Knowledge Graphs for Mechanism-Guided Natural Product Therapeutics: Integrating Compounds, Targets, Diseases, Pathways, Evidence, and Safety Signals

Omar Zayed[1]*, Leila Ibrahim[1], Ahmed Mansour[2]

[1]Department of AI for Phytochemical Isolation and Bioactivity Prediction, Faculty of Pharmacy, Cairo University, Cairo, Egypt.
[2]Department of Computational Pharmacology for Chronic Diseases, Faculty of Pharmacy, Alexandria University, Alexandria, Egypt.

ABSTRACT

Natural product therapeutics represent a chemically diverse and biologically complex discovery space in which compounds, plant sources, targets, diseases, pathways, evidence records, and safety signals are often distributed across disconnected data resources. This article proposes a mechanism-guided knowledge graph architecture for organizing these heterogeneous data into a transparent, queryable, and validation-aware framework for natural product therapeutic research. The architecture emphasizes normalized compound, target, gene, protein, disease, phenotype, pathway, evidence, and safety entities, with relation types designed to distinguish curated associations, experimental findings, computational predictions, literature-derived statements, database-derived records, and graph-inferred hypotheses. The proposed approach places evidence provenance and safety-signal integration at the same architectural level as compound–target and disease-pathway reasoning, preventing graph connectivity from being interpreted as proof of mechanism, efficacy, or safety. Query and inference logic are framed around graph traversal, rule-based reasoning, ontology-aware inference, graph embeddings, link prediction, graph neural networks, uncertainty assessment, and expert review. The main contribution is a layered architecture that supports mechanism-guided candidate prioritization while preserving interpretability, evidence traceability, and validation requirements. For natural product therapeutics, such a framework can help organize phytochemical evidence, generate compound–target–disease hypotheses, examine pathway plausibility, screen for safety concerns, and plan translational validation. Future work should focus on reproducible graph construction, natural product entity normalization, evidence-weighting strategies, safety-aware inference, benchmark development, and experimental validation workflows that connect computational hypotheses with pharmacology, toxicology, and translational research.

Key Words: *Knowledge graph, Natural products, Mechanism-guided therapeutics, Compound–target interactions, Pathway analysis, Safety signals*

eIJPPR 2025; 15(3):29-40

HOW TO CITE THIS ARTICLE: Zayed O, Ibrahim L, Mansour A. Knowledge Graphs for Mechanism-Guided Natural Product Therapeutics: Integrating Compounds, Targets, Diseases, Pathways, Evidence, and Safety Signals. Int J Pharm Phytopharmacol Res. 2025;15(3):29-40. https://doi.org/10.51847/2lfewaEoxK

INTRODUCTION

Natural product therapeutics occupy a distinctive position in drug discovery because plant-derived and other naturally occurring compounds can contain structurally diverse scaffolds, complex source metadata, variable bioactivity evidence, and safety considerations that are difficult to interpret through isolated databases or single-method computational tools. Recent knowledge graph scholarship in drug discovery shows that graph-based

Corresponding author: Omar Zayed
Address: Department of AI for Phytochemical Isolation and Bioactivity Prediction, Faculty of Pharmacy, Cairo University, Cairo, Egypt.
E-mail: ✉ omar.zayed@cu.edu.eg
Received: 17 February 2025; **Revised:** 30 May 2025; **Accepted:** 01 June 2025

integration can connect heterogeneous biomedical evidence while making dataset selection, curation quality, and source reliability central design constraints [1, 2]. For natural product research, this integration challenge is especially important because the same compound may be represented through multiple names, chemical identifiers, source organisms, assay contexts, target annotations, disease claims, and safety reports.

Biomedical knowledge graphs provide a way to represent therapeutic evidence as entities and typed relations rather than as disconnected records. Precision-medicine graph resources illustrate how normalized biomedical entities can connect diseases, genes, drugs, phenotypes, and other evidence-bearing nodes into a structure that supports interpretable traversal and hypothesis generation [3]. This principle is directly relevant to mechanism-guided natural product therapeutics, where compounds should not merely be listed beside possible targets or indications, but should be connected through evidence-tagged relations that identify how the association was generated, what biological context it reflects, and what uncertainty remains.

Natural product science increasingly requires multimodal integration across chemistry, bioactivity, taxonomy, omics, pathway biology, pharmacology, and computational prediction. Recent work on artificial intelligence and knowledge graphs in natural product science emphasizes that these technologies are most useful when they organize diverse evidence streams and support discovery reasoning rather than replace chemical authentication, pharmacological testing, or toxicological validation [4]. A knowledge graph architecture for this domain must therefore be designed as a validation-aware evidence system, not as a mechanism-confirmation engine.

The aim of this article is to propose a mechanism-guided knowledge graph architecture for natural product therapeutics that integrates compounds, targets, diseases, pathways, evidence provenance, and safety signals. The central architecture questions are: which entity layers are required, how relations should distinguish evidence types and uncertainty, how safety and ADMET information should be integrated, how graph query and inference logic can support candidate prioritization, and how graph-derived outputs should be routed toward expert review and experimental validation rather than treated as confirmed therapeutic mechanisms.

Knowledge graphs in natural product discovery

In natural product discovery, a knowledge graph can be understood as a structured representation of entities such as compounds, plant sources, targets, diseases, pathways, evidence records, and safety signals, connected through typed relations that can be queried, traversed, updated, and evaluated. Drug repurposing knowledge graphs and natural product–drug interaction graph frameworks show that therapeutic graph models can represent compound–target–disease evidence while also requiring safety-aware relation design for natural product interaction contexts [5, 6]. This distinction is important because a graph edge should not be interpreted as a universal biological fact; it should be interpreted according to its evidence type, source, context, and uncertainty.

