



Can Machine Learning Resolve Ambiguous Natural Product Structures Using Spectroscopic Evidence, Molecular Constraints, Candidate Ranking, and Automated Structure Elucidation?

Peter Nilsson^{1*}, Eva Johansson¹, Lars Andersson²

¹Department of ML-Assisted Structure Elucidation, Faculty of Pharmacy, Karolinska Institute, Stockholm, Sweden.

²Department of Spectroscopic Evidence and Candidate Ranking, Faculty of Pharmacy, Lund University, Lund, Sweden.

ABSTRACT

Natural-product structure elucidation remains a challenging inference problem because spectroscopic evidence is often incomplete, noisy, or nonunique, requiring reconciliation with molecular formulae, chemical plausibility, and stereochemical alternatives. Machine learning has increasingly been applied across this workflow, yet the term "identification" encompasses heterogeneous tasks from peak interpretation and formula prediction to full-structure generation. This scoping review mapped peer-reviewed evidence published from 2017 to 2025 concerning machine-learning-assisted structure inference for natural products and small molecules using NMR, tandem mass spectrometry, infrared, UV, and multimodal spectral data. Following a PRISMA-ScR-compatible evidence-mapping approach, we charted analytical modalities, prediction tasks, molecular constraints, candidate-generation strategies, ranking methods, validation designs, uncertainty treatments, and transferability limits. The mapped literature reveals a distributed automation landscape rather than a single solved workflow. NMR-based methods increasingly extract structure-sensitive constraints, mass-spectral approaches narrow formula and candidate spaces while inferring classes and fingerprints, and infrared or multimodal models contribute orthogonal evidence. Probabilistic and learned scoring systems prioritize competing candidates effectively. However, exact two-dimensional structure recovery, stereochemical assignment, confidence calibration, candidate-set completeness, and out-of-domain transfer remain distinct unresolved challenges. We therefore propose a constrained-inference interpretation in which machine learning progressively reduces ambiguity through staged constraint extraction and candidate prioritization, rather than replacing structural proof with a single model score. Current evidence supports machine learning as a powerful assistive tool, but fully autonomous natural-product identification is not yet universally achievable.

Key Words: Natural products, Structure elucidation, Machine learning, Spectroscopy, Candidate ranking, Multimodal analysis

eIJPPR 2025; 15(4):103-112

HOW TO CITE THIS ARTICLE: Nilsson P, Johansson E, Andersson L. Can Machine Learning Resolve Ambiguous Natural Product Structures Using Spectroscopic Evidence, Molecular Constraints, Candidate Ranking, and Automated Structure Elucidation? Int J Pharm Phytopharmacol Res. 2025;15(4):103-12. <https://doi.org/10.51847/i96x23Gau3>

INTRODUCTION

Natural-product structure elucidation is rarely a simple mapping from one spectrum to one molecule. Experimental spectra encode overlapping forms of evidence whose informativeness depends on acquisition conditions, molecular size, conformational behavior, candidate availability, and the structural distinction being

tested. Recent work therefore places machine learning across several functions rather than at a single endpoint: spectral interpretation, property prediction, formula inference, substructure extraction, candidate generation, candidate ranking, and structure validation [1]. This breadth creates a semantic problem for the field because successful class prediction, correct molecular formula, correct connectivity, and complete stereochemical

Corresponding author: Peter Nilsson

Address: Department of ML-Assisted Structure Elucidation, Faculty of Pharmacy, Karolinska Institute, Stockholm, Sweden.

E-mail: ✉ peter.nilsson@ki.se

Received: 15 April 2025; **Revised:** 10 August 2025; **Accepted:** 17 August 2025

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



assignment represent different levels of structural resolution.

The practical consequence is that “Can machine learning resolve an ambiguous structure?” is better treated as a sequence of constrained decisions. Computational NMR can expose inconsistencies in proposed natural-product structures and prioritize revisions, while direct inverse models can infer structural hypotheses from routine one-dimensional NMR under defined benchmark conditions [2, 3]. These achievements do not establish a general autonomous solution. This review therefore treats structure elucidation as progressive uncertainty reduction: spectra generate constraints; constraints delimit candidate space; candidate models assign relative support; and unresolved disagreement remains visible until independent evidence or expert interpretation is sufficient. That constrained-inference formulation is proposed here as an organizing synthesis, not as a validated universal algorithm.

