Repurposing Natural Products with Artificial Intelligence: A Framework-Oriented Review of Molecular Similarity, Target Networks, and Disease-Repositioning Logic

Mark Anderson[1]*, Lisa Wong[2], Sarah Lee[1]

[1]Department of Artificial Intelligence for Polypharmacology and Natural Products, College of Pharmacy, University of Florida, Gainesville, United States.
[2]Department of Computational Systems Pharmacology, Faculty of Pharmacy, National University of Singapore, Singapore.

ABSTRACT

Natural products remain important sources of chemically diverse and biologically active molecules, yet repositioning them toward new therapeutic contexts requires a careful distinction between computational hypothesis generation and pharmacological validation. This framework-oriented review examines how artificial intelligence and computational pharmacology can support natural-product repurposing through three connected reasoning domains: molecular similarity, target-network inference, and disease-repositioning logic. The review integrates methodological and application-oriented evidence to organize how natural-product identity can be transformed into computable molecular representations, compared within chemical space, linked to putative targets, mapped onto disease-associated networks, and integrated with omics, pharmacological, exposure, and validation evidence. The proposed framework emphasizes that molecular fingerprints, descriptors, molecular graphs, SMILES-based encodings, and learned embeddings can support analogue retrieval and chemical-neighbourhood analysis, but their outputs depend strongly on representation choice and applicability domain. Target prediction, target fishing, drug–target interaction modeling, protein-network mapping, disease-module analysis, knowledge graphs, and transcriptomic reversal can extend beyond structural resemblance, but each introduces assumptions about data provenance, target context, biological directionality, and causal relevance. Candidate prioritization is therefore framed as an evidence-calibration process rather than a proof of therapeutic utility. The review highlights the need to distinguish similarity, prediction, prioritization, validation, and translation; to identify where evidence layers converge or conflict; and to require experimental, exposure-aware, and translational assessment before repositioning claims are considered credible. Future progress depends on better natural-product data standardization, uncertainty-aware models, context-specific networks, leakage-resistant benchmarks, reproducible workflows, and prospective validation.

Key Words: *Natural products, Artificial intelligence, Drug repurposing, Molecular similarity, Chemical space, Target networks*

eIJPPR 2025; 14(2):49-59

HOW TO CITE THIS ARTICLE: Anderson M, Wong L, Lee S. Repurposing Natural Products with Artificial Intelligence: A Framework-Oriented Review of Molecular Similarity, Target Networks, and Disease-Repositioning Logic. Int J Pharm Phytopharmacol Res. 2025;15(2):49-59. https://doi.org/10.51847/MjhDTTRBdk

INTRODUCTION

Natural products occupy a distinctive position in therapeutic discovery because they combine structural diversity, stereochemical richness, biosynthetic novelty, and extensive histories of biological observation. Drug repurposing offers an attractive logic for such molecules because existing pharmacological, toxicological,

Corresponding author: Mark Anderson
Address: Department of Artificial Intelligence for Polypharmacology and Natural Products, College of Pharmacy, University of Florida, Gainesville, United States.
E-mail: ✉ mark.anderson@ufl.edu
Received: 17 December 2024; **Revised:** 12 April 2025; **Accepted:** 16 April 2025

traditional-use, target, pathway, or analogue evidence may already exist, yet this evidence must be interpreted as support for hypothesis generation rather than as confirmation of a new therapeutic indication [1-3].

The rise of AI-enabled drug repurposing has expanded the range of computational strategies available for connecting compounds to targets, pathways, phenotypes, and diseases. Deep learning methods can integrate molecular structures, drug–target associations, disease labels, omics profiles, biomedical graphs, and literature-derived relationships, but their conclusions depend on database composition, molecular representation, endpoint definition, negative sampling, and validation design [4].

Machine learning can therefore support natural-product repurposing most responsibly when the computational question is explicitly bounded. A model that predicts chemical similarity, target overlap, network proximity, or disease association is not answering the same question as a biochemical assay, cellular perturbation study, pharmacokinetic experiment, or clinical evaluation, and this distinction is essential when AI outputs are used to prioritize natural products for further investigation [5].

This review develops a framework-oriented synthesis for AI-enabled natural-product repurposing. Its central aim is to explain how molecular similarity, chemical-space positioning, target-network inference, disease-repositioning logic, evidence integration, candidate prioritization, and staged validation can be connected without treating these evidence layers as interchangeable or independently sufficient. The framework is intended to organize reasoning, identify uncertainty, and clarify when computational convergence strengthens a hypothesis and when it may instead create false confidence through shared biases or non-independent data sources.

Natural product repurposing in drug discovery

In the natural-product context, repurposing may refer to the evaluation of a known phytochemical, bioactive metabolite, natural-product-derived compound, or structurally related analogue for a disease, target, pathway, formulation, or combination context different from its original evidence base. Such repositioning is attractive because natural products may already have documented bioactivity, historical exposure, partial safety characterization, structural analogues, or mechanistic clues, but these attributes do not remove the need for indication-specific pharmacological testing [6].

