noted.

Shared training setup. All GNN models are trained end-to-end as binary classifiers using `BCEWithLogitsLoss` with positive-class weighting

$$w_+ = \frac{N_{\text{decoy}}}{N_{\text{active}}},$$

to compensate for class imbalance in each training split. We use the Adam optimizer with learning rate 10^{-3}, weight decay 10^{-4}, batch size 64, and a maximum of 100 epochs. Hyperparameter selection is performed with seed 2026 on each target. Within each run, model selection uses validation PR-AUC: after every epoch, PR-AUC is computed on the held-out validation split, the checkpoint with the best validation PR-AUC is saved, and training stops early if validation PR-AUC does not improve for 20 consecutive epochs (patience $= 20$). We additionally apply `ReduceLROnPlateau` on validation PR-AUC (`mode=max`), reducing the learning rate by a factor of 0.5 after 5 consecutive non-improving epochs..

Backbone-specific architectures. Under the shared training protocol above, all three graph backbones (GIN, GAT, and GPS) use five message-passing layers with hidden width 256, sum pooling, projection to a 32-dimensional graph embedding, and a two-layer prediction MLP with ReLU and dropout between the two linear layers.

These models differ only in the message-passing operator. GIN uses Graph Isomorphism Network layers with a two-layer update MLP and a learnable ϵ term initialized at zero (i.e., sum-based neighborhood aggregation with a scaled self-contribution), comprising a total of 672,358 trainable parameters. GAT uses four-headed graph attention in each layer, with head outputs averaged rather than concatenated so the hidden dimensionality remains 256 throughout, and it has 1,336,673 trainable parameters in total. GPS uses GPS layers that combine a local GIN-style message-passing module

[2] https://huggingface.co/ibm/MoLFormer-XL-both-10pct

(matching the GIN operator above) with a global four-head self-attention module, followed by the standard feed-forward and normalization sublayers of GPS blocks, yielding 3,310,438 trainable parameters. Dropout for all backbones is tuned over $p \in \{0.3,\ 0.5\}$.

Hybrid variants (GIN+FP, GAT+FP, GPS+FP). Each GNN encoder produces a 32-dimensional graph embedding which, for hybrid variants, is concatenated with a 32-dimensional fingerprint embedding. This embedding is obtained by passing the 2048-bit Morgan fingerprint through a two-layer projection MLP with ReLU and dropout, yielding a 64-dimensional joint representation. The classification head is widened accordingly for an input dimension of 64 instead of 32. Its hidden dimension d_{fuse} is tuned over $\{64, 128, 256\}$. The fingerprint projector hidden width d_{fp} is tuned over $\{128, 256, 512\}$. All other settings are inherited unchanged from the corresponding pure-GNN variant from the shared training setup, and the tuning values for dropout, like for the backbones, remains over $p \in \{0.3,\ 0.5\}$.

C.3 Modern Pretrained Models

D-MPNN (Chemprop). We use the Chemprop v2 reference architecture (Heid et al., 2023). Each molecule is represented as a directed atom-bond graph; messages are propagated along directed edges for $T = 3$ message-passing steps with hidden dimension 300 and ReLU activations. The molecular embedding is concatenated with 200-dimensional RDKit physicochemical descriptors (computed via the standard `normalized_features` generator and standardized with a per-fold `StandardScaler`) before being passed to a 2-layer feed-forward network of hidden size 300 with dropout $p = 0.0$. Models are trained with binary cross-entropy loss using positive-class weighting and optimized with Adam (Kingma & Ba, 2014) under the Chemprop v2 default Noam learning-rate schedule (initial LR 10^{-4}, peak LR 10^{-3} after 2 warmup epochs, final LR 10^{-4}). We use batch size 64, a maximum of 50 epochs, and early stopping on validation PR-AUC. Following Chemprop v2, edge updates use preactivation initial hidden states. The checkpoint with the highest validation PR-AUC is retained.

MolFormer. We fine-tune the publicly released `ibm/MoLFormer-XL-both-10pct` checkpoint, loaded via the HuggingFace `transformers` library (`trust_remote_code=True` required by the IBM custom modeling code). SMILES are canonicalized with `isomericSmiles=False` and tokenized with the official MolFormer tokenizer; sequences are truncated to the maximum position embedding length of 202 tokens. The encoder produces a 768-dimensional pooled molecular embedding via attention-mask-weighted mean pooling over the last hidden state. A three-layer MLP head with two residual skip connections (hidden dimensions 768/768/768, output dimension 1) is appended, matching the IBM reference fine-tuning architecture. All parameters are optimized end-to-end with the LAMB optimizer at a single learning rate of 3×10^{-5} and weight decay 0.0, with parameters grouped into a "decay" group (Linear weights) and a "no_decay" group (LayerNorm/Embedding weights and all biases) following the IBM reference. We use batch size 32, a maximum of 30 fine-tuning epochs, and early stopping on validation PR-AUC with patience $= 5$ epochs. The checkpoint with the highest validation PR-AUC is retained.

C.4 Computational Resources

All experiments were run on a shared university HPC cluster using single-GPU training runs throughout. The cluster provides a heterogeneous mix of NVIDIA GPUs, including A100 (40 GB and 80 GB), A100 MIG slices, H100 NVL, H200 NVL, A30, V100 (16 GB), P100 (16 GB), RTX 3090 (24 GB), and RTX Pro 6000. MolFormer fine-tuning was preferentially scheduled on A100 (80 GB) or H100 nodes due to its higher memory footprint; smaller GNN and fingerprint baselines ran on whatever GPU was available.

Per-target wall-clock time varies substantially with model and target size. Classical fingerprint baselines (Morgan-RF, Tanimoto-NN) complete in under a minute per target. Pure GNN backbones (GIN, GAT, GPS) and their fingerprint-augmented variants typically train in 5–30 minutes per target. D-MPNN takes roughly 10–60 minutes per target depending on dataset size. MolFormer fine-tuning is the most expensive baseline, requiring approximately 15–90 minutes per target on an A100 (80 GB). The full benchmark (all 93 targets, 10 models, 3 seeds, both Settings 1 and 2) is on the order of 2,000–3,000 GPU-hours total, dominated by D-MPNN and MolFormer fine-tuning.

Implementation Notes

Neural baselines are implemented in PyTorch (Paszke et al., 2019) with GNN models using PyTorch Geometric (Fey & Lenssen, 2019). Classical baselines (Morgan-RF, Tanimoto-NN) use scikit-learn and NumPy; fingerprint computation uses RDKit (Landrum, 2006). Chemprop v2 is used for D-MPNN, and the pretrained MolFormer weights are loaded from the official HuggingFace repository.[3] All split files, model configurations, training scripts, and random seeds are released alongside the benchmark to ensure full reproducibility. A complete description of all hyperparameters is provided in Appendix C.

D Metric Definitions

In all settings, models produce a real valued score for each test compound of a given target. Sorting compounds in descending order of this score induces a ranking. All metrics below are computed per target from this ranking, then aggregated across targets as described at the end of this section.

Let

- N be the number of test compounds for a given target,
- N^+ be the number of actives among these N compounds,
- $N^- = N - N^+$ be the number of decoys,
- $\pi = N^+/N$ be the fraction of actives in the test set.

Let the ranked list be $c_1, c_2, \ldots, c_N$, where c_1 is the top ranked compound. Let $y_i \in \{0, 1\}$ denote the label of c_i (1 for active, 0 for decoy).

D.1 Enrichment Factor at x percent (EF@x%)

For a given fraction $x \in (0, 1]$, define $k = \lceil xN \rceil$ as the number of top ranked compounds considered. Let $H(x) = \sum_{i=1}^{k} y_i$ be the number of actives found among the top k compounds.

The enrichment factor at x percent, EF@x%, is defined as $\textbf{EF@x\%} = \dfrac{\frac{H(x)}{k}}{\frac{N^+}{N}} = \dfrac{H(x)\,N}{k\,N^+}$.

Under random ranking, $\mathbb{E}[H(x)] = k\,N^+/N$, so EF@$x\% = 1$. Values greater than 1 indicate enrichment of actives relative to random.

In this work we use EF@1% and EF@10%. For smaller libraries, EF@1% can be coarse because k is small, but it still reflects the very top of the ranking. EF@10% provides a more stable summary of early enrichment.

D.2 Precision Recall AUC (PR AUC)

For a threshold τ on the model score, let

$$\mathrm{TP}(\tau), \quad \mathrm{FP}(\tau), \quad \mathrm{FN}(\tau)$$

denote the numbers of true positives, false positives, and false negatives respectively. Precision and recall at threshold τ are

$$\mathrm{Precision}(\tau) = \frac{\mathrm{TP}(\tau)}{\mathrm{TP}(\tau) + \mathrm{FP}(\tau)}, \qquad \mathrm{Recall}(\tau) = \frac{\mathrm{TP}(\tau)}{\mathrm{TP}(\tau) + \mathrm{FN}(\tau)}.$$

Sweeping τ from $+\infty$ to $-\infty$ traces out the precision recall curve. The precision recall area under the curve (PR AUC), also known as average precision, is defined as the area under this curve

$$\text{PR AUC} = \int_0^1 \mathrm{Precision}(r)\,\mathrm{d}r,$$

[3] https://huggingface.co/ibm/MoLFormer-XL-both-10pct

where r denotes recall. In practice we approximate this integral using the standard stepwise interpolation over the discrete set of thresholds induced by the ranking. PR AUC is more informative than ROC AUC in the strongly imbalanced regime relevant for LBVS.

D.3 Receiver Operating Characteristic AUC (ROC AUC)

For the same threshold τ, define the true positive rate and false positive rate as

$$\mathrm{TPR}(\tau) = \frac{\mathrm{TP}(\tau)}{N^+}, \qquad \mathrm{FPR}(\tau) = \frac{\mathrm{FP}(\tau)}{N^-}.$$

Sweeping τ produces the receiver operating characteristic (ROC) curve, which plots TPR versus FPR. The ROC area under the curve (ROC AUC) is

$$\text{ROC AUC} = \int_0^1 \mathrm{TPR}(f)\,\mathrm{d}f,$$

where f denotes the false positive rate. Equivalently, ROC AUC is the probability that a randomly chosen active is scored higher than a randomly chosen decoy. ROC AUC can remain high even when early enrichment is modest, so we treat it as a secondary, legacy metric and report it in the appendix.

D.4 BEDROC

Boltzmann enhanced discrimination of ROC (BEDROC) is a rank based metric that places an exponential weight on early retrieved actives. Let N^+ be the number of actives, and let $r_j \in \{1, \ldots, N\}$ denote the rank of the jth active in the sorted list (lower rank indicates higher score). For a given parameter $\alpha > 0$, define

$$R = \frac{1}{N^+} \sum_{j=1}^{N^+} \exp\!\left(-\alpha\, \frac{r_j}{N}\right).$$

BEDROC rescales R so that a random ranking yields a value of 0 and a perfect ranking yields 1:

$$\mathrm{BEDROC}(\alpha) = \frac{R - R_{\text{rand}}}{R_{\max} - R_{\text{rand}}}.$$

Here R_{rand} and $R_{\max}$ are constants that depend on α, N, and N^+ and correspond to the expected value under random ranking and the value under a perfect ranking respectively. Larger α places more weight on the very top of the ranked list. In our experiments we fix a single α value and use BEDROC only as an additional early enrichment summary in the appendix.

D.5 Aggregation Across Targets and Seeds

All metrics above are computed separately for each target. For each method and setting we then aggregate as follows:

- For each target, we average the metric over three predefined runs using the corresponding released splits for seeds 2026, 2027, and 2028.
- Across targets, we report class-wise averages and overall means of the per-target values in the main tables.
- The appendix reports detailed per-target results and additional evaluation metrics.

This per target then across targets aggregation avoids biasing metrics toward targets with larger test sets.

E Additional Details for Tasks and Evaluation Settings

This section expands the main protocol description in Section 4 and Table 2 with full per setting details, including split rules and per target counts.

E.1 PER TARGET LBVS TASK

Each of the 93 protein targets defines a separate LBVS task. Given a ligand (SMILES or molecular graph) and a fixed protein target, the goal is to rank compounds by the probability of being active. For target t, the model outputs a real valued score $s_t(c)$ for each test compound c, which induces a ranking. Evaluation metrics are computed per target from this ranking and then aggregated across targets.

The targets span 7 protein classes (GPCRs, kinases, nuclear receptors, ion channels, proteases, cytochrome P450s, and miscellaneous enzymes), with per target test libraries ranging from a few hundred to over ten thousand compounds. Targets classified as "Miscellaneous" in ChEMBL were excluded due to insufficient class coherence for meaningful grouping. TopU libraries are constructed at a fixed 1 to 40 active to decoy ratio using property matched, structurally challenging decoys (Section 3).

E.2 SETTING 1: CHEMBL* TO TOPU (MAIN BENCHMARK)

In Setting 1, models are trained on ChEMBL* and evaluated on the TopU libraries. This setting is intended to mimic a realistic prospective workflow: train on historical SAR for a target and then screen a new, challenging library.

For each target, we start from ChEMBL 35, apply our cleaning pipeline (Section A.1), and then remove all TopU compounds for that target to ensure zero overlap between training and test. The remaining ChEMBL* actives and decoys form the training pool. We construct per target training sets at a fixed 1 to 10 active to decoy ratio. When a target has too few decoys after removing TopU, we augment its negative pool with decoys sampled from other targets within the same protein class, using a fixed random seed and releasing the exact selections with the benchmark.

From this pool, we reserve 15 % as a stratified random validation set, preserving the 1:10 ratio, and use the remaining 85% for training. The test set for Setting 1 is the full TopU library for each target, constructed at a 1 to 40 active to decoy ratio. All models use the same fixed train, validation, and test splits.

For Setting 1, the primary test metric is EF@1%, and PR AUC is reported as a secondary ranking metric. Model selection is performed by maximizing PR AUC on the validation set. We aggregate results across the 93 targets by reporting mean and median per target EF@1%, with analogous aggregation for PR AUC.

E.3 SETTING 2: TOPU ONLY (FEW SHOT)

Setting 2 evaluates models when both training and test compounds come from the same hard, property matched TopU distribution. This setting captures a few shot regime where only a limited number of high quality TopU actives are available for training.

For each target, all splits (train, validation, test) are drawn from its TopU library at the fixed 1 to 40 active to decoy ratio. Because the number of TopU actives varies widely across targets, we use a tiered split strategy based on the number of TopU actives. Targets with fewer than 50 TopU actives (Tier 1) use a 6:2:2 split of actives into train, validation, and test, while targets with at least 50 TopU actives (Tier 2) use a 7:1:2 split. In both tiers, decoys are assigned to splits to preserve the 1 to 40 ratio per target.All splits are scaffold based and use the predefined seeds 2026, 2027, and 2028.

Because the smallest TopU libraries make EF@1% numerically unstable (only a few molecules in the top 1%), we use EF@10% as the primary test metric for Setting 2 and PR AUC as the secondary metric. We report mean and median EF@10% separately for Tier 1 and Tier 2, as well as overall. Detailed split rules and per target ranges of train, validation, and test actives are summarized in Table 10.

Table 10: Tier summary for Setting 2. Active counts are shown as min to max ranges across targets within each tier.

Tier	Cutoff	Targets	Split	Train actives	Val actives	Test actives
Tier 1	< 50	46	6:2:2	6 to 29	2 to 10	2 to 10
Tier 2	≥ 50	47	7:1:2	37 to 246	5 to 35	11 to 71

E.4 Setting 3: ChEMBL* Only (Ablation)

Setting 3 is an ablation that uses only ChEMBL* for both training and test, with randomly sampled negatives and no TopU decoy curation. The goal is to contrast performance on standard ChEMBL style splits with the harder TopU test libraries in Setting 1.

To keep the comparison focused, Setting 3 is defined on seven representative targets, one from each protein class, chosen as the targets with the largest number of TopU actives in that class. The selected targets are listed in Table 11.

Table 11: Selected targets for Setting 3, one per protein class.

Class	Target	Full Name	TopU actives	ChEMBL* actives
Cytochrome P450	cp3a4	Cytochrome P450 3A4	154	3,686
GPCR	aa2ar	Adenosine A2a receptor	127	5,159
Ion Channel	kcnh2	hERG	352	7,067
Kinase	egfr	Epidermal growth factor receptor	153	7,509
Nuclear Receptor	esr1	Estrogen receptor α	112	3,226
Misc. Enzymes	aces	Acetylcholinesterase	117	4,575
Protease	thrb	Thrombin	104	4,434

To enable a direct, controlled comparison with Setting 1, the training and validation sets in Setting 3 are constructed to be identical in size to those used in Setting 1 for the same targets. Specifically, for each selected target, the TopU compounds are first excluded from ChEMBL* to prevent any overlap between the training pool and the Setting 1 test set. The remaining ChEMBL* compounds are then split into training and validation sets using the same 85%/15% stratified random split (seed 2026) and 1 to 10 active to decoy ratio as Setting 1.

For each target, we pool eligible ChEMBL* inactives (including same-class augmentation when used in Setting 1) together with TopU inactives to form the candidate random-decoy pool. Before sampling, we exclude compounds that overlap with the fixed Setting 1 training or validation sets, as well as compounds with documented activity against the target. The remaining eligible decoys are shuffled deterministically using the per-target seed, and the Setting 3 test decoys are sampled at the same 1:40 active-to-decoy ratio as the corresponding Setting 1 TopU test set. Thus, the training and validation sets, test actives, active counts, and evaluation protocol are identical across Settings 1 and 3; the only difference is the construction of the test inactives: a focused TopU hard-decoy library in Setting 1 versus randomly sampled decoys from the leakage-filtered combined inactive pool in Setting 3. Across the seven mini targets, TopU compounds make up a small minority of the Setting 3 test inactives (typically under 10%), so the Setting 1 vs. Setting 3 comparison reported in Table 5 is conservative. The true gap between hard-decoy and random-decoy evaluation is therefore at least as large as Table 5 shows.

Table 12 reports the resulting dataset sizes for each target.

As in Settings 1 and 2, the validation set uses PR AUC (Average Precision) as the model selection metric. The primary test metric for Setting 3 is EF@1%, identical to Setting 1, enabling direct comparison between the two settings. We also report PR AUC on the ChEMBL* test sets as a secondary metric, using the same aggregation protocol as in Setting 1. Performance is reported per target as well as mean and median EF@1% across the seven selected targets.

Table 12: **Dataset sizes for Setting 3.** All counts are number of active compounds. Decoys follow a 1:10 active-to-decoy ratio in the train and validation splits and a 1:40 ratio in the test split.

Target	Class	Train actives	Val actives	Test actives
cp3a4	Cytochrome P450	3,002	530	154
aa2ar	GPCR	4,277	755	127
kcnh2	Ion Channel	5,707	1,008	352
egfr	Kinase	6,252	1,104	153
esr1	Nuclear Receptor	2,646	468	112
aces	Misc. Enzymes	3,789	669	117
thrb	Protease	3,680	650	104

F Limitations and Scope

TopU-LBVS is designed to stress-test ligand-based virtual screening methods under hard-negative conditions, and its scope reflects this focus. First, the benchmark is restricted to 2D ligand-based representations and does not incorporate protein structural information, docking, or 3D conformational effects. As a result, it does not evaluate structure-based screening methods or hybrid ligand–protein models. Second, TopU decoys are intentionally adversarial, emphasizing worst-case scenarios in which inactive compounds are highly similar to known actives. While this design exposes failure modes that standard benchmarks obscure, it may overestimate difficulty relative to early exploratory screening tasks with more diverse libraries. Third, bioactivity labels are derived from curated public data and binarized using fixed thresholds. Although this reflects common practice and improves consistency, it does not capture experimental uncertainty, assay context, or graded potency information. Finally, the benchmark evaluates a fixed panel of targets drawn from ChEMBL 35. While diverse across major protein classes, conclusions may not directly generalize to underrepresented target families or proprietary chemical spaces. These limitations are intentional trade-offs that prioritize controlled evaluation of robustness and generalization over broad task coverage. TopU-LBVS is intended to complement, rather than replace, existing LBVS benchmarks.

G Broader impacts

TopU-LBVS is intended to improve the reliability and reproducibility of ligand-based virtual screening research by providing fixed hard-negative protocols, standardized metrics, reference baselines, and released evaluation code. By exposing the gap between random-decoy performance and hard-negative screening performance, the benchmark may help reduce overoptimistic claims about molecular machine learning models and encourage methods that are more robust under realistic early-stage discovery conditions. The benchmark could support more efficient prioritization of candidate compounds, potentially reducing wasted computational and experimental effort. At the same time, TopU-LBVS is a retrospective evaluation resource and should not be interpreted as evidence that any model is ready for clinical or experimental deployment. As with other molecular screening tools, improved virtual screening methods could be misused to prioritize harmful bioactive compounds; however, TopU-LBVS does not provide generative design capabilities, synthesis routes, or prospective validation, and its release is focused on transparent evaluation rather than molecule generation or deployment. We therefore view the primary impact as improving scientific rigor and reproducibility in molecular machine learning benchmarks.

H Additional Results

This appendix provides complete per-target numerical results for all three TopU-LBVS protocols. It is organised in three parts. Part I contains the primary-metric per-target tables — one table per protocol — which are the canonical reference for comparing new methods against our baselines. Part II contains full per-model, per-metric tables for TopU-LBVS-full, reporting all seven metrics (EF@1%, EF@5%, EF@10%, PR-AUC, ROC-AUC, BEDROC, LogAUC) for each of the ten baselines individually. Part III contains the analogous full per-model, per-metric tables for TopU-LBVS-low.

A quick-reference index is given in Table 13.

Table 13: **Appendix table index.** All per-target result tables and their locations.

Content	Table(s)	Section
Part I — Primary metrics, all models, all targets		
TopU-LBVS-full: EF@1%, all 10 models × 93 targets	20–39	H.1
TopU-LBVS-low: EF@10%, all 10 models × 93 targets	40–59	H.1
TopU-LBVS-mini: EF@1%, TopU hard decoys (T), all 10 models × 7 targets	18	H.1
TopU-LBVS-mini: EF@1%, random decoys (R), all 10 models × 7 targets	19	H.1
Part II — Full metrics per model, TopU-LBVS-full		
Morgan-RF	20–21	H.2
Tanimoto-NN	22–23	H.2
GAT	24–25	H.2
GAT+FP	26–27	H.2
GIN	28–29	H.2
GIN+FP	30–31	H.2
GPS	32–33	H.2
GPS+FP	34–35	H.2
D-MPNN	36–37	H.2
MolFormer	38–39	H.2
Part III — Full metrics per model, TopU-LBVS-low		
Morgan-RF	40–41	H.3
Tanimoto-NN	42–43	H.3
GAT	44–45	H.3
GAT+FP	46–47	H.3
GIN	48–49	H.3
GIN+FP	50–51	H.3
GPS	52–53	H.3
GPS+FP	54–55	H.3
D-MPNN	56–57	H.3
MolFormer	58–59	H.3

H.1 Part I: Primary-Metric Per-Target Results

Tables 14 and 15 report per-target EF@1% for TopU-LBVS-full across all 93 targets and all 10 models. Tables 16 and 17 report per-target EF@10% for TopU-LBVS-few, grouped by protein class with tier indicated. Tables 18 and 19 report per-target EF@1% for TopU-LBVS-mini under TopU hard decoys (T) and randomly sampled ChEMBL* decoys (R), respectively; the gap between the two tables quantifies decoy difficulty per target and per model.

H.2 Part II: Full Per-Model Metrics, TopU-LBVS-full

Each pair of tables below reports all seven evaluation metrics (EF@1%, EF@5%, EF@10%, PR-AUC, ROC-AUC, BEDROC, LogAUC) for a single baseline model across all 93 targets under the ChEMBL*→TopU protocol. Results are split across two tables per model due to the number of targets; the split follows the same class ordering as the main paper (Cytochrome P450 through Protease).

Table 14: **TopU-LBVS-Full EF1% results - part 1.** Per-target EF@1% comparison across baseline models.

Target	Mor-RF	Tan-NN	GAT	GAT+FP	GIN	GIN-FP	GPS	GPS-FP	DMPNN	MolF
Cytochrome P450										
cp1a2	12.4	12.2	5.1	12.2	4.9	13.2	3.2	13.2	11.7	7.8
cp2c9	12.6	6.4	1.4	13.4	5.0	14.2	3.1	15.1	11.4	11.2
cp2d6	15.8	9.8	8.6	15.6	7.7	15.8	8.9	15.1	16.1	11.3
cp3a4	15.9	10.9	9.0	14.3	10.2	15.6	5.8	16.4	14.9	14.5
GPCR										
5ht6r	2.8	6.2	4.1	6.2	3.4	3.4	0.7	5.5	4.8	4.1
aa2ar	6.2	3.6	4.1	5.2	4.9	5.2	1.0	5.7	3.1	2.8
aa3r	4.0	7.1	3.6	4.4	2.7	5.2	3.8	4.4	6.3	4.9
adrb1	2.3	2.1	1.1	3.2	0.0	2.6	1.1	3.2	2.6	0.0
adrb2	5.2	5.2	2.3	2.9	3.3	2.3	1.0	3.9	6.8	3.9
cnr2	6.9	4.3	0.6	3.0	2.4	4.2	0.0	3.6	4.8	4.8
cxcr4	4.2	7.8	3.9	3.9	3.9	3.9	0.0	3.9	0.0	2.0
drd2	5.3	4.6	1.5	4.1	1.2	4.4	0.9	4.1	3.2	1.7
drd3	5.3	4.6	4.7	8.2	2.3	7.0	7.0	7.8	6.6	6.6
hrh3	9.1	4.6	5.1	8.1	7.1	9.6	3.0	8.6	9.6	7.1
oprd	6.0	5.7	2.5	5.4	2.5	7.0	2.2	4.8	5.4	5.4
oprk	4.2	4.5	1.6	2.7	3.1	3.5	2.3	3.5	3.5	3.9
oprm	6.2	5.7	1.7	4.9	4.1	5.2	3.2	4.1	5.8	7.3
Ion Channel										
5ht3a	8.1	9.8	4.6	4.6	2.6	4.6	3.9	3.3	2.0	10.4
cac1h	11.7	5.7	6.8	9.1	9.1	11.4	9.1	8.0	11.4	9.1
kcnh2	5.4	4.8	2.0	2.5	1.9	2.4	2.5	1.9	9.1	3.8
scn5a	4.1	9.1	2.3	2.3	3.0	3.0	0.0	3.0	0.0	6.1
trpa1	32.8	16.4	2.7	13.7	2.7	13.7	2.7	19.1	2.7	8.2
Kinase										
abl1	3.6	6.8	2.0	1.4	1.4	1.0	3.8	1.7	2.0	4.4
akt1	11.5	9.0	6.4	6.8	5.1	7.7	7.3	8.1	8.1	9.0
akt2	12.2	7.5	6.2	10.9	6.2	7.5	8.2	7.5	6.8	9.6
braf	8.7	8.5	0.0	9.1	2.3	6.3	2.3	6.8	5.7	3.4
cdk2	7.5	7.9	1.7	3.8	2.0	5.1	3.1	4.8	4.1	4.1
csf1r	3.0	7.2	3.6	9.4	5.0	7.9	4.3	8.6	10.1	7.9
egfr	6.4	5.6	2.0	2.0	2.8	3.9	2.6	3.9	7.4	3.5
fak1	10.4	10.4	6.5	11.5	9.4	10.8	5.8	13.7	12.9	7.2
fgfr1	10.0	8.1	5.6	8.1	6.0	9.0	5.1	9.8	9.0	7.7
igf1r	13.7	12.4	5.4	7.3	5.9	7.8	2.9	8.8	7.3	8.3
jak2	4.0	2.9	2.3	6.2	8.2	6.2	4.3	4.7	11.3	3.9
kit	7.0	5.0	0.0	6.2	2.5	3.1	1.9	3.7	6.2	7.5
kpcb	2.8	11.2	3.7	3.7	2.5	3.7	3.7	3.7	8.7	7.5
lck	8.2	11.6	2.5	6.3	2.1	5.6	3.2	4.6	4.2	4.2
mapk2	14.5	16.7	10.4	8.7	9.3	4.9	6.0	4.9	9.8	7.1
met	8.5	5.6	6.4	6.0	7.7	5.1	5.6	6.0	5.1	8.5
mk01	28.8	21.5	17.1	27.8	16.1	25.4	14.2	24.9	27.3	22.0
mk10	12.5	9.4	8.5	7.7	6.8	12.0	4.3	10.2	6.0	6.0
mk14	6.6	7.0	5.3	6.8	7.2	5.3	4.6	3.8	8.7	9.9
mp2k1	8.0	4.0	6.5	6.5	5.0	6.5	2.2	7.2	5.0	7.2
plk1	20.5	23.3	12.2	20.5	9.8	20.0	6.8	22.0	23.4	21.5
rock1	3.6	5.8	10.1	3.6	7.9	2.9	3.6	4.3	5.8	5.0
src	5.2	5.3	2.3	7.0	4.0	6.7	3.3	5.7	6.3	4.0
tgfr1	11.2	8.2	0.0	2.7	0.0	1.4	0.0	4.1	5.5	6.8
vgfr2	8.9	8.9	4.0	6.7	5.5	8.0	4.7	9.4	8.9	7.0
wee1	0.5	0.0	2.7	8.2	5.5	8.2	5.5	10.9	8.2	10.9

Table 15: **TopU-LBVS-Full EF1% results - part 2.** Per-target EF@1% comparison across baseline models.

Target	Mor-RF	Tan-NN	GAT	GAT+FP	GIN	GIN-FP	GPS	GPS-FP	DMPNN	MolF
Nuclear Receptor										
andr	12.2	12.7	1.5	4.9	0.5	4.4	2.4	3.9	4.4	4.9
esr1	6.2	8.0	2.4	4.2	2.1	3.3	2.4	3.3	5.9	2.7
esr2	8.0	8.3	2.4	4.8	2.4	3.6	2.4	5.9	4.2	2.4
gcr	10.1	7.1	5.9	8.9	5.3	10.7	3.0	5.9	8.3	8.3
mcr	7.0	6.2	0.0	8.2	0.0	0.0	2.7	2.7	2.7	8.2
ppara	6.5	4.9	0.0	0.0	3.2	0.0	0.8	1.6	0.8	4.8
ppard	6.3	9.9	2.7	4.6	6.4	2.7	4.6	5.5	3.6	4.6
pparg	6.7	5.3	2.3	3.6	2.7	4.6	0.5	4.1	13.2	3.2
prgr	10.1	4.6	5.3	3.6	5.9	6.5	6.5	7.1	6.5	6.5
rxra	15.1	15.2	0.0	7.6	1.5	3.0	0.0	4.6	6.1	10.6
thb	34.2	18.2	21.3	30.4	31.9	30.4	21.3	33.4	41.0	30.4
Other Enzymes										
aces	10.0	8.0	2.8	4.8	2.6	5.1	3.4	6.8	6.5	3.4
ampc	35.5	38.3	19.1	24.6	27.3	27.3	24.6	24.6	27.3	27.3
aofb	6.8	3.3	1.5	1.5	1.0	4.9	2.0	3.4	7.8	2.9
cah2	1.8	10.5	3.2	3.7	5.8	3.7	3.7	3.7	10.0	6.8
dhi1	9.9	6.3	0.0	0.0	0.0	2.3	0.0	4.6	0.0	2.3
dyr	7.1	6.8	0.0	5.1	0.0	3.4	0.0	1.7	0.0	3.4
fnta	9.2	5.0	5.5	5.5	0.0	8.2	0.0	2.7	8.2	8.2
glcm	14.0	3.7	7.6	15.2	13.7	15.2	7.6	15.2	19.7	15.2
hdac2	9.4	6.1	3.4	4.6	3.4	2.3	3.4	4.6	6.8	5.7
hdac8	4.6	2.3	3.4	0.0	3.4	0.0	2.3	0.0	3.4	1.1
hivint	6.2	6.2	0.7	0.7	2.0	0.7	0.0	0.0	2.0	1.4
hivrt	5.8	2.0	2.2	4.4	4.4	4.8	2.6	3.1	7.9	3.1
hmdh	15.0	8.2	0.0	16.4	5.5	16.4	2.7	8.2	13.7	5.5
hxk4	9.0	8.2	8.2	2.7	2.7	5.5	2.7	8.2	5.5	8.2
inha	15.9	13.7	0.0	6.8	0.0	6.8	0.0	6.8	13.7	9.1
nos1	5.0	1.3	0.0	12.0	6.8	5.1	3.4	6.8	10.2	8.5
parp1	11.5	4.6	0.0	3.0	0.0	10.6	0.0	6.1	10.6	4.6
pde5a	7.1	3.4	7.8	6.8	2.9	7.8	1.0	6.8	9.8	8.8
pgh1	9.1	6.3	3.2	3.2	2.6	3.7	2.1	3.2	4.2	2.1
pgh2	7.0	5.9	7.8	2.0	5.9	3.9	3.9	3.9	2.0	9.8
ptn1	8.2	6.1	3.0	7.6	4.0	5.6	2.5	5.6	5.6	4.6
pygm	18.4	16.4	2.7	10.9	0.0	8.2	2.7	8.2	5.5	2.7
pyrd	11.0	7.5	0.0	3.7	0.0	2.5	1.2	3.7	0.0	1.2
tysy	2.7	0.0	0.0	0.0	0.0	0.0	2.7	0.0	0.0	0.0
Protease										
ace	16.4	9.3	2.7	8.2	8.2	13.7	0.0	10.9	2.7	5.5
ada17	1.7	2.5	0.0	3.4	5.1	3.4	0.0	3.4	1.7	3.4
bace1	2.8	2.2	0.6	0.6	0.6	1.8	0.6	0.6	5.9	0.6
dpp4	7.7	5.0	0.0	3.2	1.1	5.3	1.1	5.3	8.4	4.2
fa10	10.3	5.7	1.5	0.8	1.5	2.3	0.8	0.8	2.3	5.3
hivpr	2.2	1.0	0.0	5.0	1.2	3.7	0.0	3.7	6.2	2.5
mmp13	3.6	1.7	0.0	0.0	2.0	0.0	2.0	0.0	3.9	2.0
thrb	10.1	8.3	2.2	3.5	2.2	3.5	2.5	1.9	4.5	5.7
try1	0.7	8.3	2.0	2.0	2.0	5.9	0.0	2.0	5.9	5.9
urok	11.2	0.0	2.0	3.9	2.0	5.9	0.0	7.8	2.0	9.8

Table 16: **TopU-LBVS-few: EF@10% results - part 1.** Rows are individual targets grouped by protein class; the Tier column indicates the split strategy (1: fewer than 50 TopU actives, 6:2:2 split; 2: at least 50 TopU actives, 7:1:2 split). Higher is better.

Target	Tier	Morgan-RF	Tanimoto-NN	GIN	GIN+FP	GAT	GAT+FP	GPS	GPS+FP	DMPNN	MolFormer
Cytochrome P450											
cp1a2	2	1.5	1.2	0.9	1.1	0.7	1.1	1.6	0.5	1.6	2.2
cp2c9	2	0.7	0.6	1.5	0.8	1.7	0.7	1.1	0.7	1.1	1.2
cp2d6	2	1.5	1.7	1.0	1.0	1.2	1.1	1.7	0.8	2.3	2.0
cp3a4	2	1.2	1.1	1.3	0.7	1.0	1.0	1.2	0.6	1.6	1.4
GPCR											
5ht6r	1	0.3	1.3	2.7	1.7	1.3	0.7	1.0	1.3	2.7	0.7
aa2ar	2	1.0	0.7	0.8	0.8	0.7	0.7	1.2	0.3	1.5	1.2
aa3r	2	1.4	0.0	1.2	1.0	1.0	0.7	0.7	1.0	0.8	1.2
adrb1	2	1.0	0.5	2.0	0.3	0.3	0.3	1.3	0.5	1.3	0.5
adrb2	2	1.0	1.3	1.1	0.9	0.9	0.6	1.4	0.6	2.2	1.7
cnr2	2	0.8	0.6	1.5	0.9	1.2	0.0	0.9	0.3	0.6	0.6
cxcr4	1	0.0	0.0	2.1	1.1	0.0	1.1	0.0	0.0	0.0	0.0
drd2	2	1.7	0.7	1.7	0.6	0.6	0.3	0.7	0.4	1.0	1.2
drd3	2	1.6	0.7	2.0	1.3	0.9	1.1	1.1	0.7	0.6	1.7
hrh3	2	1.5	0.5	1.0	1.3	1.5	0.8	0.3	1.0	1.0	1.3
oprd	2	0.9	0.3	0.6	0.0	1.4	0.8	1.3	1.1	0.8	1.1
oprk	2	0.5	0.0	1.2	0.4	1.6	1.4	1.0	0.6	1.0	1.6
oprm	2	1.6	0.3	1.0	1.0	0.9	1.4	1.6	1.4	1.3	0.9
Ion Channel											
5ht3a	1	0.1	1.0	0.7	1.0	1.7	1.0	1.3	1.3	1.3	1.7
cac1h	1	1.1	0.8	1.6	0.5	0.5	0.5	0.5	0.5	2.7	1.1
kcnh2	2	1.1	1.1	1.0	1.2	0.9	0.8	0.9	0.7	1.6	1.3
scn5a	1	0.6	0.0	0.4	0.4	1.2	0.4	2.5	0.8	0.4	1.2
trpa1	1	0.0	0.0	0.0	0.0	0.0	0.0	1.5	1.5	0.0	0.0
Kinase											
abl1	2	0.8	1.1	1.4	0.7	1.1	0.5	1.2	0.7	0.5	0.9
akt1	2	1.0	0.9	1.1	1.3	2.0	1.3	1.1	1.5	1.1	0.9
akt2	1	0.9	1.5	1.1	0.7	1.5	0.7	1.5	0.7	1.1	1.1
braf	2	1.1	0.3	0.0	0.6	1.8	0.3	1.2	0.3	1.2	0.9
cdk2	2	0.3	0.2	1.0	1.1	0.9	0.2	0.7	0.2	1.1	1.9
csf1r	1	0.5	0.7	1.1	0.7	0.7	0.4	1.8	0.7	1.5	1.5
egfr	2	0.6	0.2	1.1	0.9	1.1	0.6	1.3	1.0	1.2	1.1
fak1	1	1.0	0.4	1.1	1.1	0.7	0.7	1.1	0.7	1.8	1.1
fgfr1	2	0.8	0.4	0.7	0.7	1.1	0.4	1.5	0.7	0.7	0.9
igf1r	2	0.8	1.3	1.5	0.8	2.3	1.3	1.5	2.8	1.5	1.0
jak2	2	0.8	0.4	1.2	0.8	1.0	0.4	0.6	0.6	0.8	1.0
kit	2	0.8	0.9	0.0	0.3	0.3	0.3	0.3	0.3	1.5	0.3
kpcb	1	0.5	0.0	1.3	1.3	0.0	0.7	1.3	0.0	0.0	1.3
lck	2	0.9	0.7	1.8	0.9	1.2	0.9	1.6	0.5	0.5	0.5
mapk2	2	1.3	1.1	0.8	1.4	1.4	0.8	0.3	1.1	0.5	1.6
met	2	1.1	0.7	1.1	1.1	0.9	0.9	0.9	1.5	1.3	1.1
mk01	2	2.0	1.8	0.5	1.0	0.8	1.3	1.5	1.0	2.0	1.3
mk10	1	0.2	0.0	0.8	1.2	0.0	0.8	1.2	1.7	0.4	0.8
mk14	2	0.9	1.1	0.4	1.6	0.4	0.4	0.6	0.2	1.2	2.5
mp2k1	1	0.5	1.1	0.4	1.8	0.4	0.7	0.0	0.0	0.4	0.7
plk1	2	1.7	2.0	1.5	1.0	1.8	1.0	1.3	1.8	2.0	1.0
rock1	1	0.4	0.4	1.8	1.8	0.7	0.4	0.7	0.4	1.5	1.1
src	2	0.5	0.2	1.6	1.9	1.4	1.4	0.7	1.2	0.5	0.9
tgfr1	1	0.5	0.0	0.7	0.7	0.7	0.7	0.7	0.7	1.3	0.7
vgfr2	2	0.9	0.1	1.0	0.6	0.9	1.1	1.1	0.5	1.0	1.2
wee1	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.1	1.1

H.3 Part III: Full Per-Model Metrics, TopU-LBVS-few

Each pair of tables below reports all seven evaluation metrics for a single baseline model across all 93 targets under the TopU→TopU few-shot protocol. Targets are grouped by class; the Tier column indicates the split strategy (Tier 1: 6:2:2, fewer than 50 actives; Tier 2: 7:1:2, at least 50 actives).

Table 17: **TopU-LBVS-few: EF@10% results - part 2.** Rows are individual targets grouped by protein class; the Tier column indicates the split strategy (1: fewer than 50 TopU actives, 6:2:2 split; 2: at least 50 TopU actives, 7:1:2 split). Higher is better.

