



Can Machine Intelligence Solve Retrosynthesis of Complex Natural Products Respecting Stereochemistry and Synthetic Practicality?

Zsolt Kovács^{1*}, Katalin Szabo¹, Gabor Nagy², Eszter Toth³

¹Department of ML for Retrosynthesis of Complex Natural Products, Faculty of Pharmacy, Semmelweis University, Budapest, Hungary.

²Department of Stereochemistry and Selectivity in Retrosynthesis, Faculty of Pharmacy, University of Debrecen, Debrecen, Hungary.

³Department of Synthetic Practicality and Route Design, Faculty of Pharmacy, University of Pécs, Pécs, Hungary.

ABSTRACT

Machine intelligence has transformed retrosynthesis from rule retrieval into data-driven reaction prediction and search, yet structurally complex natural products expose a harder question than whether a model can recover plausible precursors. Their synthesis often depends on stereochemical information, chemoselective control, protecting-group logic, long-range strategic choices, convergent assembly, specialized building blocks, and practical execution under conditions that are only incompletely represented in reaction corpora. This critical review evaluates whether contemporary machine-learning retrosynthesis can meet those demands. We distinguish single-step prediction from route planning, formal precursor validity from experimentally meaningful selectivity, and route search from synthetic strategy. Across current approaches, template-based models retain chemically explicit transformation priors but inherit coverage limits, whereas template-free architectures broaden prediction flexibility while remaining sensitive to representation, dataset bias, and local rather than route-level objectives. Stereochemistry is a particularly revealing stress test: improved encodings and editing strategies can reduce errors, but correct atom-level configuration does not by itself guarantee stereoselective feasibility across a multistep route. We therefore propose that natural-product retrosynthesis should be evaluated as a constrained decision problem in which structural correctness, selectivity, strategic coherence, material availability, and practical risk remain separable dimensions. The literature supports substantial progress in reaction prediction and search, but evidence for autonomous solution of complex natural-product synthesis remains limited by benchmark composition, sparse experimental validation, incomplete modeling of protecting groups and conditions, and weak route-level practicality criteria. Progress will require benchmarks built around expert syntheses, prospective laboratory testing, and models that represent uncertainty and strategy explicitly.

Key Words: Retrosynthesis, Natural products, Machine learning, Stereochemistry, Synthesis planning, Synthetic practicality

eIJPPR 2026; 16(3):156-165

HOW TO CITE THIS ARTICLE: Kovács Z, Szabo K, Nagy G, Toth E. Can Machine Intelligence Solve Retrosynthesis of Complex Natural Products Respecting Stereochemistry and Synthetic Practicality? Int J Pharm Phytopharmacol Res. 2026;16(3):156-65. <https://doi.org/10.51847/8pf8pEGgz7>

INTRODUCTION

Natural-product retrosynthesis is difficult not simply because targets are large, but because molecular complexity concentrates stereochemical information, functional-group relationships, ring topology, and strategically consequential asymmetry into a limited

number of decisions. Contemporary analysis of molecular complexity emphasizes that such structural features can drive advances in synthetic strategy rather than serving only as descriptors of molecular size [1]. For machine intelligence, this means that a useful retrosynthetic proposal must preserve strategically relevant complexity

Corresponding author: Zsolt Kovács

Address: Department of ML for Retrosynthesis of Complex Natural Products, Faculty of Pharmacy, Semmelweis University, Budapest, Hungary.

E-mail: ✉ zsolt.kovacs@semmelweis.hu

Received: 01 March 2026; **Revised:** 30 May 2026; **Accepted:** 04 June 2026

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.



while deciding where it should be created, simplified, or exploited.

Computer-assisted planning has nevertheless crossed an important evidential boundary. Computational planning has contributed to syntheses of complex natural products that were subsequently examined through laboratory chemistry, establishing that automated planning can generate chemically meaningful routes for selected demanding targets [2]. At field level, contemporary AI retrosynthesis now encompasses template-based, template-free, graph-based, sequence-based, generative, and search-coupled approaches, but their performance remains strongly dependent on representation, training data, and evaluation setting [3]. More recent work explicitly treats higher-level strategy as a separate capability rather than assuming that accurate one-step prediction automatically yields strong multistep planning [4].

The central question is therefore not whether machine intelligence can produce disconnections. It is whether it can retain stereochemical intent, reason over chemoselective molecular states, organize long-horizon and convergent decisions, qualify starting materials, and rank routes according to practical constraints. This review treats those requirements as related but non-interchangeable tests of retrosynthetic competence.

Review scope and critical approach

This critical review evaluates machine-intelligence methods at the evidential level at which they are tested. Reliability and interpretability analyses show that nominal predictive performance does not by itself establish that a retrosynthesis model is trustworthy across chemically distinct cases [5]. Likewise, comparisons coupling single-step models to multistep planners demonstrate that one-step model quality and route-planning outcomes are related but not interchangeable; the planner, stopping conditions, reaction model, and search configuration jointly shape the routes recovered [6].

