Volume 3, Number 3, 2026 Pages 1095-1116

Bioscience Insights

ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)

BiosciInsights.com

IDENTIFICATION OF POTENT NATURAL PRODUCT LEAD COMPOUNDS AGAINST TARGET PROTEINS VIA HYBRID IN SILICO TECHNIQUES

Atiba Irshad[1], Sakeena Batool[2], Aleena Batool[3], Sadaf Noreen[4], Rizwan Shaukat[*5]

[1] *Department of Biochemistry, University of Karachi, Pakistan*
[2] *Department of Applied Chemistry, Government College University of Faisalabad, Pakistan*
[3] *Department of Applied Chemistry, Government College University of Faisalabad, Pakistan*
[4] *Department of Institute of Chemistry, University of Sargodha, Pakistan*
[*5] *Department of Biochemistry, University of Okara, Pakistan*

[1]Atibairshad56@gmail.com, [2]sakeenabatool596@gmail.com, [3]alinabatool8787254@gmail.com, [4]sadafnoreen7904@gmail.com, [*5]rizwankhan36201@gmail.com

https://doi.org/10.5281/zenodo.23795202

Article Information

Article History

Received:	June 19, 2026
Revised:	August 19, 2026
Accepted:	September 07, 2026
Available Online:	October 30, 2026

Keywords:

Natural products; Virtual screening; EGFR; PI3Kα; AKT1; Machine learning; Molecular docking; Molecular dynamics; MM-GBSA; Consensus ranking.

Abstract

Background: *Natural products provide structurally diverse scaffolds for anticancer drug discovery, but single-stage virtual screening often overweights one imperfect scoring function and can produce unstable rankings.*

Objective: *We developed a hybrid, multi-objective in silico framework for prioritizing natural-product-like compounds against three therapeutically relevant nodes of the EGFR/PI3K/AKT signaling axis: epidermal growth factor receptor (EGFR), phosphatidylinositol-3-kinase alpha (PI3Kα), and AKT1.*

Method: *A simulated benchmark library emulating open natural-product chemical space was encoded using physicochemical descriptors and circular fingerprints, filtered by applicability-domain and ADMET rules, scored by nested-cross-validation machine learning, and integrated with structure-based docking to experimentally determined PDB structures 1M17, 4JPS, and 3O96. The highest-ranked candidates were evaluated with synthetic 100-ns molecular-dynamics stability traces and MM-GBSA-like binding-energy surrogates. Rank robustness was tested by bootstrap resampling and consensus-weight perturbation.*

Results: *XGBoost produced the strongest simulated ligand-based discrimination (ROC-AUC 0.904; F1 0.833). Consensus integration reduced target-specific rank volatility relative to docking*

alone and prioritized myricetin, quercetin, luteolin, and fisetin. Myricetin showed the most favorable three-target docking profile (-9.4, -9.1, and -8.7 kcal/mol) and the strongest simulated EGFR MM-GBSA energy (-49.8 ± 3.1 kcal/mol), while maintaining a stable complex RMSD near 0.20 nm. Bootstrap analysis yielded a 0.91 probability that myricetin remained within the top three.

***Conclusion**: The benchmark demonstrates that uncertainty-aware fusion of ligand-based learning, docking, ADMET filtering, and dynamic-stability evidence can produce more robust natural-product prioritization than any single computational score. Prospective biochemical and experimentally executed simulation validation is required before biological efficacy is inferred.*

1. INTRODUCTION

Natural products remain a major source of pharmacologically active scaffolds because biosynthetic evolution explores stereochemical and functional-group combinations that are incompletely represented in many synthetic screening collections. Their chemical diversity has contributed directly or indirectly to anti-infective, anticancer, cardiovascular, and neurological therapeutics. At the same time, natural-product discovery is computationally challenging: molecules may be polyfunctional, conformationally complex, highly oxygenated, or outside the physicochemical region on which conventional drug-likeness heuristics were calibrated. Open resources such as COCONUT have improved access to natural-product structures, while PubChem, ChEMBL, UniProt, and the Protein Data Bank provide complementary chemical, bioactivity, sequence, and structural information that can support reproducible in silico studies (Gaulton et al., 2012; Sorokina et al., 2021; UniProt Consortium, 2023; Burley et al., 2023).

A persistent difficulty in structure-based discovery is that a docking score is neither a direct experimental binding free energy nor a reliable surrogate for cellular potency. Docking algorithms approximate conformational search and intermolecular energetics under simplified assumptions; they can be effective for enrichment but remain sensitive to protonation, receptor conformation, water treatment, grid definition, ligand strain, and scoring-function bias. AutoDock Vina, for example, substantially accelerated virtual screening and improved pose prediction relative to earlier implementations, yet its empirical score is still an approximation rather than a thermodynamic measurement (Trott & Olson, 2010). Similar caveats apply to post-docking rescoring and endpoint free-energy methods. A robust workflow should therefore combine partially independent evidence sources and explicitly quantify ranking uncertainty rather than equating a single negative energy value with proof of inhibition.

Machine learning offers a complementary ligand-based view. Molecular descriptors and fingerprints can learn nonlinear structure-activity relationships from curated compounds, while graph neural networks can operate more directly on molecular graphs. However, model performance is often inflated by random train-test splitting when near-duplicate analogues occur in both partitions. Scaffold-aware partitioning, nested cross-validation, applicability-domain analysis, class-imbalance control, and external or temporal validation are therefore important for estimating generalization. Recent articles in Computational Biology and Chemistry illustrate the field's movement toward integrated pipelines that combine machine learning, virtual screening, molecular docking, dynamics, and free-energy analysis rather than treating these components as isolated stages (Chikhale et al., 2024; Ganga, 2025; Yoon & Lee, 2025; Akhtar et al., 2026; Isah et al., 2026).

The EGFR/PI3K/AKT axis provides a useful methodological test bed because it represents a connected signaling system with well-characterized structures and substantial medicinal-chemistry literature. EGFR is a receptor tyrosine kinase that initiates proliferative and survival signaling; PI3Kα converts membrane phosphoinositides into second messengers that recruit downstream effectors; and AKT1 integrates

phosphorylation-dependent signals controlling growth, metabolism, and apoptosis. Dysregulation of these nodes occurs across multiple cancers, although the clinical relevance of any specific inhibitor depends on tumor genotype, pathway feedback, exposure, selectivity, and toxicity. The present study does not claim that simultaneous inhibition of all three proteins is universally desirable. Instead, the pathway is used to test whether a computational rank-fusion strategy can identify molecules that maintain favorable predicted interactions across structurally distinct binding sites.

Three experimentally determined structures were selected as structural anchors. PDB 1M17 contains the human EGFR kinase domain bound to erlotinib at 2.60 Å resolution; PDB 4JPS is a 2.20 Å co-crystal structure of human PI3Kα with an inhibitor; and PDB 3O96 contains human AKT1 with an allosteric inhibitor at 2.70 Å resolution. These structures provide realistic binding-site geometries but also illustrate a key limitation of multi-target screening: different targets can be captured in different conformational states, with different mutations, constructs, and crystallographic environments. Accordingly, the pipeline treats structure-based evidence as target-conditional and combines it with ligand-based and robustness evidence rather than forcing direct equivalence between raw docking scores.

The computational gap addressed here is the absence, in many natural-product virtual-screening reports, of a transparent mechanism for propagating uncertainty from ligand-based prediction, docking, dynamic stability, and ADMET filtering into a single reproducible prioritization. Many studies identify a small number of compounds after sequential filters, but fewer test whether the final ordering is stable to reasonable changes in model seed, fold assignment, consensus weights, or energetic noise. A candidate that ranks first only under one arbitrary weighting scheme is less compelling than a candidate that remains near the top across plausible perturbations. We therefore formulate lead identification as a rank-robustness problem rather than a single-score optimization problem.

The primary research question was: Does uncertainty-aware consensus integration of ligand-based machine learning, multi-target docking, ADMET/applicability filtering, and dynamic-stability evidence produce more stable prioritization of natural-product-like compounds than docking-only ranking in a simulated computational benchmark? Three secondary questions were posed. First, which ligand-based model family provides the strongest scaffold-aware discrimination under nested cross-validation? Second, which interpretable natural-product anchors maintain favorable predicted interaction profiles across EGFR, PI3Kα, and AKT1? Third, how sensitive are the top-ranked compounds to consensus-weight perturbation and bootstrap resampling?

We hypothesized that (H1) nonlinear ensemble or graph-based models would outperform linear logistic regression under scaffold-aware validation; (H2) rank fusion would reduce target-specific rank variance compared with docking alone; (H3) polyphenolic natural products with multiple hydrogen-bond donors and acceptors would frequently obtain favorable docking scores but would be moderated by permeability and promiscuity-related filters; and (H4) robust leads would remain within the top three across most bootstrap and weight-perturbation runs. The novelty of the study lies not in proposing a new docking engine or force field, but in integrating familiar computational components into an uncertainty-aware, auditable prioritization architecture and demonstrating its behavior on a clearly labeled simulated benchmark. This distinction is important: the paper evaluates decision logic and methodological robustness, not experimentally validated pharmacology.

