Ontologies for Natural Product Therapeutics: Standardizing Phytochemicals, Biological Activities, Targets, Pathways, Formulations, and Evidence Levels

Lars Petersen[1]*, Maja Jensen[1], Henrik Nielsen[2], Ursula Hoffmann[3]

[1]Department of AI for Marine Sponge and Cyanobacterial Metabolites, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark.
[2]Department of Computational Binding Free Energy Calculations, Faculty of Pharmacy, Aarhus University, Aarhus, Denmark.
[3]Department of Cheminformatics for Natural Product Libraries, Faculty of Pharmacy, University of Basel, Basel, Switzerland.

ABSTRACT

Natural product therapeutics are represented across heterogeneous chemical, biological, ethnopharmacological, pharmacological, and safety data sources, yet their interpretation often depends on inconsistent terminology, incomplete provenance, weakly defined relationships, and unclear evidence boundaries. This article proposes a conceptual ontology structure for natural product therapeutics that standardizes phytochemical entities, chemical classes, compound identifiers, synonyms, stereochemistry, natural product sources, botanical context, preparation methods, extraction context, formulation entities, biological activities, assay settings, targets, pathways, disease or phenotype contexts, safety evidence, traditional use evidence, and evidence levels. The proposed framework distinguishes controlled vocabularies, taxonomies, ontologies, knowledge graphs, database schemas, semantic classes, entity attributes, semantic relationships, provenance fields, uncertainty fields, and claim-support boundaries. It emphasizes that ontology-based standardization can improve semantic consistency, interoperability, evidence tracing, query logic, and computational reasoning, but cannot by itself validate mechanisms, efficacy, safety, therapeutic equivalence, or regulatory claims. The ontology approach links phytochemical identity to source and preparation context, connects bioactivity to assay provenance, separates target and pathway annotation from causal interpretation, and places formulation and evidence entities within explicit decision-boundary logic. The contribution is a cautious original ontology model intended to support transparent evidence organization and responsible interpretation in natural product therapeutic research.

Key Words: *Natural product therapeutics, Ontology, Phytochemicals, Bioactivity annotation, Targets, Pathways*

eIJPPR 2025; 15(6):34-46

HOW TO CITE THIS ARTICLE: Petersen L, Jensen M, Nielsen H, Hoffmann U. Ontologies for Natural Product Therapeutics: Standardizing Phytochemicals, Biological Activities, Targets, Pathways, Formulations, and Evidence Levels. Int J Pharm Phytopharmacol Res. 2025;15(6):34-46. https://doi.org/10.51847/vdn2ooX8rW

INTRODUCTION

Natural product therapeutics are knowledge-intensive because the same research object may be described as a compound, extract, plant preparation, microbial metabolite, formulation, traditional remedy, bioactivity observation, target hypothesis, pathway annotation, or safety signal. Ontology-based standardization is therefore needed because semantic interoperability, biomedical ontology governance, and awareness of heterogeneous natural-product data resources are all necessary for consistent representation across data sources [1–3].

Corresponding author: Lars Petersen
Address: Department of AI for Marine Sponge and Cyanobacterial Metabolites, Faculty of Pharmaceutical Sciences, University of Copenhagen, Copenhagen, Denmark.
E-mail: ✉ lars.petersen@sund.ku.dk
Received: 28 August 2025; **Revised:** 26 November 2025; **Accepted:** 27 November 2025

In this setting, an ontology is not merely a list of terms or a storage structure; it is a formalized representational framework that defines classes, attributes, relationships, provenance requirements, uncertainty fields, and interpretation boundaries.

Natural product research also requires explicit linkage between chemical identity, biological source, and evidence provenance. Open natural-product knowledge-management initiatives show the importance of connecting compound records with organismal source context and structured provenance, supporting the view that phytochemical identity should not be represented independently from biological origin and traceable evidence [4]. For therapeutic interpretation, this linkage matters because the same named compound, extract, or preparation may carry different evidentiary implications depending on source material, preparation conditions, assay context, or formulation state.

Fragmented terminology creates interpretive risk. A controlled vocabulary can normalize preferred names and synonyms, a taxonomy can arrange terms into hierarchical classes, a database schema can organize records, and a knowledge graph can connect entities through typed edges; however, an ontology must define the intended meaning of those entities and relationships. Without this distinction, a relation such as *has_activity*, *acts_on_target*, or *associated_with_condition* may be read as stronger than the evidence allows. In natural product therapeutics, such ambiguity can lead to unsupported mechanistic, safety, or clinical interpretations.

The aim of this article is to develop a conceptual ontology structure for natural product therapeutics. The proposed ontology model organizes phytochemicals, biological activities, targets, pathways, formulation and preparation contexts, evidence levels, provenance fields, uncertainty fields, semantic relationships, expert curation status, auditability, and claim-support boundaries. The framework is intended to support representation, interoperability, evidence integration, and cautious computational reasoning, not to establish therapeutic validity, clinical effectiveness, safety, or regulatory acceptability.

Ontology needs in natural product therapeutics

A major ontology need in natural product therapeutics is the ability to represent entities as semantically typed nodes connected by explicit relationships. Life-science knowledge graph work shows how biomedical entities can be connected through machine-readable relationships that support integration across heterogeneous data types [5]. For natural product therapeutics, this logic is useful only when relationships remain semantically constrained; for example, *found_in*, *derived_from*, *has_activity*, *tested_in_assay*, and *supported_by_evidence* should not be treated as equivalent relations.

