Multimodal AI and Multi-Omics in Natural Product Therapeutics: A Bibliometric Review of Research Frontiers, Data Modalities, and Translational Opportunities

Nora Schneider[1]*, Tobias Frank[1], Andreas Müller[2]

[1]Department of Artificial Intelligence for Cannabinoid and Terpene Profiling, Faculty of Pharmacy, University of Freiburg, Freiburg, Germany.
[2]Department of Computational Pharmacology for Pain Management, Faculty of Pharmacy, Heidelberg University, Heidelberg, Germany.

ABSTRACT

Multimodal artificial intelligence and multi-omics approaches are increasingly used to connect natural-product chemistry with biosynthetic pathways, molecular targets, biological responses, safety signals, and translational evidence, yet the intellectual structure and maturity of this research landscape remain incompletely characterized. This bibliometric review maps research frontiers, knowledge clusters, dominant data modalities, computational approaches, translational opportunity domains, and evidence gaps in multimodal AI and multi-omics research relevant to natural product therapeutics. A bounded pilot corpus was constructed through public bibliographic discovery across PubMed/PMC and DOI-indexed publisher and Crossref records, using OpenAlex-compatible metadata fields. The search covered peer-reviewed English-language journal articles published from 1 January 2017 to 11 July 2026. Forty-four candidate records underwent DOI-based deduplication and eligibility screening; one duplicate and four ineligible records were removed, producing a final corpus of 39 articles. Analyses included annual publication distribution, source distribution, reviewer-coded thematic clustering, data-modality mapping, AI-method mapping, translational-opportunity mapping, and evidence-boundary assessment. The corpus included 34 journal sources and peaked in 2022 with 11 publications; *Natural Product Reports* and *Briefings in Bioinformatics* contributed three articles each, while *Pharmaceuticals* contributed two. Four reviewer-coded clusters were identified: bibliometric methods and open metadata infrastructure, comprising eight records; natural-product omics, databases, and network knowledge, comprising nine; AI-enabled multimodal and multi-omics integration, comprising 14; and translational validation, bibliometric synthesis, and research governance, comprising eight. Dominant modalities included metabolomics, genomics, transcriptomics, chemical structures and spectra, network data, knowledge graphs, and microbiome data. Translational opportunity themes included mechanism discovery, target prediction, biomarker discovery, safety assessment, phytopharmaceutical standardization, and evidence integration. Bibliometric patterns can identify research fronts and support agenda setting, but they do not establish methodological quality, biological validity, clinical utility, therapeutic effectiveness, safety, or regulatory readiness.

Key Words: *Multimodal AI, Multi-omics, Natural product therapeutics, Bibliometric review, Phytopharmacology, Data modalities*

eIJPPR 2026; 16(1):1-18

HOW TO CITE THIS ARTICLE: Schneider N, Frank T, Müller A. Multimodal AI and Multi-Omics in Natural Product Therapeutics: A Bibliometric Review of Research Frontiers, Data Modalities, and Translational Opportunities. Int J Pharm Phytopharmacol Res. 2026;16(1):1-18. https://doi.org/10.51847/oZ8hEKjcdb

INTRODUCTION

Natural-product therapeutics increasingly sit at the intersection of computational chemistry, artificial

Corresponding author: Nora Schneider
Address: Department of Artificial Intelligence for Cannabinoid and Terpene Profiling, Faculty of Pharmacy, University of Freiburg, Freiburg, Germany.
E-mail: ✉ nora.schneider@pharmazie.uni-freiburg.de
Received: 12 October 2025; **Revised:** 09 January 2026; **Accepted:** 11 January 2026

intelligence, biosynthetic analysis, and multi-omics investigation. Machine-learning methods can support compound prioritization, molecular representation, activity prediction, and exploration of natural-product chemical space, while computational approaches to microbial natural products extend from genome mining to candidate identification. Multi-omics integration adds a complementary capacity to connect chemical constituents with molecular pathways, biological targets, and experimentally observed responses [1-3]. These developments have widened the analytical scope of phytopharmacology, although they have not removed the need for chemical authentication, biological validation, safety assessment, and reproducible experimental design.

The convergence is particularly visible in studies combining network pharmacology, systems-level pathway analysis, omics integration, and AI-assisted interpretation. Such work attempts to represent multicomponent natural products as interconnected systems rather than as isolated molecules, thereby creating computational links among compounds, targets, pathways, diseases, and phenotypic outcomes [4]. This systems orientation is potentially valuable for natural products because therapeutic preparations may contain chemically diverse constituents with context-dependent, synergistic, antagonistic, or non-specific biological effects. However, increasing model complexity may also amplify database bias, identifier inconsistency, unsupported target predictions, and overinterpretation of pathway-enrichment results.

