



Counter-Screens Before Rankings: A Falsification-First Strategy for Eliminating Artefactual, Promiscuous, and Mechanistically Implausible Natural Product Leads

Daniel Kim^{1*}, Jihoon Park¹, Min-seo Kang², Hyun-woo Jung³

¹Department of Falsification-First Lead Discovery, College of Pharmacy, Kyung Hee University, Seoul, South Korea.

²Department of Artefact and Promiscuity Elimination, College of Pharmacy, Chung-Ang University, Seoul, South Korea.

³Department of Mechanistically Implausible Lead Filtering, College of Pharmacy, Kangwon National University, Chuncheon, South Korea.

ABSTRACT

Natural-product lead discovery increasingly combines biochemical screening, phenotypic assays, virtual screening, machine learning, and target-deconvolution technologies, yet candidate ranking can still reward positive signals before plausible alternative explanations have been actively challenged. This creates an evidential asymmetry: potency, docking score, predicted affinity, phenotypic magnitude, or apparent target breadth can raise a candidate's priority even when the signal could arise from assay interference, aggregation, reactive chemistry, impurities, nonspecific activity, or an implausible target assignment. This article develops an original decision framework that reverses that order. Rather than asking first which candidate ranks highest, it asks which candidate can survive structured attempts to disprove the interpretation attached to its activity. The analysis distinguishes artefact from promiscuity and promiscuity from purposeful polypharmacology, while treating mechanistic plausibility, target engagement, and orthogonal functional reproduction as separate evidentiary requirements. The proposed framework organizes counter-screens as sequential challenges to the lead hypothesis and reserves final ranking for candidates that remain interpretable after those challenges. Importantly, failure does not always mandate rejection: discordant evidence can identify an impurity, alternative target, context-sensitive mechanism, or unresolved assay limitation that warrants investigation. The framework is a decision architecture rather than a validated universal screening algorithm. Its intended value is to make negative evidence explicit, preserve uncertainty, and prevent attractive scores from acquiring more evidential meaning than the experiments support. Prospective, multi-campaign testing is required before claims of improved attrition, efficiency, or translational success are justified.

Key Words: Natural products, Counter-screens, Assay interference, Promiscuity, Polypharmacology, Target engagement

eIJPPR 2025; 15(6):115-124

HOW TO CITE THIS ARTICLE: Kim D, Park J, Kang MS, Jung HW. Counter-Screens Before Rankings: A Falsification-First Strategy for Eliminating Artefactual, Promiscuous, and Mechanistically Implausible Natural Product Leads. Int J Pharm Phytopharmacol Res. 2025;15(6):115-24. <https://doi.org/10.51847/WZbnhhz6fC>

INTRODUCTION

Natural products remain an important source of pharmacologically active chemical matter, but their structural diversity, variable physicochemical behavior, complex target relationships, and development constraints mean that an initial activity signal is only the beginning of

lead interpretation [1]. Computational methods now shape candidate selection through docking, similarity searching, molecular modeling, machine learning, and related prioritization strategies that move compounds rapidly toward the top of a screening queue [2]. These methods compress search space, but their outputs are evidentially asymmetric: a high score supplies a reason to continue; it

Corresponding author: Daniel Kim

Address: Department of Falsification-First Lead Discovery, College of Pharmacy, Kyung Hee University, Seoul, South Korea.

E-mail: jaesung.kim@khu.ac.kr

Received: 14 August 2025; **Revised:** 24 November 2025; **Accepted:** 28 November 2025

This is an **open access** journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.



does not demonstrate that the biological explanation attached to that score is correct.

Current natural-product discovery strategies increasingly integrate computation, analytical chemistry, biological screening, target identification, and optimization rather than relying on a single technology [3]. Integration nevertheless does not resolve the ordering problem. A compound may accumulate positive evidence while a common weakness remains untested: the signal may be generated by the assay, by a concentration-dependent physical state, by an unintended chemical species, or by a biological mechanism different from the one used to justify its rank. High-throughput biochemical and cell-based screening therefore depends on confirmation and follow-up because a primary hit is not equivalent to a validated lead [4].

The argument developed here is that ranking should be conditional on deliberate exposure to counter-evidence. This does not imply that positive evidence is weak by definition. It means that discovery systems are commonly optimized to identify reasons for continuation, whereas reasons for disqualification are distributed across later assay-development, medicinal-chemistry, analytical, biophysical, and mechanistic work. The proposed falsification-first strategy reorganizes those activities into an explicit decision architecture. A candidate earns ranking eligibility only after plausible rival explanations have been challenged at the level relevant to the claim. Survival does not prove the lead hypothesis true; it means that the interpretation is harder to dismiss using the counter-tests presently available. This prevents numerical strength from outrunning the biological proposition actually supported.