Scope of the review and questions being mapped

The review was designed as a scoping review because the relevant evidence spans heterogeneous analytical modalities, computational tasks, candidate spaces, and validation endpoints. PRISMA-ScR provides the reporting framework for making the search, screening, and mapping process transparent without implying a pooled estimate of

model effectiveness [4]. Updated JBI guidance informed the use of explicit review questions, eligibility criteria, data charting, and evidence mapping rather than conversion of heterogeneous benchmark results into a meta-analytic performance claim [5].

The review asked five connected questions. First, what structural information is extracted or predicted from NMR, MS/MS, infrared, UV, or multimodal evidence? Second, how are molecular formulae, chemical validity, fragmentation behavior, retention information, and other constraints used to construct candidate sets? Third, how do machine-learning systems rank, generate, or attach confidence to competing structures? Fourth, which methods have been tested on ambiguous, difficult, library-absent, or revised structures, especially where stereochemistry matters? Fifth, what benchmark and validation gaps prevent apparent model accuracy from being interpreted as general structure-elucidation reliability? The scope therefore includes both automated and hybrid workflows but excludes claims of complete identification when the reported task stops at formula, class, fingerprint, or constitution-level prediction.

The search-and-selection logic is summarized in **Table 1**, while **Figure 1** presents the auditable evidence-selection pathway without reconstructing unavailable upstream database-yield counts.

Table 1. Search and selection architecture used to delimit the mapped evidence

Review element	Operationalization in this review	Analytical purpose	Verified outcome or boundary
Time window	Peer-reviewed journal articles published 2017–2025	Capture current ML-enabled spectroscopy while retaining necessary foundations	No earlier or later study entered the mapped set
Evidence focus	NMR, MS/MS, IR, UV, multimodal spectroscopy, candidate generation, ranking, uncertainty, or automated elucidation	Keep inclusion tied to an explicit structure-inference role	Adjacent analytical-ML papers without a structural claim were excluded
Publication eligibility	DOI-verifiable journal article in an eligible Q1 venue	Reduce metadata ambiguity and maintain source quality	Conference papers, preprints, reports, and duplicate versions were excluded
Screening unit	Unique DOI-bearing candidate article	Prevent double counting of online-first/final versions	48 candidates entered the auditable screening ledger
Selection decision	Exact contribution to a review question plus identifiable validation or limitation	Avoid inclusion based only on topical proximity	16 candidates were excluded after relevance and fit verification
Final mapping unit	Included article coded for modality, task, constraints, validation, uncertainty, and boundary conditions	Support cross-pipeline evidence mapping	32 articles formed the final mapped evidence set

Note: Numerical database yields and duplicate-removal totals preceding creation of the unique candidate ledger were not reconstructed because they were not reproducibly available from the search interfaces used.

Figure 1 is positioned immediately after **Table 1** because the verified screening ledger is most interpretable alongside the review scope and selection architecture.

