



Spectral Confidence Should Travel with Every Computationally Annotated Natural Product from Raw Signal Interpretation to Structure Assignment and Pharmacological Modeling

Santiago Morales^{1*}, Laura Rojas², Camila Vega¹, Diego Molina³

¹Department of Spectral Confidence and Annotation, Faculty of Pharmacy, University of Buenos Aires, Buenos Aires, Argentina.

²Department of Structure Assignment and Pharmacological Modeling, Faculty of Pharmacy, Pontifical Catholic University of Chile, Santiago, Chile.

³Department of Raw Signal Interpretation for Natural Products, Faculty of Pharmacy, National University of La Plata, La Plata, Argentina.

ABSTRACT

Computational mass spectrometry increasingly converts complex natural-product signals into molecular formulas, chemical classes, candidate structures, and named database records that can subsequently enter pharmacological modeling. Yet the evidential status of a molecular assignment is often compressed during this transition, so that a structure inferred from an imperfect tandem spectrum may eventually be represented downstream in the same structural format as a compound established with substantially stronger analytical evidence. This article develops an original evidence framework for treating spectral confidence as a persistent property of the molecular record rather than a temporary annotation-stage score. The analysis separates observed spectral evidence from formula inference, substructure or class inference, candidate generation, candidate ranking, and full-structure assignment, and it argues that confidence should remain linked to the analytical state, computational operation, candidate space, and provenance from which a structure was produced. The proposed framework does not assume that existing confidence scores are interchangeable or that uncertain annotations should be discarded. Instead, it conceptualizes confidence as typed, revisable evidence that can accompany computationally annotated molecules into databases and downstream modeling while remaining distinguishable from predictive uncertainty generated by pharmacological models themselves. This approach could support more defensible reuse, targeted reanalysis, and task-dependent handling of tentative structures. Its principal limitation is that no universal confidence scale, weighting function, or downstream performance benefit has yet been established across chemical classes, instruments, annotation systems, or pharmacological modeling tasks.

Key Words: *Natural products, Tandem mass spectrometry, Structural annotation, Spectral confidence, Provenance, Molecular machine learning*

eIJPPR 2025; 15(6):134-143

HOW TO CITE THIS ARTICLE: Morales S, Rojas L, Vega C, Molina D. Spectral Confidence Should Travel with Every Computationally Annotated Natural Product from Raw Signal Interpretation to Structure Assignment and Pharmacological Modeling. Int J Pharm Phytopharmacol Res. 2025;15(6):134-43. <https://doi.org/10.51847/LoiDu7ws8j>

INTRODUCTION

Computational annotation has transformed untargeted mass spectrometry from a peak-detection exercise into a route toward chemically interpretable molecular records.

Corresponding author: Santiago Morales

Address: Department of Spectral Confidence and Annotation, Faculty of Pharmacy, University of Buenos Aires, Buenos Aires, Argentina.

E-mail: ✉ santiago.morales@agro.uba.ar

Received: 08 September 2025; **Revised:** 05 December 2025; **Accepted:** 09 December 2025

That transformation is powerful precisely because a spectrum can now be connected to formulas, structural classes, candidate structures, databases, and downstream computational models. The epistemic difficulty is that each transition adds inference. Analytical identification

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.



therefore depends not only on obtaining technically acceptable data but on maintaining a defensible relationship between the evidence observed and the chemical identity eventually reported [1]. A named structure can be syntactically complete while remaining evidentially provisional.

This distinction matters because disagreement is not hypothetical. In a multilaboratory exercise using a shared untargeted mass-spectrometry dataset, participating groups recovered materially different annotations and exposed bottlenecks arising from workflow and interpretive choices [2]. Such variation does not imply that computational annotation is unreliable as a category. It shows instead that a structure label does not erase the history through which that label was produced. Once exported, however, that history may become difficult to distinguish from the canonical molecular representation itself.

Recent work proposing an identification probability demonstrates that confidence can, under defined measurement precision and candidate-search conditions, be represented explicitly rather than left as an informal judgment [3]. The present article extends this logic in a different direction. It proposes that spectral confidence should become portable evidence provenance: a typed and revisable description of how strongly the available analytical evidence supports the molecular representation being reused. Portability does not mean forcing every algorithm into one numerical scale. It means preventing evidence state from disappearing when a molecule moves from raw spectral interpretation to a named record and, eventually, to structure-dependent pharmacological computation. The purpose is therefore not to redefine identification standards, but to preserve the difference between what was measured, what was inferred, and what downstream users are entitled to assume.

The journey from raw spectrum to named molecule

A computationally annotated molecule begins neither with a database name nor with a structural fingerprint, but with measurements that have already passed through acquisition and processing decisions. Modern platforms such as MZmine 3 illustrate how feature detection, alignment, mobility information, spectral processing, and downstream annotation can be integrated into a multimodal analytical workflow [4]. The processed spectrum is consequently an evidential object shaped by instrumental and computational operations rather than an untouched representation of molecular structure.