Compound identity is another core need. Open natural-product compound collections demonstrate the value of standardized compound records, identifiers, structures, and source metadata for natural product informatics [6]. A natural product therapeutic ontology should therefore distinguish the phytochemical entity from its synonyms, external identifiers, structural annotations, stereochemical qualifiers, chemical class labels, and source links. This distinction helps prevent synonym collapse, mistaken compound merging, and unsupported transfer of evidence across structurally or contextually different entities.

Source context requires equal attention. Curated natural-product resources focused on microbially derived compounds show that source-aware records benefit from explicit organism and compound links supported by curation [7]. In the proposed ontology logic, *natural product source* should include botanical or microbial taxon, plant part where relevant, specimen or strain context where available, and preparation context when the evidence concerns an extract or formulation rather than an isolated compound. This prevents the ontology from treating a purified molecule, crude extract, and multi-ingredient preparation as interchangeable therapeutic entities.

Safety and interaction evidence further illustrate why natural product therapeutics require ontology-based standardization rather than simple terminology alignment. Knowledge graph modeling for pharmacokinetic natural product–drug interactions demonstrates that natural products, drugs, enzymes, transporters, pharmacokinetic outcomes, and evidence sources can be represented as structured interaction relations [8]. A pharmacokinetic interaction repository also shows that chemical characterization, product context, and clinical study provenance are necessary for interpreting interaction evidence [9]. These examples support a distinct interaction-evidence layer in which exposure, formulation, chemical characterization, ADMET context, study type, and uncertainty are represented before any clinical interpretation is made.

Phytochemical and bioactivity classes

Phytochemical standardization begins with separating chemical identity from chemical classification. Structural classification tools and spectral annotation methods support the organization of natural products into chemical classes and predicted structural categories, but they also show why inferred classifications and inferred structures require uncertainty metadata [10–12]. In the proposed ontology, a *phytochemical entity* should therefore carry fields for preferred name, synonyms, identifiers, structure

representation, stereochemistry where relevant, chemical class, source metadata, and identity confidence. The *chemical class* should be represented as a taxonomic or ontology class that supports grouping without implying shared pharmacological action.

Compound identifiers and synonym handling are essential for interoperability. Public chemical infrastructure provides persistent compound records and linked chemical information that can support cross-resource mapping in natural product informatics [13]. In an ontology for natural product therapeutics, identifier relationships such as *same_as*, *has_synonym*, and *has_external_identifier* should be separated from biological relationships such as *has_activity* or *acts_on_target*. This prevents representational errors in which alternative names or database links are mistaken for evidence of biological equivalence.

Bioactivity classes require a different representational logic from chemical classes. Bioassay resources show that activity data depend on assay type, endpoint, experimental system, and measurement context [14]. Therefore, a *bioactivity class* should not simply state that a natural product is "anti-inflammatory," "antioxidant," or "cytotoxic" without specifying the assay context that supports the activity statement. The ontology should represent assay type, organism or cell system, endpoint, experimental condition, potency or qualitative result where available, and evidence provenance, while also recording whether the activity relation is direct, inferred, predicted, or curated from experimental evidence.

Curated bioactivity platforms further indicate that compound, assay, and target annotations need expert curation and structured evidence links [15]. For natural product therapeutics, this means that a *has_activity* relation should be bounded by the evidence level and assay context that generated it. A cell-based activity result, target-binding assay, biochemical inhibition result, and phenotypic observation may all support different kinds of claims. The ontology should therefore preserve both chemical specificity and biological context, allowing activity claims to be queried and compared without converting assay-level evidence into therapeutic validation.

Table 1 summarises ontology classes needed to standardize phytochemical entities and biological activity annotations in natural product therapeutics.

Table 1. Ontology Classes for Phytochemical and Bioactivity Standardization in Natural Product Therapeutics: Chemical Entities, Compound Classes, Synonyms, Stereochemistry, Source Metadata, Bioactivity Types, Assay Context, Evidence Provenance, and Interpretation Boundaries

Ontology class	Definition scope	Key attributes	Example semantic relationship	Evidence requirement	Provenance requirement	Uncertainty field	Failure risk if absent
Phytochemical entity	A chemically defined natural-product constituent or annotated compound entity	Preferred name, structure representation, external identifier, synonym set, stereochemistry, source link	*has_chemical_class*; *found_in_source*	Chemical identity evidence or clearly marked inferred identity	Source database, publication, curator, version	Identity confidence; ambiguous structure status	Compounds, extracts, and mixtures may be incorrectly merged
Chemical class	A structural or biosynthetic grouping used to classify phytochemicals	Class label, parent class, classification method, hierarchy position	*subclass_of*; *classifies_compound*	Curated or computational classification evidence	Classification source and method	Predicted versus curated class status	Activity may be overgeneralized across broad classes
Compound identifier	A stable external identifier or registry-style link for a compound entity	Identifier string, identifier source, mapped compound, mapping status	*has_external_identifier*; *same_as*	Identifier mapping evidence	Database source and access version	Mapping ambiguity	Records from different resources may not interoperate
Synonym set	Alternative names associated with a phytochemical entity	Synonym, language where relevant, source, preferred-name status	*has_synonym*	Name-mapping evidence	Nomenclature source or database	Synonym confidence; homonym risk	Different compounds may be collapsed under one name