Natural products also introduce challenges that differ from those of many synthetic small-molecule libraries. Their structural complexity, stereochemical density, scaffold diversity, metabolic transformation, polypharmacology, incomplete target annotation, variable source quality, and bioavailability constraints can all complicate computational interpretation. Natural-product derivatives may broaden repositioning opportunities by improving tractability or modulating activity, but derivative similarity does not automatically preserve the same pharmacological profile or disease relevance [7].

AI-enabled natural-product research can help organize chemical, biological, and literature evidence across large and heterogeneous datasets. Molecular descriptors, fingerprints, graph representations, learned embeddings, target-prediction models, and chemical-space maps can identify plausible analogues or candidate mechanisms, yet the natural-product domain is especially sensitive to sparse annotations, uneven assay coverage, source variability, stereochemical uncertainty, and bias toward well-studied compound classes [8].

A credible repurposing claim must therefore distinguish historical or traditional use from pharmacological evidence, pharmacological evidence from computational support, computational support from validated target engagement, and validated target engagement from disease-specific therapeutic relevance. A natural product may be suitable for screening because it is chemically close to a reference drug, suitable for prioritization because molecular, target, and disease-context evidence converge, or suitable for translational assessment only after exposure, safety, formulation, and disease-model evidence are evaluated together.

Table 1 summarises the distinctive opportunities and constraints that shape the repurposing of natural products across chemical, pharmacological, biological, and translational evidence domains.

Table 1. Natural Product Repurposing in Drug Discovery: Sources of Repositioning Opportunity, Chemical and Biological Evidence, Computational Entry Points, Validation Requirements, Translational Constraints, and Risks of Overinterpretation

Repurposing dimension	Natural-product characteristic	Potential computational evidence	Repurposing opportunity	Primary uncertainty	Validation requirement	Translational constraint	Risk of overinterpretation
Chemical diversity	Diverse scaffolds, stereochemistry,	Fingerprints, descriptors, scaffold analysis,	Analogue discovery and candidate expansion	Representation may miss stereochemical or	Orthogonal representation checks and	Synthetic access, purity, source consistency	Treating chemical novelty as therapeutic relevance

	and substituent patterns	chemical-space mapping		conformational features	chemical curation		
Known bioactivity	Prior assay or pharmacological evidence may exist	Bioactivity databases, literature mining, target annotations	Reuse of existing biological clues	Assay context may not match new disease context	Repeated assays under disease-relevant conditions	Dose and exposure may differ from prior studies	Treating any bioactivity as repositioning evidence
Structural analogues	Related metabolites or derivatives may be available	Nearest-neighbour analysis, scaffold relationships, substructure matching	Candidate analogue identification	Activity cliffs and scaffold hopping	Comparative bioactivity and target-engagement testing	Analogue may have different ADMET profile	Assuming analogue similarity means shared efficacy
Polypharmacology	Multiple targets or pathways may be affected	Target prediction, target fishing, DTI modeling, network mapping	Multi-target repositioning hypotheses	Predicted target overlap may reflect promiscuity or database bias	Biochemical and cellular target confirmation	Off-target toxicity and pathway complexity	Treating target multiplicity as therapeutic advantage
Disease-pathway relevance	NP targets may intersect disease-associated pathways	Pathway enrichment, disease-gene mapping, network proximity	Disease-context hypothesis generation	Pathway overlap may be indirect or non-causal	Perturbation studies in relevant biological models	Disease heterogeneity and tissue specificity	Treating pathway overlap as mechanism
Omics response	Some compounds may have transcriptomic or phenotypic signatures	Signature reversal, phenotype similarity, multi-omics integration	Disease-state reversal hypotheses	Signature direction may be context-dependent	Independent transcriptomic and phenotypic validation	Cell-line or model mismatch	Treating signature reversal as treatment proof
Partial safety evidence	Historical exposure or early safety data may exist	Toxicity prediction, ADMET modeling, literature evidence	Early deprioritization of unsafe candidates	Historical exposure may not match therapeutic dose	Exposure-aware toxicity and metabolism studies	Bioavailability, formulation, and dose feasibility	Treating traditional exposure as clinical safety
Formulation potential	Delivery systems or derivatives may alter exposure	ADMET modeling, formulation feasibility assessment	Repositioning through improved delivery or exposure	In silico exposure predictions may be unreliable	PK/PD and formulation testing	Poor solubility, instability, metabolism	Treating formulation possibility as translational readiness

Review scope and framework orientation

This framework-oriented review integrates evidence from AI-enabled drug repurposing, natural-product discovery, molecular similarity, chemical-space analysis, target prediction, network medicine, disease-context modeling, knowledge graphs, omics-based repositioning, candidate prioritization, and validation logic. Computational repositioning literature supports the need to compare multiple evidence domains because drug–disease hypotheses may emerge from chemical resemblance, shared targets, disease modules, expression signatures, biomedical graphs, or combined data resources [9].

