



# Assay Interference in Natural Product Screening and Its Consequences for Lead Discovery, Bioactivity Modeling, and Candidate Prioritization

Sofía Martínez<sup>1\*</sup>, Carlos Gómez<sup>1</sup>, Lucia Navarro<sup>2</sup>

<sup>1</sup>Department of Assay Interference and Hit Validation, Faculty of Pharmacy, Polytechnic University of Madrid, Madrid, Spain.

<sup>2</sup>Department of Natural Product Screening Informatics, Faculty of Pharmacy, University of Barcelona, Barcelona, Spain.

## ABSTRACT

Natural products are exceptionally rich sources of chemically diverse bioactive compounds, yet their screening outputs are frequently misinterpreted due to various technical and biological interferences that confound measured signals. This challenge has expanded beyond experimental hit confirmation, as screening data now routinely feed into bioactivity databases, machine-learning models, virtual-screening platforms, and compound-prioritization pipelines. This critical review systematically examines major sources of distortion, including optical interference, reporter perturbation, redox activity, colloidal aggregation, nonspecific reactivity, promiscuous inhibition, residual complexity, impurities, and mixture effects, all of which can compromise the evidence used to nominate natural-product leads. Furthermore, it explores how such distortions propagate into computational frameworks when critical metadata—such as assay provenance, detection technology, purity levels, counter-screening results, and orthogonal validation—are inadequately documented. The central thesis posits that an activity measurement should be treated as an evidence-bearing observation rather than an automatically mechanism-qualified label. Accordingly, the review proposes a graded distinction among apparent activity, interference-qualified activity, orthogonally supported pharmacology, and evidence-qualified lead status, while avoiding the binary assumption that all structurally suspicious or promiscuous compounds are artifactual. It also evaluates computational interference filters, assay-aware modeling, data curation, virtual-screening ranking, and validation practices as complementary safeguards. The primary implication is methodological: candidate prioritization must reflect the quality and provenance of the activity evidence underpinning computational predictions. Key limitations persist, including assay-specific behaviors, chemical-space dependencies, genuine polypharmacology, incomplete interference coverage, and the lack of prospective data quantifying ranking errors specifically attributable to assay interference.

**Key Words:** Natural products, Assay interference, Bioactivity modeling, Virtual screening, Data curation, Candidate prioritization

eIJPPR 2025; 15(1):96-105

**HOW TO CITE THIS ARTICLE:** Martínez S, Gómez C, Navarro L. Assay Interference in Natural Product Screening and Its Consequences for Computational Lead Discovery, Bioactivity Modeling, and Candidate Prioritization. Int J Pharm Phytopharmacol Res. 2025;15(1):96-105. <https://doi.org/10.51847/I2YFtbhLxc>

## INTRODUCTION

Natural-product screening is vulnerable to a basic interpretive error: an instrument or reporter can register a reproducible signal even when the compound has not produced the biological mechanism that the screen was

intended to measure. Small molecules can perturb assay readouts through optical effects, colloidal behavior, chemical reactivity, reporter inhibition, redox chemistry, and other context-dependent mechanisms, so experimental activity and target-relevant pharmacology are not interchangeable categories [1]. For natural products, this

**Corresponding author:** Sofía Martínez

**Address:** Department of Assay Interference and Hit Validation, Faculty of Pharmacy, Polytechnic University of Madrid, Madrid, Spain

**E-mail:** ✉ sofia.martinez@upm.es

**Received:** 19 November 2024; **Revised:** 27 February 2025; **Accepted:** 28 February 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.



distinction is especially consequential because extracts, fractions, isolated metabolites, and structurally complex compounds often enter workflows in which chemical identity, purity, matrix composition, and assay compatibility are established with unequal depth.

The problem is increasingly computational. Artificial intelligence and related data-driven methods are now used across natural-product discovery for chemical annotation, dereplication, activity prediction, prioritization, and integration of heterogeneous experimental information [2]. At the same time, high-throughput screening data can contain recurrent artifact liabilities that are sufficiently systematic to support computational prediction [3]. These developments create a two-sided dependency: computation can help recognize suspect assay behavior, but computation can also learn from assay-derived labels whose biological meaning is uncertain. A false-positive observation is therefore not only an experimental inconvenience. If retained as a positive activity label, it can become training evidence, similarity evidence, benchmark evidence, or a ranking target.