A bibliometric perspective is useful because the field spans natural-products chemistry, pharmacology, bioinformatics, systems biology, artificial intelligence, microbiome research, and information science. Mapping publication patterns and thematic concentrations can reveal which combinations of modalities and computational methods are receiving sustained attention, which infrastructures support knowledge integration, and which translational questions remain comparatively underrepresented. Bibliometric visibility, nevertheless, measures patterns of scholarly production and connection rather than the validity of the biological claims contained within the mapped literature. A highly visible method or thematic cluster may therefore reflect intellectual momentum without demonstrating comparative performance or therapeutic usefulness.

The aim of this original bibliometric review is to map the scientific landscape of multimodal AI and multi-omics in natural product therapeutics through a transparent bounded corpus and a predefined analytical framework. The review distinguishes bibliometric indicators from biological evidence and organizes the literature around research frontiers, data modalities, computational integration methods, translational opportunities, and validation gaps. It also examines how network pharmacology, knowledge graphs, machine learning, deep learning, chemical fingerprinting, microbiomics, and other data layers contribute to the emerging evidence architecture [4]. The intended contribution is agenda setting rather than evidence confirmation: the review identifies where research activity is concentrating and what methodological, infrastructural, and translational conditions would be required for the field to advance responsibly.

Bibliometric review rationale

Bibliometric analysis offers a structured way to examine the development, composition, and intellectual organization of an interdisciplinary research field. It can combine descriptive performance indicators with science-mapping approaches that examine relationships among publications, sources, topics, and citation structures. Established bibliometric frameworks emphasize that data acquisition, cleaning, analysis, visualization, and interpretation must be specified before results are reported, while comparative reviews of bibliometric software show that different tools support different combinations of co-occurrence, citation, coupling, clustering, and visualization functions [5-7]. These principles are especially important for a field in which terminology varies across natural-products research, phytomedicine, systems pharmacology, omics science, and artificial intelligence.

A bibliometric review should therefore be treated as a sequence of explicit methodological decisions rather than as an automated production of maps. Database selection affects journal coverage, document types, affiliation data, abstracts, keywords, cited references, and export formats; query construction determines the balance between conceptual breadth and topical precision; and deduplication influences the integrity of all subsequent counts. A transparent workflow should also define which indicators are available, which analyses can be performed reliably, and which desired outputs cannot be generated from the extracted metadata [8]. This is particularly relevant when public discovery interfaces are used instead of a complete subscription-database export.

Science mapping can include keyword co-occurrence, thematic clustering, co-citation analysis, and bibliographic coupling, but these techniques describe different relationships. Keyword co-occurrence identifies terms that appear together and may indicate conceptual proximity, whereas co-citation connects works that are jointly cited and may reveal an established intellectual base. Bibliographic coupling links publications that share references and is often used to examine more recent research fronts. The availability and reliability of each method depend on the completeness and normalization of

keyword and citation metadata, and different software environments may produce non-equivalent networks or cluster assignments [7].

For multimodal AI and multi-omics in natural product therapeutics, bibliometric analysis is justified by both fragmentation and convergence. The field is fragmented because relevant studies are distributed across distinct disciplinary journals and use inconsistent labels for similar concepts, including data fusion, integrative omics, systems pharmacology, network medicine, knowledge graphs, and multimodal learning. It is convergent because these approaches increasingly address common questions concerning compound identification, mechanism discovery, target prediction, pathway interpretation, biomarker development, safety assessment, and translation. The present review consequently uses bibliometric reasoning to organize the landscape while maintaining the interpretive boundary that publication frequency, thematic clustering, and citation visibility do not demonstrate scientific quality or therapeutic validity [8].

Dataset construction and analytical strategy

The bibliometric dataset was constructed on 11 July 2026 using public bibliographic discovery across PubMed/PMC and DOI-indexed publisher and Crossref records, with OpenAlex-compatible metadata fields adopted as the organizational schema. The eligible publication period extended from 1 January 2017 to the search date, and the language filter was English. The search combined natural-product concepts—such as natural products, phytochemicals, phytomedicine, phytopharmaceuticals, herbal medicine, and traditional medicine—with multi-omics concepts and AI or computational-integration concepts. OpenAlex-compatible fields were selected because they support structured representation of work identifiers, DOIs, titles, publication years, sources, authorships, institutions, concepts, referenced works, and citation metadata where available [9].

The search and screening process produced a bounded pilot corpus rather than a claim of exhaustive global coverage. Forty-four candidate records were advanced to DOI and eligibility verification, and one duplicate was removed through normalized DOI matching and comparison of titles, authors, and publication status, leaving 43 unique records. Four records were then excluded: one retracted article, one preprint-only record, and two articles published outside the 2017–2026 eligibility period. The final corpus therefore comprised exactly 39 peer-reviewed journal articles with DOI information, including eight methodological bibliometric references and 31 topical or translational references. Because bibliographic discovery systems differ in disciplinary coverage and metadata completeness, these counts describe the bounded corpus used in this review and not the total worldwide literature [10].

