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

Supplementary Information

This Supplementary Information provides extended methods, additional results, and reproducibility details supporting the SWAN-MPO framework. Supplementary Methods S1–S13 describe workflow versioning, compound manifests, ligand preparation, docking and target calibration, binding aggregation, sensitivity analyses, curated-liability and COCONUT analyses, and the broad ChEMBL active–synthetic-profile benchmark. Supplementary Results S1–S12 provide extended rank-agreement, docking, sensitivity, benchmarking, curated-liability, and property-space results. Supplementary Tables S1–S23 and Figures S1–S4 provide machine-readable manifests, docking audits, scoring outputs, benchmark results, parameter definitions, and extended sensitivity analyses.

Supplementary Methods S1. Workflow versioning, protocol locking, and reproducibility tracking

The reproducibility package preserves the execution order, raw and processed input files, intermediate scoring tables, final analysis outputs, figure-generation scripts, software versions, command-line calls, file hashes, execution logs, and runtime information used for each analysis layer. The primary scoring implementation is locked to SHA-256

97799e91f23803de3b9a477aea3bbcfb061f1b1d9abebdecfdadcd1b537f4454, and the target-calibration resource is locked to SHA-256 e7584f10c0f970bb54ddf39cfdacbb6cb061bc6a3d2bd795e2d5f119efc7287c. Both hashes are byte-verifiable in the archived software release. After the primary framework was fixed, downstream analyses included Top2Node and Breadth aggregation, ligand-efficiency adjustment, strict-pass-only target restriction, Monte Carlo rank perturbation, external ChEMBL benchmarking, curated-liability sensitivity, and COCONUT-derived property-space contextualization.

Supplementary Methods S2. Compound-selection manifests and auditability

Supplementary Table S1 documents the internal discovery-panel compound manifest and disease-panel membership. The internal discovery panels were broader than the subsequent *A. muricata*-restricted expanded deployment and included selected Annonaceae-family comparator scaffolds. Clinical comparator selection was intended to provide disease-context references spanning established colorectal, prostate, and RCC treatment paradigms [16–21]. Representative Annonaceae scaffolds included in the internal or expanded manifests are documented in primary isolation literature [22–29]. Bullatacin, included only in the internal RCC panel, was treated as an Annonaceae-family comparator; its reported source is *Asimina triloba* [25], and it was not part of the *A. muricata*-restricted expanded deployment. Supplementary Table S2 documents the expanded *A. muricata* compound manifest, including compound names, structure and SMILES fields, PubChem identifiers where available, source-status fields, inclusion status, and structure-reconciliation fields. ChEMBL-derived record identifiers, activity measurements, clinical-status annotations, class labels, deduplication outputs, and excluded-record logs are preserved in the reproducibility package for the external ChEMBL analyses. Supplementary Table S4 documents

the retrospective benchmark compounds, their benchmark role, natural-product origin class, documented safety or developability rationale, rationale for primary inclusion or stress-test placement, and literature/source notes.

Supplementary Methods S3. Analysis-specific ligand preparation, retrospective benchmark docking, redocking validation, and calibration status

For the expanded 59-compound *A. muricata* library, curated SMILES were parsed using RDKit, and stereochemical information encoded in the input records was retained during three-dimensional structure generation. Explicit hydrogens were added, and one conformer per ligand was generated using ETKDGv3 with random seed 2026. When initial embedding failed, embedding was repeated using random-coordinate initialization with the random seed 2026. UFF geometry optimization was attempted for up to 1,000 iterations. Each structure was then written to SDF and converted to AutoDock Vina-compatible PDBQT format using Open Babel with Gasteiger partial charges.

For the retrospective benchmark, validated SMILES were parsed using RDKit, explicit hydrogens were added, and one conformer per ligand was generated using ETKDGv3 with random seed 2026. Small-ring and macrocycle torsion handling were enabled during embedding. MMFF geometry optimization was attempted when all required MMFF parameters were available, with UFF used otherwise. The initial retrospective ligand-preparation workflow used a geometry-optimization limit of 500 iterations. Midostaurin, Geldanamycin, and Illudin S were prepared subsequently using the same general workflow with a geometry-optimization limit of 1,000 iterations. Structures were written to SDF and converted to PDBQT using Open Babel with Gasteiger partial charges. Ligand-preparation summaries recorded successful parsing,

conformer generation, geometry optimization, and nonempty PDBQT generation for each retained ligand.

For native-ligand redocking, ligand coordinates extracted from the crystallographic structures were saved as PDB files. Open Babel was used to convert each native ligand from PDB to MOL2 with hydrogen addition and subsequently from MOL2 to PDBQT with Gasteiger partial charges. This procedure was applied to AGS for mTOR/4JSP, 1XJ for Bcl-2/4LVT, dihydrotestosterone for AR/2AMA, MIY for 5AR2/3V3N, M1J for PI3Kα/6OAC, and WFE for AKT1/3MVH. The crystallographic coordinates were retained for redocking. Redocked PDBQT poses were converted back to PDB only for heavy-atom RMSD comparison against the corresponding crystallographic reference coordinates.

Retrospective benchmark docking was performed only for the curated boundary-stress benchmark compounds. The docking matrix comprised 15 ligands and six unique redocking-qualified targets, including four strict top-pose validation targets and two generated-mode recovery targets. Each ligand–target pair was evaluated across 20 grid centers, producing 1,800 completed AutoDock Vina jobs. Docking used a 40.0 Å cubic search box, grid spacing of 0.375 Å, exhaustiveness of 12, nine output modes, and random seed 2026, consistent with the internal and expanded-library workflows. Native-ligand redocking was used to evaluate pose reproduction for receptors with suitable co-crystallized drug-like ligands. Strict validation was defined as a heavy-atom root-mean-square deviation (RMSD) ≤ 2.0 Å for the top-ranked redocked pose and crystallographic reference pose. Generated-mode recovery was monitored and reported separately when a native-like pose with an RMSD ≤ 2.0 Å was recovered among the generated modes but was not selected as the top-ranked pose. For receptors where the top-ranked redocking mode did not reproduce the crystallographic geometry but a native-like mode was

successfully recovered, target-calibrated binding normalization utilized the native-like generated mode energy. Targets with generated-mode recovery were retained in the calibrated target set using the validated native-like mode energy. Comparative-only targets were excluded from the primary reference-calibrated BestNode aggregation and retained only for target-level comparative or sensitivity analyses. Target-specific strict top-pose, generated-mode recovery, and comparative-only classifications, together with native-ligand identities, mode-level RMSD values, and reference-energy sources, are reported in Supplementary Table S6.

Supplementary Methods S4. Grid-center aggregation and docking-audit fields

For each completed ligand–target pair, the target-level summary retained the number of completed grid centers, best ΔG, median ΔG, mean ΔG, standard deviation, first quartile, third quartile, and interquartile range. These fields allowed the primary median-based binding value to be audited against the complete distribution of docking energies observed across the 20 grid centers. Median ΔG across 20 grid centers was locked as the primary grid aggregation statistic to reduce dependence on a single favorable grid center or stochastic Vina artifact. Internal target-level audit fields are reported in Supplementary Table S7, while the complete expanded-library grid-level deployment records are reported in Supplementary Table S3 and preserved with the corresponding target-level summaries in the reproducibility package. The same grid-center aggregation procedure was applied to the internal discovery panels, completed expanded-library screen, and retrospective docking matrix.