Knowledge graphs differ from simple databases because they encode relational meaning rather than only storing records. They also differ from static network diagrams because their nodes and edges can carry identifiers, provenance, relation types, evidence labels, and computationally actionable semantics. In natural product discovery, sample-centric computational frameworks show that compounds, biological samples, annotations, and bioactivity evidence need to remain connected so that discovery claims can be traced back to their experimental or analytical context [7]. Without such provenance, a compound–target or compound–disease connection may become detached from the sample, assay, or annotation workflow that originally supported it.

Natural product resources also show why compound-source-activity representation must be explicit. A natural product activity database can support biomedical research by linking natural products with activity information and species-source context, but architecture design must still normalize compound identities and distinguish activity records from mechanism-level claims [8]. A mechanism-guided knowledge graph should therefore treat natural product activity as one evidence layer among several, not as a direct substitute for target engagement, pathway validation, toxicology, or translational efficacy evidence.

For therapeutic hypothesis generation, the knowledge graph should function as a reasoning environment that supports questions such as whether a compound has evidence-linked targets, whether those targets are associated with disease biology, whether the targets participate in relevant pathways, and whether the compound carries safety or interaction concerns. The graph should also make negative and cautionary information visible. For example, a compound with plausible pathway links may still be deprioritized if ADMET alerts, toxicogenomic signals, herb–drug interaction evidence, or weak provenance indicate that the hypothesis is not yet suitable for experimental progression.

Compound, target, and disease entity layers

The compound layer is the entry point for most natural product therapeutics queries, but it is also one of the most error-prone layers because compounds may be represented by trivial names, synonyms, stereochemical variants, salt forms, mixtures, source-derived annotations, and assay-

specific identifiers. Bioactivity resources such as ChEMBL support assay-linked compound–target evidence that can inform relation construction [9]. Chemical identifier resources such as PubChem support normalization through structures, identifiers, synonyms, and cross-references that are necessary for reducing duplicate or ambiguous compound nodes [10]. In the proposed architecture, compound nodes should therefore include preferred names, synonyms, SMILES, InChI, InChIKey, source context, chemical class when available, and provenance-tagged activity records.

The target layer should represent genes, proteins, therapeutic targets, target classes, target identifiers, organism context, and validation status. Therapeutic target resources demonstrate the importance of connecting targets with drugs, indications, and development-relevant annotations rather than treating targets as isolated molecular labels [11]. In a natural product therapeutics knowledge graph, this means that a compound–target edge should specify whether the relation comes from experimental binding data, functional assay evidence, database curation, literature extraction, computational prediction, or graph inference. This distinction prevents predicted or weakly supported edges from being interpreted as confirmed target engagement.

The disease layer should represent diseases, phenotypes, disease classes, clinical or biological contexts, and gene-disease or target-disease associations. Disease genomics platforms show that gene-disease associations require evidence support, scoring logic, and careful interpretation because association does not necessarily imply causality or therapeutic tractability [12]. For mechanism-guided natural product research, disease nodes should therefore be connected to targets, pathways, phenotypes, and evidence records in a way that distinguishes disease relevance from therapeutic efficacy. A compound connected to a disease-associated target should be considered a candidate hypothesis only when supporting evidence and safety context are also visible.

Safety-relevant biomedical entity layers should not be postponed until late-stage interpretation. Toxicogenomic resources demonstrate that chemical, gene, and disease evidence can be linked in ways that support safety and mechanism assessment, but these associations still require careful interpretation and toxicological validation [13]. In the proposed architecture, safety nodes should include toxicogenomic associations, adverse-event concepts, toxicity alerts, ADMET properties, herb–drug interaction signals, contraindication warnings, and validation status. These nodes should be queryable alongside compound, target, and disease nodes so that candidate prioritization is mechanism-guided and safety-aware from the beginning.

Table 1 defines the core entity and relation layers required for a mechanism-guided natural product therapeutics knowledge graph.

Table 1. Core Entity and Relation Layers for a Mechanism-Guided Natural Product Therapeutics Knowledge Graph: Compound, Target, Disease, Pathway, Evidence, Safety, Identifier, Provenance, and Validation Requirements

Graph layer	Entity type	Required identifiers	Primary relation types	Evidence source	Normalization requirement	Quality-control issue	Validation need	Mechanism-guided use
Natural product compound layer	Compound, metabolite, phytochemical, extract-associated constituent	Preferred name, synonyms, SMILES, InChI, InChIKey, database accession where available	is_same_as, derived_from, has_structure, has_activity, belongs_to_chemical_class	Chemical databases, natural product databases, assay records, literature-derived annotations	Resolve synonyms, stereochemistry, salt forms, mixtures, duplicate records, and source-specific naming	Misidentified compounds, ambiguous names, incomplete structural information	Compound authentication, structural verification, assay reproducibility	Establishes the starting node for compound-centered mechanism queries
Plant or biological source layer	Plant species, organism, tissue, sample, extract, preparation	Taxonomic name, source accession, sample identifier, preparation descriptor	produces, contains, extracted_from, associated_with_sample	Natural product databases, sample-centric workflows, literature reports	Harmonize taxonomic synonyms, preparation terms, and sample metadata	Source variability, extract complexity, missing preparation context	Botanical authentication and reproducible extraction metadata	Links compounds to biological origin and sample provenance