Even nominally comparable potency values can acquire substantial noise when measurements from different experimental sources are combined without sufficient regard to assay provenance [4]. This does not mean that cross-study bioactivity integration is intrinsically invalid; protocol-dependent differences may reflect genuine biology as well as measurement error. The critical issue is that computational pipelines frequently compress several evidential states into one numerical endpoint. This review argues that natural-product lead discovery should instead distinguish the existence of an assay signal from the strength of evidence that the signal represents a chemically attributable, biologically meaningful, and computationally reusable activity. That distinction frames assay interference as a data-quality and inference problem extending from the plate reader to downstream candidate ranking.

#### *Review scope and critical approach*

This review focuses on assay interference where it materially affects the interpretation or computational reuse of natural-product screening results. Natural products remain important sources of pharmacologically relevant scaffolds and mechanisms, but their discovery workflows also involve distinctive challenges in sourcing, isolation, dereplication, structural characterization, resupply, and biological evaluation [5]. Accordingly, the scope includes purified natural products, fractions and biologically derived mixtures where the interference mechanism is directly relevant, while drawing on broader small-molecule screening literature when it establishes a

mechanism or validation principle transferable to natural-product discovery.

Bioassay choice is treated as part of evidence generation rather than as a neutral measurement step. Recent guidance for exploration of natural resources emphasizes that different assay formats provide different types of biological information and require method-appropriate interpretation [6]. The review therefore examines detection chemistry, reporter systems, aggregation, nonspecific chemistry, impurities, assay provenance, data curation, machine learning, virtual screening, and validation as connected but non-equivalent domains.

A second boundary concerns what counts as convincing activity. Chemical-probe guidance shows why potency alone is insufficient for strong mechanistic inference: selectivity, target engagement, cellular activity, controls, and orthogonal evidence can all affect whether a small molecule is fit for a particular biological claim [7]. Those standards cannot simply be transplanted into every natural-product lead program, but the underlying evidential principle is applicable. This review does not estimate a universal frequency of false positives, does not treat all promiscuity as interference, and does not assume that an interference alert invalidates a compound. Its purpose is to evaluate how different sources of uncertainty alter the evidential status of screening results and how that status should constrain downstream computational use.

#### *Sources and mechanisms of assay interference*

Assay interference is better understood as a family of mechanisms than as a single compound property. Frequent-hitter behavior may arise from colloidal aggregation, chemical reactivity, optical or reporter effects, nonspecific interactions, and other assay-dependent liabilities, while experimental datasets and modern computational tools can identify or prioritize several of these risks [8–10]. The practical consequence is that a compound can appear problematic in one assay architecture yet behave acceptably in another, and two compounds producing similar hit patterns may do so for entirely different reasons.

This mechanism plurality is important for natural products because chemically rich scaffolds are often assessed using sensitive readouts at concentrations where solubility, redox behavior, absorbance, fluorescence, or matrix effects may become relevant. A structural alert can therefore be useful as a triage signal, but it cannot establish which mechanism, if any, actually generated an observed response. Conversely, the absence of a known alert cannot certify assay validity because interference models are limited by their training data and by the mechanisms represented in those data.

A critical workflow should therefore separate three questions. First, is the primary signal analytically credible within the measurement technology? Second, is the observed response attributable to the intended compound or chemically defined sample rather than aggregation, degradation, reactive species, or minor constituents? Third, does the response survive a biologically meaningful change in assay principle? This proposed separation does not presume that every hit requires every possible counter-screen. It instead makes the burden of validation proportional to the intended downstream claim: a preliminary screening hit requires less evidence than a mechanistic label that will be reused for model training, virtual screening, or lead prioritization.

#### *Optical, redox, fluorescent, and reporter-specific artefacts*

Optical interference is particularly relevant to natural-product screening because many metabolites and complex extracts possess intrinsic chromophores or fluorophores. In a direct experimental study, medicinal-plant extracts and fractions produced signals capable of interfering with fluorescence-based measurements used to assess gastric cancer stem-cell phenotypes, including resazurin, confocal microscopy, and flow-cytometric readouts [11]. Such findings do not imply that fluorescence assays are unsuitable for natural products; they show that a biologically plausible signal can be contaminated when sample emission, absorbance, quenching, or spectral overlap is not separated from reporter-dependent signal.

Some optical liabilities are sufficiently associated with chemical structure to support computational prediction of autofluorescence and luminescence interference [12]. These tools are useful for prospective triage, particularly when large libraries make universal counterscreening impractical, but their outputs remain risk estimates rather than experimental diagnoses. Natural-product chemical space, conjugated systems, pigments, oxidation products, and mixtures can extend beyond the domains in which such models were developed.