After grid-center aggregation, median ΔG values were transformed into target-level binding desirability values before disease-panel binding aggregation. Target-specific calibration was performed separately within each receptor context so that binding desirability reflected receptor-

specific reference energies before cross-target aggregation. Redocking/reference status was used to define target-level calibration and confidence annotations, not to change compound inclusion, docking inputs, or primary safety, ADME, or liability components. Targets classified as comparative-only were excluded from the primary reference-calibrated BestNode aggregation and retained only for target-level comparative or sensitivity analyses.

Supplementary Methods S5. Expanded-library deployment and completion audit

Expanded-library eligibility, structure sources, inclusion status, and reconciliation decisions were recorded in the frozen compound manifest in Supplementary Table S2. Ambiguous entries, duplicate structures, and compounds resolved to non-*A. muricata* species were excluded or flagged in the same manifest. After structure reconciliation and exact-SMILES deduplication, the final expanded docking input contained 59 unique ligands mapped across nine unique targets. The expanded docking setup retained the same unique-target strategy and reused shared targets across panels instead of redundantly redocking identical receptor-ligand combinations by disease panel. Completion required 20 successful grid-center evaluations for each of the 531 ligand-target blocks, yielding 10,620 completed Vina jobs. The 531 target summaries were mapped to the colorectal, prostate, and RCC target panels in order to generate 177 panel-specific compound records before BestNode scoring and sensitivity analysis.

Ligand-target-grid deployment status, receptor and ligand files, grid definitions, docking outputs, runtime fields, and error fields are reported in Supplementary Table S3 and the associated repository records.

Supplementary Methods S6. BestNode, Top2Node, Breadth, and isolated-node diagnostics

BestNode was the primary disease-panel binding aggregation term. Let $\mathcal{T}_{l,p}^{\mathrm{cal}}$ denote the set of targets for which an eligible reference-calibrated binding desirability was available for ligand *l* in disease panel *p*. This calibrated target set included targets supported by either a strict top-pose redocking pass or a validated generated-mode recovery reference. Targets lacking a suitable reference ligand were retained in the docking dataset, target-level audit tables, and within-target comparative interpretation, while primary BestNode, Top2Node, and Breadth aggregation was restricted to reference-calibrated targets.

When at least one reference-calibrated target was available, BestNode selected the highest target-level calibrated binding desirability after median grid-center aggregation:

$$d_{\mathrm{Binding,BestNode},l,p} = \max_{t \in \mathcal{T}_{l,p}^{\mathrm{cal}}} d_{\mathrm{Binding},l,t}$$

If no eligible reference-calibrated target was available for a ligand–panel combination, the corresponding BestNode binding value was recorded as unavailable rather than inferred from an uncalibrated cross-receptor comparison.

To make this maximum operation auditable, the selected BestNode target, second-best node, best-to-second ratio, and isolated-node flag were retained as diagnostic fields. The isolated-node flag was retained as a descriptive diagnostic alongside the calculated SWAN scores.

The Top2Node binding sensitivity was defined as:

$$d_{\mathrm{Binding,Top2Node},l,p} = \begin{cases} \sqrt{d_{\mathrm{Binding},l,t_{(1)}} d_{\mathrm{Binding},l,t_{(2)}}} & \left|\mathcal{T}_{l,p}^{\mathrm{cal}}\right| \geq 2 \\ d_{\mathrm{Binding,BestNode},l,p} & \left|\mathcal{T}_{l,p}^{\mathrm{cal}}\right| = 1 \\ \mathrm{NA} & \left|\mathcal{T}_{l,p}^{\mathrm{cal}}\right| = 0 \end{cases}$$

Where $t_{(1)}$ and $t_{(2)}$ denote the targets with the highest and second-highest binding desirabilities within the applicable target set for panel p. When only one calibrated target was available, Top2Node reverted to the BestNode value rather than becoming unavailable solely because a second reference-calibrated target was absent.

The Breadth binding sensitivity was defined as:

$$d_{\text{Binding,Breadth},l,p} = \exp\left[\frac{1}{\left|\mathcal{T}_{l,p}^{\text{cal}}\right|}\sum_{t\in\mathcal{T}_{l,p}^{\text{cal}}}\log\left(d_{\text{Binding},l,t}\right)\right]$$

Breadth summarized binding support across all reference-calibrated targets available for ligand l in panel p. As all target-level desirabilities were bounded above zero, the logarithmic geometric-mean formulation remained numerically defined.

Using these alternative binding terms, SWAN-Top2Node and SWAN-Breadth sensitivity scores were calculated using the same equal-weight geometric integration as the primary SWAN-BestNode score:

$$\mathbf{SWAN}_{\mathbf{Top2Node},l,p} = \left(\boldsymbol{d}_{\mathbf{Safety},l,p} \times \boldsymbol{d}_{\mathbf{ADME},l,p} \times \boldsymbol{d}_{\mathbf{Binding,Top2Node},l,p} \times \boldsymbol{d}_{\mathbf{Liability},l,p}\right)^{\frac{1}{4}}$$

$$\mathbf{SWAN}_{\mathbf{Breadth},l,p} = \left(\boldsymbol{d}_{\mathbf{Safety},l,p} \times \boldsymbol{d}_{\mathbf{ADME},l,p} \times \boldsymbol{d}_{\mathbf{Binding,Breadth},l,p} \times \boldsymbol{d}_{\mathbf{Liability},l,p}\right)^{\frac{1}{4}}$$

These sensitivity scores were used to evaluate whether prioritization depended on a single selected target node or remained directionally stable when binding support was summarized across the top two nodes or across the full panel.

Under the published calibration, mTOR was the only reference-calibrated target eligible for the

colorectal panel; consequently, colorectal BestNode, Top2Node, and Breadth binding terms reduce to calibrated support from a single eligible node. Under strict-top-pose restriction, the colorectal panel contained no eligible targets and was therefore not estimable.

Supplementary Methods S7. Ligand-efficiency sensitivity

Ligand-efficiency sensitivity was performed to test whether larger natural-product scaffolds were favored by raw docking energies because of greater contact surface area or heavy-atom count. For each ligand-target pair, raw ligand efficiency was calculated as:

$$LE_{l,t} = -\frac{\widetilde{\Delta G}_{l,t}}{N_{\text{heavy},l}}$$

Here, $\widetilde{\Delta G}_{l,t}$ denotes the median docking energy for ligand *l* against target *t* and $N_{\text{heavy},l}$ denotes the ligand heavy-atom count. Heavy-atom counts were generated from ADME/SMILES-derived ligand entries using Python and RDKit-compatible parsing.

Raw ligand efficiency was transformed into a bounded desirability value before integration:

$$d_{LE,l,t} = clip_{\,0.01,1.00}\left[0.01 + 0.99\left(\frac{LE_{l,t} - LE_{worst,t}}{LE_{best,t} - LE_{worst,t}}\right)\right]$$

Here, $LE_{\text{best},t}$ and $LE_{\text{worst},t}$ denote the target-specific favorable and least-favorable ligand-efficiency reference values used in the implemented transformation. The clipping operation bounded all resulting desirabilities between 0.01 and 1.00. Target-specific ligand-efficiency reference values and transformation parameters are reported in the supplementary parameter table.