Evaluation must also distinguish aggregate success from chemically meaningful failure. Granular error analysis has shown that conventional exact-match or top-*k* summaries can combine different error classes and thereby obscure why a prediction is chemically useful, invalid, or merely different from the recorded answer [7]. This is particularly consequential for natural products, where an alternative disconnection may be strategically valuable while a nominally close prediction may still corrupt stereochemistry or functional-group logic.

Accordingly, the review focuses on stereochemical representation, template dependence, selectivity, protecting groups, search, convergence, biosynthetic and hybrid planning, building-block assumptions, route scoring, and experimental execution. It does not claim

exhaustive literature coverage or PRISMA-level completeness. Comparisons are bounded by each study's chemical space, dataset, planner configuration, and validation design, and proposed cross-domain conclusions are identified as synthesis rather than established universal rules.

Representation of stereochemistry and three-dimensional complexity

Stereochemistry is not an optional annotation in natural-product retrosynthesis. Natural-product structural datasets contain stereochemical patterns that can be learned computationally, as demonstrated by machine-learning approaches for assigning stereochemical configurations from natural-product structural information [8]. This establishes that stereochemical regularity is computationally accessible, but it does not establish that a route planner can preserve or deliberately construct that stereochemistry across multiple transformations.

Reaction prediction supplies a second level of evidence. Transfer learning has improved regio- and stereoselective prediction within carbohydrate chemistry, including experimentally examined reactions, showing that selective outcome information can be learned effectively inside a defined chemical domain [9]. Yet carbohydrate reaction transfer does not imply general stereochemical competence for arbitrary polycyclic or highly functionalized natural products. A third problem is bookkeeping: RDChiral demonstrated that retrosynthetic template extraction and application require explicit handling of stereochemical retention, inversion, and related atom-mapping behavior [10]. A model can therefore fail even before mechanistic selectivity is considered if its representation loses configuration during transformation processing.

These observations motivate a proposed distinction between stereochemical bookkeeping and stereochemical reasoning. Bookkeeping preserves encoded configuration through representations and edits. Reasoning asks whether a transformation can establish, retain, invert, or exploit configuration under a particular substrate and condition set. Three-dimensional effects such as conformational strain, facial accessibility, substrate control, and transition-state organization remain only indirectly represented in most current graph- or string-based retrosynthesis systems.

Template-based and template-free retrosynthesis

Template-based and template-free retrosynthesis place different constraints on chemical generalization. Large-scale generative pretraining, exemplified by RSGPT, demonstrates that retrosynthetic prediction can be formulated through sequence generation over very large chemical corpora rather than through explicit retrieval of a fixed reaction-template library [11]. Iterative string editing

provides another template-free formulation in which reactants are constructed through successive molecular edits, offering a different inductive bias from direct sequence generation [12]. End-to-end graph-generative architectures instead represent retrosynthesis as molecular-graph editing, allowing reaction-center changes and precursor generation to be learned in graph space [13].

These architectures broaden the ways in which a model can propose disconnections, but natural-product planning exposes a trade-off between transformation freedom and constraint visibility. Explicit templates can preserve recognizable historical transformation patterns and make local reaction logic easier to inspect, yet they inherit extraction choices, rule granularity, precedent coverage, and stereochemical-processing limits. Template-free systems can propose chemistry outside an explicit template inventory, but broader generative freedom does not by itself establish mechanistic plausibility, chemoselective compatibility, or route-level practicality.

The relevant comparison is therefore not a binary contest between templates and no templates. Model class, representation, training chemistry, stereochemical handling, and downstream search jointly determine the domain in which predictions remain credible. For structurally complex natural products, architectural flexibility is most useful when accompanied by explicit

evidence about where the model is extrapolating and which chemical constraints remain unevaluated.

Natural products provide a particularly stringent stress test because the same local transformation may acquire different strategic value as stereochemical density and functional-group coupling increase. A disconnection that looks attractive in a one-step benchmark can require several protection, redox-adjustment, or conformationally sensitive operations before it becomes executable. Conversely, an apparently uncommon disconnection may be strategically superior if it collapses complexity or creates multiple stereocentres under substrate or catalyst control. We therefore propose evaluating architecture choice against the information that must remain visible during search: transformation precedent, stereochemical state, competing reactive sites, precursor accessibility, and the downstream consequences of each edit. This does not imply that every variable must be embedded in a single model. Modular systems may instead expose uncertainty or constraint violations to the planner. What matters is that route search should not treat a high-probability local prediction as sufficient evidence that the resulting molecular state belongs on a practical long-horizon route. The resulting boundary conditions are summarized in **Table 1**. The table contains no citation numbers; its evidence basis is the cited discussion above, while the interpretive distinctions are proposed where indicated.

