Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SWAN-MPO: A Target-Calibrated Multi-Parameter Optimization Framework for Natural-Product Scaffold Triage in Oncology

Saanvii Senthilkumar

saanviisenthilkumar@gmail.com

Independent Researcher

Article

Keywords:

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10863911/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SWAN-MPO: A Target-Calibrated Multi-Parameter Optimization Framework for Natural-Product Scaffold Triage in Oncology

Saanvii Senthilkumar

Independent Researcher, Frisco, TX, USA

Corresponding author: saanviisenthilkumar@gmail.com

ORCID iD: https://orcid.org/0009-0000-2050-1070

Abstract

Natural products occupy structural and physicochemical space that is poorly represented by optimization criteria developed for conventional small molecules. We developed SWAN-MPO, a multi-parameter framework that integrates predicted target binding, absorption/distribution/metabolism/excretion (ADME) properties, general safety, and disease-context liabiliy for natural-product scaffold triage. Receptor-specific docking calibration and bounded desirability transformations place heterogeneous evidence on a common decision scale before cross-domain integration.

Applied to 59 *Annona muricata* compounds across nine oncology targets, SWAN-MPO substantially changed candidate prioritization by favoring balanced cross-domain profiles. External ChEMBL benchmarking revealed why this matters: binding almost perfectly distinguished active from inactive compounds (ROC-AUC 0.976) but was essentially uninformative for separating favorable from serious-warning profiles among already-active compounds (ROC-AUC 0.503), where ADME was more discriminative (ROC-AUC 0.620). On the combined activity/status endpoint, the integrated formulation achieved ROC-AUC 0.909. In a separate descriptor-compatible benchmark, SWAN-MPO significantly outperformed four adapted multiparameter and drug-likeness comparators in global discrimination. These results

show that natural-product prioritization depends on integrating property domains that answer different screening questions rather than optimizing any single signal.

Keywords: multi-parameter optimization; natural products; molecular docking; cheminformatics; ADME; drug discovery

Introduction

Multi-parameter optimization (MPO) is widely used to balance competing objectives in drug discovery, but many established frameworks were developed around physicochemical and developability criteria characteristic of conventional small-molecule drug space. Natural products occupy distinct structural and physicochemical space and can include larger, more lipophilic, and more conformationally complex scaffolds than conventional screening compounds [1,2]. These characteristics become particularly important when docking is incorporated into prioritization because docking scoring functions can favor larger or more interaction-rich ligands even when the same scaffold may have less favorable ADME, developability, or safety characteristics [2,3,5]. The resulting challenge is not simply to consider more properties, but to place conflicting binding and nonbinding evidence on a common decision scale so that no domain dominates merely because of its numerical representation.

Structure-based virtual screening makes the binding side of this problem particularly important. AutoDock Vina and related tools support rapid pose generation and affinity-oriented ranking across receptor structures [4]. Empirical docking scores approximate binding but remain sensitive to receptor conformation, desolvation, entropic effects, ligand size, and pocket characteristics [4,5]. In multi-target screening, these scores also arise from distinct receptor environments and therefore should not be treated as directly equivalent across heterogeneous

pockets. A framework that integrates binding across multiple receptor contexts consequently requires an explicit strategy for receptor-specific calibration before binding is combined with nonbinding evidence. Scaffolds that rank favorably under binding-only logic but move downward after safety, ADME, and liability integration are referred to as affinity traps.

Alongside binding, scaffold triage requires complementary evidence about pharmacokinetic and safety properties. SwissADME estimates physicochemical, pharmacokinetic, and drug-likeness properties; ProTox 3.0 predicts toxicity across multiple endpoints [6,8]. These tools provide complementary nonbinding evidence, but general-purpose predictions do not capture every scaffold- or disease-context-specific concern relevant to natural-product evaluation. Separating general safety from context-dependent liability allows those concerns to influence ranking explicitly rather than being collapsed into a single toxicity assessment.

Traditional workflows treat docking, ADME, and toxicity as sequential filters, which can discard viable scaffolds at arbitrary cutoffs or advance strong binders with cumulative nonbinding liabilities. Multi-parameter optimization addresses this problem by combining competing drug-discovery objectives into a single decision framework [9]. Desirability functions convert variables with different units and scales into bounded, comparable terms before aggregation [10,11]. Established approaches also illustrate the value of designing optimization frameworks around defined decision problems: QED summarizes drug-likeness from molecular-property desirabilities, CNS-MPO combines properties relevant to CNS drug development, and QEPPI emphasizes chemical features associated with protein–protein interaction-targeting compounds [11,35,37]. These approaches remain useful for their intended objectives. SWAN-MPO addresses a different integration problem: how to combine receptor-specific binding with ADME, general safety, and context-dependent liability when these evidence domains can

diverge strongly during natural-product scaffold evaluation. Its objective is balanced scaffold triage across these domains rather than optimization of any single screening signal.

SWAN-MPO was therefore designed to make these conflicting evidence streams comparable without treating them as interchangeable. Docking is summarized across 20 grid centers and calibrated to target-specific references before panel aggregation, while binding, ADME, safety, and context-dependent liability are independently transformed to bounded 0.01–1.00 desirabilities. Equal-weight geometric integration limits compensation across domains, making severe nonbinding weaknesses consequential even when predicted binding is highly favorable. General safety remains distinct from context-dependent liability, while node diagnostics expose the target responsible for the binding contribution. The methodological contribution is the specific integration architecture: receptor-specific calibration of median multi-grid binding, panel-level node aggregation, explicit separation of general safety from context-dependent liability, and compensation-limiting integration within a common bounded desirability framework.

An *Annona muricata* natural-product library was selected as the driving deployment case because compounds reported from this species combine anticancer interest with mechanism-linked safety concerns, particularly mitochondrial Complex I inhibition and neurotoxicity [12–14,30,31]. This chemical space therefore provided a demanding deployment case for the central design problem: whether favorable docking signals could be integrated with ADME, safety, and scaffold-relevant liability information without simply reproducing affinity-first rankings. The computational library was restricted to compounds with traceable species attribution and usable molecular structures.

SWAN-MPO was first evaluated in small internal panels to determine whether multi-domain integration changed affinity-first rankings, then deployed across a 59-compound *A. muricata* library and challenged by prespecified sensitivity analyses. External evaluation next used a ChEMBL activity/status benchmark to test the integrated scoring formulation across active favorable-status compounds and inactive or serious-warning compounds [38]. A secondary analysis examined whether binding and nonbinding domains contributed differently to distinguishing activity from inactivity and, among active compounds, favorable from serious-warning profiles. A separate descriptor-compatible head-to-head benchmark compared the SWAN-MPO formulation with adapted CNS-MPO, LE–QED, QED, and QEPPI approaches.

Methods

Study design and analytical scope

SWAN-MPO was evaluated as a computational framework for multi-domain scaffold prioritization. Extended methods and results are provided in Online Resource 1, with the corresponding machine-readable supplementary tables provided in Online Resource 2.

Computational workflow and reproducibility

SWAN-MPO was implemented as a deterministic file-based workflow spanning input processing, docking, descriptor transformation, score integration, sensitivity analysis, benchmarking, and manuscript reproduction.

R was used for internal component construction, selected statistical analyses, manifest generation, and figure generation, while Python was used for structure processing, docking execution, SWAN-MPO scoring, robustness analyses, and external benchmarking. The audited

SWAN-MPO command-line environment used Python 3.13.5 with NumPy 2.3.5 and NetworkX 3.6.1; structure processing used RDKit 2026.03.2 and Open Babel 3.1.0, and docking used AutoDock Vina 1.2.7. Complete environment specifications and pinned dependencies are provided with the archived software release.

Before external analyses were interpreted, the locked implementation was required to reproduce all 20 internal panel-compound scores and internal ranks. The locked scoring implementation has SHA-256: 97799e91f23803de3b9a477aea3bbcfb061f1b1d9abebdecfdadcd1b537f4454; independent reproduction matched all 20 internal panel-compound scores to within 1×10^{-12}. The published target-calibration resource is separately preserved at SHA-256: e7584f10c0f970bb54ddf39cfdacbb6cb061bc6a3d2bd795e2d5f119efc7287c. The (swan-mpo reproduce-manuscript workflow) recomputes all 531 target-level and 177 panel-level frozen scoring outputs from packaged canonical upstream predictor inputs and verifies them against expected outputs at 1×10^{-10} tolerance. Archived SwissADME, ProTox, and candidate-library docking outputs serve as canonical upstream inputs to this deterministic reproduction workflow.