The review scope includes natural products, phytochemicals, natural-product-derived molecules, bioactive metabolites, and structurally related analogues when they are relevant to repurposing-oriented reasoning. The AI and computational methods considered include fingerprints, descriptors, SMILES-based models, molecular graphs, learned embeddings, target-prediction models, drug–target interaction models, network proximity, network propagation, disease-module mapping, knowledge-graph link prediction, transcriptomic reversal, multi-omics integration, and candidate-prioritization workflows. Computational repositioning outputs are interpreted as method-dependent signals whose meaning

changes with data source, representation, endpoint, and validation design [10].

The evidence was categorized by framework function rather than by disease area or model family. Literature identification focused on peer-reviewed journal evidence relevant to at least one stage of the framework, followed by duplicate handling, relevance assessment, full-text-oriented claim matching, evidence extraction, framework-domain classification, cross-study comparison, and critical appraisal of whether each source supported a bounded methodological claim. Network-based repositioning work that connects computational prediction with external validation illustrates why framework construction should include validation logic rather than stopping at in silico ranking [11].

This approach does not claim exhaustive retrieval, formal systematic-review completeness, pooled effect estimation, or a formal risk-of-bias score. Instead, it organizes heterogeneous methodological and application-oriented evidence into a transparent reasoning framework. The framework is intended to clarify which evidence layers are useful for candidate screening, which are more appropriate for prioritization, which require experimental confirmation, and which remain speculative until exposure, target engagement, disease relevance, and translational constraints are assessed.

Molecular similarity and chemical space mapping

Molecular similarity is a foundational but representation-dependent entry point for natural-product repurposing: theoretical similarity concepts, empirical fingerprint selection, and learned natural-product features all show that the same molecule can occupy different computational neighbourhoods depending on how it is encoded and compared [12-14]. This makes similarity useful for analogue retrieval, reference-drug comparison, scaffold analysis, and chemical-neighbourhood exploration, but unsuitable as a standalone claim of shared pharmacology.

Fingerprints can represent substructures or atom environments, descriptors can summarize physicochemical and topological properties, SMILES strings can support sequence-based learning, molecular graphs can encode atom–bond relationships, and learned embeddings can place compounds in latent spaces shaped by training objectives. In natural products, these choices are especially consequential because stereochemistry, tautomerism, protonation state, glycosylation, macrocyclic structure, conformational flexibility, and scaffold novelty may be simplified or unevenly captured by common representations.

Chemical-space mapping can help identify whether a natural product is close to approved drugs, known bioactive molecules, structurally related analogues, or underexplored chemical regions. Graph-based molecular representation studies indicate that learned graph features may improve some predictive tasks compared with fixed descriptors, but they remain dependent on training data, endpoint quality, chemical-domain coverage, and evaluation design [15]. Thus, graph-based proximity should be interpreted as a model-specific signal rather than a direct measure of mechanistic equivalence.

Learned molecular embeddings, including contrastive graph-based representations, can support similarity inference by organizing molecules according to patterns learned from large datasets. However, embedding proximity depends on pretraining objectives, augmentation strategy, dataset composition, and downstream task alignment, so a natural product that appears close to a reference compound in latent space may still differ in activity, metabolism, target engagement, or disease-context relevance [16].

Similarity-based repurposing should therefore use safeguards that test whether a candidate falls within a credible applicability domain, whether similar molecules share relevant bioactivity rather than superficial structure, whether multiple representations produce convergent neighbourhoods, and whether chemical-space proximity is consistent with target, pathway, disease, ADMET, and validation evidence. High molecular similarity may fail because of activity cliffs or exposure differences, while low structural similarity may still allow functional convergence through shared targets, phenotypic effects, or disease-module relationships.

Table 2 compares major molecular representation, similarity, and chemical-space strategies used to generate natural-product repurposing hypotheses.

Table 2. Molecular Similarity and Chemical-Space Strategies for Natural Product Repurposing: Representations, Similarity Logic, Analytical Outputs, Repurposing Functions, Strengths, Failure Modes, Validation Needs, and Interpretive Safeguards