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

Stereochemical qualifier	Explicit representation of stereochemistry where relevant	Chiral center, stereochemical descriptor, known or unknown status	*has_stereochemistry*	Structural evidence sufficient to support stereochemical assignment	Analytical or database source	Unknown or partially resolved stereochemistry	Distinct stereoisomers may be treated as equivalent
Natural product source metadata	Biological or material context from which the phytochemical or preparation is derived	Taxon, organism, plant part, strain, specimen, preparation context	*derived_from*; *found_in*	Source evidence linked to compound or preparation	Source record, taxonomic authority, publication	Taxonomic or source ambiguity	Evidence may be transferred across non-equivalent sources
Bioactivity class	A categorized biological effect or measured activity	Activity term, endpoint, direction, measurement type	*has_activity*	Assay or curated bioactivity evidence	Assay record or publication	Activity confidence; endpoint ambiguity	Assay signals may be misread as therapeutic effects
Assay context	Experimental setting in which bioactivity is measured	Assay type, model system, organism or cell type, endpoint, conditions	*measured_in_assay*; *supports_activity*	Experimental or curated assay evidence	Protocol, source publication, database record	Reproducibility and context uncertainty	Incompatible assays may be compared as equivalent
Evidence provenance	Traceable source for an identity, activity, or classification assertion	Source type, citation, curator, date, version, extraction method	*supported_by_evidence*	Required for every assertion-bearing relation	DOI-bearing source or curated resource metadata	Provenance completeness	Claims become unauditable
Interpretation boundary	Constraint defining what a phytochemical or bioactivity assertion can support	Permitted claim type, excluded claim type, evidence level, curator note	*bounded_by_claim*	Evidence-level assignment	Curator and rationale	Boundary confidence	Bioactivity or class membership may be overstated as efficacy

Targets, pathways, and formulation entities

Target representation in a natural product therapeutic ontology should distinguish target identity from target confidence and disease context. Disease–gene knowledge platforms show that target interpretation depends on explicit disease or phenotype associations and evidence sources [16]. For natural product therapeutics, this means that a relation such as *acts_on_target* should not be asserted as a therapeutic mechanism unless the evidence level supports that interpretation. The ontology should allow weaker relationships such as *predicted_to_interact_with*, *experimentally_tested_against*, or *associated_with_target_context* when the available evidence is preliminary.

Pathway representation requires similar caution. Pathway resources can organize biological processes, reactions, pathway membership, and molecular events, supporting pathway-level interpretation in pharmacological research [17]. Curated pathway knowledgebases further show the value of expert-reviewed pathway hierarchies and explicit event relationships [18]. In the proposed ontology, pathway entities should include pathway identifier, pathway hierarchy, organism context, participating targets, direction of inferred effect where available, and evidence provenance. A pathway annotation may support a mechanism hypothesis, but it should not be treated as proof that a natural product produces a clinically meaningful pathway effect.

Target confidence and biosynthetic context should be represented as separate but linkable layers. Target–disease platforms demonstrate that target confidence depends on harmonized evidence streams and should remain distinct from therapeutic validation [19]. Biosynthetic gene-cluster annotation resources show that experimentally supported biosynthetic context can connect natural products to source biology and pathway-like production logic [20]. For this ontology, target confidence should qualify molecular-target assertions, whereas biosynthetic context should qualify source and production relationships. Neither should be used alone to establish clinical mechanism or therapeutic effectiveness.

Formulation and preparation entities are necessary because natural product evidence often concerns mixtures, extracts,

or traditional preparations rather than single isolated compounds. Traditional medicine mapping resources show that herbs, ingredients, targets, and symptoms can be connected through integrative semantic representations, but these links require contextual interpretation boundaries [21]. Ingredient–target databases for traditional medicine also demonstrate the need to distinguish known from predicted ingredient–target relationships within complex preparations [22]. The proposed ontology should therefore include plant part, preparation method, extraction method, formulation type, ingredient composition, dosage-form representation without dose advice, and batch or preparation context where reported.

Table 2 maps ontology entities required for representing targets, pathways, mechanisms, formulations, and preparation context in natural product therapeutics.

Table 2. Ontology Entities for Targets, Pathways, Mechanistic Context, and Formulations: Target Confidence, Pathway Membership, Disease Context, Preparation Method, Extraction Context, Dosage Form Representation, Semantic Relationships, and Validation Boundaries