The analytical strategy included annual publication distribution, normalized source-title distribution, reference-type distribution, reviewer-coded thematic clustering, data-modality mapping, AI-method mapping, translational-opportunity mapping, and citation-placement auditing. The corpus schema was designed to distinguish bibliographic metadata from thematic content and to preserve a clear record of which findings were calculated, extracted, or interpretively coded. API-based bibliometric systems can differ substantially in coverage, field availability, access conditions, and reproducibility, which makes explicit documentation of the analytical route essential [11]. **Figure 1** illustrates the bibliometric dataset construction and analytical workflow used to map multimodal AI and multi-omics research in natural product therapeutics.

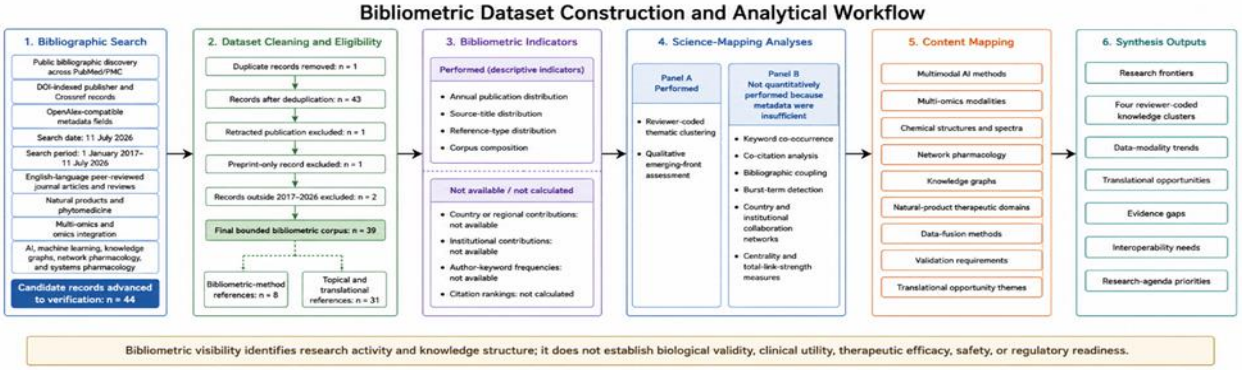

Figure 1. Bibliometric Dataset Construction and Analytical Workflow

Alt-text description

A left-to-right bibliometric workflow begins with public bibliographic discovery across PubMed/PMC and DOI-indexed publisher and Crossref records for the period 1 January 2017 to 11 July 2026. Forty-four candidate records enter dataset cleaning; one duplicate is removed, leaving

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

43 records. Four records are excluded during eligibility screening: one retracted article, one preprint-only record, and two articles outside the eligible date range. The final bounded corpus contains 39 articles. Subsequent zones show annual publication and source analyses, reviewer-coded thematic clustering, multimodal AI and multi-omics modality mapping, translational opportunity mapping, and research-agenda synthesis. Keyword co-occurrence, co-citation, bibliographic coupling, burst detection, and normalized country or institution networks are marked as not quantitatively performed because metadata were insufficient.

Use abstract icons only and do not use database logos, software logos, institutional logos, country flags, clinical images, patient data, product images, regulatory logos, or decorative botanical elements. Use a clean white background, strict geometric alignment, professional sans-serif typography, thin uniform borders and connectors, readable labels, generous white space, and a restrained colorblind-safe palette. Use equal-size analytical nodes unless node size is directly supported by the exact values above. Ensure legibility after journal-column reduction and suitability for 300 dpi export.

The review did not calculate author-keyword frequencies, keyword co-occurrence networks, total-link-strength values, modularity statistics, silhouette scores, co-citation communities, bibliographic-coupling communities, citation bursts, keyword bursts, collaboration networks, corpus h-index values, or centrality measures. These analyses require standardized keyword, affiliation, citation-link, and cited-reference fields that were not uniformly available for the bounded dataset. Search-system suitability depends not only on the apparent number of retrieved results but also on recall, precision, reproducibility, export functionality, and metadata quality [12]. Accordingly, unavailable network outputs are reported as not performed because metadata were insufficient, rather than replaced with inferred or visually estimated values.

The annual publication distribution was 2017, one article; 2018, two; 2019, two; 2020, two; 2021, four; 2022, 11; 2023, four; 2024, eight; 2025, five; and 2026 through 11 July, zero. The corpus contained 34 distinct journal sources. *Natural Product Reports* and *Briefings in Bioinformatics* each contributed three articles, *Pharmaceuticals* contributed two, and each of the remaining 31 journals contributed one. These values describe the purposefully constructed pilot corpus and must not be interpreted as global productivity rankings or prevalence estimates. **Table 1** summarises the dataset construction process, bibliometric indicators, analytical strategy, and interpretation boundaries used in this bibliometric review.

