



Sparse Labels, Rich Chemistry: Building Few-Shot Learning Strategies for Bioactivity Prediction Across Structurally Diverse and Data-Limited Natural Product Chemical Space

James Anderson¹, Maria Rossi^{2*}, William Clark³

¹Department of Cheminformatics and Few-Shot Learning, Faculty of Pharmaceutical Sciences, University of Manchester, Manchester, United Kingdom.

²Department of Natural Product Bioactivity Prediction, Faculty of Pharmacy, University of Milan, Milan, Italy.

³Department of Computational Chemistry and Chemical Space Exploration, Faculty of Pharmacy, University of Melbourne, Melbourne, Australia.

ABSTRACT

Natural products occupy chemically diverse, provenance-rich regions of molecular space, yet experimental bioactivity labels are often sparse, unevenly distributed, and assembled from heterogeneous assay sources. Conventional small-data modeling treats this primarily as a sample-size problem. This article develops an original methodological framework in which few-shot natural-product bioactivity prediction is instead treated as a joint problem of task definition, chemical-space coverage, representation choice, support-set design, transferability, leakage-controlled evaluation, and uncertainty management. The framework distinguishes scarcity of labels from reliability of labels, rarity of scaffolds from pharmacological novelty, and predictive similarity from mechanistic equivalence. It further argues that molecular representation should be selected empirically rather than hierarchically: fingerprints and descriptors remain necessary low-data controls, while self-supervised and foundation-scale representations are candidates whose value depends on source–target alignment and target-like validation. Support examples are conceptualized as part of the learning design rather than passive training observations, and subsequent adaptation must be conditional on task relatedness. The principal contribution is a proposed strategy for deciding when few-shot learning is scientifically informative in heterogeneous natural-product space and when abstention, additional measurement, or non-transfer baselines are preferable. The framework is not a validated performance standard and does not establish prospective predictive superiority, mechanistic correctness, or translational readiness. Its value lies in converting low-label prediction from an architecture-selection problem into an evidence-bounded design problem whose assumptions can be explicitly tested.

Key Words: Few-shot learning, Natural products, Bioactivity prediction, Chemical space, Molecular representation, Transfer learning

eIJPPR 2025; 15(1):77-85

HOW TO CITE THIS ARTICLE: Anderson J, Rossi M, Clark W. Sparse Labels, Rich Chemistry: Building Few-Shot Learning Strategies for Bioactivity Prediction Across Structurally Diverse and Data-Limited Natural Product Chemical Space. Int J Pharm Phytopharmacol Res. 2025;15(1):77-85. <https://doi.org/10.51847/G4sMpyl6Tz>

INTRODUCTION

Natural-product discovery is increasingly intertwined with machine learning across identification, dereplication, prioritization, target-oriented analysis, and broader drug-discovery workflows. Yet the value of artificial

intelligence in this domain depends on the quality, coverage, and interpretability of the chemical and biological information on which predictions are conditioned [1]. Natural products are therefore not simply another small-molecule benchmark class. Their structural

Corresponding author: Maria Rossi

Address: Department of Natural Product Bioactivity Prediction, Faculty of Pharmacy, University of Milan, Milan, Italy.

E-mail: ✉ maria.rossi@unimi.it

Received: 11 November 2024; **Revised:** 26 January 2025; **Accepted:** 28 January 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.



diversity, source-dependent occurrence, stereochemical richness, and uneven experimental annotation create a setting in which the effective learning problem can vary sharply from one chemical neighborhood or assay task to another.

Contemporary natural-product resources reinforce this point. COCONUT 2.0 aggregates and curates a large, heterogeneous collection of reported natural-product structures, illustrating both the breadth of accessible chemical space and the importance of structure normalization and provenance-aware curation [2]. Such repositories expand the space that can be searched computationally, but structural availability does not imply corresponding bioactivity coverage. A chemically extensive collection may therefore contain dense labels around familiar chemotypes while leaving rare scaffolds, source-restricted compounds, or newly characterized structures with only a few relevant experimental observations.

Few-shot learning is attractive precisely because it reframes this asymmetry. Rather than requiring a separately trained high-capacity model for every sparsely annotated endpoint, few-shot methods attempt to use a small support set to condition predictions for a new task. Recent ligand-based work demonstrates that activity prediction for previously unseen assay tasks can be formulated in this way [3]. The methodological opportunity, however, should not be confused with proof of natural-product generalization: an unseen assay may still contain chemistry closely related to meta-training compounds.