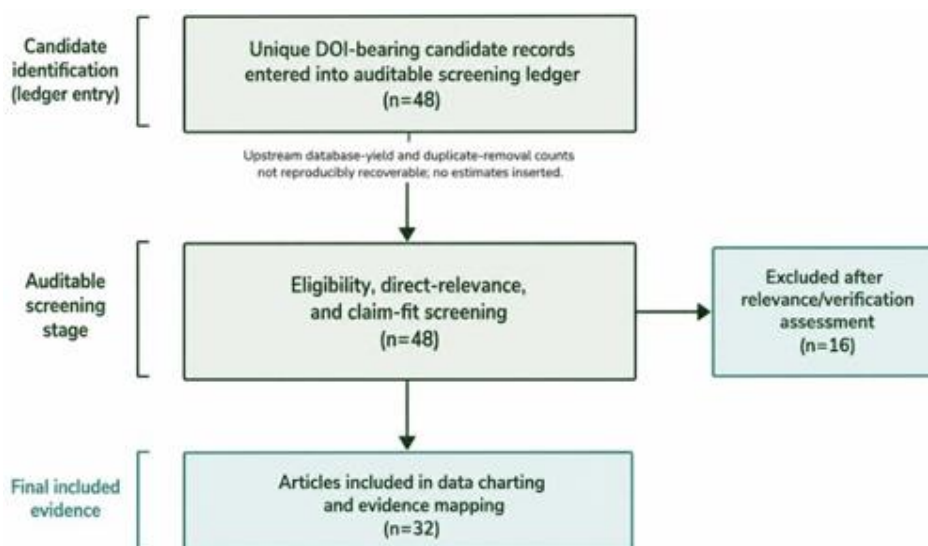


Figure 1. Auditable evidence-selection pathway for the scoping review

Protocol and prisma-scr reporting framework

The protocol defined the review as an evidence-mapping exercise centered on computational assistance for molecular-structure inference rather than a comparative effectiveness review. Reporting was organized to retain traceability from review question to eligibility decision, extraction field, synthesis category, and manuscript claim, consistent with PRISMA-ScR expectations for transparent scoping-review reporting [4]. Because the field uses different endpoints—peak recovery, formula identification, class assignment, fingerprint prediction, top-k candidate recovery, connectivity, stereochemical discrimination, or full generated structure—the protocol prohibited pooling these outcomes into a common accuracy measure.

The review process had four operational stages. First, candidate literature was located using combinations of natural-product structure elucidation, spectroscopy, machine learning, molecular formula, substructure, candidate generation, candidate ranking, confidence, stereochemistry, automated elucidation, and modality-specific terms. Second, bibliographic eligibility and DOI identity were verified. Third, eligible articles were charted against a common set of computational and analytical fields. Fourth, evidence was mapped according to where each method acted within the elucidation pipeline and the strongest structural claim its reported validation permitted. JBI scoping-review guidance supported this iterative charting and mapping logic while allowing refinement of categories as heterogeneity became visible [5]. The protocol did not assign causal effects, clinical outcomes, or universal model-performance estimates.

Eligibility criteria

Articles were eligible when they were peer-reviewed journal publications from 2017–2025, had a verifiable DOI, appeared in an eligible Q1 journal, and contributed directly to at least one structural-inference domain. Eligible contributions included spectral preprocessing with structural relevance, learned NMR or MS properties, molecular-formula or substructure inference, candidate construction, candidate scoring, de novo generation, uncertainty estimation, stereochemical discrimination, benchmark design, or natural-product validation.

Articles were excluded when they were outside the date range, duplicated another version, lacked journal or DOI verification, were conference-only or preprint-only records, or addressed spectroscopy or molecular machine learning without a direct role in structure inference. Studies were also excluded when their inclusion would require extrapolation beyond the chemical space, analytical modality, or validation regime actually examined.

Information sources and search strategy

Searches combined bibliographic indexing, publisher records, DOI metadata, and journal-level ranking verification. PubMed-indexed records and publisher pages were used to resolve titles, authorship, year, journal, volume, pages or article numbers, and DOI; DOI metadata provided additional identity checking, while journal-ranking records were used only to verify publication-venue eligibility. Search concepts were combined iteratively rather than restricted to one phrase.

Core terms covered natural products, structure elucidation, NMR, tandem mass spectrometry, infrared, UV, multimodal spectroscopy, molecular formula, substructure, candidate generation, candidate ranking, confidence, stereochemistry, de novo generation, automated elucidation, benchmark, uncertainty, and

revised or ambiguous structures. Citation chasing and close-topic searches were used to identify methodological anchors when newer papers assumed an established workflow.

Evidence selection

Screening was performed at the article level using DOI identity as the deduplication key. The auditable candidate ledger contained 48 unique papers considered sufficiently relevant for explicit screening. Each candidate was tested against publication eligibility, direct relevance to at least one review question, and whether a precise manuscript claim could be supported without enlarging the source's stated chemical or analytical scope.

Sixteen candidates were not retained because stronger sources covered the same role, the work was peripheral to structural inference, or the validation endpoint was too indirect for the planned claim. Thirty-two articles entered the final evidence map. Because the search interfaces did not preserve reproducible total-yield and duplicate-removal statistics preceding construction of the candidate ledger, those upstream quantities were left unreported rather than estimated.

Data charting and coding

For every included article, charting recorded publication type, chemical or molecular space, analytical setting, data source, sample or structure count when reported, prediction task, operationalized structural endpoint, spectral representation, algorithmic method, comparator, validation design, evaluation metric, principal finding, uncertainty source, authors' limitation, and transferability boundary. Additional coding captured whether the article informed NMR constraint extraction, MS formula or candidate inference, multimodal spectroscopy, candidate enumeration, ranking, stereochemistry, generation, confidence, benchmarking, or human-machine adjudication.

The coding scheme also separated what a model predicted from what that prediction could establish chemically. Formula prediction, chemical-class inference, structural fingerprints, constitution-level recovery, and stereochemical assignment were therefore coded as different endpoints. Missing methodological details were retained as not reported rather than inferred. This distinction was necessary because apparently similar "identification" results can represent materially different levels of structural resolution.