Table 1. Dataset Construction and Analytical Strategy for the Bibliometric Review: Database Source, Search Logic, Screening Counts, Final Corpus, Bibliometric Indicators, Network Analyses, Modality Mapping, and Interpretation Boundaries

Analytical component	Specification	Exact Part 1 result	Purpose	Interpretation boundary	Limitation
Database source	Public bibliographic discovery across PubMed/PMC and DOI-indexed publisher and Crossref records; OpenAlex-compatible metadata schema	Public multi-source discovery; no subscription-database census	Identify and verify eligible peer-reviewed records with DOI metadata	Represents the bounded review dataset, not the entire global literature	Stable database-wide raw-hit counts were not available from the interfaces used
Search date	Date on which eligibility and metadata were finalized	11 July 2026	Establish a reproducible temporal cut-off	Articles or metadata changes after this date are outside the dataset	Later indexing or corrections are not captured
Publication period	Eligible publication dates	1 January 2017–11 July 2026	Restrict the review to the predefined contemporary period	A 2026 issue date did not override the search-date cut-off	Earlier foundational literature was excluded unless represented through eligible methodological articles

Language and document filters	English-language peer-reviewed journal articles and reviews with DOI information	English; journal article or journal review; DOI mandatory	Improve metadata consistency and evidentiary traceability	Excludes potentially relevant non-English or non-DOI literature	Language and DOI restrictions may narrow disciplinary or regional representation
Search logic	Natural-product terms AND multi-omics terms AND AI or computational-integration terms, supplemented by focused combinations	Three principal concept blocks plus targeted searches	Capture interdisciplinary literature at the intersection of the review domains	Query breadth does not ensure complete recall across all disciplinary vocabularies	Terminological variation may produce missed records
Candidate records	Records advanced to DOI and eligibility verification	n = 44	Define the bounded candidate pool	Not a database-wide retrieval total	Candidate identification involved public discovery and manual verification
Deduplication	Normalized DOI matching with title, author, year, and publication-status checks	Duplicates removed: n = 1; records after deduplication: n = 43	Prevent duplicate counting of preprint and journal versions or repeated database records	Deduplication addresses record identity, not study-level conceptual overlap	Database-specific duplicate structures were not independently compared
Eligibility exclusions	Retraction, publication type, and date screening	Retracted publication: n = 1; preprint-only record: n = 1; outside period: n = 2	Retain peer-reviewed, eligible, non-retracted journal evidence	Exclusion counts apply only to the bounded candidate set	Other databases might identify additional ineligible or eligible records
Final corpus	DOI-verified records retained after screening	n = 39	Provide the analytical dataset and fixed manuscript reference set	Corpus size is not equivalent to the total volume of global research	The corpus was purposefully selected for bibliometric and narrative support
Corpus composition	Methodological versus topical or translational records	Bibliometric-method references: n = 8; topical and translational references: n = 31	Preserve methodological support while covering the review's scientific themes	Composition reflects review design rather than field-wide proportions	Purposeful balancing affects descriptive distributions
Annual publication distribution	Count of retained records by year	2017: 1; 2018: 2; 2019: 2; 2020: 2; 2021: 4; 2022: 11; 2023: 4; 2024: 8; 2025: 5; 2026 to 11 July: 0	Describe the temporal composition of the pilot corpus	Indicates the distribution of selected evidence, not global annual output	Selection strategy can influence the apparent peak year
Source distribution	Frequency of normalized journal titles	34 distinct sources; *Natural Product Reports*: 3; *Briefings in Bioinformatics*: 3; *Pharmaceuticals*: 2; 31 sources: 1 each	Characterize disciplinary dispersion	Not a worldwide journal-productivity ranking	Counts reflect the bounded corpus
Country or regional indicators	Contribution or collaboration based	Not available from extracted metadata	Would characterize geographical	No country ranking or	Affiliation fields were insufficiently normalized