Reporter interference is also platform-specific. Comparative work with firefly, Renilla, and NanoLuc luciferases shows that metal ions can influence bioluminescent signals differently across reporter systems [13]. The broader lesson is that “reporter interference” should not be encoded as a context-free property of a compound or sample. Reporter chemistry, substrate, concentration, medium, timing, wavelength, and detection geometry can alter the observed effect. Redox-active compounds introduce a related interpretive problem because they may perturb coupled detection chemistry or

cellular redox-sensitive readouts without necessarily engaging the intended target. For computational discovery, the appropriate abstraction is therefore not simply active versus inactive, but activity linked to the assay technology and to evidence that the measured signal remains after plausible detection-specific artefacts have been excluded.

#### *Aggregation, nonspecific reactivity, and promiscuous inhibition*

Colloidal aggregation is a particularly difficult interference mechanism because it can produce concentration-dependent inhibition that resembles ordinary pharmacology. Aggregates can sequester or perturb proteins nonspecifically, and the resulting activity may survive superficial reproducibility checks. Importantly, this liability is not confined to purified-protein assays: colloidal aggregation has been shown to confound cell-based antiviral screening, demonstrating that cellular context does not automatically eliminate aggregate-mediated false activity [14]. Computational filters can improve recognition of molecules with aggregation liability across broader chemical space [15], while experimentally characterized nuisance-compound collections provide a complementary way to challenge assay robustness rather than relying on prediction alone [16].

Aggregation nevertheless should not become a catch-all explanation for difficult natural-product hits. Poor solubility, precipitation, membrane perturbation, covalent or redox reactivity, cytotoxicity, and genuine multi-target pharmacology can generate superficially similar patterns. Likewise, promiscuous activity across assays is an observation, not a mechanism. A natural product acting at several biologically related targets may be pharmacologically interesting rather than artifactual. The discriminating question is whether the activity survives changes that should disrupt the suspected nuisance mechanism while preserving the intended biology.

For downstream computation, this distinction matters because a promiscuous or aggregate-prone compound may contribute many positive labels to a database. Treating those labels as independent confirmation can amplify one chemical liability into apparently broad biological evidence. Conversely, indiscriminate removal of every structurally suspicious compound can erase real pharmacology. The appropriate computational response is therefore evidence qualification rather than automatic exclusion.

The data-quality consequences of the interference mechanisms discussed above are summarized in **Table 1**.

**Table 1.** From assay disturbance to data-quality failure in natural-product screening

| Interference state                     | Immediate observation                                  | Data-quality risk                                 | Downstream computational consequence           | Required qualification                                          |
|----------------------------------------|--------------------------------------------------------|---------------------------------------------------|------------------------------------------------|-----------------------------------------------------------------|
| <b>Optical/reporter disturbance</b>    | Signal changes without secure biological attribution   | Readout mislabeled as activity                    | False-positive training or ranking evidence    | Detection-compatible control or orthogonal readout              |
| <b>Colloidal aggregation</b>           | Reproducible, often concentration-dependent inhibition | Nonspecific inhibition encoded as target activity | Inflated target or scaffold association        | Aggregation-sensitive counterscreen                             |
| <b>Reactive/nonspecific chemistry</b>  | Broad response across susceptible assay components     | Mechanism ambiguity                               | Apparent promiscuity becomes biological signal | Chemistry-appropriate control and orthogonal assay              |
| <b>Recurrent promiscuity</b>           | Activity across multiple assays                        | Correlated labels may be mistaken for replication | Overweighted compound evidence                 | Separate empirical history from mechanism                       |
| <b>Structural liability prediction</b> | Computed interference or aggregation risk              | Prediction may be treated as fact                 | Premature compound deletion                    | Confirm experimentally; retain applicability-domain uncertainty |

Note: The final column represents a proposed evidence-qualification principle; it is not a validated universal screening rule.