The LE-adjusted binding sensitivity was then:

$$d_{\text{Binding,LEadj},l,p} = \sqrt{d_{\text{Binding,BestNode},l,p} d_{LE,l,t^*_{l,p}}, p}$$

Here, $t^*_{l,p}$denotes the target selected by the primary BestNode calculation for ligand *l* in disease panel p. The LE-adjusted analysis therefore evaluated ligand efficiency at the same target selected by the primary BestNode calculation.

The corresponding LE-adjusted SWAN sensitivity score was calculated as:

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

LE-adjusted outputs are reported in Supplementary Table S17 and provide a size-normalized sensitivity assessment of the binding contribution.

Supplementary Methods S8. Strict-pass-only target-restriction sensitivity

Strict-top-pose target-restriction sensitivity was applied separately to the internal discovery-panel score matrices and the completed 59-compound expanded-library score matrices. Within each disease panel, eligible targets were restricted to receptors meeting the strict top-pose redocking criterion, thereby removing generated-mode recovery targets from the calibrated target set. For the internal analysis, BestNode, Top2Node, and Breadth scores were recalculated within the reduced target subset; the final expanded-library closeout recalculated the strict-pass BestNode score.

Target coverage for the strict-pass-only subset was recorded separately by panel, target, PDB ID, and number of available compounds. The strict-pass-only analysis preserved the same safety, ADME, and liability values as the all-target model, changing only the eligible docking target set.

Strict-pass-only target coverage and rank-shift outputs are reported in Supplementary Table S18. A disease panel was recorded as not estimable when its prespecified target set contained no strict top-pose target.

Supplementary Methods S9. Monte Carlo rank-stability analysis

Monte Carlo domain-weight perturbation was applied separately to the internal discovery panel as well as the completed 59-compound expanded-library score matrix. In each analysis, the fixed safety, ADME, BestNode binding, and liability desirabilities were recombined using the iteration-specific domain weights, and compound ranks were recalculated separately within each disease panel.

$$\mathrm{SWAN}_{l,p}^{(r)} = d_{\mathrm{Safety},l,p}^{w_{\mathrm{Safety}}^{(r)}} d_{\mathrm{ADME},l,p}^{w_{\mathrm{ADME}}^{(r)}} d_{\mathrm{Binding,BestNode},l,p}^{w_{\mathrm{Binding}}^{(r)}} d_{\mathrm{Liability},l,p}^{w_{\mathrm{Liability}}^{(r)}}$$

$$w_{\mathrm{Safety}}^{(r)} + w_{\mathrm{ADME}}^{(r)} + w_{\mathrm{Binding}}^{(r)} + w_{\mathrm{Liability}}^{(r)} = 1$$

For the internal discovery-panel analysis, weight vectors were sampled from symmetric Dirichlet distributions centered on equal weighting. The moderate perturbation regime used Dirichlet parameters (20, 20, 20, 20), and the broad perturbation regime used Dirichlet parameters (4, 4, 4, 4). Each regime consisted of 10,000 iterations, and random seed 2026 was used.

For the completed expanded-library analysis, the moderate perturbation regime used Dirichlet parameters (20, 20, 20, 20), and the broad perturbation regime used Dirichlet parameters (5, 5, 5, 5). Each regime consisted of 20,000 iterations, and random seed 2026 was used.

Only the four domain weights were varied, while compound-level desirabilities, BestNode target selection, target-calibration status, descriptor transformations, and compound inclusion were held constant. Scores were recalculated and ranked separately within each disease panel.

For the internal analysis, descending ranks were assigned using the minimum-rank rule for tied scores. Retained outputs included the median simulated score, 2.5th and 97.5th percentiles of simulated score, median simulated rank, probability of rank 1, probability of top-two placement, and probability of top-three placement.

For the expanded-library analysis, stable descending ordinal ranking was used, with a fixed compound order retained for exact numerical ties. Retained outputs included the deterministic SWAN-BestNode score and rank, median simulated score, 2.5th and 97.5th percentiles of simulated score, probability of rank 1, probability of top-three placement, median simulated rank, and 2.5th and 97.5th percentiles of simulated rank.

Complete compound-level summaries, iteration-level outputs, and sampled weight vectors are reported in Supplementary Table S19 and the associated repository files.

Supplementary Methods S10. Curated-liability sensitivity

Retrospective docking was completed for 15 prepared ligands across six unique redocking-qualified targets, including four strict top-pose validation targets and two generated-mode recovery targets. Each ligand–target pair was evaluated across 20 grid centers, producing 1,800 completed AutoDock Vina jobs. The reported benchmark comprised six favorable comparators and nine liability anchors. Docking used a 40.0 Å cubic search box, grid spacing of 0.375 Å, exhaustiveness of 12, nine output modes, and random seed 2026, consistent with the internal and expanded-library workflows.

The raw automated retrospective analysis used the predefined safety, ADME, liability, and reference-calibrated binding inputs without manual modification of the underlying ProTox or

SwissADME outputs. Curated-liability information was stored separately from these automated inputs using four permitted binary fields: chronic or developability toxicity, genotoxic-hepatotoxicity, hepatotoxicity, and nephrotoxicity. These fields were stored separately from the automated predictions and used to generate the curated-liability sensitivity score. The exact categorical mappings and aggregation rule used to construct the curated-liability desirability are reported in the supplementary parameter table.

Anchor-level ablations were performed one annotation at a time to quantify how much each documented liability annotation contributed to overall separation. As documented liabilities informed both cohort construction and curated annotations, this analysis evaluates the framework's ability to encode known literature-supported liabilities rather than prospective hazard prediction. For each ablation, the selected curated annotation was reverted to its corresponding raw automated liability value while all other compound inputs, target-level binding values, scoring transformations, and benchmark labels were held constant. Performance metrics were then recalculated using the unchanged benchmark cohort.

Compound-level raw and curated retrospective scores are reported in Supplementary Table S15. Summary metrics for the curated-liability ablation variants are reported in Supplementary Table S16.

Supplementary Methods S11. Descriptor-space principal component analyses

Principal component analysis was performed separately for the retrospective boundary-stress benchmark and the COCONUT-derived property-space analysis.

The retrospective PCA used the frozen 15-compound primary boundary-stress descriptor matrix. The 14 physicochemical, developability, and toxicological variables used for this analysis were standardized before singular value decomposition.

The COCONUT PCA used the property-descriptor matrix defined in Supplementary Methods S12. COCONUT nearest-neighbor coordinates and PCA model parameters and loadings are reported in Supplementary Table S22 and the associated repository files.