Chemical identifier layer	Identifier record, synonym record, structural representation	SMILES, InChI, InChIKey, registry-style identifiers, cross-references	maps_to, has_synonym, has_cross_reference, represents_structure	Chemical identifier resources and curated compound databases	Cross-map identifiers across resources and remove duplicate chemical nodes	Identifier conflicts, incomplete stereochemical representation	Manual curation of high-priority ambiguous compounds	Enables reliable merging of compound evidence across databases
Bioactivity and assay layer	Assay, endpoint, activity measurement, experimental context	Assay identifier, target identifier, endpoint type, activity unit when reported	tested_in, has_endpoint, active_against, inactive_against, supports_edge	Bioactivity databases, experimental literature	Preserve assay type, organism, endpoint, unit, and experimental context	Assay heterogeneity, incomparable endpoints, missing experimental conditions	Experimental replication and context-specific interpretation	Distinguishes assay evidence from inferred mechanism
Target layer	Gene, protein, therapeutic target, target class	Gene symbol, protein accession, species, target database identifier	binds_to, modulates, inhibits, activates, associated_with_disease, participates_in_pathway	Target databases, bioactivity resources, curated literature	Normalize gene/protein identifiers and species context	Ortholog confusion, obsolete symbols, target-family ambiguity	Target engagement assays and functional validation	Connects compounds to plausible molecular mechanisms
Disease and phenotype layer	Disease, indication, phenotype, clinical or biological condition	Disease ontology identifier, phenotype identifier, synonym mapping	associated_with_gene, associated_with_target, has_phenotype, candidate_for	Disease genomics resources, biomedical KGs, literature-derived associations	Harmonize disease names, phenotype terms, and indication labels	Association does not imply causality or therapeutic efficacy	Disease-model relevance and biological validation	Links target biology to disease-relevant hypotheses
Pathway and biological process layer	Pathway, reaction, biological process, mechanism module	Pathway identifier, reaction identifier, process term	participates_in, regulates, upstream_of, downstream_of, affects_process	Curated pathway resources and mechanism literature	Map target identifiers to pathway and process vocabularies	Pathway granularity mismatch and context dependence	Pathway-level experimental validation	Supports mechanism-of-action plausibility assessment
Evidence provenance layer	Evidence record, source record, assertion record, extraction method	Source DOI or database accession, evidence type, extraction method, version/date when available	supports, contradicts, derived_from, predicted_by, curated_from, mined_from	Databases, peer-reviewed literature, text mining, computational prediction, expert curation	Encode evidence type, source, date/version, and method	Unsupported edges, stale sources, weak provenance	Expert review and reproducibility checks	Makes every graph edge traceable and interpretable
Safety and ADMET layer	ADMET alert, toxicity signal, adverse-event concept, interaction warning	Safety term, toxicity endpoint, ADMET property label,	has_safety_signal, has_ADMET_alert, interacts_with, contraindicated_with, requires_review	Toxicogenomic resources, ADMET prediction tools, safety literature, interaction evidence	Separate predicted alerts from curated or experimental safety evidence	Prediction uncertainty, underreported adverse effects, context-specific risk	Toxicology, pharmacokinetics, and interaction validation	Filters mechanism-supported candidates through safety constraints

ISSN (Online) 2249-6084 (Print) 2250-1029

		interaction concept						
Validation and review layer	Validation status, expert-review record, experimental plan, translational-readiness label	Review status, validation method, evidence tier, decision record	reviewed_by, requires_validation, validated_by, deprioritized_due_to, updated_after	Expert review, experimental studies, translational development records	Standardize validation categories and decision provenance	Overinterpretation of graph connectivity or prediction outputs	Target engagement, pathway validation, toxicology, pharmacokinetics, preclinical and clinical evaluation where appropriate	Routes graph-derived hypotheses toward responsible candidate prioritization

Pathway, evidence, and safety signal integration

Pathway integration is essential because compound–target associations become more interpretable when they are connected to biological processes, reactions, and disease-relevant pathway contexts. Curated pathway resources support this layer by linking genes, proteins, reactions, pathways, genomes, and biological processes in structured forms that can be connected to compound and disease evidence [14, 15]. In a natural product therapeutics knowledge graph, pathway nodes should therefore be represented as explicit entities rather than as post hoc annotations. This allows graph queries to ask whether a compound-associated target participates in pathways that are biologically plausible for a disease context.

Evidence integration should distinguish database-derived associations, curated experimental evidence, literature-mined statements, computational predictions, and graph-inferred hypotheses. Integrated biomedical graph models show that path-based connections among compounds, genes, diseases, and biological processes can support repurposing hypotheses, but these connections remain prioritization signals rather than validated mechanisms [16]. For natural product therapeutics, this means that a compound–target–pathway–disease route should be interpreted as a mechanism-guided hypothesis only when the evidence path is visible, the relation types are clear, and uncertainty is retained.

Literature-derived evidence can expand graph coverage, especially for molecular mechanisms that are distributed across many publications, but automated extraction requires safeguards. Large-scale mechanism assembly from text mining and curated databases demonstrates the value of structured molecular relation extraction while also highlighting the need for evidence tracking and source transparency [17]. In the proposed architecture, every literature-mined relation should be attached to a source record, extraction method, relation type, and review status so that users can distinguish extracted assertions from curated or experimentally validated edges.

Safety integration should be represented as a core graph layer, not as a final manual screening step. Computational ADMET platforms can contribute absorption, distribution, metabolism, excretion, and toxicity predictions, but these predictions must be encoded as uncertain safety evidence rather than definitive toxicological outcomes [18]. The architecture should therefore connect compounds to ADMET alerts, toxicity signals, interaction warnings, and adverse-event concepts through provenance-tagged relations that can be filtered during candidate prioritization and routed toward pharmacokinetic, toxicological, and expert-review validation.

Table 2 summarizes how pathway evidence, provenance metadata, ADMET information, toxicity signals, and safety alerts can be integrated into a natural product therapeutics knowledge graph.