Entity type	Core purpose	Required attributes	Relationship types	Evidence source	Curation requirement	Interpretation boundary	Framework role
Target entity	Represent molecular targets linked to phytochemicals, extracts, or formulations	Target identifier, organism, isoform where relevant, target type	*acts_on_target*; *predicted_to_interact_with*; *tested_against*	Curated bioactivity, target database, experimental study, prediction source	Curator must distinguish known, inferred, and predicted target links	Target association does not equal therapeutic mechanism	Supports target-centered query and evidence organization
Target confidence	Qualify strength of target assertion	Evidence type, confidence category, source, curator note	*has_target_confidence*; *qualified_by_evidence*	Target–disease or bioactivity evidence	Expert review required for ambiguous evidence	Confidence is not clinical validation	Prevents overclaiming from weak target links
Pathway entity	Represent biological pathway or event context	Pathway identifier, hierarchy, organism, member targets	*participates_in_pathway*; *affects_pathway_hypothesis*	Curated pathway database or experimental study	Curator must specify pathway source and inference type	Pathway membership does not prove mechanism	Enables pathway-level reasoning with caution
Disease or phenotype context	Connect target, pathway, use, or evidence to a condition context	Disease term, phenotype term, symptom term, population where reported	*associated_with_condition*; *used_for_context*	Disease database, traditional-use source, clinical or experimental source	Mapping should preserve source terminology and normalized term	Disease association is not treatment efficacy	Anchors claims to a defined context
Mechanism hypothesis	Represent proposed mechanistic interpretation	Target, pathway, activity, evidence chain, directionality where supported	*hypothesized_to_modulate*; *supported_by_evidence*	Experimental, computational, pathway, or target evidence	High curation need to avoid causal overstatement	Hypothesis requires validation before mechanism claim	Connects evidence without asserting proof
Botanical taxon	Represent plant source identity	Genus, species, authority where available, synonym status	*source_taxon_of*; *has_part*	Taxonomic source or study report	Synonym and ambiguity review required	Taxon identity does not define formulation equivalence	Supports source traceability
Plant part	Represent anatomical source material	Leaf, root, bark, seed, flower, whole plant, other reported material	*part_of_taxon*; *used_in_preparation*	Study protocol, ethnopharmacological source, database record	Curator should preserve original reported term	Evidence from one part should not automatically transfer to another	Preserves preparation specificity

Preparation method	Represent processing before extraction or formulation	Drying, decoction, fermentation, grinding, heating, other method	*prepared_by*; *precedes_extraction*	Study method or traditional-use source	Method granularity should be recorded when available	Preparation context does not prove activity	Captures material transformation context
Extraction context	Represent extraction procedure for experimental or formulation evidence	Solvent, fraction, temperature, duration, extraction type where reported	*extracted_by*; *yields_extract*	Experimental method or product characterization source	Missing extraction details should be marked as uncertainty	Extracts are not interchangeable without comparable context	Supports evidence comparability
Formulation entity	Represent mixtures, extracts, products, or dosage-form descriptions without dose advice	Ingredients, ratios where reported, dosage form, excipients where relevant	*has_constituent*; *has_formulation*; *co_administered_with*	Product description, study protocol, traditional medicine database	Composition and evidence type should be curated	Formulation annotation does not imply therapeutic equivalence	Connects preparation, composition, and evidence

Evidence levels and semantic relationships

Evidence representation is the central safeguard against overinterpretation in a natural product therapeutic ontology. Evidence ontology work, provenance-aware knowledge representation, and nanopublication-style assertion modeling support the separation of evidence type, assertion, provenance, and claim context [23–25]. In the proposed ontology, every evidence-bearing relationship should connect an assertion to evidence level, evidence source, curator status, uncertainty field, and claim-support boundary. This design helps distinguish a computational prediction, assay observation, traditional-use statement, safety signal, and clinical finding.

Safety and supplement evidence require particularly careful semantic treatment. Integrated dietary supplement knowledge-base work shows that supplement entities, conditions, ingredients, evidence statements, and interaction or safety information can be represented through curated biomedical informatics structures [26]. For natural product therapeutics, this supports ontology classes for safety evidence, toxicity evidence, adverse-event context, interaction evidence, exposure context, and evidence provenance. A safety relation should not be reduced to a binary label; it should preserve exposure conditions, population context where available, evidence source, uncertainty, and curator interpretation.

Clinical natural product–drug interaction evidence must be separated from predicted or preclinical interaction evidence. Clinical pharmacokinetic natural product–drug interaction study-design recommendations show that evidence interpretation depends on product characterization, exposure conditions, study design, and pharmacokinetic endpoints [27]. Within the ontology, a clinical interaction assertion should therefore include the natural product entity or formulation, interacting drug, ADMET process, pharmacokinetic parameter if reported, study type, participant context where available, and provenance. This permits clinical-context representation without implying broad contraindication or regulatory acceptance.

ADMET and real-world co-use evidence should support safety-aware interpretation while remaining bounded. Recommended pharmacokinetic natural product–drug interaction research approaches emphasize the need to capture ADMET mechanisms, product characterization, and study-design details for interaction evidence [28]. Evidence on concurrent use of prescription drugs and herbal medicinal products also supports the need to represent co-use and safety context, but such evidence should not be treated as causality without appropriate assessment [29]. The ontology should therefore include semantic relationships such as *co_used_with*, *has_interaction_evidence*, *affects_pharmacokinetics*, *has_safety_signal*, *supported_by_evidence*, *has_uncertainty*, and *bounded_by_claim*.

Table 3 presents evidence-level categories and semantic relationship types for ontology-based representation of natural product therapeutic claims.