Target	Tier	Morgan-RF	Tanimoto-NN	GIN	GIN+FP	GAT	GAT+FP	GPS	GPS+FP	DMPNN	MolFormer
Nuclear Receptor											
andr	2	0.7	0.0	0.8	0.8	0.8	0.8	0.5	0.8	0.5	1.0
esr1	2	0.8	0.1	1.6	1.2	1.0	0.3	0.9	0.0	1.7	1.2
esr2	2	0.7	0.3	1.2	0.6	1.5	0.0	1.2	0.0	0.6	0.3
gcr	2	0.1	0.3	2.7	0.9	0.9	0.9	1.2	0.6	3.3	1.5
mcr	1	0.0	0.0	0.0	0.0	1.5	0.0	0.0	3.0	0.0	1.5
ppara	1	0.2	0.0	0.4	0.4	2.1	0.4	1.7	1.7	0.0	2.1
ppard	1	0.3	0.0	0.5	0.0	1.4	1.9	1.9	1.9	0.0	1.4
pparg	2	0.6	0.0	0.9	0.7	1.5	0.4	1.3	0.7	1.1	1.5
prgr	2	0.8	0.0	1.2	0.6	0.9	0.3	1.2	0.9	0.6	1.5
rxra	1	0.1	0.0	0.8	0.8	0.8	0.0	0.8	0.8	0.0	0.8
thb	1	2.0	4.8	0.8	3.2	1.6	1.6	2.4	0.8	4.8	0.8
Other Enzymes											
aces	2	0.6	0.1	1.0	0.7	1.0	0.4	1.3	0.9	1.0	0.9
ampc	1	0.7	0.0	2.1	3.2	2.1	3.2	3.2	0.0	1.1	5.3
aofb	2	0.1	0.5	0.5	1.0	0.3	0.3	0.5	0.8	0.0	0.0
cah2	2	0.7	0.0	0.5	1.5	1.3	1.0	1.0	0.8	0.3	1.3
dhi1	1	0.0	0.0	1.1	2.1	0.0	1.1	1.1	0.0	0.0	0.0
dyr	1	0.0	0.0	1.6	0.0	0.0	0.0	1.6	1.6	0.8	0.8
fnta	1	0.1	0.0	0.0	0.0	0.0	0.0	1.1	1.1	0.0	2.1
glcm	1	0.0	0.0	1.6	0.0	0.8	0.8	0.0	0.0	0.8	1.6
hdac2	1	1.3	2.0	1.3	0.0	0.7	0.0	0.7	0.7	0.0	0.7
hdac8	1	1.0	0.0	0.5	1.6	0.5	0.5	0.5	1.1	0.5	0.5
hivint	1	0.4	0.4	0.0	1.1	1.1	1.1	1.1	1.1	0.0	0.4
hivrt	2	0.4	0.0	1.8	1.5	1.3	1.1	2.0	0.9	0.9	0.4
hmdh	1	0.1	1.5	1.5	1.5	0.0	1.5	0.0	0.0	1.5	1.5
hxk4	1	0.0	1.5	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0
inha	1	0.0	2.1	1.1	0.0	0.0	1.1	1.1	0.0	0.0	3.2
nos1	1	0.1	0.8	0.0	0.8	0.8	1.6	2.4	3.2	2.4	0.8
parp1	1	0.0	0.0	0.0	0.8	1.6	0.8	1.6	1.6	0.0	0.8
pde5a	1	0.6	0.2	0.9	0.0	1.4	0.5	0.9	0.5	0.9	0.0
pgh1	2	0.6	0.0	0.5	0.8	1.0	1.8	1.5	0.8	0.5	1.3
pgh2	1	0.1	0.0	1.1	1.1	2.1	2.1	1.1	2.1	1.1	1.1
ptn1	2	0.3	0.0	1.0	0.8	1.3	1.0	1.3	1.0	0.0	0.8
pygm	1	0.0	0.0	0.0	0.0	1.5	0.0	0.0	1.5	0.0	0.0
pyrd	1	0.0	0.0	2.0	0.7	1.3	0.0	1.3	0.7	0.7	0.0
tysy	1	0.0	0.0	1.5	1.5	1.5	1.5	1.5	1.5	0.0	6.1
Protease											
ace	1	0.0	0.0	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ada17	1	0.0	0.0	1.1	0.0	3.2	1.1	1.1	0.0	0.0	3.2
bace1	2	0.9	0.3	0.9	0.3	0.6	0.3	0.6	0.6	2.1	1.2
dpp4	1	0.0	0.0	1.6	0.0	0.0	1.1	0.5	0.0	1.6	0.5
fa10	1	0.6	0.0	1.7	1.7	0.4	1.7	1.2	0.8	1.2	0.4
hivpr	1	0.1	0.0	0.7	0.7	2.0	2.0	1.3	1.3	0.7	0.0
mmp13	1	0.2	0.0	1.6	0.8	0.0	0.0	0.0	0.8	0.0	1.6
thrb	2	0.9	0.2	0.8	0.8	0.5	0.2	1.3	0.2	1.1	0.6
try1	1	0.5	0.0	1.1	0.0	0.0	1.1	0.0	2.1	2.1	0.0
urok	1	0.5	0.0	0.0	0.0	0.8	0.0	1.6	1.6	0.8	0.0

Table 18: **TopU-LBVS-mini: full metric results on TopU hard-decoy test sets.** Each block reports one metric; rows are the 7 mini targets, columns are models. PR-AUC, ROC-AUC, BEDROC, and LogAUC are on a 0–100 scale. Higher is better for all metrics.

Target	Morgan-RF	Tanimoto-NN	GAT	GAT+FP	GIN	GIN+FP	GPS	GPS+FP	DMPNN	MolFormer
EF@1%										
aa2ar	6.2	3.6	4.1	5.2	2.8	6.2	1.0	5.7	3.1	2.8
aces	10.0	8.0	2.8	4.8	2.6	5.1	3.4	6.8	6.5	3.4
cp3a4	15.9	10.9	9.0	14.3	10.3	15.6	5.8	16.4	14.9	14.5
egfr	6.4	5.6	2.0	2.0	2.8	3.9	2.6	3.9	7.4	3.5
esr1	6.2	8.0	2.4	4.2	2.1	3.3	2.4	3.3	5.9	2.7
kcnh2	5.4	4.8	2.0	2.5	1.5	2.4	2.5	1.9	9.1	3.8
thrb	10.1	8.3	2.2	3.5	2.2	3.5	2.5	1.9	4.5	5.7
Avg.	**8.1**	**6.9**	**3.4**	**5.4**	**3.1**	**5.7**	**3.3**	**5.5**	**7.5**	**4.7**
EF@5%										
aa2ar	4.0	4.0	2.8	3.5	2.8	3.8	1.3	4.0	3.4	3.5
aces	6.4	5.6	2.9	4.4	3.1	4.4	2.9	4.3	3.4	4.2
cp3a4	9.5	7.8	4.8	8.4	6.2	9.0	4.5	8.9	9.5	9.0
egfr	5.8	4.5	2.8	4.0	3.9	4.0	2.7	4.1	4.8	4.5
esr1	5.2	4.8	2.7	4.4	2.3	3.9	2.0	4.1	4.5	3.6
kcnh2	3.7	3.3	2.1	3.0	1.9	2.7	2.2	2.4	3.3	2.9
thrb	4.8	5.2	2.8	3.7	2.8	3.6	2.5	3.7	4.0	4.2
Avg.	**5.3**	**5.2**	**3.2**	**4.5**	**3.3**	**4.5**	**2.8**	**4.4**	**4.6**	**4.3**
EF@10%										
aa2ar	3.4	3.0	2.5	3.4	2.3	3.5	2.0	3.5	3.3	3.2
aces	4.3	3.6	3.4	3.8	3.1	3.9	2.6	4.0	3.7	3.7
cp3a4	6.3	5.8	4.2	5.8	4.7	6.2	3.7	6.2	6.0	6.1
egfr	4.4	4.5	3.2	4.0	3.7	3.9	2.7	3.9	4.2	4.2
esr1	4.0	4.7	2.9	3.8	2.8	3.6	2.1	3.7	3.3	3.3
kcnh2	3.0	3.1	2.0	2.7	2.1	2.6	1.9	2.5	2.5	2.6
thrb	3.5	3.7	2.5	3.1	2.9	3.5	2.2	2.9	2.8	3.4
Avg.	**4.0**	**4.1**	**3.1**	**3.8**	**3.1**	**3.8**	**2.5**	**3.7**	**3.7**	**3.7**
PR-AUC										
aa2ar	7.9	6.5	6.0	7.2	5.4	7.7	4.2	8.3	6.8	6.5
aces	12.0	10.1	6.6	8.3	6.4	8.7	6.1	9.2	9.1	7.7
cp3a4	24.5	15.0	11.2	21.7	13.5	22.8	8.9	24.0	22.4	21.0
egfr	10.4	9.3	6.3	8.2	7.2	8.2	6.1	8.3	9.2	8.4
esr1	8.8	10.2	5.9	7.3	5.7	7.1	5.3	7.0	8.1	6.6
kcnh2	6.9	6.9	4.3	5.5	4.4	5.5	4.4	5.3	5.3	5.9
thrb	10.0	9.5	5.6	7.3	5.9	7.5	5.4	6.7	7.0	7.8
Avg.	**10.3**	**9.8**	**6.9**	**9.4**	**6.8**	**9.6**	**6.1**	**9.1**	**9.4**	**8.4**
ROC-AUC										
aa2ar	75.3	73.2	72.4	74.9	71.2	74.9	68.3	74.7	74.4	74.2
aces	80.4	80.9	76.7	78.7	76.7	78.7	74.6	79.6	80.9	78.4
cp3a4	85.6	82.3	78.9	81.9	80.6	82.6	77.8	82.5	85.6	84.3
egfr	79.3	80.6	73.6	79.6	76.7	80.3	73.7	79.8	79.8	78.1
esr1	72.2	77.3	73.1	72.7	72.9	72.2	70.1	71.0	76.3	74.9
kcnh2	74.8	77.1	67.0	73.0	69.0	72.3	67.5	73.0	73.5	72.9
thrb	77.0	79.4	74.6	78.4	75.0	78.4	72.6	77.3	76.8	77.2
Avg.	**77.6**	**78.6**	**74.1**	**77.0**	**74.8**	**77.0**	**72.1**	**76.7**	**78.0**	**77.2**
BEDROC										
aa2ar	18.7	15.5	11.2	16.3	10.9	18.5	4.6	18.2	15.0	14.1
aces	28.4	23.1	13.7	19.5	13.6	20.5	11.5	20.8	18.7	17.4
cp3a4	45.4	36.4	23.4	40.8	29.3	43.2	19.9	43.2	43.7	42.5
egfr	24.9	22.2	12.0	17.7	16.2	18.5	11.0	19.1	22.6	19.5
esr1	22.4	25.0	11.5	18.5	9.8	17.5	7.1	17.7	19.3	14.9
kcnh2	15.5	15.0	6.9	11.1	6.3	10.7	7.1	9.5	15.1	11.9
thrb	22.8	22.2	10.6	15.6	12.3	16.5	9.3	14.4	13.4	17.6
Avg.	**24.1**	**23.2**	**13.7**	**20.0**	**13.9**	**20.6**	**11.0**	**19.7**	**20.7**	**18.6**
LogAUC										
aa2ar	30.2	27.6	26.0	29.1	25.3	29.9	21.9	29.9	27.9	27.8
aces	36.3	33.5	28.3	31.6	28.1	32.1	26.8	32.8	30.6	30.5
cp3a4	46.1	39.7	33.7	42.8	36.8	44.0	31.5	44.3	44.9	43.7
egfr	34.1	33.0	26.6	31.1	29.3	31.7	26.4	31.8	32.0	31.5
esr1	31.1	30.2	26.3	28.9	25.6	28.4	23.7	28.2	29.9	28.0
kcnh2	28.8	28.7	22.2	26.0	22.7	25.7	22.5	25.5	23.4	26.5
thrb	33.0	32.8	26.1	29.7	26.9	30.1	25.2	28.6	27.2	30.1
Avg.	**33.5**	**32.3**	**27.5**	**31.3**	**27.8**	**31.6**	**25.8**	**31.2**	**30.6**	**30.7**

Table 19: **TopU-LBVS-mini: full metric results on random-decoy test sets.** Each block reports one metric; rows are the 7 mini targets, columns are models. PR-AUC, ROC-AUC, BEDROC, and LogAUC are on a 0–100 scale. Higher is better for all metrics.

Target	Morgan-RF	Tanimoto-NN	GAT	GAT+FP	GIN	GIN+FP	GPS	GPS+FP	DMPNN	MolFormer
					EF@1%					
aa2ar	27.2	18.6	13.4	29.7	18.3	29.9	9.3	30.2	28.1	25.0
aces	28.8	18.7	16.5	32.2	16.2	30.2	13.1	30.5	30.5	27.3
cp3a4	16.2	8.5	9.4	18.2	9.8	16.2	6.6	16.9	15.8	13.5
egfr	32.1	17.4	16.9	31.7	22.6	32.8	15.8	30.8	28.0	28.4
esr1	24.4	17.5	15.4	23.5	18.1	25.0	10.1	22.3	24.7	19.6
kcnh2	19.2	12.0	10.2	20.5	10.2	20.7	9.4	21.1	20.1	19.1
thrb	30.6	22.6	18.1	31.8	19.1	31.5	15.9	31.8	29.6	27.7
Avg.	**25.4**	**16.7**	**14.1**	**26.3**	**16.0**	**26.4**	**11.3**	**26.0**	**25.0**	**22.5**
					EF@5%					
aa2ar	14.6	12.7	10.3	14.9	11.9	14.7	9.2	14.8	14.0	13.8
aces	15.8	15.2	12.1	15.3	12.6	15.3	10.8	15.4	14.5	14.4
cp3a4	10.6	9.8	5.6	9.7	5.8	8.9	8.8	8.8	8.2	7.0
egfr	15.1	14.8	11.2	15.4	13.1	15.0	10.8	15.4	13.7	14.1
esr1	11.7	12.2	9.2	11.8	10.7	12.0	8.9	12.1	11.9	10.9
kcnh2	12.2	10.3	7.9	11.6	7.9	11.8	7.2	11.7	11.4	10.9
thrb	15.5	14.2	12.9	16.1	13.3	15.8	12.3	15.8	15.1	14.2
Avg.	**13.5**	**12.7**	**9.9**	**13.4**	**10.8**	**13.4**	**9.2**	**13.4**	**12.7**	**12.2**
					EF@10%					
aa2ar	8.3	8.3	7.6	8.4	8.3	8.3	6.9	8.2	8.3	8.1
aces	8.8	8.9	8.0	8.7	8.5	8.7	7.9	8.7	8.6	8.1
cp3a4	5.8	6.4	4.0	5.9	4.6	5.8	3.9	5.8	5.6	5.1
egfr	8.5	8.6	6.9	8.4	7.7	8.5	7.4	8.4	7.7	7.7
esr1	6.8	7.4	6.4	6.5	6.7	6.5	6.4	6.7	6.6	6.4
kcnh2	7.7	7.9	6.4	7.6	6.4	7.5	5.9	7.5	7.4	7.3
thrb	9.0	9.0	8.6	8.9	8.6	8.7	8.0	8.9	8.9	8.3
Avg.	**7.9**	**8.0**	**6.8**	**7.8**	**7.3**	**7.7**	**6.7**	**7.7**	**7.6**	**7.3**
					PR-AUC					
aa2ar	48.8	31.6	24.3	52.8	31.7	53.3	19.8	52.8	46.8	43.1
aces	53.6	39.3	29.8	58.4	31.9	55.5	26.7	54.8	54.2	47.1
cp3a4	22.4	16.9	10.5	19.7	13.0	21.0	8.4	21.3	20.0	14.7
egfr	56.5	38.3	28.5	58.3	37.8	59.2	26.2	58.5	48.5	49.8
esr1	37.9	30.6	25.2	35.1	29.1	36.0	18.0	36.2	37.6	33.2
kcnh2	34.2	22.9	17.1	33.9	17.1	34.1	16.1	35.2	30.8	30.0
thrb	58.2	42.3	35.6	58.7	36.3	58.6	30.2	59.3	54.7	49.8
Avg.	**44.5**	**31.7**	**24.4**	**45.3**	**28.1**	**45.4**	**20.8**	**45.4**	**41.8**	**38.2**
					ROC-AUC					
aa2ar	91.9	90.6	91.2	93.0	92.3	93.0	89.5	92.9	91.7	92.2
aces	94.5	93.1	94.2	94.8	95.2	94.0	93.3	94.3	95.4	93.4
cp3a4	83.5	82.6	79.6	82.4	82.2	82.4	77.1	82.1	83.1	81.3
egfr	93.5	93.2	90.2	93.8	92.0	93.8	89.8	93.8	91.7	91.0
esr1	85.5	85.5	86.1	84.7	86.7	84.2	84.9	84.9	86.2	85.3
kcnh2	91.9	90.5	89.5	91.3	89.5	90.9	88.4	91.0	90.4	90.6
thrb	94.1	94.0	94.4	94.9	94.1	94.6	92.0	95.3	93.8	94.8
Avg.	**90.7**	**89.9**	**89.3**	**90.7**	**90.3**	**90.4**	**87.9**	**90.6**	**90.3**	**89.8**
					BEDROC					
aa2ar	69.0	58.6	49.0	70.8	57.6	70.7	43.8	70.4	66.6	65.9
aces	73.5	66.8	56.5	75.2	59.3	73.7	52.7	73.6	71.4	68.2
cp3a4	43.4	40.5	25.0	40.9	28.0	41.7	21.1	41.4	38.9	32.0
egfr	73.4	65.9	51.8	74.4	60.7	74.6	50.5	74.5	66.3	67.9
esr1	57.4	55.0	45.7	55.9	50.2	56.8	40.4	56.9	57.0	51.7
kcnh2	57.6	48.6	38.2	56.2	38.2	56.3	35.9	56.6	54.4	52.5
thrb	75.4	67.4	61.3	76.9	62.2	76.1	56.3	76.3	73.2	68.8
Avg.	**64.2**	**57.5**	**46.8**	**64.3**	**50.9**	**64.3**	**43.0**	**64.2**	**61.1**	**58.1**
					LogAUC					
aa2ar	62.6	53.6	48.9	64.6	53.9	64.8	45.6	64.7	61.6	60.2
aces	67.1	59.3	54.2	70.3	55.9	67.8	51.5	67.9	67.5	63.3
cp3a4	44.3	41.1	34.1	42.6	36.5	43.4	31.3	43.1	42.4	38.0
egfr	67.5	58.2	50.4	68.4	56.4	68.8	49.5	68.2	61.9	62.8
esr1	55.0	51.4	47.5	53.3	49.8	53.7	43.2	53.9	55.0	50.8
kcnh2	54.5	47.9	43.3	54.0	43.3	54.0	42.0	54.5	52.5	51.9
thrb	70.2	62.0	58.2	71.5	58.8	71.3	54.2	71.6	68.0	65.4
Avg.	**60.2**	**53.4**	**48.1**	**60.7**	**50.7**	**60.5**	**45.3**	**60.6**	**58.4**	**56.1**

Table 20: **TopU-LBVS-Full Morgan-RF results.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	12.7±1.2	8.0±0.3	5.3±0.1	17.5±0.3	81.4±0.4	36.7±1.0	40.2±0.3
cp2c9	12.0±0.4	7.6±0.3	5.4±0.1	16.6±0.7	84.3±0.2	36.4±0.9	41.0±0.4
cp2d6	16.3±0.3	9.5±0.2	6.2±0.2	23.4±0.9	84.6±0.2	44.3±0.4	45.2±0.4
cp3a4	15.9±0.8	9.5±0.1	6.3±0.2	24.5±0.3	85.6±0.4	45.4±0.4	46.1±0.4
GPCR							
aa2ar	6.2±0.4	4.0±0.1	3.5±0.1	8.0±0.1	75.2±0.3	19.0±0.2	30.3±0.4
aa3r	5.2±0.4	3.8±0.2	3.3±0.3	7.7±0.2	76.2±0.2	16.9±0.3	29.9±0.2
adrb1	2.1±0.7	2.7±0.2	2.2±0.1	5.3±0.0	72.1±0.9	8.8±0.2	25.0±0.2
adrb2	6.2±0.5	3.5±0.2	2.4±0.1	6.5±0.1	70.7±0.3	14.5±0.4	27.1±0.1
cnr2	7.1±1.5	4.0±0.0	3.6±0.3	9.2±0.1	79.1±0.6	19.6±0.8	31.7±0.4
cxcr4	3.9±2.8	3.1±0.6	2.0±0.0	5.8±0.5	67.7±1.2	10.2±2.3	24.4±1.3
drd2	5.8±0.4	4.5±0.5	3.5±0.1	8.3±0.2	77.0±0.3	19.5±0.3	30.9±0.2
drd3	4.7±0.0	4.1±0.7	3.9±0.2	10.1±0.4	81.0±0.4	20.1±1.6	33.0±0.6
hrh3	8.6±0.7	5.5±0.0	3.9±0.1	13.3±0.6	82.0±0.2	27.8±0.4	36.7±0.3
oprd	6.4±1.2	5.5±0.2	4.4±0.0	9.5±0.4	79.3±0.1	23.7±0.9	33.1±0.4
oprk	3.5±1.0	4.4±0.3	4.0±0.2	7.7±0.2	75.9±0.4	18.5±0.6	30.0±0.3
oprm	6.4±0.4	5.8±0.1	4.9±0.1	11.0±0.1	80.8±0.1	26.9±0.4	35.4±0.1
5ht6r	2.7±1.0	6.5±0.2	4.5±0.1	9.7±0.3	77.4±0.5	25.0±0.8	32.5±0.5
Ion Channel							
5ht3a	7.2±0.9	6.9±0.3	5.8±0.1	14.5±0.7	82.2±0.8	34.1±0.5	39.1±0.7
cac1h	13.7±2.8	11.7±0.0	8.0±0.2	25.0±1.0	89.0±0.8	51.6±1.6	48.3±0.5
kcnh2	5.4±0.5	3.6±0.1	3.0±0.0	7.0±0.2	74.9±0.2	15.6±0.5	28.9±0.2
scn5a	4.6±1.9	2.8±0.4	3.6±0.1	9.1±0.8	70.2±0.4	14.1±0.6	27.1±0.3
trpa1	32.8±0.0	9.8±0.0	5.0±0.0	39.0±3.9	71.6±0.2	50.6±0.3	47.1±0.1
Kinase							
abl1	4.1±0.8	5.7±0.3	4.3±0.0	8.6±0.1	73.9±0.3	23.1±0.5	30.9±0.2
akt1	12.0±0.6	7.3±0.2	5.2±0.1	15.7±0.7	79.6±0.1	34.2±0.1	38.6±0.2
akt2	13.0±2.6	8.0±0.2	6.1±0.1	18.5±0.2	83.3±0.2	38.1±0.1	41.9±0.2
braf	8.0±0.8	6.2±0.0	4.4±0.2	12.3±0.9	84.0±0.4	27.4±0.8	36.9±0.7
cdk2	7.2±0.8	5.4±0.3	4.9±0.1	11.9±0.5	79.2±0.3	26.8±0.1	35.4±0.2
csf1r	4.3±0.0	2.7±0.2	3.0±0.3	7.7±0.1	79.1±0.5	14.2±0.5	30.3±0.2
egfr	6.4±1.2	5.7±0.1	4.5±0.0	10.8±0.3	79.5±0.3	25.0±0.2	33.8±0.1
fak1	10.8±0.0	5.9±0.0	4.9±0.1	14.1±0.5	76.0±0.4	28.8±0.9	35.4±0.5
fgfr1	10.2±1.0	7.1±0.3	4.5±0.1	14.4±0.7	77.8±0.3	33.0±1.3	37.0±0.6
igf1r	14.2±0.7	6.8±0.4	5.2±0.1	18.3±1.3	81.7±0.7	34.2±0.2	40.4±0.2
jak2	5.1±1.1	6.0±0.1	4.8±0.3	9.9±0.1	78.7±0.8	25.3±0.2	33.5±0.3
kit	6.2±2.3	5.4±0.2	4.3±0.3	10.0±0.0	76.6±0.1	24.2±0.9	32.6±0.2
kpcb	3.7±3.0	6.1±0.0	4.3±0.2	9.7±0.4	78.0±0.6	23.4±0.5	32.5±0.7
lck	8.4±0.9	6.2±0.2	4.5±0.2	10.9±0.7	72.8±0.3	26.6±0.7	32.7±0.5
mapk2	13.7±0.8	7.3±0.4	5.3±0.1	20.2±0.7	75.9±0.2	35.7±0.6	38.7±0.1
met	8.5±0.6	5.7±0.3	4.7±0.2	13.1±0.8	75.6±0.4	28.2±0.3	34.5±0.1
mk01	28.3±0.7	12.6±0.6	7.0±0.1	40.3±1.4	84.6±0.6	60.1±1.2	56.5±0.9
mk10	13.7±2.4	5.8±0.2	4.5±0.1	14.8±0.6	76.8±0.9	29.1±0.6	36.4±0.4
mk14	6.8±0.9	7.1±0.3	5.3±0.2	12.4±0.3	79.0±0.5	31.3±0.6	36.4±0.4
mp2k1	8.6±0.0	4.2±0.2	4.0±0.3	9.4±0.4	70.8±1.2	22.0±1.1	30.4±0.8
plk1	20.5±0.0	12.1±0.5	6.8±0.1	34.5±0.3	86.7±0.9	55.1±0.2	51.5±0.5
rock1	5.0±1.0	4.2±0.4	4.3±0.4	8.6±0.2	72.4±0.4	19.1±0.9	29.5±0.4
src	5.7±0.5	4.5±0.2	4.0±0.1	9.1±0.1	78.7±0.6	20.1±0.3	32.1±0.1
tgfr1	17.8±1.9	6.8±0.4	5.0±0.4	26.0±2.3	84.7±0.0	36.6±1.3	43.0±0.9
vgfr2	9.0±1.2	6.0±0.4	4.7±0.1	12.3±0.4	81.2±0.3	28.1±1.1	36.2±0.7
wee1	0.0±0.0	3.3±0.0	4.9±0.0	9.3±0.3	75.7±2.8	20.0±0.5	30.3±0.9

Table 21: **TopU-LBVS-Full Morgan-RF results.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	11.7±0.0	5.6±0.4	4.4±0.1	13.7±1.0	82.7±0.8	28.0±1.9	37.4±1.1
esr1	6.2±0.0	5.4±0.2	4.0±0.1	9.3±0.3	72.4±0.3	22.5±0.6	30.7±0.3
esr2	7.1±0.0	5.0±0.6	4.2±0.2	13.3±1.0	75.4±0.5	23.7±2.1	32.8±1.0
gcr	8.3±0.8	8.2±0.5	6.1±0.2	15.0±0.2	84.8±0.2	36.1±0.7	40.1±0.1
mcr	8.2±0.0	3.2±1.8	3.0±0.8	12.6±4.2	76.7±1.5	17.4±5.6	29.8±2.4
ppara	7.2±2.0	5.8±0.2	3.2±0.2	9.5±0.9	78.4±0.7	21.9±1.9	32.1±1.1
ppard	7.3±1.3	6.1±0.3	4.0±0.0	12.2±0.9	78.7±0.3	27.5±1.3	35.2±0.8
pparg	7.7±0.6	5.0±0.3	4.0±0.1	10.2±0.5	80.1±0.3	23.4±0.9	33.9±0.5
prgr	9.5±1.7	5.6±0.4	4.7±0.1	12.9±1.0	82.8±0.5	28.3±2.0	37.0±0.9
rxra	16.7±2.1	7.7±0.5	4.2±0.2	21.4±2.5	85.6±0.6	37.2±0.5	42.6±0.6
thb	34.9±2.1	17.0±0.0	9.3±0.2	70.7±0.8	97.0±0.3	81.9±0.1	74.1±1.0
Other Enzymes							
aces	10.0±0.4	6.5±0.1	4.4±0.2	12.4±0.7	80.5±0.5	28.6±1.0	36.2±0.5
ada17	0.0±0.0	2.2±0.9	4.2±0.3	8.5±0.4	84.4±0.3	14.5±2.1	31.1±0.8
aofb	9.8±0.7	7.0±0.0	4.8±0.2	13.2±0.4	80.1±0.3	29.7±0.5	36.3±0.1
ampc	38.3±3.9	12.0±2.8	7.4±0.7	55.5±10.3	92.2±2.7	62.7±10.4	58.3±6.4
cah2	2.1±0.7	3.7±0.6	3.7±0.1	7.8±0.4	80.6±0.9	17.0±1.0	30.7±0.6
dhi1	9.1±3.2	4.2±0.0	3.5±0.0	9.1±0.2	73.3±0.9	19.7±0.5	28.6±0.4
dyr	10.2±0.0	5.6±0.5	3.9±0.2	12.7±2.1	79.5±0.3	24.8±0.5	34.2±0.9
fnta	8.2±0.0	6.6±0.0	3.8±0.4	18.9±0.8	76.2±0.4	28.4±2.1	34.6±0.7
glcm	12.2±2.1	9.9±0.4	7.4±0.2	24.7±1.1	83.9±0.8	50.3±0.7	46.7±0.5
hdac2	8.0±1.6	3.8±0.3	3.0±0.3	10.2±0.1	81.0±1.0	19.5±0.8	32.8±0.2
hdac8	6.8±2.8	2.5±0.3	2.5±0.3	7.2±1.2	76.2±0.3	11.8±1.2	27.5±0.7
hivint	4.1±1.7	4.4±0.5	4.3±0.1	9.0±0.6	78.8±0.8	20.2±1.6	31.9±0.9
hivrt	5.7±1.6	3.9±0.1	3.5±0.1	7.9±0.2	77.6±0.3	18.3±0.3	30.6±0.3
hmdh	16.4±0.0	5.9±0.8	4.2±0.4	17.5±0.6	84.6±0.8	31.3±1.8	38.7±0.3
hxk4	16.4±0.0	5.9±1.6	5.7±0.5	21.7±3.9	89.7±1.1	35.7±5.0	42.6±2.2
inha	15.9±3.2	7.1±0.0	4.5±0.3	23.5±1.9	89.3±0.5	33.4±1.3	42.5±1.1
nos1	5.1±0.0	4.9±0.5	5.6±0.2	13.9±2.4	84.8±0.5	27.4±1.3	36.9±0.8
parp1	16.7±2.1	6.3±0.5	4.0±0.4	16.0±1.5	77.6±0.2	30.1±1.7	36.5±0.9
pde5a	7.8±1.4	5.5±0.3	4.6±0.3	13.1±1.4	85.5±1.1	26.9±1.0	37.1±0.4
pgh1	9.5±1.3	5.8±0.3	4.0±0.1	11.0±0.8	75.6±0.1	25.1±0.4	32.9±0.4
pgh2	5.9±0.0	5.7±0.6	4.6±0.5	11.6±0.4	77.5±0.9	25.2±1.3	33.7±0.4
ptn1	8.6±1.4	5.7±0.3	4.1±0.1	10.6±0.7	75.8±0.5	24.8±1.0	32.9±0.6
pygm	13.7±10.2	7.8±0.0	5.0±0.0	20.0±3.3	74.3±1.3	36.6±2.6	38.0±2.1
pyrd	12.4±1.8	4.5±0.7	4.5±0.2	17.7±2.6	82.6±0.8	26.7±1.1	38.0±0.1
tysy	0.0±0.0	2.6±0.9	1.3±0.5	5.8±0.2	69.0±0.2	7.6±1.9	23.4±0.4
Protease							
ace	16.4±0.0	7.8±0.0	4.7±0.5	31.1±0.7	81.2±0.4	37.7±1.0	41.5±0.5
bace1	3.0±0.8	2.3±0.2	3.2±0.2	6.8±0.1	78.4±0.0	12.5±0.3	28.4±0.2
dpp4	6.3±2.6	7.9±0.3	5.7±0.2	16.7±1.1	87.8±1.2	37.7±1.5	41.6±1.2
fa10	10.6±1.1	5.4±0.2	4.4±0.2	14.3±0.8	86.0±0.3	26.6±0.8	37.8±0.2
hivpr	2.5±1.8	3.4±0.7	4.4±0.3	7.5±0.0	79.8±0.7	15.2±0.4	29.3±0.3
mmp13	2.0±2.8	2.7±0.5	3.1±0.3	6.0±0.8	69.3±1.1	10.7±1.3	24.4±0.8
thrb	10.1±0.9	4.8±0.3	3.5±0.0	11.6±0.3	77.3±0.4	23.3±0.5	33.1±0.2
try1	2.0±2.8	2.5±0.0	4.4±0.5	7.6±0.2	77.2±0.7	12.7±0.8	28.4±0.2
urok	9.8±2.8	8.6±0.5	5.9±0.0	22.0±1.9	88.5±0.5	42.5±1.2	44.5±1.2

Table 22: **TopU-LBVS-Full Tanimoto-NN results.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	12.2±1.5	8.5±0.3	5.1±0.1	15.6±1.1	75.4±1.0	37.0±1.4	37.8±1.1
cp2c9	6.4±1.0	7.7±0.3	6.1±0.1	14.1±0.5	82.8±0.6	35.4±1.0	39.0±0.5
cp2d6	9.8±0.3	8.9±0.1	6.4±0.1	16.2±0.2	81.9±0.2	39.3±0.3	40.5±0.2
cp3a4	10.9±0.0	7.8±0.2	5.8±0.1	15.0±0.5	82.3±0.3	36.4±0.9	39.7±0.5
GPCR							
5ht6r	6.2±0.0	5.6±0.4	5.0±0.3	11.8±0.5	78.0±0.4	27.5±1.1	34.8±0.6
aa2ar	3.6±0.4	4.1±0.0	3.0±0.1	6.5±0.1	73.2±0.4	15.5±0.3	27.6±0.2
aa3r	7.1±0.4	5.3±0.1	4.1±0.1	9.7±0.2	77.9±0.1	23.7±0.3	32.9±0.2
adrb1	2.1±1.5	2.1±0.4	3.1±0.1	6.5±0.3	77.6±0.3	12.1±2.0	28.0±0.7
adrb2	5.2±0.5	5.4±0.2	4.1±0.1	9.0±0.2	76.5±0.4	23.3±0.5	31.5±0.3
cnr2	4.2±2.2	3.4±0.2	4.0±0.1	7.0±0.1	77.0±0.7	16.1±0.8	27.9±0.1
cxcr4	7.8±5.5	6.2±0.6	4.6±0.0	11.1±0.5	71.8±0.6	26.4±0.7	31.0±0.0
drd2	4.7±1.5	3.2±0.1	3.0±0.1	7.1±0.1	77.9±0.1	14.9±0.4	28.5±0.1
drd3	4.7±0.0	5.1±0.1	4.8±0.3	10.2±0.3	84.2±0.2	23.8±1.1	34.6±0.4
hrh3	4.6±0.0	4.9±0.2	5.5±0.1	10.4±0.2	82.3±0.4	24.7±0.4	34.5±0.3
oprd	2.5±0.9	3.9±0.2	3.9±0.2	8.6±0.2	79.9±0.4	19.3±0.4	30.9±0.5
oprk	7.4±1.5	4.0±0.3	4.1±0.2	7.9±0.4	77.0±0.7	20.7±1.8	29.3±0.7
oprm	7.3±1.1	4.1±0.1	4.6±0.1	9.2±0.1	79.5±0.2	22.8±0.4	31.5±0.1
Ion Channel							
5ht3a	9.8±0.0	6.6±0.5	6.3±0.0	16.4±0.4	84.2±1.0	35.0±1.0	40.4±0.6
cac1h	5.7±1.6	10.0±0.3	7.5±0.2	18.8±0.3	92.3±0.8	41.7±0.6	44.7±0.2
kcnh2	4.8±0.6	3.3±0.1	3.1±0.0	6.9±0.2	77.1±0.2	15.0±0.4	28.7±0.2
scn5a	9.1±0.0	4.3±0.2	3.2±0.2	7.8±0.1	72.3±0.9	18.0±0.2	28.5±0.3
trpa1	16.4±0.0	6.5±0.9	4.0±0.0	14.0±2.5	68.6±0.8	28.2±4.6	31.2±1.5
Kinase							
abl1	6.8±0.5	6.2±0.1	4.4±0.0	9.3±0.3	75.7±0.4	24.5±0.7	32.0±0.4
akt1	9.0±0.0	7.2±0.2	5.3±0.0	14.0±0.2	78.1±0.3	32.9±0.2	37.1±0.2
akt2	7.5±1.0	8.2±0.4	6.0±0.0	15.1±0.5	82.7±0.4	36.5±1.2	39.6±0.3
braf	8.5±0.0	5.4±0.2	4.0±0.2	12.0±0.6	83.7±0.8	23.1±1.5	35.4±0.8
cdk2	7.9±0.5	5.2±0.3	4.5±0.1	10.4±0.4	78.8±0.3	25.4±1.2	33.9±0.4
csf1r	7.2±1.0	2.4±0.2	3.4±0.1	8.4±0.1	79.0±0.4	14.9±0.5	30.9±0.1
egfr	5.6±0.6	4.4±0.2	4.5±0.1	9.3±0.2	80.6±0.5	22.2±0.6	33.0±0.3
fak1	10.8±1.8	6.0±0.2	4.2±0.2	14.0±2.2	75.2±0.5	29.2±1.6	35.4±1.2
fgfr1	8.1±0.6	6.1±0.1	4.6±0.2	11.6±0.1	78.1±0.1	27.9±0.2	34.8±0.1
igf1r	12.2±0.7	6.9±0.0	5.4±0.0	16.3±0.4	81.8±0.1	34.6±0.2	39.7±0.3
jak2	1.2±0.0	5.0±0.2	5.2±0.2	9.3±0.1	80.1±0.6	23.0±0.5	32.8±0.2
kit	5.0±0.9	5.3±0.3	4.6±0.2	8.5±0.4	76.1±0.4	21.3±1.3	30.9±0.5
kpcb	11.2±0.0	6.3±0.4	4.6±0.0	15.4±0.1	71.4±0.3	31.6±0.4	35.4±0.1
lck	11.6±1.5	5.1±0.1	4.3±0.1	11.0±0.6	75.5±0.5	25.9±0.9	33.4±0.6
mapk2	17.0±0.8	8.0±0.2	5.7±0.1	18.6±0.2	79.7±0.6	39.3±0.2	41.2±0.3
met	5.5±0.6	5.7±0.2	4.3±0.1	9.6±0.1	73.8±0.2	25.4±0.3	32.1±0.1
mk01	21.5±0.7	12.8±0.3	7.2±0.2	33.8±1.5	84.5±1.1	58.3±1.5	52.2±1.3
mk10	9.4±1.2	7.7±0.2	5.3±0.1	15.6±0.8	80.5±0.5	32.4±0.7	38.1±0.4
mk14	6.8±0.0	5.3±0.2	5.7±0.1	11.1±0.3	80.3±0.3	27.4±0.7	35.2±0.4
mp2k1	3.6±1.0	5.6±0.2	4.8±0.2	9.1±0.5	73.9±1.2	23.5±1.2	31.2±0.9
plk1	23.9±0.7	12.3±0.1	6.6±0.1	33.8±0.9	84.1±0.9	56.5±0.3	52.0±0.7
rock1	5.8±1.0	4.2±0.2	3.9±0.1	8.3±0.3	73.9±0.5	20.4±1.0	29.9±0.4
src	5.3±0.5	5.8±0.1	4.3±0.1	10.2±0.1	81.2±0.5	24.9±0.6	34.2±0.1
tgfr1	8.2±0.0	6.2±0.0	4.4±0.4	18.8±0.5	86.9±0.1	29.5±1.1	39.8±0.5
vgfr2	9.0±0.6	5.1±0.3	4.5±0.3	11.1±0.5	81.5±0.4	25.7±1.2	34.8±0.6
wee1	0.0±0.0	4.9±0.0	4.9±0.0	10.1±0.2	82.3±0.9	21.0±0.3	32.3±0.4