One particularly consequential complication is spectral chimerism. Coisolated precursors can contribute fragments to the same MS/MS event, producing evidence that appears chemically coherent while containing signals from more than one molecular source. Computational cleaning can separate components and improve the interpretability of such spectra [5]. The implication for confidence is narrower but important: uncertainty exists before candidate ranking begins. A high-scoring candidate cannot retrospectively demonstrate that its input spectrum was molecularly pure.

Search strategy can also change how mixed evidence is interpreted. Reverse spectral searching has been shown to provide a practical response to particular chimeric-annotation problems by focusing matching behavior differently from conventional forward comparison [6]. This does not make reverse searching a universal solution; library coverage, acquisition quality, isomerism, and other sources of ambiguity remain. It demonstrates instead that the path from spectrum to name includes methodological contingencies that should remain recoverable after annotation.

The constructs produced by this early signal-to-record transition are defined in **Table 1** because their evidential meanings are easily compressed when represented only as prose.

Table 1. Constructs that define the early evidence transition from measured spectrum to computational molecular record

Construct	Meaning in this framework	Evidential status	Principal vulnerability
Raw spectral state	Instrument-derived precursor and fragment evidence before interpretive assignment	Observed, but measurement-conditioned	Noise, coisolation, acquisition limits
Processed spectral evidence	Signals retained after feature detection, alignment, cleaning, or deconvolution	Observed evidence after computation	Processing-dependent inclusion or distortion
Spectral interpretation	Assignment of analytical meaning to peaks, patterns, or relationships	Inferred from measured evidence	Model, rule, and library dependence
Named molecular record	Formula, class, candidate, or structure represented as a reusable chemical entity	Inferential status depends on preceding evidence	Loss of analytical context during export
Portable confidence provenance	Proposed record of evidence state, inference level, and revisability accompanying the molecule	Proposed synthesis	Requires interoperable representation and validation

note: Evidence-derived distinctions summarize relationships established in the preceding section. "Portable confidence provenance" is proposed here and is not an existing validated standard.

The early trajectory is more faithfully represented as evidence-bearing transformations than as a simple spectrum-to-structure pipeline, as shown in **Figure 1**.

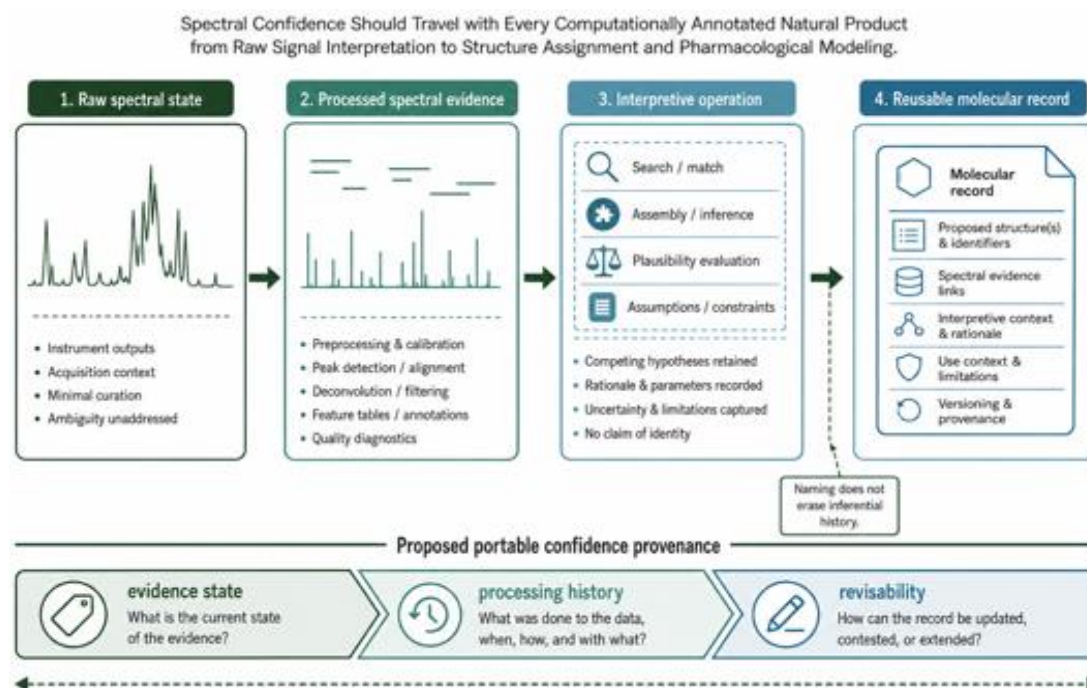


Figure 1. Evidence-bearing transitions from raw spectral state to a reusable molecular record

Distinguishing observed signal from inferred structure

The most important boundary in computational annotation is between the existence of a signal and the structural interpretation assigned to that signal. A measured fragment peak is evidence that an ion of particular mass-to-charge behavior was detected under specified conditions; the chemical structure assigned to that ion may remain inferential. This is not merely a philosophical distinction. Orthogonal spectroscopic testing of fragment-ion structures has shown that structural annotations contained in MS/MS libraries can be incorrect even when the corresponding spectral signals themselves are real [7]. Consequently, preserving the spectrum while discarding the evidential status of its structural interpretation can create false precision.