Strategy	Molecular representation	Similarity or mapping logic	Typical analytical output	Repurposing function	Main strength	Major failure mode	Validation requirement	Interpretive safeguard
Fingerprint similarity	Binary or count-based	Shared substructures	Similarity-ranked neighbours	Analogue retrieval and	Efficient comparison	Activity cliffs and feature insensitivity	Bioactivity comparison against	Do not equate high fingerprint

	molecular fingerprints	or atom environments		reference-drug comparison	across large libraries		relevant targets	similarity with equivalent pharmacology
Physicochemical descriptor analysis	Continuous molecular properties and topological descriptors	Property-space proximity	Descriptor profiles and nearest neighbours	Developability-aware candidate filtering	Links chemical structure to broad property constraints	Descriptor scaling and loss of structural detail	Orthogonal property and ADMET assessment	Interpret as coarse chemical context, not mechanism
Scaffold analysis	Core ring systems and recurring chemotypes	Shared scaffold or scaffold class	Scaffold families and analogue groups	Candidate expansion through related chemotypes	Highlights medicinal-chemistry relationships	Scaffold sharing may not preserve activity or selectivity	Comparative target and phenotypic assays	Treat scaffold overlap as a hypothesis seed
Substructure matching	Defined fragments, pharmacophores, or motifs	Presence or absence of structural motifs	Matched motifs and structural alerts	Mechanism or toxicity hypothesis generation	Interpretable chemical rationale	Motif may be non-causal or context-dependent	Mechanistic testing and negative controls	Ask whether the motif is necessary and sufficient
SMILES-based modeling	Linear molecular string encodings	Sequence-derived learned similarity	Model embeddings or predicted properties	Rapid representation for AI pipelines	Compatible with language-model approaches	Sensitivity to enumeration and stereochemical encoding	Representation robustness checks	Use canonicalization and augmentation carefully
Molecular graph modeling	Atom–bond graph with node and edge features	Learned graph representations	Graph embeddings and predicted endpoints	Target, activity, or property prediction	Captures molecular connectivity explicitly	Training-data bias and limited extrapolation	External validation on relevant chemical domains	Check applicability domain and endpoint alignment
Learned embedding similarity	Latent vectors from neural models	Proximity in learned representation space	Embedding neighbours and clusters	Discovery of non-obvious analogues	Can capture complex learned patterns	Embedding proximity may reflect training bias	Task-specific validation and interpretability checks	Treat latent proximity as model-dependent evidence
Chemical-space visualization	Fingerprints, descriptors, or embeddings projected into low dimensions	Dimensionality reduction or clustering	Maps, clusters, neighbourhoods, outliers	Library comparison and out-of-domain detection	Reveals broad chemical relationships	Projection artifacts and misleading clusters	Multiple projection and representation comparisons	Avoid interpreting map distance as a measured biological value
Nearest-neighbour analogue search	Any curated representation linked to reference libraries	Local neighbourhood ranking	Candidate analogues and reference compounds	Candidate generation for repositioning	Transparent starting point for hypothesis generation	Library bias toward well-studied molecules	Independent target, disease, and ADMET evidence	Require convergence beyond local similarity
Applicability-domain assessment	Representation plus training-library boundaries	Domain coverage or outlier detection	In-domain or out-of-domain classification	Reliability filtering before prediction	Prevents unsupported extrapolation	Domain definition may be arbitrary	External testing on structurally relevant compounds	Flag out-of-domain predictions as low confidence

Target networks and disease repositioning

Target-network analysis moves natural-product repurposing beyond the assumption that structural resemblance alone is sufficient to infer therapeutic relevance. For natural products, known target retrieval, target fishing, drug–target interaction prediction, and cross-target profiling can help connect chemically complex molecules to putative biological mechanisms, especially where direct target annotation is incomplete or fragmented [17].

Transfer-learning approaches are particularly relevant to natural products because many phytochemicals and bioactive metabolites have sparse target labels compared with approved drugs or heavily screened synthetic

compounds. Such models may extend target prediction into underannotated chemical regions, but the resulting target lists remain predictions unless supported by binding, engagement, perturbation, or pathway-level evidence [18].

Disease-repositioning logic becomes more biologically informative when predicted or known targets are interpreted within protein–protein interaction networks, disease-associated genes, disease modules, and disease-aware computational models. Disease modules, network proximity, and clinician-centered repurposing models can connect target-level evidence to disease-context prioritization, but they should be interpreted as structured prioritization tools rather than as direct demonstrations of therapeutic efficacy [19-21].

Network-based reasoning can support natural-product repurposing in several ways. Candidate targets may be mapped onto disease-associated proteins, propagated through interaction networks, compared with disease modules, or evaluated for pathway relevance. However, these outputs are sensitive to popular-target bias, hub effects, incomplete protein-interaction maps, disease-gene annotation bias, and tissue or cell-type context. A target that appears close to a disease module in a generic network may not be active, accessible, or causally relevant in the disease state of interest.

Knowledge graphs and link-prediction systems add another disease-repositioning layer by combining compound, target, pathway, phenotype, disease, literature, and sometimes omics relationships. Their value lies in organizing heterogeneous evidence and suggesting plausible drug–disease connections, but an inferred edge is not equivalent to target engagement, mechanism confirmation, or clinical usefulness. Shared targets may also have opposing effects across diseases, and pathway overlap may reflect compensatory biology rather than therapeutic opportunity.