Evidence from ultra-low-data molecular prediction further shows why the problem cannot be reduced to “use a smaller model.” Model choice, transfer strategy, and source–target relation can materially change performance, and transfer may become negative when the imported information is poorly matched to the target task [4]. The central problem addressed here is therefore not how to maximize a generic few-shot score, but how to construct a chemically and pharmacologically defensible prediction strategy when labels are scarce and chemical heterogeneity is high.

Why sparse labels are not just small data

Label count describes only one dimension of information scarcity. A support set containing ten internally consistent measurements is methodologically different from a support set containing ten nominally similar values generated under heterogeneous experimental conditions. Direct analysis of IC₅₀ and K_i measurements from different sources has shown that combining such values can introduce substantial noise, making label comparability a distinct issue from label quantity [5]. In few-shot settings,

this distinction is amplified because one discordant observation can influence a much larger fraction of the effective training signal.

Task semantics add another layer. ChEMBL spans multiple bioactivity types, target annotations, assay formats, and historical data sources, even though these records are systematically curated [6]. A “few-shot bioactivity task” must therefore specify what the label represents: binding, inhibition, functional response, cytotoxicity, or another endpoint; whether values are continuous or thresholded; how assay confidence is treated; and whether measurements from different protocols are considered exchangeable. Pooling records merely because they share a target identifier risks creating apparent sample size while weakening biological coherence.

Scarcity can also coexist with severe class imbalance. MolFeSCue was developed specifically for molecular property prediction in data-limited and imbalanced conditions, underscoring that the number and composition of positive and negative examples jointly shape low-shot learning [7]. The framework proposed here therefore separates four questions: how many labels exist, how comparable they are, how they are distributed across classes or response ranges, and how much relevant chemical variation they cover. Few-shot design begins only after those questions have been answered. A small dataset is not necessarily a few-shot task, and a large nominal dataset can remain informationally sparse after assay harmonization, chemical partitioning, and removal of weakly comparable records.

Chemical diversity, scaffold rarity, and task definition

Natural-product diversity is meaningful only relative to a specified chemical and biological reference space. LOTUS links natural-product structures to organism-level provenance across a large body of referenced literature, demonstrating that chemical identity and biological occurrence are frequently intertwined [8]. Provenance can therefore help characterize where a compound sits within reported natural-product space. It should not, however, be converted into a surrogate activity label: taxonomic relatedness may correlate with biosynthetic opportunity without implying pharmacological equivalence.

The Natural Products Atlas makes the dependence on reference collection especially visible. As a curated resource focused on microbially derived natural products, it supports structural novelty assessment within a defined source domain [9]. A scaffold that appears rare in such a resource is rare relative to that resource, not necessarily rare across all natural products. The same qualification applies to plant-derived, marine, fungal, or synthetic NP-like collections. For few-shot prediction, “rare scaffold”

should therefore mean rare under a named representation, similarity rule, and reference collection rather than functioning as an unqualified descriptor.

Task definition must be equally explicit. MoleculeNet established the broader principle that molecular-learning performance depends on endpoint, metric, dataset structure, and partitioning strategy [10]. In natural-product few-shot work, the prediction task should consequently be specified along at least three dimensions: pharmacological meaning of the label, chemical novelty of the query relative to the support and training pools, and evaluation regime under which generalization is being claimed.

This distinction prevents three common collapses. First, scaffold novelty is not equivalent to target novelty. Second, chemical dissimilarity does not demonstrate mechanistic dissimilarity. Third, a successful classifier on a randomly partitioned dataset does not establish performance for genuinely rare natural-product chemotypes. The proposed framework therefore treats task construction as the first substantive modeling decision. Before choosing a representation or adaptation algorithm, the investigator must define what is new: the assay, the target, the scaffold, the source domain, the label regime, or some combination of these.

Molecular representation under label scarcity

Representation becomes especially consequential when labels are scarce because the model has little supervised evidence with which to correct an inappropriate encoding of chemical similarity. Comparative work on molecular property prediction has shown that learned representations do not uniformly dominate conventional descriptors or fingerprints across tasks and splitting regimes [11]. Fixed representations therefore remain necessary controls in low-label natural-product prediction, particularly when interpretability, computational economy, or known substructural similarity is important.

