Specialty Journal of Pharmacognosy, Phytochemistry, and Biotechnology
ISSN: 3062-441X
2025, Volume 5, Issue 2, Page No: 22-31

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Unknown Peaks Should Carry Uncertainty, Not Forced Identities: A Confidence-Aware Computational Scheme for Annotating Plant LC–MS/MS Features Across Libraries and In-Silico Predictions

Krzysztof Wiśniewski[1]*, Małgorzata Nowak[1], Tomasz Adamczyk[2]

[1] Department of Computational Annotation and LC–MS/MS Feature Identification, Faculty of Pharmacy, Warsaw University of Life Sciences, Warsaw, Poland.
[2] Department of Confidence-Aware Spectral Interpretation, Faculty of Pharmacy, Jagiellonian University, Krakow, Poland.

***E-mail** ✉ krzysztof.wisniewski@sggw.edu.pl

Received: 03 May 2025; **Revised:** 06 August 2025; **Accepted:** 19 August 2025

ABSTRACT

Untargeted plant LC–MS/MS routinely produces far more reproducible spectral features than can be assigned to authenticated molecular structures. The resulting gap between detection and identification is increasingly filled by spectral-library search, molecular networking, formula inference, class prediction, candidate ranking, predicted spectra, retention modeling, and de novo structure generation. These tools expand chemical coverage, but their outputs are often reported at a level of structural certainty that exceeds the evidence actually available. This article develops an original computational analytical framework for preserving uncertainty across heterogeneous annotation routes rather than forcing every feature toward a single named compound. The analysis separates direct spectral evidence, analogue and neighborhood evidence, formula-level inference, class-level inference, ranked structural candidates, generated structural hypotheses, and orthogonal analytical constraints. It further identifies recurrent points at which confidence is lost, including ion-level redundancy, erroneous fragment interpretation, incomplete reference and candidate spaces, model-domain mismatch, chemically biased training data, and correlated prediction errors. The proposed scheme treats confidence as a structured evidence state rather than a single universal score and allows unresolved, class-supported, family-supported, candidate-ranked, or exact-identity outcomes to remain distinct. Its intended value is interpretive discipline: downstream plant metabolomics conclusions should inherit the uncertainty of the annotations on which they depend. The framework is not prospectively validated, does not establish universal numerical thresholds, and requires benchmarking across instruments, laboratories, chemical domains, and candidate-space conditions before operational use.

Keywords: Plant metabolomics, LC–MS/MS, Metabolite annotation, Spectral libraries, In-silico prediction, Uncertainty

How to Cite This Article: Wiśniewski K, Nowak M, Adamczyk T. Unknown Peaks Should Carry Uncertainty, Not Forced Identities: A Confidence-Aware Computational Scheme for Annotating Plant LC–MS/MS Features Across Libraries and In-Silico Predictions. Spec J Pharmacogn Phytochem Biotechnol. 2025;5(2):22-31. https://doi.org/10.51847/FlaHWfL27s

Introduction

Untargeted LC–MS/MS separates two operations that are often collapsed in reporting: detecting a reproducible analytical feature and establishing the molecular identity responsible for that feature. The first is fundamentally an observation problem; the second is an evidential inference problem. High-resolution precursor mass, isotope pattern, chromatographic behavior, and tandem-fragment information can narrow plausible chemistry, yet no single computational score converts those observations automatically into a demonstrated structure. Theodoridis *et al*. argued that metabolite identification should remain fact-based and that the terminology used for an assignment must be proportional to the evidence supporting it [1]. That principle is especially important for

workflows in which thousands of features are processed automatically and a ranked database name may be produced even when structural alternatives remain viable.

Plant metabolomics intensifies this problem because chemically dense extracts contain primary metabolites together with specialized metabolites, conjugates, isomers, adducts, degradation products, and compounds that are sparsely represented in reference resources. Guidance for plant metabolite profiling emphasizes that acquisition, annotation, interpretation, and reporting are linked parts of analytical validity rather than independent steps [2]. A feature that is correctly detected but weakly annotated can therefore become a source of interpretive error if the uncertainty attached to the annotation disappears before chemotaxonomic, pathway, ecological, or comparative conclusions are drawn. Conversely, preserving uncertainty need not make a dataset analytically useless. Formula, class, family, or analogue information may still be informative when exact structure is not defensible.