Evidence integration and candidate prioritization

No single evidence layer should dominate natural-product candidate selection. Molecular similarity can identify analogues, chemical-space proximity can locate reference neighbourhoods, target prediction can propose biological entry points, and network methods can map those targets to disease contexts, but each layer carries distinct assumptions. Gene-expression-based repurposing adds another dimension by comparing compound-induced and disease-associated signatures, although such evidence remains strongly dependent on biological context, cell type, dose, timing, and endpoint definition [22].

Transcriptomic reversal illustrates the difference between a useful prioritization signal and a validated therapeutic claim. A compound signature that opposes a disease signature may indicate a plausible direction for investigation, but it does not independently establish causal mechanism, target engagement, appropriate exposure, or disease-modifying efficacy [23].

Multi-omics integration can reduce reliance on any single biological layer by combining transcriptomic, proteomic, genomic, pathway, and disease-association evidence. At the same time, integration can increase uncertainty when datasets differ in provenance, biological context, preprocessing, disease labels, or measurement platforms. AI-guided multi-omics repositioning is therefore most useful when it makes uncertainty explicit and separates independent convergence from repeated evidence derived from the same underlying sources [24].

Latent-feature models, including autoencoder-based multi-omics approaches, can rank candidates by compressing heterogeneous biological information into lower-dimensional representations. These models can support candidate prioritization, but their latent structure requires biological interpretation, robustness checks, and external validation before the ranking can guide pharmacological claims [25].

Evidence harmonization is most credible when candidate prioritization combines chemical plausibility, target evidence, network context, disease relevance, literature support, known pharmacology, ADMET plausibility, exposure feasibility, and experimental testability. Multimodal harmonization can support rational candidate selection, but prioritization should remain a transparent decision process rather than a numerical shortcut that hides correlated evidence, database circularity, or shared-source bias [26].

Table 3 synthesizes the principal evidence layers used to prioritize natural-product repurposing candidates and distinguishes convergent support from correlated or potentially misleading computational signals.

Table 3. Evidence Integration and Candidate Prioritization for AI-Enabled Natural Product Repurposing: Molecular, Target, Network, Disease, Omics, Pharmacological, and Translational Evidence Layers

Evidence layer	Primary input	Computational output	Repositioning contribution	Evidence independence	Major uncertainty	Candidate-prioritization role	Validation requirement	Failure risk	Decision safeguard
Molecular similarity	Curated natural-product	Similarity neighbours, analogue	Generates candidate analogues	Often partly correlated with	Representation and metric dependence	Early screening	Bioactivity and target comparison	Treating similarity as	Require orthogonal chemical

	structures and reference compounds	groups, scaffold relationships	and reference-drug hypotheses	chemical-space mapping		and analogue expansion		equivalent activity	and biological evidence
Chemical-space position	Fingerprints, descriptors, embeddings, or graphs	Clusters, neighbourhoods, outlier status	Identifies chemical context and applicability domain	May overlap with molecular similarity evidence	Projection artifacts and library bias	Applicability-domain filtering	Multiple representation comparisons	Misleading visual proximity	Avoid interpreting map distance as biological proof
Target prediction	Structures, known bioactivity, target databases	Putative targets or DTI links	Proposes biological entry points	Partly independent if based on target data rather than only similarity	Sparse NP annotations and popular-target bias	Mechanism-oriented prioritization	Binding and target-engagement assays	Predicted target treated as validated target	Rank targets by evidence quality and testability
Target-network evidence	Candidate targets, PPI networks, pathway databases	Modules, proximity, propagation outputs	Connects target evidence to biological systems	Moderately independent if network source differs from target-prediction source	Hub bias, network incompleteness, tissue mismatch	Disease-context filtering	Disease-relevant perturbation studies	Network proximity treated as efficacy	Use tissue-aware networks and negative controls
Disease-module evidence	Disease genes, disease-associated proteins, pathways	Candidate–disease module relationships	Supports disease-context mapping	Depends on disease-gene curation source	Disease-label ambiguity and causal uncertainty	Disease selection and hypothesis refinement	Disease-model validation	Module overlap treated as repositioning suitability	Require directionality and disease-specific plausibility
Transcriptomic or omics evidence	Disease signatures, compound signatures, multi-omics profiles	Signature reversal, concordance, latent similarity	Adds functional biological context	Can be independent if generated from separate experiments	Cell-type, dose, timing, and platform mismatch	Prioritization and mechanism refinement	Independent omics and phenotypic validation	Signature reversal treated as therapeutic proof	Interpret with biological context and causal tests
Knowledge-graph evidence	Biomedical entity relations and literature-derived links	Predicted drug–target–disease associations	Organizes heterogeneous evidence and proposes links	May be low if graph reuses same databases as other layers	Edge provenance, relation ambiguity, circularity	Hypothesis generation and evidence discovery	External graph-free validation	Predicted link treated as confirmed relationship	Track provenance and avoid circular support
Pharmacological evidence	Prior bioactivity, mechanism studies, assay data	Known activity profile and plausible mechanisms	Grounds computational hypotheses in observed biology	More independent when derived experimentally	Assay relevance and reproducibility	Candidate strengthening or deprioritization	Replication in relevant models	Any bioactivity treated as indication evidence	Match assay context to disease mechanism
ADMET and exposure plausibility	Physicochemical properties, metabolism, toxicity, exposure data	Safety and developability flags	Filters candidates unlikely to reach relevant exposure safely	Often independent from target and disease evidence	Prediction uncertainty and formulation dependence	Translational feasibility assessment	PK, metabolism, toxicity, formulation studies	Ignoring poor exposure or toxicity	Apply exposure-aware reprioritization
Experimental feasibility	Candidate availability, assay	Testable validation plan	Determines whether	Independent from	Assay mismatch or	Final selection for validation	Biochemical, cellular, in vivo, and	Prioritizing untestable hypotheses	Require explicit validation