Self-supervised pretraining offers a different route. MolCLR demonstrates that large collections of unlabeled molecular graphs can be used for contrastive pretraining, producing embeddings that transfer to multiple downstream molecular-property tasks [12]. For natural products, the attraction is clear: unlabeled structures are much more abundant than well-harmonized activity measurements. Yet pretraining scale does not eliminate domain mismatch. If the pretraining corpus sparsely represents stereochemically complex, macrocyclic, highly oxygenated, glycosylated, or otherwise atypical natural-product chemotypes, a learned embedding may still compress distinctions that matter in the target task.

Foundation-scale graph models broaden this option further. MolE provides a recent example of transferable molecular graph representations learned at scale [13]. Such models can be included in a few-shot strategy, but they should enter as candidates rather than presumptive winners. The relevant question is whether a representation preserves distinctions that are useful for the defined bioactivity task under the target chemical shift.

The methodological implication proposed here is a representation ladder rather than a representation hierarchy. Simple fingerprints or descriptors establish a chemically intelligible floor; supervised graph models test whether learned local structure improves that floor; self-supervised or foundation representations test whether broader pretraining adds useful transfer. Advancement up the ladder should depend on target-like validation, not architectural novelty. Under extreme label scarcity, the most defensible representation is the one whose inductive assumptions remain compatible with the task after scaffold-aware evaluation.

The context-dependent logic derived from these sections is summarized in **Table 1**.

Table 1. Context-dependent representation choices for sparse-label natural-product bioactivity prediction

Prediction context	Representation choice to test	Scientific reason for inclusion	Main assumption requiring validation	Evidence status
Very small, chemically interpretable support set	Fingerprints or curated physicochemical descriptors	Provides a transparent low-data reference and preserves known substructural similarity	Chosen features capture task-relevant chemistry	Evidence-supported baseline; learned models do not uniformly dominate fixed features [11]
Large unlabeled molecular corpus but few endpoint labels	Self-supervised graph representation	Uses structural information without requiring additional activity labels	Pretraining chemistry overlaps meaningfully with target space	Evidence-supported candidate; contrastive pretraining can transfer to downstream molecular tasks [12]
Broad pretraining resources and	Foundation-scale molecular graph embedding	Tests whether large-scale transferable	Pretraining objective and chemical	Evidence-supported candidate, not presumed optimum [13]

heterogeneous downstream tasks		representations improve adaptation	coverage remain relevant to target task	
Rare or structurally shifted natural-product query set	Parallel comparison of fixed and pretrained representations	Detects representation failure under scaffold or domain shift	Similarity encoded by either representation remains pharmacologically informative	Proposed framework requirement derived from the manuscript's chemical-space and representation argument
Unclear representation advantage	Retain simplest competitive representation	Avoids adding capacity without demonstrated target-task benefit	Validation split reflects intended chemical generalization	Proposed decision rule; no universal representation ranking is asserted

Note: Table 1 is a manuscript-derived context-by-method crosswalk. Evidence-supported statements are linked to the cited comparative or pretraining studies; the decision rules are proposed methodological synthesis and require prospective validation.

Support-set construction and episodic learning

Few-shot learning changes the role of labeled examples. In conventional supervised modeling, a datum primarily contributes to parameter estimation. In an episodic formulation, the support set also defines the local task that the model must infer. Molecular one-shot learning established that graph-based representations can be paired with support/query episodes so that a model adapts to previously unseen low-data tasks [14]. For natural products, this makes support selection a scientific design decision: a support molecule is informative not simply because it carries a label, but because its chemistry and assay context help define the target relation.

Comparative few-shot drug-discovery work shows that performance varies with task construction and low-data setup [15]. Randomly drawing a few labeled molecules may therefore be unstable when the available pool contains duplicated analogue series, narrow chemotype clusters, or isolated scaffolds. The proposed strategy records support

size, class composition, within-support diversity, and the chemical distance from support compounds to the query set. Repeated episodes should be used where feasible to determine whether an apparent result depends on a fortunate support sample.