Table 2. Pathway, Evidence Provenance, ADMET, Toxicity, and Safety Signal Integration in Natural Product Therapeutics Knowledge Graphs: Data Sources, Relation Types, Confidence Constraints, Interpretation Limits, and Translational Safeguards

Integration domain	Data or evidence source	Graph relation encoded	Mechanistic interpretation	Safety or ADMET relevance	Evidence-confidence constraint	Interpretation limitation	Translational risk	Safeguard or mitigation strategy
Pathway membership	Curated pathway resources	target_participates_in_pathway; pathway_contains_process	Connects molecular targets to biological processes and pathway modules	Indirect safety relevance through pathway	Requires source and pathway-version provenance	Pathway membership does not prove disease-	Overstating pathway relevance from target	Require disease-context evidence and pathway-

				toxicity or off-target biology		modifying activity	presence alone	level validation
Reaction-level mechanism	Curated reaction and pathway knowledgebases	protein_regulates_reaction; reaction_part_of_pathway	Supports mechanistic route reconstruction from target to biological effect	May reveal mechanistic sources of toxicity or off-target effects	Requires curated reaction evidence and organism context	Reaction annotation may not reflect natural product exposure context	Treating canonical pathways as compound-specific mechanisms	Add assay context and experimental validation status
Database-derived association	Biomedical and natural product databases	database_supports_edge; record_derived_from_source	Provides structured association evidence for graph construction	May include safety, target, disease, or interaction evidence	Must preserve source database and accession	Database presence does not guarantee causality or current validity	Propagation of stale or weak associations	Version tracking, deduplication, and expert review
Literature-mined evidence	Peer-reviewed articles processed by relation extraction	text_mined_relation; sentence_supports_edge	Expands mechanism and association coverage	Can capture toxicity, interaction, and contraindication statements	Requires extraction method, source DOI, and review status	NLP extraction can misclassify negation, context, or relation direction	False-positive mechanism or safety edges	Human review of high-impact edges and relation-confidence labeling
Experimental evidence	Binding, activity, functional, pharmacology, or toxicology reports	experimental_evidence_supports_edge	Links graph relations to assay context	Can support safety or efficacy-related interpretation depending on endpoint	Requires assay type, organism, endpoint, and conditions	Assay results may not translate across models or exposure levels	Overgeneralizing in vitro evidence to therapeutic relevance	Context-aware evidence annotation and replication planning
Computational prediction	Target prediction, ADMET prediction, link prediction, or GNN output	predicted_edge; predicted_property; generated_by_model	Supports hypothesis generation and candidate ranking	Central for ADMET alerts and toxicity prioritization	Requires model type, input data, applicability limits, and uncertainty status	Prediction is not experimental confirmation	Advancing candidates based only on model output	Require orthogonal evidence and experimental validation
Toxicogenomic evidence	Chemical–gene–disease toxicology resources	chemical_associated_with_toxicity; chemical_affects_gene	Connects chemical exposure to toxicological mechanisms	Directly supports safety-signal layer	Requires curated toxicogenomic source and evidence context	Toxicogenomic association does not equal clinical toxicity	Mislabeling risk without dose or exposure context	Add exposure context and toxicology review
ADMET alert	ADMET prediction platforms or curated pharmacokinetic evidence	compound_has_ADMET_alert; property_predicted_by	Flags developability and safety constraints	Direct relevance to absorption, metabolism, toxicity, and interaction concerns	Requires prediction method and model applicability statement	Alert does not quantify clinical risk	False reassurance or false exclusion	Use as screening signal, not final safety decision
Herb–drug or natural product–	Interaction literature, pharmacokin	natural_product_interacts_with_drug;	Connects natural products to PK or	Direct relevance to contraindicat	Requires interaction mechanism	Interaction evidence may be	Undetected interaction	Add safety-filtered queries and

drug interaction	etic evidence, curated interaction records	affects_enzyme_or_transporter	PD interaction mechanisms	ions and clinical caution	, evidence type, and context	formulation- and dose-dependent	risk during prioritization	pharmacokinetic expert review
Validation status	Expert review, experimental follow-up, translational decision records	edge_reviewed_by; hypothesis_requires_validation; validated_by	Separates hypotheses from confirmed findings	Supports safe interpretation of candidate ranking	Requires explicit status labels and decision provenance	Review status may become outdated	Treating old review status as final	Periodic updating and validation feedback loops

Mechanism-guided query and inference logic

Mechanism-guided query logic should begin with transparent graph traversal. A basic query may start from a natural product compound, traverse to curated or predicted targets, connect those targets to disease-associated genes or proteins, and then examine whether the target participates in disease-relevant pathways. Knowledge graph embeddings can support discovery of protein target hypotheses by learning latent graph structure, but their outputs must be treated as candidate associations that require biological interpretation and validation [19]. In the proposed architecture, embedding-derived edges should therefore be labeled as predicted, model-generated, and uncertainty-bearing.

Heterogeneous network learning can extend this logic by integrating multiple entity and relation types into target-identification workflows. Deep learning approaches over heterogeneous biomedical networks show that graph-based models can support target identification across drug, target, disease, and biological evidence layers [20]. For natural product therapeutics, this supports an inference module in which graph traversal identifies interpretable evidence paths, while machine learning modules rank candidate compound–target or compound–disease hypotheses. The architecture should prevent rankings from being interpreted as mechanisms unless supported by independent experimental and pathway evidence.

Safety-filtered inference is also required because mechanism plausibility does not guarantee therapeutic suitability. Graph convolutional modeling of polypharmacy side effects shows that graph neural networks can represent safety-relevant drug interaction patterns [21]. Although natural product therapeutics require additional attention to plant source, extract context, pharmacokinetics, and herb–drug interaction evidence, the architecture can adapt this principle by including safety nodes in candidate-ranking queries. A safety-filtered query should ask not only whether a compound is connected to disease-relevant mechanisms, but also whether it is connected to ADMET alerts, toxicity evidence, interaction risks, or contraindication signals.