Table 23: **TopU-LBVS-Full Tanimoto-NN results.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	12.7±0.7	6.8±0.3	5.0±0.3	16.3±0.4	86.0±0.9	34.3±1.5	39.9±0.7
esr1	8.0±0.0	4.8±0.2	4.7±0.1	10.2±0.1	77.3±0.5	25.1±0.2	33.0±0.3
esr2	8.3±1.7	4.8±0.3	4.3±0.2	10.0±1.0	76.1±0.3	23.7±2.3	31.9±0.9
gcr	7.1±0.0	7.8±0.3	5.8±0.1	14.2±0.2	86.6±0.2	33.9±0.7	39.2±0.1
mcr	5.5±3.9	4.6±1.8	2.3±0.9	8.2±1.8	75.4±0.8	18.0±8.1	26.8±2.1
ppara	3.2±3.0	4.3±0.2	3.8±0.2	8.2±0.5	79.5±0.6	18.5±2.2	29.8±0.9
ppard	11.8±1.3	6.6±0.3	4.8±0.0	14.5±0.5	81.8±0.4	34.0±1.2	37.2±0.7
pparg	4.6±0.6	4.9±0.5	3.8±0.0	8.3±0.5	78.8±0.4	19.8±1.5	30.5±0.6
prgr	4.8±1.7	4.5±0.6	4.8±0.1	10.1±0.3	86.1±0.4	22.2±0.8	35.0±0.3
rxra	15.2±2.1	5.0±0.0	5.8±0.2	16.2±0.2	87.9±1.8	31.8±0.1	40.0±0.5
thb	18.2±3.7	17.3±0.5	9.2±0.2	47.5±2.3	95.9±0.1	75.6±1.0	63.4±1.2
Other Enzymes							
aces	8.0±0.8	5.6±0.2	3.6±0.2	10.1±0.4	80.9±0.7	23.0±0.9	33.2±0.5
ampc	38.3±3.9	10.4±1.6	6.0±0.8	53.8±8.6	81.2±0.5	57.4±7.8	54.7±5.2
aofb	1.9±0.7	2.7±0.3	3.2±0.2	7.2±0.2	78.0±0.4	14.6±0.9	29.0±0.3
cah2	10.5±0.7	5.2±0.7	4.7±0.3	11.7±1.1	81.4±1.6	25.6±2.4	34.6±1.4
dhi1	4.6±3.2	6.1±0.7	4.0±0.3	11.7±0.3	83.8±1.0	27.4±0.9	34.2±0.7
dyr	6.8±2.4	3.5±0.5	2.1±0.0	8.1±0.7	78.6±0.5	14.7±1.8	29.2±0.9
fnta	5.5±3.9	5.5±0.8	5.5±0.4	11.0±1.0	79.2±1.2	26.4±2.8	32.5±1.6
glcm	4.6±0.0	10.2±0.8	8.5±0.0	21.2±0.2	91.5±0.5	46.8±0.4	45.8±0.2
hdac2	6.8±2.8	4.0±0.9	4.2±0.5	8.5±0.4	76.7±0.8	19.5±2.0	29.8±0.4
hdac8	2.3±1.6	3.4±0.6	3.2±0.3	7.0±0.5	76.6±1.3	14.4±1.7	28.2±1.1
hivint	6.2±1.7	4.1±0.6	4.0±0.3	8.7±0.9	77.2±0.8	20.1±2.0	31.2±1.0
hivrt	2.6±1.1	4.0±0.2	3.1±0.1	6.7±0.2	78.0±0.5	14.4±0.4	28.3±0.4
hmdh	8.2±0.0	8.3±0.8	4.5±0.0	20.6±0.9	82.6±0.8	30.9±2.3	37.4±1.0
hxk4	8.2±0.0	7.8±0.0	5.7±0.5	15.5±1.0	87.1±0.4	32.8±2.3	39.0±0.9
inha	13.7±0.0	5.2±0.7	4.7±0.7	18.3±1.3	90.4±0.8	30.0±3.7	41.1±1.6
nos1	1.7±2.4	3.5±0.5	5.8±0.0	10.9±0.4	85.8±0.3	20.9±2.8	34.5±0.6
parp1	4.6±0.0	2.3±0.5	2.2±0.2	6.2±0.6	72.9±0.8	10.6±2.0	25.4±0.8
pde5a	1.9±2.8	4.9±0.3	6.0±0.4	11.5±0.5	87.4±1.6	25.4±1.3	35.8±0.8
pgh1	6.3±0.0	6.2±0.1	4.3±0.0	10.0±0.3	78.3±0.7	25.6±0.8	32.8±0.4
pgh2	5.9±0.0	5.3±0.0	4.4±0.3	10.7±0.7	77.8±0.2	24.1±1.2	32.2±0.5
ptn1	6.1±0.0	4.1±0.4	3.4±0.1	7.2±0.4	72.4±0.4	17.8±1.1	27.6±0.6
pygm	16.4±0.0	7.8±0.0	4.3±0.5	25.1±3.2	76.0±0.2	36.6±0.6	39.5±0.7
pyrd	7.5±0.0	2.9±0.4	3.7±0.2	11.2±0.9	83.9±0.8	20.8±1.7	34.6±1.0
tysy	0.0±0.0	1.9±0.0	1.0±0.0	4.8±0.1	71.8±0.3	1.8±0.7	21.5±0.2
Protease							
ace	5.5±3.9	5.9±0.0	4.0±0.0	14.2±1.0	83.9±1.5	29.2±3.1	34.9±1.3
ada17	1.7±2.4	6.7±0.9	5.0±0.5	11.4±0.9	87.1±0.4	25.2±3.4	35.7±1.1
bace1	4.8±2.2	3.1±0.5	3.8±0.1	7.0±0.1	78.1±0.2	15.9±1.5	28.3±0.2
dpp4	7.4±3.0	5.3±0.3	3.8±0.3	10.6±0.8	88.2±1.0	22.2±2.5	34.0±1.0
fa10	6.1±2.8	4.3±0.4	3.5±0.1	9.0±0.7	80.8±0.5	21.0±1.8	31.4±0.8
hivpr	1.2±1.8	1.6±0.0	1.5±0.2	4.3±0.1	72.4±0.5	4.9±1.0	21.0±0.3
mmp13	1.9±2.8	2.7±0.6	1.4±0.3	4.4±0.5	67.9±2.4	4.8±3.8	20.7±1.3
thrb	5.4±0.9	5.2±0.3	3.7±0.1	9.5±0.2	79.4±0.2	21.9±0.4	31.9±0.2
try1	7.8±5.5	5.4±0.6	5.0±0.0	11.5±0.5	81.0±0.4	27.1±1.6	33.6±0.6
urok	0.0±0.0	7.8±0.6	5.7±0.3	14.2±0.3	88.0±0.1	32.1±0.8	38.6±0.2

Table 24: **TopU-LBVS-Full GAT results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	5.1±1.0	4.4±0.2	3.5±0.2	7.2±0.4	72.7±0.4	18.4±1.1	28.7±0.6
cp2c9	1.4±0.8	2.9±0.1	2.6±0.2	5.6±0.2	75.2±1.4	10.7±0.4	26.3±0.7
cp2d6	8.6±0.6	6.3±0.4	4.8±0.3	13.3±1.1	81.6±0.6	30.2±1.9	37.2±1.1
cp3a4	9.0±1.0	4.8±0.1	4.2±0.2	11.2±0.5	78.9±0.9	23.4±1.0	33.7±0.7
GPCR							
5ht6r	4.1±0.0	1.4±0.5	2.1±0.1	5.3±0.1	73.5±0.7	6.1±0.9	24.6±0.4
aa2ar	4.1±0.4	2.8±0.3	2.5±0.1	6.0±0.6	72.4±0.9	11.2±1.1	26.0±0.7
aa3r	3.6±0.4	2.7±0.5	2.3±0.4	5.4±0.2	69.0±0.8	10.5±1.9	24.5±0.8
adrb1	1.1±0.7	1.4±0.4	1.3±0.1	4.0±0.0	66.0±0.4	2.9±0.2	20.8±0.1
adrb2	2.3±0.5	2.0±0.3	1.9±0.0	4.7±0.1	69.0±0.7	7.3±1.0	23.4±0.3
cnr2	0.6±0.8	2.7±0.2	2.0±0.3	4.8±0.1	72.1±0.6	7.2±0.3	23.6±0.3
cxcr4	3.9±2.8	1.3±0.0	1.1±0.3	4.5±0.5	58.3±3.3	3.8±0.7	19.3±1.3
drd2	1.5±0.4	1.9±0.4	1.9±0.2	4.6±0.3	71.2±1.1	6.2±1.5	23.2±0.9
drd3	4.7±1.0	4.0±0.4	3.6±0.2	7.9±0.4	78.2±0.4	18.4±1.5	30.7±0.7
hrh3	5.1±1.4	4.5±0.3	3.9±0.3	9.8±0.4	82.1±0.3	22.2±0.9	33.8±0.4
oprd	2.5±0.9	3.0±0.6	2.7±0.3	6.4±0.5	75.4±0.8	12.2±2.0	27.5±1.0
oprk	1.6±0.6	2.1±0.3	2.1±0.2	4.7±0.3	69.8±1.2	7.0±1.4	23.2±0.9
oprm	1.7±0.7	2.9±0.1	3.2±0.2	6.4±0.2	76.7±0.1	13.1±0.7	28.2±0.4
Ion Channel							
5ht3a	4.6±0.9	5.7±0.3	5.3±0.2	11.6±1.0	82.3±0.5	26.5±1.2	35.3±0.8
cac1h	6.8±2.8	8.8±0.6	5.9±0.8	17.2±2.5	84.4±0.7	37.2±4.4	40.2±1.8
kcnh2	2.0±0.5	2.1±0.2	2.0±0.1	4.3±0.2	67.0±0.4	6.9±0.5	22.2±0.3
scn5a	2.3±1.9	0.9±0.4	1.3±0.4	4.5±1.6	60.9±1.3	1.6±2.7	18.5±1.3
trpa1	2.7±3.9	3.9±0.0	2.7±0.9	6.9±0.5	68.3±3.3	12.2±1.0	24.9±1.6
Kinase							
abl1	2.0±0.0	2.3±0.3	2.2±0.3	4.5±0.4	67.0±1.4	8.1±1.6	22.8±1.1
akt1	6.4±1.0	6.0±0.4	4.7±0.1	9.9±0.4	75.7±0.3	25.5±0.1	32.7±0.4
akt2	6.2±0.0	4.3±0.7	3.7±0.1	9.8±0.9	72.9±2.0	20.2±1.9	30.1±1.3
braf	0.0±0.0	0.8±0.2	1.1±0.2	4.0±0.1	69.8±0.4	0.5±0.4	20.7±0.2
cdk2	1.7±0.5	2.5±0.0	2.7±0.1	5.2±0.1	69.3±1.4	9.1±0.2	24.0±0.5
csf1r	3.6±1.0	2.2±0.4	2.2±0.2	5.4±0.1	72.2±0.7	8.1±0.5	24.8±0.1
egfr	2.0±0.5	2.8±0.3	3.2±0.2	6.3±0.9	73.6±0.6	12.0±1.2	26.6±0.7
fak1	6.5±0.0	3.2±0.4	3.5±0.3	8.8±0.6	72.8±1.8	16.3±1.5	29.1±0.4
fgfr1	5.6±0.6	4.5±0.1	3.5±0.2	8.1±0.1	74.7±0.6	18.3±0.4	29.7±0.3
igf1r	5.4±0.7	4.1±0.6	4.1±0.2	8.3±0.3	76.3±1.2	19.8±0.8	30.7±0.5
jak2	2.3±0.0	3.7±0.1	3.6±0.3	6.3±0.2	72.0±0.9	15.3±0.5	27.0±0.6
kit	0.0±0.0	1.0±0.6	1.4±0.4	3.7±0.2	63.7±0.5	2.1±2.6	19.4±0.8
kpcb	3.7±0.0	4.3±0.4	3.2±0.7	7.2±0.2	71.6±1.0	16.0±1.0	27.4±0.5
lck	2.5±0.5	3.0±0.6	3.2±0.1	5.9±0.5	70.1±0.8	12.6±1.3	25.5±0.2
mapk2	10.4±2.0	4.4±0.3	3.2±0.1	10.5±2.1	67.6±0.7	20.0±1.3	29.0±0.9
met	6.4±2.8	4.6±0.1	3.8±0.2	8.3±0.8	72.4±0.9	20.1±1.4	29.9±0.8
mk01	17.1±0.7	9.7±0.4	6.3±0.1	22.3±1.1	84.8±1.0	45.0±1.5	45.2±1.1
mk10	8.5±1.2	4.0±0.3	2.9±0.3	9.5±1.1	71.8±0.9	18.9±0.9	29.5±0.8
mk14	5.3±0.5	4.6±0.6	3.8±0.2	9.0±0.7	75.7±0.3	20.9±1.4	31.3±0.4
mp2k1	6.5±0.0	3.2±0.4	2.3±0.1	5.3±0.6	57.7±1.7	11.6±0.8	21.6±0.9
plk1	12.2±2.8	6.4±0.4	4.3±0.4	15.3±2.5	79.6±0.4	31.0±2.9	37.4±1.7
rock1	10.1±1.0	4.1±0.4	2.9±0.3	11.5±0.8	71.4±0.4	19.0±0.8	30.3±0.2
src	2.3±0.5	2.1±0.1	2.2±0.2	5.0±0.1	71.0±0.2	7.9±0.1	24.1±0.2
tgfr1	0.0±0.0	1.8±1.3	1.7±0.2	4.9±0.3	73.3±1.1	3.3±1.8	22.9±0.7
vgfr2	4.0±0.4	3.4±0.1	2.5±0.1	6.1±0.1	71.2±0.9	12.9±0.5	26.1±0.3
wee1	2.7±3.9	3.3±0.0	2.5±0.7	7.0±1.3	66.3±2.3	12.7±2.5	24.8±2.3

Table 25: **TopU-LBVS-Full GAT results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	1.5±0.0	2.0±0.4	2.8±0.3	6.2±0.1	76.7±0.2	8.8±0.8	26.8±0.2
esr1	2.4±1.5	2.7±0.4	2.9±0.2	5.9±0.4	73.1±1.2	11.5±1.7	26.3±1.1
esr2	2.4±0.8	2.3±0.2	2.4±0.1	6.1±0.7	72.1±0.1	8.0±0.3	24.5±0.1
gcr	5.9±2.2	5.0±0.5	4.9±0.1	11.7±0.8	85.5±0.5	26.1±1.8	36.7±0.9
mcr	0.0±0.0	2.0±1.6	2.7±0.9	5.9±0.2	72.3±1.9	7.7±1.4	23.8±0.7
ppara	0.0±0.0	1.8±0.5	2.0±0.4	4.9±0.0	71.9±0.4	6.5±0.1	23.6±0.1
ppard	2.7±0.0	4.2±0.3	3.0±0.3	7.7±0.6	75.9±1.2	15.0±1.2	28.8±1.1
pparg	2.3±0.6	3.0±0.4	2.6±0.2	6.1±0.4	73.4±0.2	11.7±1.3	26.5±0.5
prgr	5.3±1.5	3.6±1.1	3.6±0.4	8.9±0.6	81.9±0.6	18.5±2.8	32.4±1.0
rxra	0.0±0.0	1.7±0.5	2.5±1.1	5.9±0.8	76.0±2.8	7.5±2.3	25.5±2.0
thb	21.3±2.1	12.3±0.5	7.5±0.8	43.6±4.2	91.9±0.4	61.2±4.3	58.1±2.8
Other Enzymes							
aces	2.8±1.5	2.9±0.0	3.4±0.2	6.6±0.1	76.7±0.4	13.7±0.3	28.3±0.2
ampc	19.1±3.9	13.1±1.3	7.9±0.4	42.0±2.4	95.5±1.3	59.8±4.7	56.4±2.9
aofb	1.5±0.0	1.8±0.7	1.3±0.3	4.1±0.1	68.1±0.5	3.3±1.2	21.2±0.5
cah2	3.2±0.0	2.2±0.3	2.0±0.4	5.2±0.4	72.1±1.8	7.6±1.6	24.3±1.1
dhi1	0.0±0.0	2.4±0.7	2.8±0.0	5.0±0.2	66.7±1.0	7.5±1.2	21.9±0.6
dyr	0.0±0.0	2.1±1.5	2.1±0.4	4.3±0.1	64.6±2.5	5.2±3.1	20.3±0.2
fnta	5.5±3.9	4.4±0.8	3.8±0.4	8.0±0.4	72.8±0.6	16.7±1.6	27.7±0.5
glcm	7.6±2.1	5.0±1.9	5.7±0.0	15.4±1.4	84.6±2.6	28.6±6.2	37.9±3.3
hdac2	3.4±0.0	0.9±0.3	1.8±0.5	6.2±1.5	73.6±0.3	4.1±1.1	23.9±0.6
hdac8	3.4±0.0	2.0±0.6	1.4±0.0	6.3±1.2	65.8±1.5	5.2±0.6	21.2±0.4
hivint	0.7±1.0	0.8±0.3	2.1±0.6	4.4±0.4	70.0±0.8	3.2±2.4	21.9±1.2
hivrt	2.2±0.6	2.1±0.2	2.0±0.3	5.0±0.3	70.4±0.7	7.0±1.0	23.5±0.6
hmdh	0.0±0.0	1.2±0.8	2.1±0.4	5.5±0.2	75.2±1.0	4.5±1.2	23.8±0.5
hxk4	8.2±0.0	3.3±0.9	2.3±0.9	12.6±3.6	81.7±0.7	15.9±3.8	31.6±1.6
inha	0.0±0.0	0.5±0.7	2.1±0.0	4.9±0.7	69.7±3.2	2.9±2.9	22.1±2.0
nos1	0.0±0.0	3.5±1.3	4.7±0.4	9.5±0.9	81.6±0.8	18.4±3.4	32.1±1.4
parp1	0.0±0.0	0.0±0.0	1.0±0.0	3.7±0.2	66.7±1.9	-2.1±0.7	18.5±0.8
pde5a	7.8±1.4	4.3±0.6	3.5±0.2	11.0±1.1	78.5±0.1	20.0±1.0	31.9±0.3
pgh1	3.2±0.0	1.6±0.3	1.2±0.1	5.0±0.0	58.8±0.4	3.6±0.6	18.7±0.2
pgh2	7.8±2.8	5.7±0.6	4.0±0.5	13.2±3.9	74.0±3.6	25.6±4.1	32.5±3.1
ptn1	3.0±0.0	2.3±0.1	2.6±0.1	5.2±0.1	68.8±0.4	10.1±0.3	24.1±0.2
pygm	2.7±3.9	1.3±0.9	1.7±0.5	4.7±0.9	64.8±1.4	2.3±3.8	19.7±1.0
pyrd	0.0±0.0	2.1±0.4	2.1±0.2	5.2±0.3	73.0±1.1	5.7±1.4	23.9±0.9
tysy	0.0±0.0	0.7±0.9	1.0±0.8	3.2±0.0	55.8±1.7	-2.2±2.0	15.1±0.1
Protease							
ace	2.7±3.9	3.9±0.0	3.3±0.5	8.2±0.8	74.6±0.1	14.9±2.2	27.9±0.9
ada17	0.0±0.0	3.0±0.5	2.0±0.3	6.7±0.2	79.0±0.7	10.0±2.0	27.3±0.3
bace1	0.6±0.8	2.2±0.5	2.5±0.1	5.3±0.1	74.4±0.1	8.0±0.9	25.0±0.2
dpp4	0.0±0.0	1.5±0.3	2.6±0.3	6.3±0.2	77.0±1.1	7.5±0.9	26.6±0.5
fa10	1.5±2.1	2.5±0.2	2.2±0.3	5.7±0.0	75.6±0.6	9.7±0.1	25.9±0.2
hivpr	0.0±0.0	1.6±0.6	2.1±0.4	4.5±0.2	69.3±1.0	3.7±1.4	21.7±0.7
mmp13	0.0±0.0	1.6±1.1	2.0±0.3	4.2±0.2	64.5±0.6	3.3±2.1	20.0±0.5
thrb	2.2±0.4	2.8±0.2	2.5±0.2	5.6±0.1	74.6±0.3	10.6±0.7	26.1±0.1
try1	2.0±2.8	2.1±1.2	1.9±0.5	5.3±0.3	70.5±2.3	5.8±1.7	22.9±1.2
urok	2.0±2.8	1.2±0.0	1.6±0.7	5.7±0.5	74.7±0.9	5.5±2.4	24.7±1.2

Table 26: **TopU-LBVS-Full GAT+FP hybrid model results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	12.2±1.2	7.1±0.4	4.9±0.2	16.0±1.0	77.3±1.1	33.1±1.7	37.6±1.3
cp2c9	13.4±1.4	7.5±0.2	5.3±0.1	17.2±1.3	82.8±0.3	36.3±1.1	40.5±0.6
cp2d6	15.6±0.9	8.7±0.1	5.7±0.2	21.8±0.8	81.6±1.0	41.1±0.9	42.8±0.5
cp3a4	14.3±2.1	8.4±0.6	5.8±0.1	21.7±2.3	81.9±0.5	40.8±2.0	42.8±1.2
GPCR							
5ht6r	6.2±0.0	3.8±0.3	3.7±0.3	9.1±0.6	76.7±1.5	17.8±2.6	30.5±1.1
aa2ar	5.2±1.0	3.5±0.3	3.4±0.1	7.2±0.1	74.9±0.6	16.3±0.4	29.1±0.3
aa3r	4.4±1.0	3.5±0.3	3.6±0.1	7.5±0.3	76.2±1.0	16.1±1.2	29.6±0.7
adrb1	3.2±0.0	1.3±0.3	1.8±0.3	4.7±0.1	68.9±1.2	5.5±1.0	23.0±0.3
adrb2	2.9±1.6	2.7±0.4	3.2±0.3	6.3±0.9	69.8±0.5	12.5±2.1	26.0±1.2
cnr2	3.0±0.8	3.6±0.8	3.4±0.2	7.8±0.9	80.9±1.0	16.3±2.8	30.7±1.5
cxcr4	3.9±2.8	2.6±1.1	2.9±0.8	7.4±1.6	71.8±1.4	12.8±6.0	26.8±2.4
drd2	4.1±0.8	3.2±0.3	3.0±0.1	6.7±0.1	75.6±0.8	13.6±0.3	28.2±0.3
drd3	8.2±0.0	4.8±0.3	4.2±0.1	12.2±0.6	82.8±0.3	24.0±0.3	35.4±0.3
hrh3	8.1±0.7	6.5±1.0	5.1±0.3	13.0±1.2	81.3±0.4	29.9±3.2	37.2±1.2
oprd	5.4±1.6	4.2±0.5	4.3±0.1	9.1±0.2	79.9±0.2	20.8±1.1	32.7±0.3
oprk	2.7±1.5	3.4±0.4	3.1±0.1	6.9±0.6	77.1±1.0	14.5±2.6	28.9±1.3
oprm	4.9±2.7	4.8±0.2	4.3±0.3	10.2±1.1	79.7±0.1	23.1±1.4	33.8±1.0
Ion Channel							
5ht3a	4.6±3.7	7.2±0.2	5.6±0.1	12.1±0.8	82.6±0.8	29.3±1.4	36.4±0.8
cac1h	9.1±1.6	10.0±0.3	7.6±0.5	22.9±0.8	88.2±0.3	47.7±1.9	46.6±0.6
kcnh2	2.5±0.4	3.0±0.2	2.7±0.1	5.5±0.1	73.0±0.3	11.1±0.6	26.0±0.2
scn5a	2.3±1.9	3.1±0.2	2.4±0.5	5.6±0.4	70.8±1.5	10.7±1.8	25.1±1.0
trpa1	13.7±3.9	6.5±0.9	4.7±0.5	20.8±6.3	70.7±0.6	32.5±4.8	34.6±3.0
Kinase							
abl1	1.4±0.5	5.3±0.1	4.1±0.3	7.7±0.2	75.0±0.2	19.9±1.1	29.8±0.3
akt1	6.8±0.6	6.9±0.6	5.4±0.2	12.5±0.7	80.7±0.2	30.7±2.2	36.7±0.8
akt2	10.9±4.8	7.2±0.5	5.1±0.5	16.6±2.6	79.6±0.8	34.0±3.2	38.4±1.9
braf	9.1±3.5	6.0±0.2	5.2±0.1	14.6±2.7	84.5±0.3	29.4±1.3	37.7±1.3
cdk2	3.8±0.5	4.9±0.4	4.5±0.3	9.2±0.4	78.6±0.7	22.2±0.8	32.5±0.6
csf1r	9.4±1.0	3.6±0.2	3.0±0.4	11.6±2.0	78.1±0.5	18.3±0.9	32.4±0.7
egfr	2.0±0.5	4.0±0.2	4.0±0.2	8.2±0.1	79.6±0.9	17.7±0.6	31.1±0.4
fak1	11.5±2.7	5.3±0.6	4.4±0.3	14.6±2.5	75.3±0.5	27.6±2.0	34.9±1.0
fgfr1	8.1±2.6	6.1±0.3	4.6±0.2	11.5±0.8	78.0±0.5	28.2±1.7	34.8±0.9
igf1r	7.3±0.0	5.8±0.4	5.7±0.3	12.3±0.7	80.8±1.0	28.8±0.4	36.1±0.5
jak2	6.2±0.6	5.1±0.6	4.3±0.4	10.0±0.5	80.0±1.0	23.6±1.5	33.9±0.8
kit	6.2±1.8	4.0±0.6	3.3±0.4	8.1±1.6	72.4±1.1	16.9±2.5	28.6±1.3
kpcb	3.7±0.0	5.3±0.6	3.8±0.3	9.1±0.6	77.0±0.7	21.3±1.8	31.2±0.6
lck	6.3±1.7	4.8±0.7	4.0±0.5	8.6±0.8	73.8±1.1	20.9±2.3	30.3±1.0
mapk2	8.7±2.8	6.4±0.5	4.6±0.2	13.0±1.2	75.1±0.6	28.2±0.9	34.4±0.6
met	6.0±1.6	5.7±1.0	4.8±0.2	10.8±1.1	74.4±1.1	25.4±2.3	32.9±1.1
mk01	27.8±3.2	11.4±0.4	6.8±0.1	40.6±3.7	85.3±0.8	56.6±2.1	54.7±2.3
mk10	7.7±3.6	5.2±0.5	3.7±0.5	10.8±0.9	76.8±1.7	21.8±4.2	32.3±2.2
mk14	6.8±0.0	7.0±0.4	5.6±0.4	13.0±1.0	79.5±0.6	31.2±0.5	36.6±0.4
mp2k1	6.5±0.0	3.5±0.2	3.9±0.2	8.0±0.7	70.9±1.7	18.3±1.2	28.9±0.9
plk1	20.5±1.2	11.1±0.1	6.4±0.2	32.4±0.5	83.6±1.5	51.8±0.9	49.4±0.9
rock1	3.6±1.0	3.6±0.4	3.5±0.3	9.2±1.1	74.7±0.5	17.6±1.5	29.9±0.8
src	7.0±0.8	5.3±0.2	4.0±0.1	9.9±0.2	78.5±0.8	23.5±0.9	33.1±0.4
tgfr1	2.7±1.9	3.3±0.4	3.9±0.2	10.3±2.0	83.7±0.4	16.7±0.9	32.6±0.9
vgfr2	6.7±0.6	5.2±0.7	4.1±0.1	10.5±0.4	80.8±0.8	23.9±1.6	34.4±0.7
wee1	8.2±0.0	5.5±0.8	4.4±0.4	14.4±1.6	82.4±1.3	26.3±3.0	35.9±1.7

Table 27: **TopU-LBVS-Full GAT-FP hybrid model results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	4.9±1.8	5.3±0.7	4.2±0.3	10.5±0.6	83.1±0.8	23.8±2.6	34.6±1.1
esr1	4.2±1.1	4.4±0.8	3.8±0.1	7.3±0.5	72.7±0.3	18.5±1.9	28.9±0.8
esr2	4.8±1.7	4.0±0.6	3.6±0.5	8.9±0.8	75.2±1.4	18.2±3.3	29.9±1.7
gcr	8.9±0.0	6.3±0.4	5.5±0.4	14.8±0.4	85.4±0.6	32.1±1.3	39.3±0.6
mcr	8.2±0.0	3.9±0.0	3.3±0.5	11.9±2.0	78.2±1.5	19.7±2.9	31.7±1.2
ppara	0.0±0.0	2.0±0.4	2.9±0.1	6.4±0.3	77.8±1.3	9.8±1.2	27.4±0.9
ppard	4.6±1.3	4.7±0.5	4.7±0.5	11.4±0.6	81.6±0.3	24.2±2.3	35.0±0.9
pparg	3.6±0.6	4.3±0.7	4.0±0.4	8.5±0.5	80.4±0.5	18.9±1.9	32.0±0.7
prgr	3.6±1.5	6.0±0.3	5.3±0.3	12.3±1.3	83.2±0.4	28.1±1.2	36.1±0.6
rxra	7.6±2.1	5.7±0.9	5.2±0.2	15.0±2.6	85.5±0.3	28.3±3.2	38.2±1.4
thb	30.4±2.1	14.3±0.5	8.7±0.2	62.0±3.3	93.6±0.7	74.2±0.9	68.8±1.9
Other Enzymes							
aces	4.8±0.4	4.4±0.2	3.8±0.3	8.3±0.3	78.7±0.3	19.5±0.9	31.6±0.5
ampc	24.6±6.7	13.7±1.5	8.2±0.7	49.8±1.7	94.6±1.9	65.7±5.1	59.6±2.7
aofb	1.5±1.2	4.3±0.3	3.2±0.2	6.7±0.3	76.2±0.7	15.0±1.7	28.4±0.4
cah2	3.7±0.7	2.9±0.5	3.3±0.3	8.8±0.5	82.0±0.7	14.0±2.0	30.9±0.7
dhi1	0.0±0.0	1.4±0.0	1.4±0.0	4.9±0.3	70.2±2.1	4.8±0.8	22.0±1.0
dyr	5.1±0.0	2.8±1.0	3.0±0.5	7.5±1.1	72.6±1.8	13.2±3.0	27.1±1.9
fnta	5.5±3.9	2.2±0.8	3.6±0.4	9.7±1.4	77.0±1.1	16.4±2.3	30.3±1.5
glcm	15.2±2.1	11.5±0.9	7.9±0.4	30.5±3.0	90.3±0.5	55.0±2.5	51.2±1.4
hdac2	4.6±1.6	2.8±1.2	3.1±0.6	7.7±0.2	80.3±1.1	13.5±2.0	29.7±0.6
hdac8	0.0±0.0	0.2±0.3	0.8±0.2	4.1±0.2	71.7±1.2	-2.1±0.6	20.4±0.5
hivint	0.7±1.0	1.5±0.2	2.1±0.2	5.3±0.3	75.3±0.8	5.6±1.3	24.9±0.8
hivrt	4.4±1.6	2.8±0.5	2.4±0.2	6.0±0.5	73.5±0.3	11.6±0.7	26.5±0.6
hmdh	16.4±0.0	7.1±0.0	5.6±0.4	27.0±2.8	85.3±0.5	37.1±0.2	42.5±0.6
hxk4	2.7±3.9	3.9±0.0	3.3±0.5	11.0±1.6	85.3±2.0	17.2±2.0	33.1±1.4
inha	6.8±5.6	5.7±1.2	4.2±0.6	15.5±2.4	90.0±0.9	28.5±5.5	39.6±2.3
nos1	12.0±4.8	6.0±0.5	5.8±0.4	18.3±3.2	86.8±0.4	33.5±1.0	41.2±1.2
parp1	3.0±2.1	3.3±0.5	3.2±0.5	7.5±1.2	74.7±0.8	13.3±2.2	27.7±1.3
pde5a	6.8±1.4	5.3±0.5	4.7±0.4	15.3±1.0	84.9±1.0	26.4±1.2	37.2±0.6
pgh1	3.2±2.2	3.7±0.1	3.2±0.1	7.4±1.4	75.7±0.3	14.8±2.4	28.5±1.2
pgh2	2.0±2.8	4.4±0.6	4.4±0.6	9.8±0.8	81.4±1.5	20.0±1.9	32.4±1.2
ptn1	7.6±1.2	3.5±0.5	3.3±0.5	10.1±1.0	74.5±0.8	17.3±2.2	30.2±1.0
pygm	10.9±3.9	7.8±1.6	5.7±0.5	19.3±1.6	73.3±2.3	34.1±2.8	36.4±1.5
pyrd	3.7±0.0	3.7±1.0	4.1±0.8	8.8±0.3	82.0±0.4	17.0±2.7	31.6±0.7
tysy	0.0±0.0	0.0±0.0	0.7±0.5	4.5±0.2	68.7±0.4	-2.0±1.3	20.0±0.6
Protease							
ace	8.2±6.7	7.2±0.9	5.0±0.0	18.8±4.7	77.9±0.3	32.8±1.4	36.7±1.5
ada17	3.4±2.4	3.7±1.0	3.1±0.7	10.5±1.9	84.6±0.4	15.2±1.4	32.1±0.3
bace1	0.6±0.8	1.5±0.3	2.0±0.2	5.5±0.2	76.0±1.1	6.3±0.8	25.5±0.4
dpp4	3.2±2.6	5.8±1.4	4.9±0.5	10.8±2.4	84.9±3.3	23.8±6.3	34.9±3.7
fa10	0.8±1.1	2.9±0.9	2.9±0.6	7.3±0.5	81.9±0.6	13.0±2.8	29.8±1.0
hivpr	5.0±1.8	2.9±0.4	2.5±0.2	7.3±0.3	76.6±0.6	13.3±0.6	28.0±0.1
mmp13	0.0±0.0	0.8±0.6	1.0±0.3	3.7±0.3	64.2±2.1	-0.5±0.8	18.5±1.0
thrb	3.5±2.2	3.7±0.7	3.1±0.0	7.3±0.5	78.4±0.3	15.6±1.7	29.7±0.8
try1	2.0±2.8	4.1±0.6	3.7±0.0	8.2±0.6	77.8±1.3	16.3±1.7	29.6±0.9
urok	3.9±2.8	7.4±0.6	5.1±0.3	14.0±1.4	86.0±1.1	31.2±3.5	37.6±1.5

Table 28: **TopU-LBVS-Full GIN results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	4.9±0.7	4.7±0.4	3.7±0.2	8.0±0.1	75.8±1.1	20.2±0.6	30.5±0.4
cp2c9	5.0±0.0	2.9±0.3	2.4±0.2	6.3±0.3	75.5±0.7	12.3±1.1	27.3±0.5
cp2d6	7.7±1.8	6.5±0.6	4.9±0.2	12.5±1.2	81.9±0.9	28.4±2.7	36.4±1.5
cp3a4	10.2±0.9	6.2±1.0	4.7±0.4	13.5±1.7	80.6±0.8	29.3±3.5	36.8±1.7
GPCR							
5ht6r	3.4±1.0	2.1±0.3	2.4±0.3	6.3±0.1	74.4±0.8	8.7±0.4	26.1±0.4
aa2ar	4.9±1.0	2.6±0.3	2.3±0.4	5.4±0.7	69.5±1.0	10.4±2.5	24.7±1.3
aa3r	2.7±1.0	1.9±0.2	2.2±0.2	5.7±0.3	71.2±0.5	7.9±0.7	24.7±0.4
adrb1	0.0±0.0	1.3±0.3	1.6±0.0	4.0±0.1	65.6±0.9	3.4±0.5	20.7±0.4
adrb2	3.3±0.5	2.1±0.6	2.1±0.1	5.4±0.4	71.6±0.8	8.3±2.3	25.0±1.1
cnr2	2.4±0.8	2.7±0.7	2.8±0.1	6.1±0.5	75.9±1.4	10.9±2.2	26.8±1.3
cxcr4	3.9±2.8	1.8±0.6	1.5±0.3	5.3±1.2	62.6±1.3	4.9±1.8	20.5±1.2
drd2	1.2±1.1	2.1±0.1	2.2±0.2	5.1±0.2	74.2±0.6	7.4±0.9	24.9±0.5
drd3	2.3±0.0	5.5±0.4	3.9±0.3	9.0±0.3	80.3±0.6	21.9±0.8	32.4±0.4
hrh3	7.1±3.6	5.0±0.4	4.3±0.1	10.4±1.1	81.3±0.6	23.8±2.2	34.4±1.2
oprd	2.5±1.2	2.3±0.5	2.9±0.1	6.5±0.4	76.5±0.2	10.7±1.0	27.4±0.2
oprk	3.1±1.5	3.2±0.3	2.3±0.1	6.0±0.4	72.4±0.7	11.3±0.5	25.9±0.4
oprm	4.1±1.1	3.3±0.5	3.6±0.0	7.2±0.2	77.7±0.7	15.6±0.2	29.7±0.5
Ion Channel							
5ht3a	2.6±0.9	5.4±0.7	5.3±0.2	10.7±0.6	84.1±0.8	25.0±2.2	35.1±0.8
cac1h	9.1±1.6	10.0±0.3	6.4±0.2	24.6±0.3	88.0±0.8	45.7±1.1	45.8±0.6
kcnh2	1.9±0.5	1.9±0.3	2.1±0.2	4.4±0.2	68.8±0.5	6.5±1.5	22.8±0.7
scn5a	3.0±2.8	2.5±0.8	2.0±0.1	5.0±0.9	65.5±1.8	8.6±1.5	22.4±1.2
trpa1	2.7±3.9	4.6±1.8	4.0±0.0	8.8±1.4	70.6±1.0	18.4±4.1	27.9±1.9
Kinase							
abl1	1.4±0.5	2.7±0.2	2.9±0.0	5.2±0.1	70.1±0.3	9.7±0.6	24.5±0.1
akt1	5.1±1.8	5.5±0.7	4.9±0.3	10.5±1.5	76.6±0.3	26.0±2.7	33.2±1.1
akt2	6.2±0.0	5.5±0.5	4.5±0.4	10.5±1.0	76.7±1.3	24.3±1.0	32.6±0.9
braf	2.3±0.8	1.6±0.8	1.3±0.1	4.8±0.8	70.9±1.4	4.2±2.5	22.4±1.1
cdk2	2.0±1.4	2.6±0.3	3.1±0.2	5.8±0.3	73.4±0.5	11.2±0.6	26.2±0.6
csf1r	5.0±2.7	2.7±0.5	2.7±0.5	9.4±2.5	75.5±1.4	12.7±2.8	28.5±1.9
egfr	2.8±0.3	3.9±0.3	3.7±0.1	7.2±0.3	76.7±0.0	16.2±0.2	29.3±0.1
fak1	9.4±1.0	4.5±0.4	3.9±0.4	10.7±2.1	78.1±2.1	21.9±2.3	32.4±2.0
fgfr1	6.0±0.6	4.2±0.2	3.1±0.1	8.1±0.6	74.4±0.8	17.6±0.8	29.2±0.4
igf1r	5.9±1.2	4.9±0.6	4.0±0.1	9.3±0.4	77.6±1.3	21.8±1.4	32.1±0.6
jak2	8.2±1.9	4.8±0.6	3.8±0.3	10.1±1.3	77.4±0.1	22.0±2.4	32.7±1.1
kit	2.5±1.8	2.4±0.4	2.6±0.2	5.7±0.4	71.5±0.8	9.7±0.4	25.1±0.6
kpcb	2.5±1.8	3.8±0.0	3.4±0.5	8.0±0.4	74.4±1.8	17.0±1.1	28.7±0.5
lck	2.1±0.9	3.0±0.2	2.9±0.1	5.3±0.2	70.1±0.4	11.7±0.8	24.9±0.4
mapk2	9.3±0.8	3.8±0.2	3.5±0.2	9.7±1.9	70.8±1.2	19.2±0.7	29.3±1.0
met	7.7±1.0	5.2±0.6	4.5±0.0	9.8±1.0	73.4±0.5	24.1±2.2	32.0±1.0
mk01	16.1±2.1	9.7±0.5	6.3±0.2	24.4±2.3	82.6±1.0	46.0±2.6	45.1±1.7
mk10	6.8±1.2	4.5±0.7	3.5±0.2	11.3±1.2	76.3±2.9	19.9±3.2	31.5±2.2
mk14	7.2±1.4	6.1±0.5	4.8±0.2	11.2±0.6	78.0±0.6	26.3±0.7	34.4±0.5
mp2k1	5.0±2.7	3.5±0.2	2.6±0.3	5.8±0.5	61.0±2.1	14.0±1.7	23.3±1.2
plk1	9.8±0.7	6.4±0.1	4.8±0.0	15.5±0.7	80.6±0.7	32.0±0.6	37.9±0.7
rock1	7.9±1.0	4.7±0.2	3.9±0.4	9.5±0.3	73.3±0.9	21.5±0.8	31.0±0.1
src	4.0±2.2	2.9±0.2	2.7±0.2	6.2±0.8	73.5±1.0	11.5±2.3	26.7±1.5
tgfr1	0.0±0.0	0.6±0.4	1.8±0.4	4.9±0.2	74.2±0.8	2.4±1.3	23.0±0.7
vgfr2	5.5±0.9	3.6±0.2	3.3±0.0	7.4±0.4	76.5±0.4	16.2±0.4	29.3±0.1
wee1	5.5±3.9	3.8±0.8	3.6±0.4	8.3±0.9	72.8±2.2	15.6±2.0	28.1±1.3