The rapid growth of public MS/MS repositories and propagated suspect libraries has expanded the fraction of spectra for which some chemical context can be retrieved. Repository-scale nearest-neighbor propagation can extend spectral knowledge beyond directly library-matched compounds, but it also makes clear that broader coverage and unique structural certainty are different objectives [3]. Larger search spaces can return more plausible alternatives, and a chemically related neighbor is not equivalent to an authenticated identity. The central problem is therefore no longer simply how to annotate more peaks. It is how to represent what kind of evidence supports each annotation, what uncertainty remains, and where computational extrapolation has moved beyond the domain that has actually been demonstrated. That distinction matters because modern workflows can increase annotation throughput faster than they increase direct structural evidence. Without explicit provenance, additional computational reach may make an assignment look more complete while leaving its decisive structural ambiguity unchanged.

This article proposes a confidence-aware computational analytical scheme for plant LC–MS/MS annotation. Its purpose is not to replace established annotation tools with another scoring algorithm. Instead, it organizes heterogeneous outputs so that spectral matches, formula assignments, chemical-class predictions, analogue relationships, ranked candidates, generated structures, and orthogonal evidence do not collapse into a single category of "identified metabolite." The framework is explicitly nonempirical and evidence-bounded. It synthesizes what current analytical and computational literature establishes about annotation evidence and failure modes, then proposes a structured way to preserve provenance, specificity, unresolved alternatives, and applicability concerns through downstream interpretation. A central premise is that an unknown feature is not an analytical embarrassment that must be named. It is an evidence state that may be resolved, partially resolved, or deliberately left unresolved when the available data do not justify greater structural specificity.

Why unknown features are commonly over-annotated

Over-annotation begins when precision of language outruns precision of evidence. A high-resolution spectrum can appear chemically detailed while still being compatible with several structures. Even fragment-level interpretation, often treated as mechanistically specific evidence, can be less secure than it appears. Van Tetering *et al*. used spectroscopic testing to show that structural annotations assigned to fragment ions in MS/MS libraries can be incorrect [4]. This result does not imply that library fragments are generally unreliable; it establishes a narrower but important point: apparent substructural specificity should not be treated as self-validating evidence merely because it is encoded in a reference resource.

A second route to over-annotation is the assumption that each detected feature represents a distinct metabolite. Electrospray LC–MS can generate protonated or deprotonated ions, adducts, multimers, isotope-related signals, and in-source fragments from the same underlying compound. Network-aware feature organization can preserve relationships among detected ions and MS/MS features rather than treating every signal as an independent chemical identity [5, 6]. Ion-identity molecular networking is particularly useful for linking ion forms that plausibly originate from the same neutral molecule, while feature-based molecular networking retains chromatographic and abundance information alongside spectral relationships. These approaches reduce one-feature/one-compound thinking, but a network edge remains contextual evidence rather than structural proof.

A third route is semantic collapse across annotation levels. Exact library correspondence, analogue similarity, molecular formula, compound class, database candidate rank, and a generated structure are not interchangeable outputs, yet automated reports may present all of them as named metabolites. This is a reporting problem as much as a computational one. A top-ranked candidate can be chemically plausible while still competing with

positional isomers, stereoisomers, or structures absent from the candidate database. Likewise, a predicted class may be correct while the exact molecular assignment is wrong. The critical distinction is not simply "known" versus "unknown," but the resolution at which the evidence remains defensible.

The proposed framework therefore treats over-annotation as a mismatch between evidence resolution and identity resolution. An exact name is warranted only when the analytical evidence supports that degree of specificity; otherwise, the feature should retain the highest defensible lower-resolution state. This mismatch can occur at several points: during feature consolidation, after an apparently persuasive spectral match, when a formula is translated into a single database structure, or when a model is required to emit a top-ranked candidate even though its score distribution remains diffuse. In each case, the computational system may be functioning as designed while the reporting layer overstates what its output means. This position extends the fact-based identification principle without assuming that one universal confidence hierarchy already exists. It also avoids the opposite error of discarding partially informative features simply because exact identity remains unresolved.