tractability, model relevance	hypothesis can progress	computational score	insufficient controls	orthogonal testing	path before advancement

Framework-oriented synthesis

The proposed framework begins with natural-product definition because computational inference is only as reliable as the compound identity being modeled. Ambiguous structures, mixtures, stereochemical uncertainty, inconsistent identifiers, or poorly curated derivatives can propagate error through representation, similarity analysis, target prediction, network mapping, and disease-context interpretation. A repositioning workflow should therefore begin by clarifying whether the object of analysis is a purified natural product, a derivative, a metabolite, a mixture component, or a structurally related analogue.

Once identity is established, the framework proceeds conditionally through molecular representation, similarity inference, chemical-space positioning, target inference, target-network integration, disease-context mapping, evidence convergence, candidate prioritization, experimental validation, and translational assessment. At each stage, the epistemic status of the candidate changes: structural resemblance provides a similarity signal, target prediction provides a predictive signal, network proximity provides disease-context support, multimodal agreement supports prioritization, experimental testing provides validation evidence, and exposure-aware assessment informs translational credibility. Knowledge-graph learning can help connect sparse biomedical relationships, but graph-derived hypotheses remain associative until supported by independent biological evidence [27].

Figure 1 presents the proposed framework linking molecular similarity, chemical-space positioning, target-network inference, disease-repositioning logic, multi-layer evidence integration, candidate prioritization, and staged validation for AI-enabled natural-product repurposing.

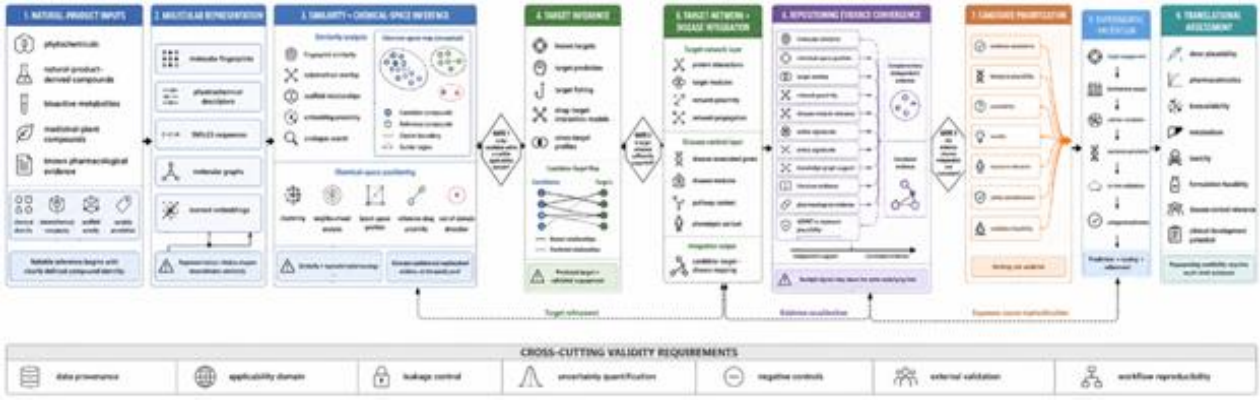

Figure 1. Integrated AI Framework for Natural Product Repurposing Through Molecular Similarity, Target Networks, and Disease-Repositioning Logic

Alt-text description

A wide left-to-right framework beginning with natural products and molecular representations. Candidates progress through molecular similarity analysis and chemical-space mapping, then into target prediction and target-network integration, followed by disease-gene and disease-module mapping. Multiple evidence streams converge in a candidate-prioritization layer incorporating chemical, target, network, omics, pharmacological, exposure, and uncertainty information. Prioritized candidates then progress through biochemical, cellular, in vivo, and translational validation. Feedback arrows return validation evidence to similarity, target, and disease-repositioning stages.