Table 3. Evidence Levels and Semantic Relationships for Natural Product Therapeutics: Computational Prediction, Experimental Evidence, Traditional Use, Safety Evidence, Clinical Evidence, Provenance, Uncertainty, Claim Boundaries, and Expert Curation

Evidence level or relationship type	Definition	Supported claim type	Insufficient-alone condition	Required provenance	Uncertainty field	Expert curation need	Interpretation boundary
Computational prediction	Algorithmically inferred compound, target, activity, or class relation	Hypothesis generation or prioritization	Clinical efficacy, safety, or mechanism proof	Algorithm, input data, version, source	Applicability domain; confidence	High	Prediction only
Network-inferred evidence	Relationship inferred from network structure or graph connectivity	Mechanistic plausibility or prioritization	Direct biological causality	Network source, edge provenance, inference method	Network incompleteness; bias	High	Association, not proof
Molecular docking evidence where relevant	Computational estimate of possible ligand–target binding mode	Possible binding hypothesis	Target engagement, in vivo effect, clinical benefit	Protein structure source, ligand preparation, software, score	Scoring and model uncertainty	High	Suggestive binding context only
In vitro bioactivity evidence	Activity measured in biochemical or cell-based assays	Assay-specific activity claim	Human efficacy or safety	Assay type, endpoint, protocol, source	Reproducibility; endpoint ambiguity	Medium to high	Context-specific biological activity
Preclinical evidence	Evidence from non-human biological models	Pharmacological plausibility or toxicity signal	Human therapeutic benefit	Model, species, intervention, endpoint, source	Translational uncertainty	High	Model-limited inference
Traditional use evidence	Documented historical or ethnopharmacological use context	Use-context claim	Clinical efficacy, safety, or mechanism	Tradition, geography where reported, preparation, source	Translation and context uncertainty	High	Traditional-use statement only
Safety evidence	Toxicity, adverse event, contraindication, tolerability, or exposure-risk evidence	Safety signal or risk-context claim	Universal safety or universal harm	Exposure, population, evidence source, method	Signal strength; confounding	High	Risk depends on context
Pharmacovigilance evidence	Real-world safety or adverse-event signal evidence	Signal detection or monitoring claim	Causality without assessment	Reporting source, coding, date, case context	Reporting bias	High	Signal, not definitive causation
Clinical evidence where available	Human study, clinical pharmacokinetic evidence, or clinical observation	Human-context effect, exposure, or interaction claim	General approval or guideline endorsement	Study design, population, intervention, endpoint, source	Bias; generalizability	High	Study-specific interpretation
Regulatory-relevant evidence where supported	Evidence that may inform claim-substantiation boundaries	Regulatory-style evidence organization	Regulatory acceptance or approval	Evidence package, jurisdiction if explicitly	Completeness and applicability uncertainty	Very high	Does not imply acceptance

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

				relevant, source			
has_constituent	Links formulation, extract, or preparation to ingredient or compound	Composition representation	Biological activity or efficacy	Composition source	Composition uncertainty	Medium	Composition only
derived_from	Links compound, extract, or formulation to biological or material source	Source traceability	Therapeutic equivalence	Source record or study method	Taxonomic/source ambiguity	Medium	Origin only
has_activity	Links entity to bioactivity class	Assay-supported activity	Therapeutic effect	Assay source and endpoint	Assay uncertainty	High	Activity within evidence context
acts_on_target	Links entity to target	Target interaction or modulation claim	Clinical mechanism without validation	Target evidence source	Target confidence	High	Target-level relation only
participates_in_pathway	Links target or process to pathway	Pathway context	Confirmed mechanism	Pathway source	Pathway inference uncertainty	High	Pathway membership or hypothesis
associated_with_condition	Links entity, target, use, or evidence to disease or phenotype	Contextual association	Treatment claim	Disease or phenotype source	Mapping uncertainty	High	Association only
has_formulation	Links preparation or product to formulation entity	Formulation representation	Therapeutic equivalence	Formulation source	Composition or preparation uncertainty	High	Formulation description only
supported_by_evidence	Links assertion to evidence entity	Evidence traceability	Claim beyond evidence type	Evidence source and curator	Evidence quality uncertainty	High	Support is evidence-level bound
has_uncertainty	Links assertion to uncertainty qualifier	Interpretation caution	Resolution of uncertainty	Curator note and reason	Explicit uncertainty category	High	Marks limitation
bounded_by_claim	Links assertion to permitted claim boundary	Responsible interpretation	Broader clinical or regulatory claim	Curator rationale	Boundary confidence	Very high	Prevents overclaiming

Proposed ontology structure

The proposed ontology structure is organized around a central *Natural Product Therapeutic Entity* that can represent a compound, extract, preparation, formulation, or evidence-linked therapeutic object depending on the level of granularity available. Because ontology and knowledge graph resources evolve over time, the structure should include explicit versioning, auditability, and change-management fields. A change language for ontologies and knowledge graphs supports the need for transparent additions, deletions, replacements, and revisions in curated semantic resources [30]. In this framework, versioning and auditability are not optional technical features; they are part of the ontology's evidence-traceability logic.

The ontology should also include a curation and quality-control module. Automated ontology workflow tools show that ontology development can be supported by repeatable release processes, validation checks, and structured maintenance practices [31]. For the proposed natural product therapeutic ontology, such workflow logic would support consistency checks for missing provenance, undefined relationships, synonym conflicts, circular class hierarchies, unqualified evidence assertions, and unsupported claim boundaries. These checks would support ontology maintenance, but they would not constitute pharmacological, clinical, or regulatory validation.

