Specialty Journal of Pharmacognosy, Phytochemistry, and Biotechnology
ISSN: 3062-441X
2026, Volume 6, Issue 1, Page No: 153-163

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

Galaxy Publication

AI-Assisted Structure Elucidation of Natural Products: A Methodological Review of Spectral Prediction, Candidate Generation, Confidence Calibration, and Human Verification

Ethan Wright[1]*, Chloe Bennett[2], Jack Turner[1], Olivia Harris[3]

[1] Department of AI-Assisted Structure Elucidation, Faculty of Pharmacy, University of Leeds, Leeds, United Kingdom.
[2] Department of Spectral Prediction and Candidate Generation, Faculty of Pharmacy, University of Sheffield, Sheffield, United Kingdom.
[3] Department of Confidence Calibration and Human Verification, Faculty of Pharmacy, University of York, York, United Kingdom.

***E-mail** ✉ ethan.wright@leeds.ac.uk

Received: 03 March 2026; **Revised:** 14 June 2026; **Accepted:** 17 June 2026

ABSTRACT

Artificial intelligence is increasingly embedded in natural-product structure elucidation, but the term encompasses methodologically distinct operations whose outputs carry different evidential meanings. This methodological review examines where AI enters the elucidation workflow, how spectral prediction and inverse inference differ, how candidates are generated and ranked, how complementary analytical modalities can be integrated, and why confidence and human verification remain separate problems from predictive accuracy. Recent methods span learned nuclear magnetic resonance representations, forward chemical-shift and tandem-mass-spectrum prediction, formula and fingerprint inference, database-assisted ranking, de novo molecular generation, multimodal spectrum-to-structure systems, and hybrid physics-guided approaches. Their capabilities are substantial but bounded by training chemistry, acquisition conditions, reference-database composition, candidate-space definition, and the distinction between simulated and experimental spectra. The central synthesis proposed here is an evidence-qualified architecture in which AI outputs progress through separable states: spectral representation or prediction, candidate construction, cross-modal reconciliation, domain-qualified confidence, and chemical verification. These states should not be collapsed into a single claim of "structure elucidation." Particular caution is required for structurally novel natural products, stereochemically dense candidates, sparse reference regions, and model outputs obtained outside the calibration domain. The review argues that methodological progress should therefore be judged not only by benchmark accuracy but also by transferability, uncertainty calibration, candidate traceability, contradiction handling, and the capacity to support or trigger expert reassessment. AI can materially reorganize structure-elucidation practice, but defensible assignment still depends on matching computational claims to the strength and scope of analytical evidence.

Keywords: Natural products, Structure elucidation, Machine learning, NMR spectroscopy, Tandem mass spectrometry, Uncertainty calibration

How to Cite This Article: Wright E, Bennett C, Turner J, Harris O. AI-Assisted Structure Elucidation of Natural Products: A Methodological Review of Spectral Prediction, Candidate Generation, Confidence Calibration, and Human Verification. Spec J Pharmacogn Phytochem Biotechnol. 2026;6(1):153-63. https://doi.org/10.51847/113Qone9P1

Introduction

Natural-product structure elucidation has traditionally depended on iterative interpretation of spectroscopic evidence, chemical reasoning, comparison with precedent, and, when necessary, synthesis or other orthogonal confirmation. Machine learning now enters this process at several points, including spectrum representation, forward prediction, structural similarity search, molecular-formula inference, candidate ranking, and inverse generation. Natural-product-specific reviews already show substantial activity across both nuclear magnetic

resonance (NMR) and mass spectrometry (MS), but also reveal that "AI-assisted structure annotation" covers tasks with markedly different evidential consequences [1]. Contemporary MS methodology similarly spans prediction, representation learning, formula inference, fingerprinting, candidate retrieval, and generative modeling rather than one homogeneous identification problem [2].

The same distinction is necessary in NMR. Machine learning has been applied to spectral processing, chemical-shift prediction, assignment-related tasks, representation learning, and structure-oriented inference, with different sources of uncertainty attached to each operation [3]. An accurate forward model that predicts a spectrum from a proposed structure, for example, answers a different question from an inverse model that proposes a structure from spectral observations. Neither result is equivalent to experimental confirmation merely because both can be evaluated against known molecules.

This review therefore treats AI-assisted natural-product elucidation as an evidence-transformation problem. The central question is not whether an algorithm can output a molecular structure, but what analytical information entered the model, what restrictions shaped the candidate space, what form the output takes, how uncertainty is expressed, and what additional evidence is needed before the output can support a defensible structural assignment. This framing is particularly important for natural products because rare scaffolds, dense stereochemistry, unusual substitution patterns, conformational effects, and incomplete reference coverage can place the most interesting structures furthest from the domains on which computational systems were optimized.