Why natural product hits accumulate multiple routes to false positivity

False positivity should not be reduced to a single structural warning. Analyses of pan-assay interference compound

behavior show that the meaning of an alert depends on activity context and molecular environment rather than on substructure alone [5]. Critical reassessment of PAINS filtering likewise supports using such motifs as warning information while cautioning against treating an alert as a self-sufficient diagnosis of assay interference [6]. For natural-product discovery, this matters because structurally unusual or multifunctional molecules may be penalized even when the relevant interference mechanism has not been demonstrated.

The problem is broader than PAINS. Modern computational assessments can flag several forms of likely screening liability, but these classifications remain predictions rather than observations of the mechanism operating in a particular experiment [7]. Conversely, characterized nuisance compounds can be deployed as positive controls to interrogate whether an assay architecture is vulnerable to known interference [8]. These approaches answer different questions: one asks whether a molecule resembles compounds with problematic behavior; the other asks whether the measurement system can be fooled.

The proposed framework therefore treats a positive natural-product hit as an evidential fork. The signal may reflect the intended biological interaction, an assay-specific distortion, a physicochemical phenomenon, a chemically altered active species, or another unresolved explanation. No single possibility should be inferred merely because a candidate carries an alert or occupies a high rank. **Table 1** organizes the early evidence objects that determine legitimate inference before mechanistic adjudication. **Figure 1** shows why continuation from a positive result should open competing explanations before it opens a ranking contest. The framework therefore treats uncertainty as structured and testable rather than as a generic reason for caution.

Table 1. Early evidence objects and permissible interpretation before mechanistic adjudication

Evidence object	Direct meaning	Not established alone	Proposed decision use
Primary positive signal	Activity in tested conditions	Mechanism or lead validity	Opens the hypothesis
Structural-context warning	Motif linked to interference risk	Actual interference	Triggers targeted challenge
PAINS/filter designation	Context-dependent warning	Universal promiscuity	Supports cautious triage
Computational liability flag	Predicted nuisance propensity	Observed artefact	Prioritizes counter-screening
Nuisance-control response	Assay vulnerability	Candidate-specific interference	Tests assay robustness
Lead hypothesis	Working biological explanation	Truth of that explanation	Remains revisable

Note: Decision uses are proposed; evidence provenance is retained in the surrounding text.

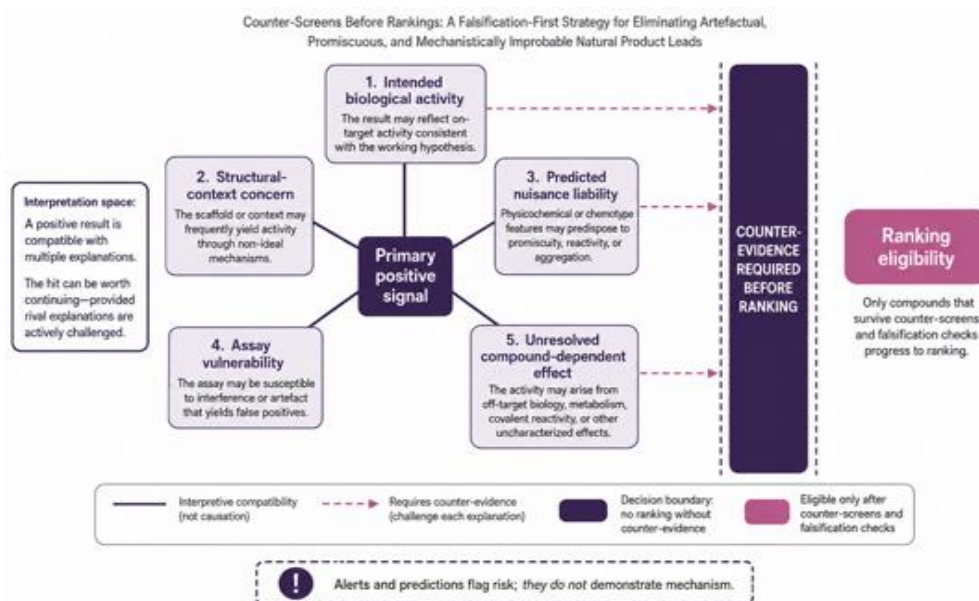


Figure 1. A positive natural-product hit as an evidential fork rather than a lead verdict

Artefact is different from promiscuity

An artefact concerns the validity of the observed signal as evidence for the claimed biological event. Promiscuity concerns the breadth and interpretation of activity across targets or assays. Those questions can overlap, but they are not interchangeable. Large-scale analysis of PAINS alerts has shown that structural flagging can generate problematic classifications when a motif is treated as direct evidence of interference [9]. The first counter-screening question should therefore be mechanistic: can the measurement be reproduced in a format or condition that excludes the suspected interference route?