### *Natural product complexity, impurities, and mixture effects*

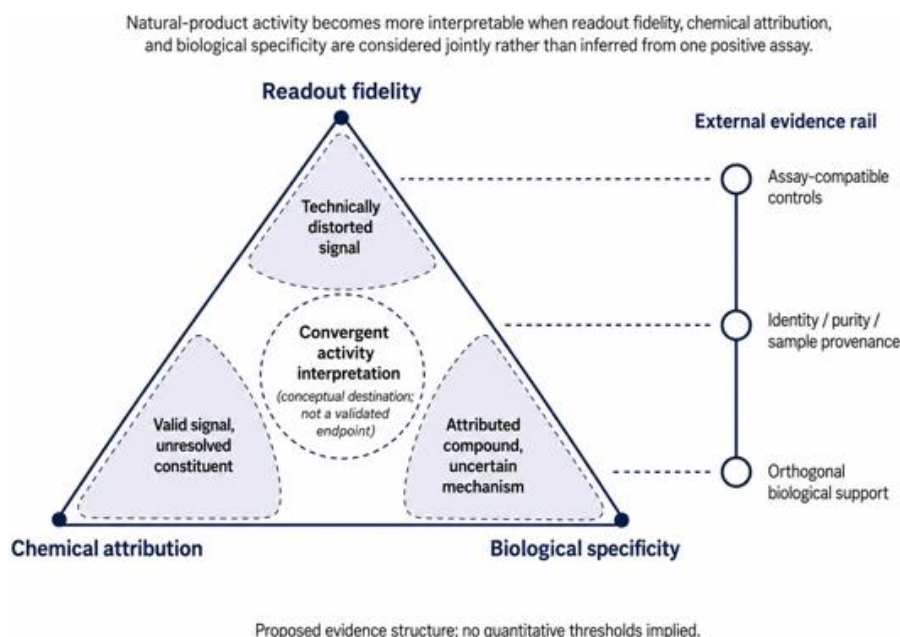
Natural-product samples introduce an additional attribution problem: the entity placed in a well is not always chemically equivalent to the structure later assigned the activity. Extracts contain many constituents, fractions retain mixtures, isolated materials may carry residual congeners, and nominally purified compounds can contain minor components with disproportionately strong biological effects. The rufomycin/rufomyazine case demonstrates the consequence directly: residual complexity altered the assignment of biological activity, showing that a trace constituent can determine the apparent pharmacology attributed to a natural-product sample [17]. Such examples establish possibility and mechanism, not a general prevalence estimate.

Mixture complexity can also distort screening efficiency without constituting classical assay interference. Closely related extracts may repeatedly contribute the same chemistry, while minor active constituents may be distributed unevenly across fractions. Mass-spectrometry-guided rationalization of natural-product libraries shows

that chemical information can be used to reduce redundancy while retaining useful biological screening space [18]. This analytical strategy addresses library composition rather than validating a particular hit, but it illustrates why computational prioritization benefits from sample-level chemical provenance.

The critical distinction is therefore between signal validity and chemical attribution. A fluorescence signal may be technically valid yet originate from the wrong constituent; a purified compound may be correctly identified yet act through aggregation; and a complex fraction may contain genuine target activity whose responsible molecule remains unresolved. Natural-product screening needs these uncertainties represented separately. Collapsing extract identity, nominal compound identity, purity, and activity into one record creates an evidential shortcut that later models cannot reconstruct.

**Figure 1** expresses the manuscript-derived relationship among readout fidelity, chemical attribution, biological specificity, and the evidence needed to move beyond an initial natural-product screening signal.



**Figure 1.** Triangulating attribution confidence in natural-product screening

#### *How interference contaminates bioactivity datasets and models*

Assay interference becomes a computational problem when an uncertain experimental observation is stored as if it were a context-free molecular property. Major bioactivity resources preserve diverse assay types, measurement formats, targets, and provenance, demonstrating that activity data are intrinsically heterogeneous rather than generated by one standardized experimental process. Data-curation confidence and assay context can materially alter machine-learning predictions, and recent work shows that assay information itself can be incorporated into bioactivity models [19–21]. The relevant question is therefore not whether public bioactivity databases are unreliable, but whether downstream modeling retains enough experimental context to distinguish comparable labels from superficially similar ones.

A useful way to describe the failure is evidence compression. In the proposed sense used here, evidence compression occurs when a multidimensional observation—compound identity, sample purity, assay technology, concentration, endpoint, interference controls, target context, and validation status—is reduced to an activity value and a binary or continuous model label. Cross-source comparison of nominally equivalent IC<sub>50</sub> or K<sub>i</sub> measurements demonstrates that combining measurements without sufficient source distinction can create substantial noise [4]. Interference adds another possible contributor to this heterogeneity because the measured endpoint may not represent the intended biological process.

Three computational consequences follow. First, misassigned positives can shift learned structure–activity relationships toward optical, aggregating, reactive, or otherwise nuisance-prone chemistry. Second, repeated records derived from related assay technologies can create apparent replication without independent mechanistic confirmation. Third, model confidence can become disconnected from evidence confidence: a model may predict a noisy label accurately because the noise itself is systematic.

A further proposed distinction is between label uncertainty and molecular uncertainty. The former concerns whether the recorded experimental outcome means what the dataset says it means; the latter concerns whether a model can generalize from molecular features to that outcome. These uncertainties can interact but should not be merged. A highly expressive model cannot correct a systematically misinterpreted endpoint simply by fitting it more accurately, while aggressive label cleaning cannot solve extrapolation into unfamiliar natural-product chemistry. Evaluation should therefore report both evidence quality and predictive uncertainty whenever the available metadata permit.