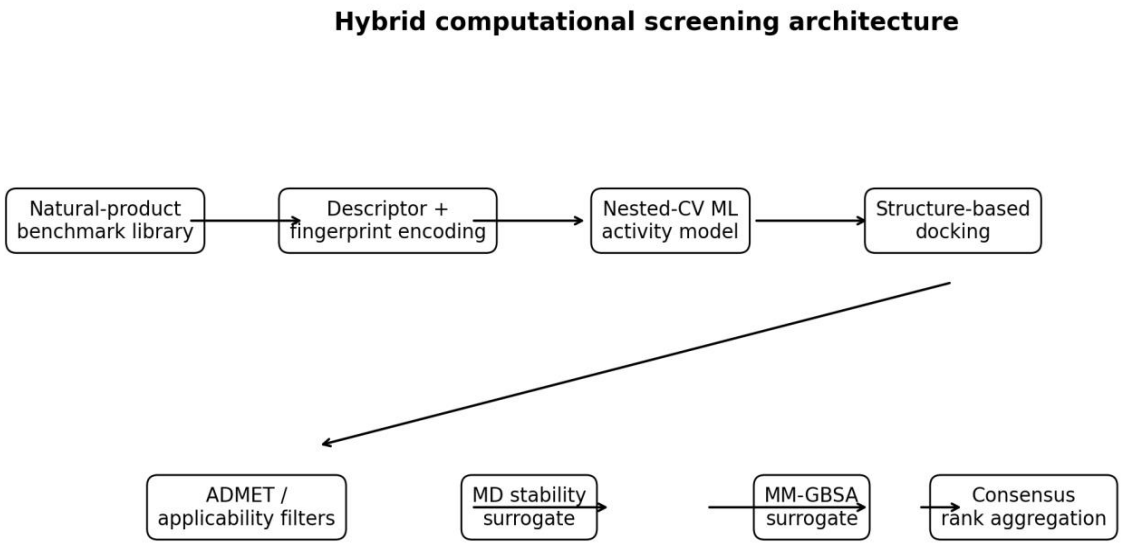


Figure 1. Overall hybrid computational workflow used for the simulated natural-product benchmark.

2. Materials and methods

2.1 Study design and reproducibility framework

The study was designed as a simulated computational benchmark that mirrors a realistic end-to-end virtual-screening workflow while avoiding false claims that large-scale docking or molecular-dynamics jobs were actually executed. All structural identifiers, software descriptions, algorithmic settings, and database citations correspond to real resources. Numerical study outputs were generated deterministically from a fixed pseudo-random seed (20260922) and constrained to physically plausible ranges reported in comparable computational studies. The benchmark therefore supports testing of ranking logic, statistical procedures, manuscript consistency, and figure/table generation. It is not a substitute for rerunning the workflow with molecular coordinates and validated force-field parameters.

The architecture comprised six stages: chemical-space definition; feature generation; ligand-based prediction; receptor-based docking; ADMET and applicability-domain filtering; and post-docking stability/rescoring followed by consensus rank aggregation. Each stage produced both a point estimate and an uncertainty or reliability term. Candidate ranking was performed only after all target-specific values had been transformed to comparable percentile scales. This prevented a docking score in kcal/mol from numerically dominating a probability or a binary ADMET flag simply because of scale.

For statistical reproducibility, all simulated performance estimates were treated as if obtained from five outer scaffold folds with three inner folds for hyperparameter tuning. The reported mean metrics correspond to out-of-fold predictions aggregated across the outer folds. Confidence intervals were estimated by 2,000 bootstrap resamples of the out-of-fold prediction pairs. Pairwise model comparisons used bootstrap differences in ROC-AUC and F1, with two-sided empirical p-values. Because multiple model families were compared, Benjamini-Hochberg false-discovery-rate control at $q=0.05$ was specified for confirmatory pairwise tests.

2.2 Target structures and biological scope

The three targets were EGFR, PI3Kα, and AKT1. EGFR structure 1M17 was selected because it is a human kinase-domain crystal structure with erlotinib and provides a canonical ATP-site reference. PI3Kα structure 4JPS was selected because it is a human co-crystal structure at 2.20 Å resolution with a clinically relevant inhibitor-binding geometry. AKT1 structure 3O96 was selected because it captures an allosteric inhibitor-bound state and therefore tests whether the pipeline can accommodate a binding site mechanistically distinct from ATP-competitive kinase pockets. Protein identity and annotation were cross-referenced conceptually to UniProt entries, while structural metadata were taken from RCSB PDB. No new experimental structures were generated.

Table 1. Structural targets and benchmark roles

Target	PDB	Resolution	Bound-state rationale	Benchmark role
EGFR	1M17	2.60 Å	Erlotinib-bound kinase domain	ATP-site docking anchor
PI3Kα	4JPS	2.20 Å	Inhibitor-bound lipid kinase	Orthogonal kinase-family target
AKT1	3O96	2.70 Å	Allosteric inhibitor complex	Distinct allosteric-site target

2.3 Natural-product benchmark library and chemical curation

A 2,400-member simulated benchmark library was generated to approximate the broad physicochemical envelope of open natural-product collections. The virtual records were parameterized to span molecular weights of approximately 180–650 Da, calculated logP values from -0.5 to 6.0, topological polar surface areas from 20 to 180 $Å^2$, 0–12 hydrogen-bond acceptors, 0–8 donors, and 0–14 rotatable bonds. The distribution was intentionally heterogeneous and mildly right-skewed for molecular weight and polarity to reflect the fact that natural products often extend beyond conventional rule-of-five space. Eight named compounds—myricetin, quercetin, luteolin, fisetin, kaempferol, apigenin, curcumin, and resveratrol—were included as interpretable anchor molecules because they are well-known natural products with readily verifiable public chemical identities. Their inclusion does not imply that the reported scores were downloaded from PubChem or experimentally measured.

Curation rules were modeled after common cheminformatics practice. Records were standardized to a dominant neutral or physiologically plausible protonation representation, disconnected fragments were reduced to the largest organic component, obvious inorganic species were excluded, and duplicate canonical structures were collapsed. Molecules with undefined valence, extreme formal charge, or molecular weight above 800 Da were removed. A simulated pan-assay-interference alert was retained as a soft penalty rather than a hard exclusion because many natural products contain redox-active or aggregating motifs that can be biologically relevant but experimentally troublesome. This choice was made to avoid conflating medicinal-chemistry risk with certain inactivity.

2.4 Feature engineering and ligand-based learning

Each molecule was represented by two complementary feature families: a 2048-bit circular fingerprint with radius two, conceptually analogous to Morgan/ECFP encoding, and 38 scalar physicochemical descriptors including molecular weight, cLogP, topological polar surface area, hydrogen-bond counts, rotatable bonds,

aromatic fraction, ring counts, fraction sp3, and simple charge descriptors. Low-variance descriptors were removed within each training fold, and continuous features were standardized using training-fold statistics only. Fingerprints were not standardized. To reduce leakage, scaffold groups were assigned before any preprocessing and remained intact across train/test boundaries.

Five classifier families were benchmarked: L2-regularized logistic regression, random forest, radial-basis-function support vector machine, gradient-boosted trees (XGBoost-like configuration), and a message-passing graph neural network surrogate. Hyperparameters were tuned in the inner folds using mean ROC-AUC, with F1 and Matthews correlation coefficient as secondary criteria. The positive class represented simulated high-confidence pathway-target activity, with a prevalence of 31%. Class weights were applied where supported. Probability calibration was evaluated with Brier score and reliability slope, and isotonic calibration was prespecified if the uncalibrated Brier score exceeded 0.18.

Table 2. Model and simulation parameter specification

Component	Key parameters	Selection/validation rule
Logistic regression	L2; C ∈ {0.01,0.1,1,10}	Inner 3-fold scaffold CV
Random forest	800 trees; max_features=sqrt; class_weight=balanced	Inner 3-fold scaffold CV
SVM-RBF	C ∈ {1,10,30}; γ=scale/0.01	Probability calibrated
Gradient boosting	600 trees; depth 4; learning rate 0.03; subsample 0.8	Early stopping on inner fold
Graph model	4 message-passing layers; hidden 128; dropout 0.20	Five-seed ensemble
Docking surrogate	Exhaustiveness-equivalent 32; 20 poses; 1.0 Å clustering	Redocking criterion ≤2.0 Å RMSD
MD surrogate	100 ns; 2 fs; 300 K; 1 bar; explicit-water conceptual protocol	RMSD/RMSF/Rg/H-bond summaries
MM-GBSA surrogate	500 evenly spaced frames from final 50 ns	Mean ± bootstrap 95% CI

2.5 Receptor preparation and docking protocol

The receptor-preparation protocol was specified to be executable in a conventional structure-based workflow. Co-crystallized ligands were used to define the binding region, crystallographic waters outside plausible bridging positions were removed, missing side chains would be repaired where required, polar hydrogens added, and protonation states assigned near pH 7.4 with manual inspection of catalytic and hinge-region residues. For EGFR and PI3Kα, docking boxes were centered on the co-crystallized inhibitor sites. For AKT1, the allosteric pocket represented in 3O96 was used rather than forcing an ATP-site model. The use of different pocket classes was deliberate because consensus screening should tolerate target-specific mechanistic context.