Promiscuity requires a different evidentiary test. Screening analyses have shown that many compounds carrying PAINS alerts do not behave as universal frequent hitters across available assay histories [10]. A structural warning may thus increase suspicion without demonstrating broad empirical activity. Conversely, a compound may appear active in multiple assays because a shared nuisance mechanism repeatedly generates signal. The distinction is whether breadth represents measurement distortion, nonspecific chemistry, or reproducible interaction with several biological targets.

Reviews of frequent hitters identify heterogeneous nuisance mechanisms, including aggregation, reactivity, reporter interference, and related assay-dependent phenomena [11]. The proposed framework therefore separates signal validity from activity breadth. A candidate can fail on the first axis without truly being promiscuous, and it can be broad on the second without the original signal being artefactual. Structural alerts and historical frequent-hit behavior are context-sensitive warning evidence rather than sufficient mechanistic diagnoses [9-

11]. Each failure mode consequently requires a different counter-screen.

Promiscuity is different from polypharmacology

Multiple-target activity is not automatically a discovery defect. Systematic investigation of dual-target-directed ligands shows that medicinal chemistry can deliberately pursue compounds intended to interact with more than one target within a coherent design strategy [12]. The issue is therefore not simply how many targets or assays respond, but whether the breadth can be explained by reproducible, chemically plausible, and biologically interpretable interactions.

Experimental design can integrate phenotype with target-directed reasoning to construct and evaluate polypharmacological compounds [13]. In such cases, breadth forms part of the intended pharmacological hypothesis. Computational generative approaches likewise demonstrate that multi-target profiles can be explicit optimization objectives rather than accidental liabilities [14]. None of these findings implies that broad activity is inherently desirable or therapeutically superior. They establish only that target multiplicity can be purposeful and experimentally interrogable.

The framework therefore distinguishes nuisance promiscuity from polypharmacology by the quality of explanation attached to the activity pattern. Nuisance promiscuity is suspected when breadth is assay-dependent, chemically nonspecific, unstable across orthogonal measurements, or unsupported by a coherent target model. Polypharmacology becomes credible when relevant interactions are reproducible, mechanistically interpretable, and aligned with the biological objective.

This is an evidential distinction, not a numerical cutoff, and a candidate may move between interpretations as counter-evidence accumulates.

Mechanistic implausibility as an independent reason for challenge

A candidate can survive artefact screening and still carry an evidentially weak mechanism. Drug-target prediction methods differ in data construction, validation design, applicability domain, and susceptibility to leakage; validation guidance therefore treats predicted associations as hypotheses requiring independent testing [15]. Mechanistic plausibility should consequently be assessed separately from whether the primary signal is technically real.

Machine-learning classification can reduce false positives within structure-based virtual screening, but improved benchmark discrimination does not establish that a selected compound engages the nominated target experimentally [16]. Consensus virtual-screening approaches can combine several computational signals without converting model agreement into mechanistic proof [17]. Shared data, representations, structures, or scoring assumptions may preserve correlated error.

Mechanistic implausibility is therefore proposed as a reason for challenge, not automatic rejection. The challenge must match the claim. A binding hypothesis requires evidence different from a pathway claim, and a pathway claim requires evidence different from a phenotype-level explanation. A surprising target assignment may be wrong, but it may also indicate an unconventional interaction or incomplete biological model. The relevant question is whether the proposed mechanism generates testable consequences that survive independent interrogation. Computational plausibility should determine what deserves testing, not what is already considered established. A mechanistically attractive prediction can therefore raise priority for validation while simultaneously increasing the importance of a decisive target-focused counter-test.

Counter-screens as sequential attempts to disprove the lead hypothesis

Counter-screening is most informative when each test addresses an explicit rival explanation. ChemFH demonstrates that several frequent-false-positive liabilities can be screened computationally before expensive follow-up [18]. STOPLIGHT shows that hit prioritization can incorporate liability evidence beyond primary activity [19]. Badapple 2.0 uses historical assay behavior to estimate promiscuity propensity and identify compounds warranting challenge [20].

These tools should not become a single rejection score. A computational warning is an instruction to test a proposition, not proof that it is true. The proposed sequence is: state the lead hypothesis; identify a plausible alternative; select a discriminating counter-screen; update evidentiary status; then address the next unresolved alternative. The ordering is intended to reduce ambiguity rather than impose a universal laboratory sequence.

Negative evidence thereby becomes decision-bearing. A counter-screen is more than a replicate or generic negative control: competing explanations should predict meaningfully different outcomes. Failure can alter eligibility for ranking, whereas failure to reproduce a suspected liability strengthens only the interpretation actually tested. It does not certify the entire lead hypothesis. Passing one challenge also does not erase uncertainty elsewhere. The framework therefore accumulates constrained interpretations rather than binary passes. This preserves the asymmetry between evidence that merely removes one rival explanation and evidence that directly strengthens the intended mechanism. Both can support continuation, but they should not be treated as interchangeable increments in confidence.