The review organizes current methods by their role in that evidential chain, then examines multimodal integration, confidence, human verification, benchmarking, and reporting. The resulting framework is proposed as a methodological synthesis rather than a validated universal standard.

Where AI enters natural-product structure elucidation

AI can influence structure elucidation before any exact structure is proposed. SMART demonstrated that learned representations of two-dimensional HSQC spectra can organize natural products according to structural similarity, making spectral pattern recognition itself a computational entry point [4]. Subsequent convolutional approaches extended rapid annotation across chemically diverse natural products, supporting the use of learned NMR features for candidate-oriented interpretation while remaining dependent on the chemical information represented in the reference domain [5]. DeepSAT further showed that NMR-derived representations can support structural-neighbor and scaffold-level retrieval from spectral information [6]. These tasks are valuable because they reduce an otherwise unstructured search problem, but a retrieved analogue or scaffold is not a unique molecular assignment.

A second entry point occurs when AI predicts an analytical observable from a proposed structure. Such forward models can supply spectra or chemical shifts against which experimental measurements are compared. A third operates in the opposite direction: inverse models infer formulae, structural features, or complete candidate molecules from measured spectra. A fourth entry point lies between them, where predicted observables, learned fingerprints, databases, and chemical constraints are combined to rank an existing candidate set.

These distinctions matter because the output vocabulary can be deceptively similar. "Prediction," "annotation," "identification," "generation," and "elucidation" may all culminate in a molecular depiction, yet the inferential route to that depiction differs. The framework proposed here therefore separates representation, forward prediction, candidate construction, candidate ranking, and verification as different evidential operations. Movement between them should be explicit rather than implied. In particular, a system that retrieves a close structural neighbor or ranks a supplied candidate list has not solved the same problem as a system required to construct a previously unseen structure without direct reference retrieval.

Methodological families in spectral prediction

Forward spectral prediction is one of the most established interfaces between molecular structure and AI-assisted elucidation. For NMR, simulated HSQC spectra can be compared with experimental spectra to support molecular identification, placing the model upstream of a matching decision rather than asking it to generate a structure directly [7]. Inverse approaches invert that relationship: routine one-dimensional NMR spectra have been used to infer and rank molecular structures within a defined chemical domain [8]. These two directions should remain methodologically separate because errors propagate differently. Forward prediction asks whether a candidate is compatible with observation; inverse prediction must additionally navigate the combinatorial ambiguity of structures compatible with the observation.

Context is also part of the prediction problem. Solvent-aware two-dimensional NMR models show that acquisition environment can be represented rather than treated as nuisance variation [9]. At the same time, unified benchmarking of NMR chemical-shift models demonstrates that apparently similar methods can be difficult to compare when data construction and evaluation protocols differ [10]. Thus, prediction error is partly a property of the model and partly a property of the domain in which the comparison is made.

MS prediction displays a parallel diversity. Recent neural methods include autoregressive fragmentation graphs, local fragmentation-event modeling, and graph-transformer architectures [11–13]. These approaches encode different assumptions about how molecular structure should be related to tandem spectra. CFM-ID illustrates a further methodological step in which predicted MS/MS behavior feeds directly into compound identification and candidate scoring [14]. Formula prediction can constrain the search earlier still: FIDDLE infers chemical formulae from tandem spectra, reducing candidate space without determining constitutional or stereochemical identity by itself [15].

The evidence characteristics of these families are summarized in **Table 1.**

Table 1. Evidence roles and principal boundaries of AI-enabled spectral prediction families relevant to structure elucidation

Methodological family	Typical analytical input	Immediate computational output	Elucidation role	Principal evidential boundary
Learned spectral representation	Experimental NMR or MS pattern	Embedding, similarity, structural neighborhood	Narrows chemical context	Similarity does not establish identity
Forward NMR prediction	Proposed molecular structure	Shifts, peaks, or simulated spectrum	Tests candidate–spectrum compatibility	Prediction error depends on environment and molecular domain
Inverse NMR inference	Experimental NMR	Formula, connectivity, or ranked structures	Constructs or ranks candidates	Search difficulty and training chemistry constrain recovery
Context-aware NMR prediction	Structure plus measurement context	Conditioned spectral prediction	Reduces systematic measurement mismatch	Represented conditions remain incomplete
Fragmentation-explicit MS prediction	Molecular graph	Predicted fragments or MS/MS spectrum	Supports spectral matching and ranking	Learned fragmentation is not complete mechanistic proof
Graph-based MS prediction	Molecular graph	Predicted tandem spectrum	Expands searchable spectral evidence	Instrument and data distributions affect transfer
Formula or candidate scoring	Experimental MS/MS and candidate information	Formula probability, fingerprint, or rank	Constrains or orders candidate space	Formula or rank alone cannot establish full structure

Note: The distinctions are analytical roles rather than a performance ranking. A single platform may combine several families.