Supplementary Methods S12. COCONUT-derived property-space implementation

COCONUT 2.0 structures [15] were processed using Python and RDKit version 2026.03.2. Structures that could not be parsed successfully were excluded, leaving 738,801 structures in the final scored universe. For each retained structure, molecular weight, calculated logP, topological polar surface area, hydrogen-bond donor count, hydrogen-bond acceptor count, rotatable-bond count, and heavy-atom count were calculated from the parsed molecular graph. The property-only score was calculated using the predefined bounded property-desirability transformations and geometric aggregation implemented in the final COCONUT scoring script. Exact descriptor membership, transformation direction, preferred or tolerated ranges, boundary values, and aggregation rules are reported in the supplementary parameter table. The workflow retained the compound-level property score, global rank, percentile position, descriptor values, nearest-neighbor outputs, and PCA model parameters used in the supplementary analyses.

Docking, ProTox predictions, curated-liability annotations, target biology, disease-panel scoring, and biological activity labels were not included in the COCONUT workflow. Muricatacin’s global-position output, nearest-neighbor results, and PCA model parameters and loadings are

reported in Supplementary Tables S22A–S22C, with the larger underlying property-space artifacts preserved in the associated repository files.

Supplementary Methods S13. Broad ChEMBL active–synthetic-profile benchmark

A supplementary broad benchmark used ChEMBL-derived bioactivity records (Kd, Ki, and IC50) spanning seven eligible targets: mTOR, EGFR, 5AR2, AR, PI3Kα, caspase-3, and AKT1; Complex I and Bcl-2 had insufficient coverage [38]. Filtering, natural-product-like selection, SMILES validation, and InChIKey deduplication yielded 1,184 compound-target records representing 1,097 unique actives. For the broad benchmark, standard pChEMBL values were converted to binding desirability using a bounded increasing transformation from 4.0 to 9.0. Of the 1,097 actives, 770 had pChEMBL values; the remaining 327 received a neutral binding desirability of 0.50 because no comparable pChEMBL value was available.

RDKit calculated molecular weight, logP, TPSA, hydrogen-bond donors, hydrogen-bond acceptors, and rotatable-bond count from SMILES to construct a reproducible ADME proxy. These descriptors were converted using the prespecified bounded desirability functions and combined geometrically to form the RDKit ADME proxy; the same transformations were applied to the synthetic comparison profiles. As compound-specific ProTox predictions and curated liability annotations were unavailable, safety and liability desirabilities were fixed at 0.50 and 0.60, respectively, for both the active and synthetic-profile classes. These class-symmetric constants therefore did not contribute discriminatory information.

Three synthetic descriptor profiles were generated per active compound, yielding 3,291 profiles under random seed 2026. Molecular weight, logP, and TPSA were sampled from bounded normal distributions centered on the active-set medians, with standard deviations equal to 1.5

times the corresponding active-set standard deviations. Hydrogen-bond donor, hydrogen-bond acceptor, and rotatable-bond counts were sampled from prespecified integer ranges. Binding desirability for the synthetic profiles was sampled uniformly from 0.01 to 1.00. The profiles were descriptor-level comparators rather than molecular structures and were not docked.

Integrated scores were calculated using the equal-weight geometric mean of binding, ADME, safety, and liability desirabilities. As safety and liability were held constant across both classes, discrimination in this supplementary benchmark arose from potency-derived binding and RDKit ADME integration under the synthetic-profile design. The integrated score was compared with binding-only and ADME-only baselines using ROC-AUC, BEDROC $\alpha = 20$, and enrichment factors at 1% and 5%. Pairwise ROC-AUC differences were evaluated using DeLong analysis.

Supplementary Results S1. Rank-agreement analysis

Kendall's τ_b between QED-only and SWAN-BestNode was 0.000 in colorectal cancer, 0.473 in prostate cancer, and 0.238 in RCC. Kendall's τ_b between QED+ADME+BestNode and SWAN-BestNode was −0.200 in colorectal cancer, 0.714 in prostate cancer, and 0.429 in RCC. These internal-panel correlations were used as descriptive rank-agreement measures preceding expanded-library evaluation. Rank agreement was panel-dependent, showing that SWAN-BestNode produced ranking behavior distinct from QED-only and QED+ADME+BestNode across the internal panels. Agreement was limited in colorectal cancer and RCC, whereas partial alignment was observed in the prostate panel. Complete coefficients, p-values, and panel sizes are reported in Supplementary Table S10.

Supplementary Results S2. Internal best-grid versus median-grid sensitivity

As a scaffold-specific sensitivity check, Muricatacin's internal binding rank was unchanged when target-level docking was summarized using median-grid rather than best-grid ΔG. It remained rank 3 in colorectal cancer, rank 8 in prostate cancer, and rank 7 in RCC under both summaries. Its internal binding-rank position was therefore preserved under median-grid aggregation. Full grid-center audit fields and the best/median docking summaries are reported in Supplementary Table S7 and visualized in Supplementary Figure S1.

Supplementary Results S3. Node-specific BestNode diagnostics

In all three internal discovery panels, mTOR was selected as Muricatacin's BestNode target. Selected BestNode target, second-best node, best-to-second ratio, and isolated-node diagnostic fields are reported in Supplementary Table S9.

Supplementary Results S4. Expanded *Annona muricata* natural-product deployment

The completed expanded screen consisted of 59 unique ligands, nine targets, 531 ligand–target blocks, and 10,620 successful Vina evaluations, producing 177 panel-specific score records.

For the representative scaffold Muricatacin, mTOR was selected as the BestNode in all three panels. Muricatacin remained rank 1 under BestNode, Top2Node, and Breadth scoring in colorectal cancer; in prostate cancer and RCC, it ranked 2 under BestNode and 3 under both Top2Node and Breadth scoring.

Expanded-screen compound inclusion, exclusion, and structure-reconciliation records are reported in Supplementary Table S2, and docking-deployment records are reported in Supplementary Table S3. Complete expanded-library primary scores are reported in Supplementary Table S8B, while binding-aggregation diagnostics, selected nodes, and associated

sensitivity outputs are reported in Supplementary Table S9 and the accompanying repository files.

Supplementary Results S5. Ligand-efficiency sensitivity

In the expanded library, LE-adjusted rankings remained significantly concordant with the primary SWAN-BestNode rankings in colorectal cancer (τ_b=0.813, p=9.34×10−20), prostate cancer (τ_b=0.804, p=2.43×10−19), and RCC (τ_b=0.804, p=2.43×10−19). Muricatacin retained rank 1 in colorectal cancer and shifted from rank 2 to rank 1 in prostate cancer and RCC, with LE-adjusted scores of 0.751863, 0.677281, and 0.677281, respectively.

Complete internal and expanded-library ligand-efficiency values, LE-adjusted binding terms, SWAN scores, ranks, and rank-concordance outputs are reported in Supplementary Table S17 and the associated repository files.

Supplementary Results S6. Strict-top-pose target-restriction results

In the internal prostate panel, Muricatacin shifted from primary rank 1 to strict-pass rank 2, while remaining at rank 2 in RCC. The colorectal analysis was not estimable because none of its prespecified targets met the strict top-pose criterion. In the expanded library, Muricatacin shifted from rank 2 to rank 4 in prostate cancer and from rank 2 to rank 3 in RCC under strict-pass-only scoring. Across the complete expanded library, strict-pass rankings remained highly concordant with the primary rankings in the two estimable panels (Kendall's τ_b = 0.871 in prostate cancer and 0.898 in RCC). Complete internal and expanded-library strict-pass scores, rank shifts, panel-level eligible-target summaries, and full-library concordance statistics are reported in Supplementary Table S18.