Table 1. Boundary conditions governing interpretation of template-based and template-free retrosynthesis for structurally complex natural products

Boundary condition	Template-based implication	Template-free implication	Critical interpretation
Stereochemical density	Rule extraction and application must preserve local chirality.	Prediction must learn and reproduce stereochemical labels and transformations.	Correct connectivity is insufficient when configuration controls intermediate identity or downstream feasibility.
Rare or scaffold-specific chemistry	Sparse precedent restricts available or applicable templates.	Broader generation may recover alternatives but increases extrapolation risk.	Coverage and chemical credibility should be evaluated separately.
Reaction-context dependence	Localized rules may omit substrate-wide steric or functional-group effects.	Richer learned context remains constrained by chemistry represented in training data.	Neither architecture alone establishes chemoselective feasibility.
Transformation novelty	Explicit rules constrain novelty to encoded or composable precedent.	Flexible generation permits less precedented disconnections.	Novelty should increase validation burden rather than count as evidence of superiority.
Route-level coupling	One-step templates do not encode global strategy by default.	One-step generators likewise require route-level objectives.	Strategic coherence and practicality require additional evaluation.

Note: The matrix is a proposed critical synthesis. It does not define validated thresholds for choosing a retrosynthesis architecture.

Selectivity, protecting groups, and chemoselective reasoning

A retrosynthetic step can be formally plausible while remaining unusable because the proposed substrate contains several competing reactive sites or because the

molecular state required for the transformation is inaccessible without protection. Recent work explicitly incorporating protection strategies and route scoring into computer-assisted synthesis planning demonstrates that protecting-group decisions can be represented as planning

operations rather than left entirely to post hoc human repair [14]. This is an important advance for natural products, where protection may determine not only compatibility with one reaction but also the order in which oxidation states, stereocentres, and ring systems can be manipulated. Selectivity nevertheless requires more than identifying protected and unprotected functional groups. Engineered and learned molecular representations differ in their ability to encode the local and non-local structural information governing reactivity and selectivity [15]. Within a defined transformation class, negative-data augmentation and multimodel integration can improve prediction of regioselective C–H halogenation, showing that site competition can be learned when the reaction domain and labels are sufficiently constrained [16]. Such evidence supports reaction-class-specific selectivity modeling, not general autonomous chemoselective reasoning across polyfunctional natural products.

The critical distinction is between reaction availability and state-conditioned reaction feasibility. A planner may know that a transformation exists while failing to represent which functional groups must be masked, which stereoelectronic arrangement is required, or whether an earlier step changes the selectivity of a later one. Protecting-group and selectivity models therefore need to participate in search rather than merely rerank a completed route. For complex targets, a useful plan must preserve a sequence of chemically compatible molecular states, not just a sequence of recognizable reactions.

Long-horizon planning and convergent synthesis

Multistep retrosynthesis is a search problem because locally attractive disconnections can lead to dead ends, excessively long routes, unavailable precursors, or strategically poor intermediates. Learning from simulated retrosynthetic experience demonstrated that search policies

can be improved by treating planning as sequential decision-making rather than independent one-step classification [17]. Experience-guided Monte Carlo tree search similarly uses information accumulated during planning to guide subsequent tree expansion, showing that the allocation of search effort materially influences which routes are discovered [18].

Natural-product synthesis adds a stronger long-horizon requirement. Early disconnections determine later stereochemical commitments, protecting-group burdens, fragment complexity, and opportunities for convergence. A search algorithm can therefore be efficient yet strategically myopic if its objective rewards rapid connection to known starting materials without representing these downstream consequences. Convergent retrosynthesis planning provides direct evidence that route-level objectives can be changed to favour shared intermediates and more convergent route structures rather than treating each target or branch independently [19]. However, algorithmic convergence should not be equated automatically with the expert concept of convergent total synthesis, where fragment complexity, branch independence, late-stage coupling risk, and material throughput interact.

We therefore propose that long-horizon natural-product planning requires two coupled evaluations: search progress, which asks whether the algorithm is moving toward reachable precursors, and strategic progress, which asks whether those moves create a coherent, controllable route. The first can be optimized by search policy; the second requires route-level chemical objectives. **Table 2** converts that distinction into validation requirements, while **Figure 1** depicts the proposed interface between computational search signals and strategic synthesis demands.

Table 2. Validation readiness of long-horizon planning claims for complex-molecule retrosynthesis

Planning claim	Evidence presently available	What it validates	Additional test required for complex natural products
Learned search improves route exploration	Simulated-experience planning	Search policy can outperform unguided decision rules in tested environments	Fixed-budget comparison on stereochemically dense targets with identical reaction models
Experience can guide tree expansion	Experience-guided MCTS	Prior search information can improve multistep exploration	Repeated runs reporting route robustness, dead ends, and chemically distinct alternatives
Convergence can be optimized explicitly	Convergent planning	Route objectives can favour shared or convergent structures	Expert assessment of fragment balance, late-stage coupling risk, and material throughput
Strategic adequacy follows from successful search	Proposed synthesis	Not established by existing evidence	Prospective comparison against expert-designed complex-molecule routes and laboratory outcomes

Note: Search success, computational convergence, and expert strategic convergence are treated as separate validation levels.