A falsification-first gatekeeping architecture

The proposed architecture organizes falsification around distinct questions. Reporter inhibition can create false-positive signals in luciferase assays and requires readout-specific challenge [21]. Colloidal aggregation represents a different physicochemical mechanism capable of misleading screening activity [22]. Natural-product screening is not exempt: work with minor cannabinoids illustrates how aggregation can complicate interpretation of target-screen activity [23].

These observations motivate four progressively mechanism-proximal gates. Gate I asks whether a physicochemical or assay artefact explains the signal. Gate II asks whether apparent breadth reflects nonspecific activity rather than interpretable target relationships. Gate III asks whether the nominated target and pathway remain plausible under direct challenge. Gate IV asks whether orthogonal functional evidence reproduces the biological consequence used to justify the mechanism. The ordering is proposed and has not been prospectively validated as universally optimal.

The gates are not identical for every program. Their operational content should depend on assay technology, compound chemistry, target class, biological context, and earlier uncertainty. Movement between gates is also not simply pass/fail. Counter-evidence may support continuation, mechanistic re-framing, temporary pause, additional analytical work, or rejection.

The architecture therefore distinguishes ranking evidence from ranking eligibility. Potency, selectivity, docking score, phenotypic magnitude, or tractability may eventually determine relative priority, but they should not automatically compensate for an unresolved explanation that undermines what the activity means. The fixed principle is narrow: ranking should not outrun the strongest unresolved alternative explanation material to the lead hypothesis. A gate is useful only when its result can change interpretation, eligibility, or the next experiment; otherwise it adds procedural burden without meaningful falsification.

Gate I: physicochemical and assay artefacts

Gate I asks whether the observed signal could arise without the claimed biological interaction. Tox21 work directly measured luciferase inhibition and autofluorescence across thousands of chemicals, demonstrating that assay interference can be characterized experimentally and can depend on readout conditions [24]. Changing reporter technology, wavelength, detergent condition, detection chemistry, or assay format may therefore be more informative than simply repeating the primary measurement.

Chemical state also matters. Fused tetrahydroquinolines have shown assay-interfering chemistry capable of

misleading hit interpretation [25]. In another campaign, apparent KRAS activity was traced to palladium salt impurities rather than the intended molecule [26]. Identity, purity, stability, aggregation state, and readout compatibility therefore belong to lead interpretation rather than merely downstream quality control.

Gate I is passed only provisionally when the relevant artefact mechanism remains unsupported under conditions material to the claim. Concentration dependence, detergent sensitivity, lot-to-lot differences, optical behavior, and chemical stability are useful only when they test a plausible route to the observed signal. A candidate does not need every conceivable interference assay; it needs tests capable of discriminating among alternatives made credible by its chemistry and assay environment. **Table 2** separates the principal Gate I failure modes by the inference each counter-screen can overturn. **Figure 2** shows how those tests progressively constrain the source of apparent activity rather than attaching a generic “false positive” label. This source-focused approach matters when artefact and genuine biology may coexist, because the aim is to determine what remains attributable to the intended biology.

Table 2. Gate I failure modes distinguished by the inference challenged

Failure mode	Counter-screen logic	Interpretation if supported
Reporter/readout interference	Change detection principle or measure interference directly	Signal may not report intended biology
Colloidal aggregation	Alter aggregation-permissive conditions and retest	Activity may arise from colloids
Chemically generated interference	Characterize species and retest defined material	Parent-compound attribution becomes uncertain
Contaminant-driven activity	Purify, quantify contaminants, and retest	Another species may account for activity
Computational liability only	Test the predicted mechanism experimentally	Prediction gains weight only if observed

Note: Counter-screen logic is a proposed synthesis; no universal rejection threshold is implied.