Prototype learning provides another route to using the support set. Attribute-guided prototype networks demonstrate that learned embeddings can be combined with molecular attributes when constructing few-shot class prototypes [16]. This supports testing chemistry-aware support summaries against purely latent similarity, while avoiding the stronger claim that the selected attributes identify biological mechanism. Across one-shot, comparative few-shot, and prototype studies, episode construction and task-conditioned similarity are therefore central design variables [14-16].

Table 2 converts these observations into failure checks that can be applied before interpreting an episode-level result.

80

Table 2. Support-set failure modes that can distort sparse-label natural-product learning

Failure mode	Diagnostic manifestation	Methodological response	Evidential status
Analogue-dominated support	Most support molecules occupy one close chemotype neighborhood	Report within-support diversity and repeat episodes after analogue control	Proposed mitigation grounded in support sensitivity [15]
Class-sparse support	One activity class or response region is weakly represented	Report support composition; use balanced episodes only as a training device, not prevalence estimate	Proposed mitigation; imbalance is a documented low-data concern [7]
Chemically remote query	Query structures lie outside support neighborhoods	Report support-query distance and permit abstention or additional measurement	Proposed framework rule
Prototype shortcut	Prototype reflects convenient attributes rather than target-relevant chemistry	Compare latent-only and chemistry-aware prototypes and inspect sensitivity	Attribute-guided prototypes are evidence-supported [16]
Single-episode dependence	Model ranking changes with a different small support sample	Repeat support sampling and summarize variability rather than one favorable episode	Proposed robustness requirement

Note: Failure modes describe methodological risks, not measured universal frequencies. Proposed mitigations require task-specific validation.

Transfer, metric learning, and adaptation strategies

A few-shot model is useful only if prior information is transferable. Deep transferable compound representations

have shown that molecular information learned across domains or tasks can improve prediction in low-data drug-discovery settings [17]. This establishes transfer as a plausible strategy, but not as a default. Pretraining may import useful chemical regularities, assay-related patterns, or task structure; it can also import irrelevant biases when the source and target problems differ.

Direct few-shot structure–activity experiments sharpen that boundary. Meta-learning benefits vary with source–target task similarity, with stronger gains reported when target tasks resemble the tasks available during meta-training [18]. The framework therefore treats transferability as a gate: before adapting a pretrained or meta-learned model, the investigator should document why the source tasks are relevant to the target endpoint and chemistry. Similarity may involve target family, activity semantics, molecular neighborhoods, or combinations of these; no universal numerical cutoff is implied.

Metric learning and adaptation should then be selected according to what information can plausibly transfer. A learned distance may be appropriate when support examples define meaningful local relationships; parameter adaptation may be preferable when task-specific adjustment is required; a no-transfer model remains necessary when the source task is weakly related. The proposed comparison is consequently triangular rather than hierarchical: transferred model, low-data task-specific baseline, and simple representation baseline. A transfer gain that disappears under a more distant scaffold split should be interpreted as conditional evidence, not as failure of few-shot learning in general. Conversely, negative transfer is informative because it identifies a boundary at which collecting additional target-specific measurements may be more defensible than importing more source-domain supervision.

Evaluation without scaffold or analogue leakage

Evaluation must distinguish memorizing a chemical neighborhood from adapting to a genuinely sparse target task. Similar compounds, repeated entities, or related molecular pairs can cross conventional partitions and make a held-out set easier than its label suggests. DataSAIL formalizes similarity- and entity-aware splitting to reduce this type of information leakage [19]. In natural-product few-shot studies, analogous controls should operate both between historical/meta-training data and the target task and between the target support and final query or test sets. Leakage control is necessary but insufficient. Molecular models can change ranking or lose performance when evaluated on out-of-distribution chemistry [20]. A scaffold-disjoint or similarity-reduced test therefore answers a different question from a random split: it probes whether the learned relation persists as chemistry shifts. The relevant shift should be described rather than hidden behind the term “OOD,” because separation by scaffold, molecular similarity, source collection, target family, or time creates different extrapolation problems.

The proposed evaluation design uses nested separation. First, meta-training information is isolated from the target task according to the claimed novelty. Second, support compounds are made available only within the target episode. Third, the final query/test set is characterized by its relationship to both support and historical training chemistry. Performance should then be interpreted together with the chemical-distance regime. Naive partitions can otherwise overstate chemically transferable performance when related entities or analogues cross partitions or when distribution shift is ignored [19, 20].