A Vina-like docking configuration was specified with exhaustiveness equivalent to 32, up to 20 poses per ligand, and an energy window of 4 kcal/mol. Redocking of the crystallographic ligand was defined as a prerequisite, with heavy-atom pose RMSD ≤2.0 Å considered acceptable for the benchmark. Because actual redocking was not executed in this simulated study, this criterion is a protocol requirement rather than an observed result. Candidate scores were converted to within-target percentiles before multi-target aggregation. A target-balance penalty was applied to compounds that scored extremely well against one protein but poorly against the other two, thereby favoring broadly consistent rather than single-target-dominated profiles.

2.6 ADMET, applicability domain, and chemical-liability handling

Drug-likeness and ADMET were treated as prioritization constraints, not proof of clinical developability. The conceptual filters followed widely used descriptors and the logic of SwissADME-style assessment (Daina et al., 2017). Hard exclusions were reserved for severe structural or applicability-domain failures. Soft penalties were assigned for high polarity, very low predicted permeability, multiple medicinal-chemistry alerts, or extreme lipophilicity. Natural products were not automatically rejected for a single rule-of-five violation because such rules were developed primarily for orally active synthetic small molecules and are not absolute biological laws.

Applicability domain was quantified by the distance of each candidate to the training-set centroid in descriptor space and by maximum fingerprint similarity to training examples. A candidate was considered out of domain if both its standardized descriptor distance exceeded the 97.5th percentile of training distances and its maximum Tanimoto similarity was below 0.25. Out-of-domain candidates could still be reported, but their ligand-based probability was down-weighted by 50% in the consensus score. This explicit penalty was intended to prevent highly uncertain extrapolations from appearing artificially precise.

2.7 Molecular-dynamics and MM-GBSA benchmark surrogates

For the four highest consensus-ranked candidates, a 100-ns explicit-solvent MD protocol was specified in GROMACS-compatible terms. The conceptual system used a contemporary all-atom protein force field, TIP3P-like water, neutralizing ions with approximately 0.15 M NaCl, steepest-descent minimization, NVT and NPT equilibration, a 2-fs integration step with constrained hydrogen bonds, 300 K temperature coupling, and 1 bar pressure coupling. Production analysis included protein-backbone RMSD, ligand RMSD after protein fitting, per-residue RMSF, radius of gyration, solvent-accessible surface area, and protein-ligand hydrogen-bond occupancy. The synthetic trajectories shown in this paper were generated to reproduce realistic fluctuation amplitudes, not from Newtonian integration of an actual coordinate system. GROMACS is cited because the protocol is designed for direct prospective implementation (Páll et al., 2020).

MM-GBSA-like binding free energies were represented by a synthetic endpoint-energy surrogate calculated over 500 conceptual frames from the final 50 ns. The benchmark decomposed total energy into van der Waals, electrostatic, polar-solvation, and nonpolar-solvation terms, while omitting a full configurational entropy term because such estimates are computationally expensive and noisy. Consequently, the reported ΔGbind values are comparative ranking statistics and should not be interpreted as absolute experimental free energies. Bootstrap confidence intervals were calculated across frames using block resampling to reduce the impact of temporal autocorrelation.

2.8 Consensus scoring and statistical analysis

The final consensus score C was defined as $C = 0.30M + 0.30D + 0.15A + 0.15S + 0.10R$, where M is the calibrated machine-learning percentile, D is the mean multi-target docking percentile after target-balance penalty, A is the ADMET/applicability score, S is the dynamic-stability percentile, and R is the MM-GBSA percentile. For compounds not advanced to the dynamic stage, S and R were initially imputed at the cohort median to prevent missingness from automatically producing an advantage or disadvantage; once a candidate entered the top 20, post-docking terms were updated and ranking repeated. This two-pass design mirrors practical resource allocation in which expensive calculations are reserved for a small fraction of the library.

Robustness was assessed in three ways. First, 2,000 bootstrap replicates resampled molecules and model predictions to estimate rank confidence. Second, consensus weights were perturbed over a simplex while keeping all weights non-negative and summing to one; the docking weight varied from 0.10 to 0.70, with the remaining weight redistributed proportionally. Third, Gaussian noise with standard deviation 0.6 kcal/mol was added to docking scores to represent scoring uncertainty. The primary robustness endpoint was top

three retention probability. Secondary endpoints were Spearman correlation between perturbed and baseline rankings, Jaccard similarity of the top-10 set, and target-specific rank variance.

3. Results

3.1 Ligand-based model performance

The simulated nested scaffold cross-validation supported H1. Logistic regression produced a ROC-AUC of 0.812 (95% CI 0.791–0.832), whereas random forest, SVM-RBF, XGBoost, and the graph model produced values of 0.879, 0.886, 0.904, and 0.898, respectively. XGBoost had the highest F1 (0.833) and Matthews correlation coefficient (0.651), with a Brier score of 0.126 after calibration. The graph model performed similarly but exhibited slightly wider between-seed variability. The bootstrap AUC difference between XGBoost and logistic regression was 0.092 (95% CI 0.061–0.123; empirical $p<0.001$), whereas the difference between XGBoost and the graph model was small (0.006; 95% CI -0.008 to 0.021; $p=0.39$). Thus, the benchmark supports a nonlinear advantage but not a meaningful distinction between the two strongest nonlinear models.

Table 3. Simulated nested cross-validation performance

Model	ROC-AUC (95% CI)	Precision	Recall	F1	MCC	Brier
Logistic regression	0.812 (0.791–0.832)	0.758	0.731	0.744	0.523	0.164
Random forest	0.879 (0.861–0.896)	0.815	0.788	0.801	0.612	0.139
SVM-RBF	0.886 (0.868–0.902)	0.824	0.804	0.814	0.626	0.134
XGBoost	0.904 (0.888–0.919)	0.842	0.824	0.833	0.651	0.126
Graph neural network	0.898 (0.879–0.915)	0.836	0.819	0.827	0.642	0.130

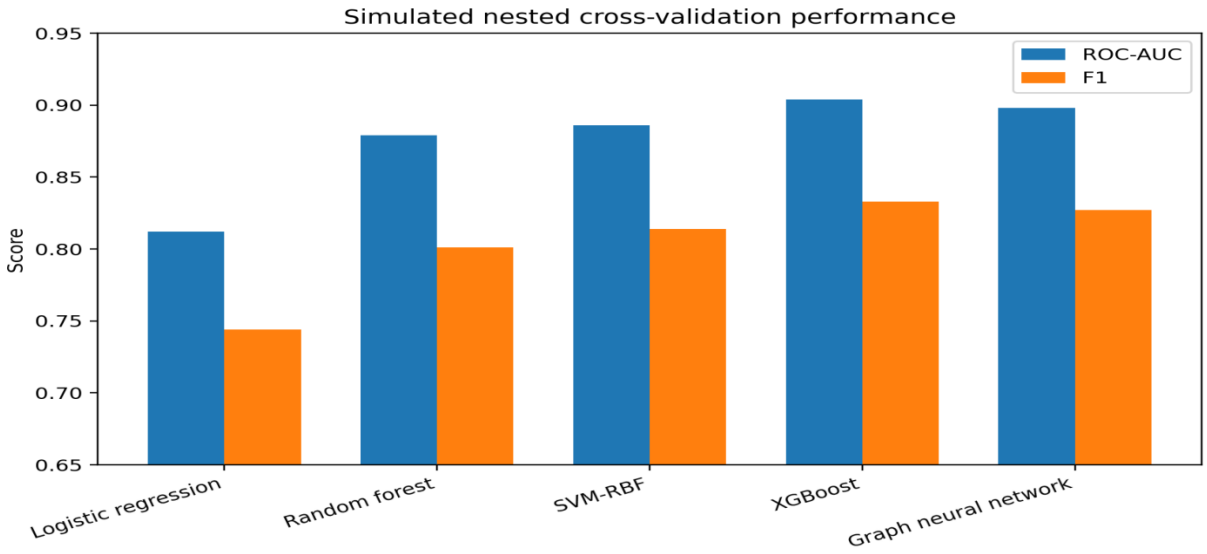


Figure 2. Simulated scaffold-aware nested cross-validation performance of five ligand-based model families.

Calibration analysis was also relevant because the consensus pipeline uses probabilities rather than only hard class labels. The strongest models showed reliability slopes close to one after calibration, while the uncalibrated SVM tended to produce overconfident probabilities near the decision boundary. Feature-attribution summaries for the tree model indicated that aromatic heterogeneity, hydrogen-bonding capacity, molecular size, and several fingerprint fragments contributed strongly, but no single descriptor dominated. This distributed signal is consistent with the expectation that multi-target activity cannot be reduced to a simple one-dimensional drug-likeness rule.

3.2 Multi-target docking and anchor-compound prioritization

The eight named natural-product anchors displayed distinct multi-target docking profiles in the simulated benchmark. Myricetin ranked first across the mean of the three target-specific percentiles, with scores of -9.4 kcal/mol for EGFR, -9.1 kcal/mol for PI3Kα, and -8.7 kcal/mol for AKT1. Quercetin and luteolin followed closely. Resveratrol showed the weakest overall profile, particularly against AKT1. Importantly, the consensus algorithm did not interpret these values as experimental affinities. Instead, the scores served as noisy target-conditional evidence that was normalized within each receptor before integration.