Sources of evidence for LC–MS/MS annotation

The most direct computational evidence in routine LC–MS/MS annotation is usually a comparison between an observed tandem spectrum and reference spectra. Similarity metrics therefore matter because they determine how spectral resemblance is quantified and ranked. Spectral entropy has been developed and operationalized as an MS/MS library-search similarity measure [7, 8]. A strong similarity score, however, remains evidence of spectral resemblance under defined conditions; it is not automatically a probability that two spectra derive from the same exact molecular structure.

When an exact reference is absent, analogue search and library expansion address different limitations. MS2Query demonstrates that spectra can be used to retrieve structurally related analogues at scale [9]. Such a result can provide useful family or neighborhood information even when no exact reference exists. By contrast, LibGen expands experimental natural-product reference coverage by generating high-quality spectra for authenticated compounds under multiple high-resolution fragmentation regimes [10]. Analogue retrieval therefore extrapolates from existing spectral knowledge, whereas authentic library expansion adds new experimentally grounded reference evidence. They should not be reported as equivalent annotation events.

Formula-level inference provides another evidence layer. SIRIUS 4 uses fragmentation-tree reasoning to convert tandem spectra into molecular-formula and structure-related information [11]. ZODIAC further shows that formula annotation can be improved by exploiting relationships across a dataset rather than scoring each feature independently [12]. Even a well-supported elemental formula leaves constitutional, positional, and stereochemical ambiguity unresolved. Formula confidence should therefore travel forward as formula confidence, not be silently promoted to compound identity.

A similar distinction applies to higher structural levels. CANOPUS predicts chemical classes for unknown metabolites from high-resolution fragmentation spectra [13], creating useful information when exact structures are unavailable. CSI:FingerID-related approaches extend further by deriving molecular fingerprints from MS/MS evidence and ranking candidate structures beyond direct spectral-library matching [14]. De novo methods such as MSNovelist can generate structural hypotheses even when the correct compound is absent from a conventional candidate database [15], while BUDDY interrogates precursor and fragment evidence from the bottom up to support molecular-formula discovery [16]. These methods answer different inferential questions.

More recent machine-learning approaches add further flexibility but not automatic certainty. Domain-inspired chemical formula transformers learn relationships between spectral information and chemically structured representations [17], and MIST-CF applies a data-driven formula-ranking strategy to tandem spectra [18]. Ensemble spectral prediction can compare measured spectra against spectra predicted for candidate structures, but its reported behavior depends on training-data composition, candidate-set size, candidate similarity, and whether the correct structure is represented among the candidates [19]. Chromatographic evidence can add another dimension: joint modeling of LC retention order and tandem MS can improve structural ranking relative to MS/MS-only use in compatible settings [20]. These complementary evidence sources motivate a framework in which agreement is informative only when the provenance, task level, and dependence structure of the contributing evidence remain visible.

Where annotation confidence is lost

Confidence is often described as if it were generated only at the final ranking step. In practice, it can be lost much earlier. One failure point is reproducibility and transfer. A model may perform well in the environment in which it was developed yet be less reliable when code, data preparation, or test chemistry changes. Independent reusability analysis of the MIST framework highlights the importance of examining reproducibility, transparency, and generalization rather than assuming that original benchmark performance transfers unchanged to new data [21]. For plant metabolomics, this is especially relevant because specialized metabolites may occupy chemical regions that are sparsely represented in the training distributions of general small-molecule models.

A second failure point is the candidate space itself. If several structures remain compatible with the available measurements, returning the highest-ranked one does not erase the alternatives. Identification probability has been proposed and operationalized as a way to preserve the ambiguity represented by multiple candidate identities compatible with available analytical or predicted properties [22, 23]. The value of this idea is not that it supplies a universal probability scale for every LC–MS/MS workflow. Rather, it makes residual ambiguity explicit. Candidate counts and probabilities remain conditional on how the search space was constructed, which properties were measured or predicted, and whether the true structure is represented at all.