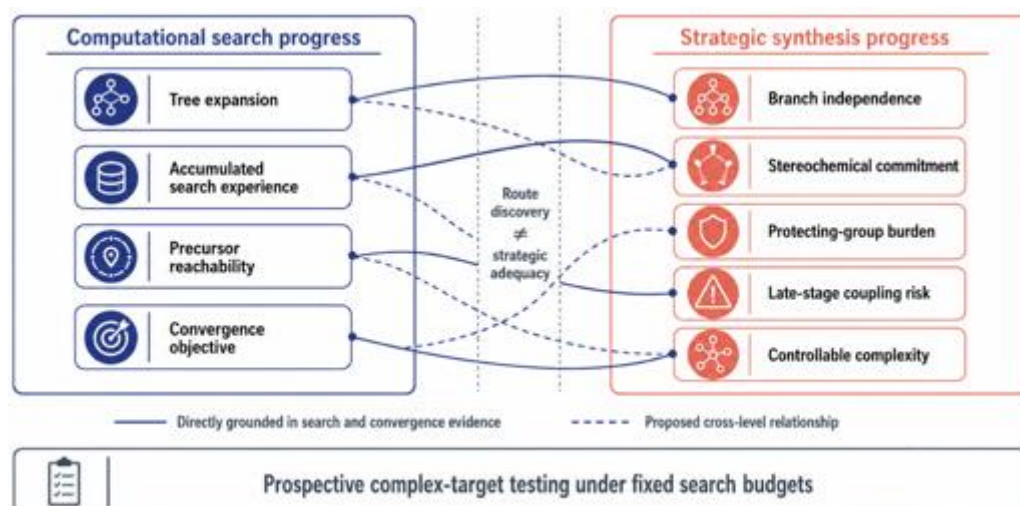


Figure 1. Search progress and strategic progress as interacting fields in long-horizon retrosynthesis

Building blocks, biosynthetic logic, and semisynthesis

The terminal nodes of a retrosynthetic search are themselves assumptions. A route may appear short because a planner treats an advanced fragment as commercially accessible, ignores the effort required to prepare a specialized building block, or fails to consider that a natural precursor could provide complexity more efficiently than de novo construction. BioNavi-NP demonstrates that biosynthetic reaction knowledge can be learned and navigated computationally for natural-product pathway reconstruction [20]. This expands retrosynthetic thinking beyond conventional synthetic reaction corpora without establishing that every inferred biosynthetic step is experimentally accessible.

Hybrid planning extends the idea. BioNavi-based methods combining biosynthetic and chemical transformations allow search across reaction domains that are usually modeled separately [21]. Automated organic–enzymatic planning further shows that enzymatic transformations can

be incorporated into computational route generation alongside conventional chemistry [22]. Such systems are especially relevant when stereoselective enzymatic steps, complex chiral-pool materials, or natural intermediates can replace a difficult sequence of protecting-group and stereocontrol operations.

The limitation is accessibility. A predicted enzyme transformation may depend on unavailable catalysts, undocumented substrate tolerance, incompatible concentrations, or untested scale. Likewise, semisynthesis is valuable only when the advanced natural precursor can be obtained reproducibly, ethically, and at useful scale. The proposed natural-product planning rule is therefore to treat building blocks as qualified resources, not terminal labels. Their provenance, preparation burden, stereochemical state, supply, and transformation compatibility should remain visible to the route scorer. **Table 3** summarizes how information quality at this interface can change the meaning of an apparently successful plan.

Table 3. Data-quality failures at the building-block and hybrid-planning boundary

Information deficit	Planning consequence	Evidence basis	Required qualification
Incomplete biosynthetic-reaction coverage	Plausible natural transformations may be missed or pathway confidence overstated	BioNavi-NP	Pathway prediction does not establish biological availability
Chemical and biosynthetic databases remain separated	Routes may overlook productive cross-domain transitions	Hybrid BioNavi	Expanded reaction space remains database dependent
Enzyme availability or substrate scope is poorly represented	A formally attractive enzymatic step may be operationally inaccessible	Organic–enzymatic planning	Predicted enzyme compatibility requires experimental confirmation
Advanced precursor treated as freely available	Route length and difficulty may be artificially compressed	Proposed synthesis	Starting-material provenance and preparation burden should be scored explicitly
Natural precursor assumed to make semisynthesis preferable	Strategic shortcut may disappear under supply or scale constraints	Proposed synthesis	Accessibility must be evaluated for the intended synthesis context

Route scoring for yield, feasibility, and practicality

Route ranking is often described as if “best” had a single computational meaning, but different scoring functions

answer different questions. Synthetic cost-aware decision frameworks show that material or route cost can be incorporated explicitly into molecular-design and planning

decisions rather than being inferred from route existence [23]. In a defined pharmaceutical-process case, computer-aided route optimization was linked to experimental implementation and greener process considerations, demonstrating that route comparison can incorporate practical objectives beyond nominal retrosynthetic success [24]. By contrast, RAScore provides a rapid learned estimate of retrosynthetic accessibility derived from planner-generated labels [25]. It is therefore a useful accessibility proxy, not a direct measurement of yield, robustness, safety, or laboratory executability.