Relationship design should draw on general biomedical knowledge graph schemas while preserving natural-product specificity. Biomedical knowledge graph schema work provides reusable entity and relationship patterns for translational science [32]. In the proposed ontology, relations such as *has_constituent*, *derived_from*, *has_activity*, *acts_on_target*, *participates_in_pathway*,

associated_with_condition, *has_formulation*, *supported_by_evidence*, *has_uncertainty*, and *bounded_by_claim* should be typed and constrained. Each relation should specify whether it is descriptive, experimental, inferred, predicted, traditional-use-based, safety-related, clinical-contextual, or claim-boundary-related.

Interoperability is a further design requirement. Biological knowledge graph exchange frameworks show that reusable construction and exchange workflows can support integration across biomedical resources [33]. A natural product therapeutic ontology should therefore be designed for mapping to external chemical identifiers, taxonomic resources, disease and phenotype vocabularies, target resources, pathway resources, assay resources, evidence ontologies, and provenance models where appropriate. Such interoperability should enable structured queries, evidence tracing, and computational reasoning, while preserving uncertainty and avoiding unsupported causal claims.

Sustainability and FAIR-oriented assessment should be part of the proposed ontology's governance model. Guidance on sustainable curated scientific resources emphasizes open data, open code, and open infrastructure as important conditions for long-term reuse [34]. FAIR-checking approaches using knowledge graphs and semantic web standards further support the value of machine-actionable validation for findability and reuse [35]. In this framework, sustainability and FAIR-oriented checking support transparent maintenance and reuse; they do not imply that the ontology has been implemented, validated, clinically accepted, or approved for regulatory use.

Table 4 presents the proposed ontology structure for natural product therapeutics.

Table 4. Proposed Ontology Structure for Natural Product Therapeutics: Phytochemicals, Biological Activities, Targets, Pathways, Formulations, Traditional Use, Safety Evidence, Evidence Levels, Semantic Relationships, Provenance, Uncertainty, and Claim-Support Boundaries

Ontology module	Core classes	Required attributes	Primary relationships	Evidence-level links	Provenance requirement	Reasoning or query use	Validation or curation need	Failure risk
Phytochemical identity	Phytochemical entity, chemical class, compound identifier, synonym set, stereochemical qualifier	Preferred name, structure, identifiers, synonyms, stereochemistry, class	*has_chemical_class*; *has_identifier*; *has_synonym*	Identity evidence, classification evidence	Database or publication source, curator, version	Retrieve compounds by class, source, identifier, synonym	Synonym resolution and structure review	Misidentification or inappropriate record merging
Natural product source	Natural product source, botanical taxon, microbial source, plant part	Taxon, organism, strain where relevant, plant part, source material	*derived_from*; *found_in*; *has_part*	Source evidence	Taxonomic/source record, study source	Trace evidence to biological origin	Taxonomic and source curation	False transfer of evidence across sources
Biological activity	Bioactivity class, assay context, endpoint, activity assertion	Activity term, assay type, endpoint, model, result qualifier	*has_activity*; *measured_in_assay*; *supported_by_evidence*	In vitro, preclinical, curated bioactivity evidence	Assay record, publication, curator	Query assay-specific activities	Assay-context review	Assay signal overstated as efficacy
Targets and pathways	Target entity, target confidence, pathway entity, disease or phenotype context, mechanism hypothesis	Target ID, pathway ID, disease term, confidence qualifier, directionality if supported	*acts_on_target*; *participates_in_pathway*; *associated_with_condition*	Target evidence, pathway evidence, disease-context evidence	Target/pathway source, evidence source, curator	Mechanistic hypothesis exploration	Evidence-level and causality review	Pathway or target annotation misread as mechanism proof

Formulation and preparation	Formulation entity, plant part, preparation method, extraction method, dosage-form representation	Ingredients, preparation, extraction, formulation type, dosage form without dose advice	*has_constituent*; *prepared_by*; *extracted_by*; *has_formulation*	Formulation evidence, preparation evidence	Product or study source, preparation protocol	Compare evidence by preparation context	Composition and method curation	Non-equivalent preparations treated as equivalent
Traditional use context	Traditional use evidence, use context, cultural or practice context where reported	Reported use, preparation, source context, indication term	*traditionally_used_for*; *associated_with_condition*	Traditional-use evidence	Ethnopharmacological or traditional medicine source	Preserve use-context evidence	High expert curation need	Traditional use converted into efficacy claim
Safety and interaction evidence	Safety evidence, toxicity evidence, ADMET evidence, interaction evidence, pharmacovigilance evidence	Exposure, interacting agent, adverse event, ADMET process, population where available	*has_safety_signal*; *co_used_with*; *has_interaction_evidence*	Safety, pharmacovigilance, clinical, ADMET evidence	Study, repository, report, curator	Safety filtering and interaction evidence tracing	High expert review required	Safety signals ignored or overgeneralized
Evidence levels	Computational prediction, network inference, docking where relevant, in vitro, preclinical, traditional use, safety, clinical evidence where available	Evidence type, source, quality qualifier, evidence context	*supports_claim*; *qualified_by_evidence*	All evidence categories	Evidence source, curator, version	Filter claims by evidence level	Evidence classification review	Evidence types treated as equivalent
Provenance and curation	Evidence provenance, expert curation status, versioning, audit trail, quality-control status	Source, curator, date, version, change record, review status	*asserted_by*; *curated_by*; *has_version*; *changed_by*	Applies to every assertion	Required for all evidence-bearing relations	Audit claim history and source traceability	Ongoing curation and release checks	Claims become unauditable
Uncertainty and claim boundaries	Uncertainty field, claim-support boundary, interpretation qualifier	Uncertainty type, claim type, permitted wording, excluded inference	*has_uncertainty*; *bounded_by_claim*	Evidence-level dependent	Curator note and rationale	Prevent unsupported therapeutic interpretation	Very high expert curation need	Ontology enables overconfident claims

Figure 1 illustrates the proposed ontology structure for natural product therapeutics, connecting phytochemicals, biological activities, targets, pathways, formulations, evidence levels, provenance, uncertainty, and claim-support boundaries.