The framework also clarifies how conflicting evidence should be handled. A natural product may show high chemical similarity to a reference drug but weak disease-module relevance; another may show low chemical similarity but stronger target-network convergence; a third may have target overlap but poor exposure plausibility; a fourth may receive knowledge-graph support while lacking experimental evidence; and a fifth may reverse a disease-associated signature while leaving causal mechanism uncertain. Knowledge-graph computational frameworks are useful for organizing such relationships, but their outputs depend on graph construction, relation semantics, and evaluation design [28].

Evidence quantity should not be confused with evidence independence. Multiple positive signals may trace back to the same curated target database, the same disease-gene annotation, the same literature-mining pipeline, or the same biased benchmark split. Reviews of knowledge-

graph methods emphasize that database design, edge provenance, negative sampling, and evaluation protocols can strongly influence drug-repurposing outputs, which means that candidate prioritization should explicitly track whether evidence is genuinely independent or merely repeated across related resources [29].

The framework therefore functions as a staged decision architecture. Similarity and chemical-space evidence are best suited to candidate generation; target prediction and target networks are suited to mechanism-oriented prioritization; disease-context mapping supports repositioning hypothesis formation; omics and knowledge-graph evidence can strengthen or challenge plausibility; ADMET and exposure evidence determine whether a candidate is worth testing; and experimental validation is required before mechanistic or therapeutic claims are made.

Future directions

Future AI-enabled natural-product repurposing should move from similarity-driven candidate generation toward evidence-calibrated repositioning. This requires better natural-product data standardization, improved stereochemical representation, metabolite-aware modeling, source and purity annotation, versioned compound identifiers, applicability-domain assessment, and out-of-distribution detection. Critical analyses of AI in drug discovery caution that impact depends less on algorithmic novelty alone than on data quality, endpoint relevance, experimental feedback, and realistic claims about what models can and cannot establish [30].

Methodological development should also strengthen molecular representation learning for natural products. Self-supervised learning, contrastive learning, graph-based encoders, multimodal molecular representations, and foundation models may help extract richer patterns from sparse or heterogeneous data, but they require careful validation against chemically diverse and stereochemically complex compounds. Uncertainty quantification, interpretability, and external validation should become routine components of similarity, target-prediction, disease-mapping, and candidate-ranking workflows.

Network and disease-context modeling should become more biologically specific. Generic protein–protein interaction networks and broad disease labels are often insufficient for natural-product repositioning because target effects may depend on tissue, cell state, disease stage, dose, metabolism, and pathway directionality. Future frameworks should incorporate context-specific networks, tissue-aware disease modules, single-cell or spatial evidence where appropriate, dynamic disease networks, causal inference, and target-engagement data that can distinguish association from mechanism.

Responsible translation will also require leakage-resistant benchmarks, temporal validation, hard-negative selection, transparent negative sampling, knowledge-graph provenance tracking, fair candidate comparison, exposure-aware ranking, pharmacokinetic integration, metabolism-aware repurposing, open code, versioned datasets, and reproducible workflows. Recent assessments of AI in drug discovery emphasize that robust progress depends on ground truth, uncertainty handling, explainability, workflow reproducibility, and human expert oversight rather than on autonomous prediction claims [31].

CONCLUSION

This framework-oriented review organized AI-enabled natural-product repurposing around molecular similarity, chemical-space mapping, target networks, disease-repositioning logic, evidence integration, and candidate prioritization. The central contribution is a staged reasoning framework that distinguishes candidate generation from prioritization, prediction from validation, and computational convergence from translational credibility. AI can support analogue discovery, molecular representation, target inference, disease-context mapping, evidence synthesis, and candidate selection, but credible repurposing depends on representation quality, applicability-domain awareness, target-evidence strength, disease-context specificity, independent evidence convergence, uncertainty reporting, exposure plausibility, experimental validation, and translational assessment.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

1. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: Progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41-58. doi:10.1038/nrd.2018.168
2. 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