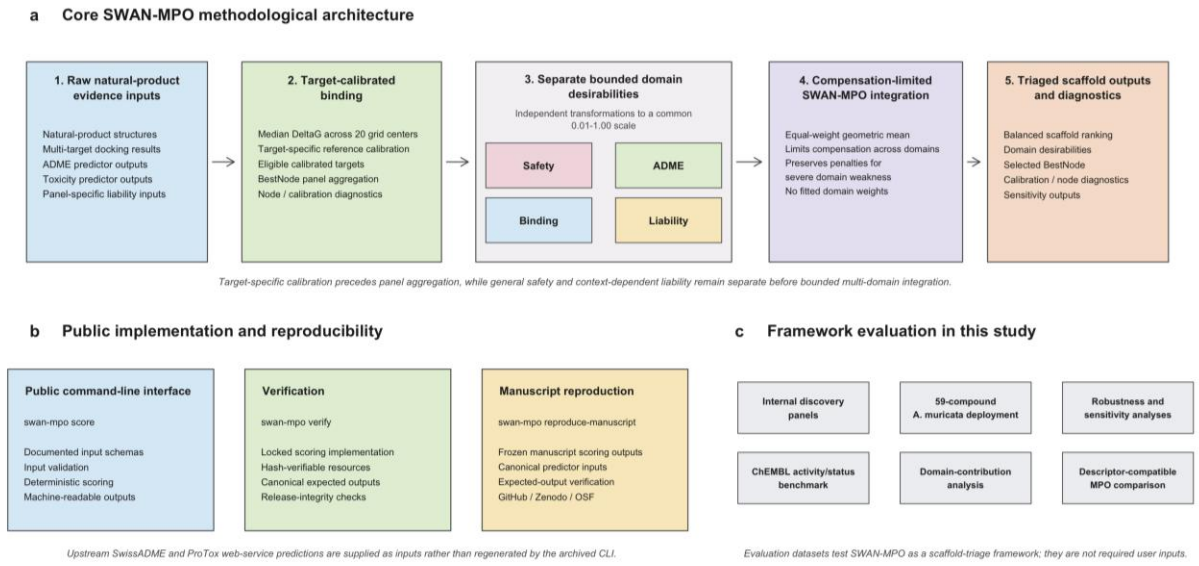


Fig. 1 SWAN-MPO methodological architecture, public implementation, and evaluation framework.
(a) Core target-calibrated multi-domain scoring architecture. (b) Public command-line implementation and reproducibility workflow. (c) Evaluation components used in this study

Target selection and PDB mapping

Target proteins were selected to represent proliferative signaling, apoptotic regulation, hormone biology, and mitochondrial vulnerability across the three cancer-oriented panels. The colorectal cancer panel comprised the mechanistic target of rapamycin (mTOR; PDB: 4JSP), epidermal growth factor receptor (EGFR; EGF-bound extracellular domain, PDB: 1IVO), and mitochondrial Complex I (PDB: 5LNK). The prostate cancer panel comprised mTOR (PDB: 4JSP), 5α-reductase type 2 (5AR2; PDB: 3V3N), the androgen receptor (AR; PDB: 2AMA), B-cell lymphoma 2 (Bcl-2; PDB: 4LVT), and EGFR (PDB: 1IVO). The RCC panel comprised mTOR (PDB: 4JSP), phosphoinositide 3-kinase alpha (PI3Kα; PDB: 6OAC), protein kinase B (AKT1; PDB: 3MVH), Bcl-2 (PDB: 4LVT), and Caspase-3 (PDB: 1QX3). Panel membership, PDB structures, binding-site definitions, native-ligand availability, and calibration classifications are provided in Supplementary Table S5.

The target panels were designed as mechanistically motivated scaffold-stress systems representing proliferative signaling, hormone response, apoptotic regulation, and mitochondrial vulnerability across the three cancer-oriented contexts [32,33,40,41]. Mitochondrial Complex I was included because its inhibition is associated with annonaceous acetogenins, with its docking outputs retained for within-target comparative interpretation [13,14,30,31].

Molecular docking and structure preparation

Molecular docking was performed using AutoDock Vina v1.2.7 with the Vina scoring function [4,42]. All primary docking calculations used rigid receptors. Configuration files and logs

recorded receptor and ligand PDBQT files, grid definitions, exhaustiveness, CPU settings, and output poses for each run.

Receptors and ligands were prepared through documented PDBQT-generation workflows. Protein receptors were prepared as rigid Vina-compatible PDBQT files using receptor-specific atom typing, hydrogen handling, and macromolecular charge assignment. For the expanded 59-compound library, one initial three-dimensional conformer per ligand was generated from curated SMILES using RDKit with explicit hydrogen addition and ETKDGv3 embedding with random seed 2026. Random-coordinate embedding was used as a fallback when required, followed by UFF geometry optimization for up to 1,000 iterations and Open Babel conversion to PDBQT with Gasteiger partial charges. Native crystallographic ligands used for redocking retained their crystallographic coordinates and were converted through PDB and MOL2 formats before PDBQT generation. Native-ligand redocking established receptor-specific reference-calibration status for the binding layer. Full preparation and redocking details are provided in Supplementary Methods S3.

The internal docking matrix comprised 15, 40, and 35 ligand–target pairs in the colorectal, prostate, and RCC panels, respectively, with each pair evaluated across 20 grid centers, yielding 1,800 completed Vina jobs. All Vina jobs used a 40.0 Å cubic search box, grid spacing of 0.375 Å, exhaustiveness of 12, nine output modes, and a fixed random seed of 2026. Each individual Vina process used one CPU thread, with the deployment runner executing up to two jobs concurrently. The same docking protocol was applied to the expanded 59-compound *A. muricata* library across nine unique targets, yielding 531 complete ligand-target blocks and 10,620 Vina jobs. Runtime metadata were preserved for 8,961 of the 10,620 expanded-screen docking jobs, totaling 584.14 serial CPU-hours, with a median runtime of 204.92 s/job and mean runtime of

234.67 s/job. Runtime metadata were unavailable for the remaining 1,659 recovered jobs, so no full-deployment runtime was extrapolated. Targets shared across disease panels were docked once per ligand and the resulting target-level summaries were reused in each applicable panel to avoid redundant receptor-ligand calculations. The 20 target-specific grid centers were prespecified before expanded-library deployment and inherited from the locked internal docking protocol; the same coordinate set was applied consistently to every ligand evaluated against a given target. Complete coordinates are reported in Supplementary Table S5.

Grid-center aggregation and target-calibrated binding

For ligand l, target t, and grid center g, the raw Vina docking energy was denoted $\Delta G_{l,t,g}$. To reduce dependence on any single grid-center configuration, each complete ligand-target pair was summarized across 20 grid centers. The primary binding layer used the median docking energy:

$$\widetilde{\Delta G}_{l,t} = median_{g=1,\ldots,G_{l,t}}(\Delta G_{l,t,g})$$

Where $G_{l,t} = 20$ for complete target ligand pairs. The best value, mean, standard deviation, quartiles, and interquartile range of the grid-center energies were also retained for auditability described in Supplementary Methods S4 and reported in Supplementary Table S7. Only the median grid-center energy entered the primary target-calibrated binding calculation; the best-grid and other distributional summaries were retained as diagnostics and for sensitivity analysis.

As raw Vina energies are receptor-context dependent, median docking energies were transformed using target-specific calibration before cross-target aggregation. The protocol converted each ligand's median docking energy into bounded desirability relative to a prespecified target-specific reference energy denoted by $\Delta G_{ref,t}$:

$$d_{\text{Binding},l,t} = \text{clip}_{0.01,1.00}\left[\min\left(1, \frac{-\widetilde{\Delta G}_{l,t}}{-\Delta G_{\text{ref},t}}\right)\right]$$

Median docking energies equal to or more favorable than the reference energy were capped at a desirability of 1.00, and all resulting values were bounded between 0.01 and 1.00.

The $\Delta G_{ref,t}$ hierarchy used the top-ranked redocked pose when its heavy-atom RMSD from the crystallographic reference pose was ≤2.0 Å; when the top-ranked pose failed this criterion but another generated mode achieved RMSD ≤2.0 Å, the energy of that recovered native-like mode was used instead. Calibration outcomes were classified as strict top-pose validation, generated-mode recovery, or comparative-only. Four targets—Bcl-2, AR, PI3Kα, and AKT1—met strict top-pose validation; mTOR and 5AR2 met generated-mode recovery; and EGFR, Caspase-3, and mitochondrial Complex I were classified as comparative-only. Targets without an accepted reference calibration—EGFR, Caspase-3, and mitochondrial Complex I—were retained because they represented prespecified panel biology and provided within-target comparative information, but they were excluded from primary BestNode aggregation to avoid combining uncalibrated energies across receptor pockets. For mTOR and 5AR2, a native-like generated mode satisfying the ≤2.0 Å RMSD criterion was recovered despite not being selected as the top-ranked Vina mode; the energy of that recovered mode was therefore used as the target-specific reference. Complete redocking RMSD values, reference-energy sources, and calibration classifications are reported in Supplementary Table S6.

BestNode aggregation and node diagnostics

For each ligand *l* in disease panel *p*, the primary panel-level binding term was BestNode:

$$d_{\text{Binding,BestNode},l,p} = \max_{t \in T_p} d_{\text{Binding},l,t}$$

BestNode was the highest binding desirability among targets supported by either strict top-pose redocking or validated generated-mode recovery. Under the published calibration, mTOR was therefore the only reference-calibrated target eligible for primary BestNode aggregation in the colorectal panel, while the prostate and RCC panels contained multiple eligible calibrated targets. Accordingly, the colorectal BestNode binding term represents calibrated support from a single eligible node. The selected target, second-best target, best-to-second desirability ratio, and isolated-node flag were retained as diagnostics of dependence on a dominant binding node alongside the primary SWAN-MPO score. Two sensitivity analyses, Top2Node–which combined the two strongest eligible targets–and Breadth–which summarized binding across all eligible targets–are detailed in Supplementary Methods S6.

ADME descriptor generation

SwissADME-derived pharmacokinetic and drug-likeness descriptors were treated as ligand-level properties shared across disease panels [6]. The locked ADME domain comprised nine inputs: molecular weight, Consensus Log P, topological polar surface area (TPSA), rotatable-bond count, hydrogen-bond acceptors, hydrogen-bond donors, PAINS alerts, predicted gastrointestinal absorption, and synthetic accessibility.