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

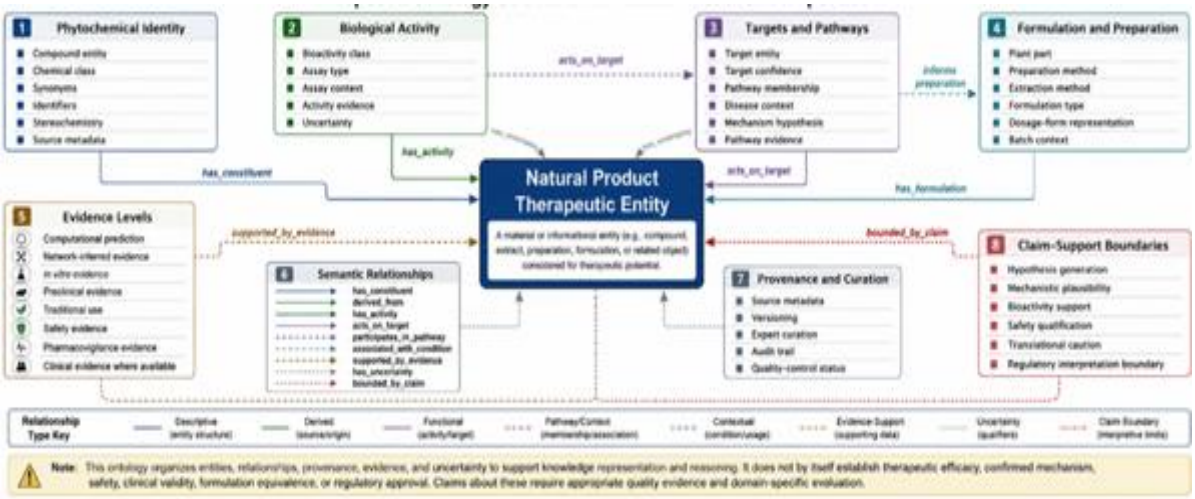

Figure 1. Proposed Ontology Structure for Natural Product Therapeutics

Alt-text description

A conceptual ontology diagram showing phytochemicals, bioactivities, targets, pathways, formulations, traditional use, safety evidence, experimental evidence, evidence levels, provenance, uncertainty, semantic relationships, expert curation, and claim boundaries as connected ontology modules.

CONCLUSION

This article proposed a conceptual ontology structure for natural product therapeutics that standardizes phytochemicals, biological activities, targets, pathways, formulations, preparation contexts, traditional use, safety evidence, evidence levels, semantic relationships, provenance, uncertainty, expert curation, auditability, and claim-support boundaries. The framework emphasizes that ontology-based standardization can improve representation, semantic interoperability, evidence tracing, query logic, and responsible computational reasoning, but it cannot by itself establish biological causality, therapeutic efficacy, safety, formulation equivalence, clinical usefulness, regulatory acceptance, or authority approval. Future development of such an ontology would require transparent curation, alignment with relevant biomedical standards, iterative validation, domain-expert review, and cautious interpretation before any therapeutic or regulatory claims are made.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Vogt L, Strömert P, Matentzoglu N, Karam N, Konrad M, Prinz M, et al. Suggestions for extending the FAIR Principles based on a linguistic perspective on semantic interoperability. Sci Data. 2025;12(1):688. doi:10.1038/s41597-025-05011-x

[2] Jackson R, Matentzoglu N, Overton JA, Vita R, Balhoff JP, Buttigieg PL, et al. OBO Foundry in 2021: Operationalizing open data principles to evaluate ontologies. Database (Oxford). 2021;2021. doi:10.1093/database/baab069

[3] Sorokina M, Steinbeck C. Review on natural products databases: Where to find data in 2020. J Cheminform. 2020;12(1):20. doi:10.1186/s13321-020-00424-9

[4] Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen EL, Gaudry A, et al. The LOTUS initiative for open knowledge management in natural products research. Elife. 2022;11. doi:10.7554/eLife.70780

[5] Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, Good BM, Griffith M, Griffith OL, et al. Wikidata as a knowledge graph for the life sciences. Elife. 2020;9. doi:10.7554/eLife.52614

[6] Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: Collection of open natural products database. J Cheminform. 2021;13(1):2. doi:10.1186/s13321-020-00478-9

[7] van Santen JA, Poynton EF, Iskakova D, McMann E, Alsup TA, Clark TN, et al. The Natural Products Atlas 2.0: A database of microbially-derived natural products. Nucleic Acids Res. 2022;50(D1). doi:10.1093/nar/gkab941

[8] 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

[9] Birer-Williams C, Gufford BT, Chou E, Alilio M, VanAlstine S, Morley RE, et al. A new data repository for pharmacokinetic natural product-drug interactions: From chemical characterization to clinical studies. Drug Metab Dispos. 2020;48(10):1104-12. doi:10.1124/dmd.120.000054