Table 29: **TopU-LBVS-Full GIN results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	0.5±0.7	2.5±0.1	2.6±0.2	6.1±0.3	79.1±0.5	9.2±1.3	27.3±0.7
esr1	2.1±0.8	2.3±0.5	2.8±0.3	5.7±0.6	72.9±1.2	9.8±1.4	25.6±0.9
esr2	2.4±0.8	2.1±0.0	2.1±0.1	5.8±0.6	73.0±1.0	6.9±0.7	24.6±0.9
gcr	5.3±1.5	5.6±0.3	5.2±0.1	11.9±0.7	85.1±0.5	27.9±1.6	37.0±0.8
mcr	0.0±0.0	2.0±1.6	2.3±0.5	6.2±0.7	72.5±0.6	7.8±3.6	24.1±1.4
ppara	3.2±1.1	2.8±0.5	2.7±0.2	6.0±0.4	74.2±1.1	12.0±2.0	26.2±0.8
ppard	6.4±1.3	4.0±0.8	3.4±0.5	8.6±0.7	76.7±0.6	18.3±3.0	30.4±1.1
pparg	2.7±1.1	3.0±0.4	3.1±0.1	6.5±0.0	74.7±0.6	13.2±0.8	27.6±0.2
prgr	5.9±2.2	5.4±1.1	4.5±0.3	10.4±0.7	82.9±0.8	24.5±1.8	34.7±0.9
rxra	1.5±2.1	3.7±0.5	3.3±0.5	7.7±0.8	79.1±0.4	14.9±3.1	29.4±1.3
thb	31.9±0.0	15.0±0.0	9.0±0.4	59.6±0.6	93.7±0.8	74.9±2.3	67.4±1.1
Other Enzymes							
aces	2.6±1.2	3.1±0.1	3.1±0.1	6.4±0.1	76.7±0.6	13.6±0.8	28.1±0.2
ampc	27.3±3.9	12.0±1.5	8.2±0.7	48.2±10.3	95.0±1.2	63.8±6.8	59.4±4.1
aofb	1.0±0.7	1.1±0.1	2.1±0.1	4.9±0.0	73.5±0.7	4.8±0.4	23.9±0.2
cah2	5.8±0.7	1.8±0.2	1.8±0.1	6.0±0.0	74.2±1.1	7.4±0.5	25.2±0.5
dhi1	0.0±0.0	2.4±0.7	2.6±0.3	4.9±0.3	66.1±1.0	7.4±1.4	21.6±0.8
dyr	0.0±0.0	1.1±0.9	1.4±0.2	4.5±0.7	67.4±1.1	3.4±4.1	20.9±1.5
fnta	0.0±0.0	2.7±0.8	3.0±0.4	6.2±0.3	71.9±1.9	9.6±1.6	24.8±0.9
glcm	13.7±3.7	7.1±0.4	6.6±0.4	22.5±4.3	86.1±1.5	40.0±2.7	44.2±2.6
hdac2	3.4±0.0	2.1±0.0	2.1±0.3	8.2±0.7	74.6±0.3	8.1±1.3	25.8±0.3
hdac8	3.4±0.0	1.6±0.3	1.4±0.3	7.3±0.2	66.4±0.9	4.9±1.0	21.4±0.6
hivint	2.0±0.0	2.3±0.4	2.6±0.2	5.7±0.3	73.1±0.7	9.6±1.0	25.2±0.2
hivrt	4.4±0.6	2.7±0.2	2.3±0.1	6.6±0.9	72.2±0.7	10.9±0.7	25.7±0.5
hmdh	5.5±3.9	3.0±2.2	3.0±0.8	8.3±2.2	75.8±2.5	12.9±8.6	27.5±3.4
hxk4	2.7±3.9	2.0±0.0	2.7±0.5	8.4±1.0	82.0±0.3	10.3±2.1	29.1±0.9
inha	0.0±0.0	2.4±0.7	3.1±0.3	8.0±0.4	85.0±0.7	9.7±0.8	30.1±0.6
nos1	6.8±2.4	2.8±0.5	4.6±0.2	11.8±0.6	84.6±1.0	21.8±0.9	35.4±0.8
parp1	0.0±0.0	0.7±0.9	1.2±0.2	3.8±0.2	67.2±0.9	-1.0±1.4	19.1±0.7
pde5a	2.9±0.0	5.3±0.8	4.5±0.3	10.9±1.1	82.8±0.3	22.5±1.7	33.7±0.7
pgh1	2.6±0.7	1.8±0.1	1.7±0.3	4.0±0.4	61.8±2.2	4.7±1.7	20.2±1.4
pgh2	5.9±0.0	4.4±1.6	3.5±0.8	10.3±4.3	68.1±2.3	18.5±7.5	27.8±2.8
ptn1	4.0±1.9	2.9±0.1	2.8±0.1	5.9±0.2	70.5±0.9	12.9±0.2	25.8±0.4
pygm	0.0±0.0	1.3±0.9	2.0±0.8	5.4±0.5	71.6±1.2	5.0±2.9	23.0±1.2
pyrd	0.0±0.0	2.4±0.6	2.1±0.2	5.5±0.3	74.9±0.6	7.7±1.7	24.9±0.8
tysy	0.0±0.0	2.0±0.0	1.3±0.5	4.3±0.4	62.0±3.7	3.4±1.5	18.6±1.6
Protease							
ace	8.2±0.0	7.2±0.9	4.0±0.0	13.5±3.7	73.2±1.2	24.8±2.3	31.0±0.8
ada17	5.1±0.0	2.6±0.5	2.8±0.5	8.5±0.3	80.6±0.5	12.0±1.3	29.6±0.3
bace1	0.6±0.8	1.6±0.6	2.4±0.3	5.3±0.2	74.7±0.5	6.8±0.9	24.9±0.5
dpp4	1.1±1.5	4.5±0.5	4.2±0.3	7.9±0.7	76.2±0.9	17.5±3.1	29.3±0.8
fa10	1.5±1.1	2.3±0.8	2.2±0.3	5.3±0.2	74.7±1.1	7.8±1.8	25.0±0.6
hivpr	1.2±1.8	2.1±0.4	1.7±0.5	5.1±0.4	71.0±1.3	7.0±1.3	23.5±0.9
mmp13	2.0±2.8	1.2±1.0	1.6±0.3	4.0±0.4	59.8±2.1	3.6±2.9	18.4±1.2
thrb	2.2±0.9	2.8±0.4	2.9±0.2	5.9±0.2	75.0±0.3	12.3±1.0	26.9±0.4
try1	2.0±2.8	2.1±1.2	2.5±0.0	7.5±2.7	71.3±1.3	8.3±2.8	24.9±1.0
urok	2.0±2.8	3.1±1.5	4.1±0.5	8.5±1.0	81.3±0.3	15.3±4.5	30.5±1.6

Table 30: **TopU-LBVS-Full GIN+FP hybrid model results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	13.2±0.6	7.4±0.4	4.9±0.4	15.7±0.2	77.6±1.0	33.8±0.9	38.0±0.5
cp2c9	14.2±1.4	7.6±0.3	5.3±0.1	17.1±0.6	83.1±0.4	35.8±0.5	40.6±0.5
cp2d6	15.8±0.0	9.2±0.5	5.8±0.3	22.2±1.9	82.4±0.4	41.8±1.6	43.5±0.8
cp3a4	15.6±0.8	9.0±0.3	6.2±0.2	22.8±1.6	82.6±0.8	43.2±1.8	44.0±0.9
GPCR							
5ht6r	3.4±1.0	3.9±0.2	3.5±0.4	8.0±0.3	77.6±0.1	18.2±0.9	30.4±0.5
aa2ar	5.2±1.6	3.7±0.5	3.5±0.2	7.4±0.6	75.2±0.4	17.1±2.1	29.5±0.9
aa3r	5.2±0.4	3.4±0.4	3.2±0.2	8.0±0.6	75.9±0.3	15.6±1.5	29.7±0.6
adrb1	2.6±0.7	1.8±0.4	1.7±0.2	4.8±0.1	69.2±0.5	6.0±0.5	23.2±0.3
adrb2	2.3±0.5	3.2±0.2	3.2±0.0	6.2±0.3	70.5±0.2	13.3±0.5	26.2±0.2
cnr2	4.2±1.7	3.7±0.9	3.4±0.5	8.5±1.7	80.6±1.5	17.8±5.2	31.3±2.6
cxcr4	3.9±2.8	2.6±0.0	2.2±0.3	6.4±0.3	71.8±0.8	9.4±1.1	25.6±0.6
drd2	4.4±0.7	3.5±0.1	2.9±0.3	6.8±0.1	76.0±0.6	15.1±0.4	28.7±0.2
drd3	7.0±0.0	4.4±0.3	4.4±0.2	10.8±0.5	83.6±0.6	22.3±1.0	34.8±0.4
hrh3	9.6±0.7	6.6±0.5	5.1±0.1	14.2±1.1	81.9±0.0	30.7±1.2	38.0±0.3
oprd	7.0±1.2	4.4±0.2	4.3±0.3	9.1±0.2	80.1±0.4	21.3±0.5	32.8±0.1
oprk	3.5±0.0	4.0±0.0	3.3±0.1	7.3±0.1	77.5±0.6	16.4±0.6	29.7±0.1
oprm	5.2±1.2	4.7±0.2	4.2±0.1	9.7±0.6	79.3±0.8	22.4±0.8	33.3±0.7
Ion Channel							
5ht3a	4.6±0.9	6.8±0.4	5.3±0.2	12.1±0.4	82.6±0.7	29.5±1.1	36.6±0.4
cac1h	11.4±1.6	9.3±0.9	7.8±0.3	23.0±1.6	88.0±1.2	48.2±2.6	46.6±1.5
kcnh2	2.4±1.3	2.7±0.1	2.6±0.2	5.5±0.3	72.3±0.4	10.7±0.6	25.7±0.4
scn5a	3.0±1.1	2.6±0.2	2.5±0.3	5.6±0.4	70.2±0.5	10.3±1.2	24.9±0.7
trpa1	13.7±7.7	7.2±1.8	4.7±0.5	22.4±4.6	69.5±1.4	33.9±5.9	35.4±3.3
Kinase							
abl1	1.0±0.0	4.9±0.7	4.2±0.3	7.8±0.4	75.1±0.7	19.6±1.5	29.8±0.5
akt1	7.7±0.0	6.6±0.2	5.1±0.1	12.8±1.0	79.9±0.5	30.7±1.8	36.8±1.0
akt2	7.5±2.6	6.6±0.7	5.1±0.4	15.5±1.6	80.0±1.0	31.8±3.4	37.6±1.5
braf	6.3±1.6	5.2±1.0	4.6±0.2	11.6±1.4	84.3±0.5	24.6±2.3	35.6±1.3
cdk2	5.1±0.0	5.2±0.4	4.7±0.1	9.8±0.7	78.9±0.3	24.2±1.9	33.5±0.9
csf1r	7.9±3.7	4.0±0.2	3.3±0.5	10.0±0.6	78.6±0.4	20.3±1.5	32.5±0.7
egfr	3.9±0.5	4.0±0.3	3.9±0.1	8.2±0.3	80.3±0.2	18.5±1.0	31.7±0.5
fak1	10.8±3.1	6.6±0.2	4.8±0.2	15.4±2.5	76.3±0.3	30.2±1.5	36.1±1.1
fgfr1	9.0±1.0	6.2±0.4	4.8±0.3	12.1±0.4	78.8±0.5	28.8±1.5	35.5±0.6
igf1r	7.8±1.8	6.2±0.4	5.6±0.1	12.7±0.5	81.0±0.3	30.2±0.9	36.8±0.6
jak2	6.2±1.5	5.1±0.4	4.7±0.3	11.0±0.5	80.9±0.5	25.6±0.9	35.2±0.4
kit	3.1±0.9	3.1±0.7	3.3±0.2	6.5±0.2	72.0±0.7	13.2±0.6	27.1±0.2
kpcb	3.7±0.0	5.1±0.4	4.1±0.2	9.3±0.3	78.0±0.5	21.7±0.7	31.7±0.3
lck	5.6±1.0	4.9±0.5	3.9±0.5	8.7±1.1	74.8±1.2	21.2±2.4	30.7±1.5
mapk2	4.9±1.3	6.1±0.0	4.7±0.2	10.9±0.5	75.4±0.5	27.2±1.1	33.4±0.5
met	5.1±1.8	5.3±0.6	5.0±0.1	10.8±1.7	74.6±1.0	26.3±2.9	33.1±1.4
mk01	25.4±0.7	11.4±0.4	6.9±0.1	39.7±1.2	86.2±1.8	56.8±1.4	54.7±0.7
mk10	12.0±1.2	4.9±0.4	3.9±0.3	12.2±0.1	76.7±0.1	23.7±1.3	33.9±0.4
mk14	5.3±1.9	6.7±0.0	5.2±0.3	12.0±0.9	79.8±0.1	29.1±0.6	35.7±0.4
mp2k1	6.5±1.8	4.1±0.6	3.4±0.3	8.1±0.6	70.8±1.2	18.8±1.7	29.0±0.7
plk1	20.0±0.7	11.1±0.1	6.4±0.1	32.3±0.2	84.2±0.4	51.8±0.8	49.4±0.4
rock1	2.9±2.0	3.6±0.4	3.9±0.3	8.9±0.9	74.6±0.6	17.4±1.5	29.5±0.4
src	6.7±0.5	4.9±0.4	4.0±0.2	10.0±1.0	78.1±0.8	22.3±1.7	32.8±0.8
tgfr1	1.4±1.9	4.2±0.8	3.8±0.2	9.7±1.3	84.0±0.3	17.2±2.8	32.9±1.6
vgfr2	8.0±1.5	5.3±0.1	4.6±0.2	11.8±0.7	81.2±0.3	25.2±1.3	35.3±0.7
wee1	8.2±0.0	4.9±1.3	4.1±0.7	13.2±2.4	80.5±1.8	22.8±6.3	33.9±2.8

Table 31: **TopU-LBVS-Full GIN+FP hybrid model results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	4.4±0.0	5.3±0.6	4.2±0.1	9.8±0.7	82.7±1.4	22.9±1.8	34.2±1.2
esr1	3.3±0.8	3.9±0.0	3.6±0.2	7.1±0.3	72.2±0.7	17.5±1.0	28.4±0.5
esr2	3.6±1.5	3.7±0.7	3.1±0.4	7.6±0.9	75.1±0.8	16.5±2.3	29.3±1.2
gcr	10.7±0.0	6.1±0.0	5.5±0.3	14.0±0.5	85.5±0.1	31.2±0.8	39.0±0.4
mcr	0.0±0.0	2.6±0.9	2.7±0.5	7.3±0.6	76.9±1.4	11.8±1.8	27.1±1.2
ppara	0.0±0.0	1.5±0.0	2.5±0.2	5.8±0.2	76.7±0.3	7.3±0.1	26.1±0.3
ppard	2.7±0.0	4.6±0.5	4.8±0.5	11.4±1.1	82.1±0.3	23.2±1.7	34.5±0.7
pparg	4.6±0.6	4.3±0.6	4.2±0.2	8.7±0.4	79.8±0.2	19.9±1.4	32.2±0.5
prgr	6.5±3.0	5.7±0.2	5.2±0.2	11.6±0.9	82.6±0.8	27.1±1.2	35.7±0.9
rxra	3.0±2.1	6.3±1.2	5.5±0.4	13.0±0.3	85.5±0.5	28.8±1.2	37.4±0.1
thb	30.4±2.1	14.3±0.5	8.7±0.5	60.9±4.8	94.1±0.8	74.3±1.9	68.3±2.1
Other Enzymes							
aces	5.1±0.7	4.4±0.5	3.9±0.2	8.7±0.3	78.7±0.8	20.5±1.2	32.1±0.3
ampc	27.3±3.9	15.3±3.4	8.5±1.4	54.8±6.6	96.5±2.2	70.7±8.3	62.8±4.6
aofb	4.9±2.5	3.6±0.6	3.5±0.2	7.9±0.5	77.8±0.3	16.7±1.8	30.3±0.8
cah2	3.7±0.7	2.8±0.4	3.1±0.1	8.2±0.4	81.3±0.9	12.5±1.1	29.9±0.4
dhi1	2.3±3.2	2.8±1.2	1.6±0.3	5.8±0.5	72.3±1.8	6.9±1.3	23.5±0.9
dyr	3.4±2.4	3.5±0.5	3.0±0.7	8.2±1.8	73.5±1.0	13.3±1.3	27.1±0.4
fnta	8.2±0.0	2.7±1.5	3.0±0.4	12.1±1.7	77.6±0.6	14.9±1.9	30.6±0.6
glcm	15.2±4.3	11.8±1.2	7.9±0.4	32.6±3.5	91.3±0.7	56.5±3.0	52.3±1.9
hdac2	2.3±1.6	2.8±0.0	2.9±0.3	7.4±0.7	79.1±0.8	12.4±1.7	29.1±1.0
hdac8	0.0±0.0	0.9±0.3	1.1±0.2	4.2±0.2	72.7±0.8	-0.7±1.1	20.9±0.6
hivint	0.7±1.0	1.1±0.5	2.3±0.0	5.5±0.3	75.2±0.6	5.9±1.8	25.1±0.8
hivrt	4.8±1.2	2.9±0.0	2.8±0.4	6.3±0.5	73.4±0.4	12.8±1.7	26.9±0.9
hmdh	16.4±0.0	7.1±0.0	5.3±0.7	21.5±2.8	85.1±2.6	36.6±1.3	41.8±1.3
hxk4	5.5±3.9	5.2±0.9	4.0±0.8	11.7±1.3	84.9±0.6	20.4±4.6	33.9±1.4
inha	6.8±5.6	4.7±1.3	5.7±1.0	14.7±1.0	90.2±1.0	27.9±4.2	39.2±1.4
nos1	5.1±4.2	5.6±0.5	5.6±0.5	14.7±2.2	86.4±0.3	29.3±3.2	38.8±1.9
parp1	10.6±2.1	3.7±0.5	3.0±0.4	9.9±0.6	74.2±0.6	17.0±1.3	29.7±0.6
pde5a	7.8±3.7	5.7±1.2	4.5±0.3	16.2±3.1	84.5±1.0	26.6±0.8	37.0±0.9
pgh1	3.7±2.0	3.8±0.4	3.4±0.3	7.2±0.8	75.8±1.3	16.3±2.6	29.0±1.3
pgh2	3.9±2.8	4.8±1.6	4.2±0.3	9.6±0.9	78.7±1.1	20.1±3.7	31.4±1.1
ptn1	5.6±0.7	3.9±0.7	3.9±0.3	10.6±1.0	76.5±0.3	18.8±1.8	31.2±0.8
pygm	8.2±0.0	9.1±2.4	5.7±0.5	17.4±3.4	74.4±3.1	35.9±6.6	36.9±3.5
pyrd	2.5±1.8	3.9±0.6	4.1±0.4	9.6±0.7	82.3±0.7	19.8±3.8	32.8±1.1
tysy	0.0±0.0	0.0±0.0	1.0±0.0	4.8±0.2	70.4±1.2	-0.7±0.7	21.0±0.5
Protease							
ace	13.7±3.9	7.2±0.9	5.0±0.0	20.0±3.3	78.0±0.4	35.2±2.0	37.7±1.0
ada17	3.4±4.8	3.0±0.5	3.5±0.3	9.1±1.5	85.0±0.6	14.3±4.0	31.8±2.0
bace1	1.8±1.5	2.1±0.2	2.2±0.2	5.7±0.4	76.1±0.6	7.7±1.7	26.0±0.8
dpp4	5.3±3.0	6.2±0.6	4.5±0.5	11.4±0.7	85.0±0.8	26.0±1.2	35.6±0.8
fa10	2.3±1.9	2.6±0.2	3.3±0.1	7.2±0.2	81.7±0.9	12.7±0.8	29.6±0.5
hivpr	3.7±0.0	3.2±0.6	3.1±0.5	6.7±0.0	77.0±0.5	12.4±0.2	27.5±0.1
mmp13	0.0±0.0	0.8±1.1	1.6±0.3	3.7±0.2	63.5±0.4	0.2±2.6	18.3±0.8
thrb	3.5±1.6	3.6±0.5	3.5±0.3	7.5±0.7	78.4±1.0	16.5±2.0	30.1±1.2
try1	5.9±0.0	4.6±0.6	3.9±0.3	9.2±0.8	79.0±1.1	18.4±1.2	31.0±1.0
urok	5.9±4.8	6.2±0.6	4.9±0.7	13.6±2.0	86.1±0.4	28.2±2.3	37.3±1.7

Table 32: **TopU-LBVS-Full GPS results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	3.2±1.4	2.9±0.9	2.4±0.6	4.9±0.9	66.1±3.2	10.5±4.3	23.3±2.4
cp2c9	3.1±0.4	2.2±0.1	1.9±0.1	4.7±0.8	66.7±2.6	7.7±0.9	22.6±1.4
cp2d6	8.9±0.3	6.0±0.1	4.1±0.1	11.1±0.3	78.0±0.6	26.6±0.3	34.4±0.4
cp3a4	5.8±1.9	4.5±0.1	3.7±0.2	8.9±0.7	77.8±1.0	19.9±1.4	31.5±1.1
GPCR							
5ht6r	0.7±1.0	0.8±0.3	1.5±0.3	4.6±0.1	72.7±0.7	2.6±0.2	22.9±0.3
aa2ar	1.0±0.4	1.3±0.2	2.0±0.3	4.2±0.1	68.3±0.4	4.6±1.0	21.9±0.3
aa3r	3.8±1.0	2.1±0.3	1.9±0.3	5.1±0.4	68.0±0.5	7.2±1.3	23.1±0.6
adrb1	1.1±0.7	2.0±0.4	1.9±0.0	4.3±0.5	62.9±1.4	5.2±1.3	20.9±0.6
adrb2	1.0±0.8	1.5±0.2	1.7±0.1	4.0±0.2	64.6±2.2	4.6±0.3	20.9±0.9
cnr2	0.0±0.0	1.2±0.7	1.5±0.3	4.2±0.2	69.9±0.8	2.2±1.4	21.4±0.6
cxcr4	0.0±0.0	1.3±0.0	1.3±0.0	3.9±0.2	63.2±1.6	2.0±0.7	18.8±0.7
drd2	0.9±0.7	1.6±0.4	1.8±0.2	4.3±0.1	70.0±0.1	4.5±0.8	22.3±0.4
drd3	7.0±1.9	3.4±0.4	2.7±0.2	7.4±0.7	75.1±0.9	15.5±1.7	28.9±1.0
hrh3	3.0±1.2	3.7±0.2	4.1±0.3	9.1±0.5	81.0±1.2	20.3±1.4	32.9±0.8
oprd	2.2±0.4	2.4±0.1	2.6±0.2	5.9±0.6	73.6±0.4	10.0±0.8	26.1±0.4
oprk	2.3±0.0	2.5±0.3	2.3±0.1	5.2±0.1	71.3±1.0	9.3±1.2	24.7±0.3
oprm	3.2±1.1	3.1±0.3	2.9±0.3	6.5±0.3	74.8±0.3	13.8±1.1	28.0±0.6
Ion Channel							
5ht3a	3.9±4.2	4.6±1.0	3.7±0.1	9.4±1.1	77.5±1.6	20.0±4.0	31.3±1.9
cac1h	9.1±1.6	8.1±0.6	5.5±0.3	16.8±0.9	80.3±2.4	33.9±2.7	38.1±1.8
kcnh2	2.5±0.4	2.2±0.3	1.9±0.1	4.4±0.2	67.5±0.8	7.1±1.2	22.5±0.6
scn5a	0.0±0.0	0.8±0.2	0.7±0.3	2.9±0.1	58.2±1.4	-1.9±2.0	16.0±0.7
trpa1	2.7±3.9	3.9±0.0	3.3±0.5	8.6±0.5	72.1±3.3	17.0±0.5	27.8±1.3
Kinase							
abl1	3.8±1.3	2.3±0.2	2.2±0.1	5.5±0.7	67.0±0.4	8.7±0.5	23.5±0.3
akt1	7.3±1.6	4.8±0.5	3.6±0.1	8.7±0.6	74.0±1.1	21.5±0.7	30.5±0.8
akt2	8.2±2.9	5.2±0.5	3.8±0.3	9.7±0.4	71.4±2.3	22.4±2.1	30.3±1.5
braf	2.3±0.8	2.1±0.3	1.7±0.2	4.7±0.2	69.6±1.5	6.1±0.5	22.8±0.4
cdk2	3.1±1.4	1.9±0.6	2.0±0.3	5.3±1.3	67.8±2.9	7.4±3.0	23.1±2.2
csf1r	4.3±0.0	3.0±0.6	2.5±0.1	6.9±0.9	72.3±0.2	11.8±1.2	26.7±0.6
egfr	2.6±0.5	2.7±0.4	2.7±0.4	6.1±0.4	73.7±1.9	11.0±2.5	26.4±1.4
fak1	5.8±1.0	3.9±0.6	2.8±0.5	7.4±1.4	73.5±2.0	15.1±3.1	28.1±2.2
fgfr1	5.1±0.0	3.7±0.7	2.8±0.5	7.7±0.8	70.7±1.1	15.0±3.2	27.1±1.5
igf1r	2.9±2.1	3.6±0.2	3.2±0.1	6.3±0.8	70.0±2.6	14.8±2.2	26.6±1.6
jak2	4.3±1.5	3.3±0.5	2.7±0.4	6.3±0.3	71.3±1.1	13.1±1.2	26.4±0.6
kit	1.9±1.5	1.6±0.5	1.8±0.1	4.0±0.4	62.2±1.4	5.3±1.7	20.1±0.8
kpcb	3.7±0.0	4.6±0.0	2.7±0.3	6.5±0.5	67.8±3.5	15.5±1.1	25.8±1.4
lck	3.2±1.5	2.7±0.5	2.3±0.3	5.4±1.0	66.8±0.2	10.1±2.3	23.5±1.0
mapk2	6.0±0.8	4.5±0.3	3.0±0.4	7.1±0.3	67.1±0.7	17.5±0.4	26.6±0.5
met	5.6±0.6	4.2±0.7	3.4±0.7	7.7±0.5	71.1±2.3	17.4±3.1	28.4±1.6
mk01	14.2±1.4	8.2±0.7	5.1±0.5	17.9±1.2	81.8±1.0	36.7±2.9	40.7±1.3
mk10	4.3±2.4	2.3±0.3	2.4±0.2	5.6±0.4	71.5±1.6	9.4±0.9	25.0±0.7
mk14	4.6±0.0	4.2±0.4	4.0±0.1	8.2±0.3	76.6±0.0	19.6±0.9	30.9±0.3
mp2k1	2.2±1.8	2.9±0.6	2.3±0.2	4.6±0.5	58.0±1.7	9.3±1.2	20.6±0.5
plk1	6.8±1.8	6.2±0.1	3.9±0.2	11.9±1.3	78.8±1.4	27.2±1.2	34.8±1.1
rock1	3.6±1.0	2.7±0.4	2.7±0.7	6.5±0.3	69.8±0.9	12.1±1.8	26.0±0.5
src	3.3±0.9	2.6±0.6	2.4±0.5	5.2±0.3	68.8±1.0	9.8±2.2	24.2±0.9
tgfr1	0.0±0.0	0.6±0.8	1.7±0.6	4.3±0.2	69.4±0.9	1.3±2.5	20.9±0.8
vgfr2	4.7±0.4	3.1±0.4	2.3±0.3	6.2±0.4	69.5±1.4	11.7±1.5	25.2±1.0
wee1	5.5±3.9	2.2±0.8	1.6±0.0	6.4±0.6	62.1±1.7	9.1±2.9	22.5±1.0

Table 33: **TopU-LBVS-Full GPS results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	2.4±2.5	2.2±0.8	2.5±0.5	6.0±0.9	74.8±1.7	9.5±4.1	26.3±2.1
esr1	2.4±0.8	2.0±0.1	2.1±0.1	5.3±0.2	70.1±1.3	7.1±0.8	23.7±0.7
esr2	2.4±0.8	1.7±0.6	1.7±0.1	4.8±0.5	69.1±1.6	5.1±0.9	22.3±0.7
gcr	3.0±0.8	4.8±0.4	4.5±0.4	10.6±0.1	84.7±1.3	22.5±2.2	34.8±1.0
mcr	2.7±3.9	2.6±2.4	3.0±0.0	8.3±3.2	73.9±2.4	12.4±7.0	26.7±4.0
ppara	0.8±1.1	2.3±0.6	1.9±0.2	4.0±0.3	63.1±2.7	6.1±1.2	20.2±1.2
ppard	4.6±1.3	2.7±1.3	2.8±0.2	7.5±0.5	73.8±1.3	11.7±2.6	26.9±1.2
pparg	0.5±0.6	2.2±0.7	2.2±0.2	5.1±0.4	72.1±1.1	7.9±2.1	24.4±1.1
prgr	6.5±2.2	3.5±0.3	3.7±0.2	10.8±0.1	81.4±1.5	18.9±0.7	33.1±0.5
rxra	0.0±0.0	2.7±0.9	3.2±0.6	7.0±0.6	78.7±1.6	11.7±2.2	28.1±1.2
thb	21.3±4.3	13.3±0.5	7.7±0.2	38.2±2.8	89.4±1.3	60.8±3.1	55.7±1.5
Other Enzymes							
aces	3.4±1.8	2.9±0.3	2.6±0.2	6.1±0.7	74.6±0.7	11.5±1.8	26.8±1.1
ampc	24.6±6.7	12.0±1.5	6.8±0.8	43.9±2.7	91.0±1.4	57.6±3.9	54.5±2.2
aofb	2.0±1.4	1.8±0.2	1.8±0.0	4.6±0.1	70.3±2.1	6.2±1.1	23.0±0.4
cah2	3.7±0.7	2.1±0.5	1.6±0.2	5.4±0.6	71.8±0.6	6.0±1.7	23.8±0.8
dhi1	0.0±0.0	2.4±0.7	1.9±0.3	4.8±0.1	67.4±1.1	6.6±0.7	21.5±0.2
dyr	0.0±0.0	0.0±0.0	1.1±0.4	3.8±0.0	66.0±1.3	-1.2±1.3	19.0±0.2
fnta	0.0±0.0	1.6±0.0	3.3±0.0	5.8±0.3	72.2±0.8	7.6±0.6	24.2±0.7
glcm	7.6±2.1	4.0±1.2	4.4±0.8	13.4±1.5	75.3±3.7	23.2±4.9	32.8±3.0
hdac2	3.4±2.8	1.9±0.7	2.5±0.5	5.9±0.6	73.8±3.1	9.0±1.7	25.6±1.5
hdac8	2.3±1.6	1.1±0.3	1.3±0.3	5.1±1.4	64.8±0.2	2.5±1.9	20.0±0.9
hivint	0.0±0.0	2.3±0.8	2.2±0.1	4.9±0.1	70.0±0.7	7.7±1.3	23.5±0.3
hivrt	2.6±0.0	2.1±0.6	2.0±0.1	4.7±0.3	70.5±1.2	6.7±2.2	23.3±1.1
hmdh	2.7±3.9	4.2±0.8	2.7±0.0	8.6±1.1	75.5±1.3	16.4±2.4	28.5±1.1
hxk4	2.7±3.9	2.0±1.6	3.0±0.0	9.9±4.6	79.0±0.6	9.6±4.8	27.7±2.5
inha	0.0±0.0	3.8±0.7	3.3±0.3	7.4±1.4	74.4±8.0	14.4±2.4	27.7±3.2
nos1	3.4±2.4	2.1±0.9	4.4±0.5	8.6±0.5	78.5±0.5	15.0±2.1	30.1±0.7
parp1	0.0±0.0	0.3±0.5	0.8±0.2	3.4±0.2	64.3±1.8	-2.0±1.2	17.7±0.9
pde5a	1.0±1.4	3.3±1.7	2.9±0.4	7.3±1.1	80.4±1.0	13.7±4.7	29.4±2.0
pgh1	2.1±0.7	1.8±0.4	1.3±0.1	4.0±0.5	57.6±1.8	3.7±1.1	18.1±0.7
pgh2	3.9±2.8	4.4±0.6	2.9±0.8	6.4±0.5	66.8±5.2	13.1±1.0	24.1±2.1
ptn1	2.5±0.7	2.8±0.4	2.0±0.1	4.9±0.2	67.4±0.5	9.1±0.5	23.3±0.3
pygm	2.7±3.9	1.3±0.9	2.0±0.0	4.4±0.6	60.1±1.9	3.9±2.7	18.5±0.8
pyrd	1.2±1.8	1.8±0.4	2.3±0.4	5.2±0.3	72.1±2.6	6.6±1.2	24.0±0.9
tysy	2.7±3.9	2.6±0.9	1.3±0.5	4.6±1.2	57.4±2.8	5.0±4.5	17.9±2.8
Protease							
ace	0.0±0.0	1.3±0.9	3.0±0.8	5.9±0.5	70.7±0.4	7.0±3.6	23.7±1.1
ada17	0.0±0.0	2.2±0.9	2.8±0.0	6.7±0.7	77.0±1.5	10.4±3.0	27.1±1.5
bace1	0.6±0.8	1.8±0.3	2.2±0.2	5.3±0.1	75.2±0.6	6.2±0.9	24.9±0.2
dpp4	1.1±1.5	3.0±1.6	2.9±0.9	6.2±1.2	74.5±2.8	10.6±4.8	26.3±2.5
fa10	0.8±1.1	1.4±0.8	2.1±0.2	4.6±0.2	71.6±1.1	4.7±1.4	22.7±0.9
hivpr	0.0±0.0	1.1±0.7	1.1±0.5	4.6±0.4	71.4±1.1	1.7±2.9	22.0±1.3
mmp13	2.0±2.8	1.6±0.6	1.2±0.0	3.8±0.5	62.5±2.1	1.3±2.6	18.4±1.3
thrb	2.5±1.6	2.5±0.6	2.2±0.3	5.4±0.5	72.6±0.9	9.3±2.1	25.2±1.1
try1	0.0±0.0	0.4±0.6	1.0±0.6	4.2±0.3	68.0±1.5	0.2±2.3	20.1±1.3
urok	0.0±0.0	2.0±1.5	2.3±1.0	5.7±1.4	72.4±4.8	6.8±5.2	24.1±3.5

Table 34: textbfTopU-LBVS-Full GPS+FP hybrid model results - part 1. All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	13.2±1.8	7.1±0.6	5.0±0.3	15.3±1.2	77.8±0.4	33.4±1.4	37.7±0.6
cp2c9	15.1±1.2	7.8±0.5	5.3±0.1	17.5±0.8	83.2±0.1	36.9±0.9	40.9±0.4
cp2d6	15.1±0.6	8.4±0.6	5.4±0.4	19.2±1.6	81.8±0.9	38.5±2.5	41.4±1.4
cp3a4	16.4±0.3	8.9±0.3	6.2±0.3	24.0±0.4	82.5±0.6	43.2±0.3	44.3±0.3
GPCR							
5ht6r	5.5±1.0	4.6±0.9	3.8±0.2	8.4±0.6	77.4±0.2	19.9±2.3	30.8±0.8
aa2ar	5.7±1.0	4.0±0.2	3.5±0.2	8.3±0.6	74.7±0.2	18.2±0.4	29.9±0.2
aa3r	4.4±0.8	3.2±0.3	3.4±0.3	7.6±0.4	75.7±0.1	15.7±1.7	29.6±0.8
adrb1	3.2±1.3	2.0±0.2	1.6±0.1	5.0±0.3	68.9±1.0	6.3±0.5	23.3±0.2
adrb2	3.9±1.4	2.7±0.4	3.2±0.1	6.6±0.8	70.4±0.7	12.9±1.5	26.4±0.8
cnr2	3.6±1.5	4.5±0.2	3.7±0.2	8.1±0.3	81.4±1.0	18.1±0.9	31.2±0.7
cxcr4	3.9±2.8	2.2±0.6	2.9±0.6	6.7±0.0	72.5±0.5	10.4±0.5	26.3±0.2
drd2	4.1±1.5	3.2±0.0	3.0±0.1	6.9±0.1	76.2±0.9	14.4±0.2	28.8±0.3
drd3	7.8±0.6	5.4±0.6	4.5±0.1	12.8±0.9	84.0±0.6	25.1±1.4	36.2±0.8
hrh3	8.6±1.4	6.2±0.4	5.1±0.1	13.5±0.8	81.3±0.2	30.0±1.5	37.3±0.8
oprd	4.8±1.6	4.1±0.1	3.8±0.3	8.9±0.2	79.5±0.3	19.0±0.5	31.8±0.1
oprk	3.5±1.0	3.4±0.4	3.1±0.6	7.1±0.5	78.3±1.3	14.8±2.1	29.4±1.1
oprm	4.1±1.1	5.1±0.1	4.5±0.0	10.4±0.4	80.4±0.2	23.0±0.4	33.8±0.2
Ion Channel							
5ht3a	3.3±2.4	6.9±0.3	5.2±0.2	11.6±0.2	82.3±0.4	28.8±0.5	35.9±0.2
cac1h	8.0±1.6	10.5±0.3	7.8±0.3	23.5±0.8	87.7±0.7	48.7±0.3	46.8±0.5
kcnh2	1.9±0.1	2.4±0.2	2.5±0.2	5.3±0.2	73.0±0.4	9.5±0.8	25.5±0.5
scn5a	3.0±1.1	2.9±0.2	2.3±0.2	5.7±0.3	70.9±0.4	10.9±1.0	25.4±0.5
trpa1	19.1±3.9	7.8±0.0	4.7±0.5	27.5±2.9	72.2±1.0	36.6±1.3	37.5±1.7
Kinase							
abl1	1.7±0.5	5.0±0.3	3.9±0.0	7.6±0.5	74.7±0.3	19.0±1.8	29.6±0.8
akt1	8.1±1.6	6.5±0.4	5.1±0.1	13.8±0.8	80.5±1.0	30.9±0.3	36.9±0.3
akt2	7.5±1.9	6.9±0.2	5.0±0.2	15.1±1.5	80.2±0.4	32.6±1.2	37.7±0.8
braf	6.8±2.4	6.1±0.3	4.7±0.2	13.2±1.2	85.2±0.5	28.3±1.6	37.2±0.8
cdk2	4.8±0.5	5.4±0.3	4.3±0.3	9.8±0.2	78.8±0.2	23.4±1.2	33.2±0.3
csf1r	8.6±3.1	3.9±0.6	3.5±0.3	11.4±1.6	79.0±0.4	20.2±2.0	33.1±0.9
egfr	3.9±0.5	4.1±0.3	3.9±0.2	8.3±0.1	79.8±0.3	19.1±0.5	31.8±0.0
fak1	13.7±3.7	6.6±0.4	4.9±0.2	19.3±2.0	77.1±0.2	32.3±1.4	38.1±1.3
fgfr1	9.8±1.2	6.0±0.2	4.5±0.2	13.0±0.2	77.5±0.6	28.3±0.5	35.6±0.2
igf1r	8.8±0.0	6.2±0.6	5.7±0.4	13.4±1.2	81.0±0.4	30.8±2.0	37.3±1.0
jak2	4.7±1.0	5.0±0.6	4.8±0.3	10.8±1.0	80.9±0.5	24.3±1.6	34.4±0.7
kit	3.7±0.0	3.1±0.2	3.6±0.2	6.9±0.2	73.4±1.6	15.2±1.3	27.9±0.8
kpcb	3.7±0.0	5.6±0.7	4.1±0.2	9.3±0.0	78.0±0.7	21.4±0.5	31.6±0.3
lck	4.6±1.3	4.9±0.4	4.4±0.3	9.0±0.4	75.7±0.6	23.0±0.9	31.7±0.4
mapk2	4.9±1.3	6.2±0.7	4.6±0.1	10.3±0.5	75.3±0.6	26.3±1.2	33.0±0.6
met	6.0±1.6	5.8±0.4	4.7±0.3	10.9±1.4	74.2±0.3	25.5±1.8	32.7±0.8
mk01	24.9±3.2	12.0±0.2	7.0±0.2	39.0±1.7	86.4±1.4	57.3±0.3	54.6±0.7
mk10	10.2±0.0	5.0±0.3	3.5±0.4	13.0±0.7	76.4±0.4	24.5±1.3	34.1±0.7
mk14	3.8±1.1	6.6±0.1	5.4±0.1	11.7±0.5	79.6±0.4	29.4±0.9	35.7±0.7
mp2k1	7.2±1.0	3.8±0.4	4.0±0.1	8.6±0.6	72.4±1.0	19.7±0.9	30.0±0.6
plk1	22.0±1.2	11.0±0.1	6.3±0.1	33.8±1.0	83.7±0.4	52.6±0.6	50.3±0.5
rock1	4.3±3.1	3.6±0.6	3.5±0.4	7.8±1.1	74.4±0.8	16.8±3.4	29.2±1.6
src	5.7±0.5	5.6±0.2	4.4±0.2	10.1±0.1	79.3±0.3	25.3±0.5	33.8±0.1
tgfr1	4.1±0.0	4.2±0.8	3.2±0.6	10.5±0.5	83.9±1.3	18.5±3.5	33.5±1.2
vgfr2	9.4±1.3	4.8±0.2	4.3±0.3	11.8±0.4	81.1±0.5	24.8±0.9	35.4±0.2
wee1	10.9±3.9	5.5±1.5	4.4±0.4	15.3±3.0	81.4±0.7	27.5±4.5	36.3±2.0