The proposed remedy is not to discard all uncertain observations. Dataset construction should preserve provenance and, where feasible, distinguish unqualified primary activity from activity that has survived relevant interference controls or orthogonal validation. Assay-aware modeling provides one route for representing such context, but it does not repair missing metadata or retrospectively establish mechanism. Evidence qualification therefore belongs upstream of model fitting as well as inside model architecture.

*Consequences for virtual screening and candidate ranking*  
 Candidate ranking converts imperfect evidence into an allocation decision: which structures receive synthesis, isolation, resupply, mechanistic experiments, or medicinal-chemistry effort. Structure-based virtual screening already illustrates that false-positive burden can be modified computationally; machine-learning classification has reduced false positives in evaluated docking workflows [22]. This shows that ranking is not a passive reflection of molecular affinity hypotheses. It depends on how positives, negatives, decoys, and features are constructed. The same caution applies to learned scoring functions. Robust interrogation of machine-learning-based scoring has shown that performance can depend on learnable dataset similarities and benchmark properties rather than exclusively on transferable protein–ligand recognition [23]. In an interference-sensitive natural-product pipeline, an analogous risk arises when nuisance-associated chemical features correlate with positive assay labels. A

model can then elevate compounds for reproducing the historical labeling process rather than for possessing better validated pharmacology.

This does not justify replacing potency ranking with an interference penalty. A strong filter can wrongly demote genuine polypharmacology, unusual scaffolds, or chemistry outside its applicability domain. Instead, candidate ranking should separate predicted biological promise from confidence in the evidence that trained or supports that prediction. The resulting ranking problem is multi-evidential: chemical novelty, predicted activity, assay provenance, interference liability, analytical attribution, and validation status may all matter, but their appropriate weighting remains a prospective research question.

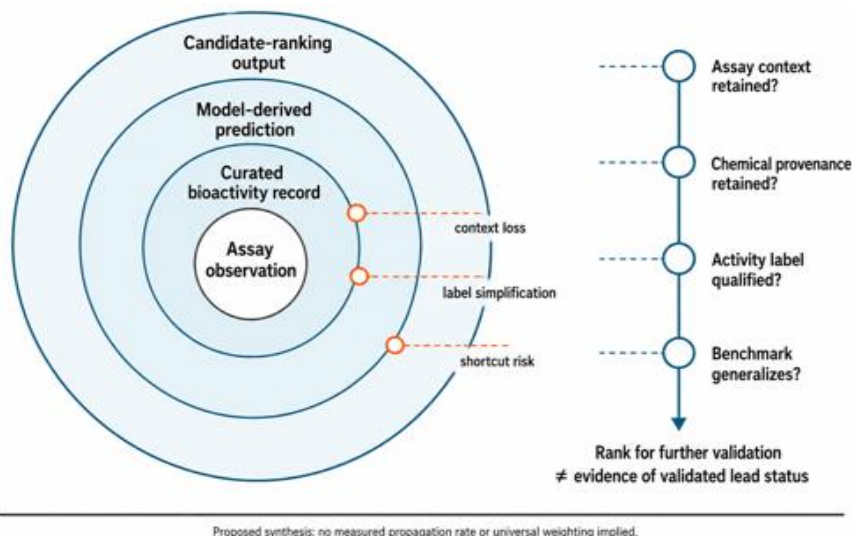
**Table 2** compares computational methods according to the specific decision problem they can address without treating any method as a substitute for experimental validation.

**Table 2.** Computational safeguards and their appropriate use in interference-sensitive candidate ranking

| Method                                   | Primary use case             | Useful input                   | Decision value                          | Principal boundary                  |
|------------------------------------------|------------------------------|--------------------------------|-----------------------------------------|-------------------------------------|
| Interference/structural-liability filter | Early triage                 | Molecular structure            | Flags assay-risk chemistry              | Risk is not proof                   |
| Aggregation predictor                    | Aggregator prioritization    | Structure/descriptors          | Selects candidates for counterscreening | Domain dependent                    |
| Assay-aware bioactivity model            | Context-sensitive prediction | Molecule plus assay metadata   | Separates some assay-dependent signal   | Cannot restore missing provenance   |
| False-positive classifier                | Virtual-screen refinement    | Docking/structural features    | Reorders likely screening artefacts     | Target and benchmark dependent      |
| Scoring-function interrogation           | Model audit                  | Benchmark/similarity structure | Detects shortcut-learning risk          | Does not establish biological truth |

Note: The table defines complementary use cases derived from the reviewed evidence; it does not prescribe a validated universal ranking algorithm.