A third failure point arises when several imperfect evidence streams appear to agree. Agreement can be compelling, but it is not necessarily independent corroboration. A predicted spectrum and a predicted retention property may both inherit bias from related training data, while a database search and a generative model may share the same underlying structural universe. Likewise, a confident formula can constrain candidate structures without validating their connectivity. The proposed framework therefore distinguishes evidence accumulation from evidence independence. Confidence should increase most strongly when orthogonal observations constrain different aspects of the same hypothesis; it should increase less when multiple scores reproduce the same assumptions. Apparent consensus among dependent predictors is therefore recorded as concordance, not treated automatically as independent confirmation.

Finally, confidence can be lost through overextension beyond the demonstrated domain. A method validated on familiar, database-rich chemistry may be least secure precisely where plant metabolomics needs it most: unusual scaffolds, rare conjugates, poorly represented stereochemistry, or compounds with atypical fragmentation. This does not justify rejecting computational annotation. It justifies recording where the inference depends on extrapolation. The analytical objective is therefore not maximal assignment coverage, but maximal defensible information: preserve the strongest supported resolution, retain unresolved alternatives, and expose the assumptions that would otherwise disappear behind a single compound name.

A confidence-aware annotation scheme

A useful confidence scheme should preserve the type, resolution, and provenance of evidence rather than compress heterogeneous outputs into one numerical rank. Contemporary annotation practice draws on complementary analytical and informatic strategies, but those strategies answer different questions and carry different failure modes [24]. The scheme proposed here therefore treats an LC–MS/MS feature as an evidence record that can move among defensible resolution states without being forced toward exact identity. At minimum, the record should distinguish observed feature evidence, molecular-formula support, chemical-class or family support, ranked structural candidates, and exact-identity evidence. Each state should retain the method that produced it, the relevant search or model domain, and any unresolved competing explanations.

Residual ambiguity is a positive piece of information, not an inconvenience to delete. Identification-probability approaches illustrate why several structures compatible with the available properties should remain visible rather than disappear behind the highest-ranked candidate [22]. In the proposed scheme, a feature may therefore terminate provisionally at the highest structural resolution justified by its evidence. A confident class prediction can remain a class-level annotation; a strong formula can remain a formula assignment; and several similarly plausible structures can remain an unresolved candidate set. Exact naming is reserved for cases in which the evidential package supports that specificity.

Evidence concordance is also separated from evidence independence. Predicted spectra can strengthen candidate ranking, but their performance depends on the chemical and candidate spaces represented during training and evaluation [19]. Retention-order evidence can add a complementary constraint when the chromatographic setting is sufficiently comparable [20]. The framework therefore records whether agreement arises from genuinely different observations or from computational streams that may share databases, representations, or training

biases. Agreement among dependent predictors is treated as concordance, whereas orthogonal experimental evidence can provide stronger constraint.

Confidence is not conceived as a one-way ladder. New evidence may increase resolution, expose an alternative, or downgrade an earlier assignment when reference or domain assumptions change. The proposed record should therefore remain revisable and provenance-preserving rather than erase the evidence path by which its current state was reached. This matters when a strong candidate is later challenged by an authentic spectrum, a competing isomer, or evidence of model-domain mismatch.

Finally, the proposed scheme includes explicit abstention. A pipeline should be allowed to report that no exact structure is defensible, that the feature is outside the demonstrated model domain, or that additional measurement is required. This is not equivalent to analytical failure: it preserves usable information while preventing unsupported structural specificity. The manuscript-derived states and corresponding reporting actions are summarized in **Table 1**.

Table 1. Proposed evidence states and reporting actions for confidence-aware LC–MS/MS annotation

Evidence state	Defensible report	Main unresolved issue	Proposed action
Observed feature	Reproducible analytical signal	Chemical identity	Retain as unknown
Formula-supported	Molecular formula	Connectivity/isomerism	Do not assign exact name
Class/family-supported	Chemical class or related family	Unique structure	Report partial resolution
Candidate-ranked	Ordered structural hypotheses	Candidate multiplicity	Retain alternatives
Orthogonally constrained	Candidate set narrowed independently	Remaining nonunique candidates	Escalate only if warranted
Exact-identity supported	Specific molecular structure	Residual evidential limits	Report with provenance
Outside demonstrated domain	Model output with applicability concern	Transfer validity	Flag or abstain

Note: These states and reporting actions are proposed framework elements, not universally validated confidence categories or numerical thresholds.