Table 35: textbfTopU-LBVS-Full GPS+FP hybrid model results - part 2. All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	3.9±0.7	6.1±0.3	4.5±0.2	10.7±0.4	82.7±0.7	25.4±0.9	35.0±0.5
esr1	3.3±1.5	4.1±0.4	3.7±0.2	7.0±0.5	71.0±0.9	17.7±1.2	28.2±0.9
esr2	5.9±2.2	4.0±0.2	3.5±0.7	10.3±0.9	75.7±0.7	19.2±0.8	30.7±0.2
gcr	5.9±0.8	7.0±0.7	5.5±0.2	14.5±1.3	85.9±0.5	32.5±3.0	39.3±1.4
mcr	2.7±3.9	3.9±0.0	3.7±0.5	7.8±0.5	77.7±0.8	13.6±1.4	28.0±0.7
ppara	1.6±1.1	2.0±0.4	2.9±0.2	6.1±0.1	77.0±0.7	9.2±0.3	26.9±0.3
ppard	5.5±2.2	4.0±0.5	4.6±0.2	10.1±1.3	81.4±1.0	21.5±2.6	33.7±1.7
pparg	4.1±1.9	3.5±0.5	4.1±0.4	8.4±1.0	79.4±1.2	17.6±3.1	31.3±1.6
prgr	7.1±0.0	6.0±0.7	5.1±0.1	11.8±0.1	82.6±0.6	27.5±1.1	36.0±0.4
rxra	4.6±3.7	6.0±0.8	6.3±0.6	14.9±2.7	85.9±0.4	29.6±1.8	38.3±1.2
thb	33.4±7.7	14.7±0.5	8.8±0.2	66.8±5.2	94.5±0.5	76.6±1.7	71.7±3.5
Other Enzymes							
aces	6.8±1.4	4.3±0.2	4.0±0.1	9.2±0.2	79.6±0.5	20.8±0.7	32.8±0.2
ampc	24.6±0.0	11.5±1.3	7.7±0.8	48.0±1.6	92.1±4.2	61.1±3.9	57.0±3.0
aofb	3.4±1.4	3.7±0.5	3.8±0.1	8.2±0.9	77.5±1.0	18.1±2.2	30.5±1.4
cah2	3.7±0.7	2.5±0.5	2.9±0.2	8.1±0.4	81.2±1.1	12.9±1.9	30.2±1.0
dhi1	4.6±3.2	1.9±0.7	1.6±0.3	7.6±3.0	72.2±2.0	6.3±3.4	23.8±2.2
dyr	1.7±2.4	3.5±0.5	3.2±0.9	6.6±0.7	74.0±1.5	12.1±2.0	26.6±1.5
fnta	2.7±3.9	2.2±0.8	3.6±0.8	8.6±1.4	75.3±1.5	14.0±3.6	28.5±1.0
glcm	15.2±4.3	10.6±0.4	7.5±0.0	28.8±3.3	90.1±0.4	50.2±1.4	49.2±1.3
hdac2	4.6±1.6	4.0±0.7	3.3±0.4	9.7±1.9	80.1±1.2	18.2±2.2	31.5±1.4
hdac8	0.0±0.0	0.7±0.0	1.1±0.2	4.3±0.1	72.7±0.5	-0.6±0.8	21.1±0.4
hivint	0.0±0.0	1.4±0.5	2.6±0.1	5.4±0.1	75.4±1.0	5.9±1.5	25.1±0.5
hivrt	3.1±1.2	3.3±0.5	2.7±0.1	5.9±0.0	73.5±0.3	12.2±0.6	26.4±0.1
hmdh	8.2±0.0	7.1±0.0	5.1±0.4	21.2±3.0	84.3±2.3	34.0±2.5	40.0±1.9
hxk4	8.2±0.0	4.6±0.9	4.3±0.5	10.9±0.7	84.3±1.7	21.2±1.5	33.4±1.3
inha	6.8±0.0	7.1±1.2	4.5±0.9	15.9±0.7	90.1±1.2	32.3±2.3	40.5±1.2
nos1	6.8±4.8	6.3±0.0	5.8±0.4	14.5±1.6	86.7±0.2	29.2±2.3	38.7±1.5
parp1	6.1±2.1	3.0±0.0	2.5±0.4	9.1±2.7	73.7±1.0	14.0±1.9	28.0±1.8
pde5a	6.8±2.8	6.1±1.0	4.3±0.1	12.3±2.0	84.4±0.9	26.2±3.5	36.5±2.0
pgh1	3.2±2.2	4.0±0.5	3.5±0.1	7.6±0.1	75.6±0.9	17.1±1.4	29.3±0.4
pgh2	3.9±2.8	3.5±0.6	4.0±1.1	9.7±0.9	80.9±1.8	19.0±2.9	32.1±1.3
ptn1	5.6±0.7	4.0±0.3	3.6±0.1	11.4±0.4	75.9±0.3	19.2±0.1	31.0±0.1
pygm	8.2±0.0	7.8±1.6	5.7±0.5	20.7±2.1	74.1±1.8	34.7±3.6	36.7±1.9
pyrd	3.7±0.0	3.4±0.4	4.1±0.4	11.3±1.1	82.1±0.9	18.5±0.7	32.8±0.4
tysy	0.0±0.0	0.7±0.9	1.7±0.5	4.8±0.3	70.0±0.8	0.7±2.8	21.2±1.1
Protease							
ace	10.9±3.9	7.8±0.0	4.3±0.5	19.4±2.2	77.2±1.6	33.2±1.3	36.8±0.9
ada17	3.4±2.4	4.8±0.5	4.2±0.5	10.5±0.8	86.6±0.2	20.6±3.1	34.3±1.0
bace1	0.6±0.8	2.0±0.2	2.2±0.2	5.3±0.3	75.0±1.2	6.7±1.2	25.0±1.0
dpp4	5.3±3.9	6.4±0.9	4.8±0.5	11.9±1.0	86.3±0.5	26.7±2.7	36.5±1.0
fa10	0.8±1.1	3.2±0.8	3.6±0.6	7.3±0.5	81.6±0.9	13.6±2.6	29.8±1.0
hivpr	3.7±0.0	3.2±0.6	2.8±0.3	7.9±1.5	76.7±0.3	12.5±0.4	27.8±0.4
mmp13	0.0±0.0	1.6±1.1	1.6±0.3	3.9±0.1	63.3±1.8	2.9±0.9	19.0±0.4
thrb	1.9±0.8	3.7±0.9	2.9±0.3	6.7±0.6	77.3±0.9	14.4±2.8	28.6±1.3
try1	2.0±2.8	3.3±0.6	4.3±0.5	8.6±0.7	79.5±0.8	16.6±3.3	30.5±1.0
urok	7.8±2.8	7.0±1.0	4.7±0.5	12.9±0.9	85.8±0.9	28.0±1.3	36.8±1.0

Table 36: **TopU-LBVS-Full DMPNN results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	11.7±1.8	7.5±0.7	5.1±0.4	16.7±1.5	80.6±1.0	34.3±2.6	39.0±1.6
cp2c9	11.4±2.4	7.7±0.7	5.5±0.3	16.7±3.1	83.3±1.2	35.6±2.8	40.2±2.2
cp2d6	16.1±1.8	8.6±0.5	5.8±0.3	20.5±2.4	84.4±0.3	41.0±3.0	43.3±1.6
cp3a4	14.9±1.3	9.5±0.5	6.0±0.1	22.4±1.8	85.6±0.5	43.7±1.6	44.9±1.0
GPCR							
5ht6r	4.8±1.0	3.7±0.5	3.6±0.3	7.6±0.4	75.5±1.0	17.2±0.6	29.1±0.9
aa2ar	3.1±0.6	3.4±0.3	3.3±0.5	6.8±0.4	74.4±1.0	15.0±1.4	27.9±0.3
aa3r	6.3±0.4	4.3±0.2	3.4±0.0	7.6±0.2	77.4±0.2	18.4±0.5	29.4±0.8
adrb1	2.6±0.7	1.9±0.3	2.0±0.1	5.7±0.9	72.3±1.2	7.0±0.9	24.5±0.9
adrb2	6.8±1.4	3.7±0.3	3.2±0.2	7.5±0.3	75.3±0.5	17.4±1.1	29.3±0.3
cnr2	4.8±5.5	2.9±0.3	3.1±0.5	6.7±0.4	78.7±1.6	13.3±3.7	26.3±0.9
cxcr4	0.0±0.0	1.8±0.6	2.2±0.3	5.2±0.2	67.8±2.2	5.2±2.8	22.7±0.3
drd2	3.2±1.1	3.6±0.8	3.3±0.1	7.3±0.4	77.4±1.0	15.1±2.0	28.5±1.0
drd3	6.6±1.5	5.2±0.4	4.5±0.3	11.1±0.6	82.6±0.7	25.5±1.4	35.5±0.5
hrh3	9.6±1.9	6.4±0.2	5.1±0.5	13.6±0.6	82.5±0.5	30.5±2.3	37.5±0.5
oprd	5.4±0.9	5.1±0.6	4.2±0.1	9.4±0.6	79.2±0.4	22.6±1.7	31.9±1.1
oprk	3.5±2.5	3.1±0.2	3.2±0.3	6.6±0.3	76.0±0.3	14.3±0.9	28.1±0.6
oprm	5.8±1.6	5.4±0.2	4.7±0.1	10.4±0.4	81.1±0.8	26.1±0.8	33.9±0.3
Ion Channel							
5ht3a	1.9±1.6	6.0±0.4	5.8±0.1	12.4±1.1	84.3±1.3	28.6±0.6	36.3±0.9
cac1h	11.4±1.6	10.5±1.5	6.9±0.5	24.0±3.4	87.9±1.8	46.9±2.9	45.9±2.0
kcnh2	9.1±3.4	3.3±0.9	2.5±0.1	5.3±0.2	73.5±0.2	15.1±4.9	23.4±1.6
scn5a	0.0±0.0	2.5±0.9	3.1±0.2	5.4±0.2	71.4±0.9	10.4±1.2	24.8±0.5
trpa1	2.7±3.9	1.9±0.0	3.0±0.8	7.4±2.0	64.2±2.1	11.9±3.0	24.4±1.8
Kinase							
abl1	2.0±1.7	5.0±0.6	3.7±0.1	7.8±0.5	73.7±1.2	19.3±0.5	29.0±0.3
akt1	8.1±0.6	6.8±0.7	5.4±0.1	12.7±1.4	79.1±0.4	30.3±1.6	36.0±0.8
akt2	6.8±1.9	5.8±0.3	5.3±0.6	11.0±1.4	80.4±2.3	27.0±3.2	34.9±2.1
braf	5.7±4.0	5.6±1.1	4.6±0.4	11.2±2.0	82.5±0.9	26.0±5.4	34.4±2.2
cdk2	4.1±0.8	5.0±0.1	3.9±0.3	8.7±0.0	76.6±0.9	20.4±0.8	31.3±0.1
csf1r	10.1±2.7	5.6±0.9	4.3±0.2	13.4±1.9	81.9±0.3	27.5±2.6	36.5±1.2
egfr	7.4±2.2	4.8±0.5	4.2±0.2	9.2±1.0	79.8±0.6	22.6±2.5	32.0±1.4
fak1	12.9±6.3	6.0±0.2	4.8±0.2	14.7±4.7	78.0±1.6	29.0±4.1	35.3±3.8
fgfr1	9.0±1.8	6.2±0.6	4.4±0.1	11.7±2.0	77.1±1.1	26.9±1.4	33.9±1.4
igf1r	7.3±1.2	5.3±0.4	4.9±0.3	12.0±1.5	80.1±0.9	26.7±1.3	34.9±0.8
jak2	11.3±4.9	5.5±1.0	4.6±0.2	10.2±0.8	79.5±1.0	27.5±3.6	31.7±1.1
kit	6.2±4.4	4.3±0.2	3.6±0.4	8.6±1.2	73.9±1.0	18.0±4.1	28.8±1.7
kpcb	8.7±1.8	4.0±0.4	3.8±0.6	11.4±1.8	72.7±1.2	19.9±1.4	30.3±0.5
lck	4.2±0.0	5.2±0.3	4.0±0.1	8.1±0.2	74.1±0.1	21.0±0.8	30.1±0.3
mapk2	9.8±1.3	6.2±0.4	4.5±0.4	11.1±0.4	76.4±2.5	26.9±2.0	32.5±1.1
met	5.1±3.6	7.0±0.8	5.0±0.2	12.7±1.5	77.7±0.6	29.3±2.6	34.0±1.6
mk01	27.3±1.4	11.0±0.4	6.2±0.2	39.6±3.2	85.4±0.5	53.7±1.7	53.3±1.9
mk10	6.0±2.4	4.9±0.0	4.1±0.1	10.2±0.8	78.6±1.6	21.5±1.9	32.6±1.2
mk14	8.7±3.8	6.5±0.6	5.7±0.1	13.9±0.9	79.6±0.6	32.4±0.9	36.9±1.1
mp2k1	5.0±2.0	3.5±0.2	3.1±0.2	7.5±0.4	69.6±3.0	16.5±0.8	27.4±1.0
plk1	23.4±2.1	10.3±0.8	6.4±0.6	36.5±3.1	86.2±0.6	52.0±3.7	51.7±2.5
rock1	5.8±1.0	4.4±0.6	3.9±0.3	8.7±0.7	75.9±0.8	19.9±1.6	30.8±1.0
src	6.3±0.9	4.3±0.2	3.2±0.3	8.6±0.7	76.6±0.4	19.3±1.3	30.6±1.0
tgfr1	5.5±3.9	4.8±0.4	4.2±0.8	12.7±2.3	85.1±1.1	22.9±2.3	35.5±1.9
vgfr2	8.9±2.6	5.7±0.5	4.1±0.2	11.2±1.9	80.0±0.5	25.5±0.8	33.5±1.6
wee1	8.2±0.0	3.8±0.8	3.3±0.0	10.6±0.5	70.6±1.7	18.2±1.9	29.1±0.4

Table 37: **TopU-LBVS-Full DMPNN results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	4.4±1.2	5.2±0.6	4.1±0.0	9.6±0.3	82.2±0.6	22.8±0.9	33.6±0.4
esr1	5.9±1.5	4.5±0.6	3.3±0.2	8.1±0.8	76.3±0.2	19.3±1.7	29.9±1.3
esr2	4.2±1.7	3.2±0.8	2.7±0.4	7.4±1.4	75.9±0.6	13.9±3.0	28.2±1.3
gcr	8.3±2.2	7.2±0.9	5.8±0.2	15.3±2.0	87.6±0.3	34.1±3.4	40.5±1.7
mcr	2.7±3.9	3.2±0.9	2.3±0.5	6.4±1.7	71.2±2.7	8.9±3.7	23.9±2.5
ppara	0.8±1.1	2.0±0.8	2.9±0.5	6.3±0.5	77.3±1.0	9.9±2.5	27.0±1.1
ppard	3.6±1.3	4.4±0.5	4.7±0.3	9.5±0.9	80.0±0.3	20.9±1.3	32.6±0.8
pparg	13.2±4.5	4.7±0.2	4.1±0.4	10.7±2.1	79.4±1.0	24.3±2.6	31.8±2.7
prgr	6.5±2.2	5.9±0.8	4.8±0.2	12.2±1.1	84.2±0.2	28.2±2.3	36.3±0.5
rxra	6.1±5.7	5.0±0.8	5.2±0.6	12.9±0.5	87.0±0.6	27.3±1.6	37.4±0.5
thb	41.0±0.0	16.0±1.4	9.2±0.2	75.5±2.8	97.6±0.2	82.4±2.9	77.4±1.6
Other Enzymes							
aces	6.5±2.9	3.4±1.3	3.7±0.4	9.1±0.4	80.9±0.3	18.7±0.8	30.6±1.2
ampc	27.3±3.9	12.6±0.8	8.2±0.0	47.1±5.5	94.7±2.7	63.1±2.7	58.9±2.6
aofb	7.8±6.6	3.5±0.2	2.6±0.7	6.0±0.5	74.1±1.0	15.1±0.6	24.2±3.3
cah2	10.0±5.2	3.8±0.7	3.0±0.2	7.7±0.7	80.9±2.0	18.7±2.5	28.4±2.4
dhi1	0.0±0.0	2.4±0.7	2.6±0.3	5.6±0.2	70.3±1.6	9.1±1.1	23.7±0.3
dyr	0.0±0.0	2.1±0.9	1.4±0.5	4.4±0.4	67.5±0.9	4.0±3.3	21.0±0.8
fnta	8.2±0.0	1.6±0.0	1.4±0.4	9.5±2.8	72.8±0.4	7.8±1.5	25.7±0.8
glcm	19.7±9.4	9.9±1.2	7.1±0.4	34.3±8.5	86.7±1.3	51.0±6.1	50.5±5.2
hdac2	6.8±4.8	4.5±1.8	3.3±1.2	9.0±1.6	79.6±2.1	19.4±6.9	29.7±1.9
hdac8	3.4±2.8	2.5±0.8	2.2±0.2	5.3±0.6	71.5±1.3	8.9±3.6	23.9±0.9
hivint	2.0±1.7	2.1±0.6	2.1±0.5	5.2±0.4	73.5±1.6	7.7±3.1	24.4±0.9
hivrt	7.9±2.9	2.5±0.3	1.6±0.6	4.7±0.2	72.2±0.6	9.5±0.9	20.1±2.4
hmdh	13.7±3.9	7.1±2.5	4.8±1.5	21.0±3.7	87.0±4.2	35.3±10.0	41.9±4.8
hxk4	5.5±3.9	3.2±0.9	3.0±0.8	10.1±0.3	84.8±2.0	15.4±0.5	32.1±0.5
inha	13.7±0.0	3.8±0.7	3.3±0.3	21.3±1.4	86.7±1.3	22.8±0.2	37.8±0.6
nos1	10.2±8.4	7.0±0.5	5.6±0.5	18.2±5.0	88.4±2.7	35.6±7.7	42.0±4.7
parp1	10.6±12.0	4.3±3.1	3.0±0.7	7.4±1.9	71.9±1.0	20.4±12.1	24.2±4.6
pde5a	9.8±3.6	6.8±0.3	4.4±0.2	15.1±2.9	83.7±1.3	30.4±3.0	37.8±2.1
pgh1	4.2±0.7	3.0±0.4	2.7±0.6	6.4±0.1	74.7±0.7	12.5±0.8	27.2±0.4
pgh2	1.9±2.8	7.5±0.6	6.0±0.5	14.7±1.7	85.6±2.7	32.6±3.1	38.4±2.3
ptn1	5.6±2.9	3.6±0.8	2.9±0.3	7.3±1.1	75.0±2.0	14.4±3.6	27.6±1.3
pygm	5.5±3.9	2.6±0.9	2.3±1.2	7.0±1.4	72.2±2.3	10.7±4.9	25.4±2.8
pyrd	0.0±0.0	2.6±1.0	3.5±0.5	7.4±0.9	80.2±2.1	12.5±3.1	29.2±1.8
tysy	0.0±0.0	0.0±0.0	0.7±0.5	3.7±0.3	64.0±1.5	−2.8±1.7	17.3±1.2
Protease							
ace	2.7±3.9	5.2±1.8	4.7±1.2	10.6±1.5	77.3±2.5	21.1±4.7	31.4±2.2
ada17	1.7±2.4	4.4±1.8	3.5±0.3	9.7±2.6	82.2±2.2	18.2±5.7	31.9±3.3
bace1	5.9±4.7	2.0±0.6	1.9±0.7	5.9±0.6	77.0±1.1	10.6±3.6	24.8±1.4
dpp4	8.4±3.0	5.1±1.4	4.3±0.4	11.8±2.1	83.7±1.4	25.5±4.4	35.5±1.9
fa10	2.3±1.9	3.1±0.2	3.5±0.4	7.2±0.6	80.7±2.0	14.2±2.1	28.9±1.0
hivpr	6.2±1.8	3.4±1.0	2.9±0.2	8.9±2.0	76.2±1.1	14.1±1.7	28.5±0.6
mmp13	3.9±2.8	2.7±1.1	2.3±0.0	6.0±0.9	66.2±1.9	9.6±2.2	23.3±1.4
thrb	4.5±3.9	4.0±0.3	2.9±0.3	7.0±0.5	76.8±0.7	13.4±3.0	27.2±0.8
try1	5.9±4.8	4.1±0.6	3.5±0.3	10.9±3.2	76.6±1.0	18.3±4.2	31.2±2.9
urok	1.9±2.8	3.5±0.0	3.5±0.0	10.1±2.6	82.4±1.4	15.7±1.1	31.3±1.2

Table 38: **TopU-LBVS-Full Molformer results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450							
cp1a2	7.8±3.1	6.1±0.4	4.6±0.2	11.4±1.9	79.1±0.4	26.7±3.2	34.7±1.8
cp2c9	11.2±1.4	7.8±0.9	5.6±0.4	16.7±1.4	84.8±0.1	36.5±3.8	41.0±1.3
cp2d6	11.3±0.9	7.5±0.5	5.4±0.2	16.0±1.4	84.4±0.6	34.8±2.2	40.2±1.2
cp3a4	14.5±1.8	9.0±0.3	6.1±0.1	21.0±0.7	84.3±0.1	42.5±0.9	43.7±0.4
GPCR							
5ht6r	4.1±1.7	3.0±0.4	2.8±0.2	7.8±1.0	75.6±0.3	13.0±1.3	27.9±0.5
aa2ar	2.8±1.0	3.5±1.0	3.2±0.3	6.5±0.8	74.2±1.2	14.1±3.3	27.8±1.6
aa3r	4.9±1.2	3.1±0.2	3.4±0.4	7.2±0.4	76.7±0.3	15.2±1.4	29.3±0.7
adrb1	0.0±0.0	2.5±0.3	1.9±0.3	4.6±0.4	69.1±1.4	6.9±1.9	22.9±1.2
adrb2	3.9±1.4	3.2±0.3	3.1±0.2	6.5±0.2	71.7±0.8	13.9±2.1	26.7±0.8
cnr2	4.8±1.7	3.2±0.6	2.5±0.2	7.4±1.4	77.5±0.4	12.8±3.0	28.1±1.4
cxcr4	2.0±2.8	2.7±0.0	1.8±0.3	6.0±0.7	67.7±3.5	9.9±2.0	23.9±2.0
drd2	1.8±0.0	2.5±0.7	3.0±0.4	6.0±0.5	76.4±1.0	11.3±2.4	27.2±1.2
drd3	6.6±1.5	5.4±0.4	4.0±0.3	10.6±1.0	82.5±0.6	24.5±2.4	34.9±1.2
hrh3	7.1±4.0	5.5±0.9	4.7±0.2	12.6±1.6	81.9±0.8	27.1±2.9	36.2±1.5
oprd	5.4±2.5	4.6±0.6	3.6±0.1	10.1±1.1	78.4±0.7	19.6±2.4	31.7±0.9
oprk	3.9±1.5	3.6±0.5	3.2±0.2	6.9±0.4	76.5±0.5	15.0±1.4	28.7±0.7
oprm	7.3±0.4	4.8±0.7	4.2±0.1	9.9±0.5	79.7±0.3	23.2±1.1	33.7±0.6
Ion Channel							
5ht3a	10.4±2.4	5.8±0.5	5.3±0.3	15.0±1.7	79.8±1.2	29.5±1.0	37.0±0.8
cac1h	9.1±3.2	9.5±0.0	7.3±0.2	20.7±0.9	88.6±1.0	42.5±1.7	44.3±1.2
kcnh2	3.8±0.4	2.9±0.1	2.6±0.1	5.9±0.2	72.9±0.2	11.9±0.7	26.5±0.3
scn5a	6.1±1.1	3.1±0.4	3.2±0.2	8.6±1.6	73.5±1.1	14.7±1.7	28.1±1.2
trpa1	8.2±0.0	5.2±0.9	3.3±0.5	13.0±4.0	67.2±3.0	22.3±3.4	29.0±2.1
Kinase							
abl1	4.4±1.3	4.4±0.6	3.5±0.4	7.7±0.5	73.6±0.8	18.3±1.9	29.3±0.9
akt1	9.0±1.8	7.2±0.7	4.9±0.3	12.9±0.8	77.5±0.4	31.9±1.7	36.1±0.8
akt2	9.6±1.0	6.6±0.7	5.1±0.3	13.8±2.8	78.4±0.9	30.5±2.6	36.0±1.6
braf	3.4±2.4	4.4±0.8	3.7±0.4	8.7±1.1	79.2±2.0	19.1±3.3	31.1±1.6
cdk2	4.1±1.7	4.7±0.4	4.3±0.2	8.5±0.4	77.1±0.6	20.7±1.2	31.4±0.7
csf1r	7.9±1.0	3.5±0.4	3.0±0.2	11.3±0.2	77.7±0.5	17.8±1.4	31.6±0.7
egfr	3.5±1.3	4.5±0.4	4.2±0.3	8.4±0.8	78.1±1.1	19.5±1.9	31.5±1.3
fak1	7.2±1.0	5.6±0.6	4.8±0.3	11.8±1.6	77.8±0.7	25.7±2.8	33.7±1.5
fgfr1	7.7±2.1	5.5±0.4	4.4±0.4	11.1±1.2	78.2±0.2	24.7±2.3	33.6±1.2
igf1r	8.3±0.7	7.1±0.9	5.3±0.1	14.4±1.3	81.8±0.5	31.2±2.9	38.1±1.3
jak2	3.9±2.4	5.5±0.7	4.4±0.3	9.3±1.6	76.3±1.3	23.0±3.6	32.1±2.2
kit	7.5±4.0	4.8±0.6	3.5±0.2	11.1±4.2	71.5±1.9	20.0±3.9	29.8±2.8
kpcb	7.5±3.0	4.8±0.4	4.9±0.2	11.9±1.0	81.1±0.5	25.4±0.7	35.1±0.7
lck	4.2±1.5	4.3±1.0	3.6±0.7	7.5±1.4	72.1±0.8	18.8±4.5	28.8±2.1
mapk2	7.1±2.1	5.4±0.3	4.0±0.1	9.3±0.8	72.9±0.3	22.0±0.3	30.5±0.5
met	8.5±1.2	6.3±0.9	4.7±0.2	12.7±1.2	76.1±1.2	29.5±2.8	35.0±1.2
mk01	22.0±2.1	11.0±0.1	6.6±0.1	35.0±1.3	85.4±0.3	52.9±0.9	51.0±0.5
mk10	6.0±2.4	4.3±0.8	4.1±0.3	10.8±0.3	76.0±1.0	21.0±2.9	31.7±1.1
mk14	9.9±0.5	6.2±0.3	5.3±0.0	13.3±0.3	79.4±0.3	30.3±0.8	36.5±0.4
mp2k1	7.2±1.0	3.2±0.4	2.7±0.3	8.5±0.7	69.4±1.0	15.4±1.4	27.7±0.7
plk1	21.5±2.8	8.9±0.5	5.6±0.2	30.4±4.6	84.0±0.1	45.6±3.3	47.5±2.4
rock1	5.0±1.0	4.8±0.6	4.5±0.2	9.2±0.6	75.0±0.1	21.3±1.4	31.4±0.6
src	4.0±0.0	3.6±0.4	3.6±0.1	7.5±0.6	77.6±0.9	15.8±0.2	29.6±0.4
tgfr1	6.8±1.9	5.4±1.3	4.2±0.8	12.2±2.2	84.3±2.3	24.7±4.1	35.7±2.7
vgfr2	7.0±0.4	4.6±0.3	4.2±0.2	9.2±0.4	79.9±0.4	21.8±0.9	32.9±0.5
wee1	10.9±3.9	3.3±0.0	2.7±0.8	11.2±2.9	72.6±2.0	16.0±2.3	29.5±2.6

Table 39: **TopU-LBVS-Full Molformer results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC ($\alpha = 20$), and LogAUC are reported on a 0–100 scale.

Target	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor							
andr	4.9±1.8	3.2±0.6	3.6±0.6	8.3±1.1	81.1±1.4	15.9±2.8	31.2±1.7
esr1	2.7±1.5	3.6±0.8	3.3±0.2	6.6±0.8	74.9±0.9	14.9±3.2	28.0±1.6
esr2	2.4±1.7	3.0±0.3	3.0±0.3	6.3±0.3	74.8±0.5	12.3±1.0	27.2±0.5
gcr	8.3±2.2	6.5±1.0	5.3±0.2	15.7±2.1	86.5±0.3	32.4±2.8	40.4±1.7
mcr	8.2±0.0	3.9±1.6	3.3±0.9	12.1±2.1	73.5±1.5	18.3±4.3	29.0±2.0
ppara	4.8±2.0	4.2±0.2	2.9±0.2	7.2±0.7	75.7±0.5	15.4±2.1	28.5±0.9
ppard	4.6±1.3	5.1±0.5	4.9±0.9	10.8±1.3	80.7±1.1	23.5±2.5	34.1±1.5
pparg	3.2±1.7	3.7±0.3	3.5±0.2	7.9±0.5	77.4±0.8	16.7±1.0	30.0±0.6
prgr	6.5±1.7	6.5±0.8	4.5±0.3	12.3±0.9	81.9±1.1	26.6±3.2	35.2±1.5
rxra	10.6±4.3	6.0±2.2	4.2±0.2	15.0±3.6	84.0±1.8	28.0±7.5	37.6±3.5
thb	30.4±2.1	16.7±0.9	9.3±0.2	58.5±0.6	98.1±0.4	79.1±1.9	70.8±0.5
Other Enzymes							
aces	3.4±1.2	4.2±0.4	3.7±0.3	7.7±0.4	78.4±0.7	17.4±1.2	30.5±0.7
ampc	27.3±3.9	15.3±1.6	8.5±0.4	51.3±4.6	95.6±2.7	70.5±3.8	63.0±2.4
aofb	2.9±0.0	3.0±0.6	2.6±0.1	6.5±0.2	73.9±0.5	11.9±1.8	26.5±0.5
cah2	6.8±0.7	5.0±0.8	4.1±0.2	11.2±0.9	81.8±1.0	23.7±2.7	34.8±1.3
dhi1	2.3±3.2	2.4±0.7	2.6±0.7	6.0±0.5	70.3±0.4	10.4±1.5	24.5±0.6
dyr	3.4±2.4	2.1±0.9	2.8±1.0	7.0±1.2	75.9±2.5	10.1±3.7	26.6±2.3
fnta	8.2±0.0	3.8±0.8	3.0±0.4	10.3±0.2	75.4±1.3	17.7±2.7	29.9±0.6
glcm	15.2±7.7	9.3±0.0	6.8±0.2	23.3±4.1	87.7±0.6	45.0±3.2	45.8±2.7
hdac2	5.7±3.2	4.7±0.3	3.8±0.3	10.6±1.8	79.1±1.9	21.1±1.7	32.3±1.2
hdac8	1.1±1.6	0.9±0.3	1.6±0.4	4.4±0.6	69.6±2.5	2.8±2.4	21.4±1.8
hivint	1.4±1.0	2.5±0.7	2.4±0.3	5.5±0.4	72.8±1.5	8.1±1.9	25.1±0.9
hivrt	3.1±0.6	1.7±0.2	1.7±0.2	4.8±0.5	69.1±0.5	5.2±0.7	22.4±0.2
hmdh	5.5±3.9	7.7±0.8	5.7±1.1	20.1±3.6	89.1±0.6	36.0±4.4	41.9±2.3
hxk4	8.2±0.0	3.2±0.9	2.7±0.5	13.1±2.6	79.1±0.2	15.5±2.4	30.7±0.6
inha	9.1±6.4	6.6±1.3	5.0±1.5	17.2±4.0	89.9±1.3	30.9±7.1	40.9±3.8
nos1	8.5±2.4	6.3±0.0	6.3±0.0	16.7±2.6	83.5±0.4	31.9±2.6	39.0±1.3
parp1	4.6±0.0	3.3±0.5	2.8±0.8	6.7±0.8	71.4±1.4	12.5±2.5	26.2±1.3
pde5a	8.8±2.4	5.3±0.0	4.4±0.4	12.5±2.8	82.7±0.6	24.8±1.4	34.9±0.7
pgh1	2.1±0.7	3.7±0.9	3.4±0.4	7.1±0.8	73.9±0.8	15.4±2.7	28.0±1.2
pgh2	9.8±2.8	5.7±0.6	4.6±0.5	12.9±1.5	77.0±2.5	27.1±0.8	34.4±1.5
ptn1	4.6±1.2	2.7±1.1	2.4±0.6	6.3±1.5	72.0±2.3	10.6±4.2	25.6±2.6
pygm	2.7±3.9	3.2±0.9	3.7±0.5	8.4±1.5	72.0±1.2	12.8±0.8	27.1±0.9
pyrd	1.2±1.8	3.1±0.6	3.9±0.7	7.3±0.4	80.7±0.8	12.8±1.2	29.2±0.7
tysy	0.0±0.0	1.3±0.9	2.3±1.2	4.7±0.5	63.5±0.9	4.2±1.8	20.3±1.1
Protease							
ace	5.5±3.9	5.2±2.4	4.0±0.8	10.2±1.8	75.8±2.6	20.5±5.6	30.4±2.5
ada17	3.4±2.4	3.3±0.0	2.8±0.5	7.8±1.1	79.9±0.6	12.9±2.2	29.1±1.3
bace1	0.6±0.8	1.5±0.3	2.4±0.1	5.4±0.1	76.1±0.2	6.0±1.1	25.2±0.4
dpp4	4.2±1.5	5.3±1.1	4.2±0.7	10.9±1.2	82.5±1.5	24.2±5.0	34.2±2.3
fa10	5.3±2.1	4.2±0.7	3.5±0.0	8.1±1.4	78.5±1.3	15.5±2.9	29.7±1.7
hivpr	2.5±1.8	2.6±0.4	2.9±0.2	7.0±0.9	77.3±1.8	11.3±0.4	27.7±1.1
mmp13	1.9±2.8	2.7±0.6	2.3±0.0	5.7±0.4	68.0±1.5	10.3±1.4	23.4±1.0
thrb	5.7±0.8	4.2±0.5	3.4±0.4	7.8±0.6	77.2±0.5	17.6±2.3	30.1±1.0
try1	5.9±4.8	2.5±1.0	2.5±0.5	7.1±1.4	74.9±1.5	11.1±5.0	26.9±2.4
urok	9.8±5.5	6.2±1.1	4.3±0.7	16.2±3.4	86.7±0.8	27.8±3.8	38.4±2.4

Table 40: **TopU-LBVS-few Morgan-RF results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	2.3±1.6	1.6±0.6	1.5±0.3	3.7±0.2	62.2±3.6	3.6±1.5	25.7±1.7
cp2c9	2	0.0±0.0	0.3±0.2	0.7±0.4	2.5±0.3	48.6±5.1	-4.3±1.4	20.2±4.5
cp2d6	2	5.2±2.6	1.9±0.9	1.5±0.3	3.7±1.0	53.9±2.5	5.6±4.4	23.8±3.5
cp3a4	2	1.1±1.5	0.9±0.4	1.2±0.2	3.6±1.6	49.8±3.0	-0.4±2.5	19.9±3.0
GPCR								
5ht6r	1	0.0±0.0	0.1±0.2	0.3±0.3	2.4±0.1	45.4±2.0	-5.7±1.1	23.0±2.7
aa2ar	2	0.2±0.3	0.8±0.7	1.0±0.5	2.8±0.1	54.1±3.2	-0.5±1.9	21.9±1.3
aa3r	2	0.0±0.0	1.6±0.2	1.4±0.2	3.0±0.5	55.6±6.6	1.8±0.9	24.0±3.7
adrb1	2	1.1±1.6	1.5±1.2	1.0±0.7	3.3±0.7	54.9±6.3	1.3±5.7	25.8±4.7
adrb2	2	2.3±1.9	1.1±0.6	1.0±0.3	2.9±0.6	49.7±2.4	0.0±3.5	22.0±1.4
cnr2	2	0.0±0.0	0.6±0.7	0.8±0.1	2.5±0.1	47.5±2.0	-3.2±1.9	24.1±0.6
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.5	44.1±8.0	-6.7±0.1	25.6±4.4
drd2	2	0.0±0.0	1.0±1.1	1.7±0.7	3.1±0.7	53.5±4.1	1.3±4.9	23.5±3.4
drd3	2	1.7±2.4	2.2±0.8	1.6±0.5	3.9±0.3	58.0±6.6	5.6±1.8	27.1±1.8
hrh3	2	1.5±2.1	1.8±0.9	1.5±0.5	3.2±0.6	53.2±6.7	2.3±4.9	27.7±7.0
oprd	2	0.0±0.0	0.3±0.4	0.9±0.4	2.6±0.3	52.0±3.4	-2.8±1.0	20.0±2.0
oprk	2	0.0±0.0	0.1±0.1	0.5±0.1	2.4±0.1	48.5±2.5	-4.9±0.1	21.0±0.5
oprm	2	0.0±0.0	1.7±0.4	1.6±0.1	3.4±0.9	55.8±5.8	3.4±3.1	25.3±3.5
Ion Channel								
5ht3a	1	0.0±0.0	0.0±0.0	0.1±0.1	2.4±0.1	43.0±1.7	-6.2±0.3	18.0±1.8
cac1h	1	0.0±0.0	1.1±1.5	1.1±1.5	3.6±1.5	50.1±7.8	-0.7±6.9	27.8±7.3
kcnh2	2	1.2±1.0	1.0±0.3	1.1±0.3	2.7±0.2	52.3±2.0	0.2±1.7	16.7±1.2
scn5a	1	0.0±0.0	0.0±0.0	0.6±0.5	2.4±0.2	44.9±6.9	-4.6±1.6	19.7±4.9
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	39.6±4.6	-6.7±0.0	30.0±7.6
Kinase								
abl1	2	1.7±2.4	1.0±0.7	0.8±0.4	2.6±0.1	46.7±3.9	-1.2±2.8	19.8±2.7
akt1	2	0.0±0.0	1.0±0.7	1.0±0.3	2.9±0.1	54.7±3.0	-0.4±1.9	23.0±0.6
akt2	1	0.4±0.6	0.8±0.9	0.9±0.6	2.8±0.3	50.1±2.2	-1.0±3.8	25.2±1.4
braf	2	2.7±3.9	1.2±0.8	1.1±0.3	4.5±1.8	51.9±7.7	1.9±2.6	25.4±4.5
cdk2	2	0.0±0.0	0.3±0.5	0.3±0.3	2.3±0.2	44.3±4.2	-4.2±2.3	16.3±2.7
csf1r	1	0.0±0.0	0.3±0.4	0.5±0.6	2.5±0.1	44.6±3.6	-4.7±2.8	22.4±1.2
egfr	2	0.0±0.0	0.4±0.3	0.6±0.3	2.5±0.1	50.5±1.9	-3.5±0.3	19.2±2.3
fak1	1	0.0±0.0	0.7±1.0	1.0±0.9	3.6±0.6	58.8±3.0	-1.5±5.0	28.5±4.1
fgfr1	2	0.0±0.0	0.9±1.2	0.8±0.8	2.6±0.5	47.0±8.1	-2.5±4.7	20.4±5.7
igf1r	2	0.0±0.0	1.0±1.4	0.8±0.6	2.8±0.5	44.3±6.0	-0.9±5.2	18.7±4.7
jak2	2	0.0±0.0	0.4±0.6	0.8±0.9	2.7±0.7	50.2±7.7	-2.6±4.0	19.8±5.1
kit	2	0.0±0.0	0.6±0.8	0.8±0.2	2.3±0.1	42.2±3.6	-3.3±2.1	19.2±3.4
kpcb	1	0.0±0.0	0.1±0.2	0.5±0.4	3.1±1.0	50.3±11.9	-4.3±2.0	27.1±10.0
lck	2	0.0±0.0	0.6±0.7	0.9±0.7	2.8±0.6	54.5±7.1	-2.3±3.7	22.8±4.6
mapk2	2	0.3±0.4	1.0±0.9	1.3±1.0	2.9±0.5	49.9±7.6	0.2±4.9	23.7±5.6
met	2	2.0±2.8	1.1±0.3	1.1±0.3	4.9±2.9	51.7±5.6	0.0±1.6	24.2±3.0
mk01	2	6.1±1.1	2.6±1.4	2.0±1.0	5.5±2.2	60.4±9.1	9.4±6.7	32.2±7.5
mk10	1	0.0±0.0	0.0±0.0	0.2±0.3	2.4±0.1	41.6±4.6	-5.9±0.9	17.8±1.4
mk14	2	0.0±0.0	0.7±0.5	0.9±0.3	2.8±0.1	54.1±1.2	-2.1±1.4	24.3±2.2
mp2k1	1	0.0±0.0	0.8±1.0	0.5±0.5	2.6±0.6	42.1±10.1	-3.0±4.2	19.4±7.0
plk1	2	3.4±4.8	2.3±2.7	1.7±1.6	5.3±3.9	55.7±9.3	5.7±14.6	30.0±11.0
rock1	1	0.0±0.0	0.1±0.1	0.4±0.3	2.4±0.1	45.9±4.1	-5.4±1.2	22.4±1.3
src	2	0.0±0.0	0.3±0.4	0.5±0.4	2.6±0.2	50.7±5.5	-3.5±1.1	19.3±1.6
tgfr1	1	0.0±0.0	0.3±0.5	0.5±0.6	2.5±0.1	42.6±7.5	-4.8±2.4	24.7±9.0
vgfr2	2	0.1±0.2	0.8±0.7	0.9±0.3	2.8±0.6	52.8±8.5	-1.4±2.2	21.2±4.3
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.1	37.0±2.9	-6.7±0.0	20.3±3.3