Candidate generation and ranking

Once spectra have been represented or predicted, the central computational problem becomes how candidate structures enter the search. Conditional molecular generation from ^13C NMR demonstrates one route in which spectral evidence and prior chemical information jointly constrain the generative space [16]. More recent multispectral systems extend the input side of this problem by combining several forms of spectroscopic evidence rather than treating each modality independently [17]. Multimodal NMR image models provide another inverse strategy, learning directly from multiple spectral images to propose structures [18]. NMR-Solver takes a hybrid route, combining large-scale spectral matching with physics-guided fragment optimization rather than requiring either retrieval or unconstrained generation to carry the entire inference [19].

MS-based workflows likewise occupy a continuum. SIRIUS operationalizes a hierarchical route from tandem spectra through molecular-formula and structural information toward candidate ranking [20]. Domain-inspired formula transformers provide learned intermediate representations that connect spectra to formulae, fingerprints, and candidate-level inference [21]. At the more open end of the search space, de novo systems can generate

molecular hypotheses without restricting the final proposal step to direct retrieval of a pre-existing reference spectrum [22–24]. Their methodological advantage is expanded candidate reach; their corresponding liability is that generation introduces questions of chemical validity, stereochemical completeness, ranking, and domain transfer that library search partly avoids.

Foundation representations add another layer. DreaMS shows that large collections of tandem spectra can support self-supervised molecular representations transferable to downstream tasks [25]. Such pretraining may reduce dependence on exhaustively labeled datasets, but it does not remove the need to establish whether the learned representation covers the structural and instrumental domain of a particular natural product.

The resulting architecture is therefore not "spectrum in, structure out." **Figure 1** presents the proposed evidence-qualified synthesis that distinguishes the computational operations producing a candidate from the later operations required to make that candidate defensible.

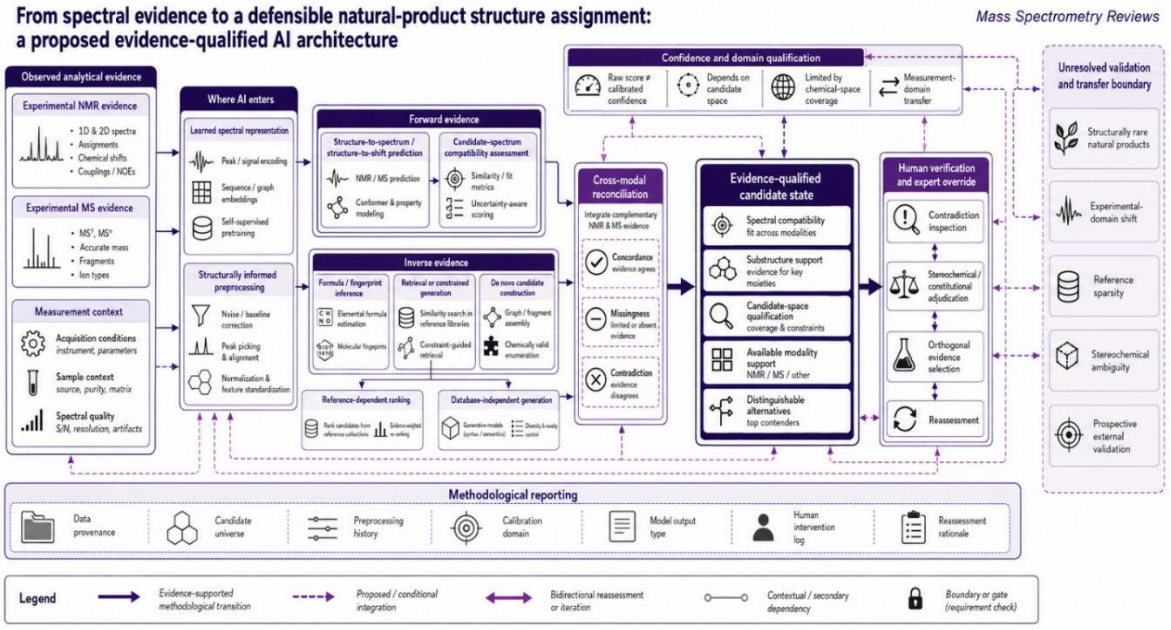

Figure 1. From spectral evidence to a defensible natural-product structure assignment: a proposed evidence-qualified AI architecture

Integration across analytical modalities

Natural-product elucidation rarely depends on a single observable, yet multimodal AI is not simply a matter of concatenating spectra. Integration first requires compatible data provenance. NMRexp illustrates the scale now attainable for experimental NMR resources, with 3.3 million experimental spectra, but a large corpus does not by itself guarantee balanced representation of solvents, nuclei, acquisition conditions, structural classes, or unusual natural-product chemistry [26]. The distinction between simulated and experimental domains is therefore methodological rather than merely logistical: a model can perform well against internally consistent synthetic data while remaining vulnerable to shifts introduced by real acquisition and curation.