Machine learning increases the scale and sophistication of this interpretive layer. Contemporary approaches can learn spectral similarity, structural fingerprints, formulas, fragments, and candidate relationships that would be difficult to derive manually at repository scale. A recent review of machine learning in small-molecule mass spectrometry nevertheless emphasizes that these models operate across distinct tasks, datasets, representations, and validation settings rather than as a single identification mechanism [8]. Their outputs should therefore be read according to the inference they actually perform.

The framework proposed here treats observation confidence and structural-inference confidence as related but nonidentical. Clean, reproducible spectral evidence may still support several structures, while a computationally decisive ranking may be based on incomplete or contaminated input evidence. Conversely, imperfect spectra may remain informative at class or substructure level even when unique structure assignment is unjustified. The key requirement is not maximal certainty at every stage. It is preservation of the level at which uncertainty occurs, so that later users do not mistake confidence in the existence of a signal for confidence in a particular molecular structure.

Confidence at formula, substructure, and full-structure levels

Computational annotation does not move directly from unknown spectrum to confirmed molecule. Formula inference, chemical-class inference, structural-fingerprint inference, and candidate ranking answer different questions. SIRIUS, ZODIAC, and CANOPUS exemplify this separation: tandem spectra can support molecular-formula reasoning and systematic compound-class inference without thereby supplying independent experimental confirmation of a unique molecular structure [9–11]. Treating these outputs as one generic “annotation

confidence” obscures what information has actually been resolved.

Full candidate ranking adds another evidential layer. COSMIC demonstrates that sophisticated scoring can produce high-confidence structural annotations even for compounds absent from spectral libraries by combining predicted spectral behavior and candidate-level evidence [12]. Such performance is analytically valuable, but a ranking remains conditional on the candidate universe, available spectral information, scoring procedure, and chemical representation. The framework therefore distinguishes confidence in an inference operation from confirmation of the inferred structure.

Measurement precision further complicates attempts to translate confidence between levels. Analysis of multidimensional, reference-free metabolite identification shows that candidate ambiguity depends on measurement precision and on the number of structures remaining compatible with the measured properties [13]. Formula-level certainty can thus coexist with unresolved constitutional, positional, or stereochemical alternatives.

For natural products, this distinction is especially consequential because structurally related metabolites may share exact formulas, closely related fragmentation, and common biosynthetic scaffolds while differing at positions that matter for later computational modeling. The proposed record should therefore specify the inference level to which confidence applies—formula, class or substructure, candidate structure, or stronger full-structure assignment—rather than allowing a confidence score generated at one level to be silently interpreted as certainty at another. No cross-tool numerical equivalence is assumed.

Candidate ranking is not equivalent to structural identification

Candidate-ranking systems can greatly reduce structural search space without eliminating the distinction between plausibility and identification. MIST, for example, uses domain-informed transformer architectures to infer structural information from tandem mass spectra and support candidate retrieval, demonstrating that learned representations can extract substantial chemical information from fragmentation patterns [14]. The appropriate interpretation is that a model can rank or retrieve structures under its training and evaluation conditions, not that the highest-ranked molecular graph automatically becomes experimentally established.

Analogue-oriented methods make the distinction still clearer. MS2Query is explicitly designed to retrieve structurally related molecules from spectral information, including situations in which an exact library match may be unavailable [15]. A useful analogue can support chemical-class reasoning, dereplication, or prioritization

while remaining different from the compound that generated the query spectrum. Recording only the retrieved structure without preserving analogue-level status would convert relational evidence into an identity claim that the method itself does not make.

De novo generation moves one step further by producing chemically plausible candidates beyond direct database retrieval. MSNovelist demonstrates that tandem spectral evidence can be translated into candidate molecular structures through generative modeling [16]. Generation expands the hypothesis space; it does not collapse it. The distinction becomes particularly important for natural products, where isomeric alternatives, stereochemistry, uncommon substituent patterns, and incompletely represented chemical space can leave multiple structures compatible with available evidence.

Additional analytical evidence can subsequently change that situation. Online chemical labeling coupled with tandem mass spectrometry illustrates how experimentally introduced chemical constraints can sharpen annotation beyond the information supplied by unmodified MS/MS alone [17]. This possibility makes confidence intrinsically revisable. A candidate accepted provisionally today may be strengthened, demoted, or replaced when orthogonal evidence appears.