The corresponding architecture in **Figure 1** integrates these states with over-annotation routes, heterogeneous evidence sources, confidence-loss mechanisms, unresolved-feature handling, downstream plant interpretation, and the validation boundary.

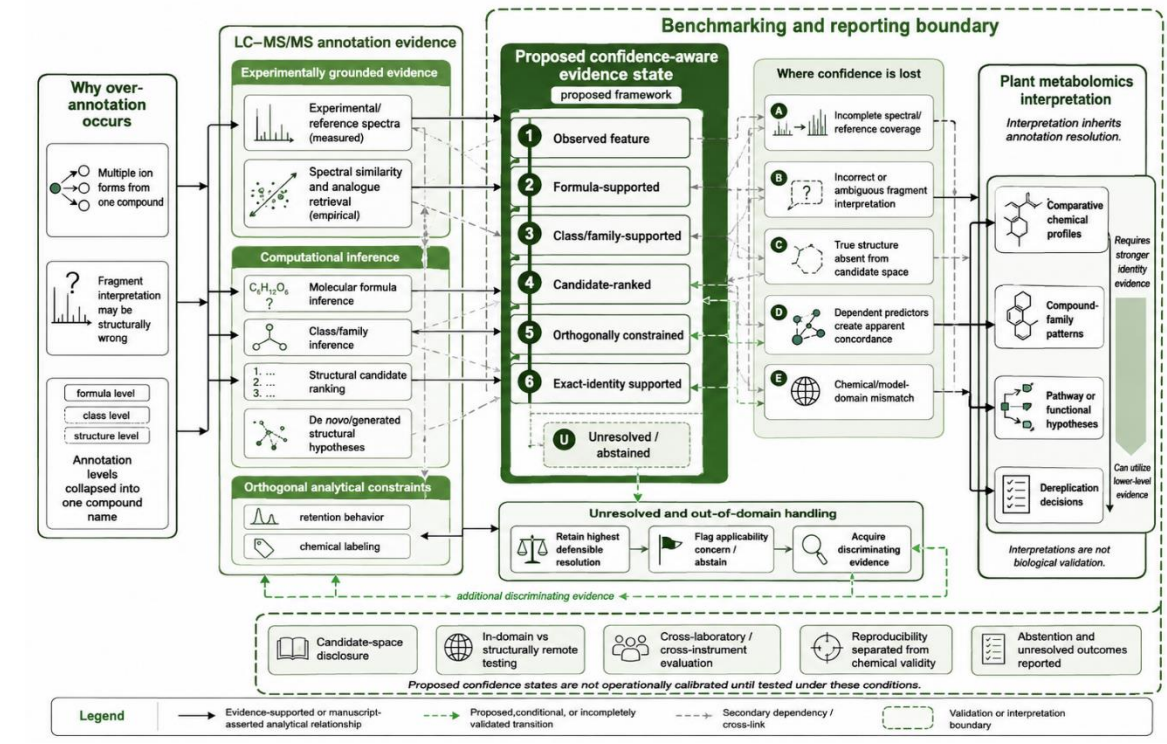

Figure 1. Evidence-preserving architecture for confidence-aware annotation of plant LC–MS/MS features

Handling unresolved and out-of-domain features

The central discipline of a confidence-aware system is to stop where the evidence stops. For an unresolved feature, this does not mean reducing the output to "unknown" when useful lower-resolution information exists. Natural-product molecular networking can recover compound-family relationships from network topology and structural similarity even when unique structures are not available [25]. The proposed rule is therefore to retain the highest defensible resolution: exact structure when justified, otherwise candidate set, family or class, formula, or unresolved feature. Each lower-resolution state should remain available for downstream analysis without being silently promoted.