Multimodal inference can nevertheless change the structure-search problem by allowing one modality to constrain ambiguities left by another. SECS combines raw spectroscopic modalities with learned representations, candidate search, confidence estimation, and a pilot comparison with expert chemists, providing evidence that integration can occur across more than one computational stage rather than only at final candidate scoring [27]. This architecture is important because disagreement between modalities is itself informative. A candidate compatible with one spectrum but inconsistent with another should not simply accumulate a reduced aggregate score; the contradiction may indicate an incorrect constitution, stereochemistry, measurement assignment, or model-domain failure.

The synthesis proposed here therefore distinguishes fusion from reconciliation. Fusion combines information; reconciliation asks whether independently informative observations support the same structural hypothesis and what should happen when they do not. Missing modalities, correlated errors, unequal model quality, and acquisition-specific biases can all limit the value of nominally richer input. Multimodal systems are consequently

strongest when they preserve evidence provenance and expose disagreement rather than hiding it within a single latent score.

Table 2 compares the principal ways current computational technologies combine or preserve analytical evidence.

Table 2. Comparison of computational strategies for combining analytical evidence during structure elucidation

Strategy	Evidence combined	Where integration occurs	Main methodological advantage	Important failure condition
Sequential single-modality analysis	Separate NMR or MS outputs	After independent analysis	Preserves modality-specific interpretability	Late fusion may miss informative interactions
Hierarchical candidate constraint	Formula, fingerprint, predicted spectrum, database candidates	During candidate reduction	Efficiently shrinks search space	Early incorrect constraint can exclude the true structure
Conditional molecular generation	Spectrum plus chemical prior	During structure construction	Uses prior knowledge to control combinatorial search	Prior may dominate when spectra are weak
Multispectral neural inference	Multiple spectrum types	Within shared representation/model	Allows complementary constraints to interact	Correlated or missing modalities can distort inference
Hybrid matching and optimization	Reference matching plus chemical/physical constraints	During candidate refinement	Combines empirical similarity with structural constraints	Reference coverage and optimization assumptions remain limiting
Self-supervised spectral representation	Large unlabeled spectral corpora plus downstream labels	Before task-specific inference	Reduces dependence on exhaustive annotation	Pretraining coverage may not include rare target chemistry
Raw multimodal search with confidence	Multiple raw spectra, candidate search, confidence estimation	Across representation, search, and ranking	Preserves a route from evidence to candidate and uncertainty	Confidence can still fail under chemical-space or measurement-domain shift

Note: These strategies are not mutually exclusive. The table describes where evidence is integrated and what methodological vulnerability remains, not a universal ranking of technologies.

Confidence calibration and chemical-space limitations

A ranked candidate list does not by itself communicate how much confidence should be placed in the leading structure. This distinction becomes critical when models are used beyond simple retrieval. Uncertainty-calibrated NMR confirmation methods illustrate that a probability intended to describe candidate correctness is methodologically different from an uncalibrated similarity or ranking score [28]. Confidence-aware MS annotation likewise shows that candidates absent from direct spectral libraries can be prioritized with explicit confidence-oriented procedures, but such confidence remains conditional on the candidate universe and scoring assumptions used to construct it [29]. Identification Probability further formalizes the need to distinguish interpretable probabilistic confidence from arbitrary algorithmic scores [30].

Calibration is also inseparable from chemical space. A model can be well calibrated on structures resembling its training and validation data while becoming overconfident on unusual chemistry. Evaluation of MS prediction at the edges of represented chemical space demonstrates that structurally atypical compounds can expose weaknesses hidden by conventional in-domain benchmarks [31]. Natural products present an especially demanding version of this problem because structural novelty may coincide with sparse training representation, unusual fragmentation, conformational effects, and stereochemical ambiguity.

The synthesis proposed here therefore treats confidence as domain-qualified rather than intrinsic to the model. A candidate should be interpreted through both its model-derived score and the evidence that the relevant chemical, spectral, instrumental, and candidate spaces are sufficiently represented. **Table 3** separates the major confidence states and the evidential gaps that remain between them.

Table 3. Evidence states, uncertainty sources, and unresolved gaps in AI-assisted structural confidence

Evidence state	What can legitimately be concluded	Dominant uncertainty	Additional evidence needed
High similarity or rank	Candidate is favored within a defined comparison set	Candidate-set composition	Broader candidate challenge or orthogonal evidence
Calibrated model probability	Confidence has statistical meaning within the validated domain	Calibration transfer	External calibration under relevant domain shift
Multimodal concordance	Different analytical inputs support the same candidate	Correlated errors or shared model bias	Independent modality-specific checks
Out-of-domain prediction	Model has produced an answer for unfamiliar chemistry	Unknown error inflation	Novelty-aware evaluation and cautious interpretation
Experimental compatibility	Predicted and measured observables are consistent	Nonunique spectral compatibility	Discriminating experiments or stereochemical evidence
Proposed structural assignment	Evidence supports a working structure hypothesis	Residual constitutional or stereochemical alternatives	Expert adjudication and, where necessary, orthogonal confirmation

Note: The progression is proposed as an interpretive hierarchy; it does not imply that every case must pass through identical stages.