	on normalized affiliation metadata		participation if reliable	collaboration pattern is reported	
Institutional indicators	Organization-level productivity or collaboration	Not available from extracted metadata	Would identify organizational participation where possible	No institution ranking is reported	Institutional naming and affiliation data were inconsistent
Keyword and network analyses	Keyword co-occurrence, co-citation, coupling, bursts, and network metrics	Not quantitatively performed because metadata were insufficient	Would support algorithmic science mapping if complete data were available	No network communities, frequencies, centrality scores, or burst terms are claimed	Standardized keywords, cited references, and citation links were incomplete
Thematic clustering	Reviewer-coded classification based on article scope, titles, abstracts, and evidence roles	Four clusters: 8, 9, 14, and 8 records	Organize the interdisciplinary evidence landscape	Clusters are thematic synthesis categories, not algorithmically detected communities	Manual coding is interpretive and may not reproduce software-derived clustering
Modality mapping	Mapping of omics, chemical, network, graph, phenotypic, and safety data	Dominant modalities: metabolomics; genomics and transcriptomics; chemical structures and spectra; network data; knowledge graphs; microbiome data	Identify recurring evidence types and analytical roles	Modality prominence is qualitative rather than frequency-ranked	Exact modality frequencies were not calculated
AI-method mapping	Mapping of computational approaches	Machine learning; deep learning; network-based integration; knowledge graphs; literature mining; generative representation learning	Characterize dominant and emerging computational families	Method visibility does not establish comparative accuracy or readiness	Heterogeneous task definitions prevent direct performance comparison
Translational-opportunity mapping	Synthesis of recurring application domains	Mechanism discovery; target prediction; biomarker discovery; safety and ADMET; standardization; evidence integration	Structure agenda-setting opportunities	Opportunities are not validated translational outcomes	Clinical and implementation evidence remained limited
Overall interpretation boundary	Rule governing all bibliometric and thematic findings	Research activity and knowledge structure are mapped; therapeutic validity is not inferred	Prevent overstatement	Bibliometric prominence does not prove efficacy, safety, clinical utility, or regulatory relevance	Requires separate experimental, clinical, and implementation evidence

Research frontiers and knowledge clusters

The bounded corpus was organized into four reviewer-coded thematic clusters rather than algorithmically derived network communities. Cluster 1, bibliometric methods and open metadata infrastructure, comprised eight records; Cluster 2, natural-product omics, databases, and network knowledge, comprised nine; Cluster 3, AI-enabled multimodal and multi-omics integration, comprised 14; and Cluster 4, translational validation, bibliometric synthesis, and research governance, comprised eight. The first cluster provides the methodological foundation for data acquisition, science mapping, metadata assessment,

and interpretation, whereas the second begins with the biological and informational infrastructures needed to investigate natural-product biosynthesis and therapeutic mechanisms. Omics approaches to secondary-metabolite biosynthesis illustrate how genomic, transcriptomic, and metabolomic evidence can be combined to study the production and regulation of natural compounds [13].

The natural-product omics, databases, and network-knowledge cluster is defined by the integration of biosynthetic information, molecular structures, spectra, organismal provenance, compound–target relationships, and pathway-level evidence. Natural-product databases are increasingly expected to operate within multi-omics environments rather than as isolated catalogues of structures, yet interoperability, identifier normalization, annotation quality, and provenance remain uneven [14]. Within the same cluster, integrated omics methods are used to prioritize candidate targets and biological processes associated with natural compounds or multicomponent preparations [15]. The translational opportunity is improved mechanism prioritization and standardization, while the principal evidence gap is the limited experimental confirmation of computationally inferred targets and pathways.

A related research frontier concerns plant metabolic gene clusters and network-pharmacology models. Multi-omics analysis of metabolic gene clusters can connect genomic organization with transcriptional regulation, enzymatic activity, and metabolite production, supporting the discovery or interpretation of specialized plant metabolites [16]. Network pharmacology, by contrast, connects compounds, targets, pathways, and disease concepts to generate systems-level hypotheses about medicinal plants and traditional formulations [17]. Its bibliometric visibility reflects sustained research attention, but dense interaction networks may be strongly shaped by database incompleteness, promiscuous target prediction, enrichment-background choices, and the tendency to treat predicted relations as mechanistic confirmation.

The remaining frontiers connect natural-product knowledge integration with AI-enabled multimodal analysis and translational governance. Computational literature-based discovery can identify relations dispersed across publications and can support entity linking, evidence retrieval, and hypothesis generation [18]. Open linked-data initiatives similarly seek to connect compounds, organisms, taxonomic information, references, and structural identifiers in reusable knowledge environments [19]. These infrastructures create opportunities for knowledge graphs and multimodal AI, but their outputs remain dependent on entity resolution, provenance, database coverage, and explicit validation. **Table 2** presents the major research frontiers and knowledge clusters identified in the bibliometric corpus.

Table 2. Research Frontiers and Knowledge Clusters in Multimodal AI and Multi-Omics for Natural Product Therapeutics: Cluster Keywords, Representative Evidence, Data Modalities, AI Methods, Therapeutic Focus, Translational Opportunities, and Evidence Gaps

Cluster label	Core keywords	Representative evidence	Dominant data modalities	Dominant AI or computational methods	Therapeutic focus	Translational opportunity	Evidence gap	Interpretation boundary
Bibliometric methods and open metadata infrastructure, n = 8	Bibliometrics; science mapping; APIs; OpenAlex; Crossref; retrieval; deduplication; reproducibility	Comprehensive science-mapping methodology and software infrastructure [5]	Bibliographic metadata, DOI records, publication years, journal sources, citation links, keywords, author and affiliation fields where available	Performance analysis, science mapping, API extraction, metadata normalization, deduplication	Methodological support for mapping natural-product AI and multi-omics research	Reproducible landscape analysis, updateable evidence surveillance, and transparent indicator reporting	Incomplete metadata, database-dependent coverage, inconsistent export fields, and limited reproducibility across platforms	Bibliometric capability does not guarantee complete retrieval or reliable biological interpretation
Natural-product omics,	Secondary metabolites;	Multi-omics investigatio	Genomics, transcriptomics,	Genome mining, pathway	Mechanism discovery, biosyntheti	Better target hypotheses,	Inconsistent identifiers,	Database or network association