[10] Kim HW, Wang M, Leber CA, Nothias LF, Reher R, Kang KB, et al. NPClassifier: A deep neural network-based structural classification tool for natural products. J Nat Prod. 2021;84(11):2795-807. doi:10.1021/acs.jnatprod.1c00399

[11] Dührkop K, Nothias LF, Fleischauer M, Reher R, Ludwig M, Hoffmann MA, et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nat Biotechnol. 2021;39(4):462-71. doi:10.1038/s41587-020-0740-8

[12] Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, et al. SIRIUS 4: A rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods. 2019;16(4):299-302. doi:10.1038/s41592-019-0344-8

[13] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: New data content and improved web interfaces. Nucleic Acids Res. 2021;49(D1). doi:10.1093/nar/gkaa971

[14] Wang Y, Bryant SH, Cheng T, Wang J, Gindulyte A, Shoemaker BA, et al. PubChem BioAssay: 2017 update. Nucleic Acids Res. 2017;45(D1). doi:10.1093/nar/gkw1118

[15] Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL Database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024;52(D1). doi:10.1093/nar/gkad1004

[16] 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

[17] Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: Integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1). doi:10.1093/nar/gkaa970

[18] Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The Reactome Pathway Knowledgebase 2024. Nucleic Acids Res. 2024;52(D1). doi:10.1093/nar/gkad1025

[19] Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, et al. The next-generation Open Targets Platform: Reimagined, redesigned, rebuilt. Nucleic Acids Res. 2023;51(D1). doi:10.1093/nar/gkac1046

[20] Terlouw BR, Blin K, Navarro-Muñoz JC, Avalon NE, Chevrette MG, Egbert S, et al. MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res. 2023;51(D1). doi:10.1093/nar/gkac1049

[21] Wu Y, Zhang F, Yang K, Fang S, Bu D, Li H, et al. SymMap: An integrative database of traditional Chinese medicine enhanced by symptom mapping. Nucleic Acids Res. 2019;47(D1). doi:10.1093/nar/gky1021

[22] Kong X, Liu C, Zhang Z, Cheng M, Mei Z, Li X, et al. BATMAN-TCM 2.0: An enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins. Nucleic Acids Res. 2024;52(D1). doi:10.1093/nar/gkad926

[23] Nadendla S, Jackson R, Munro J, Quaglia F, Mészáros B, Olley D, et al. ECO: The Evidence and Conclusion Ontology, an update for 2022. Nucleic Acids Res. 2022;50(D1). doi:10.1093/nar/gkab1025

[24] Sikos LF, Philp D. Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs. Data Sci Eng. 2020;5(3):293-316. doi:10.1007/s41019-020-00118-0

[25] Kuhn T, Taelman R, Emonet V, Antonatos H, Soiland-Reyes S, Dumontier M. Semantic micro-contributions with decentralized nanopublication services. PeerJ Comput Sci. 2021;7. doi:10.7717/peerj-cs.387

[26] Rizvi RF, Vasilakes J, Adam TJ, Melton GB, Bishop JR, Bian J, et al. iDISK: The integrated DIetary Supplements Knowledge base. J Am Med Inform Assoc. 2020;27(4):539-48. doi:10.1093/jamia/ocz216

[27] Cox EJ, Rettie AE, Unadkat JD, Thummel KE, McCune JS, Paine MF. Adapting regulatory drug-drug interaction guidance to design clinical pharmacokinetic natural product-drug interaction studies: A NaPDI Center recommended approach. Clin Transl Sci. 2022;15(2):322-9. doi:10.1111/cts.13172

[28] Paine MF, Shen DD, McCune JS. Recommended approaches for pharmacokinetic natural product-drug interaction research: A NaPDI Center commentary. Drug Metab Dispos. 2018;46(7):1041-5. doi:10.1124/dmd.117.079962

[29] Agbabiaka TB, Wider B, Watson LK, Goodman C. Concurrent use of prescription drugs and herbal medicinal products in older adults: A systematic review. Drugs Aging. 2017;34(12):891-905. doi:10.1007/s40266-017-0501-7

[30] Hegde H, Vendetti J, Goutte-Gattat D, Caufield JH, Graybeal JB, Harris NL, et al. A change language for

ontologies and knowledge graphs. Database (Oxford). 2025;2025. doi:10.1093/database/baae133

[31] Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. ROBOT: A tool for automating ontology workflows. BMC Bioinformatics. 2019;20(1):407. doi:10.1186/s12859-019-3002-3

[32] Unni DR, Moxon SAT, Bada M, Brush M, Bruskiewich R, Clemons P, et al. Biolink Model: A universal schema for knowledge graphs in clinical, biomedical, and translational science. Clin Transl Sci. 2022;15(8):1848-55. doi:10.1111/cts.13302

[33] 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

[34] Hoyt CT, Gyori BM. The O3 guidelines: Open data, open code, and open infrastructure for sustainable curated scientific resources. Sci Data. 2024;11(1):547. doi:10.1038/s41597-024-03406-w

[35] Gaignard A, Rosnet T, de Lamotte F, Lefort V, Devignes MD. FAIR-Checker: Supporting digital resource findability and reuse with knowledge graphs and Semantic Web standards. J Biomed Semantics. 2023;14(1):7. doi:10.1186/s13326-023-00289-5