Human verification and expert override

Human verification is sometimes described as a temporary safeguard that will become unnecessary as models improve. The available evidence supports a more durable methodological role. Hybrid NMR approaches can accelerate discrimination among supplied structures by combining quantum-mechanical calculations with machine learning, as demonstrated by ML-J-DP4 [32]. Such systems change the scale and speed of candidate adjudication, but they still operate on explicit structural alternatives and assumptions about the relationship between calculated and experimental observables.

Computational verification can also challenge previously accepted structures. DU8ML was applied to high-throughput validation and revision of complex alkaloid structures using machine-learning-augmented NMR calculations [33]. This is particularly relevant to natural products because plausible but incorrect assignments may survive when multiple constitutions or stereoisomers explain much of the available evidence. A computational discrepancy is therefore not merely a model failure; it can also be a signal that the proposed structure, the assignment, or the underlying experimental interpretation requires reconsideration.

Expert override should consequently be defined more precisely than "a chemist checks the answer." In the framework proposed here, expert verification includes identifying which spectral observation most strongly discriminates alternatives, recognizing chemically implausible model outputs, evaluating whether stereochemical information is sufficient, deciding when conflicting modalities should trigger reassignment, and selecting the next orthogonal experiment. Override is justified when the evidence supporting a model output is weaker than the evidence contradicting it, not simply because human judgement differs from an algorithm.

Human judgement itself is fallible. The methodological objective is therefore not to place the expert above the model, but to create a traceable interaction in which machine-derived candidates, confidence, contradictory evidence, and expert decisions remain inspectable. For difficult natural products, the strongest workflow is likely to be one in which disagreement initiates targeted reassessment rather than being suppressed into a single consensus score.

Benchmarking and method comparison

Benchmarking should distinguish whether a method predicts spectra accurately, retrieves the correct structure, generates chemically plausible candidates, ranks the correct candidate highly, remains calibrated, or transfers to new chemical space. These outcomes are related but not interchangeable. A model may achieve low spectral prediction error yet perform poorly when a database contains many near-isomeric candidates; conversely, a ranking system may perform well because the true structure is surrounded by few credible alternatives.

Data provenance is therefore part of model evaluation. Spectraverse demonstrates the methodological importance of curating and harmonizing small-molecule MS/MS libraries before spectra from different repositories are treated as comparable benchmark material [34]. In NMR, comparable concerns arise from solvent, acquisition,

assignment quality, simulated-versus-experimental spectra, and structural similarity between training and evaluation sets. Benchmark results should consequently be interpreted as properties of a model–dataset–candidate-space combination rather than as context-free measures of "elucidation performance."

Natural-product relevance also requires deliberate stress testing. Random molecular splits can reward interpolation among close analogues while understating the difficulty of rare scaffolds. Candidate sets generated from a known molecular formula may be much easier than open-ended de novo generation. Multimodal methods may benefit from inputs unavailable in routine isolation workflows. Human comparisons can also be misleading if experts and algorithms receive different information.

The interpretation priorities derived from these distinctions are summarized in **Table 4**.

Table 4. Benchmark questions that should precede interpretation of AI structure-elucidation performance

Benchmark question	Favorable result demonstrates	It does not establish	Priority stress test
Are spectra predicted accurately?	Forward-model fidelity in the tested domain	Correct inverse structural assignment	Novel scaffold and experimental-domain transfer
Is the correct candidate highly ranked?	Effective discrimination within that candidate set	Ability to generate an absent structure	Hard negative and near-isomer challenges
Can the model generate valid molecules?	Learned structural constraints	Spectral correctness of generated candidates	Spectral consistency and stereochemical completeness
Is confidence calibrated?	Score–outcome correspondence in validation data	Calibration after domain shift	External chemical-space calibration
Do modalities agree?	Cross-modal consistency	Independence of their errors	Missing or contradictory modality tests
Does AI match expert performance?	Comparative performance under a defined protocol	General expert replacement	Blinded, information-matched prospective comparison
Is performance reproducible?	Stability under specified data and preprocessing	Transfer across laboratories or instruments	Independent external replication

Note: The priorities are proposed methodological tests derived from the review synthesis rather than a validated universal benchmark standard.

Figure 2 integrates these benchmark distinctions with the evidence boundaries that determine how far a computational result can support a structural claim.