For complex natural products, predicted reaction probability, search score, accessibility, expected yield, material cost, step count, protecting-group burden, and

operational risk should remain separate until an explicit decision rule combines them. Collapsing them prematurely into one scalar obscures why one route outranks another and can allow a high-confidence sequence of individually familiar reactions to dominate despite weak global practicality.

Figure 2 therefore proposes a nested interpretation of route quality. Model-defined reachability occupies the innermost level; route-level resource and efficiency objectives provide a second level; and chemically conditioned practicality forms an outer level. Laboratory testing sits outside all three because execution can invalidate assumptions at any inner layer.

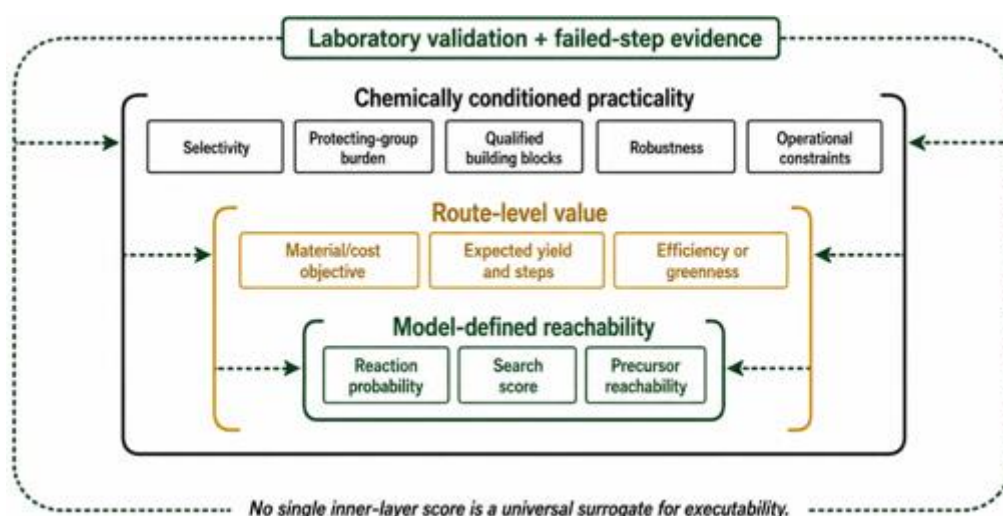


Figure 2. Nested evidence levels for interpreting retrosynthetic route quality

Benchmarking against expert natural product synthesis

The strongest test of synthesis planning is not agreement with a hidden reaction in a patent dataset but survival of a route under chemical execution. Computer-generated routes have been taken into the laboratory for selected medically relevant targets, establishing that computational planning can cross the boundary from virtual proposal to experimentally realized synthesis [26]. That achievement is important, but target selection and human intervention mean it should not be interpreted as evidence that arbitrary natural products can be solved autonomously.

Expert natural-product synthesis also supplies a richer benchmark than route completion. Contemporary reflection on total synthesis emphasizes that the discipline can pursue strategic simplicity, conceptual insight, reactivity discovery, structural understanding, and access to broader chemical space in addition to obtaining the target itself [27]. A machine route that reaches a

purchasable precursor in fewer graph-search steps may therefore be inferior to an expert route that deliberately creates complexity, exploits symmetry or biosynthetic logic, minimizes fragile operations, or opens access to analogues.

The syntheses of paspaline A and emindole PB illustrate a complementary relationship: computational analysis augmented a transform-guided expert retrosynthetic strategy rather than replacing strategic chemical judgment [28]. This is a particularly relevant benchmark model for complex natural products because it measures whether computation resolves a difficult decision inside an expert synthesis process.

Table 4 therefore compares method classes by the questions they can legitimately answer. The comparison rejects a single leaderboard: one-step prediction, route search, laboratory execution, and expert-augmented synthesis test different levels of competence.

Table 4. Method–use-case comparison for benchmarking complex natural-product retrosynthesis

Evaluation mode	Best-supported use	Strongest inference	Critical limitation
One-step reaction benchmark	Diagnose precursor-prediction behavior	Model can recover or propose transformations within the tested distribution	Does not establish coherent multistep synthesis
Multistep computational search	Compare route discovery under controlled planner settings	Search/model combination can reach defined terminal precursors	Route discovery does not demonstrate execution
Computer plan followed by laboratory synthesis	Test whether selected recommendations survive experiment	Some machine-generated routes are experimentally realizable	Human intervention and target selection constrain generalization
Expert natural-product synthesis criteria	Assess strategic and scientific value beyond endpoint access	Route quality can include objectives omitted from standard CASP metrics	Criteria are multidimensional and not reducible to one benchmark score
Computationally augmented expert total synthesis	Test decision support on chemically difficult targets	Computation can strengthen specific expert strategic choices	Does not demonstrate autonomous machine planning
Prospective machine-versus-expert complex-target benchmark	Proposed use case	Could test route-level competence directly	Requires blinded assessment, fixed resources, and experimental follow-through

Critical synthesis: what machine intelligence still misses

The evidence does not support a binary conclusion that machine retrosynthesis either “works” or “fails.” The more consequential pattern is a mismatch between what current systems optimize and what complex natural-product synthesis demands. Computer-assisted synthesis-planning efficiency depends on model and search behavior, including computational effort and route-exploration characteristics [29]. A planner can therefore fail practically because it is too slow, branches excessively, or repeatedly explores similar chemistry even when its reaction model is individually competent.