Supplementary Results S7. Monte Carlo rank-stability analysis

Monte Carlo perturbation generated compound-level score and rank distributions for every compound in the internal discovery panels and completed expanded library. Complete compound-level probabilities, simulated score and rank summaries, and iteration-level outputs are reported in Supplementary Table S19 and the associated repository files.

In the internal panels, Muricatacin's median ranks under moderate perturbation were 1 in colorectal cancer, 1 in prostate cancer, and 2 in RCC. The corresponding probabilities of rank 1 were 0.9940, 0.7538, and 0.0074; top-two probabilities were 1.0000, 0.8674, and 0.9997; and top-three probabilities were 1.0000 in all three panels. Under broader internal weight stress, median ranks remained 1, 1, and 2, with top-three probabilities of 1.0000, 0.9986, and 0.9999.

In the completed expanded library, Muricatacin had median ranks of 1, 2, and 2 under moderate perturbation, with rank-1 probabilities of 0.96085, 0.28910, and 0.28910 and top-three probabilities of 0.98925, 0.59785, and 0.59785. Under broad perturbation, median ranks remained 1, 2, and 2; rank-1 probabilities were 0.79905, 0.40235, and 0.40235; and top-three probabilities were 0.87510, 0.57090, and 0.57090. The corresponding 95% simulated-rank intervals were 1–30 in colorectal cancer and 1–5 in both prostate cancer and RCC.

Supplementary Results S8. Complete DeLong pairwise comparisons

Pairwise ROC-AUC comparisons for the broad ChEMBL active-synthetic-profile benchmark and descriptor-compatible MPO comparator benchmark were evaluated using DeLong analysis. Complete model-pair outputs included ΔAUC, standard error, z-statistic, p-value, and 95%

confidence interval. Complete pairwise DeLong outputs are reported in Supplementary Tables S12 and S14.

Supplementary Results S9. Curated-Liability benchmark descriptor-space PCA

The first two principal components explained 62.41% of the total variance, with 37.19% explained by PC1 and 25.23% by PC2. PCA loadings, primary boundary-stress compound coordinates, and explained-variance outputs are reported in Supplementary Table S20.

Supplementary Results S10. Broad ChEMBL active–synthetic-profile benchmark

Using the ChEMBL target set defined in Methods, the framework was first evaluated on 1,097 natural-product-like ChEMBL actives and 3,291 synthetic descriptor profiles. Of the 1,097 active compounds, 770 had observed pChEMBL values and 327 received the prespecified neutral binding desirability of 0.50 because no comparable pChEMBL value was available. Safety and liability desirabilities were fixed identically across both classes to prevent these unavailable domains from contributing class-specific information. Performance metrics for the broad active–synthetic-profile benchmark are shown in Supplementary Table S11. The integrated score achieved ROC-AUC = 0.810, compared with 0.612 for binding-only scoring and 0.811 for ADME-only scoring.

Early-enrichment behavior differed from global ROC-AUC: ADME-only scoring led BEDROC and EF@5%, whereas binding-only scoring led EF@1% (Supplementary Table S11).

DeLong analysis showed that the integrated score significantly outperformed binding-only scoring (ΔAUC = 0.1980, SE = 0.00921, z = 21.51, $p < 10^{-100}$, 95% confidence interval 0.1800 to 0.2160) but did not differ significantly from ADME-only scoring (ΔAUC = −0.0010, SE =

0.00410, z = −0.245, p = 0.806, 95% confidence interval −0.0091 to 0.0070). ADME-only scoring also significantly outperformed binding-only scoring (ΔAUC = 0.1990, SE = 0.01183, z = 16.82, p = 1.85×10^{-63}, 95% confidence interval 0.1758 to 0.2222).

The integrated score therefore substantially improved global discrimination relative to binding-only ranking but performed similarly to ADME-only scoring, primarily reflecting binding-plus-ADME integration, since safety and liability were held constant.

Supplementary Results S11. Curated-liability boundary-stress analysis

The boundary-stress analysis compared six translationally favorable comparators with nine compounds carrying documented safety or developability liabilities (Supplementary Table S4).

Raw automated SWAN-BestNode showed little class separation with ROC-AUC values of 0.537 in colorectal cancer and 0.519 in both prostate cancer and RCC. These near-random values indicate that the automated safety, ADME, liability, and calibrated-binding inputs did not independently distinguish the favorable comparators from the documented liability anchors in the small benchmark.

Binding also provided limited discrimination in this setting: fourteen of the fifteen compounds reached or approached the upper bound of BestNode binding desirability across the six redocking-supported targets (Supplementary Table S21). Binding operated mainly as a pocket-exploitation compatibility term rather than a mechanism-specific classifier among these compounds.

Adding literature-curated liability annotations produced complete separation of the six favorable comparators and nine documented liability anchors in this prespecified boundary-stress cohort.

As documented liabilities informed both cohort construction and curated annotation, the complete-separation result reflects encoding of known literature-supported hazards. The largest single-anchor reversions reduced ROC-AUC to 0.907 and BEDROC $\alpha = 20$ to 0.806, demonstrating the contribution of individual curated annotations to the resulting separation (Supplementary Tables S15–S16).

Supplementary Results S12. Global COCONUT property-space contextualization

Among 738,801 RDKit-parseable COCONUT compounds, Muricatacin achieved a property-only desirability score of 0.964, ranked 404,582nd, and occupied the 45.24th percentile of the scored distribution. Complete descriptor and chemical-space outputs are reported in Supplementary Tables S22A–S22C.

This slightly below-median property-space placement indicates that Muricatacin's integrated panel ranking was not explained by exceptional generic physicochemical favorability alone. Global property-space contextualization is shown in Supplementary Figure S4.

TABLES

Supplementary Table S1. Internal discovery-panel compound manifest and disease-panel membership.

This table reports the compounds included in each internal discovery panel together with normalized compound identifiers. Internal component scores and final ranks are reported separately in Supplementary Table S8A, with binding-node and ligand-efficiency diagnostics reported in Supplementary Tables S9A and S17A. Provided as: Table_S1_internal_discovery_manifest.csv.

Supplementary Table S2. Expanded *Annona muricata* natural-product compound manifest.

This table reports the expanded *A. muricata* natural-product compound list, including compound names, SMILES/source fields, PubChem identifiers when available, inclusion status, and structure-reconciliation notes. Provided as: Table_S2_expanded_annona_compound_manifest.csv.

Supplementary Table S3. Expanded *Annona muricata* docking-deployment matrix.

This table reports the expanded *A. muricata* docking deployment across ligand-target-grid combinations, including receptor and ligand files, grid definitions, docking status, best affinity values, runtime, and error fields where applicable. Provided as: Table_S3_expanded_annona_docking_deployment.csv.

Supplementary Table S4. Retrospective benchmark compound manifest.