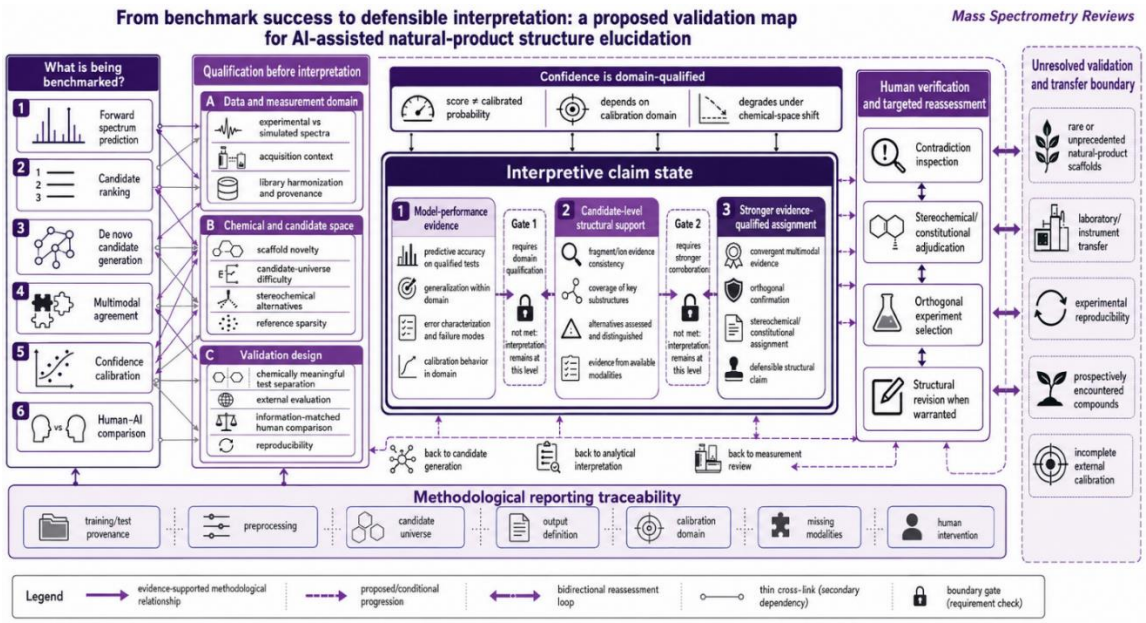

Figure 2. From benchmark success to defensible interpretation: a proposed validation map for AI-assisted natural-product structure elucidation

Methodological reporting standards

Methodological reporting should make it possible to reconstruct not only the model but also the evidential conditions under which a structural output was produced. Established metabolomics guidance already emphasizes that annotation and identification should be reported according to the evidence actually available rather than collapsed into an undifferentiated claim of identity [35]. AI-assisted elucidation requires the same discipline, extended to computational provenance.

At minimum, reports should state the analytical input used by the model, whether spectra were experimental or simulated, preprocessing and peak-handling procedures, training and evaluation data provenance, structural overlap between data partitions, candidate-generation rules, candidate-database restrictions, and the exact output being evaluated. A formula prediction, structural fingerprint, spectral similarity, top-ranked candidate, generated molecule, and calibrated probability are different objects and should be named accordingly.

The manuscript further proposes that calibration information should include the domain on which probability-like outputs were validated. Reporting a confidence value without describing chemical-space coverage, candidate-set construction, or calibration data risks giving numerical precision to an unknown transfer problem. Multimodal studies should additionally disclose which modalities were available for each sample, how missing information was handled, whether spectra were fused before or after candidate generation, and how contradictions between modalities were resolved.

Human involvement should also be traceable. Expert removal of implausible candidates, correction of assignments, selection of priors, reinterpretation of spectra, or choice of orthogonal experiments can materially affect apparent system performance. These interventions should therefore be reported as part of the method rather than hidden as informal post-processing.

The resulting standard is proposed rather than established consensus: a reproducible AI-assisted elucidation report should enable another investigator to determine what the model observed, what candidate space it searched, what uncertainty it expressed, what the human changed, and which evidence ultimately justified the structural claim.

Unresolved problems and future directions

The principal unresolved problem is not simply whether models can achieve higher benchmark accuracy. It is whether performance remains trustworthy when the compound is chemically unusual, analytical measurements are imperfect, reference spectra are sparse, candidate alternatives are stereochemically dense, or the experimental setting differs from training data. These conditions are common precisely where natural-product elucidation is most difficult.

Future benchmarks should therefore emphasize prospective and chemically separated evaluation rather than further optimization on repeatedly reused random splits. Newly isolated compounds, chronology-based holdouts, rare scaffolds, cross-laboratory spectra, and deliberately difficult constitutional or stereochemical alternatives would provide stronger tests of transfer. Multimodal systems also require ablation studies that determine which modality changes the result and whether agreement reflects independent evidence or correlated modeling error.