Mapping and synthesis approach

Synthesis proceeded by locating each study within a proposed constrained-inference sequence rather than by averaging incompatible performance metrics. The

sequence begins with spectral evidence, proceeds through learned or computed constraints, candidate-space construction, candidate scoring or generation, ambiguity assessment, and final structural adjudication. This organization is a review-derived synthesis: individual studies establish particular components, whereas the integrated sequence is used here to compare where information is gained, where alternatives are discarded, and where uncertainty remains.

Within each stage, evidence was mapped on three dimensions. The first was structural resolution: formula, class, substructure, constitution, relative stereochemistry, or more complete structural assignment. The second was computational role: preprocessing, forward prediction, inverse inference, candidate enumeration, ranking, generative search, or confidence estimation. The third was validation context: benchmark dataset, experimental spectrum, library-absent candidate set, ambiguous structure, retrospective revision, or multimodal comparison. Studies were interpreted only at the strongest level directly supported by their validation.

This approach intentionally treats disagreement as informative. A high-ranking candidate can remain uncertain when the correct structure is absent from the search space, when stereoisomers are weakly distinguished by the available modality, or when model confidence is not demonstrably calibrated outside the training distribution. The later synthesis therefore asks not only whether machine learning produces a candidate, but what evidence constrains that candidate, what alternatives survive, and what additional measurement would be required to resolve the ambiguity.

PRISMA-scr evidence-selection results

The auditable screening ledger began with 48 unique DOI-bearing candidate articles. Sixteen were excluded because they duplicated an evidence role more directly supported elsewhere, were peripheral to molecular-structure inference, or lacked sufficient claim-level fit. Thirty-two studies entered charting and mapping. Upstream database yields and duplicate-removal totals were not reconstructed because those values were not reproducibly retained. Accordingly, the selection results describe the verified screening stage rather than presenting estimated identification counts. No quantitative synthesis was attempted because the included studies differed substantially in analytical modality, prediction target, candidate universe, and validation endpoint.

Characteristics of included studies

The included literature clustered around NMR constraint prediction, MS-based structural inference, inverse or multimodal spectroscopy, and candidate ranking or

generation. Designs ranged from benchmarked machine-learning models to natural-product validation studies and data resources. Natural-product-specific evidence was less abundant than general small-molecule evidence, making chemical-space transfer an explicit boundary. Each study was therefore mapped to the structural level it actually addressed. Forward models were distinguished from inverse models because accurate spectrum prediction does not by itself demonstrate that an unknown structure can be uniquely recovered. Likewise, candidate-ranking studies were separated from de novo generation studies, since ranking presupposes a candidate universe whereas generation attempts to construct one.

Spectroscopic inputs used for structure elucidation

From Raw Spectra to Candidate Molecular Structures

Spectral automation begins by converting measurements into machine-readable constraints. NMRNet illustrates why chemical-shift models require standardized benchmarks before comparisons are interpretable [6]. DEEP Picker showed that neural networks can deconvolve overlapping two-dimensional NMR signals, improving peak extraction without itself determining molecular structure [7].

NMR-based machine learning for structural constraint extraction

Inverse NMR models move from peak interpretation toward chemically meaningful constraints. Neural networks can infer molecular substructure information from NMR correlations [8], while machine-learning correction of lower-cost quantum calculations accelerates structure-sensitive chemical-shift prediction [9]. Recent systems extend this role through solvent-aware two-dimensional prediction, fast multi-parameter modeling, and quantified uncertainty [10-12]. Forward prediction strengthens candidate checking; inverse inference reconstructs constraints from observed spectra. Their errors also propagate differently. A biased forward predictor may mis-score an otherwise correct candidate, whereas an inverse model may omit a necessary fragment or connectivity hypothesis before ranking begins. For ambiguous natural products, these failure modes should therefore be diagnosed separately rather than summarized by one endpoint accuracy.

Natural-product evidence also shows why candidate comparison remains necessary. Direct NMR annotation can operate in complex mixtures, although overlap constrains transferability [13]. Automated DP4+ accelerates probabilistic comparison [14], and ML-J-DP4 couples quantum-mechanical and learned components for stereochemical discrimination [15]. Candidate definition and conformational treatment remain consequential. This

is particularly important when several diastereomers have similar one-dimensional signatures or when conformational averaging changes calculated observables. In such cases, machine learning reduces the computational burden of comparison but does not remove the need to ask whether the chosen NMR observables are sufficiently discriminating for the stereochemical question.