Molecular weight, consensus logP, and hydrogen-bond donor and acceptor counts were included because molecular size, lipophilicity, and hydrogen-bonding capacity are established determinants of solubility and passive permeability in oral drug-like chemical space [43]. TPSA and rotatable-bond count captured molecular polarity and flexibility, both of which influence intestinal permeation and oral bioavailability [44]. Predicted gastrointestinal absorption, PAINS alerts, and synthetic accessibility provided complementary estimates of absorption potential,

assay-interference risk, and practical synthetic tractability, respectively [6,7,45–46]. PAINS status was treated as a graded screening alert over a hard compound-exclusion rule.

Each input was transformed into a bounded desirability value and aggregated by geometric mean to form the ligand-level ADME desirability:

$$d_{\mathrm{ADME},l} = \mathrm{GM}\left(d_{\mathrm{MW},l}, d_{\mathrm{logP},l}, d_{\mathrm{TPSA},l}, d_{\mathrm{RotB},l}, d_{\mathrm{HBA},l}, d_{\mathrm{HBD},l}, d_{\mathrm{PAINS},l}, d_{\mathrm{GI},l}, d_{\mathrm{SA},l}\right)$$

The resulting ADME domain score entered the final SWAN-MPO calculation. Internal component scores are reported in Supplementary Tables S8, while transformation directions, ranges, bounds, and rationales are reported in Supplementary Table S23.

ProTox-based safety descriptor generation

ProTox 3.0 predictions were treated as ligand-level inputs shared across disease panels [8]. The primary ProTox-based safety domain incorporated predicted median lethal dose (LD50), toxicity class, hepatotoxicity, neurotoxicity, nephrotoxicity, respiratory toxicity, and cardiotoxicity. Predicted LD50 and toxicity class captured acute systemic hazard, while hepatotoxicity and nephrotoxicity represented target-organ injury and neurotoxicity, cardiotoxicity, and respiratory toxicity covered vital-system liabilities emphasized in nonclinical safety assessment [8].

For retained confidence-coded organ-toxicity predictions ($c \in [0.5,1.0]$), predicted inactive classes received desirability (c), whereas predicted active classes received (1-c). Predicted LD50 values were transformed logarithmically between 50 and 5,000 mg/kg, and toxicity classes 1–6 were mapped linearly, with higher LD50 values and higher toxicity classes receiving greater desirability. All transformed safety values were bounded between 0.01 and 1.00:

$$d_{\mathrm{LD50},l} = \mathrm{clip}_{0.01,1}\left[\frac{\log_{10}(\mathrm{LD50}_l) - \log_{10}(50)}{\log_{10}(5000) - \log_{10}(50)}\right]$$

$$d_{\mathrm{ToxClass},l} = \mathrm{clip}_{0.01,1}\left[\frac{\mathrm{ToxClass}_l - 1}{5}\right]$$

Acute and organ level safety were first aggregated separately:

$$d_{\mathrm{Acute},l} = \mathrm{GM}\left(d_{\mathrm{LD50},l}, d_{\mathrm{ToxClass},l}\right)$$

$$d_{\mathrm{Organ},l} = \mathrm{GM}\left(d_{\mathrm{Hepato},l}, d_{\mathrm{Neuro},l}, d_{\mathrm{Nephro},l}, d_{\mathrm{Respi},l}, d_{\mathrm{Cardio},l}\right)$$

The final ligand-level safety desirability was:

$$d_{\mathrm{Safety},l} = \mathrm{GM}\left(d_{\mathrm{Acute},l}, d_{\mathrm{Organ},l}\right)$$

Geometric-mean integration followed established desirability-based multi-parameter optimization practice [10,11]. The two-stage structure assigned equal group-level influence to acute and organ-level safety, preventing the five organ endpoints from outweighing the two acute-toxicity measures because they were more numerous. Internal safety component scores and endpoint-level desirabilities are reported in Supplementary Table S8; exact transformations and missing-value rules are provided in the supplementary parameter table.

Panel-specific liability descriptor generation

The liability domain captured context-specific risks distinct from the general safety domain. For these non-CNS cancer panels, predicted blood-brain barrier permeability was evaluated together with neurotoxicity to capture the specific concern created when predicted central exposure coincided with neurological liability.

Complex I activity was penalized only in panels in which it was not an intended docking target. It was therefore not penalized in the colorectal panel, avoiding treatment of the same mechanistic signal as both favorable target engagement and an adverse liability. In the prostate and RCC panels, predicted Complex I activity was treated as an off-panel mitochondrial liability [13,14,30,31]. Accordingly, the colorectal liability term combined BBB permeability and neurotoxicity, whereas the prostate and RCC liability terms combined BBB permeability, neurotoxicity, and Complex I activity.

Panel-specific liability desirabilities were defined as:

$$d_{\mathrm{Liability},l,p} = \begin{cases} \mathrm{GM}\left(d_{\mathrm{BBB},l}, d_{\mathrm{Neuro},l}\right), & p = \mathrm{Colorectal}, \\ \mathrm{GM}\left(d_{\mathrm{BBB},l}, d_{\mathrm{ComplexI},l}, d_{\mathrm{Neuro},l}\right), & p \in \{\mathrm{Prostate}, \mathrm{RCC}\} \end{cases}$$

BBB permeability and Complex I predictions used the same confidence-aware transformation as the categorical organ-toxicity endpoints, with active predictions receiving lower desirability and inactive predictions receiving higher desirability according to the reported confidence.

Desirability transformations and geometric aggregation

All desirability values were bounded using:

$$\mathrm{clip}_{(0.01,1.00)}(x) = \min(\max(x, 0.01), 1.00)$$

The lower bound of 0.01 prevented zero-valued collapse and undefined logarithms during geometric aggregation while preserving strong penalties for unfavorable properties.

For descriptors where higher values were more favorable, the monotonic increasing desirability function was:

$$d_{\mathrm{inc}}(x;L,U)=\begin{cases}0.01, & x\le L,\\ 0.01+0.99\left(\dfrac{x-L}{U-L}\right), & L<x<U,\\ 1.00, & x\ge U.\end{cases}$$

Here, L and U are the predefined lower and upper reference bounds.

For descriptors where lower values were more favorable, the monotonic decreasing desirability function was:

$$d_{\mathrm{dec}}(x;L,U)=\begin{cases}1.00, & x\le L,\\ 1.00-0.99\left(\dfrac{x-L}{U-L}\right), & L<x<U,\\ 0.01, & x\ge U.\end{cases}$$

For descriptors with an optimal interval, the windowed desirability function was:

$$\boldsymbol{d_{\mathrm{win}}(x;L,A,B,U)=\begin{cases}0.01, & x\le L,\\ 0.01+0.99\left(\dfrac{x-L}{A-L}\right), & L<x<A,\\ 1.00, & A\le x\le B,\\ 1.00-0.99\left(\dfrac{x-B}{U-B}\right), & B<x<U,\\ 0.01, & x\ge U.\end{cases}}$$

The interval [A,B] defined the preferred range, while desirability declined linearly toward 0.01 across the tolerated range [L,U]. Binary alerts received higher desirability when absent and lower desirability when present, according to the prespecified mapping. The exact direction, preferred interval, tolerated range, and categorical mapping used for each descriptor are reported in Supplementary Table S23.

The equal-weight geometric mean used for SWAN-MPO aggregation was:

$$\mathrm{GM}(d_1, \ldots, d_k) = \exp\left[\left(\frac{1}{k}\right)\sum_{i=1}^{k} \ln(d_i)\right]$$

As all transformed desirabilities were bounded above zero, the geometric mean remained defined and limited compensation for low-performing terms [10,11].

Primary SWAN-MPO score

The primary integrated score was SWAN-BestNode:

$$\mathrm{SWAN}_{\mathrm{BestNode},l,p} = \left(d_{\mathrm{Safety},l,p} \times d_{\mathrm{ADME},l,p} \times d_{\mathrm{Binding,BestNode},l,p} \times d_{\mathrm{Liability},l,p}\right)^{\frac{1}{4}}$$

Safety, ADME, BestNode binding, and panel-specific liability received equal weights. Equal domain weighting was prespecified and held fixed across deployment, with alternative weight structures evaluated through Monte Carlo perturbation.

Internal ranking and affinity-trap analysis

Internal panel rankings were compared across binding-only, QED-only, QED/ADME-style, and SWAN-BestNode scoring. An affinity trap was defined as a compound that ranked highly by BestNode binding but declined after safety, ADME, and liability integration. Comparisons used binding-only rank, component desirabilities, final SWAN rank, and descriptive Kendall's τ_b, given the small panel sizes.

Expanded Annona muricata screening

After model lock, the same workflow was applied to 59 unique *A. muricata* natural products retained after species verification, structure reconciliation, and exact SMILES deduplication. The

compound manifest and docking-deployment matrix are reported in Supplementary Tables S2 and S3, respectively.

Sensitivity Analyses

Robustness analyses evaluated alternative binding aggregation (Top2Node and Breadth), ligand-efficiency adjustment, strict-pass-only target restriction, Monte Carlo domain-weight perturbation, and best-grid versus median-grid aggregation while holding the primary scoring architecture fixed. Methods are provided in Supplementary Methods S4 and S6–S9; outputs are reported in Supplementary Tables S7, S9, and S17–S19.