Out-of-domain features require a different warning. Small-molecule machine-learning datasets can contain substantial chemical-coverage bias, meaning that conventional train/test performance may not represent structurally remote chemistry [26]. For plant metabolomics, uncommon scaffolds, conjugation patterns, or fragmentation behavior may therefore produce predictions that look numerically precise while resting on weak domain support. An out-of-domain flag should not be interpreted as proof that a prediction is wrong. It should indicate that the model's demonstrated applicability is insufficient to justify ordinary confidence semantics.

The proposed domain check is deliberately broader than a single structural-distance criterion. Chemical coverage is one component, but the annotation record should also note when inference depends on unverified candidate inclusion, unfamiliar ion forms, poorly matched fragmentation conditions, or an analytical context materially different from that used for model development. These dimensions should not be collapsed into a universal distance threshold. Their purpose is to expose reasons why a numerical score may be less interpretable, including the especially important closed-world case in which the correct structure is absent from the search space entirely.

Uncertainty can also trigger additional measurement rather than immediate abandonment. Online chemical labeling shows that orthogonal reaction-based information can constrain metabolite annotation by adding structural evidence not contained in the original tandem spectrum [27]. Comparable escalation could involve authentic standards, altered fragmentation conditions, chromatographic comparison, or another genuinely independent measurement when available. In the proposed scheme, such escalation is conditional: additional evidence is sought when the scientific importance of the feature justifies the effort and when the new measurement can discriminate among remaining alternatives. The result may still remain unresolved. A confidence-aware workflow therefore treats abstention, partial resolution, and targeted remeasurement as legitimate analytical outcomes.

Consequences for plant metabolomics interpretation

Annotation uncertainty should propagate into biological interpretation rather than being stripped away at export. Plant metabolite profiling depends on disciplined acquisition, annotation, and reporting because the chemical identities assigned to features shape subsequent comparisons and narratives [2]. If a tentative candidate is converted into a definitive metabolite name, downstream analyses may treat that label as measured fact. The proposed framework instead requires interpretive claims to inherit the structural resolution of the underlying evidence. A formula-supported feature can contribute to compositional reasoning at that level, whereas pathway or mechanism claims requiring an exact structure need stronger support.

Partial annotation can be scientifically useful when its limits are explicit. Natural-product family annotation can group chemically related features even when exact identities remain unresolved [25]. In ecological, taxonomic, or comparative plant studies, such family-level organization may preserve meaningful patterns without inventing molecular precision. The relevant question becomes whether the biological interpretation actually requires a unique structure. If it does not, lower-resolution evidence may be sufficient; if it does, unresolved identity should constrain the conclusion rather than being hidden.

Forced identities are most consequential when they propagate across multiple layers of analysis. Fact-based identification principles caution against reporting a molecular assignment more strongly than the evidence permits [1]. The experimental demonstration that fragment structural annotations can themselves be incorrect further shows how apparently specific upstream information may contain structural error [4]. A mistaken exact name can therefore influence dereplication decisions, pathway placement, enrichment analyses, chemotaxonomic comparisons, or hypotheses about plant function. The selected evidence does not quantify how often these downstream distortions occur across plant metabolomics, so the framework does not assign a universal error burden.

A practical consequence is that data structures and figures should preserve annotation state alongside the feature, rather than exporting only a preferred compound name. Comparative analyses can then distinguish patterns supported at formula, class/family, or candidate level from those supported by exact identity. When several uncertain features map to the same biological interpretation, their numerical abundance should not be mistaken for multiple independent confirmations of that interpretation. This is a proposed reporting discipline, not evidence that uncertainty-aware aggregation will necessarily improve every downstream model.

Confidence-aware interpretation is therefore not synonymous with conservative under-annotation. It is a resolution-matching strategy. Strong evidence should permit strong structural statements; partial evidence should support partial chemical statements; and uncertain evidence should remain visibly uncertain. This preserves more information than a binary identified/unidentified system because it keeps defensible class, family, formula, and candidate relationships available while preventing them from masquerading as exact molecular observations. The downstream benefit is methodological transparency rather than guaranteed biological correctness.