Mass-spectral evidence in candidate generation

High-resolution tandem MS contributes a different form of constraint. Fragmentation-tree analysis, bottom-up molecular-formula discovery, and transformer-based interpretation each reduce the candidate universe without making molecular formula synonymous with unique structure [16-18]. MIST-CF similarly learns formula inference from tandem spectra [19], while graph-transformer prediction of tandem spectra provides a forward-model signal for comparing structures against observed fragmentation [20].

The evidence then branches into inverse and generative tasks. End-to-end learning can produce de novo molecular hypotheses [21]; fragment-tree transformers demonstrate annotation in a miniaturized-MS setting [22]; and FIORA models local fragmentation for forward prediction [23]. CANOPUS predicts chemical classes without exact reference spectra [24], COSMIC ranks library-absent structures with confidence estimation [25], MSNovelist generates structures de novo [26], and LC-MS²Struct adds retention order as an orthogonal ranking constraint [27]. These outputs represent different structural levels. Formula inference constrains elemental composition; class prediction supplies broader chemical context; fingerprint or fragmentation models support constitutional ranking; and de novo generators enlarge the candidate universe. Because each task discards a different set of alternatives, a reported gain at one level cannot be assumed to propagate unchanged to the next. The review therefore treats MS evidence as a hierarchy of constraints rather than a single identification operation.

Infrared, uv, and multimodal spectroscopic contributions

Recent work treats infrared spectra as learnable inverse fingerprints. Transformer models can generate candidate structures from IR under defined conditions [28], while benchmarking shows that molecular-formula constraints can materially alter inverse decoding [29]. An equivariant representation has also predicted IR, Raman, UV, and NMR spectra from structure [30].

A paired computational IR–NMR resource for 177,000 patent-derived molecules expands cross-modal training but does not reproduce the full natural-product distribution or experimental artifacts [31]. Bidirectional deep learning further shows that structure-to-spectrum and spectrum-to-

structure mappings can be learned within defined domains [32]. Prospective natural-product generalization remains unresolved. UV evidence is less prominent in the selected inverse-elucidation literature and is therefore treated mainly as a complementary multispectral modality, not an independently validated route to complete assignment.

Molecular constraints and candidate generation

Candidate enumeration under chemical and biosynthetic constraints

Candidate generation determines what downstream ranking can recover. Formula inference restricts elemental composition [17], while de novo generators extend beyond database lookup [26]. This review proposes staged enumeration: apply hard constraints such as formula and valence before softer evidence such as spectral compatibility, retention behavior, or biosynthetic plausibility. Biosynthetic plausibility should prioritize candidates, not rescue a spectroscopically inconsistent structure.

A ranking model cannot recover a molecule excluded from its candidate universe, and a chemically plausible generator can still be wrong. Candidate completeness should therefore be reported separately from ranking quality. This is a proposed synthesis, not a validated universal enumeration algorithm. Candidate-generation

reports should distinguish mandatory chemical constraints from probabilistic or contextual evidence.

Machine-learning candidate ranking and confidence ranking competing structures

Once multiple structures survive, ranking asks which candidate is best supported. Automated DP4+ supplies NMR-based discrimination [14]; learned tandem-spectrum prediction supplies compatibility evidence [20]; COSMIC adds confidence-aware ranking for library-absent structures [25]; and retention order contributes an orthogonal LC-MS/MS discriminator [27]. These scores arise from different observables and are not numerically interchangeable.

Ranking should preserve score provenance and distinguish accumulated support from proof. Rank 1 can remain ambiguous when alternatives are indistinguishable, acquisition conditions shift, or the correct structure was never enumerated. **Table 2** contrasts representative ranking studies. The comparison emphasizes validation target rather than headline performance because a method can be strong at relative ranking while remaining unable to state whether the winning candidate is uniquely correct. Confidence is most informative when tied to a defined candidate universe, acquisition domain, and structural resolution.