Confidence evaluation should move with the model into those harder domains. Calibration curves, coverage measures, abstention behavior, and failure detection should be assessed under increasing chemical and measurement shift rather than only within the training distribution. Human–AI comparisons should likewise be blinded, information matched, and designed to measure complementarity, error correction, and escalation decisions rather than merely which participant produces the highest top-one accuracy.

The most useful future systems may therefore be those that know when to narrow a hypothesis, when to expose alternatives, when to request additional evidence, and when not to make a stronger structural claim. That objective is less dramatic than autonomous elucidation, but methodologically more compatible with the evidential demands of natural-product chemistry.

Conclusion

AI-assisted natural-product structure elucidation is best understood as a collection of evidence-transforming methods rather than a single automated task. Learned spectral representations, forward prediction, formula and fingerprint inference, candidate retrieval, generative modeling, multimodal integration, confidence estimation, and computational verification each contribute different information and inherit different failure modes.

The central contribution of this review is a proposed evidence-qualified interpretation in which structural outputs become progressively stronger only when the computational operation, candidate space, analytical modalities, calibration domain, and verification evidence are made explicit. A high-ranking candidate is not equivalent to a confirmed structure; multimodal agreement is not automatically independent corroboration; and calibrated confidence cannot be presumed to survive substantial chemical-space or measurement-domain shift.

Human verification remains important not because computational methods must remain subordinate to traditional practice, but because difficult structural assignments require decisions about contradiction, stereochemistry, missing evidence, alternative hypotheses, and the next discriminating experiment. The strongest role for AI is therefore not merely to accelerate candidate production, but to make structural reasoning more searchable, comparable, uncertainty-aware, and auditable.

Progress should ultimately be judged by external transfer, reproducible evidence handling, calibrated uncertainty, transparent human–AI interaction, and successful resolution of compounds that are genuinely difficult rather than merely familiar to the benchmark. Until those conditions are demonstrated prospectively across diverse natural-product chemistry, AI-assisted elucidation should be interpreted as powerful methodological support for structural assignment, not as universal autonomous proof.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Hu G, Qiu M. Machine learning-assisted structure annotation of natural products based on MS and NMR data. Nat Prod Rep. 2023;40(11):1735-53. doi:10.1039/D3NP00025G
2. 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
3. Klukowski P, Riek R, Güntert P. Machine learning in NMR spectroscopy. Prog Nucl Magn Reson Spectrosc. 2025;148-149:101575. doi:10.1016/j.pnmrs.2025.101575
4. Zhang C, Idelbayev Y, Roberts N, Tao Y, Nannapaneni Y, Duggan BM, et al. Small molecule accurate recognition technology (SMART) to enhance natural products research. Sci Rep. 2017;7(1):14243. doi:10.1038/s41598-017-13923-x
5. Reher R, Kim HW, Zhang C, Mao HH, Wang M, Nothias LF, et al. A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. J Am Chem Soc. 2020;142(9):4114-20. doi:10.1021/jacs.9b13786
6. Kim HW, Zhang C, Reher R, Wang M, Alexander KL, Nothias LF, et al. DeepSAT: Learning molecular structures from nuclear magnetic resonance data. J Cheminform. 2023;15(1):71. doi:10.1186/s13321-023-00738-4
7. Priessner M, Lewis RJ, Johansson MJ, Goodman JM, Janet JP, Tomberg A. HSQC spectra simulation and matching for molecular identification. J Chem Inf Model. 2024;64(8):3180-91. doi:10.1021/acs.jcim.3c01735
8. Hu F, Chen MS, Rotskoff GM, Kanan MW, Markland TE. Accurate and efficient structure elucidation from routine one-dimensional NMR spectra using multitask machine learning. ACS Cent Sci. 2024;10(11):2162-70. doi:10.1021/acscentsci.4c01132
9. Li Y, Xu H, Kumar A, Wang DS, Heiss C, Azadi P, et al. TransPeakNet for solvent-aware 2D NMR prediction via multi-task pre-training and unsupervised learning. Commun Chem. 2025;8(1):51. doi:10.1038/s42004-025-01455-9
10. Xu F, Guo W, Wang F, Yao L, Wang H, Tang F, et al. Toward a unified benchmark and framework for deep learning-based prediction of nuclear magnetic resonance chemical shifts. Nat Comput Sci. 2025;5(4):292-300. doi:10.1038/s43588-025-00783-z