Interpretability and validation safeguards should be built into the inference workflow. Neuro-symbolic representation learning on biological knowledge graphs supports the idea that ontology-aware constraints can make graph inference more biologically structured [22]. Evaluation studies of knowledge graph embeddings in drug discovery further show that embedding performance depends on graph structure, task definition, and benchmark design [23]. Broader biomedical embedding evaluations also indicate that link prediction and downstream tasks must be interpreted in a task-specific way rather than assumed to be universally useful [24]. Therefore, graph-derived natural product hypotheses should be reviewed through provenance, path explainability, uncertainty, safety filtering, and expert assessment before experimental validation is planned.

Figure 1 illustrates the mechanism-guided query and inference workflow that connects compounds, targets, diseases, pathways, evidence provenance, and safety filters in a natural product therapeutics knowledge graph.

ISSN (Online) 2249-6084 (Print) 2250-1029

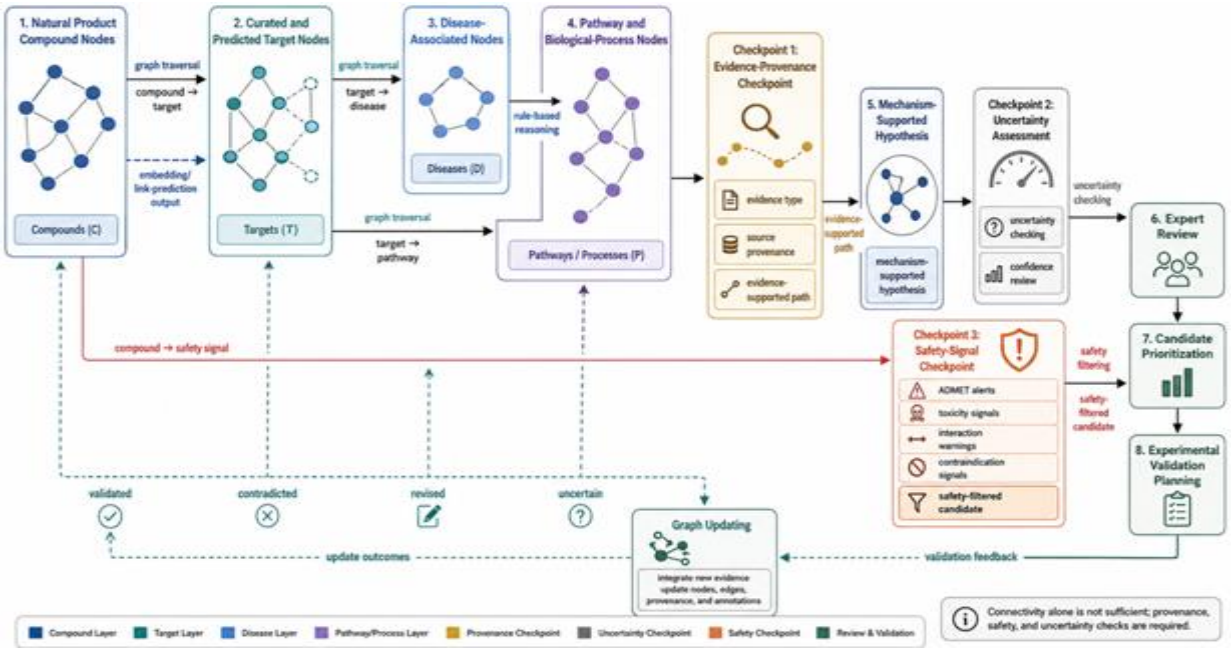

Figure 1. Mechanism-Guided Query and Inference Workflow for Natural Product Therapeutics Knowledge Graphs

Table 3 presents mechanism-guided query and inference logic for using knowledge graphs to prioritize natural product therapeutic hypotheses while preserving evidence transparency and validation requirements.

Table 3. Mechanism-Guided Query and Inference Logic for Natural Product Therapeutics Knowledge Graphs: Query Purpose, Graph Pattern, Inference Method, Evidence Requirement, Safety Filter, Uncertainty Source, Validation Need, and Candidate Action

Query or inference goal	Graph pattern queried	Inference method	Required evidence layer	Safety or ADMET filter	Uncertainty source	Interpretation boundary	Validation requirement	Candidate or research action
Identify candidate targets for a natural product compound	Compound → target → evidence	Graph traversal; embedding-assisted target ranking	Compound–target evidence; provenance record	Remove or flag compounds with severe safety alerts	Assay variability; prediction uncertainty	Candidate target, not confirmed target engagement	Binding, biochemical, or cellular target engagement assays	Prioritize targets for experimental testing
Connect targets to disease biology	Target → disease → phenotype → evidence	Rule-based traversal; ontology-aware mapping	Target–disease and gene–disease evidence	Check disease-specific contraindication or toxicity context	Association strength and disease-context mismatch	Disease relevance, not therapeutic efficacy	Disease-model validation and functional assays	Select disease contexts for mechanistic follow-up
Explore pathway plausibility	Compound → target → pathway → biological process	Pathway traversal; mechanism path reconstruction	Curated pathway and reaction evidence	Screen for pathway-linked toxicity or off-target concerns	Pathway granularity and biological context	Pathway plausibility, not pathway modulation proof	Pathway activity assays and omics validation	Generate pathway-level mechanism hypotheses
Rank compound–disease hypotheses	Compound → target → disease; compound → pathway → disease	Link prediction; graph embeddings; graph neural networks	Multi-source evidence with provenance labels	Exclude or downgrade safety-flagged candidates	Model bias; incomplete graph coverage	Prioritized hypothesis, not clinical indication	Orthogonal computational review and experimental validation	Create ranked candidate list for expert review
Detect safety concerns early	Compound → ADMET alert; compound → toxicity signal;	Safety-filtered graph query	ADMET, toxicogenomic, adverse-event,	Required safety layer query	Prediction uncertainty and incomplete	Safety signal, not definitive clinical risk	Toxicology, pharmacokinetic, and interaction assessment	Flag, deprioritize, or route to