Table S4A reports the frozen 15-compound primary boundary-stress benchmark, including benchmark-group assignments, curated liability annotations, and inclusion rationale. Table S4B

reports the corresponding 15-compound primary structure inventory, including compound class, benchmark role, analysis SMILES, and structure source. Provided as: Table_S4A_retrospective_benchmark_manifest_PRIMARY.csv and Table_S4B_retrospective_primary_structure_inventory.csv.

Supplementary Table S5. Target/PDB master table.

This table reports the disease-panel targets, PDB structures, docking-site definitions, validation status, native ligand availability, grid-center sensitivity status, and target rationale used in the docking workflow. Provided as: Table_S5_target_pdb_master.csv.

Supplementary Table S6. Redocking and reference-calibration table.

This table reports redocking validation outcomes, strict top-pose recovery, generated-mode recovery, RMSD values, reference docking energies, and the reference-calibration status used for target-level binding desirability. Provided as: Table_S6_redocking_reference_calibration.csv.

Supplementary Table S7. Internal target-level docking and grid-center audit.

This table reports target-level docking summaries for the internal discovery-panel compounds, including best, median, mean, standard deviation, interquartile range, reference-calibrated binding desirability, and ligand-efficiency fields. Provided as: Table_S7_internal_target_level_grid_audit.csv.

Supplementary Table S8. Final SWAN-MPO scores and ranks.

This table family reports the final SWAN-MPO scoring outputs. Table S8A reports internal discovery-panel scores and ranks, including safety, ADME, liability, BestNode binding, Top2Node, Breadth, ligand-efficiency-adjusted scores, and associated rank-shift fields. Table

S8B reports the final panel-level outputs for the completed 59-compound *A. muricata* screen across the colorectal, prostate, and RCC panels, including selected BestNode target, BestNode binding desirability, safety, ADME, liability, SWAN-BestNode score, binding-only rank, and SWAN-BestNode rank. Provided as: Table_S8A_internal_final_SWAN_scores.csv and Table_S8B_expanded_library_final_SWAN_scores.csv.

Supplementary Table S9. BestNode, Top2Node, and Breadth diagnostics.

Table S9A reports internal discovery-panel binding diagnostics, and Table S9B reports the corresponding expanded-library diagnostics. Reported fields include selected BestNode target, second-best node, best-to-second ratio, Top2Node binding, Breadth binding, isolated-node annotations, and associated SWAN scores and ranks where applicable. Provided as: Table_S9A_internal_bestnode_top2_breadth_diagnostics.csv and Table_S9B_expanded_bestnode_top2_breadth_diagnostics.csv.

Supplementary Table S10. QED/simple-MPO Kendall rank-agreement analysis.

This table reports Kendall rank-agreement comparisons between SWAN-MPO and simpler comparator scoring schemes across internal disease panels. Provided as: Table_S10_kendall_rank_agreement.csv.

Supplementary Table S11. Broad ChEMBL active–synthetic-profile benchmark metrics.

This table reports ROC-AUC, BEDROC $\alpha = 20$, EF@1%, EF@5%, and cohort counts for SWAN-MPO, binding-only, and ADME-only scoring in the benchmark of 1,097 natural-product-like ChEMBL actives and 3,291 synthetic descriptor profiles. Of the 1,097 actives, 770 had observed pChEMBL values and 327 received a neutral binding desirability of 0.50 because no comparable pChEMBL value was available. Safety and liability desirabilities were held

constant across classes and therefore contributed no class-specific discriminatory information. Provided as:

Table_S11_broad_ChEMBL_metrics.csv.

Supplementary Table S12. Broad ChEMBL active–synthetic-profile DeLong comparisons.

This table reports pairwise DeLong comparisons among SWAN-MPO, binding-only, and ADME-only ROC-AUC values in the broad ChEMBL active–synthetic-profile benchmark, including AUC differences, standard errors, z-statistics, p-values, and 95% confidence intervals. Provided as: Table_S12_broad_ChEMBL_DeLong_pairwise_comparisons.csv.

Supplementary Table S13. ChEMBL activity/status benchmark and domain-contribution analyses.

This table family reports the complete ChEMBL activity/status benchmark, including benchmark records and group assignments, primary performance metrics, domain-contribution analyses across the A vs C, A vs B, and class-balanced A vs B+C contrasts, associated bootstrap comparisons, and target-stratified results. Provided as: Table_S13A_ChEMBL_activity_status_records.csv, Table_S13B_ChEMBL_activity_status_performance.csv, Table_S13C_ChEMBL_domain_contribution.csv, and Table_S13D_ChEMBL_target_stratified_metrics.csv.

Supplementary Table S14. Descriptor-compatible MPO and drug-likeness comparator benchmark.

This table family reports shared-descriptor benchmark results for the descriptor-compatible SWAN formulation, CNS-MPO-inspired oncology proxy, LE-QED composite proxy, QED, and QEPPI-inspired proxy, including ROC-AUC, BEDROC $\alpha = 20$, EF@1%, EF@5%, pairwise

DeLong comparisons, and compound-level scores. The adapted comparator implementations were evaluated in a common descriptor space to support direct comparison under shared benchmark inputs. Provided as: Table_S14A_MPO_comparator_metrics.csv, Table_S14B_MPO_comparator_DeLong.csv, and Table_S14C_MPO_comparator_compound_level_scores.csv.

Supplementary Table S15. Curated retrospective benchmark scores.

This table reports the corrected panel-specific scores for the 15 primary boundary-stress compounds across the colorectal, prostate, and RCC panels, including benchmark group, selected BestNode target, BestNode binding desirability, safety, ADME, raw and curated liability desirabilities, raw and curated SWAN-BestNode scores, and corresponding ranks. Provided as: Table_S15_curated_retrospective_scores.csv.

Supplementary Table S16. Curated-liability ablation summary.

This table reports one row per curated-liability ablation variant across the frozen 15-compound primary boundary-stress benchmark, including ablation scope, affected compound where applicable, ROC-AUC, BEDROC $\alpha = 20$, score range, and metric basis. Provided as: Table_S16_curated_liability_ablation_summary.csv.

Supplementary Table S17. Ligand-efficiency sensitivity analysis.

This table family reports internal and expanded-library ligand-efficiency sensitivity analyses. Table S17A reports internal LE-adjusted SWAN-MPO sensitivity scores and ranks, Table S17B reports internal target-level ligand-efficiency values, Table S17C reports expanded-library LE-adjusted SWAN-MPO scores and rank shifts, and Table S17D reports expanded-library target-level ligand-efficiency values. Across the expanded library, LE-adjusted rankings remained

highly concordant with primary SWAN-BestNode rankings (Kendall's τb = 0.804–0.813); Muricatacin retained rank 1 in colorectal cancer and shifted from rank 2 to rank 1 in prostate cancer and RCC. Provided as: Table_S17A_internal_LE_adjusted_SWAN_sensitivity.csv, Table_S17B_internal_target_level_ligand_efficiency.csv, Table_S17C_expanded_LE_adjusted_SWAN_sensitivity.csv, and Table_S17D_expanded_target_level_ligand_efficiency.csv.

Supplementary Table S18. Strict-pass-only sensitivity analysis.