Table 3 makes explicit what different evaluation representations assume before a generalization claim is attached to them.

Table 3. Evaluation representations and the assumptions behind few-shot generalization claims

Evaluation representation	What it primarily tests	Implicit assumption	Claim that remains unsafe
Random or weakly constrained split	Interpolation among sampled molecules	Train and test chemistry are exchangeable	Generalization to rare scaffolds
Scaffold-aware separation	Transfer across defined core structures	Scaffold partition captures the relevant chemical shift	Target or assay novelty
Similarity/entity-aware separation	Reduced analogue or entity leakage	Chosen similarity representation identifies meaningful relatedness	Complete absence of information leakage [19]
Explicit OOD chemical shift	Robustness beyond the training distribution	Constructed shift resembles the intended future use	Prospective discovery performance [20]
Episode-aware support/query separation	Adaptation from designated support labels	Query labels are unavailable during adaptation	Generalization beyond the represented support neighborhood

Note: The comparison is methodological. No partition is treated as a universal validity test.

These evaluation decisions form the outer methodological gates around a few-shot claim, as synthesized in **Figure 1**.

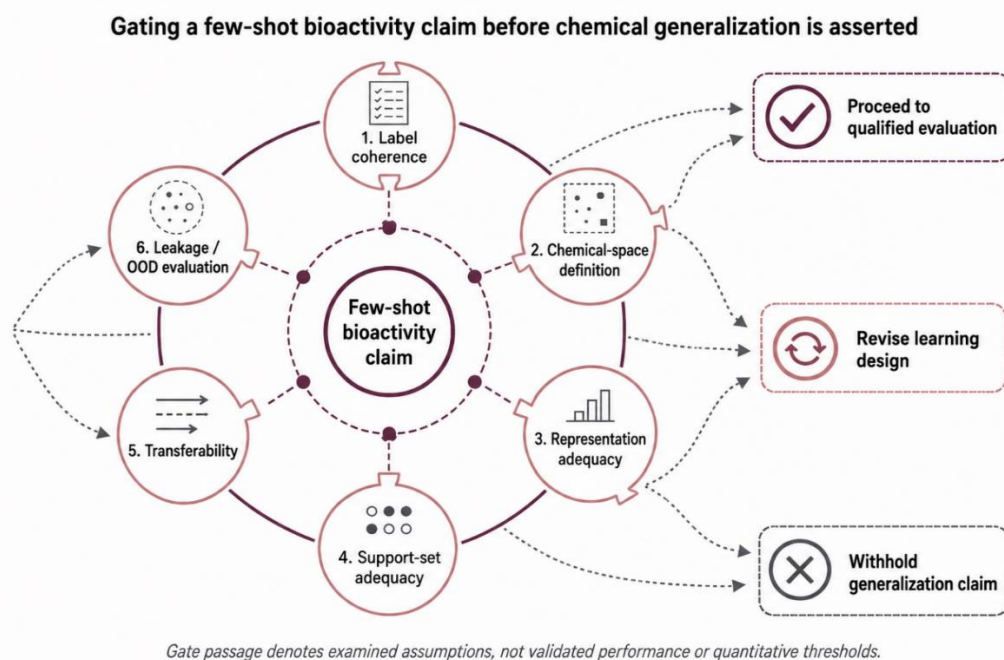


Figure 1. Gating a few-shot bioactivity claim before chemical generalization is asserted

The few-shot bioactivity strategy framework

The evidence above suggests that no single architecture defines a defensible few-shot strategy. The framework proposed here instead organizes decisions around the claim that is intended. Its first stage fixes task semantics and the target novelty: endpoint, assay context, chemical reference space, and whether novelty concerns scaffold, target, source domain, or some combination. The second stage compares representations against simple baselines and characterizes whether the support set covers the chemistry on which predictions will be made.

Adaptation is then selected conditionally. Evidential meta-learning demonstrates that few-shot molecular prediction can be coupled to uncertainty estimation and imbalance handling [21]. In-context bioactivity prediction with MHNfs shows that support examples may instead condition predictions without full task-specific retraining [22]. Property-aware relation networks provide a third option in which learned molecular relationships are conditioned on the prediction property and selective adaptation [23]. These methods establish different adaptation families; they do not establish a universal winner or show that learned relational features are biological mechanisms.