The proposed framework therefore treats ranking status, analogue status, generative status, and later evidence-supported revision as properties that should remain attached to the molecular record. Candidate ranking is scientifically useful precisely because it orders uncertainty; it should not be made to appear more definitive by stripping that uncertainty away.

Where spectral uncertainty is lost during data export and reuse

Uncertainty becomes most vulnerable when an annotation changes computational context. A spectrum may initially coexist with precursor information, processing parameters, candidate lists, scores, and analyst-visible ambiguity, yet export can reduce that rich evidential object to a molecular name, identifier, or structure string. Existing infrastructure shows that this compression is not technically inevitable. mzTab-M provides a standardized mechanism for exchanging metabolomics results, the Universal Spectrum Identifier makes individual spectra persistently addressable, and ReDU demonstrates that harmonized public metadata can enable systematic rediscovery and reanalysis of deposited mass-spectrometry data [18–20]. Together, these systems establish components of portable provenance, although they do not themselves constitute an end-to-end confidence standard.

Reuse also introduces a second problem: the evidential quality of the reference material may be heterogeneous.

Reproducible cleaning of public MS/MS libraries has shown why spectra and metadata require harmonization before they are treated as reusable computational resources [21]. A library entry is therefore not merely a lookup target; it is another evidence-bearing record whose own quality can constrain downstream annotation.

The distinction becomes sharper when annotations are intentionally propagated. Repository-scale nearest-neighbor approaches can extend suspect spectral libraries far beyond directly measured reference compounds, increasing analytical coverage while explicitly operating at a different evidential level [22]. Propagation is useful precisely because it permits inference beyond authenticated standards. Its value would be undermined if the resulting suspect assignments subsequently became indistinguishable from measured reference identities.

Natural-product databases further show that chemical records are curated and revisable rather than epistemically frozen. The Natural Products Atlas documents continued curation of microbially derived natural-product structures [23]. LIPID MAPS similarly demonstrates the need for maintained structural nomenclature and controlled levels of molecular description in a chemically demanding domain [24]. The proposed implication is that confidence provenance should survive export with the structure: what source evidence supported the record, which inference level produced it, and whether later evidence has modified that status.

The downstream cost of treating tentative structures as certain

The principal downstream risk is not that every tentative annotation will produce an incorrect pharmacological

conclusion. It is that a structure-dependent model cannot know that its molecular representation was provisional unless that information is retained. Once a tentative assignment is converted into an ordinary molecular graph, fingerprint, descriptor vector, or standardized structure string, downstream computation generally receives the structure rather than the history of how the structure was inferred.

This distinction matters because molecular machine-learning predictions can respond materially to data curation and confidence stratification. Xerxa and colleagues showed that compound predictions differed with data curation and confidence level in the evaluated modeling settings [25]. That result does not quantify the penalty caused by spectral misannotation specifically, but it establishes the more defensible premise that the evidential quality of molecular records can influence learned predictions.

Model performance also depends on factors beyond molecular identity. A systematic study of molecular-property prediction found substantial dependence on dataset, representation, model, and task choices rather than on model architecture alone [26]. Spectral confidence should therefore not be mistaken for a master variable explaining predictive validity. It is one upstream evidence dimension that may interact with an already complex modeling system.

The distinction developed in this section is summarized in **Table 2**. The table separates information that is lost when tentative structures are flattened into hard molecular labels from information that could be retained under the proposed framework.

Table 2. Distinguishing confidence-flattened molecular reuse from confidence-preserving reuse

Reuse condition	What downstream computation receives	Evidential consequence	Appropriate interpretation
Tentative assignment exported as ordinary structure	Molecular graph or identifier without annotation status	Source uncertainty becomes invisible	Prediction applies to the supplied structure, not necessarily the true analyte
Curated structure treated as hard training label	Structure plus measured or modeled endpoint	Errors may enter model fitting if structural identity is wrong; magnitude is task dependent [25]	Data curation remains part of model validity
Structure used in property or activity prediction	Representation selected by the modeling pipeline	Predictive behavior depends additionally on dataset, representation, and model design [26]	Structure confidence and model performance are distinct
Confidence-preserving reuse	Structure plus source-evidence and inference metadata	Original uncertainty remains inspectable	Proposed: use may be adapted to evidential status
Revised annotation propagated downstream	Updated structure plus version history	Dependent computations can be identified for reconsideration	Proposed: reanalysis should be selective rather than automatic

note: The last two rows are proposed framework operations. No universal weighting coefficient, exclusion threshold, or quantitative performance benefit is implied.

Propagation into docking, qsar, network pharmacology, and machine learning

Once an annotated molecule enters a pharmacological workflow, at least two uncertainty systems coexist. The first concerns whether the molecular representation accurately reflects the analyte that generated the spectrum. The second concerns whether a downstream model predicts the relevant property or activity reliably. These uncertainties arise from different evidence and should remain distinguishable.