ChEMBL activity/status external benchmark

For external evaluation, a ChEMBL activity/status benchmark was constructed independently from the production candidate-library workflow. Bioactivity records for 5AR2, AKT1, AR, Caspase-3, EGFR, PI3Kα, and mTOR were harmonized across IC50, Ki, and Kd measurements reported in nanomolar units [38]. Parent–target records were classified as active when qualifying potency was ≤1,000 nM and as weak/inactive when no active evidence was present and qualifying potency was ≥10,000 nM; intermediate records and parent–target groups with discordant activity assignments were excluded. Molecule records were resolved to the ChEMBL parent level before clinical status was assigned. Favorable status required a maximum clinical phase of 4 without a parent-level Black Box Warning or Withdrawn annotation, whereas either warning defined serious-warning status.

Three benchmark groups were defined: active favorable-status compounds (A), active serious-warning compounds (B), and weak/inactive favorable-status compounds (C). The primary endpoint distinguished group A from the combined B+C class, directly testing whether the

integrated scoring formulation separated compounds combining activity and favorable status from compounds failing either the activity or favorable-status criterion. The final primary benchmark comprised 190 parent–target records, including 46 group-A, 25 group-B, and 119 group-C records.

Measured median potency was converted to pActivity as $9 - \log_{10}(\text{nM})$ and transformed to a bounded 0.01–1.00 binding desirability over the 4.0–9.0 activity range. The benchmark-compatible integrated score combined binding desirability, an RDKit-derived ADME desirability, and a structure-derived liability desirability by equal-weight geometric mean. The ADME desirability summarized transformed molecular weight, logP, TPSA, hydrogen-bond donor and acceptor counts, and rotatable-bond count, whereas the structure-derived liability desirability summarized extreme lipophilicity, extreme molecular size, extreme flexibility, fraction sp3, aromatic-ring burden, and structural-alert desirability.

Performance was evaluated using ROC-AUC, PR-AUC, BEDROC $\alpha = 20$ [34], and tie-aware enrichment factors at 1% and 5%. Pairwise ROC-AUC differences were evaluated using 2,000 bootstrap replicates clustered by ChEMBL parent molecule with random seed 2026.

Domain-contribution analysis

A secondary post hoc analysis examined whether the relative contributions of binding and nonbinding information differed across screening decisions. Activity discrimination was evaluated by comparing active favorable-status compounds with weak/inactive favorable-status compounds (A vs C), while status discrimination among active compounds compared active favorable-status with active serious-warning compounds (A vs B). Binding-only, ADME-only,

structure-derived liability, reduced-domain combinations, the integrated benchmark formulation, and QED were evaluated where applicable.

Since group C was substantially larger than group B, an additional sensitivity analysis assigned equal total negative-class weight to groups B and C before recomputing A-versus-B+C discrimination. Group-A records together received unit total weight, while groups B and C each contributed one-half of the total negative weight. Pairwise ROC-AUC differences in the domain-contribution analyses were evaluated using 5,000 parent-molecule-clustered bootstrap replicates with random seed 2026.

Descriptor-compatible MPO comparator benchmark

A shared-descriptor benchmark compared the descriptor-compatible SWAN-MPO formulation with adapted MPO and drug-likeness approaches under a common input space. Comparators comprised a CNS-MPO-inspired oncology formulation, an LE–QED composite formulation, RDKit QED, and a QEPPI-inspired property-window formulation [11,35–37]. The adapted labels distinguish these implementations from exact reproductions of the original frameworks in their published application settings.

For comparator calculations requiring a pChEMBL-derived potency term, the 327 active records without a comparable pChEMBL value were deterministically assigned a value of 6.0. Performance was evaluated using ROC-AUC, area under the precision–recall curve (PR-AUC), Boltzmann-enhanced discrimination of receiver operating characteristic (BEDROC; $\alpha = 20$) [34], and tie-aware enrichment factors (EF) at 1% and 5%. Comparator metrics, pairwise comparisons, and compound-level scores are reported in Supplementary Table S14.

Exact score ties were handled independently of input row order. When an EF@1% or EF@5% cutoff intersected an exact-score tie block, the expected number of favorable records was calculated from the favorable fraction within that tied block.

Code availability

The SWAN-MPO source code, documented command-line implementation, and reproducibility workflows are publicly available at https://github.com/saanviisenthilkumar/swan-mpo. SWAN-MPO v1.3.2 is permanently archived at Zenodo: https://doi.org/10.5281/zenodo.21988306.

Results

Internal discovery-panel rankings

Within the internal colorectal, prostate, and RCC panels, SWAN-MPO produced a ranking structure distinct from binding-only, QED-only, and QED/ADME-style scoring.

Muricatacin served as one example of this rank shift, ranking third by BestNode binding in colorectal, eighth in prostate, and seventh in RCC. After SWAN-BestNode integration however, it ranked first in colorectal cancer (0.733), first in prostate cancer (0.659), and second in RCC (0.683). As mTOR was the only reference-calibrated target eligible for colorectal BestNode aggregation, the colorectal result represents single-node calibrated binding integrated with the non-binding domains. Under QED-only and QED/ADME-style scoring, Muricatacin ranked #4/#4 in colorectal, #3/#3 in prostate, and #4/#4 in RCC.

The internal panels served as illustrative mechanism checks, showing how multi-domain integration reordered individual compounds before evaluation across the expanded library.

Illustrative affinity-trap rank shifts

In the colorectal panel, Annomuricin C ranked first by BestNode binding (1.000), but fell to third after full SWAN-BestNode integration (0.652). By comparison, Muricatacin had lower binding desirability (0.820) but a higher integrated score (0.733).

A similar rank reversal appeared in prostate cancer where Annopentocin A ranked first by BestNode binding but eighth of eight after SWAN-BestNode integration because its favorable binding was accompanied by lower non-binding desirabilities. These internal-panel rank shifts are illustrated in Fig. 2.

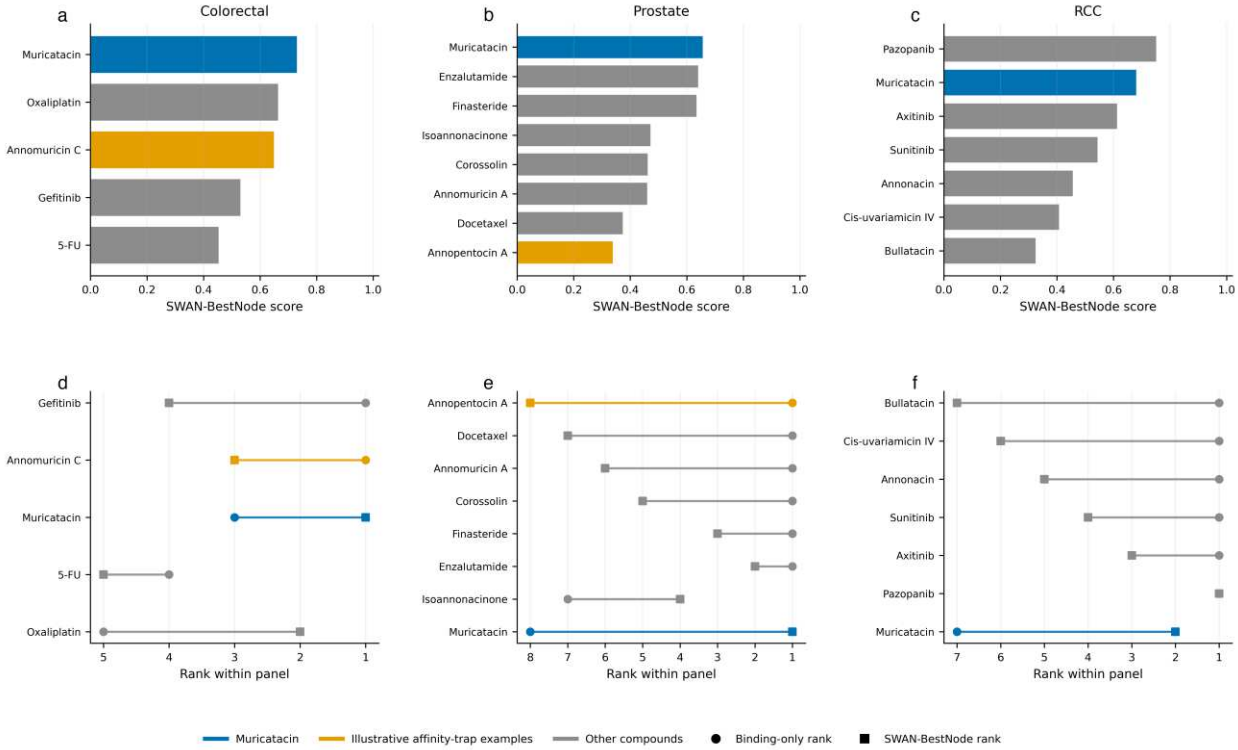


Fig. 2 Illustrative affinity-trap rank shifts across disease panels
(a–c) Final SWAN-BestNode scores for colorectal, prostate, and renal cell carcinoma (RCC) discovery panels. Muricatacin is highlighted as a high-ranked scaffold, with Annomuricin C and Annopentocin A shown as affinity-trap examples. (d–f) Binding-only and final SWAN-BestNode ranks are compared, illustrating downward reprioritization after safety, ADME, and liability information were incorporated

The expanded-library analysis next tested whether this reordering persisted across a larger candidate set.