The proposed framework therefore asks three sequential questions after task and representation definition. First, is transfer justified by relevant source information? Task similarity must be treated as evidence for this decision rather than assumed [18]. Second, does the chosen evaluation prevent the intended novelty from leaking across partitions? Similarity-aware separation provides

one way to test that condition [19]. Third, is the resulting prediction sufficiently supported to be acted on, or should the system abstain, request measurement, or revert to a simpler baseline? Uncertainty-aware prediction can inform that final decision, but its reliability must itself be tested [21].

Failure modes, uncertainty, and applicability boundaries

Failure can arise even when average predictive performance appears favorable. Explainable uncertainty methods for molecular property prediction distinguish sources such as aleatoric and epistemic uncertainty and can localize contributions to model uncertainty [24]. For few-shot natural-product prediction, this distinction matters because an uncertain query may reflect noisy labels, insufficient support coverage, representation mismatch, or extrapolation into chemistry not represented during learning. These possibilities call for different responses.

Uncertainty estimates are not self-validating. Comparative work across neural molecular prediction tasks shows that uncertainty methods differ in calibration and error-ranking behavior and that no single approach dominates every setting [25]. A high-confidence prediction should therefore not be interpreted as evidence of target engagement, mechanism, or biological correctness. The uncertainty method should be evaluated on the intended task, preferably under the same scaffold or distribution shifts used for performance testing.

The framework defines an applicability boundary as the region in which the combination of task semantics, support coverage, representation, transfer relation, and uncertainty

behavior has been examined sufficiently to justify a qualified prediction. It is not a fixed similarity threshold. Outside that region, the appropriate output may be abstention, wider predictive intervals, additional experimental labels, or a statement that the compound is outside the evaluated regime. Uncertainty estimates should be calibrated and evaluated rather than treated as intrinsic confidence [24, 25].

Hard cases remain especially important: activity cliffs can make close analogues pharmacologically divergent, scaffold hopping can produce similar activity without obvious structural proximity, and heterogeneous assay biology can make chemically sensible neighborhoods biologically inconsistent. These cases limit any interpretation of molecular distance as pharmacological equivalence.

Translational implications for natural product prioritization

The immediate translational role of the proposed framework is prioritization, not replacement of experiment. Machine-learning analysis of natural-product LC-MS/MS metabolomics has shown that computational models can be integrated with experimental bioactivity workflows to classify and prioritize natural-product samples [26]. That evidence demonstrates practical coupling between prediction and experimental natural-product discovery, although spectral or extract-level features should not be conflated with purified structure-only few-shot models.

A structure-based few-shot system could therefore be used to rank compounds for confirmatory measurement when a target or assay has only a small coherent label set. The output should include the task definition, support chemistry, evaluation regime, and uncertainty or abstention status alongside the predicted activity. Compounds could be prioritized not solely by estimated potency but by expected informational value: for example, a chemically distant candidate with high uncertainty may be valuable as the next experiment because it tests the boundary of the current model.

This role fits the broader natural-product discovery process, in which screening, isolation, structural characterization, mechanism determination, supply, and optimization remain distinct challenges [27]. A predicted bioactivity score cannot substitute for those stages. Binding prediction does not establish functional response; an in vitro label does not establish exposure; and prioritization does not establish therapeutic effectiveness. Natural-product-derived candidates nevertheless remain relevant to downstream drug development. A recent retrospective analysis reported favorable observed clinical-development outcomes for natural-product-related

candidates relative to comparator classes [28]. That association supplies context for why efficient natural-product prioritization remains worthwhile, but it does not show that natural-product provenance causes those outcomes and cannot validate the present framework. The appropriate translational claim is narrower: when experimental resources are limited, a chemically explicit few-shot strategy may help decide what to test next, while preserving a clear boundary between computational plausibility and experimental evidence.

Limitations

The framework inherits the limitations of the data from which low-shot tasks are constructed. Denoising studies in drug-discovery property prediction show that data-quality and curation choices can materially alter model behavior [29]. Removing questionable observations may improve internal consistency, but it can also discard difficult yet real chemistry; curation should therefore be documented and treated as a sensitivity decision rather than a route to assumed ground truth.