Table 2. Representative study characteristics shaping candidate ranking and structural confidence

Study	Evidence used	Ranking target	Validation emphasis	Principal boundary
Franco <i>et al.</i> [14]	Calculated and experimental NMR	Candidate structures	DP4+ discrimination	Candidate and conformer dependence
Tsai <i>et al.</i> [15]	NMR shifts/couplings with QM-ML	Stereochemical alternatives	Relative candidate discrimination	Does not guarantee absolute configuration
Young <i>et al.</i> [20]	Predicted versus observed MS/MS	Molecular candidates	Forward-spectrum agreement	Instrument and collision-domain shift
Hoffmann <i>et al.</i> [25]	Fragmentation-derived structural evidence	Library-absent candidates	Confidence-aware ranking	Confidence is model- and domain-dependent
Bach <i>et al.</i> [27]	MS/MS plus retention order	Joint candidate list	Cross-modal ranking	Chromatographic-condition dependence
Alberts <i>et al.</i> [29]	IR plus molecular constraints	Generated candidates	Top-k inverse recovery	Benchmark space and sim-to-real limits

The manuscript-derived relationships underlying those heterogeneous ranking routes are integrated visually in **Figure 2**.

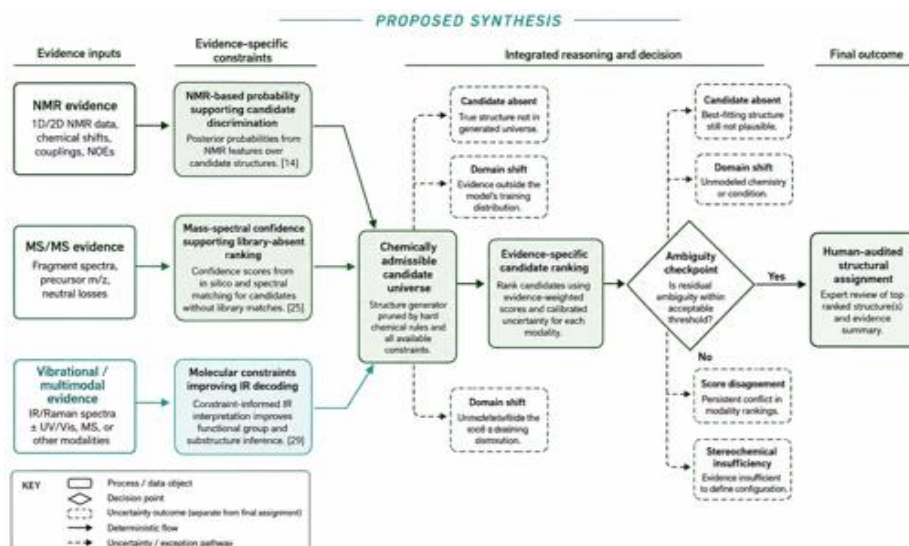


Figure 2. Proposed evidence map for progressive ambiguity reduction in machine-assisted structure elucidation

Automated and hybrid structure-elucidation workflows

Generative approaches to automated structure elucidation

Generative systems alter the candidate problem by constructing molecular hypotheses rather than only ranking a fixed database. MSNovelist demonstrates de novo generation conditioned by mass-spectral information [26], while spectrum-to-molecule deep learning similarly explores direct inverse generation [21]. IR-based inverse models extend this logic to vibrational fingerprints [28]. Their common strength is escape from strict library dependence; their common limitation is that chemical validity or high spectral compatibility does not establish that a generated molecule is the experimental structure.

Hybrid workflows are therefore more defensible than a single generative endpoint. Formula, spectral constraints, forward prediction, and subsequent ranking can be used to restrict generation and challenge its outputs. Bidirectional spectral models [32] make this iterative architecture increasingly plausible, but the integrated natural-product workflow remains a proposed synthesis requiring prospective validation.

Validation against ambiguous or revised structures

stereochemical assignment as the harder remaining problem

Retrospective structure revision provides a stringent test because the task is not simply retrieval of a familiar database entry. DU8ML showed that machine-learning-augmented NMR calculations can flag inconsistencies and support revision of complex alkaloid structures [2]. DP4+ automation [14] and ML-J-DP4 [15] further show that candidate-dependent NMR comparison can discriminate closely related stereochemical alternatives.

The boundary is crucial: constitutional recovery is not equivalent to stereochemical resolution. Fragmentation-

based MS and retention evidence can constrain and rank structures, but stereoisomers may remain weakly distinguishable by those data alone [27]. Validation should therefore report the structural level actually resolved and distinguish retrospective candidate discrimination from prospective elucidation when the correct answer is not pre-enumerated. A stronger challenge would conceal the final structure, include plausible constitutional and stereochemical decoys, and allow the workflow to fail explicitly when available evidence is insufficient. Such abstention would be scientifically preferable to forced rank-1 assignment.