Generalization is a second fault line. Reaction-prediction models challenged with chemistry that departs from training precedent can lose reliability, showing why random or historically similar benchmark splits may overstate readiness for genuinely novel transformations [30]. Natural products amplify this problem because unusual ring constructions, rearrangements, selective oxidations, and densely functionalized substrates may occupy precisely the regions where precedent is sparse.

Third, aggregate metrics suppress chemically heterogeneous errors. Direct analysis of one-step retrosynthesis failure modes shows that predictions can fail for distinguishable chemical reasons that are poorly summarized by a single accuracy value [31]. Stereochemical corruption, implausible bond changes, missing reagents or contextual dependencies, and alternative but potentially reasonable disconnections should not all be counted identically.

The principal synthesis proposed here is therefore a hierarchy of evidential competence. A model may demonstrate reaction prediction without route competence; route competence without strategic adequacy; strategic adequacy without practical executability; or executability for selected targets without transferable natural-product

competence. Progress at an inner level is valuable, but it does not logically validate the next. Machine intelligence currently appears strongest as an expanding decision-support system whose local predictions and search capabilities can increasingly assist synthesis planning. The evidence remains insufficient to treat it as a generally validated autonomous solver of stereochemically and strategically demanding natural-product synthesis.

Research priorities for complex-molecule retrosynthesis

Future progress requires changing both models and the experiments used to judge them. Integrated platforms such as ASKCOS demonstrate that reaction prediction, condition recommendation, tree search, scoring, and related planning components can coexist in modular synthesis-planning infrastructure [32]. Such modularity is advantageous for complex-molecule research because stereochemistry, protection logic, selectivity prediction, building-block qualification, and route scoring can be tested independently rather than hidden inside one opaque success metric.

Higher-level chemical reasoning is another emerging direction. Language-mediated systems can incorporate strategic instructions and chemical reasoning into synthesis-planning tasks, suggesting that constraints ordinarily expressed by chemists in natural language may become model inputs rather than external commentary [33]. This evidence is promising but early; verbalized strategy should not be confused with reliable mechanistic reasoning, and hallucination or chemically unsupported explanations require direct checking.

Route diversity also needs explicit treatment. Analysis in reaction space shows that successful planners can differ in robustness and diversity even when conventional success criteria appear similar [34]. For natural products, generating several genuinely distinct strategic hypotheses

may be more useful than returning many minor variants of the same transformation sequence.

A decisive benchmark should therefore be prospective and adversarial rather than retrospective and convenience sampled. We propose targets with complete stereochemical annotation, rare transformations, competing chemoselective sites, explicit protection decisions, and strategically meaningful branch points. Starting-material catalogs and search budgets should be frozen before prediction. Machine-generated and expert routes should undergo blinded assessment for chemical plausibility, stereochemical integrity, convergence, building-block burden, risk, and route diversity. Selected plans should then be executed experimentally, with failed steps retained as evidence rather than removed from the record. Such a benchmark would permit rejection as well as confirmation of claims and would distinguish improvement in reaction prediction from improvement in complex synthesis planning.

Limitations

This review is critical rather than exhaustive, and the available literature is heterogeneous in datasets, reaction domains, search algorithms, terminal-material catalogs, and evaluation criteria. Uncertainty-calibrated reaction prediction demonstrates that confidence information can be estimated computationally, but forward-model uncertainty is not equivalent to uncertainty for an entire synthetic route [35]. Historical-precedent dependence also remains structurally embedded in some planning approaches; molecular-similarity retrosynthesis is explicitly conditioned on reactions related to known precedent [36].

Finally, commonly used synthesizability and complexity scores encode operational definitions rather than direct laboratory truth. SCScore learns synthetic complexity from patterns in a reaction corpus and therefore captures a historical data-derived signal, not an empirical measurement of whether a particular natural-product route will be practical under specified conditions [37]. These constraints limit cross-study comparison and prevent strong claims about universal machine-versus-human performance. The proposed evidential hierarchy and validation architecture are consequently organizing constructs for testing, not validated standards.

CONCLUSION

Machine intelligence has made retrosynthesis substantially more capable, but structurally complex natural products reveal why reaction prediction and synthesis strategy cannot be treated as synonyms. Stereochemical bookkeeping must become stereochemical reasoning;

reaction availability must be conditioned by selectivity and molecular state; search progress must be distinguished from strategic progress; and route scores must not substitute automatically for experimental practicality. Template-free generation, protection-aware planning, learned search, hybrid biosynthetic–synthetic routes, and multidimensional scoring each solve meaningful parts of this problem, yet none establishes autonomous general competence.