	compound → interaction warning		and interaction evidence		safety reporting			safety review
Distinguish evidence types	Edge → evidence record → source/method	Provenance query; evidence-type filtering	Curated, experimental, literature-mined, database-derived, predicted, or inferred evidence	Identify safety source type and reliability	Missing provenance or weak extraction metadata	Evidence support level, not truth status	Manual review of high-impact edges	Retain, downgrade, or exclude edges
Explain predicted links	Predicted edge → supporting paths → evidence nodes	Path-based explainability; rule inspection	Path evidence and relation provenance	Include safety paths in explanations	Non-causal paths and graph shortcuts	Explanation of graph rationale, not biological proof	Expert review and targeted experiments	Use explainable paths to guide validation design
Update graph after validation	Hypothesis → validation result → graph edge status	Feedback loop; rule-based status update	Validation-status evidence	Update safety and ADMET status where relevant	Changing evidence and version drift	Updated graph state, not permanent conclusion	Reproducible documentation and periodic review	Promote, revise, or retire graph hypotheses

Proposed knowledge graph architecture

The proposed architecture is a layered system that begins with data sources and moves through entity normalization, relation construction, evidence provenance, safety-signal integration, graph storage, query logic, inference modules, explainability, uncertainty assessment, expert review, and validation feedback. Open-source biomedical knowledge graph ecosystems demonstrate the importance of ontology-aligned construction, source transparency, and reusable graph components for life-science applications [25]. In a natural product therapeutics setting, this means that compounds, plant sources, targets, diseases, pathways, evidence records, and safety signals should be treated as interoperable graph layers rather than as disconnected tables or isolated prediction outputs.

Figure 2 presents the proposed knowledge graph architecture for mechanism-guided natural product therapeutics, integrating entity layers, relation layers, evidence provenance, safety signals, inference modules, and validation safeguards.

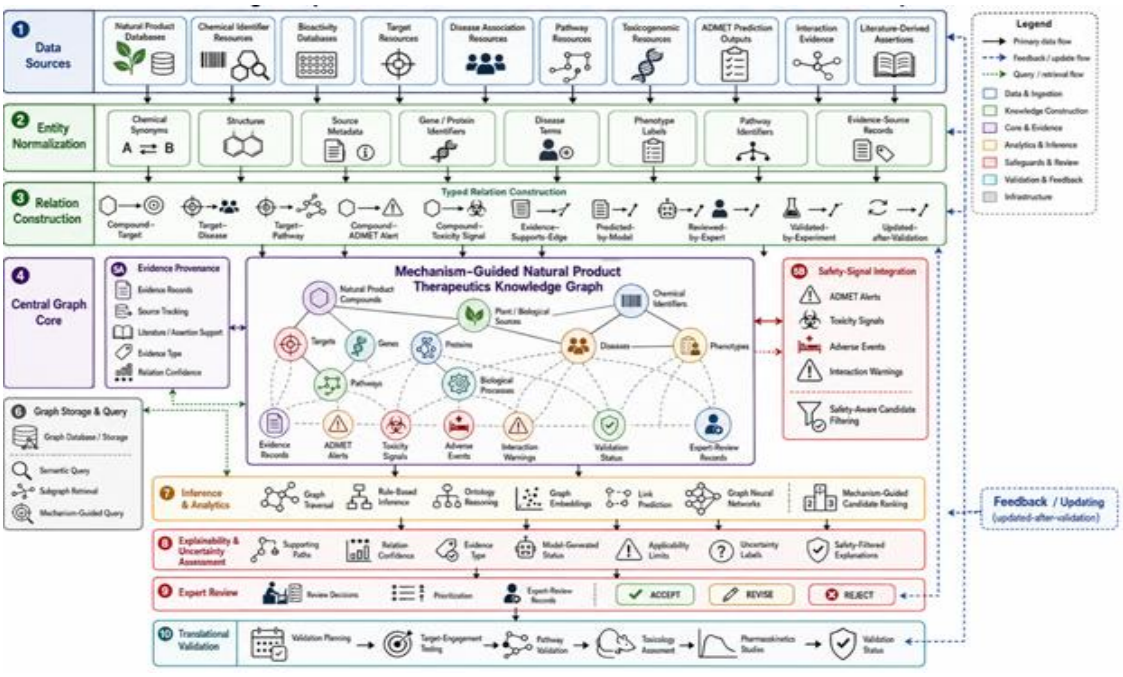

Figure 2. Knowledge Graph Architecture for Mechanism-Guided Natural Product Therapeutics

The data-source layer should include natural product databases, chemical identifier resources, bioactivity databases, target resources, disease association resources, pathway resources, toxicogenomic resources, ADMET prediction outputs, interaction evidence, and literature-derived assertions. Biomedical knowledge graph construction literature emphasizes that graph design should specify how entities are selected, how relations are constructed, how sources are integrated, and how applications shape the representation [26]. For this architecture, the entity-normalization layer should resolve chemical synonyms, structural identifiers, gene and protein names, disease terms, phenotype labels, pathway identifiers, and evidence-source records before relation construction begins.