Evidence gaps and benchmarking needs

Confidence estimation and candidate ambiguity

Uncertainty-aware NMR prediction demonstrates that model outputs can carry explicit uncertainty estimates [12], whereas COSMIC illustrates confidence-oriented candidate ranking in MS annotation [25]. These are not interchangeable quantities, and neither should automatically be read as the posterior probability that a complete molecular structure is correct. Calibration under chemical-space or instrument shift remains an independent problem.

Benchmark datasets and their coverage of natural product complexity

Benchmark standardization is improving for NMR [6], but modality-specific datasets still differ in molecular size, elemental composition, acquisition conditions, and task definition. Large computational multimodal resources broaden scale [31], yet simulated spectra and patent-derived molecules cannot be assumed to reproduce natural-product complexity. A decisive next benchmark would contain blinded, experimentally measured multimodal

spectra, difficult stereochemical alternatives, revised structures, candidate-absent cases, and predefined confidence reporting. That benchmark design is proposed here; existing studies do not yet establish its prospective performance. Benchmark reporting should additionally separate candidate recall, rank of the correct structure, calibration or abstention behavior, and the structural level counted as correct. Without those distinctions, improvements in candidate retrieval can be mistaken for improvements in complete elucidation, and cross-study comparisons become chemically misleading.

RESULTS AND DISCUSSION

The evidence supports a qualified answer: machine learning can reduce ambiguity through constraint extraction, forward prediction, candidate construction, and ranking. It does not establish that one architecture can convert arbitrary natural-product spectra into universally reliable, stereochemically complete structures. Ambiguity reduction and ambiguity elimination must remain distinct. The proposed constrained-inference view therefore changes the success criterion: a useful model need not

always name one molecule, but it should make clear which alternatives were removed, which remain plausible, why a candidate is preferred, and what additional evidence would most efficiently separate the survivors.

Human-machine collaboration in final structure assignment

Current evidence favors a division of labor in which computation makes large candidate spaces tractable while expert judgment adjudicates evidence conflict and requests additional measurements. Natural-product mixture annotation demonstrates the value of automated assistance in complex spectral settings [13], whereas retrospective revision shows that computational disagreement can itself be diagnostically useful [2]. The proposed role of the human is consequently not to override models intuitively, but to inspect unresolved alternatives, verify whether model assumptions apply, and decide when additional experimental evidence is necessary.

Table 3 maps the principal residual uncertainty states to the computational contribution and the corresponding need for adjudication.

Table 3. Evidence map of residual ambiguity and the corresponding human-machine task boundary

Residual uncertainty	Computational contribution	Evidence anchor	Appropriate adjudication
Overlapping spectral features	Automated deconvolution and feature extraction	DEEP Picker [7]	Confirm questionable peaks against raw data
Candidate-set omission	Confidence/ranking among available structures	COSMIC [25]	Expand enumeration rather than overtrust rank 1
Stereochemical alternatives	QM-ML/NMR discrimination	ML-J-DP4 [15]	Inspect conformers and obtain orthogonal stereochemical evidence
Instrument/domain shift	Context-specific learned annotation	TeFT [22]	Check acquisition compatibility before transfer
Synthetic-to-experimental gap	Large paired computational spectra	IR-NMR dataset [31]	Validate with experimental natural-product spectra
Model-score uncertainty	Predictive uncertainty estimation	Jonas and Kuhn [12]	Retain ambiguity when calibration is unsupported

The proposed constrained-inference view therefore treats automation as evidence management rather than authority transfer. A machine-generated structure becomes progressively more credible as independent constraints converge, but discordant evidence remains a reason for investigation rather than a quantity to hide inside a composite score.

Review limitations

This review maps heterogeneous computational and analytical evidence rather than estimating a pooled performance effect. Search interfaces did not preserve reproducible upstream identification and deduplication totals, so only the auditable candidate-screening stage is reported. Natural-product-specific validation was less extensive than general small-molecule benchmarking, and several included methods were developed on simulated

spectra, curated databases, or particular acquisition settings. Comparisons across studies are therefore conceptual rather than numerical. The review also focuses on spectroscopy-centered computational structure inference and does not claim exhaustive coverage of every quantum-chemical, crystallographic, biosynthetic, or chemical-derivatization strategy that may contribute to conventional structure elucidation.