Table 41: **TopU-LBVS-few Morgan-RF results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	0.0±0.0	0.7±0.6	2.5±0.4	46.0±9.0	-3.8±2.0	17.8±5.3
esr1	2	0.0±0.0	0.6±0.2	0.8±0.2	2.6±0.1	51.1±2.2	-2.7±0.9	21.6±2.0
esr2	2	0.0±0.0	0.3±0.3	0.7±0.7	2.6±0.5	48.1±7.7	-3.9±2.1	23.1±7.4
gcr	2	0.0±0.0	0.0±0.0	0.1±0.1	2.4±0.1	45.4±2.6	-5.7±0.8	21.0±0.8
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	39.6±2.6	-6.7±0.0	28.9±5.0
ppara	1	0.0±0.0	0.0±0.0	0.2±0.2	2.4±0.0	43.6±1.9	-6.1±0.9	23.1±1.1
ppard	1	0.0±0.0	0.1±0.1	0.3±0.3	2.4±0.1	45.5±2.8	-5.5±1.1	24.9±2.6
pparg	2	0.0±0.0	0.1±0.2	0.6±0.5	2.6±0.4	49.3±6.1	-4.1±2.5	20.0±4.5
prgr	2	0.0±0.0	0.9±1.3	0.8±0.9	2.7±0.5	48.9±4.1	-2.0±5.8	23.6±5.6
rxra	1	0.0±0.0	0.0±0.0	0.1±0.1	2.7±0.3	49.7±1.4	-5.8±0.6	25.9±2.9
thb	1	0.0±0.0	1.9±1.9	2.0±1.5	4.7±2.0	57.8±4.5	3.2±8.1	38.5±7.6
Other Enzymes								
aces	2	0.0±0.0	0.7±0.5	0.6±0.4	2.5±0.2	50.9±3.1	-3.2±2.4	18.6±3.1
ampc	1	0.4±0.5	0.6±0.9	0.7±0.9	2.7±0.4	48.0±7.5	-4.0±3.8	30.3±6.6
aofb	2	0.0±0.0	0.0±0.0	0.1±0.2	2.4±0.0	43.7±3.8	-6.1±0.9	20.3±2.6
cah2	2	2.3±3.2	1.1±0.6	0.7±0.1	3.3±0.6	52.7±4.3	-0.1±3.3	24.8±1.5
dhi1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	42.6±1.5	-6.7±0.0	30.5±3.4
dyr	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	30.1±9.6	-6.7±0.0	15.2±7.5
fnta	1	0.0±0.0	0.0±0.0	0.1±0.1	2.4±0.0	43.3±2.7	-6.7±0.0	32.9±6.1
glcm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.6	45.5±11.4	-5.9±0.9	23.0±8.0
hdac2	1	0.0±0.0	1.2±1.8	1.3±1.8	4.2±1.6	53.3±10.3	0.3±8.0	30.8±8.0
hdac8	1	0.0±0.0	1.1±1.6	1.0±0.8	3.0±0.6	52.8±3.7	-0.6±5.9	30.2±5.5
hivint	1	0.0±0.0	0.0±0.0	0.4±0.5	2.3±0.1	35.5±2.0	-5.6±1.6	14.0±2.2
hivrt	2	0.0±0.0	0.1±0.1	0.4±0.3	2.4±0.1	43.7±3.9	-5.3±1.3	19.6±3.8
hmdh	1	0.0±0.0	0.0±0.0	0.1±0.1	2.4±0.0	44.8±0.3	-6.7±0.0	40.8±1.0
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	39.2±5.2	-6.7±0.0	29.4±7.9
inha	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.4	47.7±9.4	-6.4±0.5	30.7±5.6
nos1	1	0.0±0.0	0.0±0.0	0.1±0.1	3.1±0.9	51.9±9.1	-5.3±1.3	31.0±4.4
parp1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	41.0±0.5	-6.7±0.0	22.2±3.2
pde5a	1	0.0±0.0	0.0±0.0	0.6±0.5	2.6±0.3	48.4±8.0	-4.5±1.8	24.0±6.1
pgh1	2	0.0±0.0	0.7±0.6	0.6±0.4	2.4±0.0	41.9±6.4	-4.1±1.9	19.7±5.1
pgh2	1	0.0±0.0	0.0±0.0	0.1±0.1	2.6±0.2	46.0±6.2	-6.1±0.9	32.3±6.3
ptn1	2	0.0±0.0	0.0±0.1	0.3±0.5	2.3±0.2	40.6±8.7	-5.3±1.6	16.9±7.1
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	42.1±2.8	-6.7±0.0	34.3±6.0
pyrd	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.1	39.6±4.4	-6.6±0.1	20.4±5.5
tysy	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	36.2±5.9	-6.7±0.0	25.1±8.3
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	40.4±3.1	-6.7±0.0	30.8±6.4
ada17	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.0	31.2±7.5	-6.7±0.0	17.1±7.9
bace1	2	0.0±0.0	0.8±0.8	0.9±0.0	3.4±0.7	57.8±9.4	-1.1±2.4	26.7±4.1
dpp4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.4	51.5±7.3	-5.3±1.5	26.3±3.1
fa10	1	0.0±0.0	0.8±1.1	0.6±0.9	2.8±0.5	47.7±6.1	-2.1±4.9	23.2±6.0
hivpr	1	0.0±0.0	0.0±0.0	0.1±0.1	2.6±0.2	43.1±2.9	-5.9±1.2	19.5±2.7
mmp13	1	0.0±0.0	0.0±0.0	0.2±0.2	2.5±0.0	45.1±3.6	-6.2±0.8	28.8±2.6
thrb	2	0.0±0.0	0.2±0.2	0.9±0.4	2.6±0.2	50.6±4.1	-3.4±1.5	22.0±0.9
try1	1	0.0±0.0	0.0±0.0	0.5±0.3	2.7±0.2	50.3±5.3	-5.2±1.1	35.1±3.2
urok	1	0.0±0.0	0.0±0.0	0.5±0.7	3.0±0.9	48.4±10.9	-5.1±2.1	28.8±10.2

Table 42: **TopU-LBVS-few Tanimoto-NN results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	1.1±1.6	1.2±0.3	1.2±0.5	3.0±0.2	50.4±2.2	2.1±0.4	19.3±1.0
cp2c9	2	0.0±0.0	0.3±0.4	0.6±0.2	2.6±0.1	51.9±1.9	-3.5±0.2	17.7±0.5
cp2d6	2	4.6±3.2	3.3±0.7	1.7±0.3	4.6±0.5	54.1±3.1	8.6±2.5	23.6±2.0
cp3a4	2	1.1±1.5	0.9±0.3	1.1±0.2	2.5±0.3	46.6±4.1	-0.6±1.3	16.3±2.2
GPCR								
5ht6r	1	0.0±0.0	0.7±0.9	1.3±0.5	3.2±0.7	54.3±8.0	-0.9±1.9	23.3±5.0
aa2ar	2	0.0±0.0	0.5±0.4	0.7±0.2	2.3±0.2	45.9±4.1	-2.4±1.7	15.7±1.6
aa3r	2	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.0	46.9±1.0	-6.1±0.2	14.5±0.4
adrb1	2	0.0±0.0	0.0±0.0	0.5±0.4	2.9±0.1	56.4±0.7	-4.0±1.1	21.7±0.7
adrb2	2	0.0±0.0	0.6±0.4	1.3±0.4	3.3±0.5	59.3±4.0	-0.7±2.5	22.8±2.9
cnr2	2	0.0±0.0	0.0±0.0	0.6±0.4	2.1±0.2	36.1±7.1	-4.4±0.9	13.8±2.9
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.3	33.6±6.8	-6.7±0.0	14.9±3.4
drd2	2	0.0±0.0	0.3±0.4	0.7±0.4	2.8±0.1	55.6±1.3	-3.6±0.4	19.2±0.3
drd3	2	0.0±0.0	0.7±0.5	0.7±0.3	3.2±0.5	56.6±4.2	-0.6±2.5	22.8±3.0
hrh3	2	0.0±0.0	0.0±0.0	0.5±0.7	3.7±0.2	65.1±1.2	-3.7±1.9	26.1±1.0
oprd	2	0.0±0.0	0.6±0.4	0.3±0.2	2.8±0.2	56.3±3.1	-4.0±1.4	19.5±1.8
oprk	2	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.1	47.3±2.0	-6.1±0.3	15.4±1.2
oprm	2	0.0±0.0	0.3±0.4	0.3±0.4	2.8±0.5	54.3±5.9	-3.4±3.1	19.1±3.9
Ion Channel								
5ht3a	1	0.0±0.0	0.0±0.0	1.0±1.4	3.4±0.2	59.0±2.4	-2.9±3.1	25.3±1.1
cac1h	1	0.0±0.0	0.0±0.0	0.8±1.2	3.7±0.3	56.5±7.5	-3.6±2.3	27.7±2.8
kcnh2	2	1.8±0.6	0.9±0.5	1.1±0.5	2.8±0.4	53.0±3.6	0.2±2.5	16.7±2.2
scn5a	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.3	42.4±8.3	-6.5±0.3	15.0±4.0
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±0.6	43.6±7.0	-6.5±0.3	23.8±6.5
Kinase								
abl1	2	0.0±0.0	0.7±0.5	1.1±0.4	3.0±0.7	52.0±6.8	-0.5±3.5	20.4±4.5
akt1	2	0.0±0.0	0.9±0.6	0.9±0.3	2.6±0.2	46.2±4.0	-1.9±1.8	18.3±1.6
akt2	1	0.0±0.0	1.4±1.0	1.5±0.5	4.1±0.8	58.0±6.3	2.2±3.0	28.4±5.1
braf	2	0.0±0.0	0.0±0.0	0.3±0.4	2.5±0.3	48.3±7.0	-5.5±0.5	18.4±2.8
cdk2	2	0.0±0.0	0.3±0.5	0.2±0.2	2.5±0.3	50.6±5.3	-5.1±1.1	17.1±2.2
csf1r	1	0.0±0.0	0.0±0.0	0.7±0.5	3.2±0.6	53.7±3.5	-3.7±2.1	23.2±4.1
egfr	2	0.0±0.0	0.0±0.0	0.2±0.3	2.4±0.2	51.9±4.1	-5.7±0.6	15.6±1.8
fak1	1	0.0±0.0	0.7±1.0	0.4±0.5	4.2±1.6	54.3±12.4	-0.2±4.9	27.7±8.7
fgfr1	2	0.0±0.0	0.0±0.0	0.4±0.3	2.1±0.1	39.6±2.7	-5.4±0.9	13.5±1.1
igf1r	2	0.0±0.0	1.5±0.0	1.3±0.7	4.0±0.2	62.3±3.7	1.2±2.1	28.2±1.2
jak2	2	0.0±0.0	0.0±0.0	0.4±0.6	2.3±0.4	45.0±7.4	-5.2±1.9	15.2±3.7
kit	2	0.0±0.0	0.0±0.0	0.9±0.7	2.4±0.2	43.7±1.3	-3.8±2.1	17.1±1.9
kpcb	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±1.3	45.3±16.1	-5.3±2.0	21.1±10.8
lck	2	0.0±0.0	0.0±0.0	0.7±0.2	2.8±0.1	54.9±2.7	-3.4±0.8	20.1±0.5
mapk2	2	0.0±0.0	0.5±0.8	1.1±0.8	2.9±0.6	50.3±6.6	-1.9±3.2	21.1±4.8
met	2	0.0±0.0	0.4±0.6	0.7±0.5	3.0±0.7	53.5±11.0	-3.3±2.5	20.9±5.8
mk01	2	4.6±6.4	3.5±1.4	1.8±0.7	6.3±3.2	56.1±7.2	9.1±8.7	29.1±6.9
mk10	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.3	43.8±6.3	-5.9±0.6	16.8±4.1
mk14	2	0.0±0.0	0.4±0.6	1.1±0.8	3.5±1.0	59.4±9.3	-0.8±4.2	23.5±6.6
mp2k1	1	0.0±0.0	0.7±1.0	1.1±0.9	2.6±0.7	40.7±9.4	-2.9±3.3	17.2±6.6
plk1	2	6.8±5.6	2.9±1.9	2.0±0.7	7.2±3.9	61.0±3.0	10.2±9.6	31.4±5.5
rock1	1	0.0±0.0	0.0±0.0	0.4±0.5	2.4±0.4	44.2±6.8	-5.1±2.0	17.4±4.1
src	2	0.0±0.0	0.3±0.5	0.2±0.2	2.5±0.1	51.3±1.9	-4.6±1.1	17.5±0.5
tgfr1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.2	54.2±1.9	-6.1±0.7	24.1±2.2
vgfr2	2	0.0±0.0	0.2±0.3	0.1±0.2	2.1±0.1	43.9±2.2	-5.4±1.5	12.9±1.5
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	4.0±0.3	51.9±6.7	-4.8±0.9	30.1±3.0

Table 43: **TopU-LBVS-few Tanimoto-NN results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.2	44.3±4.1	-6.2±0.2	15.1±1.8
esr1	2	0.0±0.0	0.0±0.0	0.1±0.2	2.4±0.2	50.4±3.8	-5.0±0.8	16.5±2.3
esr2	2	0.0±0.0	0.0±0.0	0.3±0.4	2.4±0.2	47.2±4.2	-5.7±1.1	16.6±2.7
gcr	2	0.0±0.0	0.0±0.0	0.3±0.4	3.3±0.3	60.0±3.3	-4.5±1.1	24.4±1.5
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.3±1.1	36.0±21.2	-6.6±0.1	20.8±13.2
ppara	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.1	44.3±2.7	-6.6±0.0	16.0±1.2
ppard	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.4	49.9±8.3	-6.2±0.1	20.4±3.6
pparg	2	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.1	46.5±2.0	-5.7±0.2	15.9±1.1
prgr	2	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.3	47.4±6.4	-6.1±0.3	17.1±3.0
rxra	1	0.0±0.0	0.0±0.0	0.0±0.0	2.9±0.3	49.4±5.8	-6.5±0.1	21.9±2.5
thb	1	0.0±0.0	2.3±1.9	4.8±2.0	9.2±2.7	74.4±9.6	11.8±5.1	53.4±10.6
Other Enzymes								
aces	2	0.0±0.0	0.0±0.0	0.1±0.2	2.3±0.2	47.8±3.3	-5.4±0.5	15.5±1.6
ampc	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.8	34.3±11.2	-5.4±1.8	18.0±8.7
aofb	2	0.0±0.0	0.0±0.0	0.5±0.7	2.7±0.8	49.2±11.1	-4.7±1.7	18.8±6.1
cah2	2	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.1	42.8±2.1	-6.1±0.1	15.1±0.8
dhi1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.3	36.9±7.0	-6.7±0.0	16.4±3.7
dyr	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.4	39.9±5.4	-6.5±0.2	17.6±4.1
fnta	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.1	31.9±2.9	-6.7±0.0	14.2±1.2
glcm	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.5	48.0±6.3	-5.7±1.0	23.1±5.1
hdac2	1	0.0±0.0	1.2±1.8	2.0±1.6	4.7±2.1	43.9±16.0	3.3±7.2	27.2±14.7
hdac8	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.1	43.1±3.0	-6.5±0.1	17.1±1.2
hivint	1	0.0±0.0	0.0±0.0	0.4±0.5	2.2±0.3	39.9±7.0	-5.7±1.4	14.5±3.9
hivrt	2	0.0±0.0	0.0±0.0	0.0±0.0	1.9±0.1	36.0±3.0	-6.6±0.1	10.9±0.9
hmdh	1	13.7±19.3	2.7±3.9	1.5±2.1	19.7±22.8	51.5±12.3	13.5±28.0	35.6±16.8
hxk4	1	0.0±0.0	0.0±0.0	1.5±2.1	4.5±2.3	46.0±18.5	-3.9±4.0	28.7±17.3
inha	1	0.0±0.0	2.0±2.8	2.1±1.5	5.7±1.8	54.8±10.9	2.9±7.0	36.9±10.7
nos1	1	0.0±0.0	0.0±0.0	0.8±1.1	3.8±1.2	55.5±11.2	-3.8±2.8	28.7±9.4
parp1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.1	41.4±5.9	-6.2±0.4	18.8±1.7
pde5a	1	0.0±0.0	0.0±0.0	0.2±0.2	3.8±0.7	59.7±3.1	-4.0±1.7	27.9±4.0
pgh1	2	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.1	37.4±3.7	-6.2±0.2	13.2±1.3
pgh2	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.2	28.9±9.8	-6.7±0.0	11.9±4.3
ptn1	2	0.0±0.0	0.0±0.0	0.0±0.0	2.1±0.2	38.8±5.8	-5.9±0.5	13.5±2.2
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.9±0.4	38.4±8.0	-6.7±0.0	18.6±4.8
pyrd	1	0.0±0.0	0.0±0.0	0.0±0.0	2.1±0.3	33.0±7.5	-6.6±0.2	12.3±4.0
tysy	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.1	22.2±3.4	-6.7±0.0	10.0±2.5
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	4.3±1.4	54.3±12.2	-6.1±0.9	31.2±12.0
ada17	1	0.0±0.0	0.0±0.0	0.0±0.0	3.8±1.4	52.2±13.4	-5.3±2.1	28.0±11.1
bace1	2	0.0±0.0	0.6±0.8	0.3±0.4	2.7±0.1	52.1±2.9	-4.4±2.7	19.6±1.2
dpp4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.2	45.5±4.7	-6.7±0.0	17.4±2.3
fa10	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.3	44.3±4.9	-5.7±0.8	17.2±3.1
hivpr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.7	51.4±10.2	-6.1±0.6	22.9±6.1
mmp13	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±1.0	39.7±13.2	-5.9±1.0	19.2±9.2
thrb	2	0.0±0.0	0.3±0.4	0.2±0.2	2.4±0.1	50.2±2.7	-5.0±1.1	16.2±1.3
try1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.5	41.9±10.5	-6.7±0.0	18.8±6.1
urok	1	0.0±0.0	0.0±0.0	0.0±0.0	3.9±1.1	59.9±9.6	-5.8±0.9	29.2±7.8

Table 44: **TopU-LBVS-few GAT results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	0.5±0.7	1.0±0.9	0.7±0.5	2.5±0.2	48.3±1.4	-2.1±2.8	16.4±1.7
cp2c9	2	2.7±1.9	1.4±0.4	1.7±0.6	4.1±0.9	60.1±2.7	3.5±2.9	24.3±3.0
cp2d6	2	0.0±0.0	1.2±0.3	1.2±0.7	3.0±0.4	52.5±2.2	1.4±3.1	19.9±1.8
cp3a4	2	0.0±0.0	0.9±0.8	1.0±0.7	3.0±0.5	54.6±2.7	-0.4±4.5	19.4±2.6
GPCR								
5ht6r	1	2.7±3.9	2.0±0.0	1.3±0.5	3.8±0.5	45.4±3.6	3.8±1.4	23.5±1.3
aa2ar	2	0.0±0.0	0.5±0.4	0.7±0.2	2.8±0.1	54.5±1.1	-2.2±1.0	19.2±0.6
aa3r	2	0.0±0.0	1.1±0.8	1.0±0.9	2.7±0.5	46.3±4.8	0.0±4.3	17.8±3.4
adrb1	2	0.0±0.0	0.5±0.7	0.3±0.4	2.5±0.5	46.9±10.2	-3.9±2.1	18.0±4.8
adrb2	2	1.5±2.1	0.6±0.9	0.9±0.8	4.5±2.2	53.5±6.5	-0.3±4.3	20.5±4.6
cnr2	2	0.0±0.0	0.6±0.8	1.2±0.4	3.1±0.4	53.0±4.8	-1.3±2.7	22.3±3.1
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.6	44.6±11.9	-6.6±0.2	20.3±7.4
drd2	2	0.0±0.0	0.3±0.4	0.6±0.5	2.5±0.2	49.9±2.7	-3.3±1.8	17.5±2.1
drd3	2	3.4±2.4	1.5±0.5	0.9±0.3	3.7±0.8	53.8±3.0	2.0±2.7	22.5±2.8
hrh3	2	0.0±0.0	0.5±0.7	1.5±0.6	4.3±0.8	62.3±4.6	2.0±2.3	28.8±3.5
oprd	2	1.5±2.1	0.9±0.8	1.4±0.4	3.3±0.7	52.4±5.7	2.0±3.8	21.3±2.9
oprk	2	2.0±2.8	2.3±2.5	1.6±1.5	5.4±4.2	54.5±11.4	5.2±12.7	24.2±11.7
oprm	2	0.0±0.0	0.9±0.7	0.9±0.4	2.7±0.8	48.3±11.4	-1.4±2.8	17.8±5.6
Ion Channel								
5ht3a	1	2.7±3.9	2.6±0.9	1.7±0.5	5.7±3.0	54.4±5.3	7.7±6.2	28.3±5.6
cac1h	1	0.0±0.0	0.0±0.0	0.5±0.8	3.2±1.1	47.7±13.7	-3.8±2.4	22.1±9.5
kcnh2	2	0.0±0.0	0.8±0.6	0.9±0.2	2.4±0.5	47.5±5.9	-1.3±2.7	14.3±3.0
scn5a	1	3.4±4.8	1.6±1.1	1.2±0.0	7.3±5.5	52.3±7.3	3.5±4.8	25.8±3.7
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.6	39.6±11.4	-6.7±0.0	20.5±6.3
Kinase								
abl1	2	1.7±2.4	0.7±0.5	1.1±0.7	3.5±0.9	47.2±6.3	-0.4±2.6	18.5±2.8
akt1	2	3.9±2.8	2.6±1.1	2.0±0.9	6.6±2.2	56.2±8.2	8.0±5.1	27.1±6.1
akt2	1	0.0±0.0	1.4±2.0	1.5±0.5	4.2±1.9	56.9±9.3	2.2±7.4	27.6±9.1
braf	2	2.7±3.9	1.8±1.5	1.8±1.3	4.6±1.5	56.3±0.6	4.8±8.5	27.4±5.4
cdk2	2	1.7±2.4	0.7±0.5	0.9±0.5	2.9±0.7	51.2±5.5	-1.4±3.5	19.6±3.9
csf1r	1	0.0±0.0	0.7±1.0	0.7±1.0	3.5±1.3	52.7±12.5	-1.4±5.1	24.1±9.1
egfr	2	0.1±0.2	1.0±0.7	1.1±0.7	2.8±0.8	51.5±11.1	-0.5±3.7	18.0±5.9
fak1	1	0.0±0.0	1.4±2.0	0.7±1.0	2.8±1.1	41.4±10.0	-1.6±6.4	17.8±7.5
fgfr1	2	0.0±0.0	0.9±0.6	1.1±1.1	3.0±1.0	50.4±9.3	-1.3±4.8	20.4±6.6
igf1r	2	6.8±0.0	2.5±0.7	2.3±0.6	5.5±1.2	56.2±3.2	9.2±4.4	29.0±3.5
jak2	2	0.0±0.0	1.2±0.0	1.0±0.6	2.9±0.8	51.4±5.8	-0.3±3.8	20.0±4.6
kit	2	0.0±0.0	0.6±0.8	0.3±0.4	2.6±0.3	50.0±2.1	-4.4±2.9	18.8±2.7
kpcb	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.2	38.6±4.6	-6.5±0.2	16.0±3.2
lck	2	0.0±0.0	0.3±0.5	1.2±0.7	2.9±0.5	50.6±5.0	-0.6±3.5	20.1±3.5
mapk2	2	0.0±0.0	0.5±0.8	1.4±0.4	3.3±0.1	57.1±4.3	-2.2±2.0	23.9±0.6
met	2	2.0±2.8	0.9±0.6	0.9±0.6	4.9±3.1	47.2±11.4	-0.5±4.7	19.2±7.3
mk01	2	2.3±3.2	0.5±0.7	0.8±0.0	3.6±1.8	43.1±6.3	-1.2±3.2	17.8±3.3
mk10	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.3	49.9±4.2	-6.2±0.3	19.5±2.5
mk14	2	0.0±0.0	0.8±1.1	0.4±0.6	2.3±0.7	37.2±6.7	-2.7±4.9	13.3±4.8
mp2k1	1	0.0±0.0	0.0±0.0	0.4±0.5	2.5±0.7	41.5±11.4	-4.4±1.5	17.3±6.3
plk1	2	2.3±3.2	2.0±0.7	1.8±0.7	4.4±0.4	58.3±2.8	5.2±2.5	27.4±2.2
rock1	1	0.0±0.0	1.4±1.0	0.7±0.5	3.1±0.3	45.0±9.3	-0.3±2.4	21.0±3.6
src	2	0.0±0.0	1.1±0.9	1.4±0.7	2.9±0.6	51.8±5.5	-0.5±3.8	20.0±4.0
tgfr1	1	0.0±0.0	0.0±0.0	0.7±0.9	3.8±0.7	60.8±6.0	-4.0±1.8	29.5±5.4
vgfr2	2	1.0±1.4	1.2±1.2	0.9±0.8	2.9±0.4	50.7±4.0	0.2±4.7	18.4±1.6
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.5±0.5	48.5±5.3	-5.4±1.3	25.8±5.3

Table 45: **TopU-LBVS-few GAT results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	0.5±0.7	0.8±0.6	2.7±0.2	49.8±0.8	-3.3±3.3	18.8±1.7
esr1	2	0.0±0.0	0.6±0.4	1.0±0.7	3.0±0.3	57.4±2.3	-2.0±3.1	20.4±2.3
esr2	2	0.0±0.0	0.6±0.8	1.5±0.8	3.6±1.8	49.5±12.7	1.0±5.8	23.4±9.6
gcr	2	2.7±3.9	1.2±0.8	0.9±0.7	4.0±2.5	44.9±8.3	0.2±6.4	19.4±6.3
mcr	1	0.0±0.0	2.7±3.9	1.5±2.1	6.4±4.1	38.8±25.6	2.1±9.8	32.2±23.2
ppara	1	6.8±4.8	2.4±0.0	2.1±0.6	7.2±1.4	56.5±8.1	9.4±2.9	31.4±3.5
ppard	1	9.1±6.4	1.8±1.3	1.4±0.0	6.4±2.8	40.6±5.8	6.9±5.9	23.4±5.4
pparg	2	0.0±0.0	0.9±0.6	1.5±0.6	3.0±0.4	50.4±4.0	0.6±2.1	21.6±2.8
prgr	2	0.0±0.0	1.2±0.8	0.9±0.7	3.1±0.5	51.0±4.8	-1.1±4.1	21.9±4.0
rxra	1	0.0±0.0	0.0±0.0	0.8±1.1	3.0±0.6	47.2±4.0	-5.0±2.3	22.1±4.8
thb	1	13.7±19.3	3.0±4.3	1.6±2.3	19.7±23.9	56.9±21.3	13.8±27.9	35.4±23.6
Other Enzymes								
aces	2	2.7±3.9	0.9±0.7	1.0±0.5	3.3±0.9	49.4±3.5	0.6±4.7	19.3±2.7
ampc	1	0.0±0.0	2.0±2.8	2.1±1.5	7.1±4.5	46.3±7.8	7.0±12.4	31.2±11.4
aofb	2	2.3±3.2	0.5±0.7	0.3±0.4	2.9±1.0	49.0±7.2	-3.0±4.4	18.6±5.4
cah2	2	2.3±3.2	1.5±2.1	1.3±0.7	5.2±3.6	54.2±8.6	3.9±10.1	25.2±8.6
dhi1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.7±0.8	51.9±3.6	-5.4±1.7	27.5±6.2
dyr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.2±0.5	48.5±6.2	-5.4±0.9	23.8±4.9
fnta	1	0.0±0.0	0.0±0.0	0.0±0.0	4.2±1.5	53.8±7.9	-4.6±3.0	30.0±10.1
glcm	1	0.0±0.0	0.0±0.0	0.8±1.1	3.2±0.3	48.8±6.5	-4.9±1.9	23.7±3.3
hdac2	1	0.0±0.0	0.0±0.0	0.7±0.9	3.3±0.9	47.1±13.7	-3.6±3.5	22.7±8.1
hdac8	1	0.0±0.0	0.0±0.0	0.5±0.8	3.2±0.8	53.9±10.6	-5.0±1.6	24.2±6.8
hivint	1	0.0±0.0	2.2±1.8	1.1±0.9	3.6±1.3	47.2±12.6	2.5±6.0	23.8±8.8
hivrt	2	0.0±0.0	0.4±0.6	1.3±0.5	2.7±0.6	45.9±7.9	-1.0±2.5	18.9±4.2
hmdh	1	0.0±0.0	0.0±0.0	0.0±0.0	5.6±2.5	57.1±17.7	-5.0±1.6	38.7±16.7
hxk4	1	0.0±0.0	0.0±0.0	1.5±2.1	4.6±3.7	25.6±24.0	-0.9±8.3	20.8±24.2
inha	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.6	33.2±11.7	-5.9±0.8	17.7±8.4
nos1	1	0.0±0.0	1.5±2.1	0.8±1.1	3.9±1.5	44.4±12.0	-0.9±5.8	25.5±11.1
parp1	1	0.0±0.0	1.5±2.1	1.6±1.1	3.9±1.5	44.4±17.7	0.0±5.0	24.8±13.2
pde5a	1	0.0±0.0	0.0±0.0	1.4±1.2	4.7±1.5	66.5±10.6	-1.2±2.9	33.2±8.0
pgh1	2	0.0±0.0	1.5±0.0	1.0±0.4	3.5±0.5	53.5±7.7	2.1±1.6	24.4±3.5
pgh2	1	0.0±0.0	0.0±0.0	2.1±1.5	5.6±1.5	56.6±7.1	1.8±5.6	37.6±7.9
ptn1	2	2.3±3.2	2.0±1.9	1.3±0.9	4.6±3.3	44.4±12.7	3.5±9.3	21.1±10.5
pygm	1	0.0±0.0	2.7±3.9	1.5±2.1	7.0±3.5	55.4±14.8	1.2±10.3	39.9±16.0
pyrd	1	0.0±0.0	1.2±1.8	1.3±0.9	4.3±2.4	45.7±10.0	1.1±8.4	23.4±10.1
tysy	1	0.0±0.0	2.7±3.9	1.5±2.1	7.5±6.3	55.4±21.3	2.9±13.7	37.4±25.2
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±0.5	46.0±8.7	-6.7±0.1	24.2±6.3
ada17	1	0.0±0.0	3.9±5.5	3.2±2.6	8.9±6.2	60.4±15.0	9.7±15.4	45.6±20.3
bace1	2	2.7±3.9	0.6±0.8	0.6±0.4	4.0±2.0	46.4±7.3	-0.5±5.1	19.8±5.0
dpp4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.3	47.1±6.2	-5.7±1.0	20.1±3.6
fa10	1	0.0±0.0	0.8±1.1	0.4±0.6	3.0±0.9	49.3±13.5	-3.3±3.8	21.3±8.3
hivpr	1	0.0±0.0	2.5±1.8	2.0±0.0	5.3±1.5	54.3±7.4	6.2±5.4	31.9±6.3
mmp13	1	0.0±0.0	0.0±0.0	0.0±0.0	3.5±0.7	55.1±8.2	-5.9±0.6	27.0±6.1
thrb	2	0.0±0.0	0.6±0.4	0.5±0.0	2.2±0.3	43.0±7.8	-3.9±0.8	14.1±2.8
try1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.5±0.2	48.4±6.2	-5.2±1.5	26.2±1.2
urok	1	0.0±0.0	1.5±2.1	0.8±1.1	4.3±1.8	50.2±9.2	-0.2±7.1	28.6±9.6

Table 46: **TopU-LBVS-few GAT results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	1.1±1.6	1.5±0.6	1.1±0.3	2.6±0.4	46.7±3.1	0.0±2.4	17.0±2.4
cp2c9	2	0.0±0.0	0.3±0.4	0.7±0.2	2.3±0.1	45.7±2.2	-3.1±1.3	16.0±0.9
cp2d6	2	1.1±1.6	1.2±0.3	1.1±0.3	2.6±0.4	42.1±1.7	0.2±1.9	15.6±1.4
cp3a4	2	2.1±1.5	1.1±0.8	1.0±0.0	3.0±0.3	47.2±6.6	0.2±2.2	17.4±1.7
GPCR								
5ht6r	1	0.0±0.0	0.7±0.9	0.7±0.9	2.7±0.6	45.5±5.8	-2.3±4.1	19.5±3.8
aa2ar	2	0.0±0.0	0.0±0.0	0.7±0.2	2.1±0.2	42.0±3.9	-4.7±0.7	13.5±2.1
aa3r	2	0.0±0.0	0.8±0.7	0.7±0.4	2.1±0.3	39.0±3.6	-2.4±2.6	13.0±2.5
adrb1	2	0.0±0.0	0.5±0.7	0.3±0.4	2.2±0.2	41.9±1.9	-4.4±2.3	14.6±1.3
adrb2	2	0.0±0.0	0.6±0.4	0.6±0.4	2.8±0.4	52.9±6.6	-2.0±1.1	19.6±2.7
cnr2	2	0.0±0.0	0.0±0.0	0.0±0.0	2.1±0.1	41.2±2.7	-6.3±0.3	14.0±1.2
cxcr4	1	0.0±0.0	2.0±2.8	1.1±1.5	5.3±3.0	49.5±17.9	1.3±8.1	32.7±17.8
drd2	2	0.0±0.0	0.6±0.4	0.3±0.2	2.5±0.5	46.2±10.1	-3.3±1.0	16.0±3.9
drd3	2	3.4±2.4	2.2±0.0	1.1±0.0	4.1±0.4	53.0±2.3	5.2±0.5	23.9±0.9
hrh3	2	0.0±0.0	1.0±0.7	0.8±0.0	3.4±0.4	59.5±4.7	-1.1±2.4	24.6±2.7
oprd	2	0.0±0.0	0.9±1.3	0.8±0.8	3.1±0.5	54.4±4.1	-1.2±5.6	20.4±3.2
oprk	2	3.9±2.8	2.0±2.0	1.4±1.5	6.3±5.4	50.4±12.4	5.4±11.8	23.2±11.2
oprm	2	0.0±0.0	1.1±0.4	1.4±0.4	3.4±0.5	57.3±5.3	1.2±2.8	22.8±2.9
Ion Channel								
5ht3a	1	2.7±3.9	1.3±1.8	1.0±0.8	4.0±1.6	56.3±6.9	1.1±7.6	25.9±6.8
cac1h	1	0.0±0.0	1.1±1.5	0.5±0.8	3.3±0.3	46.6±6.1	-1.6±3.8	23.4±2.2
kcnh2	2	0.5±0.6	0.7±0.1	0.8±0.2	2.3±0.1	45.2±0.6	-1.1±0.9	13.6±0.5
scn5a	1	0.0±0.0	0.0±0.0	0.4±0.6	2.6±0.3	46.2±4.5	-4.3±2.0	18.7±4.1
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.9±1.6	41.7±19.8	-5.8±1.1	25.9±15.2
Kinase								
abl1	2	0.0±0.0	0.3±0.5	0.5±0.4	2.3±0.3	42.4±6.9	-3.2±2.1	15.2±2.8
akt1	2	2.0±2.8	1.3±1.1	1.3±0.5	4.3±1.5	48.4±3.5	2.9±4.9	22.4±1.1
akt2	1	0.0±0.0	0.0±0.0	0.7±0.5	3.0±0.2	54.3±1.9	-4.4±1.4	22.4±1.8
braf	2	0.0±0.0	0.6±0.8	0.3±0.4	2.5±0.5	42.2±7.3	-4.2±3.2	16.8±4.6
cdk2	2	0.0±0.0	0.0±0.0	0.2±0.2	2.0±0.1	39.6±3.1	-5.9±0.4	12.4±1.1
csf1r	1	0.0±0.0	0.7±1.0	0.4±0.5	2.5±0.3	42.5±2.2	-3.6±2.9	17.3±2.4
egfr	2	0.0±0.0	0.4±0.3	0.6±0.5	2.4±0.5	47.6±7.7	-3.2±2.2	15.8±3.8
fak1	1	0.0±0.0	0.0±0.0	0.7±0.5	2.7±0.0	46.0±3.3	-2.9±1.6	20.0±0.1
fgfr1	2	0.0±0.0	0.0±0.0	0.4±0.3	2.1±0.4	35.8±7.1	-4.2±1.6	13.1±3.8
igf1r	2	2.3±3.2	0.5±0.7	1.3±0.4	3.1±0.7	48.3±5.2	-0.3±3.3	20.7±3.0
jak2	2	0.0±0.0	0.4±0.6	0.4±0.3	2.3±0.1	46.1±2.5	-4.2±1.4	15.6±1.1
kit	2	0.0±0.0	0.0±0.0	0.3±0.4	2.0±0.4	34.4±12.3	-5.9±0.8	12.2±5.5
kpcb	1	0.0±0.0	0.0±0.0	0.7±0.9	2.9±0.4	45.8±5.9	-4.9±1.8	21.7±3.9
lck	2	0.0±0.0	1.4±1.3	0.9±0.5	2.7±0.8	44.6±5.5	0.1±4.7	17.5±4.5
mapk2	2	0.0±0.0	0.0±0.0	0.8±0.0	2.5±0.4	47.7±7.0	-3.7±0.4	18.3±3.1
met	2	0.0±0.0	1.3±1.1	0.9±0.3	3.2±0.7	53.3±7.6	-0.2±2.6	22.5±4.6
mk01	2	0.0±0.0	1.5±1.2	1.3±0.4	3.6±0.7	56.6±6.1	1.4±3.5	25.3±3.8
mk10	1	0.0±0.0	0.8±1.1	0.8±0.6	3.3±1.1	43.0±11.6	-0.6±4.7	20.7±8.7
mk14	2	0.0±0.0	0.0±0.0	0.4±0.3	2.3±0.7	42.6±13.2	-5.0±0.7	15.3±5.8
mp2k1	1	0.0±0.0	0.7±1.0	0.7±1.0	2.9±1.2	41.0±9.4	-1.6±7.1	17.9±8.2
plk1	2	2.3±3.2	1.0±1.4	1.0±1.4	3.2±1.5	47.4±6.6	-0.4±8.0	20.2±7.4
rock1	1	0.0±0.0	0.0±0.0	0.4±0.5	3.4±1.3	49.1±15.1	-2.8±3.0	22.8±10.0
src	2	0.0±0.0	1.1±0.0	1.4±0.2	3.0±0.3	50.5±3.0	1.3±0.7	20.6±1.6
tgfr1	1	0.0±0.0	0.0±0.0	0.7±0.9	3.0±0.7	49.1±11.2	-5.1±1.7	22.3±6.7
vgfr2	2	0.0±0.0	0.7±0.6	1.1±0.8	2.8±0.5	50.6±3.1	-1.1±4.3	18.1±3.1
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±0.5	51.8±6.7	-6.5±0.2	25.4±4.5

Table 47: **TopU-LBVS-few GAT results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	1.5±1.2	0.8±0.6	2.8±0.2	45.8±4.9	-0.6±3.4	19.2±1.7
esr1	2	0.0±0.0	0.3±0.4	0.3±0.4	2.2±0.4	41.8±5.5	-3.8±2.8	14.2±3.2
esr2	2	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.1	35.3±3.7	-5.6±0.2	12.7±1.7
gcr	2	0.0±0.0	0.6±0.8	0.9±0.0	2.9±0.3	50.9±5.1	-2.3±1.3	21.2±2.1
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.4	21.0±13.5	-6.7±0.0	9.3±6.5
ppara	1	0.0±0.0	0.0±0.0	0.4±0.6	2.7±0.6	45.4±7.0	-4.9±2.0	18.8±5.4
ppard	1	4.6±6.4	1.8±2.6	1.9±1.8	5.5±4.2	42.7±8.0	6.0±13.8	24.8±12.0
pparg	2	0.0±0.0	0.0±0.0	0.4±0.6	2.3±0.5	43.6±9.1	-5.0±1.5	15.2±4.6
prgr	2	0.0±0.0	0.0±0.0	0.3±0.4	2.7±0.1	51.0±1.7	-5.4±0.7	19.7±0.8
rxra	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.3	35.5±6.3	-6.7±0.0	14.4±3.8
thb	1	0.0±0.0	1.5±2.1	1.6±2.3	6.1±4.9	55.3±20.2	3.0±13.4	32.4±20.4
Other Enzymes								
aces	2	0.0±0.0	0.6±0.8	0.4±0.4	2.3±0.3	44.3±3.4	-3.2±3.3	15.0±2.2
ampc	1	6.8±9.7	5.9±4.8	3.2±2.6	15.9±11.7	71.9±15.8	20.0±23.4	54.1±22.9
aofb	2	0.0±0.0	0.0±0.0	0.3±0.4	2.5±0.4	47.2±5.8	-4.7±1.2	17.3±3.4
cah2	2	0.0±0.0	0.5±0.7	1.0±0.9	3.1±1.3	50.0±10.7	-1.7±5.5	20.5±7.8
dhi1	1	0.0±0.0	2.0±2.8	1.1±1.5	4.7±2.5	48.1±8.0	-0.3±8.6	30.1±12.1
dyr	1	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.6	38.0±11.8	-6.1±0.8	16.8±7.1
fnta	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.2	33.6±5.6	-6.7±0.0	14.7±2.2
glcm	1	0.0±0.0	1.5±2.1	0.8±1.1	6.2±3.2	57.7±3.4	3.0±10.8	32.2±4.3
hdac2	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.3	41.5±7.5	-6.7±0.0	16.0±3.3
hdac8	1	0.0±0.0	0.0±0.0	0.5±0.8	3.7±0.4	59.4±2.9	-3.9±2.0	28.1±2.7
hivint	1	0.0±0.0	0.7±1.0	1.1±0.9	3.0±1.0	42.8±3.6	-1.0±5.5	19.3±5.2
hivrt	2	3.9±5.5	1.3±1.9	1.1±0.8	3.9±2.5	41.1±7.7	1.3±8.3	17.6±6.8
hmdh	1	0.0±0.0	2.7±3.9	1.5±2.1	6.9±4.9	44.6±4.6	3.0±13.4	32.3±13.6
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.9±1.0	30.8±17.7	-6.6±0.3	16.4±12.0
inha	1	0.0±0.0	0.0±0.0	1.1±1.5	4.2±0.7	56.6±0.7	-4.2±3.1	31.3±4.0
nos1	1	0.0±0.0	1.5±2.1	1.6±1.1	5.2±1.4	57.0±8.8	2.1±4.4	35.7±7.8
parp1	1	0.0±0.0	0.0±0.0	0.8±1.1	2.8±0.8	39.2±13.5	-4.5±3.0	18.6±8.7
pde5a	1	0.0±0.0	0.0±0.0	0.5±0.7	3.3±0.6	57.9±8.3	-4.8±0.7	25.1±4.0
pgh1	2	0.0±0.0	1.0±0.7	1.8±0.9	3.5±1.3	49.1±11.0	2.6±5.2	23.1±7.6
pgh2	1	0.0±0.0	0.0±0.0	2.1±1.5	5.5±0.6	58.5±4.9	-0.2±1.1	38.9±3.8
ptn1	2	2.3±3.2	1.5±1.2	1.0±0.4	3.2±1.4	44.0±6.9	0.5±5.8	18.9±5.3
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	3.7±1.4	41.3±15.5	-5.7±1.5	24.3±13.2
pyrd	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.4	33.9±8.5	-5.8±1.3	14.3±5.8
tysy	1	0.0±0.0	2.7±3.9	1.5±2.1	6.3±4.7	50.6±22.0	0.8±10.6	34.8±23.2
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.2	33.7±6.2	-6.7±0.0	15.7±3.8
ada17	1	0.0±0.0	2.0±2.8	1.1±1.5	5.4±2.6	58.1±13.5	-0.5±8.7	35.0±13.8
bace1	2	0.0±0.0	0.6±0.8	0.3±0.4	2.5±0.4	45.2±7.2	-4.0±2.7	17.4±4.3
dpp4	1	0.0±0.0	1.1±1.5	1.1±1.5	3.8±1.4	53.0±6.0	-1.0±6.7	26.2±7.7
fa10	1	0.0±0.0	1.6±1.1	1.7±0.6	3.9±0.3	53.7±5.4	2.2±3.7	27.2±0.3
hivpr	1	0.0±0.0	2.5±1.8	2.0±0.0	4.8±0.9	54.9±5.9	3.7±4.0	32.6±4.2
mmp13	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.4	40.3±5.7	-5.9±0.9	18.7±4.1
thrb	2	0.0±0.0	0.3±0.4	0.2±0.2	2.2±0.4	39.5±7.3	-4.2±2.2	13.3±3.6
try1	1	6.8±9.7	2.0±2.8	1.1±1.5	7.9±7.3	40.4±5.4	5.1±16.5	24.8±9.5
urok	1	0.0±0.0	0.0±0.0	0.0±0.0	3.2±0.5	50.5±5.8	-5.8±0.6	24.5±4.2