This table family reports strict-top-pose sensitivity results. Table S18A reports internal compound-level strict-pass scores and ranks, Table S18B reports expanded-library strict-top-pose compound scores and rank shifts, and Table S18C reports panel-level strict-target composition, estimability, and full-library rank concordance. Provided as: Table_S18A_strict_pass_only_internal_scores.csv, Table_S18B_expanded_strict_top_pose_sensitivity_scores.csv, and Table_S18C_expanded_strict_top_pose_sensitivity_summary.csv.

Supplementary Table S19. Monte Carlo rank-stability analysis.

This table family reports Monte Carlo domain-weight perturbation outputs. Table S19A reports internal all-compound rank-stability summaries, Table S19B reports expanded-library all-compound rank-stability summaries, Table S19C reports the expanded-library sampled weight vectors, and Table S19D preserves the expanded-library Muricatacin iteration-level audit. Under broad expanded-library weight perturbation, Muricatacin retained median ranks of 1, 2, and 2 in colorectal, prostate, and RCC, respectively; top-three probabilities were 0.8751, 0.5709, and 0.5709, indicating greater weight sensitivity in the prostate and RCC rankings. Provided as: Table_S19A_monte_carlo_rank_stability_all_compounds.csv,

Table_S19B_expanded_monte_carlo_weight_stability_summary.csv,
Table_S19C_expanded_monte_carlo_weight_draws.csv, and
Table_S19D_expanded_monte_carlo_muricatacin_iterations.csv.

Supplementary Table S20. Retrospective descriptor-space PCA outputs.

Table S20A reports PCA coordinates and descriptor values for the frozen 15-compound primary boundary-stress benchmark, Table S20B reports descriptor loadings, and Table S20C reports explained variance. Provided as: Table_S20A_retrospective_PCA_coordinates.csv, Table_S20B_retrospective_PCA_loadings.csv, and Table_S20C_retrospective_PCA_explained_variance.csv.

Supplementary Table S21. Retrospective binding-saturation diagnostics.

This table family reports the corrected panel-specific retrospective binding-saturation analysis. Table S21A reports target-level reference-calibrated binding and diagnostic-formula audit fields, Table S21B reports panel-specific compound-level primary and binding-sensitivity scores and rank shifts, Table S21C reports panel-level performance and rank-concordance summaries, and Table S21D reports panel-level binding-saturation diagnostics. Provided as:
Table_S21A_retrospective_target_level_binding_saturation.csv,
Table_S21B_retrospective_compound_level_binding_saturation.csv,
Table_S21C_retrospective_binding_saturation_summary.csv, and
Table_S21D_retrospective_binding_saturation_diagnostics.csv.

Supplementary Table S22. COCONUT property-only global scan.

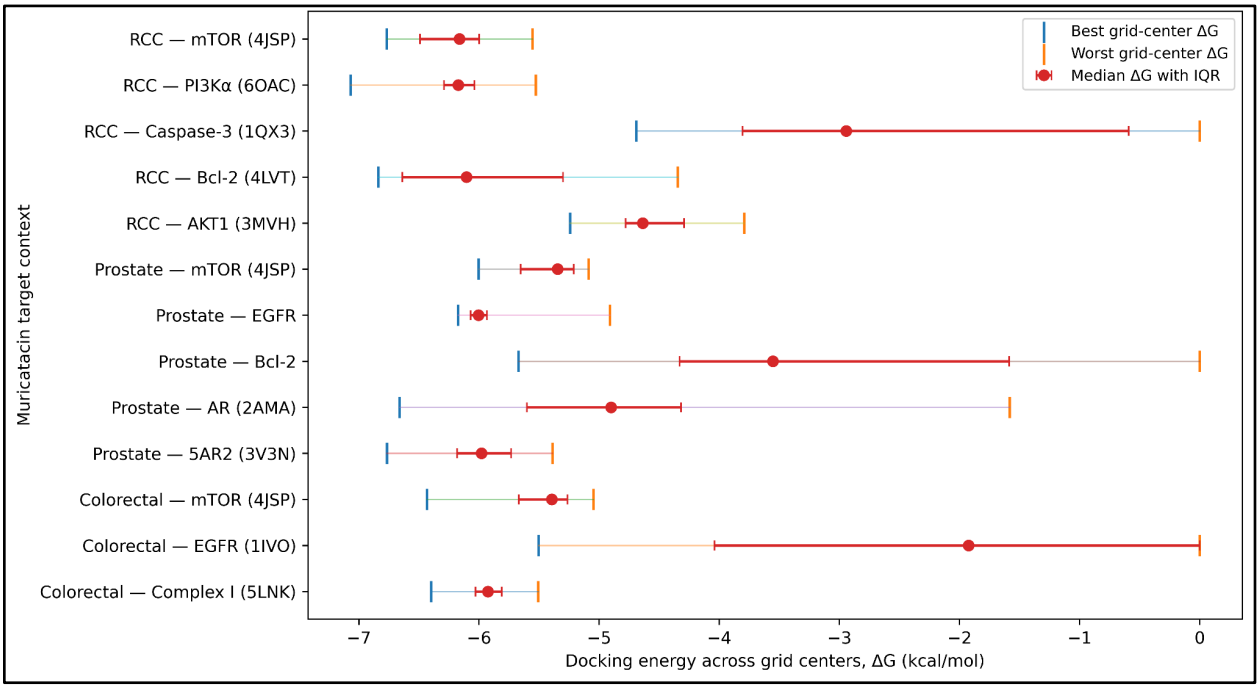
Table S22A reports Muricatacin's full-precision and manuscript-reported property-only desirability, global rank, percentile, and total scored universe. Table S22B reports the nearest

compounds in the COCONUT PCA representation, including PC1/PC2 coordinates and distance from Muricatacin, and Table S22C reports the PCA scaler parameters and PC1/PC2 descriptor loadings. Provided as: Table_S22A_COCONUT_muricatacin_global_position.csv, Table_S22B_COCONUT_nearest_neighbors.csv, and Table_S22C_COCONUT_PCA_model.csv.