**Figure 2** maps where evidential context can be lost as an assay result moves toward a computational ranking decision.



**Figure 2.** Evidence distance from assay observation to computational priority

*Counter-screens, orthogonal validation, and data curation*  
Interference control is strongest when the follow-up experiment changes the property that could have generated the original signal. Orthogonal reassessment of reported histone acetyltransferase inhibitors has shown that compounds accepted as inhibitors can display assay interference and off-target liabilities when examined with complementary biochemical, biophysical, and cellular methods [24]. The lesson is not that every natural-product hit needs an identical validation panel, but that repeating the same detection principle with minor procedural changes provides weaker evidence than testing the claim through a genuinely different measurement route.

The appropriate counter-screen depends on the suspected failure mode. Fluorescent samples require controls that separate sample-derived signal from reporter-derived signal; suspected aggregators require conditions or measurements sensitive to colloidal behavior; complex natural-product samples require identity, purity, or fraction-level evidence; and mechanism claims require a biologically meaningful orthogonal test. Negative results also need interpretation: failure in one assay can reflect permeability, stability, target conformation, concentration, or protocol differences rather than proving that the primary result was artifactual.

Data curation should capture this reasoning rather than retaining only the final activity number. A systematic review of chemical-probe use found substantial shortcomings in probe selection and control practice in cell-based biomedical research, illustrating that experimentally convenient compounds can acquire strong biological interpretations without commensurate qualification [25]. For computational datasets, curation should therefore distinguish reported activity from evidential status. A record can remain valuable even when uncertain, provided its assay provenance, sample identity, controls, and validation state are visible. Deletion removes information; qualification preserves both the observation and the reason for caution.

#### *Critical synthesis: from apparent activity to evidence-qualified leads*

The reviewed evidence supports a distinction between detecting activity and deciding how much inferential weight that activity should carry. Computational methods can help identify difficult interference classes: expert-guided augmentation has been proposed to improve classification of scarce and imbalanced assay-interference data [26], while empirically derived promiscuity models use accumulated assay histories to estimate how broadly compounds have appeared active [27]. These approaches can enrich triage, but neither an interference prediction nor

a promiscuity estimate establishes why an individual natural product was active.

This review therefore proposes a four-state evidence qualification continuum. Apparent activity denotes a reproducible primary readout with unresolved interference and attribution. Interference-qualified activity denotes an observation that has survived controls appropriate to plausible assay liabilities and has adequate chemical identity or sample provenance. Orthogonally supported pharmacology adds evidence from a distinct assay principle supporting the intended biological interpretation. Evidence-qualified lead status adds sufficient chemical, biological, and computational provenance to justify preferential advancement. These are proposed evidential states, not validated thresholds, and a project may legitimately advance a compound before every state is satisfied if the decision is explicitly exploratory.

The continuum also prevents the opposite error of overfiltering. Expert-reviewed chemical-probe resources demonstrate the value of integrating potency, selectivity, target engagement, controls, and independent evidence when judging small-molecule fitness for a biological purpose [28]. Natural-product lead discovery requires its own context-specific standards, but the transferable principle is convergence. No single structural alert, counterscreen, potency estimate, docking score, or model probability should dominate simply because it is easy to compute.

Candidate prioritization can thus be reframed as ranking under evidence uncertainty. The strongest candidate is not necessarily the compound with the largest primary signal or highest predicted score; it is the candidate whose apparent promise remains persuasive after uncertainty in assay behavior, chemical attribution, data provenance, and model generalization has been made visible. Prospective testing is required before this proposed approach can be treated as superior to conventional ranking.

#### *Research priorities and reporting standards*

The first priority is to preserve experimental context at the point where bioactivity becomes reusable data. The continuing role of ChEMBL in cheminformatics illustrates the value of curated, structured bioactivity information and sustained attention to data provenance [29]. For natural-product records, useful context extends beyond target and potency to sample type, isolation state, purity or compositional evidence, assay technology, reporter, concentration regime, interference controls, and whether orthogonal confirmation was performed. Not every historical record will contain these fields, but absence should be represented as missing evidence rather than silently converted into confidence.

Second, interference-aware modeling requires benchmarks that test more than random interpolation. Validation guidance for drug–target prediction emphasizes stronger computational evaluation and orthogonal experimental confirmation [30]. Applied here, future studies should compare scaffold-aware and assay-held-out splits, external chemical spaces, calibrated uncertainty, and prospective rankings generated before confirmation. A decisive experiment would compare models trained on minimally curated activity labels with models trained on evidence-qualified labels while holding chemical space and target task as constant as practical.