The relation-construction and graph-update layers should be reproducible and version-aware. Knowledge graph exchange frameworks show that reusable build pipelines, standardized transformation logic, and shareable graph artifacts are important for reproducibility and interoperability [27]. In natural product therapeutics, this means that each graph build should preserve source versions, transformation steps, mapping decisions, relation definitions, and quality-control decisions. When new evidence becomes available, the graph should update relation status rather than silently overwrite prior evidence, allowing users to see whether an edge has been strengthened, contradicted, revised, or retired.

The graph storage and query layer should support both direct traversal and federated retrieval. Federated biomedical query engines demonstrate that graph-based discovery can connect multiple biomedical APIs and resources through query logic rather than requiring all data to be permanently centralized [28]. A natural product therapeutics graph can therefore use a hybrid approach: core high-value entities and relations can be stored in the graph, while additional evidence can be retrieved through controlled query interfaces when needed. This design supports extensibility while preserving provenance and avoiding unnecessary duplication of external resources.

The evidence-provenance and explainability layers are central to responsible architecture design. Biomedical relation extraction methods can improve the coverage of literature-derived graph edges, but extracted relations require classification, source tracking, and review because automated extraction can misrepresent context or relation direction [29]. Explainable path-based knowledge graph completion further supports the use of reviewable paths for drug-repurposing hypotheses, provided that path explanations are interpreted as graph rationales rather than biological proof [30]. In this architecture, every inferred hypothesis should therefore be accompanied by its supporting paths, source evidence, relation types, safety signals, uncertainty labels, and validation requirements.

Use cases and research implications

A first use case is compound–target hypothesis generation for natural product therapeutics. Graph neural network approaches to drug repurposing show that graph-based models can support candidate generation across biomedical networks, but such outputs should remain hypothesis-generating until validated by target engagement and functional assays [31]. In a natural product context, this use case would begin with a normalized compound node, retrieve curated and predicted targets, inspect supporting evidence, check disease relevance, and then apply safety filters before selecting targets for experimental testing.

A second use case is integrated natural product repurposing and therapeutic exploration. Unified biomedical knowledge graphs can support drug discovery by combining heterogeneous biomedical entities and relations into a searchable structure [32]. For natural product therapeutics, this implies that repurposing hypotheses should not be based only on compound–disease similarity or target prediction. Instead, the graph should require interpretable paths that connect compound identity, target biology, disease association, pathway plausibility, evidence provenance, and safety context.

A third use case is hidden indication discovery and pathway-based prioritization. Graph neural network approaches using heterogeneous data can uncover candidate therapeutic indications, but these indications remain computational hypotheses that require biological and translational validation [33]. In the proposed architecture, such outputs should be ranked not only by graph inference but also by evidence type, source reliability, pathway interpretability, safety-signal burden, and feasibility of validation. This helps avoid prioritizing highly connected but weakly supported candidates.

A fourth use case is translational-readiness planning for natural product candidates. Disease-specific graph neural network repurposing research shows that harmonizing multiple evidence streams can support urgent therapeutic exploration, but the same logic requires caution because graph-derived evidence remains dependent on data quality, disease context, and validation design [34]. For natural product therapeutics, research implications include the need for better compound normalization, natural product–specific benchmarks, standardized safety-signal representation, interoperable pathway mapping, expert-review workflows, reproducible graph builds, and validation pipelines that connect graph hypotheses with pharmacology, toxicology, pharmacokinetics, and translational development.

CONCLUSION

This article proposes a mechanism-guided knowledge graph architecture for natural product therapeutics that integrates compounds, targets, diseases, pathways, evidence provenance, and safety signals into a transparent and validation-aware discovery framework. The architecture is designed to support hypothesis generation, candidate prioritization, pathway plausibility assessment, safety-aware filtering, and translational validation planning without treating graph connectivity, model prediction, or inferred paths as proof of mechanism, safety, efficacy, or clinical utility. Its practical value depends on entity normalization, relation quality, evidence provenance, uncertainty representation, reproducible graph construction, expert review, and integration with pharmacology, toxicology, medicinal chemistry, pharmacokinetics, preclinical testing, and clinical development. Future progress will require natural product-specific graph benchmarks, improved evidence-weighting methods, safety-aware inference, interoperable graph resources, and feedback loops that update graph knowledge after experimental and translational validation.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Serra A, Fratello M, Federico A, Greco D. An update on knowledge graphs and their current and potential applications in drug discovery. Expert Opin Drug Discov. 2025;20(5):599-619. doi:10.1080/17460441.2025.2490253

[2] Bonner S, Barrett IP, Ye C, Swiers R, Engkvist O, Bender A, et al. A review of biomedical datasets relating to drug discovery: A knowledge graph perspective. Brief Bioinform. 2022;23(6). doi:10.1093/bib/bbac404

[3] Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. Sci Data. 2023;10(1):67. doi:10.1038/s41597-023-01960-3

[4] Meijer D, Beniddir MA, Coley CW, Mejri YM, Öztürk M, van der Hooft JJJ, et al. Empowering natural product science with AI: Leveraging multimodal data and knowledge graphs. Nat Prod Rep. 2025;42(4):654-62. doi:10.1039/D4NP00008K

[5] Boudin M, Diallo G, Drancé M, Mougin F. The OREGANO knowledge graph for computational drug repurposing. Sci Data. 2023;10(1):871. doi:10.1038/s41597-023-02757-0

[6] Taneja SB, Callahan TJ, Paine MF, Kane-Gill SL, Kilicoglu H, Joachimiak MP, et al. Developing a knowledge graph for pharmacokinetic natural product-drug interactions. J Biomed Inform. 2023;140:104341. doi:10.1016/j.jbi.2023.104341