Expanded Annona muricata screening results

Application of the SWAN-MPO workflow to 59 unique *A. muricata* natural products produced 177 panel-specific rankings across the colorectal, prostate, and RCC panels. The internal discovery panels and expanded deployment were scored from separate analysis-specific docking matrices; consequently, compounds present in both analyses can have slightly different calibrated binding and integrated SWAN-MPO values.

The expanded candidate pool illustrated similar rank reversals to the internal screening. Muricatacin, for example, ranked 47th by BestNode binding in all three panels but ranked first by SWAN-BestNode in colorectal cancer (0.734) and second in both prostate cancer and RCC (0.661 in each panel). This illustrates that high integrated placement did not require high binding-only rank. Library-wide rank reordering in the expanded deployment is summarized in Fig. 3.

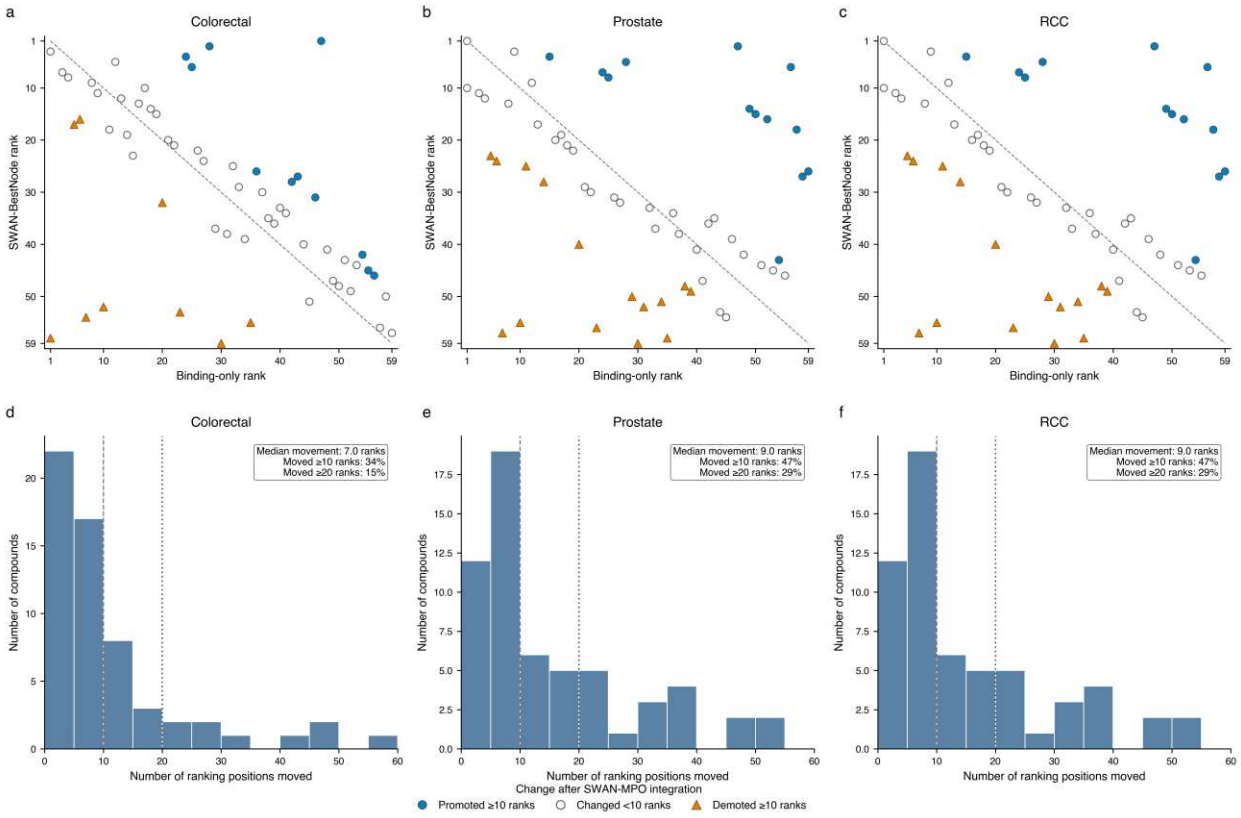


Fig. 3 Expanded 59-compound screening reveals framework-wide rank reordering. (a–c) Binding-only rank is compared with SWAN-BestNode rank across colorectal, prostate, and RCC panels; compounds shifting by at least 10 positions are distinguished by direction. (d–f) Distributions of absolute rank movement summarize overall reordering, with annotations reporting median movement and the proportions shifting by at least 10 and 20 positions

We next tested whether these results persisted under alternative binding, calibration, and weighting assumptions.

Robustness and sensitivity analyses

Robustness analyses evaluated ligand-efficiency adjustment, strict-pass-only target restriction, alternative binding aggregation, domain-weight perturbation, and best-grid versus median-grid aggregation.

Ligand-efficiency adjustment retained substantial full-library agreement with the primary rankings. Kendall's τ_b was 0.813 in colorectal cancer and 0.804 in both prostate cancer and RCC,

with all three associations statistically significant. Substantial full-library ranking structure was therefore retained after adjustment for ligand size.

Restriction to targets that passed strict top-pose redocking also preserved strong full-library agreement in the two estimable panels. Kendall's τ_b was 0.871 in prostate cancer and 0.898 in RCC. Prostate and RCC rankings remained highly concordant when calibration was restricted to strict top-pose targets. The colorectal panel contained no strict-top-pose-eligible target and was therefore not estimable under this sensitivity.

Alternative binding aggregation using Top2Node and Breadth and Monte Carlo domain-weight perturbation further supported broad ranking stability while identifying greater sensitivity for selected compounds and the colorectal configuration. For Muricatacin specifically, replacing best-grid with median-grid aggregation did not change its internal binding rank in any of the three panels.

Full methods and outputs are reported in Supplementary Methods S4 and S6–S9 and Supplementary Results S2 and S4–S7.

These analyses tested robustness within the deployment library. ChEMBL-derived benchmarks next evaluated the scoring architecture externally.

External ChEMBL activity/status benchmark

The primary ChEMBL activity/status benchmark comprised 46 active favorable-status records and 144 records failing either the activity or serious-warning criterion. The integrated benchmark formulation achieved a ROC-AUC of 0.909, with PR-AUC of 0.680 and BEDROC $\alpha = 20$ of 0.677. EF@1% and EF@5% were 2.07 and 2.89, respectively. These results showed strong

global discrimination of records combining activity with favorable clinical status from records with inactivity or serious-warning status.

Domain contributions across screening decisions

Domain-level ROC-AUC values across the three screening contrasts are summarized in Table 1.

Table 1. Domain contributions across ChEMBL screening decisions

Score formulation	A vs C: activity ROC-AUC	A vs B: warning-status ROC-AUC	Class-balanced A vs B+C weighted ROC-AUC
Binding only	0.976	0.503	0.740
ADME only	0.503	0.620	0.562
Structure-derived liability	0.567	0.560	0.564
Binding + ADME	0.969	0.612	0.790
Integrated three-domain	0.973	0.605	0.789
QED	0.553	0.502	0.527

A vs C compared active favorable-status with weak/inactive favorable-status records. A vs B compared active favorable-status with active serious-warning records. The balanced sensitivity assigned equal total negative-class weight to groups B and C.

The relative contribution of each domain depended strongly on the screening decision. For active favorable-status versus weak/inactive favorable-status compounds, binding alone achieved ROC-AUC 0.976, while the integrated formulation achieved 0.973, indicating that activity discrimination was already dominated by the potency-derived binding signal.

Among active compounds, however, binding no longer distinguished favorable from serious-warning status (ROC-AUC 0.503). ADME provided greater discrimination in this comparison (ROC-AUC 0.620), while the integrated formulation achieved 0.605. Thus, once activity was

established, nonbinding developability information contributed more strongly to separation of favorable and serious-warning profiles.

The class-balanced sensitivity produced the same overall pattern. Binding alone achieved ROC-AUC 0.740, whereas binding plus ADME increased discrimination to 0.790 and the full three-domain formulation achieved 0.789; QED achieved 0.527. Together, these analyses showed that binding and nonbinding domains contributed complementary information across different screening decisions.

Descriptor-compatible MPO comparator analysis

The descriptor-compatible benchmark next compared the SWAN formulation with comparator scoring schemes derived from existing MPO, ligand-efficiency, and drug-likeness concepts (Table 2).

In this descriptor-compatible comparison, the SWAN formulation achieved the highest ROC-AUC at 0.606, compared with 0.543 for the CNS-MPO-inspired oncology proxy, 0.563 for the LE–QED composite proxy, 0.518 for QED, and 0.514 for the QEPPI-inspired proxy.

Table 2. Performance of descriptor-compatible MPO and drug-likeness implementations in the broad ChEMBL active–synthetic-profile benchmark.