The central contribution of this review is a proposed hierarchy separating reaction prediction, route competence, strategic adequacy, and practical executability. A convincing claim that machine intelligence can solve complex natural-product retrosynthesis will require prospective, stereochemically explicit, resource-controlled benchmarks linked to expert assessment and laboratory execution, including failed routes. Until then, the strongest evidence supports machine intelligence as increasingly powerful synthesis decision support rather than a universally validated replacement for expert synthetic judgment.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Wright BA, Sarpong R. Molecular complexity as a driving force for the advancement of organic synthesis. *Nat Rev Chem*. 2024;8(10):776-92. doi:10.1038/s41570-024-00645-8
- [2] Mikulak-Klucznik B, Gołębiowska P, Bayly AA, Popik O, Klucznik T, Szymkuć S, et al. Computational planning of the synthesis of complex natural products. *Nature*. 2020;588(7836):83-8. doi:10.1038/s41586-020-2855-y
- [3] Long L, Li R, Zhang J. Artificial intelligence in retrosynthesis prediction and its applications in medicinal chemistry. *J Med Chem*. 2025;68(3):2333-55. doi:10.1021/acs.jmedchem.4c02749
- [4] Roh J, Joung JF, Yu K, Tu Z, Bartholomew GL, Santiago-Reyes OA, et al. Higher-level strategies for computer-aided retrosynthesis. *ACS Cent Sci*. 2026;12(3):345-57. doi:10.1021/acscentsci.5c02014
- [5] Hastedt F, Bailey RM, Hellgardt K, Yaliraki SN, del Rio Chanona EA, Zhang D. Investigating the reliability and interpretability of machine learning

- frameworks for chemical retrosynthesis. *Digit Discov.* 2024;3(6):1194-212. doi:10.1039/D4DD00007B
- [6] Torren-Peraire P, Hassen AK, Genheden S, Verhoeven J, Clevert DA, Preuss M, et al. Models matter: The impact of single-step retrosynthesis on synthesis planning. *Digit Discov.* 2024;3(3):558-72. doi:10.1039/D3DD00252G
- [7] Tadanki AS, Rao HSP, Priyakumar UD. Dissecting errors in machine learning for retrosynthesis: A granular metric framework and a transformer-based model for more informative predictions. *Digit Discov.* 2025;4(3):831-45. doi:10.1039/D4DD00263F
- [8] Orsi M, Reymond JL. Assigning the stereochemistry of natural products by machine learning. *J Cheminform.* 2026;18(1):76. doi:10.1186/s13321-026-01205-6
- [9] Pesciullesi G, Schwaller P, Laino T, Reymond JL. Transfer learning enables the molecular transformer to predict regio- and stereoselective reactions on carbohydrates. *Nat Commun.* 2020;11(1):4874. doi:10.1038/s41467-020-18671-7
- [10] Coley CW, Green WH, Jensen KF. RDChiral: An RDKit wrapper for handling stereochemistry in retrosynthetic template extraction and application. *J Chem Inf Model.* 2019;59(6):2529-37. doi:10.1021/acs.jcim.9b00286
- [11] Deng Y, Zhao X, Sun H, Chen Y, Wang X, Xue X, et al. RSGPT: A generative transformer model for retrosynthesis planning pre-trained on ten billion datapoints. *Nat Commun.* 2025;16(1):7012. doi:10.1038/s41467-025-62308-6
- [12] Han Y, Xu X, Hsieh CY, Ding K, Xu H, Xu R, et al. Retrosynthesis prediction with an iterative string editing model. *Nat Commun.* 2024;15(1):6404. doi:10.1038/s41467-024-50617-1
- [13] Zhong W, Yang Z, Chen CYC. Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing. *Nat Commun.* 2023;14(1):3009. doi:10.1038/s41467-023-38851-5
- [14] Westerlund AM, Sigmund LM, Mijangos MV, Kannas C, Genheden S, Kabeshov M. Toward lab-ready AI synthesis plans with protection strategies and route scoring. *J Chem Inf Model.* 2026;66(11):6361-75. doi:10.1021/acs.jcim.6c01147
- [15] Gallegos LC, Luchini G, St John PC, Kim S, Paton RS. Importance of engineered and learned molecular representations in predicting organic reactivity, selectivity, and chemical properties. *Acc Chem Res.* 2021;54(4):827-36. doi:10.1021/acs.accounts.0c00745
- [16] Zhang Z, Qiu J, Zheng J, Yu Z, Su L, Lin Q, et al. Enhanced regioselectivity prediction of sp² C-H halogenation via negative data augmentation and multimodel integration. *J Chem Inf Model.* 2025;65(7):3420-30. doi:10.1021/acs.jcim.5c00281
- [17] Schreck JS, Coley CW, Bishop KJM. Learning retrosynthetic planning through simulated experience. *ACS Cent Sci.* 2019;5(6):970-81. doi:10.1021/acscentsci.9b00055
- [18] Hong S, Zhuo HH, Jin K, Shao G, Zhou Z. Retrosynthetic planning with experience-guided Monte Carlo tree search. *Commun Chem.* 2023;6(1):120. doi:10.1038/s42004-023-00911-8
- [19] Torren-Peraire P, Verhoeven J, Herman D, Ceulemans H, Tetko IV, Wegner JK. Improving route development using convergent retrosynthesis planning. *J Cheminform.* 2025;17(1):26. doi:10.1186/s13321-025-00953-1
- [20] Zheng S, Zeng T, Li C, Chen B, Coley CW, Yang Y, et al. Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP. *Nat Commun.* 2022;13(1):3342. doi:10.1038/s41467-022-30970-9
- [21] Zeng T, Jin Z, Zheng S, Yu T, Wu R. Developing BioNavi for hybrid retrosynthesis planning. *JACS Au.* 2024;4(7):2492-502. doi:10.1021/jacsau.4c00228
- [22] Wang X, Yin X, Zhang X, Zhao H, Gu S, Wu Z, et al. A virtual platform for automated hybrid organic-enzymatic synthesis planning. *Nat Commun.* 2025;16(1):10929. doi:10.1038/s41467-025-65898-3
- [23] Fromer JC, Coley CW. An algorithmic framework for synthetic cost-aware decision making in molecular design. *Nat Comput Sci.* 2024;4(6):440-50. doi:10.1038/s43588-024-00639-y
- [24] Teixeira RI, Andresini M, Luisi R, Benyahia B. Computer-aided retrosynthesis for greener and optimal total synthesis of a helicase-primase inhibitor active pharmaceutical ingredient. *JACS Au.* 2024;4(11):4263-72. doi:10.1021/jacsau.4c00624
- [25] Thakkar A, Chadimová V, Bjerrum EJ, Engkvist O, Reymond JL. Retrosynthetic accessibility score (RAscore)—rapid machine learned synthesizability classification from AI driven retrosynthetic planning. *Chem Sci.* 2021;12(9):3339-49. doi:10.1039/D0SC05401A
- [26] Klucznik T, Mikulak-Klucznik B, McCormack MP, Lima H, Szymkuć S, Bhowmick M, et al. Efficient syntheses of diverse, medically relevant targets planned by computer and executed in the laboratory. *Chem.* 2018;4(3):522-32. doi:10.1016/j.chempr.2018.02.002
- [27] Shenvi RA. Natural product synthesis in the 21st century: Beyond the mountain top. *ACS Cent Sci.* 2024;10(3):519-28. doi:10.1021/acscentsci.3c01518