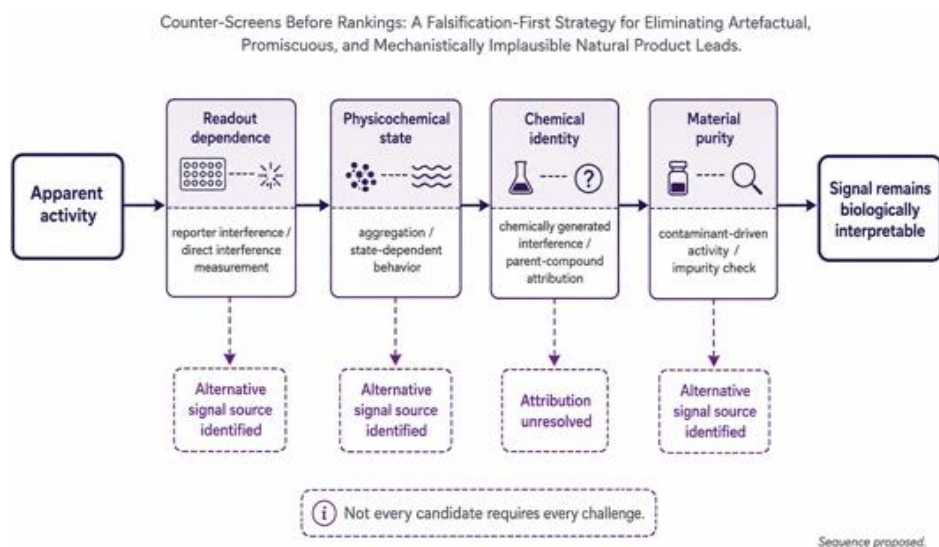


Figure 2. From apparent activity to an experimentally constrained signal source

Gate II: nonspecific and promiscuous activity

Once the signal is technically credible, breadth must be interpreted rather than counted. Chemical-proteomic profiling of a large kinase-inhibitor collection demonstrates that compounds centered on intended targets can possess substantial measured target landscapes [27]. Breadth is therefore not an assay artefact simply because it is broad; it raises questions about selectivity, target dependence, and mechanism.

Clinically relevant kinase inhibitors can also inhibit non-kinase proteins, showing that genuine off-target pharmacology can cross conventional target-family boundaries [28]. Gate II therefore asks whether broad activity is reproducible and interpretable or whether it collapses under controls for aggregation, reactive chemistry, common assay format, or generalized cellular injury. A broad profile that survives these challenges may still be undesirable for a particular program, but that is a pharmacological prioritization decision rather than proof of nuisance behavior.

No universal target-count threshold is proposed. Biologically coherent multi-target activity can remain credible, whereas breadth generated by a shared nuisance mechanism should lose weight. The relevant comparison is between explanations of breadth: direct interaction with several meaningful targets, concentration-dependent nonspecific effects, assay-common interference, or generalized toxicity. Gate II is provisionally satisfied when the activity pattern has an interpretable molecular basis and remains stable under counter-screens directed at the suspected nonspecific mechanism. The result should still be reported as context-dependent breadth rather than converted automatically into a claim of beneficial polypharmacology.

Gate III: target and pathway plausibility

Gate III moves from “the signal appears real” to “the proposed mechanism is credible.” Target-engagement assays provide ways to test whether a compound engages its nominated target in relevant experimental contexts [29]. Biochemical activity, docking compatibility, transcriptional association, and phenotype do not themselves demonstrate cellular engagement.

For natural products, CETSA and thermal proteome profiling provide complementary routes for target identification and validation [30]. Recent analyses further indicate that no single method resolves every natural-product target-deconvolution problem and that complementary strategies may be necessary [31]. Gate III therefore treats mechanism as a convergence problem rather than a verdict issued by the method that produced the first target assignment.

The strongest feasible test should address the target claim supporting rank. A candidate justified by direct binding should face a binding-sensitive challenge; one justified by pathway modulation should also confront whether engagement plausibly explains that pathway response. Failure may weaken the nominated mechanism without eliminating the compound because the phenotype may arise through another target or because the assay has a detection boundary. Passing Gate III preserves plausibility; it does not establish causal sufficiency, therapeutic relevance, or exclusivity. The distinction is critical where several proteins can produce similar pathway or phenotypic consequences and where target engagement may occur without being the dominant driver of the observed effect.

Gate IV: orthogonal functional reproduction

Credible engagement can still be pharmacologically incomplete. Chemoproteomics provides broad approaches to target deconvolution and can expose interactions not anticipated by the original hypothesis [32]. Cellular mapping of small-molecule binding sites can resolve interaction evidence at greater molecular detail [33]. CeTEAM demonstrates that cellular target engagement can be connected experimentally to downstream pharmacology [34].

Gate IV asks whether the functional consequence attributed to the mechanism reproduces through a genuinely different evidentiary route. Merely changing instrument, laboratory, or endpoint label is insufficient if the second assay preserves the same interference chemistry or inferential dependency. Orthogonality should reduce dependence on the assumptions, readout chemistry, or inferential step that generated the original claim. Target-engagement and target-identification strategies operating on different biophysical or proteomic principles can provide complementary mechanistic evidence [29, 30, 32]. The proposed standard is convergence without equivalence. Agreement among engagement, binding-site evidence, pathway perturbation, and phenotype strengthens interpretation but does not prove causality. Discordance is also informative: engagement without the expected phenotype may indicate pharmacological insufficiency, while phenotype without measurable engagement should reopen the target hypothesis. Gate IV therefore supports an evidence-qualified mechanism, not clinical effectiveness, safety, or translational readiness. Its purpose is narrower: to test whether the functional interpretation survives a change in evidentiary route rather than merely a repeat of the same inferential chain.

Ranking only the candidates that survive falsification

Ranking becomes meaningful after the decision object changes. Before counter-screening, a “lead” may be a positive score attached to an interpretation. After counter-screening, it is an evidence-qualified candidate with characterized liabilities, alternative explanations, target support, and residual uncertainty. Machine-learning-assisted hit prioritization shows that candidates can be ranked using information richer than primary-screen response alone [35]. The framework extends that logic by proposing that some counter-evidence should alter eligibility before positive attributes determine relative priority.