[7] Gaudry A, Pagni M, Mehl F, Moretti S, Quiros-Guerrero LM, Cappelletti L, et al. A sample-centric and knowledge-driven computational framework for natural products drug discovery. ACS Cent Sci. 2024;10(3):494-510. doi:10.1021/acscentsci.3c00800

[8] Zhao H, Yang Y, Wang S, Yang X, Zhou K, Xu C, et al. NPASS database update 2023: Quantitative natural product activity and species source database for biomedical research. Nucleic Acids Res. 2023;51(D1). doi:10.1093/nar/gkac1069

[9] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1). doi:10.1093/nar/gkw1074

[10] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. Nucleic Acids Res. 2023;51(D1). doi:10.1093/nar/gkac956

[11] Wang Y, Zhang S, Li F, Zhou Y, Zhang Y, Wang Z, et al. Therapeutic Target Database 2020: Enriched resource for facilitating research and early development of targeted therapeutics. Nucleic Acids Res. 2020;48(D1). doi:10.1093/nar/gkz981

[12] Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1). doi:10.1093/nar/gkz1021

[13] Davis AP, Grondin CJ, Johnson RJ, Sciaky D, McMorran R, Wiegers J, et al. Comparative Toxicogenomics Database: Update 2021. Nucleic Acids Res. 2021;49(D1). doi:10.1093/nar/gkaa891

[14] Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1). doi:10.1093/nar/gkac963

[15] Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2020;48(D1). doi:10.1093/nar/gkz1031

[16] Himmelstein DS, Lizee A, Hessler C, Brueggeman L, Chen SL, Hadley D, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. Elife. 2017;6. doi:10.7554/eLife.26726

[17] Bachman JA, Gyori BM, Sorger PK. Automated assembly of molecular mechanisms at scale from text mining and curated databases. Mol Syst Biol. 2023;19(5). doi:10.15252/msb.202211325

[18] Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021;49(W1). doi:10.1093/nar/gkab255

[19] Mohamed SK, Nováček V, Nounu A. Discovering protein drug targets using knowledge graph embeddings. Bioinformatics. 2020;36(2):603-10. doi:10.1093/bioinformatics/btz600

[20] Zeng X, Zhu S, Lu W, Liu Z, Huang J, Zhou Y, et al. Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci. 2020;11(7):1775-97. doi:10.1039/C9SC04336E

[21] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 2018;34(13). doi:10.1093/bioinformatics/bty294

[22] Alshahrani M, Khan MA, Maddouri O, Kinjo AR, Queralt-Rosinach N, Hoehndorf R. Neuro-symbolic representation learning on biological knowledge graphs. Bioinformatics. 2017;33(17):2723-30. doi:10.1093/bioinformatics/btx275

[23] Bonner S, Barrett IP, Ye C, Swiers R, Engkvist O, Hoyt CT, et al. Understanding the performance of knowledge graph embeddings in drug discovery. Artif Intell Life Sci. 2022;2:100036. doi:10.1016/j.ailsci.2022.100036

[24] Gema AP, Grabarczyk D, De Wulf W, Borole P, Alfaro JA, Minervini P, et al. Knowledge graph embeddings in the biomedical domain: Are they useful? A look at link prediction, rule learning, and downstream polypharmacy tasks. Bioinform Adv. 2024;4(1). doi:10.1093/bioadv/vbae097

[25] Callahan TJ, Tripodi IJ, Stefanski AL, Cappelletti L, Taneja SB, Wyrwa JM, et al. An open source knowledge graph ecosystem for the life sciences. Sci Data. 2024;11(1):363. doi:10.1038/s41597-024-03171-w

[26] Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J. 2020;18:1414-28. doi:10.1016/j.csbj.2020.05.017

[27] Caufield JH, Putman T, Schaper K, Unni DR, Hegde H, Callahan TJ, et al. KG-Hub—building and exchanging biological knowledge graphs. Bioinformatics. 2023;39(7). doi:10.1093/bioinformatics/btad418

[28] Callaghan J, Xu CH, Xin J, Alvarado Cano M, Riutta A, Zhou E, et al. BioThings Explorer: A query engine for a federated knowledge graph of biomedical APIs. Bioinformatics. 2023;39(9). doi:10.1093/bioinformatics/btad570

[29] Ming S, Zhang R, Kilicoglu H. Enhancing the coverage of SemRep using a relation classification approach. J Biomed Inform. 2024;155:104658. doi:10.1016/j.jbi.2024.104658

[30] Jiménez A, Merino MJ, Parras J, Zazo S. Explainable drug repurposing via path based knowledge graph completion. Sci Rep. 2024;14(1):16587. doi:10.1038/s41598-024-67163-x

[31] Doshi S, Chepuri SP. A computational approach to drug repurposing using graph neural networks. Comput Biol Med. 2022;150:105992. doi:10.1016/j.compbiomed.2022.105992

[32] Kumari M, Chauhan R, Garg P. MedKG: Enabling drug discovery through a unified biomedical knowledge graph. Mol Divers. 2025;29(4):3465-83. doi:10.1007/s11030-025-11164-z

[33] Ayuso-Muñoz A, Prieto-Santamaría L, Ugarte-Carro E, Serrano E, Rodríguez-González A. Uncovering hidden therapeutic indications through drug repurposing with graph neural networks and heterogeneous data. Artif Intell Med. 2023;145:102687. doi:10.1016/j.artmed.2023.102687

[34] Hsieh K, Wang Y, Chen L, Zhao Z, Savitz S, Jiang X, et al. Drug repurposing for COVID-19 using graph neural network and harmonizing multiple evidence. Sci Rep. 2021;11(1):23179. doi:10.1038/s41598-021-02353-5