Real-world compound-activity prediction illustrates why downstream uncertainty cannot be reduced to structural identity alone. Benchmarking across drug-discovery assay settings shows that prediction performance varies with assay context, available data, and modeling task [27]. A structurally correct molecule can therefore still receive an unreliable prediction; conversely, an apparently confident model output cannot retrospectively validate the upstream spectral assignment.

Bioactivity databases add their own provenance requirements. ChEMBL integrates heterogeneous compound and assay information across multiple bioactivity types and time periods [28]. Such records provide valuable pharmacological evidence, but the evidential meaning of a bioactivity measurement remains separate from the confidence that a mass-spectrometric annotation represents the correct molecule. Chemical machine-learning studies likewise show that uncertainty-quantification methods require explicit characterization and calibration rather than interpretation of a point prediction as self-validating confidence [29].

Applicability adds another boundary. Distance-aware molecular-property prediction demonstrates that a molecule's position relative to learned structure-property space can influence prediction reliability [30]. The proposed framework therefore treats downstream use as a two-axis problem: source-structure confidence describes the evidential support for the molecular identity entering the model, whereas model applicability and predictive uncertainty describe confidence in the model's response to that representation. Docking scores, QSAR predictions, network-pharmacology links, and machine-learning outputs may be informative, but none should be used as circular evidence that the original spectral assignment was correct.

A persistent spectral-confidence provenance framework

A usable confidence framework must persist beyond the annotation interface. Public-data infrastructures increasingly make repository-scale reanalysis technically possible, demonstrating the value of preserving source data and harmonized metadata rather than retaining only final interpretations [31]. Building on that principle, this article

proposes a traveling spectral-confidence record linked to each computationally annotated molecule.

The proposed record contains six conceptual components. First, source evidence points back to the spectrum or spectra from which the annotation arose. Second, analytical state records relevant processing or deconvolution status. Third, inference level specifies whether the claim concerns formula, class, substructure, analogue, candidate structure, or stronger structural assignment. Fourth, method context identifies the computational operation that generated or ranked the claim. Fifth, confidence state records the tool-appropriate score, category, or evidence status without assuming cross-platform numerical equivalence. Sixth, revision lineage records later strengthening, weakening, or replacement of the assignment.

Relational evidence should also remain distinguishable from identity evidence. Feature-based molecular networking can organize MS features according to spectral relationships while preserving feature-level context [32]. Such relationships are scientifically useful, but a network edge need not establish unique molecular identity. The proposed record would therefore preserve whether support derives from exact evidence, analogue relationships, propagation, or network context.

Finally, confidence must be updateable because new analytical dimensions can refine the structural question itself. MS-DIAL 5 illustrates how multimodal fragmentation and related analytical evidence can improve fine-grained lipid structural characterization [33]. Although lipid-specific results cannot be generalized to all natural products, they demonstrate the broader principle: later evidence may change the appropriate structural granularity. The molecular record should therefore be versioned rather than silently overwritten.

Confidence-weighted use of computationally annotated molecules

Retaining confidence is useful only if it changes how evidence is handled. The proposed framework does not prescribe a universal numerical threshold. Instead, it links admissible use to the consequence of being wrong. This is justified by the broader modeling literature showing that chemical machine-learning outputs depend on data curation, dataset construction, and uncertainty characterization [25, 26, 29].

For exploratory analysis, lower-confidence structures may remain informative when the research question concerns chemical neighborhoods, class enrichment, hypothesis generation, or prioritization for additional measurement. Their inclusion should be visible rather than disguised as certainty. For model training, the admissibility question becomes stricter: an uncertain structure used as a hard label

can inject an unrecorded error into the relationship between molecular representation and outcome. Appropriate responses could include exclusion, sensitivity analysis, stratification, or model-visible confidence metadata. These are proposed options, not validated universal rules.

For lead prioritization, the evidential burden should reflect decision consequence. A candidate whose structural identity materially determines docking pose, predicted target class, synthetic follow-up, or experimental resource allocation warrants stronger scrutiny than a candidate used merely to explore a chemical neighborhood. Confidence should therefore function as a decision variable alongside assay quality, model applicability, and pharmacological evidence rather than replacing them.

This approach also prevents the opposite error: treating uncertainty as grounds for universal deletion. Analogue-level and propagated annotations can remain scientifically useful when their status is preserved. The proposed decision rule is therefore confidence-aware admissibility, not confidence-maximization. A structure is used according to what its evidence supports and the consequence attached to downstream error.

Reanalysis when new spectral evidence appears

Persistent provenance converts structural annotation from a terminal label into a revisable record. Addressable spectra and public-data harmonization make it possible to return to source evidence rather than reconstructing provenance from a molecular name alone [19]. Repository-scale reanalysis further shows that existing data can be systematically revisited when methods, metadata, or computational resources improve [31].