Survivors need not have identical evidence packages, and a composite score should not erase a decisive unresolved liability. Aggregation, uncertain material identity, or failure of target engagement is not necessarily

compensable by exceptional potency elsewhere. Comparison should remain campaign-specific: potency, cellular engagement, mechanistic selectivity, tractability, or exposure feasibility may receive different weight. What should not occur is compensation without scrutiny.

Table 3 summarizes the proposed dimensions available once earlier challenges have clarified what positive evidence means. Re-ranking is not a return to the original leaderboard. It is comparison among candidates whose scores now carry different evidentiary meanings because some rival explanations have weakened, others remain unresolved, and some mechanisms have been re-framed. A lower-potency survivor may therefore deserve higher priority than a stronger primary hit whose mechanism remains materially ambiguous.

Table 3. Proposed dimensions for re-ranking evidence-qualified natural-product candidates

Dimension	Ranking question	Stronger evidence state	Residual caution
Signal integrity	Robust to relevant artefact challenges?	Orthogonal controls agree	Untested interference may remain
Mechanistic specificity	Is breadth explainable?	Interpretable target relationships	Off-target activity may matter
Target plausibility	Does the target survive challenge?	Engagement or complementary target identification	Engagement is not causal sufficiency
Functional coherence	Does biology fit the mechanism?	Orthogonal phenotype or pathway response	Parallel mechanisms may mimic coherence
Residual uncertainty	Which rival explanation remains?	Uncertainty is explicitly bounded	No detected failure is not proof
Decision status	Advance, pause, reject, or re-frame?	Evidence fits the next decision	No universal threshold

Note: All decision dimensions are proposed.

When a failed counter-screen should trigger investigation rather than rejection

Falsification-first does not mean rejection-first. Natural-product target identification remains technically heterogeneous, with complementary strengths and limitations across available approaches [36]. A failed counter-screen can mean that the current hypothesis is undermined, that the test is not decisive, or that the candidate is interesting for a different mechanism.

The palladium/KRAS example shows how identifying another active species can invalidate attribution to the intended compound [26]. Genuine non-kinase activities of kinase inhibitors show that unexpected activity can instead represent real off-target pharmacology [28]. Natural-product target-identification strategies likewise explain why discordance may sometimes justify a different method rather than immediate abandonment [31]. A failed lead interpretation is therefore not always identical to a failed molecule.

The proposed escalation rule is interpretability-based. Reject when counter-evidence directly explains the result through a disqualifying mechanism. Investigate when discordance is technically plausible and another

experiment can discriminate an alternative species, target, or mechanism. Pause when available methods cannot resolve the discrepancy proportionately. Re-frame when evidence supports biological activity but not the original target explanation. Escalation should reduce a defined uncertainty rather than protect an attractive candidate from disconfirming evidence [36]. The practical stopping rule is whether another experiment can discriminate among remaining explanations strongly enough to change the decision.

Implementation boundaries

The framework is modular rather than a fixed counter-screen panel. A luciferase campaign, phenotypic screen, covalent program, and extract-derived natural-product campaign should not share an identical sequence because nuisance mechanisms differ with assay design and compound behavior [11]. Implementation should identify the smallest set of rival explanations capable of reversing the immediate decision, then select counter-screens that distinguish them. The sequence is an evidentiary discipline, not a rigid laboratory recipe.

Computational filters also require careful placement. Validation principles for drug-target prediction keep model

output distinct from experimental evidence [15], while false-positive prediction tools can prioritize liabilities for testing [18]. Prediction can accelerate falsification but cannot substitute for observation of the predicted mechanism.

More testing has costs: material consumption, delay, false negatives, platform discordance, and inconclusive results. Target-engagement methods have technique-dependent limitations [29], and natural-product target identification may remain unresolved despite complementary approaches [36]. The objective is proportionate testing of explanations capable of changing the decision.

Prospective validation remains necessary. The architecture has not been shown to reduce attrition, cost, development time, or false-positive progression. Evidence-rich hit prioritization is feasible [35], but comparative multi-campaign studies are required before performance claims about this framework can be justified. Such studies should compare interpretability, reproducibility, downstream attrition, and decision resources.

CONCLUSION

Natural-product lead discovery should distinguish being interesting from being difficult to falsify. Primary activity, computational scores, apparent selectivity, target predictions, and phenotypic magnitude can justify investigation, but none should receive more evidential authority than the relevant counter-tests survived.

The proposed framework makes rival explanations decision-bearing before final ranking. It separates artefact from promiscuity, promiscuity from polypharmacology, target plausibility from target engagement, and engagement from functional reproduction. Candidates advance because their interpretations remain credible after challenge, not because uncertainty has disappeared.