Framework	ROC-AUC	BEDROC α = 20	EF@1%	EF@5%
Descriptor-compatible SWAN-MPO formulation	0.606	0.170	0.00	0.40
CNS-MPO-inspired oncology proxy	0.543	0.335	1.72	1.42
LE–QED composite proxy	0.563	0.330	0.91	1.20
QED (RDKit)	0.518	0.161	1.09	0.55
QEPPI-inspired proxy	0.514	0.510	2.55	2.33

DeLong analysis [39] showed that SWAN-MPO formulation achieved a significantly higher ROC-AUC than each comparator: CNS-MPO-inspired oncology proxy (ΔAUC = 0.0625, $p = 7.04 \times 10^{-10}$), LE–QED composite proxy (ΔAUC = 0.0425, $p = 4.66 \times 10^{-6}$), QED (ΔAUC = 0.0879, $p = 1.30 \times 10^{-16}$), and QEPPI-inspired proxy (ΔAUC = 0.0918, $p = 9.78 \times 10^{-13}$). Complete statistics are reported in Supplementary Table S14.

Early-enrichment behavior differed from global ROC-AUC performance. The QEPPI-inspired formulation led BEDROC, EF@1%, and EF@5%, while the CNS-MPO-inspired formulation also exceeded the descriptor-compatible SWAN-MPO formulation on these early-ranking metrics. The benchmark therefore distinguished two ranking behaviors: SWAN-MPO provided the strongest global class separation, whereas comparator formulations concentrated active records more strongly at the extreme top of the ranked list.

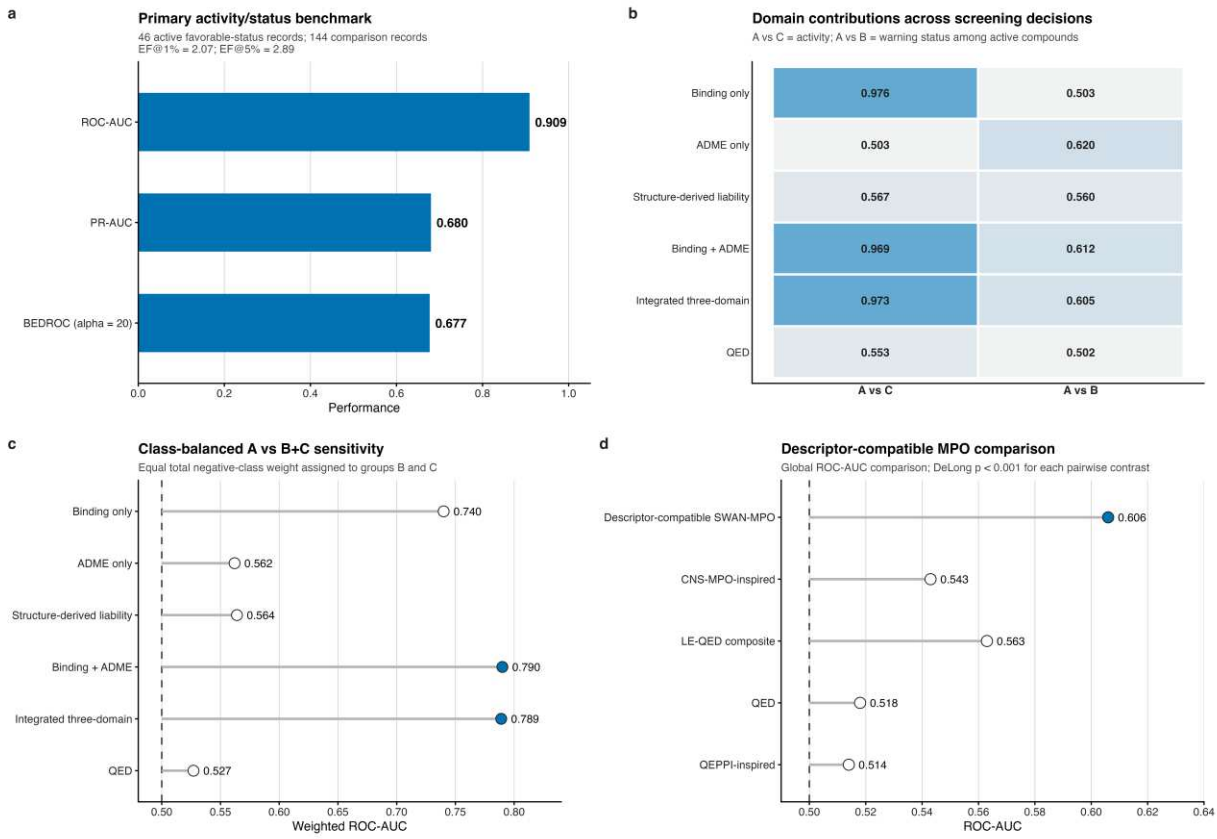


Fig. 4 External ChEMBL benchmarking and domain-contribution analysis.
(a) Primary activity/status benchmark performance. (b) Domain-level discrimination for A vs C activity status and A vs B warning status among active compounds. (c) Class-balanced A vs B+C sensitivity analysis. (d) Descriptor-compatible MPO comparison by global ROC-AUC; SWAN-MPO exceeded each comparator by DeLong analysis ($p < 0.001$)

Discussion

SWAN-MPO reframes screening as multi-domain scaffold triage

Affinity-centered virtual screening can prioritize compounds with favorable docking scores even when their safety, ADME, or scaffold-liability profiles are less favorable. This limitation becomes more pronounced when heterogeneous inputs are combined without placing them on a comparable scale. SWAN-MPO was therefore designed to integrate these signals within a common desirability framework while limiting the influence of any single favorable domain.

Unlike generic drug-likeness scores or descriptor-only MPO approaches, SWAN-MPO combines target-calibrated median multi-grid docking, panel-level node aggregation, general safety, and panel-specific liability within one auditable architecture. Its methodological contribution therefore lies in the calibration, separation, and diagnostic integration of these evidence layers rather than in geometric averaging alone.

Reproducibility was treated as part of the methodology rather than as a separate release step. The scoring implementation and published target-calibration resource are independently hash-verifiable, and the manuscript-reproduction workflow recomputes all 177 panel-level and 531 target-level outputs against frozen expected results at 1×10^{-10} tolerance. This architecture enables third-party audit of the scoring path without requiring independent reimplementation of the framework.

SWAN-MPO complements established MPO approaches by addressing a different decision problem: integration of receptor-specific binding with ADME, general safety, and context-dependent liability during natural-product scaffold triage. Its added value is the prespecified architecture linking target-calibrated multi-grid docking to panel-level node diagnostics and distinct nonbinding evidence domains while retaining an auditable path from upstream inputs to final scaffold ranks.

Across the internal panels and expanded library, SWAN-MPO produced rankings distinct from affinity-only and generic property-based approaches. Muricatacin is used throughout as an illustrative worked example; framework-level conclusions are based on library-wide rank reordering, sensitivity analyses, and external benchmark results rather than on the ranking of this single scaffold.

Cross-domain integration reshapes affinity-first rankings

The internal panels illustrated affinity traps as a rank-based phenomenon in which compounds favored by binding alone moved downward after safety, ADME, and liability integration.

Annomuricin C and Annopentocin A illustrated this behavior in the internal panels. Both ranked favorably by binding but moved downward after the non-binding domains were incorporated. As the four domains enter the same bounded geometric mean, low desirability in one domain reduces the final score even when another domain is favorable. The framework therefore preserves binding information without allowing affinity alone to determine the final ordering. This is especially relevant for acetogenins and other flexible, lipophilic natural products where favorable docking scores may partly reflect extensive hydrophobic contact surfaces.

Target calibration supports receptor-aware binding integration

Docking scores are sensitive to receptor conformation, search-space definition, and scoring-function limitations [4,5]. SWAN-MPO therefore uses docking as a calibrated comparative signal, with median aggregation across grid centers reducing dependence on a single favorable search region and target-specific desirability mapping accounting for receptor-specific score ranges.

Calibration coverage varied across targets, with mTOR representing the only reference-calibrated target eligible for primary BestNode aggregation in the colorectal panel. This target-level classification makes calibration confidence explicit and prevents comparative-only receptor scores from entering the primary cross-target aggregation.

BestNode identity and related node diagnostics made dependence on a dominant target auditable while preserving the primary scoring formulation.

Sensitivity analyses support ranking stability across alternative assumptions

The sensitivity analyses showed substantial rank concordance under ligand-efficiency adjustment, strict-top-pose calibration restriction where estimable, alternative binding aggregation, and domain-weight perturbation. Full-library concordance remained high after ligand-efficiency correction and strict-pass restriction, supporting the stability of the primary ranking architecture under alternative binding assumptions. The analyses also identified greater model sensitivity in the colorectal configuration and for selected compounds, providing compound- and panel-level information about ranking uncertainty. Complete definitions, rank shifts, concordance statistics, and Monte Carlo stability outputs are reported in Supplementary Methods S4 and S6–S9 and Supplementary Tables S17–S19.

External benchmarking reveals complementary domain contributions

The ChEMBL activity/status benchmark provided external evidence that the integrated scoring formulation can distinguish compounds combining activity with favorable status from compounds failing either the activity or serious-warning criterion. The resulting ROC-AUC of 0.909 demonstrated strong global discrimination in this external benchmark.

The domain-contribution analysis clarified why multi-domain integration is useful. Binding strongly separated active from inactive compounds, whereas among compounds with established activity, ADME became more informative for distinguishing favorable from serious-warning profiles. No individual domain therefore dominated across both screening decisions. By integrating these complementary signals, the SWAN-MPO formulation achieved strong discrimination on the primary combined activity/status endpoint (ROC-AUC 0.909), in which compounds could be deprioritized for either insufficient activity or serious-warning status.