Table 48: **TopU-LBVS-few GIN results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	0.0±0.0	1.2±0.7	0.9±0.2	2.9±0.4	53.9±6.4	-0.6±2.3	19.4±2.5
cp2c9	2	1.4±1.9	1.9±1.0	1.5±0.5	3.6±0.4	55.9±2.7	3.3±3.4	23.1±1.7
cp2d6	2	0.0±0.0	0.5±0.3	1.0±0.4	3.0±0.2	55.2±1.4	-0.9±0.4	20.5±0.9
cp3a4	2	0.0±0.0	1.1±0.3	1.3±0.3	3.2±0.6	56.7±7.7	2.0±1.6	21.4±3.6
GPCR								
5ht6r	1	2.7±3.9	2.0±1.6	2.7±0.5	5.8±0.8	60.9±11.6	8.8±2.4	33.1±5.3
aa2ar	2	1.2±1.8	0.8±0.6	0.8±0.7	3.0±0.3	54.5±3.5	-1.4±3.1	19.8±2.3
aa3r	2	0.0±0.0	1.1±0.8	1.2±0.0	3.1±0.1	50.2±4.1	1.8±2.0	20.1±1.1
adrb1	2	0.0±0.0	0.0±0.0	2.0±1.3	3.5±1.1	54.8±8.1	0.5±3.7	23.9±6.0
adrb2	2	0.0±0.0	0.0±0.0	1.1±0.2	3.7±0.2	64.9±2.3	-0.9±0.8	25.3±1.2
cnr2	2	2.7±3.9	1.2±0.8	1.5±0.4	3.5±0.6	48.8±3.8	2.1±1.9	22.9±1.1
cxcr4	1	0.0±0.0	2.0±2.8	2.1±3.0	7.4±6.6	51.6±27.9	5.0±16.4	35.5±28.0
drd2	2	1.4±1.9	1.4±0.4	1.7±0.6	3.6±0.2	59.6±0.6	2.9±2.4	24.1±1.0
drd3	2	3.4±4.8	1.8±1.0	2.0±0.5	5.1±1.8	62.4±3.6	6.3±6.3	28.4±2.9
hrh3	2	2.3±3.2	1.0±1.4	1.0±1.4	6.2±4.8	57.6±5.5	1.2±9.0	25.5±8.1
oprd	2	1.5±2.1	0.6±0.9	0.6±0.6	3.4±0.9	57.6±5.7	-0.5±4.1	22.8±4.5
oprk	2	0.0±0.0	1.2±1.0	1.2±1.0	3.3±0.7	53.8±4.8	0.9±4.7	22.1±4.3
oprm	2	1.4±1.9	0.9±1.2	1.0±0.9	3.5±1.3	56.7±8.1	0.4±7.2	21.8±6.9
Ion Channel								
5ht3a	1	0.0±0.0	0.7±0.9	0.7±0.5	3.2±0.6	50.7±2.5	-1.3±3.6	22.4±2.3
cac1h	1	0.0±0.0	1.1±1.5	1.6±0.0	6.2±1.8	68.8±10.7	4.0±5.0	38.2±7.5
kcnh2	2	0.9±0.6	1.0±0.1	1.0±0.2	2.7±0.0	52.4±1.7	-1.0±1.0	20.1±5.1
scn5a	1	0.0±0.0	0.8±1.1	0.4±0.6	2.6±0.2	43.0±3.5	-4.2±2.7	18.6±1.9
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.9±0.4	34.2±6.2	-6.6±0.1	18.0±5.8
Kinase								
abl1	2	3.4±2.4	1.8±1.3	1.4±1.1	4.1±1.5	54.3±9.1	3.7±7.2	23.5±7.4
akt1	2	0.0±0.0	1.3±1.1	1.1±0.3	2.8±0.5	48.3±5.3	-0.2±2.3	19.7±3.2
akt2	1	6.8±4.8	2.2±0.0	1.1±0.0	5.7±1.9	51.5±1.8	5.7±2.1	25.5±1.7
braf	2	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.4	48.2±5.9	-5.9±0.6	17.7±3.4
cdk2	2	3.4±2.4	1.1±0.9	1.0±0.4	3.4±0.7	49.5±3.5	1.0±3.2	24.1±7.1
csf1r	1	3.4±4.8	0.7±1.0	1.1±0.0	6.8±4.9	57.0±2.9	1.2±4.9	26.1±2.1
egfr	2	1.1±1.5	0.9±0.8	1.1±0.4	2.8±0.3	48.5±5.1	-0.3±2.7	17.2±2.3
fak1	1	0.0±0.0	1.4±1.0	1.1±0.0	2.9±0.4	39.1±3.7	0.5±2.7	18.8±1.9
fgfr1	2	2.0±2.8	0.9±0.6	0.7±0.5	3.4±1.1	49.7±2.9	-0.4±3.6	20.6±2.3
igf1r	2	2.3±3.2	1.0±0.7	1.5±1.2	3.9±0.9	54.7±4.0	2.2±5.5	24.8±4.7
jak2	2	0.0±0.0	1.2±1.0	1.2±0.8	2.9±0.6	49.4±4.8	-0.4±4.2	19.8±3.7
kit	2	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.3	44.3±4.2	-5.9±1.0	15.7±3.4
kpcb	1	4.6±6.4	2.5±3.5	1.3±1.8	10.8±11.5	53.9±15.4	5.8±17.1	30.7±16.8
lck	2	0.0±0.0	1.4±0.5	1.8±0.2	3.3±0.4	52.5±2.2	2.9±2.1	22.3±1.5
mapk2	2	0.0±0.0	1.1±0.8	0.8±0.7	2.7±0.6	42.6±3.0	-1.2±3.4	17.8±3.0
met	2	0.0±0.0	0.4±0.6	1.1±0.3	2.9±0.4	49.4±6.5	-1.3±2.1	20.8±3.1
mk01	2	2.3±3.2	1.0±1.4	0.5±0.7	5.4±4.2	51.9±4.3	-1.3±7.1	21.1±6.2
mk10	1	0.0±0.0	1.6±2.3	0.8±1.2	4.1±2.8	49.3±14.3	1.0±10.8	23.8±14.6
mk14	2	0.0±0.0	0.0±0.0	0.4±0.6	2.2±0.2	44.0±3.4	-4.8±1.4	15.0±2.0
mp2k1	1	0.0±0.0	0.0±0.0	0.4±0.5	2.5±0.2	44.2±1.3	-4.0±0.8	18.5±1.3
plk1	2	0.0±0.0	2.0±1.9	1.5±0.6	3.5±1.4	45.8±6.1	3.2±7.2	21.3±5.8
rock1	1	3.4±4.8	2.2±0.0	1.8±0.5	7.5±4.4	60.1±3.2	5.7±4.0	33.0±3.5
src	2	3.4±4.8	1.4±1.3	1.6±0.4	5.6±3.3	51.9±4.6	4.5±5.3	23.4±3.0
tgfr1	1	0.0±0.0	0.0±0.0	0.7±0.9	3.9±1.3	57.8±11.6	-3.1±3.9	28.3±10.2
vgfr2	2	0.0±0.0	1.5±1.0	1.0±0.2	3.1±0.3	54.2±2.4	0.6±2.6	20.6±1.5
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.4	40.2±9.1	-6.6±0.1	18.6±5.0

Table 49: **TopU-LBVS-few GIN results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	0.5±0.7	0.8±0.6	2.8±0.6	51.3±10.3	-3.4±1.9	20.2±4.9
esr1	2	1.4±1.9	1.7±0.7	1.6±0.4	3.6±1.2	55.4±8.9	2.7±4.3	23.3±5.7
esr2	2	0.0±0.0	1.2±0.8	1.2±0.8	3.1±0.2	48.8±1.9	0.1±3.0	22.5±1.4
gcr	2	0.0±0.0	3.0±0.8	2.7±0.0	4.9±0.4	61.7±5.1	7.6±0.6	31.5±2.6
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±1.6	31.5±20.4	-5.7±1.5	18.7±15.9
ppara	1	0.0±0.0	0.8±1.1	0.4±0.6	3.0±0.9	48.1±6.7	-2.8±4.9	20.8±5.7
ppard	1	0.0±0.0	0.9±1.3	0.5±0.7	3.1±1.2	43.5±10.9	-1.9±5.7	19.9±7.9
pparg	2	0.0±0.0	0.4±0.6	0.9±0.6	3.0±0.3	52.4±1.6	-1.5±2.5	21.2±1.8
prgr	2	0.0±0.0	1.8±1.5	1.2±0.4	4.1±0.5	62.9±2.6	1.7±3.8	28.4±2.2
rxra	1	6.8±9.7	1.5±2.1	0.8±1.1	11.7±11.2	55.3±5.2	4.5±14.7	31.6±6.2
thb	1	0.0±0.0	0.0±0.0	0.8±1.1	3.0±0.6	39.1±10.1	-4.1±3.3	20.3±6.2
Other Enzymes								
aces	2	0.0±0.0	0.9±0.7	1.0±0.5	2.7±0.4	47.5±4.9	0.2±3.4	18.7±2.3
ampc	1	6.8±9.7	3.9±5.5	2.1±3.0	17.6±21.4	50.2±19.2	14.3±29.8	34.5±25.7
aofb	2	0.0±0.0	0.5±0.7	0.5±0.4	2.6±0.3	49.6±2.6	-3.5±2.5	18.9±2.6
cah2	2	0.0±0.0	0.0±0.0	0.5±0.4	2.5±0.1	49.8±2.2	-5.0±0.8	18.1±1.0
dhi1	1	0.0±0.0	2.0±2.8	1.1±1.5	6.2±2.3	58.3±1.2	2.0±9.4	38.2±7.7
dyr	1	6.8±9.7	1.5±2.1	1.6±1.1	13.1±12.7	63.3±12.3	8.3±14.8	39.6±13.8
fnta	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.6	43.3±7.7	-6.1±0.9	21.3±6.4
glcm	1	6.8±9.7	1.5±2.1	1.6±1.1	14.2±10.7	69.4±2.6	9.8±13.3	45.7±5.4
hdac2	1	0.0±0.0	0.0±0.0	1.3±0.9	3.8±1.3	51.7±10.6	-2.1±3.7	26.9±8.8
hdac8	1	0.0±0.0	0.0±0.0	0.5±0.8	3.5±0.7	53.7±11.7	-2.5±2.4	26.1±5.7
hivint	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.7	47.4±10.2	-5.7±1.1	18.4±6.3
hivrt	2	0.0±0.0	1.8±1.6	1.8±1.1	3.3±1.1	47.9±6.6	2.6±5.7	21.4±5.4
hmdh	1	0.0±0.0	0.0±0.0	1.5±2.1	7.8±4.7	61.5±19.0	1.3±8.0	47.3±24.9
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.9±1.0	50.8±15.4	-6.5±0.2	28.6±11.3
inha	1	0.0±0.0	2.0±2.8	1.1±1.5	5.8±3.1	51.5±3.2	2.4±11.5	31.8±5.7
nos1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.6±0.4	52.9±7.9	-5.2±0.8	27.3±3.9
parp1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.4	44.8±6.2	-6.3±0.3	20.4±4.1
pde5a	1	0.0±0.0	0.0±0.0	0.9±0.7	2.9±0.3	49.7±7.6	-4.3±1.0	21.9±2.6
pgh1	2	0.0±0.0	0.0±0.0	0.5±0.4	2.4±0.5	43.8±7.5	-4.0±1.5	17.1±4.2
pgh2	1	0.0±0.0	2.0±2.8	1.1±1.5	7.3±4.2	57.1±11.8	3.7±13.8	35.5±8.1
ptn1	2	0.0±0.0	2.0±1.4	1.0±0.7	3.7±0.9	49.8±8.0	2.4±5.8	22.5±4.6
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.1	31.2±3.1	-6.7±0.0	14.5±1.2
pyrd	1	0.0±0.0	1.2±1.8	2.0±1.6	4.3±1.4	48.7±6.5	2.0±5.9	28.0±8.2
tysy	1	0.0±0.0	2.7±3.9	1.5±2.1	8.7±7.2	55.2±12.3	5.6±17.3	37.5±16.5
Protease								
ace	1	0.0±0.0	2.7±3.9	1.5±2.1	6.0±4.6	48.5±18.7	0.8±10.6	32.6±22.2
ada17	1	0.0±0.0	0.0±0.0	1.1±1.5	5.2±2.2	57.0±18.7	-2.2±3.9	35.3±15.5
bace1	2	0.0±0.0	1.2±1.7	0.9±0.7	3.2±1.3	41.7±1.3	0.0±7.8	18.0±4.1
dpp4	1	0.0±0.0	0.0±0.0	1.6±1.3	3.7±1.3	53.3±13.7	-1.7±3.8	26.1±10.2
fa10	1	0.0±0.0	0.8±1.1	1.7±0.6	3.3±0.6	44.9±9.9	1.0±2.7	22.7±5.0
hivpr	1	0.0±0.0	1.2±1.8	0.7±0.9	3.3±0.7	47.2±5.6	-2.9±4.8	23.4±4.6
mmp13	1	0.0±0.0	1.5±2.1	1.6±2.3	4.8±2.7	50.9±20.1	-0.4±8.2	29.6±17.3
thrb	2	0.0±0.0	0.9±0.8	0.8±0.6	2.8±0.3	51.1±3.3	-1.3±3.3	19.3±2.4
try1	1	0.0±0.0	2.0±2.8	1.1±1.5	6.5±4.0	59.4±9.8	2.3±12.1	37.0±13.5
urok	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±0.4	48.1±9.8	-5.0±0.6	25.1±4.4

Table 50: **TopU-LBVS-few GIN-FP results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	2.3±1.6	1.5±0.6	1.1±0.3	3.0±0.3	52.6±7.2	0.9±1.9	19.8±2.4
cp2c9	2	2.7±1.9	0.5±0.4	0.8±0.6	3.3±0.9	48.8±6.7	-0.1±3.4	19.4±4.4
cp2d6	2	0.0±0.0	0.7±0.6	1.0±0.3	2.3±0.1	45.6±0.6	-1.9±0.9	15.6±0.3
cp3a4	2	2.1±1.5	1.1±0.3	0.7±0.4	3.0±0.4	45.6±2.2	0.7±2.3	17.0±2.0
GPCR								
5ht6r	1	2.7±3.9	2.6±1.8	1.7±1.2	5.1±2.2	51.3±3.5	6.4±9.5	26.0±6.7
aa2ar	2	1.2±1.8	0.8±0.6	0.8±0.6	3.1±0.6	52.0±3.8	-0.9±3.4	19.1±3.2
aa3r	2	0.0±0.0	0.8±0.7	1.0±0.2	2.4±0.2	46.1±3.5	-1.8±1.5	16.6±1.6
adrb1	2	0.0±0.0	0.5±0.7	0.3±0.4	2.2±0.3	40.1±3.9	-4.6±1.8	14.6±2.7
adrb2	2	0.0±0.0	0.9±0.8	0.9±0.4	2.5±0.5	46.5±7.2	-1.3±3.1	17.2±4.1
cnr2	2	0.0±0.0	0.0±0.0	0.9±1.3	2.8±0.7	47.2±5.2	-2.6±4.1	19.7±4.7
cxcr4	1	0.0±0.0	2.0±2.8	1.1±1.5	5.1±3.2	53.6±20.8	-0.5±8.7	31.6±19.4
drd2	2	0.0±0.0	0.6±0.4	0.6±0.2	2.3±0.3	45.3±6.0	-3.6±0.7	15.0±2.5
drd3	2	1.7±2.4	1.1±0.9	1.3±0.3	5.2±2.2	57.3±7.1	1.4±2.6	24.0±3.0
hrh3	2	0.0±0.0	1.0±0.7	1.3±0.4	3.2±0.1	57.2±1.6	-0.8±0.6	23.3±1.0
oprd	2	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.3	49.4±5.7	-5.5±0.6	16.4±2.7
oprk	2	0.0±0.0	0.0±0.0	0.4±0.3	2.4±0.3	47.0±4.7	-4.7±0.8	16.7±2.7
oprm	2	1.4±1.9	0.6±0.8	1.0±0.7	2.6±0.5	46.5±3.4	-1.7±3.6	16.9±3.2
Ion Channel								
5ht3a	1	0.0±0.0	1.3±0.9	1.0±0.8	2.9±0.4	44.4±4.7	-0.4±3.9	19.9±3.0
cac1h	1	0.0±0.0	0.0±0.0	0.5±0.8	2.7±0.6	43.0±9.9	-4.4±2.3	19.6±5.6
kcnh2	2	0.5±0.6	1.2±0.6	1.2±0.5	2.7±0.5	52.2±5.5	0.3±2.5	16.6±2.5
scn5a	1	0.0±0.0	0.0±0.0	0.4±0.6	2.5±0.1	42.3±5.2	-4.1±2.3	17.8±1.1
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.2±0.8	31.2±13.7	-6.4±0.4	19.2±10.5
Kinase								
abl1	2	0.0±0.0	0.3±0.5	0.7±0.2	2.3±0.4	44.4±9.3	-3.5±1.3	15.8±3.6
akt1	2	0.0±0.0	1.3±1.1	1.3±1.1	2.8±0.6	44.5±8.4	0.3±3.9	18.6±4.4
akt2	1	0.0±0.0	0.7±1.0	0.7±0.5	3.8±1.1	58.3±7.9	-1.2±4.5	26.2±6.9
braf	2	0.0±0.0	0.6±0.8	0.6±0.4	2.7±0.5	46.1±9.3	-3.2±2.3	19.1±5.1
cdk2	2	1.7±2.4	1.4±1.3	1.1±0.7	3.3±1.1	50.9±5.5	1.5±6.6	21.0±4.9
csf1r	1	0.0±0.0	0.0±0.0	0.7±0.5	3.3±0.5	55.5±6.3	-2.7±1.8	24.6±3.8
egfr	2	0.0±0.0	1.1±0.6	0.9±0.3	2.5±0.4	48.0±4.1	-0.0±2.1	16.6±2.1
fak1	1	0.0±0.0	1.4±1.0	1.1±0.9	3.3±0.5	48.7±5.8	0.6±4.3	22.9±3.4
fgfr1	2	0.0±0.0	0.4±0.6	0.7±0.9	2.5±0.6	42.1±5.5	-2.5±4.4	16.8±4.8
igf1r	2	0.0±0.0	1.0±0.7	0.8±0.6	3.0±0.4	51.5±5.2	-1.2±2.6	21.5±3.2
jak2	2	0.0±0.0	0.0±0.0	0.8±0.3	2.5±0.3	48.0±4.5	-3.4±0.6	17.8±2.6
kit	2	0.0±0.0	0.6±0.8	0.3±0.4	2.4±0.8	37.1±14.5	-3.7±3.8	14.8±7.9
kpcb	1	4.6±6.4	2.5±3.5	1.3±1.8	11.8±12.2	57.8±17.6	7.2±18.4	34.7±19.0
lck	2	0.0±0.0	0.0±0.0	0.9±0.2	2.4±0.3	47.9±4.9	-3.6±1.2	16.9±2.6
mapk2	2	0.0±0.0	1.6±1.3	1.4±0.4	3.8±1.1	52.2±10.1	2.9±4.4	25.0±6.1
met	2	2.0±2.8	0.9±1.2	1.1±0.3	3.6±0.8	47.6±9.8	2.5±4.1	22.3±4.1
mk01	2	0.0±0.0	2.0±0.7	1.0±0.4	3.7±1.3	55.8±9.9	1.5±4.1	25.0±7.3
mk10	1	3.4±4.8	0.8±1.1	1.2±1.0	5.6±2.9	54.5±12.1	2.5±6.6	27.1±10.1
mk14	2	2.0±2.8	1.6±0.6	1.6±0.6	4.2±1.5	54.8±9.2	4.8±5.1	24.7±5.8
mp2k1	1	0.0±0.0	2.2±1.8	1.8±1.0	4.5±1.3	58.7±3.6	3.7±5.9	29.9±5.3
plk1	2	2.3±3.2	1.0±1.4	1.0±0.4	4.0±2.2	49.4±4.9	0.2±5.5	20.9±4.6
rock1	1	0.0±0.0	2.2±0.0	1.8±0.5	5.0±1.0	58.1±3.8	6.2±2.6	30.2±3.8
src	2	0.0±0.0	2.5±1.0	1.9±0.2	3.6±0.5	51.7±6.0	5.3±1.6	22.9±3.1
tgfr1	1	0.0±0.0	0.0±0.0	0.7±0.9	4.2±0.6	64.1±3.0	-2.6±3.0	32.4±3.4
vgfr2	2	0.0±0.0	0.7±0.0	0.6±0.3	2.4±0.2	45.8±3.8	-1.7±0.2	16.3±1.8
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.2	38.9±6.0	-6.7±0.0	17.2±2.9

Table 51: **TopU-LBVS-few GIN-FP results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	2.3±3.2	0.5±0.7	0.8±0.6	2.9±1.0	42.7±6.9	-0.6±4.5	17.9±4.7
esr1	2	1.4±1.9	1.1±0.4	1.2±0.5	2.8±0.3	48.0±4.1	-0.1±2.1	18.2±2.2
esr2	2	0.0±0.0	0.0±0.0	0.6±0.8	2.5±0.9	43.2±13.2	-4.0±3.4	16.8±7.9
gcr	2	0.0±0.0	1.2±0.8	0.9±0.0	3.1±0.6	50.2±2.8	-0.6±3.3	21.9±3.2
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.2±1.4	34.2±24.1	-6.7±0.1	18.2±15.6
ppara	1	0.0±0.0	0.8±1.1	0.4±0.6	2.8±0.9	44.4±7.8	-3.1±3.9	19.6±6.5
ppard	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.2	39.3±3.3	-5.6±0.7	16.0±2.3
pparg	2	0.0±0.0	0.4±0.6	0.7±0.5	2.5±0.2	47.8±2.0	-3.2±1.7	18.1±1.5
prgr	2	0.0±0.0	0.0±0.0	0.6±0.8	2.8±0.3	51.0±2.0	-3.8±3.5	19.7±2.7
rxra	1	0.0±0.0	1.5±2.1	0.8±1.1	6.8±1.7	65.3±11.4	2.6±9.2	40.4±5.5
thb	1	13.7±9.7	4.6±3.7	3.2±2.3	22.7±13.8	71.6±13.4	25.1±22.8	50.0±18.7
Other Enzymes								
aces	2	1.4±1.9	0.3±0.4	0.7±0.2	2.6±0.4	48.1±1.7	-2.8±2.9	16.9±2.2
ampc	1	6.8±9.7	3.9±5.5	3.2±2.6	20.8±24.4	55.1±33.5	18.3±31.3	43.5±34.1
aofb	2	0.0±0.0	1.5±1.2	1.0±0.4	3.3±0.7	54.8±6.0	0.5±3.2	23.8±4.1
cah2	2	0.0±0.0	0.0±0.0	1.5±1.1	3.3±0.8	53.0±7.1	0.4±4.2	23.0±5.4
dhi1	1	0.0±0.0	0.0±0.0	2.1±1.5	4.5±1.5	48.4±10.6	-1.5±4.0	30.9±10.9
dyr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.3±0.2	52.3±1.0	-5.6±1.0	25.2±1.7
fnta	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.4	42.0±7.6	-6.6±0.2	19.4±5.1
glcm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.9	36.2±21.1	-6.4±0.3	16.4±10.9
hdac2	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.2	39.0±3.5	-6.5±0.2	16.5±2.4
hdac8	1	0.0±0.0	2.1±1.5	1.6±0.0	4.9±0.8	58.1±2.9	5.1±4.4	32.1±1.7
hivint	1	0.0±0.0	0.0±0.0	1.1±0.9	2.9±0.4	47.0±6.9	-1.5±3.1	21.0±3.0
hivrt	2	3.9±2.8	0.9±0.6	1.5±0.8	4.2±1.5	50.2±6.3	2.1±4.9	21.8±5.1
hmdh	1	13.7±19.3	2.7±3.9	1.5±2.1	19.4±22.6	48.8±7.1	13.2±28.2	33.1±13.5
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	4.7±1.5	57.1±18.3	-6.4±0.3	33.9±14.4
inha	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.2	47.5±3.4	-6.6±0.1	22.9±2.0
nos1	1	0.0±0.0	0.0±0.0	0.8±1.1	4.0±0.3	55.6±8.2	-3.5±1.9	30.5±3.3
parp1	1	0.0±0.0	1.5±2.1	0.8±1.1	3.7±1.6	48.7±17.6	-2.6±5.6	24.7±12.9
pde5a	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.5	45.6±6.0	-4.3±1.1	20.2±3.9
pgh1	2	0.0±0.0	0.5±0.7	0.8±0.6	2.9±0.5	46.9±4.1	-1.4±4.1	19.8±2.2
pgh2	1	0.0±0.0	2.0±2.8	1.1±1.5	5.9±5.1	39.4±12.1	3.3±14.1	23.2±13.6
ptn1	2	4.6±3.2	1.5±1.2	0.8±0.6	5.6±3.7	38.6±9.7	1.7±6.2	17.1±7.1
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.6	32.7±14.9	-6.7±0.0	16.4±8.2
pyrd	1	4.6±6.4	1.2±1.8	0.7±0.9	9.8±10.4	49.8±15.7	3.0±12.8	26.8±13.9
tysy	1	0.0±0.0	2.7±3.9	1.5±2.1	6.3±4.1	53.1±11.1	0.8±10.6	35.6±17.4
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±1.1	42.3±13.3	-6.3±0.6	22.9±11.0
ada17	1	0.0±0.0	0.0±0.0	0.0±0.0	4.0±0.8	57.6±10.2	-6.1±0.2	30.8±6.4
bace1	2	0.0±0.0	0.0±0.0	0.3±0.4	2.6±0.7	46.1±12.6	-5.2±1.0	18.2±5.9
dpp4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.4	47.1±6.5	-6.2±0.4	19.5±4.2
fa10	1	0.0±0.0	1.6±1.1	1.7±0.6	3.2±0.7	41.9±9.5	2.0±2.3	21.4±5.4
hivpr	1	0.0±0.0	0.0±0.0	0.7±0.9	2.8±0.5	39.2±3.1	-2.8±2.9	19.9±3.4
mmp13	1	0.0±0.0	1.5±2.1	0.8±1.1	4.1±2.4	46.7±13.3	-1.1±8.0	25.3±14.9
thrb	2	0.0±0.0	0.6±0.4	0.8±0.4	2.3±0.2	43.6±3.8	-2.4±1.3	15.7±1.8
try1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.5	39.0±9.7	-6.3±0.5	18.6±6.3
urok	1	0.0±0.0	0.0±0.0	0.0±0.0	3.8±1.5	51.8±10.8	-4.8±2.7	26.8±10.5

Table 52: **TopU-LBVS-few GPS results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	2.3±1.6	2.0±0.9	1.6±0.6	4.3±0.6	62.6±1.7	5.5±4.0	25.5±2.2
cp2c9	2	1.4±1.9	1.1±1.0	1.1±0.9	4.4±2.7	51.3±8.3	0.7±6.3	19.7±6.7
cp2d6	2	2.3±1.6	2.1±0.6	1.7±0.6	3.8±0.6	52.6±7.6	5.7±2.5	22.3±3.9
cp3a4	2	0.0±0.0	1.5±0.3	1.2±0.5	3.2±0.4	56.1±2.3	1.6±2.4	21.2±2.1
GPCR								
5ht6r	1	0.0±0.0	1.3±0.9	1.0±0.8	3.3±1.1	47.3±14.2	0.1±4.8	21.8±8.4
aa2ar	2	0.0±0.0	0.8±0.6	1.2±0.3	3.3±0.4	57.8±2.3	0.6±3.7	22.0±2.2
aa3r	2	1.4±1.9	0.8±0.7	0.7±0.2	2.6±0.4	44.0±4.8	-1.2±2.2	16.2±2.0
adrb1	2	2.3±3.2	1.0±0.7	1.3±0.7	3.8±1.0	52.4±6.4	2.2±4.3	23.3±4.0
adrb2	2	1.5±2.1	0.3±0.4	1.4±0.8	3.1±1.1	48.5±9.0	-0.1±4.0	18.9±6.0
cnr2	2	0.0±0.0	1.2±1.7	0.9±0.7	3.3±1.1	49.7±2.4	-0.1±6.6	22.0±5.0
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.6±0.8	53.8±7.0	-6.1±0.9	26.4±7.3
drd2	2	0.0±0.0	1.1±1.6	0.7±0.7	2.8±0.6	49.8±2.4	-1.6±5.5	18.0±3.8
drd3	2	3.4±2.4	1.5±0.5	1.1±0.0	3.6±0.4	53.4±3.4	2.7±1.1	22.5±1.8
hrh3	2	0.0±0.0	0.0±0.0	0.3±0.4	3.4±0.5	61.7±5.5	-3.9±1.4	24.6±3.1
oprd	2	0.0±0.0	1.2±0.4	1.3±0.4	3.1±0.5	53.6±7.1	1.1±2.4	21.3±3.0
oprk	2	2.0±2.8	1.6±1.1	1.0±0.7	3.8±1.9	47.0±8.1	1.6±5.0	20.0±5.6
oprm	2	0.0±0.0	0.9±1.2	1.6±0.5	3.2±0.2	54.5±5.1	0.6±3.3	21.3±1.0
Ion Channel								
5ht3a	1	0.0±0.0	2.0±0.0	1.3±0.5	3.8±0.8	51.2±5.3	3.6±2.5	24.8±3.7
cac1h	1	0.0±0.0	0.0±0.0	0.5±0.8	3.3±1.5	47.5±18.0	-3.0±3.4	22.2±12.1
kcnh2	2	0.5±0.6	0.7±0.6	0.9±0.5	2.4±0.5	48.6±6.0	-1.9±2.8	14.3±3.1
scn5a	1	3.4±4.8	3.2±1.1	2.5±1.0	9.4±6.9	54.9±11.9	12.0±6.5	33.5±9.4
trpa1	1	0.0±0.0	2.7±3.9	1.5±2.1	11.1±10.8	49.8±8.3	8.9±22.1	34.2±13.5
Kinase								
abl1	2	0.0±0.0	1.1±0.9	1.2±0.2	2.8±0.5	47.2±4.1	0.2±3.4	19.0±2.9
akt1	2	5.9±4.8	1.8±0.6	1.1±0.6	5.2±2.5	54.3±6.3	4.9±5.0	25.1±4.9
akt2	1	0.0±0.0	1.4±2.0	1.5±1.4	4.6±2.3	59.1±10.6	2.8±9.6	29.1±10.4
braf	2	0.0±0.0	0.6±0.8	1.2±1.1	3.0±0.3	49.2±0.3	-1.3±3.5	21.5±1.7
cdk2	2	0.0±0.0	1.1±0.0	0.7±0.2	2.6±0.2	47.9±4.4	-1.5±0.6	18.1±2.0
csf1r	1	0.0±0.0	1.4±2.0	1.8±0.5	4.5±1.0	63.5±5.2	3.1±4.4	31.4±4.2
egfr	2	1.1±1.5	1.1±0.3	1.3±0.3	3.4±0.2	57.9±5.0	2.1±2.3	21.9±1.4
fak1	1	0.0±0.0	1.4±1.0	1.1±0.9	3.3±0.8	45.5±6.3	1.2±4.3	22.0±4.9
fgfr1	2	0.0±0.0	1.3±1.9	1.5±0.8	3.8±1.4	56.4±5.7	2.3±7.0	24.8±5.8
igf1r	2	4.6±3.2	2.5±0.7	1.5±0.6	5.4±1.6	54.5±6.3	6.6±2.4	26.8±3.3
jak2	2	0.0±0.0	0.8±0.6	0.6±0.5	2.7±0.5	48.9±4.8	-1.7±3.3	18.5±4.0
kit	2	0.0±0.0	0.0±0.0	0.3±0.4	2.3±0.4	42.0±5.7	-5.1±2.2	15.4±4.3
kpcb	1	0.0±0.0	1.2±1.8	1.3±0.9	4.0±1.6	45.4±19.5	0.9±5.9	25.2±13.5
lck	2	1.7±2.4	1.8±0.5	1.6±0.4	3.6±0.2	54.3±1.7	3.2±0.9	23.0±1.1
mapk2	2	0.0±0.0	0.0±0.0	0.3±0.4	2.4±0.3	46.0±3.5	-5.2±1.1	16.9±3.2
met	2	2.0±2.8	0.9±0.6	0.9±0.6	2.9±0.7	47.2±6.1	-0.3±4.4	19.2±4.9
mk01	2	4.6±6.4	1.5±1.2	1.5±1.1	7.4±6.1	52.8±9.0	4.5±9.0	25.1±8.2
mk10	1	3.4±4.8	1.6±1.1	1.2±1.0	5.1±2.4	59.4±13.0	4.2±7.2	30.9±9.8
mk14	2	0.0±0.0	0.4±0.6	0.6±0.5	2.9±0.4	55.3±5.2	-2.8±2.5	20.7±3.1
mp2k1	1	0.0±0.0	0.0±0.0	0.0±0.0	1.9±0.1	30.8±3.8	-6.4±0.2	11.0±1.9
plk1	2	2.3±3.2	2.0±1.4	1.3±0.9	4.6±1.8	57.1±7.0	3.8±7.1	26.6±6.6
rock1	1	0.0±0.0	0.7±1.0	0.7±1.0	3.8±1.0	56.8±5.9	-0.1±5.9	26.6±5.6
src	2	1.7±2.4	1.4±1.0	0.7±0.5	3.0±0.7	48.6±5.9	0.2±4.4	19.2±4.2
tgfr1	1	0.0±0.0	0.0±0.0	0.7±0.9	3.6±1.0	57.8±9.0	-4.3±2.6	27.6±7.5
vgfr2	2	1.1±1.6	1.0±0.7	1.1±0.5	3.2±0.6	56.4±3.6	0.6±3.5	21.2±3.1
wee1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.3±0.8	43.2±8.9	-5.0±1.3	23.5±8.0

Table 53: **TopU-LBVS-few GPS results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	0.5±0.7	0.5±0.4	2.5±0.3	48.9±2.8	-4.3±1.7	17.9±2.3
esr1	2	0.0±0.0	0.3±0.4	0.9±0.7	2.9±0.2	55.7±1.3	-1.8±2.8	20.0±1.4
esr2	2	2.7±3.9	1.2±0.8	1.2±0.8	3.8±1.3	47.6±4.3	1.4±5.4	22.1±2.5
gcr	2	0.0±0.0	0.6±0.8	1.2±1.1	3.9±1.4	57.2±10.9	0.5±4.9	26.2±8.5
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.8±1.3	42.5±24.8	-6.3±0.4	25.5±16.0
ppara	1	3.4±4.8	1.6±1.1	1.7±0.6	5.5±2.4	54.4±6.2	3.6±3.8	27.9±3.5
ppard	1	9.1±6.4	2.7±2.2	1.9±0.7	9.8±5.3	54.1±3.4	9.3±7.5	31.0±3.1
pparg	2	2.0±2.8	1.8±1.2	1.3±1.1	3.9±0.9	54.4±2.8	4.0±6.3	24.7±3.9
prgr	2	2.7±3.9	1.2±0.8	1.2±0.4	5.7±4.1	49.5±3.1	0.9±3.8	21.7±2.7
rxra	1	0.0±0.0	1.5±2.1	0.8±1.1	5.1±4.1	48.2±22.3	0.9±10.7	26.7±19.1
thb	1	13.7±19.3	3.0±4.3	2.4±2.0	20.6±24.3	55.5±26.8	17.3±27.0	39.7±25.0
Other Enzymes								
aces	2	0.0±0.0	1.4±1.1	1.3±0.7	3.1±0.5	51.9±2.5	1.8±4.1	20.4±2.9
ampc	1	0.0±0.0	2.0±2.8	3.2±2.6	11.3±5.8	68.8±18.4	12.2±13.3	52.0±22.5
aofb	2	0.0±0.0	0.0±0.0	0.5±0.4	2.6±0.3	50.9±5.3	-4.3±1.2	18.9±3.0
cah2	2	0.0±0.0	1.0±0.7	1.0±0.9	3.3±0.3	54.4±2.2	0.1±3.2	23.6±2.1
dhi1	1	6.8±9.7	2.0±2.8	1.1±1.5	13.6±15.4	45.8±7.0	7.1±19.6	25.7±11.9
dyr	1	0.0±0.0	0.0±0.0	1.6±1.1	4.3±1.5	54.7±11.4	-1.1±4.5	30.2±10.0
fnta	1	0.0±0.0	2.0±2.8	1.1±1.5	6.2±3.4	61.1±11.9	1.0±9.7	39.1±17.0
glcm	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.1	42.8±3.1	-6.5±0.2	19.7±1.6
hdac2	1	0.0±0.0	1.2±1.8	0.7±0.9	4.6±1.6	57.8±11.5	-0.2±6.7	30.9±9.6
hdac8	1	0.0±0.0	0.0±0.0	0.5±0.8	3.2±0.5	50.8±5.0	-3.1±2.4	23.7±3.7
hivint	1	0.0±0.0	0.7±1.0	1.1±0.0	3.9±1.1	54.4±9.7	0.4±3.5	25.7±6.8
hivrt	2	5.9±0.0	2.6±1.1	2.0±0.5	6.2±1.4	57.1±2.0	10.0±3.7	28.5±2.5
hmdh	1	0.0±0.0	0.0±0.0	0.0±0.0	5.9±2.1	65.2±11.3	-5.2±1.3	42.8±12.9
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	4.9±3.0	50.4±22.7	-5.4±1.9	31.7±21.7
inha	1	0.0±0.0	2.0±2.8	1.1±1.5	5.1±3.3	38.4±23.7	2.1±9.3	26.4±18.9
nos1	1	0.0±0.0	4.6±3.7	2.4±2.0	10.6±6.3	60.5±20.4	14.9±16.0	41.3±20.8
parp1	1	6.8±9.7	1.5±2.1	1.6±1.1	7.1±5.6	48.0±17.4	4.3±12.2	26.8±13.0
pde5a	1	0.0±0.0	0.9±1.3	0.9±0.7	3.6±1.3	52.3±11.0	-2.1±4.0	25.7±8.2
pgh1	2	0.0±0.0	1.0±0.7	1.5±0.0	2.8±0.4	43.8±5.9	1.0±2.1	19.4±3.2
pgh2	1	6.8±9.7	2.0±2.8	1.1±1.5	14.2±15.5	54.7±9.4	7.2±19.5	32.4±12.7
ptn1	2	2.3±3.2	1.5±1.2	1.3±0.9	3.9±2.0	48.4±9.8	2.2±7.2	21.9±7.7
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	4.4±1.0	54.6±11.1	-6.3±0.3	32.2±9.4
pyrd	1	0.0±0.0	0.0±0.0	1.3±0.9	2.5±0.5	33.1±6.3	-3.6±2.2	15.8±5.7
tysy	1	0.0±0.0	2.7±3.9	1.5±2.1	8.5±6.9	51.7±7.5	5.7±17.2	36.5±13.3
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±0.6	42.5±7.4	-6.6±0.2	23.3±6.4
ada17	1	0.0±0.0	2.0±2.8	1.1±1.5	5.4±2.5	51.6±10.4	1.9±9.4	32.1±10.6
bace1	2	0.0±0.0	0.6±0.8	0.6±0.4	2.7±0.5	46.3±5.7	-3.1±3.2	19.1±4.5
dpp4	1	0.0±0.0	0.0±0.0	0.5±0.8	3.6±0.4	58.6±2.7	-4.0±2.0	27.6±3.0
fa10	1	0.0±0.0	1.6±1.1	1.2±1.0	3.6±1.0	49.9±9.5	0.5±5.1	24.3±7.1
hivpr	1	0.0±0.0	1.2±1.8	1.3±0.9	3.8±1.4	50.0±11.7	-0.3±4.9	27.0±9.2
mmp13	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.1	38.9±2.5	-5.5±1.0	18.8±1.0
thrb	2	3.0±2.1	2.2±0.9	1.3±0.2	3.3±0.5	46.7±4.8	3.9±2.9	19.8±1.8
try1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.2	45.7±2.7	-6.2±0.6	22.8±2.1
urok	1	0.0±0.0	1.5±2.1	1.6±2.3	4.8±2.3	55.1±5.1	0.2±8.8	31.7±10.4