Under the proposed framework, a revised structural assignment should not automatically invalidate every downstream result. Instead, the record should expose dependencies. If the change is only a confidence upgrade for the same structure, some analyses may require no recalculation. If the molecular graph changes, structure-dependent descriptors, fingerprints, docking calculations, network assignments, or predictive outputs become candidates for reanalysis. The distinction is operationally important because confidence revision and identity revision are not equivalent events.

Library improvement provides another route to revision. Reproducible spectral-library cleaning can alter the quality of reusable evidence [21], while newer analytical modalities may resolve structural detail previously unavailable [33]. The framework therefore treats reanalysis as triggered by evidential change, not as periodic repetition for its own sake.

A practical implementation would retain the prior assignment, the revised assignment, the evidence motivating revision, and the downstream analyses

dependent on the affected structure. This proposed versioned approach preserves scientific history while making correction possible without pretending that the earlier record never existed.

Reporting standards for machine-generated annotations

Machine-generated annotation should report enough information for a reader or downstream system to reconstruct what kind of claim is being made. Existing standards demonstrate that metabolomics results and source spectra can be represented in machine-addressable forms [18]. The proposed reporting minimum therefore combines structural representation with provenance rather than replacing established formats.

At minimum, a machine-generated record should specify the source spectrum identifier where available; analytical and processing context relevant to interpretation; annotation method and version; inference level; candidate or analogue status; the confidence output actually produced by that method; candidate-space information when material; and whether orthogonal evidence or an authentic reference contributes to the assignment. The key requirement is semantic: a generated candidate, an analogue match, and a stronger structure assignment should not be exported under an indistinguishable evidential label. MS2Query illustrates why analogue retrieval requires its own interpretation [15].

Database reuse should additionally retain revision state. Curated chemical resources demonstrate that molecular records evolve, and pan-repository reanalysis shows why durable links to earlier evidence remain valuable [31]. Reporting should therefore permit a later user to determine not only *what structure is currently represented* but *why that structure was represented, at what inferential level, and whether its status has changed*.

These fields are proposed reporting requirements, not an assertion that present metabolomics repositories or journals mandate them uniformly.

Boundaries between useful uncertainty and unusable ambiguity

Not every uncertain annotation is equally problematic. Some uncertainty arises from correctable analytical mixtures. DecoID, for example, demonstrates that computational deconvolution of coisolated MS/MS signals can improve identification in settings affected by chimeric spectra [34]. This supports a distinction between reducible uncertainty, for which additional processing or measurement can materially clarify the evidence, and residual ambiguity, for which multiple structures remain plausible after reasonable analysis.

Residual ambiguity can still be useful. MS2DeepScore shows that learned spectral similarity can capture structural

relatedness between tandem spectra [35]. Such similarity can support neighborhood analysis, analogue search, or prioritization without uniquely identifying a molecule. The appropriate response is therefore not to transform structural relatedness into identity, nor to discard it because identity is unresolved.

The practical boundary should depend on the intended decision. Class-level interpretation may tolerate ambiguity that would be unacceptable for stereochemistry-sensitive docking or for assigning a specific pharmacological mechanism. Conversely, demanding authentic-standard confirmation for every exploratory feature could erase useful chemical information and bias analysis toward compounds already represented in reference collections. The proposed framework therefore rejects both false certainty and indiscriminate skepticism. Useful uncertainty is uncertainty whose level, source, and consequence are visible. Ambiguity becomes unusable when the remaining structural alternatives are large enough, decision-relevant enough, or poorly characterized enough that downstream conclusions can no longer be attributed to a defensible molecular representation.

Limitations

This framework integrates analytical annotation, chemical-information provenance, and downstream modeling, but it has not been prospectively validated as an integrated system. Existing studies establish components of the argument—annotation disagreement, chimeric spectra, candidate ambiguity, library curation, revisability, data-quality effects, and predictive uncertainty—rather than demonstrating that the proposed traveling-confidence record improves pharmacological decision making.

No universal confidence ontology or conversion function is justified by the evidence reviewed here. Tool-specific scores arise from different models, candidate spaces, measurements, and validation regimes; treating them as numerically interchangeable would reproduce the simplification the framework is intended to prevent. Natural-product chemistry also creates difficult cases involving stereochemistry, tautomerism, ion forms, conjugates, mixtures, and incompletely represented chemical space.

The framework additionally does not determine when an annotation becomes sufficiently secure for a specific regulatory, clinical, synthetic, or investment decision. Those thresholds would require task-specific empirical validation. Nor does preserved provenance repair low-quality source measurements automatically.

Its immediate contribution is consequently conceptual and infrastructural: to define what information should remain attached to a computational molecular record so that later

users can judge, revise, or restrict its use without confusing analytical confidence with pharmacological truth.

CONCLUSION

Computational annotation is most useful when it converts difficult spectral evidence into chemically actionable hypotheses without erasing the uncertainty that made inference necessary. The central proposal of this framework is therefore simple but consequential: spectral confidence should travel with the molecular record.