In the descriptor-compatible head-to-head benchmark, the SWAN-MPO formulation achieved the strongest global ROC-AUC, whereas selected adapted comparators performed better on early-enrichment metrics [11,35–37]. This difference is consistent with SWAN-MPO's objective of balanced scaffold triage across competing domains rather than optimization of a single ranking signal.

Practical scope and generalizability

The practical value of SWAN-MPO is moving natural-product scaffold triage beyond affinity- or property-dominated prioritization toward transparent multi-domain comparison. Structurally complex natural products can appear favorable by docking or generic physicochemical criteria while carrying less favorable ADME, safety, or context-dependent liability profiles [2,3].

SWAN-MPO provides an auditable framework for evaluating these competing signals on a common decision scale before candidates are advanced to higher-cost experimental investigation [2,8].

Operationally, application of SWAN-MPO to a new compound library requires candidate docking outputs, ADME and toxicity predictions, and a prespecified panel-liability configuration. Extension to new targets additionally requires target-specific reference calibration before those targets can enter primary BestNode aggregation. The framework therefore separates reusable scoring logic from analysis-specific upstream prediction and target-calibration steps, allowing unavailable or lower-confidence inputs to remain explicit rather than being hidden within the final score.

Together, these findings support SWAN-MPO as a transparent computational framework designed for natural-product scaffold triage in oncology, with an architecture that can be configured for other compound and target-panel settings requiring analogous binding, safety, ADME, and liability integration.

Limitations

The docking layer retains the limitations of rigid-receptor Vina scoring and simplified ligand preparation, including incomplete treatment of receptor flexibility, induced fit, solvent effects, desolvation, and binding entropy. One initial ETKDGv3 conformer was generated per compound, with subsequent torsional exploration delegated to Vina rather than an independent multi-conformer ensemble. Median aggregation across 20 inherited target-specific grid centers reduces dependence on a single search-center placement but does not address these structural

limitations; prioritized scaffold–target pairs therefore require higher-fidelity structural refinement before mechanistic interpretation.

The target panels were mechanistically motivated scaffold-stress systems rather than comprehensive models of colorectal, prostate, or renal cell carcinoma biology. Calibration confidence also varied by target, and the colorectal primary BestNode depended on a single eligible calibrated node. These constraints limit disease-specific and multi-target mechanistic interpretation even where ranking behavior is reproducible.

SwissADME and ProTox 3.0 were developed from broader chemical datasets that may incompletely represent annonaceous acetogenins and other structurally specialized natural products. The endpoint confidence values returned by ProTox were incorporated into scoring, but they do not constitute a formal applicability-domain assessment for this scaffold class. SWAN-MPO therefore inherits uncertainty from its input predictors, and favorable automated ADME or toxicity desirabilities should be treated as screening estimates requiring experimental confirmation.

The ChEMBL activity/status benchmark evaluated the benchmark-compatible integrated scoring formulation using measured activity together with ADME and structure-derived liability descriptors. It therefore provides external evidence for the ranking formulation under the available ChEMBL activity/status data, while the production candidate-library workflow additionally incorporates target-calibrated docking and the full SWAN-MPO predictor inputs.

The expanded 59-compound *A. muricata* deployment was species restricted. Although the architecture is configurable to other compound libraries, target panels, and liability definitions,

broader transferability across natural-product classes and other chemical spaces remains to be evaluated empirically.

Future evaluation should include biochemical target-engagement assays, cellular activity studies, toxicity and pharmacokinetic testing, and flexible-residue docking or molecular-dynamics refinement for experimentally justified scaffold–target pairs.

Conclusion

SWAN-MPO addresses natural-product scaffold triage as a target-calibrated multi-domain integration problem. It moves prioritization beyond affinity- or property-dominated ranking by integrating receptor-calibrated median multi-grid docking with ADME, general safety, and context-dependent liability on a common bounded desirability scale. This architecture is particularly relevant to structurally complex natural products, where favorable docking or physicochemical signals can coexist with less favorable developability, safety, or scaffold-specific liability profiles.

Across the internal panels and expanded 59-compound *A. muricata* deployment, SWAN-MPO substantially reordered affinity-first rankings while retaining broad stability under ligand-efficiency adjustment, alternative binding aggregation, calibration restriction, and domain-weight perturbation. The resulting rankings therefore reflected cross-domain balance.

In the external ChEMBL activity/status benchmark, the integrated scoring formulation achieved ROC-AUC 0.909 in distinguishing active favorable-status compounds from inactive or serious-warning compounds. Domain-level analysis showed that binding dominated activity discrimination whereas ADME was more informative among active compounds differing in warning status, while their integration achieved strong overall discrimination when both failure

modes were considered together. The descriptor-compatible head-to-head benchmark further showed significantly higher global ROC-AUC for the SWAN-MPO formulation than adapted CNS-MPO, LE–QED, QED, and QEPPI formulations. Together with its hash-verifiable scoring implementation, target-calibration resource, and deterministic reproduction workflow, these results support SWAN-MPO as a reproducible framework for balanced multi-domain natural-product scaffold triage.

Acknowledgments

The author gratefully acknowledges Dr. Kunal Rai of The University of Texas MD Anderson Cancer Center for his guidance and feedback.

Abbreviations

5AR2, 5α-reductase type 2; ADME, absorption, distribution, metabolism, and excretion; AKT1, protein kinase B; AR, androgen receptor; BBB, blood–brain barrier; Bcl-2, B-cell lymphoma 2; BEDROC, Boltzmann-enhanced discrimination of receiver operating characteristic; CLI, command-line interface; CNS-MPO, central nervous system multiparameter optimization; EF, enrichment factor; EGFR, epidermal growth factor receptor; ETKDGv3, Experimental-Torsion-Knowledge Distance Geometry version 3; GI, gastrointestinal; HBA, hydrogen-bond acceptor; HBD, hydrogen-bond donor; IC50, half-maximal inhibitory concentration; Kd, dissociation constant; Ki, inhibition constant; LD50, median lethal dose; LE, ligand efficiency; logP, octanol–water partition coefficient (logarithm); MMFF, Merck Molecular Force Field; MPO, multi-parameter optimization; mTOR, mechanistic target of rapamycin; PAINS, pan-assay interference compounds; pChEMBL, negative logarithm of ChEMBL-standardized activity value; PDB, Protein Data Bank; PDBQT, Protein Data Bank, partial charge, atom type (AutoDock format);

PI3Kα, phosphoinositide 3-kinase alpha; QED, quantitative estimate of drug-likeness; QEPPI, quantitative estimate of protein–protein interaction targeting; RCC, renal cell carcinoma; RMSD, root-mean-square deviation; ROC-AUC, area under the receiver operating characteristic curve; SA, synthetic accessibility; SMILES, simplified molecular-input line-entry system; TPSA, topological polar surface area; UFF, universal force field.

Statements and Declarations

Ethics approval and consent to participate

Not applicable. This study did not involve human participants, human data, human tissue, or animals.

Consent for publication

Not applicable.

Availability of data and materials

The data and materials supporting the findings of this study are provided in the Supplementary Information, accompanying machine-readable supplementary data files, and associated public repositories. Canonical expanded-screen manuscript-reproduction data and supporting provenance materials are archived in the Open Science Framework (OSF): https://doi.org/10.17605/OSF.IO/35QF2. The SWAN-MPO source code, documented command-line implementation, and reproducibility packages for the ChEMBL activity/status benchmark, domain-contribution analysis, and descriptor-compatible MPO comparator analysis are available on GitHub: https://github.com/saanviisenthilkumar/swan-mpo. SWAN-MPO v1.3.2 is permanently archived on Zenodo: https://doi.org/10.5281/zenodo.21988306. The archived reproduction workflow recomputes frozen scoring outputs from packaged canonical upstream

predictor inputs; SwissADME and ProTox web-service predictions and candidate-library docking are not regenerated by the archived software. SWAN-MPO v1.3.2 is implemented in Python 3.13.x and distributed under the MIT License; dependency, installation, and tested-environment details are provided in the public GitHub repository.

Competing interests

The author declares no competing interests.

Funding

The author received no specific funding for this work.

Authors' contributions

SS conceptualized the study; developed the SWAN-MPO methodology and software; curated and processed the data; performed the computational, docking, benchmarking, sensitivity, and statistical analyses; generated the figures and supplementary materials; and wrote and revised the manuscript. SS reviewed and approved the final manuscript and accepts responsibility for the work.

Supplementary Information

Online Resource 1 (PDF): SWAN-MPO Supplementary Information. Extended methods, supplementary results, descriptions of Supplementary Tables S1–S23, and Supplementary Figures S1–S4.

Online Resource 2 (ZIP): SWAN-MPO machine-readable supplementary tables. CSV files corresponding to Supplementary Tables S1–S23 together with the supplementary-table file manifest.