11. Goldman S, Li J, Coley CW. Generating molecular fragmentation graphs with autoregressive neural networks. Anal Chem. 2024;96(8):3419-28. doi:10.1021/acs.analchem.3c04654
12. Nowatzky Y, Russo FF, Lisec J, Kister A, Reinert K, Muth T, et al. FIORA: Local neighborhood-based prediction of compound mass spectra from single fragmentation events. Nat Commun. 2025;16(1):2298. doi:10.1038/s41467-025-57422-4
13. Young A, Röst H, Wang B. Tandem mass spectrum prediction for small molecules using graph transformers. Nat Mach Intell. 2024;6(4):404-16. doi:10.1038/s42256-024-00816-8
14. Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS. CFM-ID 4.0: More accurate ESI-MS/MS spectral prediction and compound identification. Anal Chem. 2021;93(34):11692-700. doi:10.1021/acs.analchem.1c01465
15. Hong Y, Li S, Ye Y, Tang H. FIDDLE: A deep learning method for chemical formulas prediction from tandem mass spectra. Nat Commun. 2025;16(1):11102. doi:10.1038/s41467-025-66060-9
16. Yao L, Yang M, Song J, Yang Z, Sun H, Shi H, et al. Conditional molecular generation net enables automated structure elucidation based on 13C NMR spectra and prior knowledge. Anal Chem. 2023;95(12):5393-401. doi:10.1021/acs.analchem.2c05817
17. Priessner M, Lewis RJ, Lemurell I, Johansson MJ, Goodman JM, Janet JP, et al. Advancing structure elucidation with a flexible multi-spectral AI model. Angew Chem Int Ed Engl. 2026;65(2):e17611. doi:10.1002/anie.202517611
18. Han C, Pan X, Zhang Y. End-to-end molecular structure elucidation from multimodal NMR spectra images using vision transformers. Chem Sci. 2026;17(26):13072-82. doi:10.1039/D6SC02352E
19. Jin Y, Wang JJ, Xu F, Ji X, Gao Z, Zhang L, et al. NMR-Solver: Automated structure elucidation via large-scale spectral matching and physics-guided fragment optimization. Nat Commun. 2026;17(1):4740. doi:10.1038/s41467-026-71315-0
20. 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
21. 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
22. 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
23. Litsa EE, Chenthamarakshan V, Das P, Kavraki LE. An end-to-end deep learning framework for translating mass spectra to de-novo molecules. Commun Chem. 2023;6(1):132. doi:10.1038/s42004-023-00932-3
24. Wang H, Zhang H, Zhu ZJ. Structure-informed deep generation enables de novo metabolite annotation in untargeted metabolomics. Nat Commun. 2026;17(1):5426. doi:10.1038/s41467-026-72149-6
25. Bushuiev R, Bushuiev A, Samusevich R, Brungs C, Sivic J, Pluskal T. Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS. Nat Biotechnol. 2026;44(4):630-40. doi:10.1038/s41587-025-02663-3
26. Wang JJ, Jin Y, Zhi CY, Liu YJ, Huang XH, Xu F, et al. NMRexp: A database of 3.3 million experimental NMR spectra. Sci Data. 2025;12(1):1954. doi:10.1038/s41597-025-06245-5
27. Mirza A, Patiny L, Jablonka KM. End-to-end multimodal structure elucidation from raw spectra combining contrastive learning and evolutionary algorithms. Nat Commun. 2026;17(1):5013. doi:10.1038/s41467-026-73846-y
28. Kotlyarov R, Howarth A, Goodman JM. DP5 without DFT: Uncertainty-calibrated graph neural net accelerates structure confirmation via NMR. Chem Sci. 2026;17(18):9138-47. doi:10.1039/D5SC06988B
29. 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
30. 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

31. Brown SM, Allgair E, Kryštůfek R. Mapping the edges of mass spectral prediction: Evaluation of machine learning EIMS prediction for xeno amino acids. Anal Chem. 2025;97(19):10282-8. doi:10.1021/acs.analchem.5c00286
32. Tsai YH, Amichetti M, Zanardi MM, Grimson R, Hernández Daranas A, Sarotti AM. ML-J-DP4: An integrated quantum mechanics-machine learning approach for ultrafast NMR structural elucidation. Org Lett. 2022;24(41):7487-91. doi:10.1021/acs.orglett.2c01251
33. Novitskiy IM, Kutateladze AG. DU8ML: Machine learning-augmented density functional theory nuclear magnetic resonance computations for high-throughput in silico solution structure validation and revision of complex alkaloids. J Org Chem. 2022;87(7):4818-28. doi:10.1021/acs.joc.2c00169
34. Gupta V, Qiang H, Chung HH, Herbst E, Skinnider MA. Comprehensive curation and harmonization of small-molecule MS/MS libraries in Spectraverse. Anal Chem. 2026;98(5):3934-43. doi:10.1021/acs.analchem.5c06256
35. Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'Auria JC, Ewald J, et al. Mass spectrometry-based metabolomics: A guide for annotation, quantification and best reporting practices. Nat Methods. 2021;18(7):747-56. doi:10.1038/s41592-021-01197-1