That confidence should identify the level of inference being supported, preserve links to source evidence and analytical history, remain distinguishable from downstream predictive uncertainty, and permit later revision when new spectra, standards, algorithms, or orthogonal measurements change the structural interpretation. The purpose is not to impose one score across incompatible platforms or to exclude every tentative natural product from computational analysis.

Instead, the proposed framework makes uncertainty usable. Tentative structures can remain available for exploration when their status is visible, while higher-consequence applications can demand stronger evidence without pretending that docking, QSAR, network pharmacology, or machine learning can validate an upstream identity by themselves. Structural confidence is therefore best treated not as a transient annotation output but as persistent evidence provenance: a property of the record that determines what subsequent computation is entitled to assume.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Theodoridis G, Gika H, Raftery D, Goodacre R, Plumb RS, Wilson ID. Ensuring fact-based metabolite identification in liquid chromatography–mass spectrometry-based metabolomics. *Anal Chem*. 2023;95(8):3909-16. doi:10.1021/acs.analchem.2c05192
- [2] Houriet J, Manwill PK, Magaña AA, Anderson VM, Beniddir MA, Bertrand S, et al. Multilaboratory untargeted mass spectrometry metabolomics collaboration to identify bottlenecks and comprehensively annotate a single dataset. *Anal*

- Chem. 2025;97(30):16110-22. doi:10.1021/acs.analchem.4c05577
- [3] Metz TO, Chang CH, Gautam V, Anjum A, Tian S, Wang F, et al. Introducing "Identification Probability" for automated and transferable assessment of metabolite identification confidence in metabolomics and related studies. *Anal Chem.* 2025;97(1):1-11. doi:10.1021/acs.analchem.4c04060
- [4] Schmid R, Heuckeroth S, Korf A, Smirnov A, Myers O, Dyrland TS, et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat Biotechnol.* 2023;41(4):447-9. doi:10.1038/s41587-023-01690-2
- [5] Zhao T, Xing S, Yu H, Huan T. De novo cleaning of chimeric MS/MS spectra for LC-MS/MS-based metabolomics. *Anal Chem.* 2023;95(35):13018-28. doi:10.1021/acs.analchem.3c00736
- [6] Xing S, Charron-Lamoureux V, Zhao HN, El Abiead Y, Wang M, Dorrestein PC. Reverse spectral search reimaged: A simple but overlooked solution for chimeric spectral annotation. *Anal Chem.* 2025;97(33):17926-30. doi:10.1021/acs.analchem.5c02047
- [7] van Tetering L, Spies S, Wildeman QDK, Houthuijs KJ, van Outersterp RE, Martens J, et al. A spectroscopic test suggests that fragment ion structure annotations in MS/MS libraries are frequently incorrect. *Commun Chem.* 2024;7(1):30. doi:10.1038/s42004-024-01112-7
- [8] Hong Y, Ye Y, Tang H. Machine learning in small-molecule mass spectrometry. *Annu Rev Anal Chem (Palo Alto Calif).* 2025;18(1):193-215. doi:10.1146/annurev-anchem-071224-082157
- [9] Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods.* 2019;16(4):299-302. doi:10.1038/s41592-019-0344-8
- [10] Ludwig M, Nothias LF, Dührkop K, Koester I, Fleischauer M, Hoffmann MA, et al. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. *Nat Mach Intell.* 2020;2(10):629-41. doi:10.1038/s42256-020-00234-6
- [11] Dührkop K, Nothias LF, Fleischauer M, Reher R, Ludwig M, Hoffmann MA, et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. *Nat Biotechnol.* 2021;39(4):462-71. doi:10.1038/s41587-020-0740-8
- [12] Hoffmann MA, Nothias LF, Ludwig M, Fleischauer M, Gentry EC, Witting M, et al. High-confidence structural annotation of metabolites absent from spectral libraries. *Nat Biotechnol.* 2022;40(3):411-21. doi:10.1038/s41587-021-01045-9
- [13] Chang CH, Schwartz SC, Im AK, Bloodworth KJ, Webb-Robertson BJM, Ewing RG, et al. Assessing the impact of measurement precision on metabolite identification probability in multidimensional mass spectrometry-based, reference-free metabolomics. *Anal Chem.* 2025;97(26):13861-71. doi:10.1021/acs.analchem.5c01067
- [14] Goldman S, Wohlwend J, Stražar M, Haroush G, Xavier RJ, Coley CW. Annotating metabolite mass spectra with domain-inspired chemical formula transformers. *Nat Mach Intell.* 2023;5(9):965-79. doi:10.1038/s42256-023-00708-3
- [15] de Jonge NF, Louwen JJR, Chekmeneva E, Camuzeaux S, Vermeir FJ, Jansen RS, et al. MS2Query: Reliable and scalable MS2 mass spectra-based analogue search. *Nat Commun.* 2023;14(1):1752. doi:10.1038/s41467-023-37446-4
- [16] Stravs MA, Dührkop K, Böcker S, Zamboni N. MSNovelist: De novo structure generation from mass spectra. *Nat Methods.* 2022;19(7):865-70. doi:10.1038/s41592-022-01486-3
- [17] Vitale GA, Xia SN, Dührkop K, Zare Shahneh MR, Brötz-Oesterhelt H, Mast Y, et al. Enhancing tandem mass spectrometry-based metabolite annotation with online chemical labeling. *Nat Commun.* 2025;16(1):6911. doi:10.1038/s41467-025-61240-z
- [18] Hoffmann N, Rein J, Sachsenberg TT, Hartler J, Haug K, Mayer G, et al. mzTab-M: A data standard for sharing quantitative results in mass spectrometry metabolomics. *Anal Chem.* 2019;91(5):3302-10. doi:10.1021/acs.analchem.8b04310
- [19] Deutsch EW, Perez-Riverol Y, Carver J, Kawano S, Mendoza L, Van Den Bossche T, et al. Universal Spectrum Identifier for mass spectra. *Nat Methods.* 2021;18(7):768-70. doi:10.1038/s41592-021-01184-6
- [20] Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguirre S, Aksenov AA, et al. ReDU: A framework to find and reanalyze public mass spectrometry data. *Nat Methods.* 2020;17(9):901-4. doi:10.1038/s41592-020-0916-7
- [21] de Jonge NF, Hecht H, Strobel M, Wang M, van der Hoof JJJ, Huber F. Reproducible MS/MS library cleaning pipeline in matchms. *J Cheminform.* 2024;16(1):88. doi:10.1186/s13321-024-00878-1
- [22] Bittremieux W, Avalon NE, Thomas SP, Kakhkhorov SA, Aksenov AA, Gomes PWP, et al. Open access repository-scale propagated nearest neighbor suspect spectral library for untargeted metabolomics. *Nat Commun.* 2023;14(1):8488. doi:10.1038/s41467-023-44035-y