References

[1] Newman DJ, Cragg GM (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod 83:770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

[2] Atanasov AG, Zotchev SB, Dirsch VM et al (2021) Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov 20:200–216. https://doi.org/10.1038/s41573-020-00114-z

[3] Sun D, Gao W, Hu H, Zhou S (2022) Why 90% of clinical drug development fails and how to improve it? Acta Pharm Sin B 12:3049–3062. https://doi.org/10.1016/j.apsb.2022.02.002

[4] Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31:455–461. https://doi.org/10.1002/jcc.21334

[5] Guedes IA, Pereira FSS, Dardenne LE (2018) Empirical scoring functions for structure-based virtual screening: applications, critical aspects, and challenges. Front Pharmacol 9:1089. https://doi.org/10.3389/fphar.2018.01089

[6] Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717

[7] Ertl P, Schuffenhauer A (2009) Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminform 1:8. https://doi.org/10.1186/1758-2946-1-8

[8] Banerjee P, Kemmler E, Dunkel M, Preissner R (2024) ProTox 3.0: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res 52:W513–W520. https://doi.org/10.1093/nar/gkae303

[9] Segall MD (2012) Multi-parameter optimization: identifying high quality compounds with a balance of properties. Curr Pharm Des 18:1292–1310. https://doi.org/10.2174/138161212799436430

[10] Derringer G, Suich R (1980) Simultaneous optimization of several response variables. J Qual Technol 12:214–219. https://doi.org/10.1080/00224065.1980.11980968

[11] Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL (2012) Quantifying the chemical beauty of drugs. Nat Chem 4:90–98. https://doi.org/10.1038/nchem.1243

[12] Moghadamtousi SZ, Fadaeinasab M, Nikzad S, Mohan G, Mohd Ali H, Abdul Kadir H (2015) *Annona muricata*(Annonaceae): a review of its traditional uses, isolated acetogenins and biological activities. Int J Mol Sci 16:15625–15658. https://doi.org/10.3390/ijms160715625

[13] Jacobo-Herrera N, Pérez-Plasencia C, Castro-Torres VA, Martínez-Vázquez M, González-Esquinca AR, Zentella-Dehesa A (2019) Selective acetogenins and their potential as anticancer agents. Front Pharmacol 10:783. https://doi.org/10.3389/fphar.2019.00783

[14] Kojima N, Tanaka T (2009) Medicinal chemistry of Annonaceous acetogenins: design, synthesis, and biological evaluation of novel analogues. Molecules 14:3621–3661. https://doi.org/10.3390/molecules14093621

[15] Chandrasekhar V, Rajan K, Kanakam SRS, Sharma N, Weißenborn V, Schaub J, Steinbeck C (2025) COCONUT 2.0: a comprehensive overhaul and curation of the collection of open natural products database. Nucleic Acids Res 53:D634–D643. https://doi.org/10.1093/nar/gkae1063

[16] Goldberg RM, Sargent DJ, Morton RF, Fuchs CS, Ramanathan RK, Williamson SK et al (2004) A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. J Clin Oncol 22:23–30. https://doi.org/10.1200/JCO.2004.09.046

[17] Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A et al (2004) Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med 351:337–345. https://doi.org/10.1056/NEJMoa033025

[18] Scher HI, Fizazi K, Saad F, Taplin M-E, Sternberg CN, Miller K et al (2012) Increased survival with enzalutamide in prostate cancer after chemotherapy. N Engl J Med 367:1187–1197. https://doi.org/10.1056/NEJMoa1207506

[19] Tannock IF, de Wit R, Berry WR, Horti J, Pluzanska A, Chi KN et al (2004) Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N Engl J Med 351:1502–1512. https://doi.org/10.1056/NEJMoa040720

[20] Motzer RJ, Hutson TE, Tomczak P, Michaelson MD, Bukowski RM, Rixe O et al (2007) Sunitinib versus interferon alfa in metastatic renal-cell carcinoma. N Engl J Med 356:115–124. https://doi.org/10.1056/NEJMoa065044

[21] Sternberg CN, Davis ID, Mardiak J, Szczylik C, Lee E, Wagstaff J et al (2010) Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. J Clin Oncol 28:1061–1068. https://doi.org/10.1200/JCO.2009.23.9764

[22] Rieser MJ, Kozlowski JF, Wood KV, McLaughlin JL (1991) Muricatacin: a simple biologically active acetogenin derivative from the seeds of *Annona muricata* (Annonaceae). Tetrahedron Lett 32:1137–1140. https://doi.org/10.1016/S0040-4039(00)92027-6

[23] McCloud TG, Smith DL, Chang CJ, Cassady JM (1987) Annonacin, a novel, biologically active polyketide from *Annona densicoma*. Experientia 43:947–949. https://doi.org/10.1007/BF01951681

[24] Wu FE, Gu ZM, Zeng L, Zhao GX, Zhang Y, McLaughlin JL et al (1995) Two new cytotoxic monotetrahydrofuran Annonaceous acetogenins, annomuricins A and B, from the leaves of *Annona muricata*. J Nat Prod 58:830–836. https://doi.org/10.1021/np50120a002

[25] Zhao G, Hui Y, Rupprecht JK, McLaughlin JL, Wood KV (1992) Additional bioactive compounds and trilobacin, a novel highly cytotoxic acetogenin, from the bark of *Asimina triloba*. J Nat Prod 55:347–356. https://doi.org/10.1021/np50081a011

[26] Roblot F, Laugel T, Lebœuf M, Cavé A, Laprévote O (1993) Two acetogenins from *Annona muricata* seeds. Phytochemistry 34:281–285. https://doi.org/10.1016/S0031-9422(00)90820-3

[27] Myint SH, Cortes D, Laurens A, Hocquemiller R, Lebœuf M, Cavé A et al (1991) Solamin, a cytotoxic mono-tetrahydrofuranic γ-lactone acetogenin from *Annona muricata* seeds. Phytochemistry 30:3335–3338. https://doi.org/10.1016/0031-9422(91)83204-X

[28] Wu FE, Zeng L, Gu ZM, Zhao GX, Zhang Y, Schwedler JT et al (1995) New bioactive monotetrahydrofuran Annonaceous acetogenins, annomuricin C and muricatocin C, from the leaves of *Annona muricata*. J Nat Prod 58:909–915. https://doi.org/10.1021/np50120a014

[29] Gleye C, Duret P, Laurens A, Hocquemiller R, Cavé A (1998) cis-Monotetrahydrofuran acetogenins from the roots of *Annona muricata*. J Nat Prod 61:576–579. https://doi.org/10.1021/np970494m

[30] Champy P, Melot A, Guérineau Eng V, Gleye C, Fall D, Höglinger GU et al (2005) Quantification of acetogenins in *Annona muricata* linked to atypical parkinsonism in Guadeloupe. Mov Disord 20:1629–1633. https://doi.org/10.1002/mds.20632

[31] Potts LF, Luzzio FA, Smith SC, Hetman M, Champy P, Litvan I (2012) Annonacin in *Asimina triloba* fruit: implication for neurotoxicity. Neurotoxicology 33:53–58. https://doi.org/10.1016/j.neuro.2011.10.009

[32] Hoxhaj G, Manning BD (2020) The PI3K–AKT network at the interface of oncogenic signalling and cancer metabolism. Nat Rev Cancer 20:74–88. https://doi.org/10.1038/s41568-019-0216-7

[33] Czabotar PE, Lessene G, Strasser A, Adams JM (2014) Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. Nat Rev Mol Cell Biol 15:49–63. https://doi.org/10.1038/nrm3722

[34] Truchon JF, Bayly CI (2007) Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem. J Chem Inf Model 47:488–508. https://doi.org/10.1021/ci600426e

[35] Wager TT, Hou X, Verhoest PR, Villalobos A (2010) Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. ACS Chem Neurosci 1:435–449. https://doi.org/10.1021/cn100008c

[36] Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection. Drug Discov Today 9:430–431. https://doi.org/10.1016/S1359-6446(04)03069-7

[37] Kosugi T, Ohue M (2021) Quantitative estimate index for early-stage screening of compounds targeting protein-protein interactions. Int J Mol Sci 22:10925. https://doi.org/10.3390/ijms222010925

[38] Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S et al (2024) The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res 52:D1180–D1192. https://doi.org/10.1093/nar/gkad1004

[39] DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845. https://doi.org/10.2307/2531595

[40] Mostaghel EA, Nelson PS, Lange P, Lin DW, Taplin M-E, Balk S et al (2014) Targeted androgen pathway suppression in localized prostate cancer: a pilot study. J Clin Oncol 32:229–237. https://doi.org/10.1200/JCO.2012.48.6431

[41] Audet-Walsh É, Yee T, Tam IS, Giguère V (2017) Inverse regulation of DHT synthesis enzymes 5α-reductase types 1 and 2 by the androgen receptor in prostate cancer. Endocrinology 158:1015–1021. https://doi.org/10.1210/en.2016-1926

[42] Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. J Chem Inf Model 61:3891–3898. https://doi.org/10.1021/acs.jcim.1c00203

[43] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/S0169-409X(00)00129-0

[44] Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623. https://doi.org/10.1021/jm020017n

[45] Daina A, Zoete V (2016) A BOILED-Egg to predict gastrointestinal absorption and brain penetration of small molecules. ChemMedChem 11:1117–1121. https://doi.org/10.1002/cmdc.201600182

[46] Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53:2719–2740. https://doi.org/10.1021/jm901137j

[47] Senthilkumar S (2026) SWAN-MPO study data and reproducibility materials. Open Science Framework. https://doi.org/10.17605/OSF.IO/35QF2. Accessed 17 Aug 2026.

[48] Senthilkumar S (2026) SWAN-MPO. GitHub. https://github.com/saanviisenthilkumar/swan-mpo. Accessed 17 Aug 2026.

[49] Senthilkumar S (2026) SWAN-MPO v1.3.2 [Computer software].

Zenodo. https://doi.org/10.5281/zenodo.21988306. Accessed 17 Aug 2026.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- OnlineResource1.pdf
- OnlineResource2.zip