databases, and network knowledge, n = 9	biosynthetic pathways; databases; network pharmacology; target discovery; open knowledge; interoperability	n of secondary-metabolite biosynthesis [13]	metabolomics, molecular structures, spectra, organismal provenance, compound–target and pathway data	enrichment, network pharmacology, literature mining, linked-data integration	c interpretation, target prioritization, natural-product standardization	interoperable databases, improved chemical and biological annotation	incomplete annotations, weak negative controls, limited experimental target confirmation	s remain hypotheses until chemically and biologically validated
AI-enabled multimodal and multi-omics integration, n = 14	Machine learning; deep learning; data fusion; knowledge graphs; multimodal AI; microbiome; representation learning	Natural-product AI and integrative multi-omics research introduced in the review's opening evidence base [1-3]	Molecular structures, chemical fingerprints, spectra, genomic and transcriptomic profiles, proteomic and metabolomic data, microbiome data, literature, graphs, and limited phenotypic information	Machine learning, deep learning, multilayer networks, knowledge graphs, representation learning, generative integration	Candidate prioritization, target prediction, mechanism modelling, bioactivity prediction, response stratification	Integrated discovery models and cross-modal evidence synthesis	Missing data, small datasets, batch effects, domain shift, information leakage, limited interpretability, and weak external validation	Computational sophistication does not establish comparative performance, causality, or clinical utility
Translational validation, bibliometric synthesis, and research governance, n = 8	Safety; ADMET; validation; interoperability; interpretability; bibliometric fronts; clinical translation; evidence standards	Critical assessment of methodological limitations in medicinal-plant network pharmacology [17]	Toxicity descriptors, bibliometric indicators, clinical or phenotypic data where available, pathway-informed omics, structured evidence and provenance records	In silico toxicity assessment, bibliometric synthesis, evidence mapping, interpretable or biologically structured modelling	Safety prioritization, standardization, readiness assessment, evidence integration, translation planning	Validation pipelines, responsible decision support, reproducible evidence standards, and implementation planning	Scarcity of prospective studies, clinical outcomes, regulatory-aligned evidence, calibrated models, and reproducible benchmarks	Translational opportunity is agenda setting; it does not demonstrate therapeutic effectiveness, safety, or regulatory readiness

Multimodal AI and multi-omics trends

Machine learning and network-based integration constitute the principal computational families through which heterogeneous omics layers are combined in the mapped literature. These approaches include early or intermediate feature concatenation, latent-factor modelling, similarity-based integration, supervised prediction, multilayer network construction, and representation learning. Their common analytical objective is to extract cross-modal patterns that cannot be observed reliably within a single data layer, although high dimensionality, missing values, batch effects, small sample sizes, and non-independent features remain persistent constraints [20-22]. In natural-product therapeutics, these limitations are compounded by

variable chemical composition, inconsistent product identity, incomplete negative examples, and differences among experimental systems.

Deep generative integration represents an emerging methodological frontier because it can learn compressed representations across heterogeneous omics datasets and potentially model complex nonlinear relationships. Such models may support cross-modal imputation, subtype discovery, molecular-response prediction, and the generation of integrated latent spaces, but their apparent flexibility does not eliminate risks related to overfitting, information leakage, unstable representations, or weak external transportability [23]. Foundation-model concepts and graph neural networks may eventually contribute to this frontier where sufficiently large, standardized, and biologically contextualized datasets exist. In the present corpus, however, these approaches are better interpreted as developing research directions than as established platforms for natural-product translation.

Knowledge graphs offer a complementary integration strategy by representing compounds, targets, pathways, diseases, organisms, assays, publications, and other biomedical entities as structured relations. Graph-based inference can facilitate candidate prioritization, evidence retrieval, repurposing hypotheses, and the identification of paths connecting natural products with therapeutic mechanisms [24]. The reliability of these outputs depends on the quality of source databases, entity resolution, relation provenance, temporal consistency, and protection against circular validation. Reviews of biomedical datasets further indicate that coverage gaps, schema differences, uncertain labels, and uneven curation can be inherited by downstream graph models [25].

The convergence of multimodal data and knowledge-graph infrastructure is therefore a notable research front in natural-product science. Chemical structures, mass-spectral features, genomic context, metabolomic profiles, bioactivity measurements, taxonomic provenance, textual evidence, and pathway information can potentially be linked within a shared computational environment [26]. This convergence may support more traceable hypothesis generation than isolated prediction models, particularly when evidence paths and provenance are retained. Nevertheless, multimodal integration should not be equated with multimodal validation: combining several uncertain data sources may increase apparent coherence without increasing biological truth. **Table 3** maps multimodal AI and multi-omics data modalities identified across the bibliometric landscape.