This is neither a universal assay panel nor a validated attrition model. Some failures expose assay artefacts or unsuitable materials; others reveal impurities, unexpected targets, context dependence, or limitations of the counter-screen itself. The central discipline is narrower: positive evidence should earn weight only in proportion to the strongest unresolved alternative explanation. Ranking thus becomes a consequence of surviving informed attempts at disproof rather than the starting point of lead belief. The practical question changes from “Which hit looks best?” to “Which interpretation has earned the right to be compared?”

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, et al. Natural products in drug discovery: Advances and opportunities. *Nat Rev Drug Discov.* 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z
- [2] Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature.* 2023;616(7958):673-85. doi:10.1038/s41586-023-05905-z
- [3] Wang Y, Wang F, Liu W, Geng Y, Shi Y, Tian Y, et al. New drug discovery and development from natural products: Advances and strategies. *Pharmacol Ther.* 2024;264:108752. doi:10.1016/j.pharmthera.2024.108752
- [4] Blay V, Tolani B, Ho SP, Arkin MR. High-throughput screening: Today's biochemical and cell-based approaches. *Drug Discov Today.* 2020;25(10):1807-21. doi:10.1016/j.drudis.2020.07.024
- [5] Siramshetty VB, Preissner R, Gohlke BO. Exploring activity profiles of PAINS and their structural context in target-ligand complexes. *J Chem Inf Model.* 2018;58(9):1847-57. doi:10.1021/acs.jcim.8b00385
- [6] Baell JB, Nissink JWM. Seven year itch: Pan-assay interference compounds (PAINS) in 2017—utility and limitations. *ACS Chem Biol.* 2018;13(1):36-44. doi:10.1021/acscchembio.7b00903
- [7] Alves VM, Yasgar A, Wellnitz J, Rai G, Rath M, Braga RC, et al. Lies and liabilities: Computational assessment of high-throughput screening hits to identify artifact compounds. *J Med Chem.* 2023;66(18):12828-39. doi:10.1021/acs.jmedchem.3c00482
- [8] Hu H, Yi X, Xue L, Baell JB. A collection of useful nuisance compounds (CONS) for interrogation of bioassay integrity. *JACS Au.* 2024;4(12):4883-91. doi:10.1021/jacsau.4c00851
- [9] Capuzzi SJ, Muratov EN, Tropsha A. Phantom PAINS: Problems with the utility of alerts for pan-assay interference compounds. *J Chem Inf Model.* 2017;57(3):417-27. doi:10.1021/acs.jcim.6b00465
- [10] Jasial S, Hu Y, Bajorath J. How frequently are pan-assay interference compounds active? Large-scale analysis of screening data reveals diverse activity profiles, low global hit frequency, and many consistently inactive compounds. *J Med Chem.* 2017;60(9):3879-86. doi:10.1021/acs.jmedchem.7b00154