Third, natural-product pipelines need paired analytical and biological validation. Extract-to-fraction-to-compound lineage should remain machine-readable so that a model does not treat activity associated with a mixture as though it were established for an isolated structure. Where interference predictions are used, reports should identify the model, applicability domain, version, threshold rationale, and follow-up test rather than merely stating that compounds “passed” a filter.

Finally, negative and contradictory results should be retained. A compound active in one reporter system but inactive in an orthogonal assay contains information about both the molecule and the measurement process. Such discordance is not failed data; it is precisely the evidence needed to distinguish transferable pharmacology from assay-contingent signal.

#### Limitations

This review integrates several evidence traditions that do not have identical natural-product coverage. Direct demonstrations exist for natural-product optical interference and residual complexity, but much mechanistic and computational evidence derives from broader small-molecule screening. Transfer to natural-product chemical space is therefore reasoned and bounded rather than universally established.

A second limitation concerns structural nuisance alerts. Large-scale analyses show that PAINS-based flags can have limited specificity for experimentally relevant interference, so a structural alert should not be treated as a mechanistic verdict [31]. Conversely, this criticism does not establish that alerted structures are harmless. Extensively screened PAINS-containing compounds display heterogeneous activity profiles, including many compounds that are not globally frequent hitters [32]. The evidence therefore supports neither blanket rejection nor blanket acceptance.

The proposed evidence-qualification continuum and evidence-sensitive ranking interpretation have not been prospectively validated. Their value is organizational: they separate assay signal, chemical attribution, biological

support, and computational reuse. The review also does not estimate the prevalence of interference in natural-product libraries or the fraction of model error attributable to contaminated labels. Those questions require purpose-built prospective datasets with linked assay, analytical, counterscreen, and modeling information.

## CONCLUSION

Assay interference should be treated as a discovery-inference problem rather than a late-stage nuisance confined to hit confirmation. In natural-product screening, uncertainty can enter through the measurement technology, the chemical identity of the tested material, nonspecific physicochemical behavior, mixture complexity, or the biological interpretation of a reproducible signal. Once converted into a bioactivity label, these uncertainties can travel into computational models and candidate rankings even when the calculations themselves are technically correct.

The principal contribution of this review is a proposed evidence-sensitive view of activity: experimental signal, chemical attribution, interference control, orthogonal pharmacology, and computational reuse should remain distinguishable. This approach does not justify automatic rejection of promiscuous, structurally flagged, or unconventional natural products, nor does it establish a universal ranking formula. It instead makes uncertainty explicit before scarce discovery resources are allocated. Progress will depend on richer assay provenance, analytical traceability, mechanism-appropriate counter-screens, context-aware modeling, and prospective validation of whether evidence-qualified labels actually improve lead prioritization across diverse natural-product chemical space.

**Acknowledgments:** None

**Conflict of interest:** None

**Financial support:** None

**Ethics statement:** None

## REFERENCES

- [1] Tan L, Hirte S, Palmacci V, Stork C, Kirchmair J. Tackling assay interference associated with small molecules. *Nat Rev Chem.* 2024;8(5):319-39. doi:10.1038/s41570-024-00593-3
- [2] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hoof JJJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. *Nat*