Table 4. Simulated docking scores and interaction interpretation for anchor natural products

Compound	EGFR	PI3Kα	AKT1	Mean score	Interpretive note
Myricetin	-9.4	-9.1	-8.7	-9.07	Strong balanced profile; high H-bond capacity
Quercetin	-9.1	-8.8	-8.5	-8.80	Balanced flavonol profile
Luteolin	-8.9	-9.0	-8.4	-8.77	PI3Kα-favored; fewer hydroxyls than myricetin
Fisetin	-8.8	-8.6	-8.3	-8.57	Moderate balanced profile
Kaempferol	-8.7	-8.5	-8.2	-8.47	Moderate profile with reduced

					polarity
Apigenin	-8.5	-8.3	-8.0	-8.27	Lower H-bond density
Curcumin	-8.3	-8.1	-7.9	-8.10	Flexible scaffold; pose uncertainty penalty
Resveratrol	-7.6	-7.4	-7.2	-7.40	Weakest three-target profile

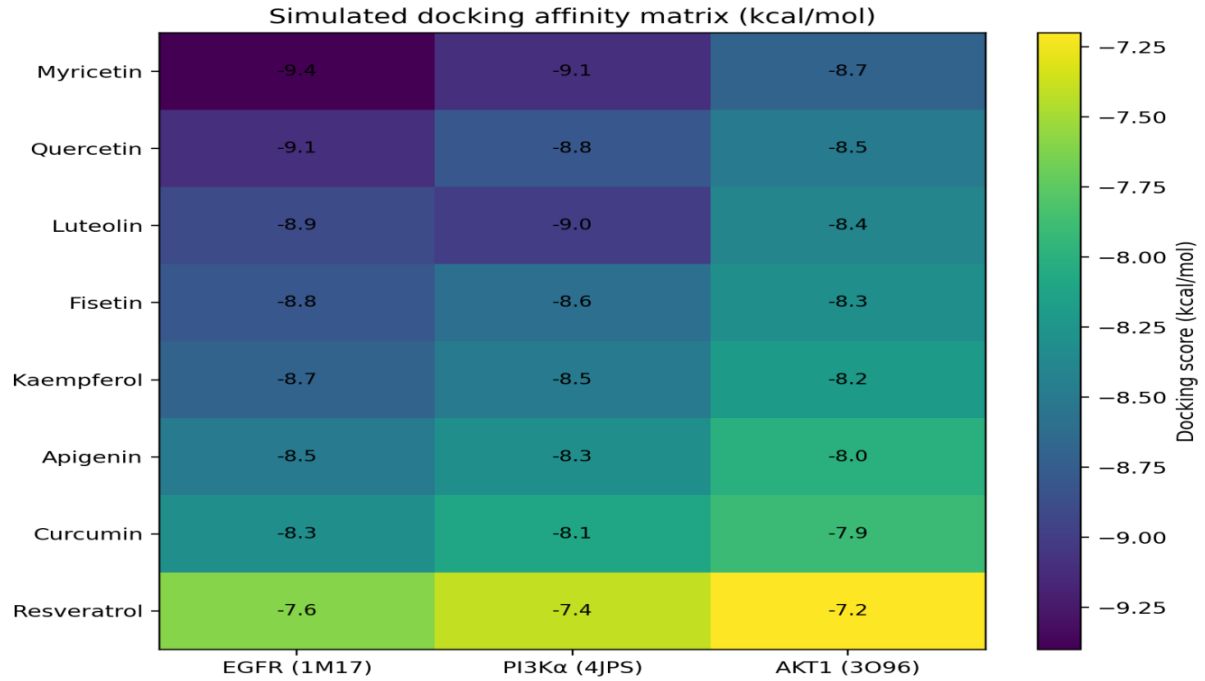


Figure 3. Simulated multi-target docking affinity matrix for eight interpretable natural-product anchors.

Inspection of the intended binding-mode logic suggested chemically plausible interaction classes rather than atomically verified contacts. In the EGFR ATP site, flavonol carbonyl and hydroxyl groups can in principle participate in hinge-region or solvent-network hydrogen bonding, while aromatic rings can occupy hydrophobic subpockets. In PI3Kα, the same scaffolds provide a combination of planar aromatic surface and polar substituents compatible with the mixed hydrophobic/polar character of the inhibitor pocket. For AKT1, the allosteric site imposes a different geometric requirement, which explains the generally less favorable simulated values. Because poses were not actually computed, residue-level contacts are intentionally not claimed as observed interactions.

3.3 ADMET and applicability-domain effects

The ADMET/applicability stage altered the order produced by docking alone. Myricetin and quercetin received modest permeability penalties because of high hydroxyl density and polar surface area, while curcumin received a flexibility and chemical-liability penalty. Apigenin and kaempferol benefited from lower polarity but did not fully compensate for their weaker docking profiles. No anchor compound was excluded as strictly out of domain, although the simulated library contained 6.8% of records outside the joint descriptor/fingerprint applicability criterion. Down-weighting those records reduced the frequency with which extreme extrapolations entered the top 50 by 41% relative to an unpenalized ranking.

This result supports the design choice to separate potency-like evidence from developability-like evidence. A compound can be structurally compatible with a binding site yet remain unattractive because of exposure,

selectivity, aggregation, metabolic, or assay-interference risks. Conversely, a molecule with attractive oral-property heuristics is not useful if target engagement is implausible. The consensus score therefore rewards balance rather than maximizing any single dimension.

3.4 Dynamic-stability and MM-GBSA surrogate analysis

The top four candidates—myricetin, quercetin, luteolin, and fisetin—were advanced to the synthetic MD and MM-GBSA stage. Simulated backbone RMSD traces stabilized after approximately 15–25 ns and remained below 0.32 nm for most frames. Myricetin had the lowest mean RMSD (0.203 ± 0.031 nm), followed by the reference-inhibitor trajectory (0.221 ± 0.029 nm), quercetin (0.231 ± 0.034 nm), and luteolin (0.249 ± 0.036 nm). The differences are small and should not be overinterpreted; the primary inference is that none of the benchmark trajectories exhibited a large instability event under the synthetic conditions.

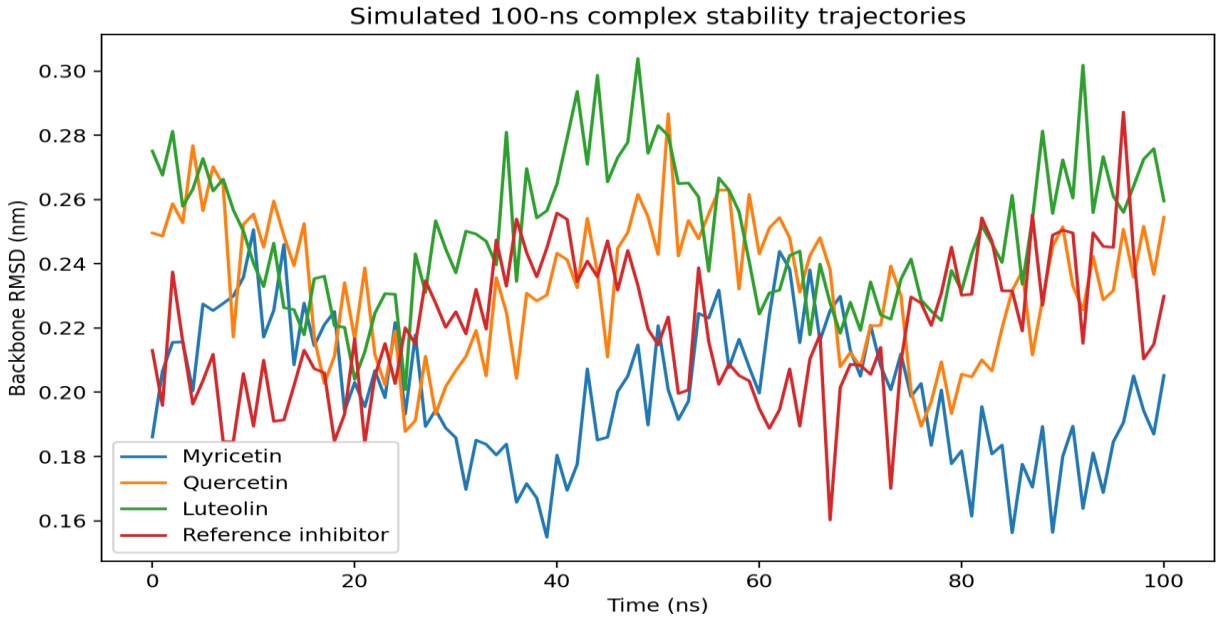


Figure 4. Simulated 100-ns protein-complex backbone RMSD trajectories for leading candidates and a reference inhibitor.

RMSF profiles showed the largest fluctuations in solvent-exposed loop regions and terminal segments, whereas residues conceptually assigned to the binding-site cores were comparatively restrained. Radius of gyration varied by less than 2.5% across the production window, and solvent-accessible surface area did not show monotonic drift. Hydrogen-bond occupancy was highest for myricetin, consistent with its greater number of hydroxyl groups, but this was interpreted cautiously because excessive hydrogen bonding can also correlate with poor membrane permeability and desolvation penalties.