- [23] van Santen JA, Poynton EF, Iskakova D, McMann E, Alsup TA, Clark TN, et al. The Natural Products Atlas 2.0: A database of microbially-derived natural products. *Nucleic Acids Res.* 2022;50(D1):D1317-23. doi:10.1093/nar/gkab941
- [24] Conroy MJ, Andrews RM, Andrews S, Cockayne L, Dennis EA, Fahy E, et al. LIPID MAPS: Update to databases and tools for the lipidomics community. *Nucleic Acids Res.* 2024;52(D1):D1677-82. doi:10.1093/nar/gkad896
- [25] 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
- [26] Deng J, Yang Z, Wang H, Ojima I, Samaras D, Wang F. A systematic study of key elements underlying molecular property prediction. *Nat Commun.* 2023;14(1):6395. doi:10.1038/s41467-023-41948-6
- [27] Tian T, Li S, Zhang Z, Chen L, Zou Z, Zhao D, et al. Benchmarking compound activity prediction for real-world drug discovery applications. *Commun Chem.* 2024;7(1):127. doi:10.1038/s42004-024-01204-4
- [28] 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):D1180-92. doi:10.1093/nar/gkad1004
- [29] Heid E, McGill CJ, Vermeire FH, Green WH. Characterizing uncertainty in machine learning for chemistry. *J Chem Inf Model.* 2023;63(13):4012-29. doi:10.1021/acs.jcim.3c00373
- [30] Kim JY, Vlachos DG. Distance-aware molecular property prediction in nonlinear structure-property space. *J Chem Inf Model.* 2025;65(13):6744-56. doi:10.1021/acs.jcim.5c01037
- [31] El Abiead Y, Strobel M, Payne T, Fahy E, O'Donovan C, Subramamiam S, et al. Enabling pan-repository reanalysis for big data science of public metabolomics data. *Nat Commun.* 2025;16(1):4838. doi:10.1038/s41467-025-60067-y
- [32] Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods.* 2020;17(9):905-8. doi:10.1038/s41592-020-0933-6
- [33] Takeda H, Matsuzawa Y, Takeuchi M, Takahashi M, Nishida K, Harayama T, et al. MS-DIAL 5 multimodal mass spectrometry data mining unveils lipidome complexities. *Nat Commun.* 2024;15(1):9903. doi:10.1038/s41467-024-54137-w
- [34] Stancliffe E, Schwaiger-Haber M, Sindelar M, Patti GJ. DecoID improves identification rates in metabolomics through database-assisted MS/MS deconvolution. *Nat Methods.* 2021;18(7):779-87. doi:10.1038/s41592-021-01195-3
- [35] Huber F, van der Burg S, van der Hooft JJJ, Ridder L. MS2DeepScore: A novel deep learning similarity measure to compare tandem mass spectra. *J Cheminform.* 2021;13(1):84. doi:10.1186/s13321-021-00558-4