Table 3. Multimodal AI and Multi-Omics Data Modalities in Natural Product Therapeutics: Metabolomics, Transcriptomics, Proteomics, Genomics, Microbiomics, Chemical Fingerprints, Bioactivity Profiles, Knowledge Graphs, Network Pharmacology, Clinical Data, and Safety Signals

Data modality	Analytical role	AI or integration method	Bibliometric trend signal	Natural product therapeutic relevance	Typical output	Validation need	Interpretation boundary
Metabolomics	Characterizes chemical phenotypes, exposure profiles, and biological responses	Multivariate machine learning, latent-factor integration, pathway mapping, joint embeddings	Strong qualitative signal	Supports constituent identification, response profiling, metabolic-pathway interpretation, and product standardization	Metabolite signatures, discriminant features, pathway-enrichment outputs	Authentic standards, batch correction, independent cohorts, dose–time replication	Metabolite association does not establish causal mechanism or therapeutic efficacy
Transcriptomics	Measures gene-expression responses and pathway-level perturbations	Clustering, network inference, supervised learning, multiview integration	Strong qualitative signal	Links natural products or formulations to gene programs and disease-relevant processes	Differential-expression signatures, gene modules, enriched pathways	Cell-type resolution, experimental replication, perturbation and target-engagement studies	Expression changes may be indirect, non-specific, or context dependent

Proteomics	Assesses protein abundance, modification, interaction, and pathway state	Feature selection, network integration, multilayer learning	Moderate qualitative signal	Supports target-engagement assessment and protein-level mechanism interpretation	Protein panels, interaction modules, pathway-state profiles	Orthogonal protein assays, independent replication, functional experiments	Network centrality or abundance change does not establish therapeutic necessity
Genomics	Supports biosynthetic discovery, genetic-context analysis, and variant interpretation	Genome mining, sequence modelling, graph learning, comparative genomics	Strong qualitative signal	Identifies biosynthetic gene clusters, candidate enzymes, and genetic determinants of response	Candidate clusters, enzymes, variants, biosynthetic pathways	Functional expression, gene disruption, metabolite confirmation	A predicted gene cluster may not produce an expressed, detectable, or active natural product
Microbiomics	Examines host–microbe, product–microbe, and metabolism-related relationships	Compositional machine learning, multilayer networks, joint omics modelling	Moderate emerging signal	Relevant to herbal metabolism, bioavailability, response variability, and toxicity	Taxonomic profiles, microbial functions, host–microbiome signatures	Longitudinal sampling, dietary and medication control, functional validation	Microbiome associations are highly confounded and context dependent
Chemical fingerprinting	Supports authentication, batch comparison, dereplication, and structural representation	Chemometrics, similarity learning, spectral modelling, representation learning	Strong established signal	Central to natural-product identity, quality control, scaffold analysis, and consistency assessment	Spectral or chromatographic fingerprints, chemical clusters, similarity scores	Reference compounds, interlaboratory reproducibility, validated analytical procedures	Chemical similarity or batch consistency does not prove equivalent biological activity
Bioactivity profiling	Connects compounds, fractions, or extracts with functional assay outcomes	Supervised learning, multitask models, matrix factorization, phenotypic clustering	Moderate qualitative signal	Supports candidate prioritization, activity-spectrum analysis, and synergy investigation	Ranked candidates, activity matrices, phenotypic profiles	Dose-response testing, orthogonal assays, negative controls, independent replication	Screening activity is not equivalent to therapeutic efficacy or safety
Network pharmacology data	Integrates compounds, predicted targets, pathways, and disease concepts	Network propagation, enrichment analysis, link prediction, graph analysis	Strong visibility signal	Generates systems-level hypotheses for multicomponent natural products	Compound–target networks, pathway maps, ranked mechanisms	Curated inputs, null models, experimental target confirmation, sensitivity analysis	Dense networks may reproduce database bias and promiscuous target predictions