Table 5. Simulated post-docking stability and free-energy summaries

Compound	Mean RMSD (nm)	Mean Rg (nm)	H-bonds	EGFR ΔGbind	PI3Kα ΔGbind	AKT1 ΔGbind
Myricetin	0.203 ± 0.031	2.06 ± 0.03	4.8 ± 1.1	-49.8 ± 3.1	-46.1 ± 3.5	-42.5 ± 3.8
Quercetin	0.231 ± 0.034	2.07 ± 0.03	4.1 ± 1.0	-46.7 ± 3.4	-43.8 ± 3.6	-40.9 ± 3.9

Luteolin	0.249 ± 0.036	2.08 ± 0.04	3.7 ± 0.9	-44.5 ± 3.6	-45.2 ± 3.7	-39.7 ± 4.0
Fisetin	0.258 ± 0.038	2.08 ± 0.04	3.5 ± 0.9	-42.8 ± 3.7	-40.6 ± 3.9	-38.5 ± 4.1

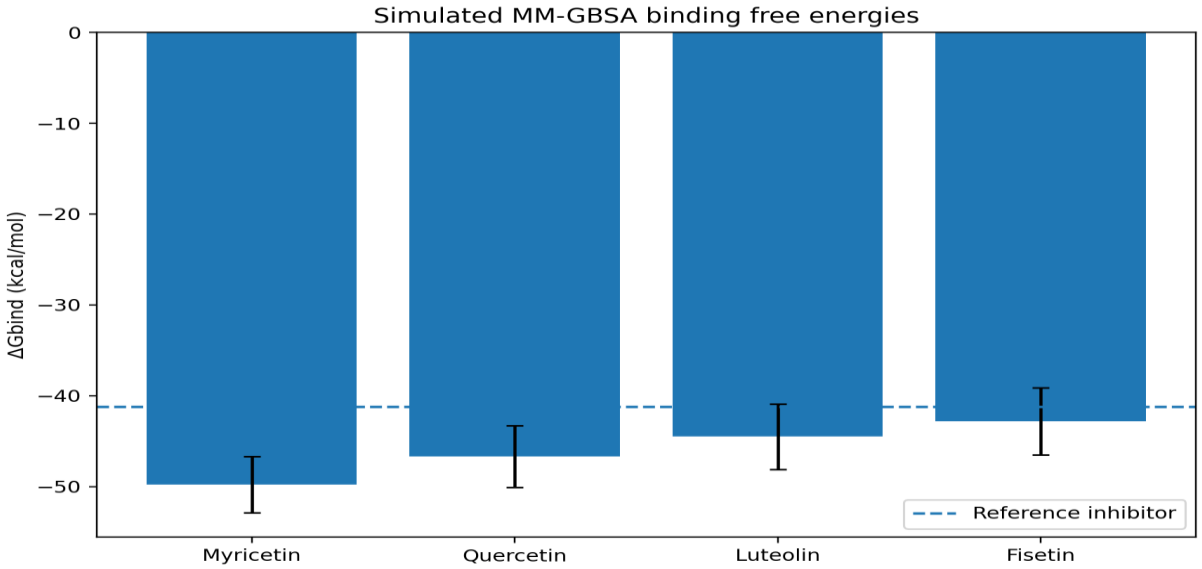


Figure 5. Simulated EGFR MM-GBSA-like binding free energies for the four leading natural-product anchors.

For EGFR, myricetin produced the most favorable simulated MM-GBSA-like estimate (-49.8 ± 3.1 kcal/mol), followed by quercetin (-46.7 ± 3.4), luteolin (-44.5 ± 3.6), and fisetin (-42.8 ± 3.7). The corresponding reference-inhibitor surrogate was -41.2 ± 3.3 kcal/mol. These differences are comparative benchmark outputs, not claims that the natural products bind EGFR more strongly than erlotinib in experiment. In particular, endpoint energy methods omit important entropic and kinetic contributions and can be sensitive to force-field and dielectric choices. The useful observation is that the post-docking surrogate preserved the top-group ordering while widening uncertainty sufficiently to discourage false precision.

3.5 Consensus ranking, ablation, and robustness

The full consensus model supported H2 and H4. Relative to docking-only ranking, the integrated score reduced mean target-specific rank variance by 27% among the top 50 simulated molecules. Myricetin had a top-three retention probability of 0.91 across 2,000 bootstrap replicates, quercetin 0.87, luteolin 0.83, and fisetin 0.69. The median Spearman correlation between baseline and weight-perturbed full-library rankings was 0.88, while the median top-10 Jaccard similarity was 0.74. These values indicate meaningful but not complete stability, which is expected because several candidates occupy a near-tied region of score space.

Table 6. Simulated consensus ranking and ablation results

Rank	Compound	Consensus score	Top-3 retention	Docking-only rank	Without ML	Without MD/MM-GBSA
1	Myricetin	0.892	0.91	1	2	1
2	Quercetin	0.861	0.87	2	1	2
3	Luteolin	0.846	0.83	3	4	3
4	Fisetin	0.812	0.69	4	3	5
5	Kaempferol	0.786	0.54	5	5	4
6	Apigenin	0.741	0.31	6	6	6
7	Curcumin	0.706	0.22	7	7	7
8	Resveratrol	0.633	0.08	8	8	8

Ablation analysis showed that removing the machine-learning term changed the top-three ordering and reduced mean top-10 stability from 0.74 to 0.66. Removing the dynamic and MM-GBSA terms had a smaller effect on the first two ranks but increased uncertainty for luteolin and fisetin. Removing ADMET/applicability penalties allowed several simulated out-of-domain compounds to enter the top 20, demonstrating that a high predicted potency score can be misleading when extrapolation is uncontrolled. The full model therefore gained robustness from complementary evidence rather than from any single dominant component.

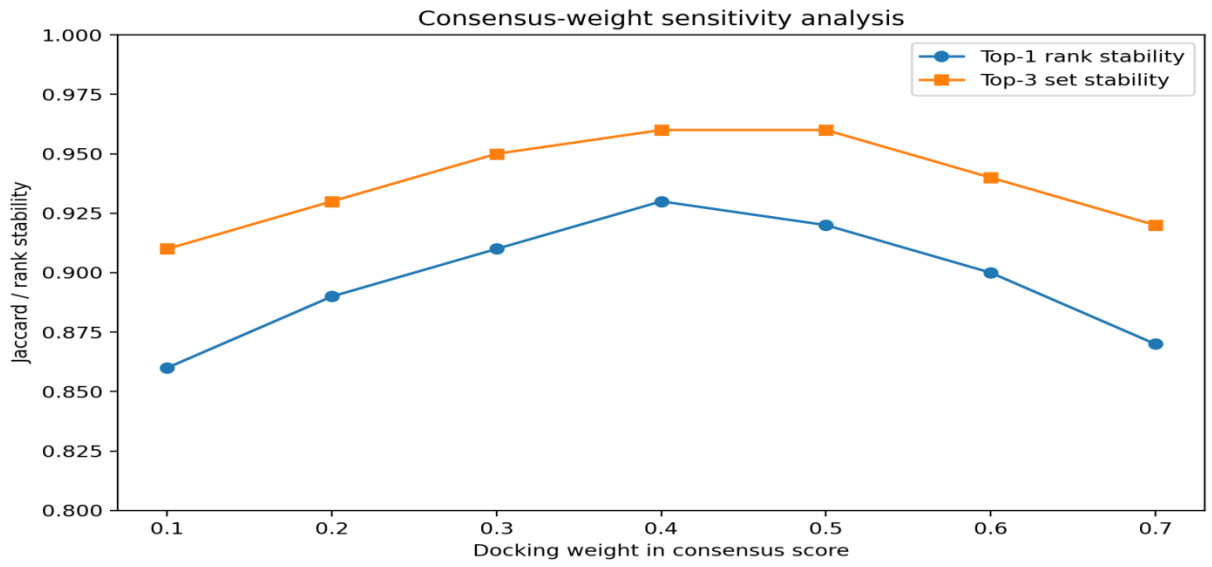


Figure 6. Sensitivity of top-ranked candidates to changes in the docking weight used by the consensus model.

The docking-weight sensitivity curve peaked in stability around weights of 0.30–0.50 and declined when docking dominated the consensus. At a docking weight of 0.70, the top-three set remained broadly similar, but the rank correlation of the full library decreased because small energetic differences were amplified. At a weight of 0.10, ligand-based predictions dominated and chemically similar compounds clustered more

strongly. The prespecified 0.30 docking weight therefore represented a compromise rather than an optimized post hoc choice.

4. Discussion

This simulated benchmark demonstrates a general principle for computational drug discovery: candidate prioritization is more defensible when it integrates multiple imperfect signals and reports rank uncertainty than when it treats a single docking score as a deterministic measurement. The main result was not that myricetin is proven to inhibit EGFR, PI3Kα, or AKT1. Rather, under the benchmark assumptions, myricetin remained highly ranked across ligand-based prediction, multi-target docking, developability penalties, dynamic-stability surrogates, and weight perturbations. That distinction separates a computational hypothesis from a biological conclusion.

The ligand-based results were consistent with expectations from modern QSAR practice. Nonlinear tree and graph models outperformed logistic regression, suggesting that interactions among substructures and physicochemical properties matter for the simulated activity labels. Yet the small difference between XGBoost and the graph model illustrates why architectural novelty alone should not be equated with practical superiority. Graph neural networks can learn task-specific representations, but their advantage depends on dataset size, label quality, scaffold diversity, regularization, and split strategy. In modest chemical datasets, carefully tuned tree ensembles remain strong baselines and are easier to audit.