Benchmarking and reporting needs

A confidence framework is useful only if its categories can be challenged under conditions resembling real analytical use. A multilaboratory untargeted metabolomics exercise has demonstrated that annotation remains subject to practical bottlenecks when a shared dataset is processed and interpreted across different groups [28]. Such comparisons are important because a workflow that is internally consistent in one laboratory may behave differently when preprocessing, spectral resources, instruments, or analyst decisions change. Prospective evaluation of the proposed scheme should therefore include cross-laboratory and, where feasible, cross-instrument tests rather than relying only on internal benchmarks.

Reproducibility and analytical validity should also be reported separately. Independent reusability assessment shows why computational metabolomics methods need transparent evaluation of implementation, reproducibility, and generalization [21]. Reproducing a model's output does not prove that an annotation is chemically correct, while disagreement between implementations does not by itself reveal which structural assignment is true. Benchmark reports should therefore distinguish software reproducibility, ranking performance, calibration or selective-prediction behavior, and experimentally supported identification.

Chemical-domain coverage is another necessary reporting dimension. Coverage-bias analysis indicates that apparent small-molecule machine-learning performance can depend strongly on how chemical space is partitioned between training and evaluation sets [26]. Confidence-aware benchmarks should therefore include structurally remote test cases and explicitly describe how out-of-domain status is determined. Random splits alone are insufficient to demonstrate transfer to rare plant-specialized chemistry.

Candidate-space design must also be disclosed. Predicted-spectrum annotation changes as candidate number and similarity change [19], while identification-probability reasoning depends on the set of candidates compatible with the properties being considered [22]. Benchmarks should report whether the true structure is guaranteed to be present, whether stereoisomers or constitutional isomers are represented, and how databases are filtered. Formula, class, candidate-ranking, and exact-structure tasks should be evaluated separately rather than merged into a single annotation-accuracy statistic.

Reporting should expose the evidence path behind each final state, including reference source, computational method, candidate-space construction, domain flag, orthogonal evidence, unresolved alternatives, and any reason for abstention. These are proposed reporting targets, not a universal metadata standard; they distinguish identical labels reached through materially different evidential routes.

Finally, abstention should be benchmarked as an outcome rather than discarded before analysis. A useful system should be judged not only by how often its assigned names are correct, but also by whether it withholds exact identity when evidence is inadequate and whether useful lower-resolution information is retained. Prospective tests should compare always-predict pipelines with selective pipelines under known, withheld, and out-of-domain chemistry. Until such evaluation is performed, the proposed confidence states should be treated as a reporting and reasoning architecture, not as calibrated operational thresholds.

Conclusion

Untargeted plant LC–MS/MS increasingly combines experimental spectra with library search, molecular networking, formula inference, class prediction, candidate ranking, predicted spectra, chromatographic evidence,

and generative models. These methods expand the fraction of chemical space that can be described, but they do not convert every detected feature into an experimentally established molecule. The central contribution of this framework is therefore a change in what an annotation system is asked to preserve: not merely its preferred identity, but the resolution, provenance, competing alternatives, domain conditions, and unresolved uncertainty associated with that identity.

The proposed scheme treats exact identification as one possible endpoint rather than the mandatory destination of every feature. Formula-supported, class-supported, family-supported, candidate-ranked, unresolved, and out-of-domain outcomes can remain analytically useful when their limits are explicit. It further distinguishes agreement from independent corroboration and allows abstention or additional measurement when computational evidence cannot defensibly discriminate among alternatives. Downstream plant interpretations should inherit these confidence states instead of silently converting them into categorical molecular observations.

This framework is intentionally nonquantitative and nonvalidated. It does not establish universal confidence scores, numerical cutoffs, or superiority over existing annotation systems. Its value is conceptual and analytical: it specifies where evidence resolution should constrain identity resolution and where computational reach should remain visibly conditional on candidate space, model domain, experimental support, and transferability. It also makes falsifiability operationally important: future evaluation should test whether preserving partial states and abstentions actually reduces unsupported exact assignments without discarding useful chemical information.