CONCLUSION

Machine learning is increasingly capable of converting spectroscopic evidence into useful structural constraints, candidate molecules, rankings, and uncertainty signals. The evidence does not support treating these outputs as interchangeable forms of identification or as a universal replacement for experimental structural reasoning. This

scoping review proposes constrained inference as a unifying interpretation: automation progressively reduces candidate ambiguity while preserving the structural level, evidential provenance, and uncertainty attached to each decision. The most important remaining challenges are stereochemical discrimination, candidate completeness, confidence calibration, multimodal experimental benchmarking, and transfer to natural-product chemical space. Progress should therefore be judged by whether integrated workflows expose and resolve difficult ambiguity, not merely whether they recover familiar structures in retrospective benchmarks.

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] 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
- [3] 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
- [4] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann Intern Med.* 2018;169(7):467-73. doi:10.7326/M18-0850
- [5] Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. *JBIM Evid Synth.* 2020;18(10):2119-26. doi:10.11124/JBIES-20-00167
- [6] 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
- [7] Li DW, Hansen AL, Yuan C, Bruschweiler-Li L, Bruschweiler R. DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra. *Nat Commun.* 2021;12(1):5229. doi:10.1038/s41467-021-25496-5
- [8] Liu S, Cole JM. Automated determination of the molecular substructure from nuclear magnetic resonance spectra using neural networks. *J Chem Inf Model.* 2025;65(16):8435-47. doi:10.1021/acs.jcim.5c00499
- [9] Li J, Liang J, Wang Z, Ptaszek AL, Liu X, Ganoe B, et al. Highly accurate prediction of NMR chemical shifts from low-level quantum mechanics calculations using machine learning. *J Chem Theory Comput.* 2024;20(5):2152-66. doi:10.1021/acs.jctc.3c01256
- [10] 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
- [11] Yiu C, Honoré B, Gerrard W, Napolitano-Farina J, Russell D, Trist IML, et al. IMPRESSION generation 2—accurate, fast and generalised neural network model for predicting NMR parameters in place of DFT. *Chem Sci.* 2025;16(19):8377-82. doi:10.1039/D4SC07858F
- [12] Jonas E, Kuhn S. Rapid prediction of NMR spectral properties with quantified uncertainty. *J Cheminform.* 2019;11:50. doi:10.1186/s13321-019-0374-3
- [13] Egan JM, van Santen JA, Liu DY, Linington RG. Development of an NMR-based platform for the direct structural annotation of complex natural products mixtures. *J Nat Prod.* 2021;84(4):1044-55. doi:10.1021/acs.jnatprod.0c01076
- [14] Franco BA, Luciano ER, Sarotti AM, Zanardi MM. DP4+App: Finding the best balance between computational cost and predictive capacity in the structure elucidation process by DP4+: Factors analysis and automation. *J Nat Prod.* 2023;86(10):2360-7. doi:10.1021/acs.jnatprod.3c00566
- [15] Tsai YH, Amichetti M, Zanardi MM, Grimson R, Daranas AH, 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
- [16] 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



- [17] 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
- [18] 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
- [19] 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
- [20] Young A, Röst HL, 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
- [21] 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
- [22] Yang Y, Sun S, Yang S, Yang Q, Lu X, Wang X, et al. Structural annotation of unknown molecules in a miniaturized mass spectrometer based on a transformer enabled fragment tree method. *Commun Chem*. 2024;7(1):109. doi:10.1038/s42004-024-01189-0
- [23] 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
- [24] 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
- [25] 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
- [26] 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
- [27] 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
- [28] Alberts M, Laino T, Vaucher AC. Leveraging infrared spectroscopy for automated structure elucidation. *Commun Chem*. 2024;7(1):268. doi:10.1038/s42004-024-01341-w
- [29] Alberts M, Zipoli F, Laino T. Setting new benchmarks in AI-driven infrared structure elucidation. *Digit Discov*. 2025;4(7):1936-43. doi:10.1039/D5DD000131E
- [30] Zou Z, Zhang Y, Liang L, Wei M, Leng J, Jiang J, et al. A deep learning model for predicting selected organic molecular spectra. *Nat Comput Sci*. 2023;3(11):957-64. doi:10.1038/s43588-023-00550-y
- [31] Zipoli F, Alberts M, Laino T. IR-NMR multimodal computational spectra dataset for 177K patent-extracted organic molecules. *Sci Data*. 2025;12(1):1375. doi:10.1038/s41597-025-05729-8
- [32] Hu T, Zou Z, Li B, Zhu T, Gu S, Jiang J, et al. Deep learning for bidirectional translation between molecular structures and vibrational spectra. *J Am Chem Soc*. 2025;147(31):27525-36. doi:10.1021/jacs.5c05010