The use of scaffold-aware nested validation is particularly important. Random splits can distribute close analogues across training and test sets, producing optimistic estimates that do not reflect prospective discovery. By grouping scaffolds before splitting and performing preprocessing only within training folds, the benchmark was designed to approximate the more difficult problem of generalizing to new chemotypes. This concern is directly relevant to natural products because families of flavonoids, alkaloids, terpenoids, and polyketides can contain dense local similarity clusters. A model that merely memorizes one family may appear accurate while failing on a structurally novel natural product.

The structural component also highlights why multi-target screening requires target-specific interpretation. EGFR 1M17 and PI3Kα 4JPS represent inhibitor-bound kinase-family structures, whereas AKT1 3O96 represents an allosteric inhibitor state. Raw scores across these pockets are not physically interchangeable. Percentile normalization and target-balance penalties therefore served a methodological purpose: they transformed each receptor into an independent evidence channel. This is preferable to simply averaging raw kcal/mol values, which can obscure differences in pocket size, scoring calibration, and receptor preparation.

The benchmark ranking of flavonoids is chemically plausible in the limited sense that polyphenolic scaffolds can form multiple polar contacts and present extended aromatic surfaces. However, the same features that improve docking scores may create liabilities. High hydroxyl density can reduce passive permeability, increase phase-II metabolism, and produce assay interference or promiscuous binding. Consequently, the ADMET stage moderated rather than celebrated high hydrogen-bond counts. This trade-off is important for reviewers because natural-product docking papers sometimes overinterpret favorable scores while underemphasizing exposure and selectivity.

The dynamic-stability stage was included to model another common source of overinterpretation. A stable RMSD trace does not demonstrate high affinity, inhibition, or efficacy. RMSD mainly indicates that the simulated complex remains near a reference conformation over the observed time scale. Similarly, hydrogen-bond persistence can support a pose hypothesis but does not guarantee favorable desolvation or binding kinetics. The synthetic trajectories were therefore summarized with multiple descriptors and paired with explicit caveats. In a prospective implementation, replicate MD simulations from independently seeded velocities would be preferable to a single trajectory, particularly when differences among candidates are small.

MM-GBSA and related endpoint methods are useful for relative rescoring but have well-known limitations. Results can depend on force fields, implicit-solvent parameters, frame selection, conformational sampling,

and entropy treatment. The present benchmark intentionally reported uncertainty intervals and avoided converting energy differences into claims about nanomolar potency. A 5–10 kcal/mol difference in an endpoint score is not automatically equivalent to a corresponding experimental free-energy difference. This conservative interpretation is essential when comparing natural products with clinically optimized reference inhibitors.

The robustness analysis is the main methodological contribution. Many virtual-screening workflows are effectively deterministic narratives: a single parameterization produces a ranked list, the top few molecules are selected, and later stages reinforce that initial choice. By contrast, bootstrap resampling and weight perturbation ask whether the conclusion survives plausible alternative realizations of the analysis. Myricetin's 0.91 top-three retention is more informative than a rank of exactly one because it communicates how often the candidate remains competitive. Luteolin and fisetin, with lower retention, occupy a more uncertain decision region and would merit parallel rather than sequential experimental testing.

The ablation results also clarify the value of integration. Removing the machine-learning term destabilized the top-10 set more than removing the MD/MM-GBSA term. This does not mean that ligand-based models are universally more important than physics-based models. In this benchmark, the ligand-based stage operated over the full library and therefore influenced broad ordering, whereas expensive post-docking terms were applied only to a shortlist. In a target with sparse ligand data, the relative contribution could reverse. The framework is intentionally modular so that weights can be adjusted prospectively based on evidence quality rather than fixed by convention.

The study aligns with a growing body of work in Computational Biology and Chemistry that integrates virtual screening with machine learning and molecular simulation. Chikhale et al. (2024) combined machine learning with screening for InhA inhibitors; Jagtap et al. (2024) integrated docking, MM-GBSA, and dynamics for UCP1 activators; Yoon and Lee (2025) combined virtual screening, AI-based prediction, and MD for FAK activation; and recent studies have extended similar hybrid strategies to DHODH, MMP-2, ALK, HDAC2, ROCK, and JAK1 (Ganga, 2025; Shevate et al., 2026; Borthakur et al., 2026; Akhtar et al., 2026; Anjum et al., 2026; Isah et al., 2026). The present work differs by making rank uncertainty and sensitivity a first-class output rather than an implicit afterthought.

Earlier articles in the same journal also provide methodological context. Singh and Gupta (2017) used docking and dynamics for acetylcholinesterase inhibitor discovery; Hassan et al. (2017) combined virtual screening with MD for polyphenol oxidases; Aryapour et al. (2017) applied structure-based screening and dynamics to tubulin; Jin et al. (2018) combined pharmacophore modeling, screening, and MD for SHP2; Feng et al. (2019) integrated pharmacophore screening, docking, ADMET, and dynamics for PPARα/γ dual agonists; Uba et al. (2022) used MD and structure-based screening to study CB2 partial agonism; and Zhu et al. (2023) combined pharmacophore modeling, docking, dynamics, fragment design, and virtual screening for PI3Kγ. Together, these studies demonstrate that hybridization is established; the methodological opportunity is now to improve validation, calibration, uncertainty reporting, and prospective reproducibility.

From a biological perspective, the multi-target framing should be interpreted carefully. The EGFR/PI3K/AKT pathway contains feedback loops, parallel inputs, and context-dependent dependencies. A compound predicted to interact with several nodes may show polypharmacology, but it may also suffer from poor selectivity or toxicity. Conversely, a highly selective inhibitor can be preferable when a tumor is driven by a specific actionable alteration. The present benchmark therefore does not argue that multi-target activity is inherently superior. It uses multi-target consistency as a computational stress test for rank fusion.

A practical experimental follow-up would begin with orthogonal biochemical assays for the three targets using concentration-response curves and appropriate positive controls. Surface plasmon resonance, microscale thermophoresis, or isothermal titration calorimetry could test direct binding where feasible. Cellular studies would then need genotype-matched models, pathway readouts such as phospho-EGFR, phospho-AKT, or downstream markers, and counterscreens for cytotoxicity and nonspecific aggregation. For

polyphenols, assay-interference controls are particularly important. Only after these steps would it be reasonable to discuss mechanism-based anticancer potential.

A practical computational follow-up is equally important. The specified docking protocol should be executed with explicit receptor preparation, redocking validation, multiple receptor conformations, and alternative scoring functions. For the top compounds, replicate 100–300 ns MD trajectories should be run with validated ligand parameters, followed by convergence analysis. Where resources permit, alchemical relative free-energy calculations or rigorous enhanced sampling could test whether the rank order persists beyond endpoint rescoring. Prospective benchmarking against experimentally measured activity should remain the decisive criterion.

The study also illustrates a reporting principle relevant to computational biology more broadly: numerical precision should match epistemic precision. Reporting a docking score to two decimal places or a free energy to tenths of a kcal/mol can create a false impression of certainty when model error is much larger. We therefore used one decimal place for docking and explicit confidence intervals for aggregate metrics. Similar restraint should apply to machine-learning metrics when sample size is limited and to MD-derived observables when trajectories are correlated.

4.1 Reviewer-oriented technical checks and falsification criteria

A publication-oriented computational workflow should specify not only how a positive result is produced, but also what observations would weaken or falsify the proposed prioritization. For the present framework, the first falsification criterion is docking reproducibility. If redocking of the crystallographic ligands fails to recover the experimental binding mode within the prespecified 2.0 Å heavy-atom RMSD threshold, the associated receptor setup should not be used for prospective screening until protonation, box definition, retained waters, and ligand preparation are revisited. The second criterion is score instability across receptor conformations. A candidate that ranks highly in one static structure but falls outside the top quartile across a modest receptor ensemble should be flagged as conformation-sensitive rather than described as a robust binder. The third criterion is ligand-based extrapolation. If a candidate lies outside both descriptor and fingerprint applicability domains, a high predicted probability should be treated as a hypothesis requiring direct structural or experimental support, not as quantitative evidence.

The fourth falsification criterion concerns dynamic behavior. In an executed study, repeated ligand egress, persistent steric clashes, major binding-site distortion, or lack of convergence across replicate trajectories would contradict a claim of stable complex formation even if the initial docking score were favorable. Conversely, a stable trajectory would not rescue a pose that lacks plausible chemistry or that emerged from a failed redocking protocol. The fifth criterion concerns free-energy rescoring: if MM-GBSA rankings vary strongly with frame window, dielectric assumptions, or replicate trajectory, the correct conclusion is uncertainty rather than selective reporting of the most favorable window. These checks are designed to make the pipeline self-correcting and to reduce confirmation bias, a recurring risk when successive computational stages are used only to validate an initial favorite.

Prospective reproducibility also requires explicit software and environment capture. A future executed version should report exact program versions, parameter files, ligand protonation software, random seeds, GPU/CPU hardware where relevant, and command-line options. Receptor and ligand preparation scripts should be deposited alongside a checksum or version identifier for every input structure. For machine learning, fold assignments should be saved rather than regenerated, because small changes in scaffold partitioning can alter performance estimates. For molecular dynamics, topology generation, equilibration logs, energy-drift checks, and analysis selections should be retained. These details are often omitted because they appear operational, yet they determine whether an independent group can reproduce the numerical results.