- [11] Yang ZY, He JH, Lu AP, Hou TJ, Cao DS. Frequent hitters: Nuisance artifacts in high-throughput screening. *Drug Discov Today*. 2020;25(4):657-67. doi:10.1016/j.drudis.2020.01.014
- [12] Lembo V, Bottegoni G. Systematic investigation of dual-target-directed ligands. *J Med Chem*. 2024;67(12):10374-85. doi:10.1021/acs.jmedchem.4c00838
- [13] Li S, Li X, Ma L, Luo Z, Yin F, Zhang Y, et al. Polypharmacological drug design guided by integrating phenotypic and restricted fragment docking strategies. *J Med Chem*. 2024;67(23):21049-69. doi:10.1021/acs.jmedchem.4c01731
- [14] Munson BP, Chen M, Bogosian A, Kreisberg JF, Licon K, Abagyan R, et al. De novo generation of multi-target compounds using deep generative chemistry. *Nat Commun*. 2024;15(1):3636. doi:10.1038/s41467-024-47120-y
- [15] Tanoli Z, Schulman A, Aittokallio T. Validation guidelines for drug-target prediction methods. *Expert Opin Drug Discov*. 2025;20(1):31-45. doi:10.1080/17460441.2024.2430955
- [16] Adeshina YO, Deeds EJ, Karanicolas J. Machine learning classification can reduce false positives in structure-based virtual screening. *Proc Natl Acad Sci U S A*. 2020;117(31):18477-88. doi:10.1073/pnas.2000585117
- [17] Moshawih S, Bu ZH, Goh HP, Kifli N, Lee LH, Goh KW, et al. Consensus holistic virtual screening for drug discovery: A novel machine learning model approach. *J Cheminform*. 2024;16(1):62. doi:10.1186/s13321-024-00855-8
- [18] Shi S, Fu L, Yi J, Yang Z, Zhang X, Deng Y, et al. ChemFH: An integrated tool for screening frequent false positives in chemical biology and drug discovery. *Nucleic Acids Res*. 2024;52(W1). doi:10.1093/nar/gkae424
- [19] Wellnitz J, Martin HJ, Hossain MA, Rath M, Fox C, Popov KI, et al. STOPLIGHT: A hit scoring calculator. *J Chem Inf Model*. 2024;64(11):4387-91. doi:10.1021/acs.jcim.4c00412
- [20] Ringer JA, Lambert CG, Bradfute SB, Bologna CG, Yang JJ. Badapple 2.0: An empirical predictor of compound promiscuity, updated, modernized, and enhanced for explainability. *J Chem Inf Model*. 2025;65(23):12641-47. doi:10.1021/acs.jcim.5c02297
- [21] Yang ZY, Dong J, Yang ZJ, Lu AP, Hou TJ, Cao DS. Structural analysis and identification of false positive hits in luciferase-based assays. *J Chem Inf Model*. 2020;60(4):2031-43. doi:10.1021/acs.jcim.9b01188
- [22] Yang ZY, Yang ZJ, Dong J, Wang LL, Zhang LX, Ding JJ, et al. Structural analysis and identification of colloidal aggregators in drug discovery. *J Chem Inf Model*. 2019;59(9):3714-26. doi:10.1021/acs.jcim.9b00541
- [23] Santillo MF, Glenn IS, Paris L, Sprando RL. In vitro biological target screening and colloidal aggregation of minor cannabinoids. *J Nat Prod*. 2024;87(12):2881-6. doi:10.1021/acs.jnatprod.4c00608
- [24] Borrel A, Huang R, Sakamuru S, Xia M, Simeonov A, Mansouri K, et al. High-throughput screening to predict chemical-assay interference. *Sci Rep*. 2020;10(1):3986. doi:10.1038/s41598-020-60747-3
- [25] Bashore FM, Annor-Gyamfi J, Du Y, Katis V, Nwogbo F, Flax RG, et al. Fused tetrahydroquinolines are interfering with your assay. *J Med Chem*. 2023;66(21):14434-46. doi:10.1021/acs.jmedchem.3c01277
- [26] Gerstberger T, Berger H, Büttner FH, Gmachl M, Kessler D, Koegl M, et al. Chasing red herrings: Palladium metal salt impurities feigning KRAS activity in biochemical assays. *J Med Chem*. 2024;67(14):11701-11. doi:10.1021/acs.jmedchem.3c02381
- [27] Reinecke M, Brear P, Vornholz L, Berger BT, Seefried F, Wilhelm S, et al. Chemical proteomics reveals the target landscape of 1,000 kinase inhibitors. *Nat Chem Biol*. 2024;20(5):577-85. doi:10.1038/s41589-023-01459-3
- [28] Brauer NR, Kempen AL, Hernandez D, Sintim HO. Non-kinase off-target inhibitory activities of clinically-relevant kinase inhibitors. *Eur J Med Chem*. 2024;275:116540. doi:10.1016/j.ejmech.2024.116540
- [29] St John-Campbell S, Bhalay G. Target engagement assays in early drug discovery. *J Med Chem*. 2025;68(12):12331-68. doi:10.1021/acs.jmedchem.4c03115
- [30] Tu Y, Tan L, Tao H, Li Y, Liu H. CETSA and thermal proteome profiling strategies for target identification and drug discovery of natural products. *Phytomedicine*. 2023;116:154862. doi:10.1016/j.phymed.2023.154862
- [31] Pan X, Jiang S, Zhang X, Wang Z, Wang X, Cao L, et al. Recent strategies in target identification of natural products: Exploring applications in chronic inflammation and beyond. *Br J Pharmacol*. 2025;182(20):4841-60. doi:10.1111/bph.17356
- [32] Gao Y, Ma M, Li W, Lei X. Chemoproteomics, a broad avenue to target deconvolution. *Adv Sci (Weinh)*. 2024;11(8). doi:10.1002/advs.202305608
- [33] Wozniak JM, Li W, Governa P, Chen LY, Jadhav A, Dongre A, et al. Enhanced mapping of small-molecule binding sites in cells. *Nat Chem Biol*. 2024;20(7):823-34. doi:10.1038/s41589-023-01514-z



- [34] Valerie NCK, Sanjiv K, Mortusewicz O, Zhang SM, Alam S, Pires MJ, et al. Coupling cellular drug-target engagement to downstream pharmacology with CeTEAM. *Nat Commun.* 2024;15(1):10347. doi:10.1038/s41467-024-54415-7
- [35] Boldini D, Friedrich L, Kuhn D, Sieber SA. Machine learning assisted hit prioritization for high throughput screening in drug discovery. *ACS Cent Sci.* 2024;10(4):823-32. doi:10.1021/acscentsci.3c01517
- [36] Liu TT, Zeng KW. Recent advances in target identification technology of natural products. *Pharmacol Ther.* 2025;269:108833. doi:10.1016/j.pharmthera.2025.108833