Representation choice is another unresolved boundary. Comparative analysis of pretrained molecular models has shown that pretrained embeddings are not uniformly superior to canonical fingerprints across downstream tasks [30]. The proposed representation ladder accordingly avoids assuming that model scale or pretraining establishes relevance to natural-product chemistry.

More broadly, the framework has not been validated prospectively as an integrated pipeline. Its gates and decision rules organize existing methodological evidence but do not establish optimal support size, chemical-distance thresholds, transfer criteria, uncertainty cutoffs, or universal natural-product applicability. Performance may differ by assay class, target family, natural-product source, stereochemical complexity, and available historical tasks. External and prospective tests capable of rejecting individual framework decisions are therefore required.

CONCLUSION

Sparse-label natural-product bioactivity prediction is best treated as a design problem rather than an architecture contest. The proposed framework links label coherence, explicit chemical-space definition, representation benchmarking, support-set construction, conditional transfer, leakage-reduced evaluation, and uncertainty-aware prioritization into a single testable strategy. Its central principle is that a prediction becomes scientifically stronger when the assumptions supporting its generalization claim are made visible and challenged.

This approach preserves several boundaries that are especially important in natural-product discovery.

Structural similarity is not pharmacological equivalence; assay labels are not mechanisms; pretrained representations are not intrinsically transferable; uncertainty is not proof of correctness; and computational priority is not therapeutic readiness. The framework therefore permits abstention and additional measurement as legitimate outcomes rather than treating every sparse-label task as solvable by model complexity. Future value depends on prospective, externally separated natural-product tasks that can test, refine, or reject these proposed decisions. Until then, the framework should be understood as an evidence-bounded methodology for asking when few-shot prediction is informative, what it is allowed to claim, and where experimental evidence must take over.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

- [1] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov.* 2023;22(11):895-916. doi:10.1038/s41573-023-00774-7
- [2] Chandrasekhar V, Rajan K, Kanakam SRS, Sharma N, Weißenborn V, Schaub J, et al. COCONUT 2.0: A comprehensive overhaul and curation of the collection of open natural products database. *Nucleic Acids Res.* 2025;53(D1). doi:10.1093/nar/gkae1063
- [3] Eckmann P, Anderson J, Yu R, Gilson MK. Ligand-based compound activity prediction via few-shot learning. *J Chem Inf Model.* 2024;64(14):5492-9. doi:10.1021/acs.jcim.4c00485
- [4] Eraqi BA, Khizbullin D, Nagaraja SS, Sarathy SM. Molecular property prediction in the ultra-low data regime. *Commun Chem.* 2025;8:201. doi:10.1038/s42004-025-01592-1
- [5] Landrum GA, Riniker S. Combining IC50 or Ki values from different sources is a source of significant noise. *J Chem Inf Model.* 2024;64(5):1560-7. doi:10.1021/acs.jcim.4c00049
- [6] Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* 2024;52(D1). doi:10.1093/nar/gkad1004
- [7] Zhang R, Wu C, Yang Q, Liu C, Wang Y, Li K, et al. MolFeSCue: Enhancing molecular property prediction in data-limited and imbalanced contexts using few-shot and contrastive learning. *Bioinformatics.* 2024;40(4). doi:10.1093/bioinformatics/btae118
- [8] Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, et al. The LOTUS initiative for open knowledge management in natural products research. *Elife.* 2022;11. doi:10.7554/eLife.70780
- [9] van Santen JA, Jacob G, Singh AL, Aniebok V, Balunas MJ, Bunsko D, et al. The Natural Products Atlas: An open access knowledge base for microbial natural products discovery. *ACS Cent Sci.* 2019;5(11):1824-33. doi:10.1021/acscentsci.9b00806
- [10] Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: A benchmark for molecular machine learning. *Chem Sci.* 2018;9(2):513-30. doi:10.1039/C7SC02664A
- [11] Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. *J Chem Inf Model.* 2019;59(8):3370-88. doi:10.1021/acs.jcim.9b00237
- [12] Wang Y, Wang J, Cao Z, Barati Farimani A. Molecular contrastive learning of representations via graph neural networks. *Nat Mach Intell.* 2022;4(3):279-87. doi:10.1038/s42256-022-00447-x
- [13] Méndez-Lucio O, Nicolaou CA, Earnshaw B. MolE: A foundation model for molecular graphs using disentangled attention. *Nat Commun.* 2024;15(1):9431. doi:10.1038/s41467-024-53751-y
- [14] Altae-Tran H, Ramsundar B, Pappu AS, Pande V. Low data drug discovery with one-shot learning. *ACS Cent Sci.* 2017;3(4):283-93. doi:10.1021/acscentsci.6b00367
- [15] Vella D, Ebejer JP. Few-shot learning for low-data drug discovery. *J Chem Inf Model.* 2023;63(1):27-42. doi:10.1021/acs.jcim.2c00779
- [16] Hou L, Xiang H, Zeng X, Cao D, Zeng L, Song B. Attribute-guided prototype network for few-shot molecular property prediction. *Brief Bioinform.* 2024;25(5). doi:10.1093/bib/bbae394
- [17] Abbasi K, Poso A, Ghasemi J, Amanlou M, Masoudi-Nejad A. Deep transferable compound representation across domains and tasks for low data drug discovery. *J Chem Inf Model.* 2019;59(11):4528-39. doi:10.1021/acs.jcim.9b00626
- [18] Kötter A, Allenspach S, Grebner C, Matter H, Hiss JA, Schneider G, et al. Task-similarity is a crucial factor for few-shot meta-learning of structure-activity relationships. *ChemBioChem.* 2024;25(19). doi:10.1002/cbic.202400095