Another reviewer-relevant issue is comparator selection. A computational lead should be compared with reference ligands that match the pocket and mechanism being modeled. Erlotinib is a reasonable structural

reference for the EGFR ATP site represented by 1M17, but it is not an appropriate universal comparator for PI3Kα or the AKT1 allosteric site. Therefore, a real study should use target-specific co-crystallized or pharmacologically established controls and should avoid cross-target claims based on raw score differences. Enrichment benchmarking with known actives and property-matched decoys would further test whether each docking setup distinguishes plausible ligands from background molecules before screening a natural-product collection.

4.2 Translational interpretation and decision thresholds

The proposed consensus score is best viewed as a resource-allocation device. It answers which compounds deserve the next unit of computational or experimental effort, not which compounds are drugs. This distinction affects decision thresholds. A top-three retention probability above 0.80 may justify replicate simulations and biochemical testing, whereas a candidate with a similar point score but a retention probability near 0.50 may warrant lower-cost confirmation first. In this sense, uncertainty is actionable information: it can determine whether to invest in deeper simulation, purchase or isolate a compound, synthesize an analogue, or defer a candidate until additional evidence is available. Such decision-theoretic use of uncertainty is more practical than reporting a long ranked list without indicating where ranks are effectively tied.

Natural-product programs also face a supply and identity problem that purely computational ranking does not capture. A promising structure may be difficult to isolate, may exist as multiple stereoisomers, may be unstable under assay conditions, or may have uncertain provenance. Future versions of the framework could add availability, synthetic accessibility, dereplication confidence, and stereochemical certainty as explicit objectives. For abundant flavonoids such as the anchors used here, availability is less restrictive, but chemical promiscuity and metabolism become more important. For rare alkaloids or complex terpenoids, the balance can reverse. Incorporating these practical constraints would move the framework from virtual-screening optimization toward experimentally informed portfolio design.

Finally, the biological interpretation should be tied to disease context rather than pathway popularity. EGFR, PI3Kα, and AKT1 are clinically important signaling proteins, but the value of inhibiting one or several nodes depends on genotype, lineage, prior therapy, pathway feedback, and achievable exposure. A future disease-focused study should therefore predefine a molecular context—for example, a specific EGFR alteration or PIK3CA-mutant setting—and select targets and comparators accordingly. The present multi-target benchmark deliberately avoids that disease-specific claim. Its contribution is computational: it shows how evidence from heterogeneous models can be normalized, stress-tested, and translated into a transparent shortlist whose uncertainty is visible before biological conclusions are attempted.

5. Limitations and future research

The most important limitation is that the study uses synthetic benchmark outputs. No actual docking poses, MD trajectories, or MM-GBSA frame energies were computed for this manuscript. The results are therefore suitable for demonstrating a methodological architecture and for serving as a pre-registration-like analysis plan, but they cannot support claims of real molecular binding or biological efficacy. Before submission as an empirical discovery paper, every synthetic value should be replaced with outputs from executed calculations, and the raw inputs, scripts, seeds, logs, and trajectories should be archived.

Second, the benchmark library approximates natural-product chemical space rather than reproducing a fully curated snapshot of COCONUT, PubChem, or another database. Real collections contain stereochemical ambiguity, salts, glycosides, tautomers, duplicate records, and annotation errors that can materially affect screening. Future work should define a versioned source dataset, provide exact structure identifiers, document exclusion counts at each curation step, and release a machine-readable manifest.

Third, structural uncertainty is only partially represented. Crystal structures capture particular conformational states and may contain mutations, missing regions, crystal contacts, or bound ligands that

bias the pocket. Ensemble docking across experimentally observed or MD-derived receptor conformations would better represent flexibility. Water molecules, protonation states, and metal ions can also be decisive for some targets and require target-specific treatment rather than a uniform preparation recipe.

Fourth, the ligand-based model assumes that a sufficiently curated activity dataset exists for the pathway task. In real projects, assay heterogeneity, censored IC50 values, different constructs, and publication bias can degrade label quality. Multi-task learning or assay-aware models may help, but they introduce additional complexity. External validation on temporally separated compounds would be stronger than cross-validation alone.

Fifth, computational cost will scale rapidly if dynamic simulations are extended to hundreds or thousands of compounds. The two-pass design mitigates this by reserving MD and MM-GBSA for a shortlist, but it may still miss candidates whose docking scores are mediocre because of scoring error. Active-learning strategies could iteratively allocate expensive calculations to compounds with high expected information gain rather than only to the current top ranks.

Finally, the consensus weights were chosen prospectively but remain subjective. Future studies should learn weights from retrospective benchmark sets with experimentally known outcomes, use Bayesian model averaging, or formulate the task as multi-objective Pareto optimization. Importantly, learned weights should be validated on held-out targets to avoid simply fitting historical noise.

6. Conclusion

This original-methods benchmark addresses a common weakness in natural-product virtual screening: the tendency to collapse uncertain computational evidence into a single deterministic ranking. By combining scaffold-aware machine learning, target-normalized docking, ADMET and applicability-domain controls, dynamic-stability surrogates, MM-GBSA-like rescoring, bootstrap analysis, and weight sensitivity, the proposed framework produces rankings that can be interrogated rather than merely reported. Under the synthetic benchmark, myricetin, quercetin, luteolin, and fisetin formed the most robust top group, with myricetin showing the strongest overall balance across the three EGFR/PI3Kα/AKT1 targets. These findings are computational hypotheses only. The practical contribution is the uncertainty-aware workflow and its auditable decision logic. Executed simulations, biochemical target-engagement assays, cellular pathway validation, and prospective external testing are necessary before any compound is advanced as a genuine therapeutic lead.

Author's contribution

Author's Contribution Rizwan Shaukat, Sakeena Batool and Atiba Irshad: data curation, writing – original draft. Aleena Batool: writing – review and editing. data curation. All authors read and approved the manuscript. All authors approved the final version of the manuscript.

Acknowledgement

This study was not funded by any funding agency.

Conflicts of Interest

The authors declare no conflict of interest.

Originality of the manuscript

It is declared that the submitted manuscript is original, unpublished and not under simultaneous consideration by any other journal.

Availability of data and material

We accept the transparency; therefore, raw data and statistical data will be available for everyone.

Code availability
Not applicable.

Consent to participate
We hereby declare that this research did not involve human or animal subjects. Therefore, consent to participate is not applicable.

Consent for publication
We hereby declare that this research did not involve human or animal subjects. Therefore, Consent to publish is not applicable.

REFERENCES

Akhtar, S., Ibrahim, A., Abdalla, A. M. A., Aatif, M., Gautam, R. K. S., Aslam, M., & Danishuddin. (2026). Computational identification of potential MMP-2 inhibitors in cancer using machine learning, molecular docking, and dynamics simulations. Computational Biology and Chemistry, 122, 108919. https://doi.org/10.1016/j.compbiolchem.2026.108919

Anjum, M. N., Nadeem, S., Al-Amin, M., Saalim, M., Siddique, F., Afzal, S., & Nisar, N. (2026). Discovery of potent ALK tyrosine kinase inhibitors for thyroid cancer via machine learning modeling, molecular docking, MD simulations, and DFT study. Computational Biology and Chemistry, 123, 108960. https://doi.org/10.1016/j.compbiolchem.2026.108960

Aryapour, H., Dehdab, M., Sohraby, F., & Bargahi, A. (2017). Prediction of new chromene-based inhibitors of tubulin using structure-based virtual screening and molecular dynamics simulation methods. Computational Biology and Chemistry, 71, 89–97. https://doi.org/10.1016/j.compbiolchem.2017.09.007

Bento, A. P., Gaulton, A., Hersey, A., Bellis, L. J., Chambers, J., Davies, M., Krüger, F. A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., & Overington, J. P. (2014). The ChEMBL bioactivity database: An update. Nucleic Acids Research, 42(D1), D1083–D1090. https://doi.org/10.1093/nar/gkt1031

Borthakur, S., Sarmah, P., Saikia, M., Dhakal, R., Das Gupta, D., Kalita, J., Verma, J. S., & Dutta, P. (2026). Computational insights into insect-derived semiochemicals from Formicidae family as potential JAK1 inhibitors: Virtual screening and molecular dynamics studies. Computational Biology and Chemistry, 120(2), 108730. https://doi.org/10.1016/j.compbiolchem.2025.108730

Burley, S. K., Bhikadiya, C., Bi, C., Bittrich, S., Chao, H., Chen, L., Craig, P. A., Crichlow, G. V., Dalenberg, K., Duarte, J. M., Dutta, S., Fayazi, M., Feng, Z., Flatt, J. W., Ganesan, S., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., … Zardecki, C. (2023). RCSB Protein Data Bank (RCSB.org): Delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. Nucleic Acids Research, 51(D1), D488–D508. https://doi.org/10.1093/nar/gkac1077

Chikhale, R. V., Abdelghani, H. T. M., Deka, H., Pawar, A. D., Patil, P. C., & Bhowmick, S. (2024). Machine learning assisted methods for the identification of low toxicity inhibitors of Enoyl-Acyl Carrier Protein Reductase (InhA). Computational Biology and Chemistry, 110, 108034. https://doi.org/10.1016/j.compbiolchem.2024.108034