- Rev Drug Discov. 2023;22(11):895-916. doi:10.1038/s41573-023-00774-7
- [3] 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
- [4] Landrum GA, Riniker S. Combining IC50 or Ki values from different sources is a source of significant noise. *J Chem Inf Model.* 2024;64(5):1560-7. doi:10.1021/acs.jcim.4c00049
- [5] 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
- [6] Sabotič J, Bayram E, Ezra D, Gaudêncio SP, Haznedaroğlu BZ, Janež N, et al. A guide to the use of bioassays in exploration of natural resources. *Biotechnol Adv.* 2024;71:108307. doi:10.1016/j.biotechadv.2024.108307
- [7] Mader MM, Rudolph J, Hartung IV, Uehling D, Workman P, Zuercher W. Which small molecule? Selecting chemical probes for use in cancer research and target validation. *Cancer Discov.* 2023;13(10):2150-65. doi:10.1158/2159-8290.CD-23-0536
- [8] 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
- [9] 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
- [10] 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):W439-W49. doi:10.1093/nar/gkae424
- [11] Tapia-Rojas S, García-Paitán M, Del Rosario-Chavarri J, Santiani A, Alvarez-Vega S, Amiel-Pérez J, et al. Medicinal plant extracts interfere in gastric cancer stem cells fluorescence-based assays. *Saudi J Biol Sci.* 2024;31(6):104000. doi:10.1016/j.sjbs.2024.104000
- [12] Borrel A, Mansouri K, Nolte S, Saddler T, Conway M, Schmitt C, et al. InterPred: A webtool to predict chemical autofluorescence and luminescence interference. *Nucleic Acids Res.* 2020;48(W1):W586-W90. doi:10.1093/nar/gkaa378
- [13] Canyelles i Font F, Żukowski K, Khan MA, Kwiatek D, Kolanowski JL. Interference of metal ions on the bioluminescent signal of firefly, Renilla, and NanoLuc luciferases in high-throughput screening assays. *Front Chem.* 2024;12:1436389. doi:10.3389/fchem.2024.1436389
- [14] Glenn IS, Hall LN, Khalid MM, Ott M, Shoichet BK. Colloidal aggregation confounds cell-based COVID-19 antiviral screens. *J Med Chem.* 2024;67(12):10263-74. doi:10.1021/acs.jmedchem.4c00597
- [15] Abou Hajal A, Bryce RA, Ben Amor B, Atatreh N, Ghattas MA. Boosting the accuracy and chemical space coverage of the detection of small colloidal aggregating molecules using the BAD molecule filter. *J Chem Inf Model.* 2024;64(13):4991-5005. doi:10.1021/acs.jcim.4c00363
- [16] 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
- [17] Choules MP, Klein LL, Lankin DC, McAlpine JB, Cho SH, Cheng J, et al. Residual complexity does impact organic chemistry and drug discovery: The case of rufomyazine and rufomycin. *J Org Chem.* 2018;83(12):6664-72. doi:10.1021/acs.joc.8b00988
- [18] Ness M, Peramuna T, Wendt KL, Collins JE, King JB, Paes R, et al. Rationally minimizing natural product libraries using mass spectrometry. *mSystems.* 2025;10(3):e0084424. doi:10.1128/msystems.00844-24
- [19] Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* 2024;52(D1):D1180-D92. doi:10.1093/nar/gkad1004
- [20] Xerxa E, Vogt M, Bajorath J. Influence of data curation and confidence levels on compound predictions using machine learning models. *J Chem Inf Model.* 2024;64(24):9341-9. doi:10.1021/acs.jcim.4c01573
- [21] Schoenmaker L, Sastrokarijo EG, Heitman LH, Beltman JB, Jespers W, van Westen GJP. Toward assay-aware bioactivity model(er)s: Getting a grip on biological context. *J Chem Inf Model.* 2025;65(13):7013-23. doi:10.1021/acs.jcim.5c00603
- [22] 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
- [23] Durant G, Boyles F, Birchall K, Marsden B, Deane CM. Robustly interrogating machine learning-based

- scoring functions: What are they learning? *Bioinformatics*. 2025;41(2):btaf040. doi:10.1093/bioinformatics/btaf040
- [24] Dahlin JL, Nelson KM, Strasser JM, Barsyte-Lovejoy D, Szewczyk MM, Organ S, et al. Assay interference and off-target liabilities of reported histone acetyltransferase inhibitors. *Nat Commun*. 2017;8(1):1527. doi:10.1038/s41467-017-01657-3
- [25] Sterling J, Baker JR, McCluskey A, Munoz L. Systematic literature review reveals suboptimal use of chemical probes in cell-based biomedical research. *Nat Commun*. 2023;14(1):3228. doi:10.1038/s41467-023-38952-1
- [26] Palmacci V, Nahal Y, Welsch M, Engkvist O, Kaski S, Kirchmair J. E-GuARD: Expert-guided augmentation for the robust detection of compounds interfering with biological assays. *J Cheminform*. 2025;17(1):64. doi:10.1186/s13321-025-01014-3
- [27] Ringer JA, Lambert CG, Bradfute SB, Bologa 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):12641-7. doi:10.1021/acs.jcim.5c02297
- [28] Sanfelice D, Antolin AA, Crisp A, Chen Y, Bellenie B, Brennan PE, et al. The Chemical Probes Portal – 2024: Update on this public resource to support best-practice selection and use of small molecules in biomedical research. *Nucleic Acids Res*. 2025;53(D1):D1663-D9. doi:10.1093/nar/gkae1062
- [29] Zdrazil B. Fifteen years of ChEMBL and its role in cheminformatics and drug discovery. *J Cheminform*. 2025;17(1):32. doi:10.1186/s13321-025-00963-z
- [30] 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
- [31] 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
- [32] 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