Future work should compare this architecture with always-predict workflows using authenticated and deliberately withheld compounds, chemically remote test sets, varied candidate spaces, and cross-laboratory or cross-instrument analyses. Evaluation should address structural accuracy, calibration, selective prediction, retention of lower-resolution information, reproducibility, and downstream uncertainty propagation. Until then, uncertainty should be treated as preserved analytical information rather than a defect annotation software is expected to conceal.

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. Çiçek SS, Mangoni A, Hanschen FS, Agerbirk N, Zidorn C. Essentials in the acquisition, interpretation, and reporting of plant metabolite profiles. Phytochemistry. 2024;220:114004. doi:10.1016/j.phytochem.2024.114004
3. 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
4. van Tetering L, Spies S, Wildeman QDK, Houthuijs KJ, van Outersterp REV, 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
5. Schmid R, Petras D, Nothias LF, Wang M, Aron AT, Jagels A, et al. Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment. Nat Commun. 2021;12(1):3832. doi:10.1038/s41467-021-23953-9
6. 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

7. Li Y, Kind T, Folz J, Vaniya A, Mehta SS, Fiehn O. Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. Nat Methods. 2021;18(12):1524-31. doi:10.1038/s41592-021-01331-z
8. Li Y, Fiehn O. Flash entropy search to query all mass spectral libraries in real time. Nat Methods. 2023;20(10):1475-8. doi:10.1038/s41592-023-02012-9
9. 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
10. Kong F, Keshet U, Shen T, Rodriguez E, Fiehn O. LibGen: Generating high quality spectral libraries of natural products for EAD-, UVPD-, and HCD-high resolution mass spectrometers. Anal Chem. 2023;95(46):16810-8. doi:10.1021/acs.analchem.3c02263
11. 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
12. 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
13. 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
14. 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
15. 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
16. Xing S, Shen S, Xu B, Li X, Huan T. BUDDY: Molecular formula discovery via bottom-up MS/MS interrogation. Nat Methods. 2023;20(6):881-90. doi:10.1038/s41592-023-01850-x
17. 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
18. Goldman S, Xin J, Provenzano J, Coley CW. MIST-CF: Chemical formula inference from tandem mass spectra. J Chem Inf Model. 2024;64(7):2421-31. doi:10.1021/acs.jcim.3c01082
19. Li X, Chen YZ, Kalia A, Zhu H, Liu LP, Hassoun S. An ensemble spectral prediction (ESP) model for metabolite annotation. Bioinformatics. 2024;40(8). doi:10.1093/bioinformatics/btae490
20. Bach E, Schymanski EL, Rousu J. Joint structural annotation of small molecules using liquid chromatography retention order and tandem mass spectrometry data. Nat Mach Intell. 2022;4(12):1224-37. doi:10.1038/s42256-022-00577-2
21. Heirman J, Bittremieux W. Reusability report: Annotating metabolite mass spectra with domain-inspired chemical formula transformers. Nat Mach Intell. 2024;6(11):1296-302. doi:10.1038/s42256-024-00909-4
22. 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
23. Ngan HL, Turkina V, van Herwerden D, Yan H, Cai Z, Samanipour S. Machine learning for enhanced identification probability in RPLC/HRMS nontargeted workflows. Anal Chem. 2025;97(33):18028-35. doi:10.1021/acs.analchem.5c01873
24. Cai Y, Zhou Z, Zhu ZJ. Advanced analytical and informatic strategies for metabolite annotation in untargeted metabolomics. TrAC Trends Anal Chem. 2023;158:116903. doi:10.1016/j.trac.2022.116903
25. Morehouse NJ, Clark TN, McMann EJ, van Santen JA, Haeckl FPJ, Gray CA, et al. Annotation of natural product compound families using molecular networking topology and structural similarity fingerprinting. Nat Commun. 2023;14(1):308. doi:10.1038/s41467-022-35734-z
26. Kretschmer F, Seipp J, Ludwig M, Klau GW, Böcker S. Coverage bias in small molecule machine learning. Nat Commun. 2025;16(1):554. doi:10.1038/s41467-024-55462-w

27. 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
28. Houriet J, Manwill PK, Alcázar Magaña A, 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