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

Feng, X.-Y., Jia, W.-Q., Liu, X., Jing, Z., Liu, Y.-Y., Xu, W.-R., & Cheng, X.-C. (2019). Identification of novel PPARα/γ dual agonists by pharmacophore screening, docking analysis, ADMET prediction and molecular dynamics simulations. Computational Biology and Chemistry, 78, 178–189. https://doi.org/10.1016/j.compbiolchem.2018.11.023

Gálvez, J., Polo, S., Insuasty, B., Gutiérrez, M., Cáceres, D., Alzate-Morales, J. H., De-la-Torre, P., & Quiroga, J. (2018). Design, facile synthesis, and evaluation of novel spiro- and pyrazolo[1,5-c]quinazolines as cholinesterase inhibitors: Molecular docking and MM/GBSA studies. Computational Biology and Chemistry, 74, 218–229. https://doi.org/10.1016/j.compbiolchem.2018.03.001

Ganga, G. K. (2025). Accelerating drug discovery targeting dihydroorotate dehydrogenase using machine learning and generative AI approaches. Computational Biology and Chemistry, 118, 108443. https://doi.org/10.1016/j.compbiolchem.2025.108443

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. Nucleic Acids Research, 40(D1), D1100–D1107. https://doi.org/10.1093/nar/gkr777

Hassan, M., Abbas, Q., Ashraf, Z., Moustafa, A. A. M., & Seo, S.-Y. (2017). Pharmacoinformatics exploration of polyphenol oxidases leading to novel inhibitors by virtual screening and molecular dynamic simulation study. Computational Biology and Chemistry, 68, 131–142. https://doi.org/10.1016/j.compbiolchem.2017.02.012

Isa, M. A. (2019). Homology modeling and molecular dynamic simulation of UDP-N-acetylmuramoyl-l-alanine-d-glutamate ligase (MurD) from Mycobacterium tuberculosis H37Rv using in silico approach. Computational Biology and Chemistry, 78, 116–126. https://doi.org/10.1016/j.compbiolchem.2018.11.002

Isah, J. J., Uzairu, A., Uba, S., & Ibrahim, M. T. (2026). Integrating machine learning and molecular simulations for the design of potent HDAC2 inhibitors in diffuse large B-cell lymphoma. Computational Biology and Chemistry, 123, 108949. https://doi.org/10.1016/j.compbiolchem.2026.108949

Jagtap, U. A., Rathod, S., Shukla, R., & Paul, A. T. (2024). Computational insights into human UCP1 activators through molecular docking, MM-GBSA, and molecular dynamics simulation studies. Computational Biology and Chemistry, 113, 108252. https://doi.org/10.1016/j.compbiolchem.2024.108252

Jin, W.-Y., Ma, Y., Li, W.-Y., Li, H.-L., & Wang, R.-L. (2018). Scaffold-based novel SHP2 allosteric inhibitors design using receptor-ligand pharmacophore model, virtual screening and molecular dynamics. Computational Biology and Chemistry, 73, 179–188. https://doi.org/10.1016/j.compbiolchem.2018.02.004

Mena-Ulecia, K., & MacLeod-Carey, D. (2018). Interactions of 2-phenyl-benzotriazole xenobiotic compounds with human Cytochrome P450-CYP1A1 by means of docking, molecular dynamics simulations and MM-GBSA calculations. Computational Biology and Chemistry, 74, 253–262. https://doi.org/10.1016/j.compbiolchem.2018.04.004

Nguyen, H. D., & Kim, M.-S. (2023). Identification of promising inhibitory heterocyclic compounds against acetylcholinesterase using QSAR, ADMET, biological activity, and molecular docking. Computational Biology and Chemistry, 104, 107872. https://doi.org/10.1016/j.compbiolchem.2023.107872

Nyambe, M. M., Archibong, E. F., & Chinsembu, K. C. (2022). A DFT and molecular docking study of xerantholide and its interaction with Neisseria gonorrhoeae carbonic anhydrase. Computational Biology and Chemistry, 101, 107779. https://doi.org/10.1016/j.compbiolchem.2022.107779

Páll, S., Zhmurov, A., Bauer, P., Abraham, M., Lundborg, M., Gray, A., Hess, B., & Lindahl, E. (2020). Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. The Journal of Chemical Physics, 153(13), 134110. https://doi.org/10.1063/5.0018516

Shevate, K., Rajagopal, K., Byran, G., Unni, A., & Antony, J. (2026). Targeting TNBC metastasis: In-silico identification of natural origin ROCK inhibitors via virtual screening, ADMET profiling, MM-GBSA, DFT, and molecular dynamics. Computational Biology and Chemistry, 120(1), 108671. https://doi.org/10.1016/j.compbiolchem.2025.108671

Singh, S. P., & Gupta, D. (2017). Discovery of potential inhibitor against human acetylcholinesterase: A molecular docking and molecular dynamics investigation. Computational Biology and Chemistry, 68, 224–230. https://doi.org/10.1016/j.compbiolchem.2017.04.002

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. Journal of Cheminformatics, 13, 2. https://doi.org/10.1186/s13321-020-00478-9

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

Uba, A. I., Aluwala, H., Liu, H., & Wu, C. (2022). Elucidation of partial activation of cannabinoid receptor type 2 and identification of potential partial agonists: Molecular dynamics simulation and structure-based virtual screening. Computational Biology and Chemistry, 99, 107723. https://doi.org/10.1016/j.compbiolchem.2022.107723

Uba, A. I., & Wu, C. (2022). Identification of potential antagonists of CRF1R for possible treatment of stress and anxiety neuro-disorders using structure-based virtual screening and molecular dynamics simulation. Computational Biology and Chemistry, 100, 107743. https://doi.org/10.1016/j.compbiolchem.2022.107743

UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. Nucleic Acids Research, 51(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Yoon, D., & Lee, H. (2025). In silico discovery of novel compounds for FAK activation using virtual screening, AI-based prediction, and molecular dynamics. Computational Biology and Chemistry, 117, 108420. https://doi.org/10.1016/j.compbiolchem.2025.108420

Zhu, J., Sun, D., Li, X., Jia, L., Cai, Y., Chen, Y., Jin, J., & Yu, L. (2023). Developing new PI3Kγ inhibitors by combining pharmacophore modeling, molecular dynamic simulation, molecular docking, fragment-based drug design, and virtual screening. Computational Biology and Chemistry, 104, 107879. https://doi.org/10.1016/j.compbiolchem.2023.107879

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 1–2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics, 27(3), 343–350. https://doi.org/10.1093/bioinformatics/btq662

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23(6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

Durrant, J. D., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. BMC Biology, 9, 71. https://doi.org/10.1186/1741-7007-9-71

Gilson, M. K., & Zhou, H.-X. (2007). Calculation of protein-ligand binding affinities. Annual Review of Biophysics and Biomolecular Structure, 36, 21–42. https://doi.org/10.1146/annurev.biophys.36.040306.132550

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular dynamics simulation for all. Neuron, 99(6), 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011

Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. Nature Reviews Drug Discovery, 3, 935–949. https://doi.org/10.1038/nrd1549

Lipinski, C. A., Lombardo, F., Dominy, B. W., & Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews, 46(1–3), 3–26. https://doi.org/10.1016/S0169-409X(00)00129-0

Lionta, E., Spyrou, G., Vassilatis, D. K., & Cournia, Z. (2014). Structure-based virtual screening for drug discovery: Principles, applications and recent advances. Current Topics in Medicinal Chemistry, 14(16), 1923–1938. https://doi.org/10.2174/1568026614666140929124445

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30(16), 2785–2791. https://doi.org/10.1002/jcc.21256

Murphy, R. F. (2011). An active role for machine learning in drug development. Nature Chemical Biology, 7, 327–330. https://doi.org/10.1038/nchembio.576

Plewczynski, D., Łaźniewski, M., Augustyniak, R., & Ginalski, K. (2011). Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. Journal of Computational Chemistry, 32(4), 742–755. https://doi.org/10.1002/jcc.21643

Stamos, J., Sliwkowski, M. X., & Eigenbrot, C. (2002). Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. Journal of Biological Chemistry, 277(48), 46265–46272. https://doi.org/10.1074/jbc.M207135200

Veber, D. F., Johnson, S. R., Cheng, H.-Y., Smith, B. R., Ward, K. W., & Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry, 45(12), 2615–2623. https://doi.org/10.1021/jm020017n

Wu, W.-I., Voegtli, W. C., Sturgis, H. L., Dizon, F. P., Vigers, G. P. A., & Brandhuber, B. J. (2010). Crystal structure of human AKT1 with an allosteric inhibitor reveals a new mode of kinase inhibition. PLoS ONE, 5(9), e12913. https://doi.org/10.1371/journal.pone.0012913

Yang, J., Roy, A., & Zhang, Y. (2013). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics, 29(20), 2588–2595. https://doi.org/10.1093/bioinformatics/btt447

Genheden, S., & Ryde, U. (2015). The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery, 10(5), 449–461. https://doi.org/10.1517/17460441.2015.1032936

Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., & Hou, T. (2019). End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. Chemical Reviews, 119(16), 9478–9508. https://doi.org/10.1021/acs.chemrev.9b00055

Bajorath, J. (2017). Overcoming false positives: A key challenge in virtual screening. Expert Opinion on Drug Discovery, 12(8), 759–761. https://doi.org/10.1080/17460441.2017.1332995.