Table 54: **TopU-LBVS-few GPS-FP results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	1.1±1.6	0.5±0.7	0.5±0.5	2.6±0.7	46.2±3.0	-2.5±3.8	15.5±2.8
cp2c9	2	0.0±0.0	0.8±0.7	0.7±0.4	2.9±0.7	53.8±9.3	-1.5±3.5	19.6±4.9
cp2d6	2	2.3±1.6	0.9±0.3	0.8±0.2	2.4±0.4	39.4±2.2	-0.4±2.5	14.5±2.1
cp3a4	2	0.0±0.0	0.2±0.3	0.6±0.3	2.0±0.1	40.0±3.5	-3.7±1.4	12.8±0.8
GPCR								
5ht6r	1	0.0±0.0	0.7±0.9	1.3±0.5	3.0±0.5	45.0±3.3	0.0±2.7	21.1±3.4
aa2ar	2	1.2±1.8	0.5±0.4	0.3±0.2	2.4±0.5	45.8±4.8	-3.4±2.2	15.5±3.0
aa3r	2	0.0±0.0	1.4±1.0	1.0±0.2	2.6±0.5	44.9±7.5	0.1±2.6	17.3±3.7
adrb1	2	2.3±3.2	0.5±0.7	0.5±0.4	2.7±0.9	43.6±5.6	-2.4±4.0	16.9±4.1
adrb2	2	0.0±0.0	0.3±0.4	0.6±0.2	2.6±0.4	50.1±7.3	-2.8±1.1	17.8±3.0
cnr2	2	2.7±3.9	0.6±0.8	0.3±0.4	3.5±1.9	37.5±8.8	-2.5±5.1	15.1±3.4
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.6	46.9±12.2	-6.7±0.1	21.4±7.1
drd2	2	0.0±0.0	0.3±0.4	0.4±0.4	2.5±0.4	47.0±8.2	-3.2±0.7	16.4±3.7
drd3	2	0.0±0.0	0.4±0.5	0.7±0.3	2.3±0.1	44.2±5.5	-3.2±2.0	16.0±1.5
hrh3	2	2.3±3.2	1.0±0.7	1.0±0.4	4.0±1.6	53.1±0.4	0.1±2.6	21.7±0.9
oprd	2	1.5±2.1	1.2±1.2	1.1±0.8	3.5±0.7	56.1±5.3	0.9±5.7	22.3±3.2
oprk	2	2.0±2.8	0.8±1.1	0.6±0.8	2.7±1.1	42.4±10.2	-2.1±5.7	16.0±7.0
oprm	2	1.4±1.9	1.4±0.4	1.4±0.7	3.5±1.1	51.8±10.2	3.1±4.9	21.8±5.8
Ion Channel								
5ht3a	1	0.0±0.0	1.3±0.9	1.3±0.5	3.4±0.3	52.4±5.7	1.1±1.7	24.1±2.4
cac1h	1	0.0±0.0	0.0±0.0	0.5±0.8	2.9±0.4	47.8±11.6	-5.1±1.1	21.6±3.9
kcnh2	2	0.0±0.0	0.5±0.3	0.7±0.3	2.5±0.5	49.7±7.7	-2.2±2.0	14.9±3.2
scn5a	1	3.4±4.8	0.8±1.1	0.8±0.6	6.5±5.9	42.4±7.2	0.8±6.7	19.4±4.6
trpa1	1	0.0±0.0	0.0±0.0	1.5±2.1	4.0±2.1	29.6±19.7	-3.9±4.0	21.7±17.2
Kinase								
abl1	2	1.7±2.4	1.1±0.0	0.7±0.2	2.6±0.3	44.5±3.6	-0.7±1.5	17.3±1.5
akt1	2	3.9±2.8	2.2±1.6	1.5±0.6	5.7±3.6	45.4±9.6	5.7±7.2	21.7±7.1
akt2	1	0.0±0.0	0.7±1.0	0.7±0.5	3.5±1.1	54.4±5.4	-1.7±5.1	24.0±6.0
braf	2	0.0±0.0	0.0±0.0	0.3±0.4	2.4±0.4	44.2±7.0	-5.2±1.5	16.9±4.2
cdk2	2	0.0±0.0	0.3±0.5	0.2±0.2	2.4±0.3	44.6±2.8	-3.5±2.4	16.0±1.9
csf1r	1	0.0±0.0	0.7±1.0	0.7±0.5	4.2±1.4	60.2±11.3	0.0±4.6	28.2±8.7
egfr	2	2.1±1.5	1.1±0.3	1.0±0.5	2.8±0.5	47.0±7.3	1.0±2.7	17.2±3.9
fak1	1	0.0±0.0	0.7±1.0	0.7±0.5	2.8±0.3	49.9±5.2	-3.3±2.0	20.6±3.0
fgfr1	2	0.0±0.0	0.4±0.6	0.7±0.5	2.6±0.3	44.2±6.8	-2.0±1.7	17.9±3.0
igf1r	2	2.3±3.2	2.5±0.7	2.8±0.9	5.5±2.4	56.2±5.4	9.5±7.9	29.5±6.8
jak2	2	2.0±2.8	0.8±0.6	0.6±0.0	2.8±0.5	46.6±2.9	-1.8±2.4	17.3±1.5
kit	2	0.0±0.0	0.0±0.0	0.3±0.4	2.3±0.6	37.3±15.3	-5.0±1.3	14.5±6.9
kpcb	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.6	41.0±13.0	-5.5±0.9	18.7±8.1
lck	2	0.0±0.0	0.0±0.0	0.5±0.4	2.3±0.2	44.6±3.6	-4.6±1.4	15.4±2.6
mapk2	2	0.0±0.0	0.0±0.0	1.1±0.4	2.4±0.3	44.8±2.6	-3.4±1.3	17.5±2.3
met	2	0.0±0.0	0.9±0.6	1.5±0.3	3.5±0.1	57.1±1.0	1.0±0.8	24.7±0.6
mk01	2	0.0±0.0	2.0±0.7	1.0±0.4	3.7±1.0	54.0±9.0	2.7±3.4	24.7±5.2
mk10	1	0.0±0.0	1.6±1.1	1.7±1.2	4.8±2.4	56.4±15.8	3.9±8.0	29.4±12.6
mk14	2	0.0±0.0	0.0±0.0	0.2±0.3	2.3±0.1	45.9±1.3	-5.8±1.0	15.3±1.2
mp2k1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.2	32.2±5.7	-6.0±0.6	12.0±2.7
plk1	2	0.0±0.0	1.5±2.1	1.8±0.9	3.6±1.5	47.2±6.7	3.5±7.2	22.4±6.0
rock1	1	3.4±4.8	0.7±1.0	0.4±0.5	3.3±1.7	45.6±11.8	-2.4±6.0	18.5±8.1
src	2	5.1±4.2	2.1±0.9	1.2±0.2	4.2±0.9	48.4±6.1	5.6±3.2	22.1±3.4
tgfr1	1	0.0±0.0	1.2±1.8	0.7±0.9	5.2±2.7	59.1±8.3	0.7±8.6	31.5±9.0
vgfr2	2	1.1±1.6	0.5±0.7	0.5±0.3	2.3±0.2	41.5±5.0	-2.6±2.6	14.4±1.5
wee1	1	0.0±0.0	0.0±0.0	1.1±1.5	3.6±1.1	45.6±5.5	-3.9±3.8	24.7±8.3

Table 55: **TopU-LBVS-few GPS-FP results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	0.5±0.7	0.8±0.6	2.5±0.7	39.5±10.0	-2.2±4.6	15.4±5.7
esr1	2	0.0±0.0	0.0±0.0	0.0±0.0	2.1±0.5	39.5±9.8	-5.5±0.9	12.9±4.1
esr2	2	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.1	36.2±5.1	-6.1±0.3	12.3±1.7
gcr	2	0.0±0.0	0.6±0.8	0.6±0.4	2.5±0.4	45.0±8.9	-3.5±2.2	17.7±3.5
mcr	1	0.0±0.0	0.0±0.0	3.0±2.1	7.0±3.8	45.8±30.6	2.8±7.3	38.6±27.2
ppara	1	0.0±0.0	2.4±2.0	1.7±1.2	4.2±1.8	40.4±9.0	4.8±8.4	22.2±8.9
ppard	1	9.1±6.4	2.7±2.2	1.9±1.3	10.1±6.4	51.8±10.8	10.1±11.7	30.3±10.9
pparg	2	2.0±2.8	1.3±0.0	0.7±0.0	3.8±1.4	49.9±2.0	0.7±1.7	20.6±1.3
prgr	2	0.0±0.0	0.6±0.8	0.9±0.7	3.0±0.6	52.4±6.0	-2.6±4.1	21.3±4.8
rxra	1	0.0±0.0	0.0±0.0	0.8±1.1	2.9±0.7	42.3±4.7	-5.2±2.1	20.4±5.7
thb	1	0.0±0.0	1.5±2.1	0.8±1.1	3.9±1.9	39.9±11.5	-0.2±7.7	24.3±10.0
Other Enzymes								
aces	2	0.0±0.0	0.3±0.4	0.9±0.9	2.4±0.7	43.5±8.7	-3.2±3.8	15.1±5.1
ampc	1	0.0±0.0	0.0±0.0	0.0±0.0	3.5±0.8	48.2±6.6	-5.0±2.3	25.8±6.1
aofb	2	2.3±3.2	0.5±0.7	0.8±1.1	3.6±1.2	57.8±5.4	-0.5±5.9	24.3±5.6
cah2	2	0.0±0.0	0.0±0.0	0.8±0.6	2.9±0.5	52.2±7.2	-3.6±0.9	21.1±3.2
dhi1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.3±0.9	45.6±13.9	-5.5±1.7	23.1±9.3
dyr	1	0.0±0.0	1.5±2.1	1.6±1.1	4.3±1.6	50.9±13.0	0.3±4.6	30.0±10.4
fnta	1	0.0±0.0	2.0±2.8	1.1±1.5	5.5±3.0	51.8±20.3	0.0±8.2	34.4±18.2
glcm	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.7	45.3±11.2	-5.8±0.7	22.2±7.1
hdac2	1	0.0±0.0	1.2±1.8	0.7±0.9	3.7±1.4	47.8±9.7	-0.7±6.2	24.6±9.2
hdac8	1	0.0±0.0	0.0±0.0	1.1±0.8	4.3±1.3	59.9±12.4	-0.7±2.9	30.7±8.1
hivint	1	0.0±0.0	1.4±1.0	1.1±0.0	3.1±0.5	51.0±6.0	-0.9±2.1	22.7±3.3
hivrt	2	2.0±2.8	0.9±1.2	0.9±1.2	3.7±2.3	47.7±12.1	-0.7±8.1	19.6±10.0
hmdh	1	0.0±0.0	0.0±0.0	0.0±0.0	6.7±2.4	57.1±18.9	-3.3±0.9	44.6±14.9
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.6±1.4	42.3±16.5	-6.1±0.9	24.2±13.7
inha	1	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.4	29.5±5.7	-6.3±0.6	14.1±5.3
nos1	1	0.0±0.0	1.5±2.1	3.2±1.1	6.4±2.6	62.8±10.8	5.3±7.7	40.6±11.2
parp1	1	6.8±9.7	3.0±2.1	1.6±1.1	7.4±5.5	48.3±16.5	6.6±11.5	28.1±13.3
pde5a	1	0.0±0.0	0.9±1.3	0.5±0.7	4.0±2.0	51.9±10.6	-0.9±7.3	24.0±8.6
pgh1	2	0.0±0.0	0.5±0.7	0.8±0.6	2.6±0.3	46.5±2.7	-2.2±2.7	18.1±2.2
pgh2	1	0.0±0.0	2.0±2.8	2.1±1.5	8.2±3.8	68.3±7.1	6.1±11.1	47.3±10.5
ptn1	2	0.0±0.0	1.5±2.1	1.0±0.9	3.3±1.4	47.9±6.7	0.7±7.3	21.0±7.0
pygm	1	0.0±0.0	0.0±0.0	1.5±2.1	5.0±3.5	38.1±16.1	-0.8±8.3	26.8±20.1
pyrd	1	4.6±6.4	1.2±1.8	0.7±0.9	5.5±4.4	39.9±3.1	1.1±11.0	19.1±4.5
tysy	1	0.0±0.0	2.7±3.9	1.5±2.1	8.4±7.0	48.5±10.3	5.7±17.2	34.1±15.5
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.2	30.6±5.0	-6.7±0.0	13.6±2.6
ada17	1	0.0±0.0	0.0±0.0	0.0±0.0	4.6±0.3	61.1±2.6	-4.0±1.4	35.3±2.1
bace1	2	0.0±0.0	0.6±0.8	0.6±0.8	2.7±0.7	47.1±5.9	-3.0±4.7	18.9±5.5
dpp4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.2	37.6±5.3	-6.7±0.0	13.7±2.4
fa10	1	0.0±0.0	0.8±1.1	0.8±0.6	3.0±0.3	47.8±6.3	-1.3±2.9	22.1±2.5
hivpr	1	4.6±6.4	2.5±1.8	1.3±0.9	6.5±4.2	48.1±4.2	5.0±9.7	27.7±4.1
mmp13	1	0.0±0.0	0.0±0.0	0.8±1.1	3.5±0.9	46.8±7.9	-3.6±3.3	25.6±6.5
thrb	2	0.0±0.0	0.0±0.0	0.2±0.2	2.1±0.1	43.4±1.7	-5.6±0.7	14.2±0.4
try1	1	0.0±0.0	0.0±0.0	2.1±1.5	4.8±1.5	52.6±8.1	-0.2±5.3	33.3±8.8
urok	1	0.0±0.0	0.0±0.0	1.6±2.3	4.5±1.8	56.0±6.8	-1.3±6.4	31.0±9.5

Table 56: **TopU-LBVS-few DMPNN results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	1.1±1.6	1.5±1.0	1.6±0.3	3.8±0.9	59.1±3.4	3.6±4.3	23.5±2.9
cp2c9	2	0.0±0.0	1.6±0.7	1.1±0.5	3.7±0.5	62.3±2.6	1.6±3.1	24.3±2.5
cp2d6	2	3.4±0.0	2.4±0.3	2.3±0.7	5.1±1.3	66.0±6.5	8.8±4.0	28.9±4.8
cp3a4	2	2.1±1.5	1.1±0.3	1.6±0.7	4.2±0.7	63.9±5.4	3.7±2.9	25.4±3.4
GPCR								
5ht6r	1	5.5±7.7	3.9±1.6	2.7±0.5	7.9±4.2	59.8±4.6	12.7±7.2	33.9±5.6
aa2ar	2	0.0±0.0	1.1±0.7	1.5±0.4	3.7±0.3	62.1±1.8	1.7±0.6	24.6±1.3
aa3r	2	1.4±1.9	1.1±0.8	0.8±0.6	3.6±0.2	61.2±1.7	0.7±3.7	23.6±1.3
adrb1	2	0.0±0.0	1.5±0.0	1.3±0.4	4.0±0.8	59.6±7.2	2.3±2.5	27.7±4.4
adrb2	2	3.0±4.3	2.5±0.4	2.2±0.8	5.5±2.1	67.9±4.9	8.7±6.2	30.8±6.0
cnr2	2	0.0±0.0	0.6±0.8	0.6±0.8	2.5±0.6	41.9±2.8	-3.1±5.1	16.0±3.6
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.1±0.6	40.3±8.7	-5.6±0.8	22.1±6.3
drd2	2	1.4±1.9	0.9±0.7	1.0±0.5	3.0±0.7	53.9±6.5	-0.9±3.6	19.8±4.4
drd3	2	1.7±2.4	1.1±0.0	0.6±0.0	3.4±0.2	56.1±4.1	0.2±0.7	22.8±2.3
hrh3	2	0.0±0.0	1.0±0.7	1.0±0.7	4.4±0.6	66.9±2.1	0.9±4.4	29.6±3.1
oprd	2	1.5±2.1	0.9±0.8	0.8±0.2	3.8±1.3	56.4±3.3	0.1±3.2	21.5±2.5
oprk	2	0.0±0.0	0.8±1.1	1.0±0.3	2.8±0.4	51.7±5.8	-1.7±2.0	19.7±3.3
oprm	2	0.0±0.0	1.1±1.1	1.3±0.4	3.6±0.5	60.7±3.6	1.8±2.0	23.8±2.4
Ion Channel								
5ht3a	1	0.0±0.0	1.3±0.9	1.3±0.5	3.8±0.5	59.2±5.1	1.4±3.3	27.6±2.6
cac1h	1	0.0±0.0	1.1±1.5	2.7±0.8	6.1±1.5	64.7±15.1	4.8±0.8	39.3±8.2
kcnh2	2	2.3±1.3	2.2±0.5	1.6±0.2	4.0±0.5	61.1±1.5	5.6±1.4	21.6±1.0
scn5a	1	0.0±0.0	0.8±1.1	0.4±0.6	3.0±0.9	49.3±10.6	-2.5±4.4	21.1±7.2
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.1	19.4±5.4	-6.7±0.0	8.3±2.6
Kinase								
abl1	2	0.0±0.0	1.1±0.0	0.5±0.0	2.5±0.3	49.8±5.6	-2.9±0.1	17.6±2.4
akt1	2	0.0±0.0	1.8±0.6	1.1±0.3	3.2±0.4	51.1±7.2	1.7±0.8	21.5±3.3
akt2	1	0.0±0.0	1.4±2.0	1.1±0.9	4.4±2.5	51.9±8.5	2.5±9.2	26.0±9.8
braf	2	2.7±3.9	1.2±1.7	1.2±1.1	6.3±5.6	46.4±10.5	2.5±10.0	21.8±9.2
cdk2	2	1.7±2.4	0.7±1.0	1.1±0.7	4.1±1.8	57.4±4.2	0.1±6.0	22.6±4.3
csf1r	1	10.2±8.4	2.2±1.8	1.5±0.5	10.2±7.7	55.4±6.1	9.7±9.7	30.0±8.0
egfr	2	0.0±0.0	1.1±0.6	1.2±0.5	2.9±0.4	53.8±3.5	0.0±2.8	19.3±2.3
fak1	1	0.0±0.0	0.7±1.0	1.8±0.5	3.7±0.1	52.8±1.2	0.6±2.0	26.4±0.3
fgfr1	2	0.0±0.0	0.9±0.6	0.7±0.0	2.7±0.5	45.3±7.3	-1.0±1.4	18.4±4.1
igf1r	2	2.3±3.2	1.5±1.2	1.5±0.0	4.6±0.1	59.4±5.5	4.0±2.3	28.5±2.3
jak2	2	0.0±0.0	0.8±0.6	0.8±0.3	2.4±0.3	46.3±8.3	-2.9±1.2	16.9±3.2
kit	2	2.7±3.9	0.6±0.8	1.5±0.4	3.9±1.1	56.1±2.1	1.9±4.8	25.9±3.6
kpcb	1	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.2	41.1±2.8	-5.9±1.1	17.5±2.5
lck	2	0.0±0.0	0.0±0.0	0.5±0.4	2.6±0.2	53.4±3.5	-4.6±1.2	18.5±1.8
mapk2	2	2.7±3.9	1.1±1.5	0.5±0.8	4.0±2.7	45.7±7.2	-0.3±8.1	19.2±7.1
met	2	3.9±2.8	2.2±1.2	1.3±0.5	4.5±0.7	52.8±3.9	5.1±1.9	24.6±1.9
mk01	2	6.8±5.6	3.0±2.1	2.0±0.4	7.5±2.6	66.4±2.9	9.5±7.8	33.3±4.0
mk10	1	0.0±0.0	0.0±0.0	0.4±0.6	3.5±1.0	58.3±13.0	-3.8±3.1	25.3±8.6
mk14	2	0.0±0.0	0.4±0.6	1.2±0.5	3.4±0.2	59.3±1.0	-1.6±2.2	23.7±1.2
mp2k1	1	0.0±0.0	0.0±0.0	0.4±0.5	2.3±0.5	38.9±8.9	-5.4±1.0	15.1±5.4
plk1	2	2.3±3.2	2.0±1.4	2.0±0.9	5.9±2.2	66.5±11.5	7.3±7.0	34.4±9.0
rock1	1	0.0±0.0	1.4±1.0	1.5±1.4	4.1±1.6	54.4±6.2	2.1±6.1	26.7±7.4
src	2	0.0±0.0	0.3±0.5	0.5±0.4	2.9±0.3	54.4±4.9	-3.1±1.1	20.1±1.8
tgfr1	1	0.0±0.0	1.2±1.8	1.3±1.8	5.1±2.3	64.9±9.2	0.3±8.5	34.8±11.6
vgfr2	2	2.3±1.6	1.2±0.9	1.0±0.5	3.1±0.4	50.5±2.3	1.6±3.8	19.6±2.0
wee1	1	6.8±9.7	2.0±2.8	1.1±1.5	8.4±7.0	48.4±4.5	5.6±16.3	29.5±7.1

Table 57: **TopU-LBVS-few DMPNN results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	1.0±0.7	0.5±0.4	3.2±0.6	53.8±5.6	-1.1±3.3	22.2±4.8
esr1	2	1.4±1.9	1.7±0.7	1.7±0.7	4.2±0.8	61.8±4.0	4.0±4.0	25.7±3.7
esr2	2	0.0±0.0	0.6±0.8	0.6±0.8	3.3±0.8	53.4±1.7	-1.5±6.0	22.1±3.1
gcr	2	0.0±0.0	3.0±2.2	3.3±0.4	6.4±1.7	70.0±6.6	10.5±5.8	37.4±6.2
mcr	1	0.0±0.0	0.0±0.0	0.0±0.0	3.4±0.6	43.3±9.0	-6.6±0.2	23.8±6.8
ppara	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.4	42.1±9.1	-5.7±0.8	16.3±5.1
ppard	1	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.1	38.9±2.4	-6.4±0.1	15.3±1.7
pparg	2	0.0±0.0	1.3±0.0	1.1±0.3	3.0±0.2	50.7±2.4	0.8±0.1	21.1±0.9
prgr	2	0.0±0.0	0.0±0.0	0.6±0.8	3.6±0.6	62.9±2.1	-2.8±4.5	25.7±3.9
rxra	1	0.0±0.0	0.0±0.0	0.0±0.0	4.1±1.6	53.8±12.9	-3.7±2.3	29.1±10.7
thb	1	0.0±0.0	6.1±2.1	4.8±2.0	17.5±11.0	74.8±15.3	28.2±19.0	62.0±19.6
Other Enzymes								
aces	2	1.4±1.9	1.4±0.4	1.0±0.2	3.0±0.1	53.1±1.4	0.8±0.3	19.9±0.7
ampc	1	0.0±0.0	0.0±0.0	1.1±1.5	4.3±1.1	53.2±8.1	-3.5±2.5	31.7±8.6
aofb	2	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.1	39.5±1.8	-6.6±0.1	12.3±1.0
cah2	2	0.0±0.0	0.5±0.7	0.3±0.4	2.3±0.2	41.7±0.4	-4.0±2.8	15.6±1.8
dhi1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.5	39.8±7.5	-6.7±0.1	17.9±5.6
dyr	1	0.0±0.0	1.5±2.1	0.8±1.1	6.0±2.9	53.2±9.2	2.1±11.3	30.3±1.2
fnta	1	0.0±0.0	0.0±0.0	0.0±0.0	2.3±0.2	32.3±6.7	-6.7±0.0	13.3±3.1
glcm	1	0.0±0.0	0.0±0.0	0.8±1.1	4.3±0.7	58.9±2.6	-2.5±4.6	31.5±3.2
hdac2	1	0.0±0.0	0.0±0.0	0.0±0.0	2.5±0.5	37.6±7.4	-5.8±0.7	17.2±5.3
hdac8	1	0.0±0.0	0.0±0.0	0.5±0.8	3.1±0.6	48.8±5.4	-3.4±1.8	23.5±4.2
hivint	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.2	51.5±1.0	-5.6±0.5	20.4±1.5
hivrt	2	0.0±0.0	0.4±0.6	0.9±0.3	2.5±0.1	47.0±1.3	-2.3±1.1	18.3±0.6
hmdh	1	0.0±0.0	0.0±0.0	1.5±2.1	7.4±1.4	72.7±4.6	-2.2±4.5	52.0±6.9
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.7	35.4±10.8	-6.6±0.2	19.3±7.5
inha	1	0.0±0.0	0.0±0.0	0.0±0.0	2.7±0.6	39.4±12.8	-6.7±0.1	17.5±7.6
nos1	1	6.8±9.7	1.5±2.1	2.4±2.0	9.6±4.4	64.4±6.1	9.7±10.1	42.5±7.5
parp1	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±1.0	40.0±17.6	-5.4±1.9	18.4±10.8
pde5a	1	0.0±0.0	0.9±1.3	0.9±0.7	4.1±1.0	61.3±5.4	-0.6±5.7	29.1±5.6
pgh1	2	0.0±0.0	1.0±0.7	0.5±0.4	2.3±0.1	41.3±2.5	-3.3±2.0	16.4±1.1
pgh2	1	0.0±0.0	0.0±0.0	1.1±1.5	3.6±1.2	46.9±12.7	-4.6±2.7	25.9±9.3
ptn1	2	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.2	37.8±5.9	-6.5±0.2	12.5±2.0
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	3.7±1.9	40.0±23.0	-6.6±0.2	22.3±17.2
pyrd	1	0.0±0.0	1.2±1.8	0.7±0.9	2.9±0.9	39.0±11.1	-2.9±4.9	19.0±6.6
tysy	1	0.0±0.0	0.0±0.0	0.0±0.0	2.8±0.4	34.4±11.0	-6.7±0.0	16.2±6.5
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	3.9±0.8	51.7±10.5	-6.6±0.1	29.0±7.8
ada17	1	0.0±0.0	0.0±0.0	0.0±0.0	4.4±0.9	62.1±6.2	-5.2±2.0	33.4±6.8
bace1	2	0.0±0.0	1.2±1.7	2.1±0.8	4.0±1.2	54.0±6.1	3.7±6.6	26.1±6.0
dpp4	1	0.0±0.0	0.0±0.0	1.6±0.0	3.6±0.5	54.6±6.6	-1.7±0.3	27.3±3.8
fa10	1	0.0±0.0	0.8±1.1	1.2±1.0	3.6±1.0	52.0±10.4	0.2±3.5	25.8±6.8
hivpr	1	0.0±0.0	1.2±1.8	0.7±0.9	3.4±1.3	47.1±11.1	-2.2±4.8	23.9±9.6
mmp13	1	0.0±0.0	0.0±0.0	0.0±0.0	2.4±0.7	34.8±16.3	-6.6±0.2	14.1±8.6
thrb	2	4.6±0.0	0.9±0.0	1.1±0.2	4.9±2.0	53.7±5.8	2.2±0.5	22.3±2.3
try1	1	0.0±0.0	0.0±0.0	2.1±1.5	4.6±1.2	51.9±6.2	-2.2±3.1	32.5±8.5
urok	1	0.0±0.0	1.5±2.1	0.8±1.1	4.7±1.2	57.2±3.9	-1.0±5.6	33.0±6.3

Table 58: **TopU-LBVS-few Molformer results - part 1.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Cytochrome P450								
cp1a2	2	2.3±1.6	2.0±0.9	2.2±0.3	4.5±0.7	58.3±3.9	6.3±2.5	24.4±1.3
cp2c9	2	0.0±0.0	1.4±0.4	1.2±0.3	3.7±0.4	63.5±4.4	1.2±0.9	24.7±2.3
cp2d6	2	3.4±0.0	2.1±0.0	2.0±0.3	4.1±0.7	60.1±6.7	5.8±1.5	25.1±3.2
cp3a4	2	0.0±0.0	1.1±0.8	1.4±0.4	3.7±0.3	63.7±1.2	1.4±3.7	23.9±1.9
GPCR								
5ht6r	1	0.0±0.0	0.7±0.9	0.7±0.5	3.6±0.1	63.4±2.2	-3.4±2.1	26.6±0.8
aa2ar	2	2.5±1.8	1.6±0.0	1.2±0.0	3.6±0.5	57.9±2.2	1.6±1.3	22.5±1.5
aa3r	2	4.1±3.3	1.4±1.4	1.2±0.6	4.0±1.5	54.2±3.1	3.5±6.7	22.3±4.3
adrb1	2	0.0±0.0	0.5±0.7	0.5±0.7	2.9±0.1	54.8±2.6	-3.8±2.8	21.1±1.0
adrb2	2	1.5±2.1	1.9±0.8	1.7±0.4	5.1±0.6	70.9±2.6	6.6±3.2	31.3±2.0
cnr2	2	0.0±0.0	0.6±0.8	0.6±0.8	3.3±0.9	52.3±3.6	-1.3±5.9	22.4±4.1
cxcr4	1	0.0±0.0	0.0±0.0	0.0±0.0	2.2±0.3	27.5±7.7	-6.7±0.0	10.8±3.2
drd2	2	0.0±0.0	0.9±0.7	1.2±0.2	2.9±0.3	54.0±3.5	-0.3±1.1	20.4±1.8
drd3	2	3.4±2.4	2.2±1.6	1.7±0.9	4.9±1.8	59.8±1.6	6.3±6.9	26.7±4.2
hrh3	2	2.3±3.2	1.5±1.2	1.3±1.3	5.2±1.7	63.6±4.2	2.8±6.5	28.2±5.7
oprd	2	1.5±2.1	1.9±0.8	1.1±0.2	3.3±0.1	54.8±3.6	2.4±1.7	22.4±0.7
oprk	2	2.0±2.8	1.6±0.6	1.6±0.6	4.1±1.7	54.3±8.7	2.2±2.1	23.2±4.7
oprm	2	2.7±1.9	1.1±0.4	0.9±0.4	3.7±0.6	60.3±1.8	1.3±3.3	23.8±2.5
Ion Channel								
5ht3a	1	0.0±0.0	0.7±0.9	1.7±0.5	4.3±0.4	61.5±1.6	2.3±3.9	30.0±1.5
cac1h	1	0.0±0.0	2.1±1.5	1.1±0.8	4.4±1.4	54.3±9.3	2.8±5.9	28.4±7.8
kcnh2	2	0.5±0.6	1.5±0.3	1.3±0.3	3.4±0.5	60.4±1.7	2.4±2.5	19.8±1.6
scn5a	1	0.0±0.0	0.8±1.1	1.2±1.0	3.5±1.3	50.7±9.2	-0.4±5.1	24.1±7.7
trpa1	1	0.0±0.0	0.0±0.0	0.0±0.0	5.8±1.5	62.5±8.4	-3.7±2.6	42.6±9.0
Kinase								
abl1	2	3.4±4.8	1.1±0.9	0.9±0.2	3.5±1.4	48.7±3.3	0.8±5.0	20.0±3.7
akt1	2	0.0±0.0	0.9±0.6	0.9±0.3	3.0±0.7	51.8±9.2	-1.1±2.7	21.1±5.4
akt2	1	0.0±0.0	2.2±1.8	1.1±0.9	4.1±0.8	62.5±2.6	0.2±5.1	29.2±3.9
braf	2	0.0±0.0	0.0±0.0	0.9±0.0	2.6±0.2	48.2±5.0	-3.8±0.8	18.7±2.0
cdk2	2	0.0±0.0	1.8±0.5	1.9±0.2	3.6±0.3	55.5±2.9	4.5±1.6	24.1±1.5
csf1r	1	6.8±4.8	2.9±1.0	1.5±0.5	6.8±3.2	44.2±12.2	8.2±6.0	24.4±8.9
egfr	2	1.1±1.5	1.5±0.3	1.1±0.3	3.0±0.5	52.5±5.3	1.8±2.3	19.5±2.6
fak1	1	0.0±0.0	1.4±1.0	1.1±0.9	3.7±1.3	54.4±11.7	0.2±4.4	25.5±8.7
fgfr1	2	2.0±2.8	1.3±1.1	0.9±0.3	3.2±0.6	49.3±3.8	0.3±3.4	20.8±2.9
igf1r	2	0.0±0.0	0.5±0.7	1.0±0.4	3.3±0.7	53.2±4.3	-0.1±3.7	23.3±4.0
jak2	2	0.0±0.0	0.8±0.6	1.0±0.3	2.4±0.2	44.4±2.2	-2.1±2.0	16.5±1.8
kit	2	0.0±0.0	0.0±0.0	0.3±0.4	2.4±0.1	47.2±1.8	-5.1±1.0	17.6±1.2
kpcb	1	0.0±0.0	1.2±1.8	1.3±0.9	4.3±1.4	52.4±8.6	0.5±5.2	29.6±7.7
lck	2	0.0±0.0	0.0±0.0	0.5±0.4	2.4±0.0	47.5±3.2	-4.2±0.7	16.9±0.5
mapk2	2	0.0±0.0	0.0±0.0	1.6±0.7	3.2±0.8	51.1±8.5	-0.2±2.2	22.6±5.1
met	2	2.0±2.8	2.2±1.6	1.1±0.8	4.1±1.6	50.5±0.9	3.6±7.3	23.0±3.7
mk01	2	2.3±3.2	1.0±1.4	1.3±0.7	5.9±4.0	56.1±2.4	1.5±6.8	25.1±4.1
mk10	1	3.4±4.8	1.6±1.1	0.8±0.6	3.9±1.2	46.1±3.7	1.6±5.9	22.5±4.8
mk14	2	3.9±5.5	2.3±2.5	2.5±1.9	7.9±6.8	59.8±8.1	9.6±15.0	29.9±11.4
mp2k1	1	0.0±0.0	0.7±1.0	0.7±0.5	3.1±0.7	49.5±11.6	-1.6±2.9	22.0±5.6
plk1	2	0.0±0.0	1.0±0.7	1.0±0.7	4.1±0.7	61.1±5.2	1.4±3.8	27.6±3.3
rock1	1	0.0±0.0	0.0±0.0	1.1±0.0	2.6±0.3	41.6±9.8	-2.8±1.6	17.7±3.6
src	2	1.7±2.4	1.1±1.5	0.9±0.9	3.0±0.9	47.6±3.4	-0.1±6.4	18.7±3.4
tgfr1	1	0.0±0.0	1.2±1.8	0.7±0.9	4.2±2.7	44.7±13.8	0.1±8.9	22.8±11.5
vgfr2	2	1.1±1.6	1.7±1.2	1.2±0.6	3.1±0.9	51.1±8.3	2.0±5.3	19.8±5.3
wee1	1	0.0±0.0	2.0±2.8	1.1±1.5	6.2±4.7	45.5±2.9	3.3±14.1	27.3±8.7

Table 59: **TopU-LBVS-few Molformer results - part 2.** All values are reported as mean ± std over three random seeds. EF = Enrichment Factor. PR-AUC, ROC-AUC, BEDROC, and LogAUC are reported on a 0–100 scale.

Target	Tier	EF@1%	EF@5%	EF@10%	PR-AUC	ROC-AUC	BEDROC	LogAUC
Nuclear Receptor								
andr	2	0.0±0.0	1.5±1.2	1.0±0.9	3.1±0.8	45.8±3.8	0.8±5.6	20.1±4.7
esr1	2	0.0±0.0	1.4±0.4	1.2±0.4	2.8±0.4	50.7±4.0	0.6±2.4	19.3±2.7
esr2	2	0.0±0.0	0.0±0.0	0.3±0.4	2.5±0.4	46.3±7.0	-4.9±1.8	17.6±4.1
gcr	2	2.7±3.9	1.2±1.7	1.5±0.8	8.0±5.5	68.8±5.4	5.0±9.1	33.9±6.5
mcr	1	0.0±0.0	0.0±0.0	1.5±2.1	6.4±1.5	61.5±8.2	-1.5±3.4	44.8±8.7
ppara	1	0.0±0.0	3.2±4.5	2.1±2.1	5.7±4.6	51.0±14.4	6.8±15.3	28.7±16.8
ppard	1	0.0±0.0	0.9±1.3	1.4±0.0	3.0±0.4	43.3±4.9	0.0±2.4	21.3±3.0
pparg	2	2.0±2.8	1.3±1.9	1.5±0.8	4.0±2.5	45.8±8.3	2.4±8.1	19.8±7.0
prgr	2	0.0±0.0	1.2±1.7	1.5±1.1	4.1±1.5	55.5±6.1	2.5±8.0	26.4±6.2
rxra	1	0.0±0.0	1.5±2.1	0.8±1.1	4.9±2.3	46.4±16.4	1.0±10.1	26.2±9.6
thb	1	0.0±0.0	1.5±2.1	0.8±1.1	6.4±2.0	70.7±7.9	1.9±7.0	43.1±8.6
Other Enzymes								
aces	2	0.0±0.0	0.9±0.7	0.9±0.4	2.8±0.3	52.8±2.9	-1.5±2.3	19.5±1.8
ampc	1	6.8±9.7	5.9±0.0	5.3±1.5	23.1±12.7	83.4±5.3	27.2±11.1	68.5±8.1
aofb	2	0.0±0.0	0.0±0.0	0.0±0.0	2.1±0.0	42.3±0.6	-6.2±0.4	14.5±0.6
cah2	2	0.0±0.0	0.5±0.7	1.3±0.9	3.1±0.6	51.8±1.0	-0.6±5.0	21.5±3.1
dhi1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.8±1.2	46.7±9.4	-5.2±1.2	26.6±9.8
dyr	1	0.0±0.0	0.0±0.0	0.8±1.1	3.6±1.5	44.6±13.0	-3.9±3.7	24.1±11.3
fnta	1	0.0±0.0	2.0±2.8	2.1±1.5	5.6±3.0	49.8±12.5	2.5±9.2	32.9±14.2
glcm	1	0.0±0.0	1.5±2.1	1.6±2.3	5.8±4.7	49.0±20.7	2.5±13.0	29.0±20.5
hdac2	1	0.0±0.0	0.0±0.0	0.7±0.9	3.6±0.5	55.9±4.5	-4.7±1.5	27.8±4.1
hdac8	1	0.0±0.0	0.0±0.0	0.5±0.8	4.0±2.0	54.0±19.8	-3.2±2.5	26.6±13.3
hivint	1	0.0±0.0	0.0±0.0	0.4±0.5	2.5±0.4	44.1±2.0	-4.5±3.0	17.5±3.6
hivrt	2	2.0±2.8	0.9±1.2	0.4±0.6	3.6±2.0	43.4±6.3	-1.0±6.3	17.7±5.2
hmdh	1	13.7±19.3	2.7±3.9	1.5±2.1	20.7±23.4	63.5±16.9	13.8±27.9	44.2±21.0
hxk4	1	0.0±0.0	0.0±0.0	0.0±0.0	4.2±0.3	57.7±3.0	-6.7±0.0	31.2±2.7
inha	1	0.0±0.0	2.0±2.8	3.2±0.0	6.5±1.2	61.0±9.1	4.9±4.8	42.1±6.0
nos1	1	0.0±0.0	0.0±0.0	0.8±1.1	4.8±1.2	65.3±6.7	-3.2±3.1	35.5±8.2
parp1	1	0.0±0.0	1.5±2.1	0.8±1.1	4.9±2.9	47.3±10.4	1.6±9.8	27.3±10.7
pde5a	1	0.0±0.0	0.0±0.0	0.0±0.0	2.6±0.6	45.1±11.5	-6.3±0.3	18.4±5.6
pgh1	2	2.3±3.2	1.0±0.7	1.3±0.4	5.0±3.9	42.8±10.3	0.2±4.3	18.3±6.1
pgh2	1	6.8±9.7	2.0±2.8	1.1±1.5	8.6±7.6	55.8±10.7	5.1±16.6	31.9±12.5
ptn1	2	0.0±0.0	0.5±0.7	0.8±0.6	3.0±0.3	53.2±5.8	-2.4±2.4	21.1±2.2
pygm	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.7	35.6±17.0	-6.7±0.0	17.8±9.3
pyrd	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.1	48.8±3.5	-5.3±0.8	23.2±1.4
tysy	1	0.0±0.0	5.5±3.9	6.1±2.1	14.2±5.1	76.3±13.1	15.8±7.0	70.3±17.3
Protease								
ace	1	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.5	37.1±8.9	-6.7±0.1	19.9±6.5
ada17	1	6.8±9.7	5.9±4.8	3.2±2.6	17.8±10.9	64.1±18.8	23.9±22.6	50.6±21.9
bace1	2	2.7±3.9	1.2±0.8	1.2±1.1	3.8±1.8	46.3±12.0	2.0±6.7	21.6±8.5
dpp4	1	0.0±0.0	0.0±0.0	0.5±0.8	3.5±0.2	55.6±3.8	-4.0±1.5	26.6±1.1
fa10	1	0.0±0.0	0.0±0.0	0.4±0.6	2.9±0.4	49.9±3.7	-3.9±1.1	21.8±2.8
hivpr	1	0.0±0.0	0.0±0.0	0.0±0.0	1.9±0.0	25.6±2.3	-6.7±0.0	9.6±0.8
mmp13	1	0.0±0.0	1.5±2.1	1.6±2.3	4.8±3.0	49.7±11.2	1.3±10.3	29.8±14.0
thrb	2	0.0±0.0	0.9±0.0	0.6±0.2	2.3±0.0	45.3±2.1	-2.5±0.3	15.9±0.2
try1	1	0.0±0.0	0.0±0.0	0.0±0.0	3.2±0.2	41.6±9.5	-5.7±0.6	23.0±3.1
urok	1	0.0±0.0	0.0±0.0	0.0±0.0	4.0±1.4	56.7±17.8	-5.3±1.1	29.1±12.2