- [28] Kim DE, Zweig JE, Newhouse TR. Total synthesis of paspaline A and emindole PB enabled by computational augmentation of a transform-guided retrosynthetic strategy. *J Am Chem Soc.* 2019;141(4):1479-83. doi:10.1021/jacs.8b13127
- [29] Hartog PBR, Westerlund AM, Tetko IV, Genheden S. Investigations into the efficiency of computer-aided synthesis planning. *J Chem Inf Model.* 2025;65(4):1771-81. doi:10.1021/acs.jcim.4c01821
- [30] Bradshaw J, Zhang A, Mahjour B, Graff DE, Segler MHS, Coley CW. Challenging reaction prediction models to generalize to novel chemistry. *ACS Cent Sci.* 2025;11(4):539-49. doi:10.1021/acscentsci.5c00055
- [31] Tran SBA, Roh J, Coley CW. Quantifying the failure modes of current one-step retrosynthesis models. *Chem Sci.* 2026;17(28):14054-63. doi:10.1039/D6SC01323F
- [32] Tu Z, Choure SJ, Fong MH, Roh J, Levin I, Yu K, et al. ASKCOS: Open-source, data-driven synthesis planning. *Acc Chem Res.* 2025;58(11):1764-75. doi:10.1021/acs.accounts.5c00155
- [33] Bran AM, Neukomm TA, Armstrong DP, Jončev Z, Schwaller P. Chemical reasoning in LLMs unlocks strategy-aware synthesis planning and reaction mechanism elucidation. *Matter.* 2026;9(5):102812. doi:10.1016/j.matt.2026.102812
- [34] Hassen AK, Lai H, Genheden S, Preuss M, Clevert DA. Synthesis planning in reaction space: A study on success, robustness and diversity. *Digit Discov.* 2026;5(4):1623-34. doi:10.1039/D5DD00280J
- [35] Schwaller P, Laino T, Gaudin T, Bolgar P, Hunter CA, Bekas C, et al. Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Cent Sci.* 2019;5(9):1572-83. doi:10.1021/acscentsci.9b00576
- [36] Coley CW, Rogers L, Green WH, Jensen KF. Computer-assisted retrosynthesis based on molecular similarity. *ACS Cent Sci.* 2017;3(12):1237-45. doi:10.1021/acscentsci.7b00355
- [37] Coley CW, Rogers L, Green WH, Jensen KF. SCScore: Synthetic complexity learned from a reaction corpus. *J Chem Inf Model.* 2018;58(2):252-61. doi:10.1021/acs.jcim.7b00622