3. 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
4. Pan X, Lin X, Cao D, Zeng X, Yu PS, He L, et al. Deep learning for drug repurposing: Methods, databases, and applications. Wiley Interdiscip Rev Comput Mol Sci. 2022;12(4):e1597. doi:10.1002/wcms.1597
5. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5
6. Rastelli G, Pellati F, Pinzi L, Gamberini MC. Repositioning natural products in drug discovery. Molecules. 2020;25(5):1154. doi:10.3390/molecules25051154
7. Huang B, Zhang Y. Teaching an old dog new tricks: Drug discovery by repositioning natural products and their derivatives. Drug Discov Today. 2022;27(7):1936-44. doi:10.1016/j.drudis.2022.02.007
8. Saldívar-González FI, Aldas-Bulos VD, Medina-Franco JL, Plisson F. Natural product drug discovery in the artificial intelligence era. Chem Sci. 2022;13(6):1526-46. doi:10.1039/D1SC04471K
9. Jarada TN, Rokne JG, Alhajj R. A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. J Cheminform. 2020;12(1):46. doi:10.1186/s13321-020-00450-7
10. Ko Y. Computational drug repositioning: Current progress and challenges. Appl Sci (Basel). 2020;10(15):5076. doi:10.3390/app10155076
11. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun. 2018;9(1):2691. doi:10.1038/s41467-018-05116-5
12. López-Pérez K, Avellaneda-Tamayo JF, Chen L, López-López E, Juárez-Mercado KE, Medina-Franco JL, et al. Molecular similarity: Theory, applications, and perspectives. Artif Intell Chem. 2024;2(2):100077. doi:10.1016/j.aichem.2024.100077
13. Boldini D, Ballabio D, Consonni V, Todeschini R, Grisoni F, Sieber SA. Effectiveness of molecular fingerprints for exploring the chemical space of natural products. J Cheminform. 2024;16(1):35. doi:10.1186/s13321-024-00830-3
14. Menke J, Massa J, Koch O. Natural product scores and fingerprints extracted from artificial neural networks. Comput Struct Biotechnol J. 2021;19:4593-602. doi:10.1016/j.csbj.2021.07.032
15. Jiang D, Wu Z, Hsieh CY, Chen G, Liao B, Wang Z, et al. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. J Cheminform. 2021;13(1):12. doi:10.1186/s13321-020-00479-8
16. 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
17. Moumbock AFA, Li J, Mishra P, Gao M, Günther S. Current computational methods for predicting protein interactions of natural products. Comput Struct Biotechnol J. 2019;17:1367-76. doi:10.1016/j.csbj.2019.08.008
18. Qiang B, Lai J, Jin H, Zhang L, Liu Z. Target prediction model for natural products using transfer learning. Int J Mol Sci. 2021;22(9):4632. doi:10.3390/ijms22094632
19. Sadegh S, Skelton J, Anastasi E, Bernett J, Blumenthal DB, Galindez G, et al. Network medicine for disease module identification and drug repurposing with the NeDRex platform. Nat Commun. 2021;12(1):6848. doi:10.1038/s41467-021-27138-2
20. Morselli Gysi D, do Valle Í, Zitnik M, Ameli A, Gan X, Varol O, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proc Natl Acad Sci U S A. 2021;118(19):e2025581118. doi:10.1073/pnas.2025581118
21. Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. Nat Med. 2024;30(12):3601-13. doi:10.1038/s41591-024-03233-x
22. He H, Duo H, Hao Y, Zhang X, Zhou X, Zeng Y, et al. Computational drug repurposing by exploiting large-scale gene expression data: Strategy, methods and applications. Comput Biol Med. 2023;155:106671. doi:10.1016/j.compbiomed.2023.106671
23. Koudijs KKM, Böhringer S, Guchelaar HJ. Validation of transcriptome signature reversion for drug repurposing in oncology. Brief Bioinform. 2023;24(1):bbac490. doi:10.1093/bib/bbac490
24. Cong Y, Endo T. Multi-omics and artificial intelligence-guided drug repositioning: Prospects, challenges, and lessons learned from COVID-19.

OMICS. 2022;26(7):361-71. doi:10.1089/omi.2022.0068

25. Mohammadzadeh-Vardin T, Ghareyazi A, Gharizadeh A, Abbasi K, Rabiee HR. DeepDRA: Drug repurposing using multi-omics data integration with autoencoders. PLoS One. 2024;19(7):e0307649. doi:10.1371/journal.pone.0307649
26. Schultz B, Zaliani A, Ebeling C, Reinshagen J, Bojkova D, Lage-Rupprecht V, et al. A method for the rational selection of drug repurposing candidates from multimodal knowledge harmonization. Sci Rep. 2021;11(1):11049. doi:10.1038/s41598-021-90296-2
27. Bang D, Lim S, Lee S, Kim S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. Nat Commun. 2023;14(1):3570. doi:10.1038/s41467-023-39301-y
28. Gao Z, Ding P, Xu R. KG-Predict: A knowledge graph computational framework for drug repurposing. J Biomed Inform. 2022;132:104133. doi:10.1016/j.jbi.2022.104133
29. Perdomo-Quinteiro P, Belmonte-Hernández A. Knowledge graphs for drug repurposing: A review of databases and methods. Brief Bioinform. 2024;25(6):bbae461. doi:10.1093/bib/bbae461
30. Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009
31. Hasselgren C, Oprea TI. Artificial intelligence for drug discovery: Are we there yet? Annu Rev Pharmacol Toxicol. 2024;64:527-50. doi:10.1146/annurev-pharmtox-040323-040828