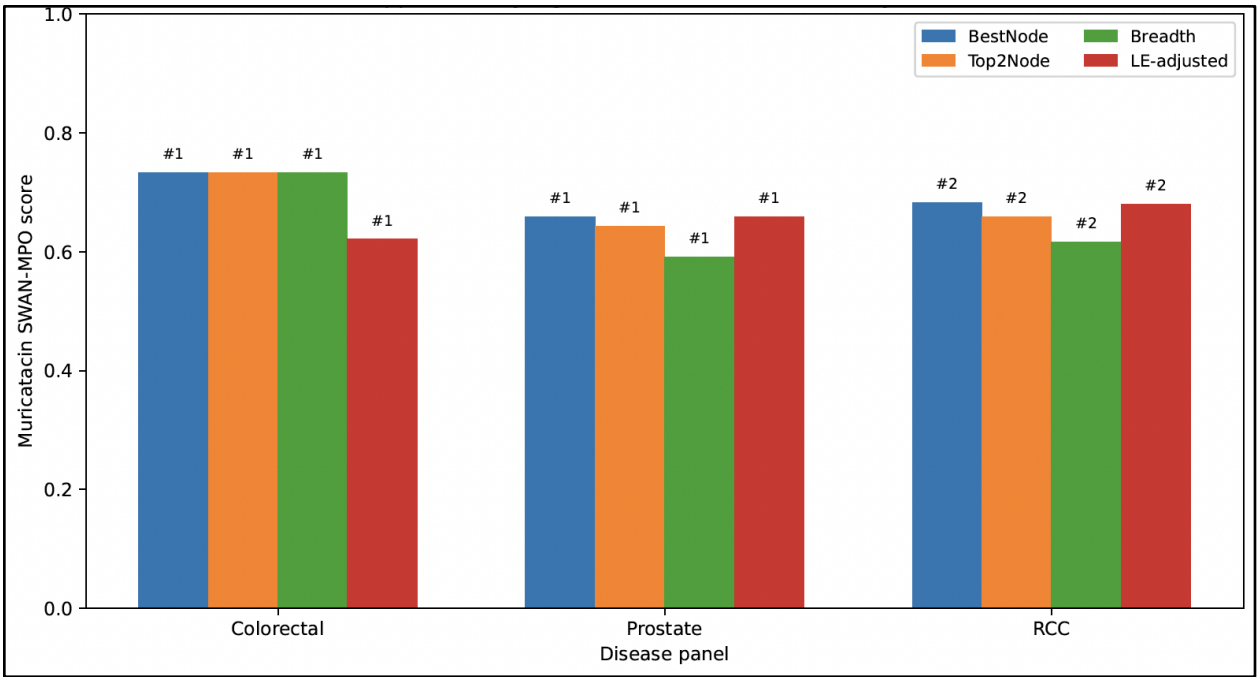
Supplementary Table S23. Descriptor-transformation, missing-value, and aggregation parameter dictionary.

This table reports the transformation direction, preferred and tolerated ranges, categorical mappings, desirability bounds, missing-value rules, and aggregation procedures used for the primary SWAN-MPO domains, binding sensitivities, curated-liability sensitivity, and COCONUT property-only contextualization. Primary and sensitivity-analysis parameters are identified separately.
Provided as: Table_S23_parameter_dictionary.csv.

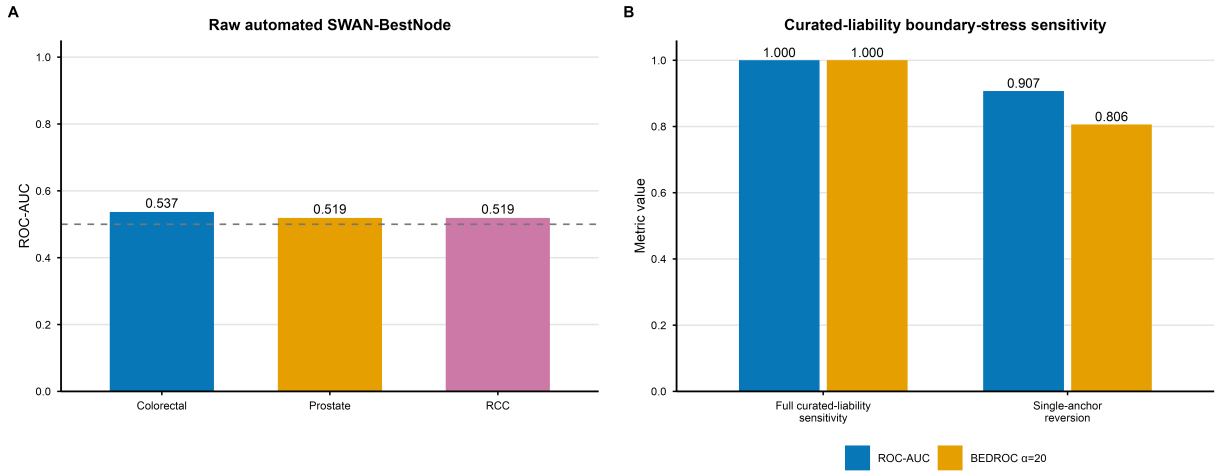
FIGURES



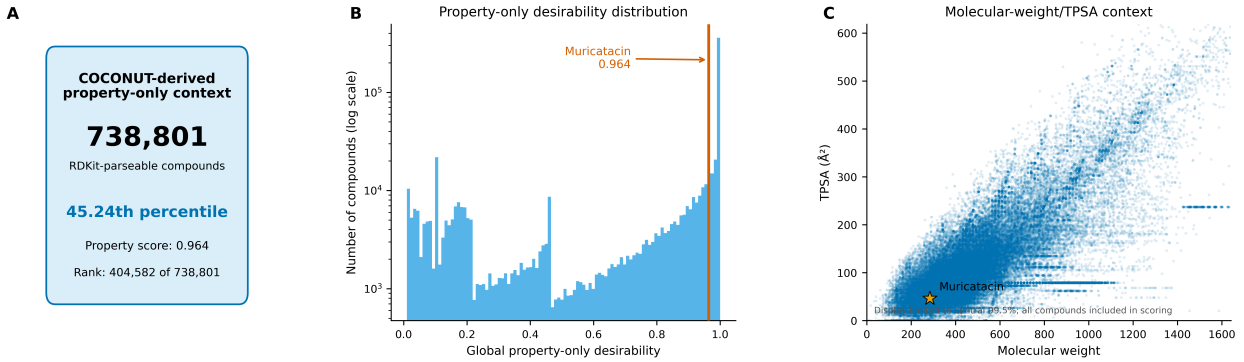
Supplementary Fig. S1 Grid-center docking sensitivity
Muricatacin docking outputs were summarized across grid centers for each target context. Median ΔG values are shown with interquartile ranges, while best and worst grid-center ΔG values indicate the observed grid-center spread. This analysis supports use of the median docking statistic by showing the target-level docking distribution rather than relying on a single favorable grid-center result. Muricatacin's internal binding rank was unchanged when target-level docking was summarized using median rather than best grid-center ΔG, remaining rank 3 in colorectal, rank 8 in prostate, and rank 7 in RCC



Supplementary Fig. S2 SWAN-MPO sensitivity scores
Muricatacin SWAN-MPO scores are shown across the internal rerun using BestNode, Top2Node, Breadth, and ligand-efficiency-adjusted binding formulations. Rank labels indicate the within-panel rank produced by each formulation. The preserved rank pattern across sensitivity formulations shows that Muricatacin's prioritization remained stable across alternative binding-aggregation and ligand-efficiency treatments



Supplementary Fig. S3 Retrospective boundary-stress performance with curated-liability sensitivity
(a) Raw automated SWAN-BestNode showed little discrimination between six favorable comparators and nine documented liability anchors across colorectal, prostate, and RCC panels. (b) Literature-curated liability annotations produced complete separation in the constructed cohort, while single-anchor reversions reduced performance.



Supplementary Fig. S4 Global natural-product property-space contextualizes framework prioritization

(a) Muricatacin's property-only desirability and global position among 738,801 RDKit-parseable COCONUT compounds. (b) Distribution of property-only desirability scores with Muricatacin indicated. (c) Molecular-weight/TPSA context within the broader natural-product chemical space. Muricatacin's 45.24th-percentile placement indicates that its high integrated panel ranking was not explained by exceptional generic physicochemical favorability alone