- [19] Joeres R, Blumenthal DB, Kalinina OV. Data splitting to avoid information leakage with DataSAIL. *Nat Commun.* 2025;16(1):3337. doi:10.1038/s41467-025-58606-8
- [20] Fooladi H, Vu TNL, Mathea M, Kirchmair J. Evaluating machine learning models for molecular property prediction: Performance and robustness on out-of-distribution data. *J Chem Inf Model.* 2025;65(19):9871-91. doi:10.1021/acs.jcim.5c00475
- [21] Ham KP, Sael L. Evidential meta-model for molecular property prediction. *Bioinformatics.* 2023;39(10). doi:10.1093/bioinformatics/btad604
- [22] Schimunek J, Luukkonen S, Klambauer G. MHNfs: Prompting in-context bioactivity predictions for low-data drug discovery. *J Chem Inf Model.* 2025;65(9):4243-50. doi:10.1021/acs.jcim.4c02373
- [23] Yao Q, Shen Z, Wang Y, Dou D. Property-aware relation networks for few-shot molecular property prediction. *IEEE Trans Pattern Anal Mach Intell.* 2024;46(8):5413-29. doi:10.1109/TPAMI.2024.3368090
- [24] Yang CI, Li YP. Explainable uncertainty quantifications for deep learning-based molecular property prediction. *J Cheminform.* 2023;15(1):13. doi:10.1186/s13321-023-00682-3
- [25] Hirschfeld L, Swanson K, Yang K, Barzilay R, Coley CW. Uncertainty quantification using neural networks for molecular property prediction. *J Chem Inf Model.* 2020;60(8):3770-80. doi:10.1021/acs.jcim.0c00502
- [26] Brittin NJ, Anderson JM, Braun DR, Rajsiki SR, Currie CR, Bugni TS. Machine learning-based bioactivity classification of natural products using LC-MS/MS metabolomics. *J Nat Prod.* 2025;88(2):361-72. doi:10.1021/acs.jnatprod.4c01123
- [27] Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. Natural products in drug discovery: Advances and opportunities. *Nat Rev Drug Discov.* 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z
- [28] Domingo-Fernández D, Gadiya Y, Preto AJ, Krettlér CA, Mubeen S, Allen A, et al. Natural products have increased rates of clinical trial success throughout the drug development process. *J Nat Prod.* 2024;87(7):1844-51. doi:10.1021/acs.jnatprod.4c00581
- [29] Adrian M, Chung Y, Cheng AC. Denoising drug discovery data for improved absorption, distribution, metabolism, excretion, and toxicity property prediction. *J Chem Inf Model.* 2024;64(16):6324-37. doi:10.1021/acs.jcim.4c00639
- [30] Zhang Z, Bian Y, Xie A, Han P, Zhou S. Can pretrained models really learn better molecular representations for AI-aided drug discovery? *J Chem Inf Model.* 2024;64(7):2921-30. doi:10.1021/acs.jcim.3c01707