Knowledge graph data	Represents heterogeneous entities, relations, provenance, and literature evidence	Graph embeddings, link prediction, retrieval, reasoning, path analysis	Strong emerging signal	Connects compounds, organisms, targets, diseases, assays, pathways, and publications	Ranked relations, inferred links, evidence paths, integrated queries	Provenance auditing, benchmark questions, temporal validation, leakage control	An inferred graph edge is a hypothesis rather than verified biological evidence
Clinical or phenotypic data	Links molecular patterns with patient characteristics, responses, or observable phenotypes	Multimodal fusion, supervised learning, survival analysis, representation learning	Limited signal in the natural-product subset	May support responder stratification or outcome modelling where validated data exist	Risk estimates, response probabilities, phenotype-linked biomarkers	Prospective cohorts, external validation, calibration, confounder control	Patient-level utility cannot be inferred from preclinical or cross-sectional omics data
Imaging data	Captures spatial, morphological, histological, microscopic, or cellular phenotypes	Convolutional models, vision transformers, multimodal embeddings	Weak or underrepresented signal	Potentially supports spatial mechanism analysis, phenotypic screening, or plant and cellular characterization	Image features, morphology classes, spatial-response maps	Standardized acquisition, annotated external datasets, reproducible preprocessing	Imaging is an opportunity area rather than a dominant validated modality in the corpus
Safety and pharmacovigilance data	Supports toxicity prediction, adverse-event surveillance, and interaction detection	Quantitative structure–activity models, anomaly detection, natural-language processing, graph models	Weak-to-moderate signal	Relevant to early toxicity prioritization and post-use monitoring of natural products	Toxicity scores, interaction alerts, adverse-event signals	Exposure-relevant toxicology, causality assessment, reporting-quality controls	A computational safety signal neither confirms harm nor establishes absence of risk

Translational opportunities

Safety assessment and ADMET prediction represent immediate but bounded translational opportunities because computational methods can prioritize compounds or extracts that warrant closer toxicological examination. In silico profiling of natural-product libraries illustrates how chemical descriptors and predictive models can be used to identify potential toxicity liabilities before more resource-intensive testing [27]. These predictions are most useful as triage instruments rather than substitutes for toxicology, particularly because natural-product chemical space may differ from the synthetic compounds that dominate many training datasets. Applicability domains, exposure levels, mixture effects, metabolite formation, and uncertainty estimates must therefore be reported explicitly.

The broader drug-discovery literature indicates that AI translation is constrained by data quality, reproducibility, prospective evaluation, and the distance between retrospective model performance and decision-relevant utility [28]. For natural products, these constraints include uncertain taxonomy, variable extraction procedures, incomplete structural annotation, inconsistent assay conditions, and insufficiently characterized formulations. Curated natural-product databases may reduce some of these problems by organizing chemical structures, biological sources, activities, and references, but database scope and annotation quality vary [29]. Machine learning in microbial natural-product research nevertheless demonstrates opportunities spanning genome mining,

biosynthetic prioritization, structural prediction, dereplication, and activity assessment [30].

Patient stratification and biomarker discovery are longer-term opportunities rather than established natural-product applications. Clinical multi-omics integration can combine molecular layers with phenotypic and outcome data to identify patient subgroups or response-associated signatures, but reliable stratification requires independent validation, calibration, appropriate comparator groups, and sufficient sample size [31]. Applying these methods to natural-product therapeutics would additionally require standardized interventions, verified composition, adherence information, concomitant-medication data, and clinically meaningful endpoints. Without these conditions, molecular subgroups may be analytically reproducible yet therapeutically uninterpretable.

Microbiome integration may be particularly relevant because microbial communities can influence natural-product metabolism, bioavailability, transformation, and host response. Host–microbiome multi-omics methods provide a framework for combining microbial composition, microbial function, host molecular profiles, and clinical variables [32]. This creates agenda-setting opportunities for investigating response heterogeneity, herb–drug–microbiome interactions, and microbiome-associated safety signals. Pharmacovigilance integration, phytopharmaceutical standardization, clinical translation planning, and regulatory evidence alignment remain underdeveloped, however, and should be framed as future evidence requirements rather than as outcomes already demonstrated by the bibliometric landscape.

Bibliometric synthesis

The bibliometric landscape indicates that multimodal AI and multi-omics research in natural-product therapeutics is developing through overlapping methodological and domain-specific streams. Bibliometric studies of traditional Chinese medicine network pharmacology have documented visible publication growth and thematic concentration around compound–target–pathway analysis, mechanism interpretation, and computational pharmacology [33]. Within the bounded corpus, this visibility is reflected in the natural-product omics and network-knowledge cluster, but it does not establish that network-derived mechanisms are experimentally valid. The most defensible interpretation is that network pharmacology constitutes a prominent hypothesis-generation paradigm whose evidentiary strength depends on source-database quality and downstream validation.

Multi-omics drug-discovery bibliometrics similarly identify integrative omics, biomarkers, precision medicine, and computational analysis as active research themes [34]. The present review extends this perspective to natural products by emphasizing chemical fingerprints, biosynthetic information, metabolomics, network relations, knowledge graphs, microbiomics, and multimodal evidence integration. The dominant pilot modalities are not equally mature: chemical and metabolomic data are comparatively established, whereas clinical multimodality, imaging integration, pharmacovigilance, and regulatory evidence structures are less visible. Consequently, bibliometric prominence should be interpreted as a distribution of research attention rather than a hierarchy of translational readiness.

Bibliometric analyses in other natural-product-related domains demonstrate that co-citation, keyword co-occurrence, and burst-detection methods can reveal intellectual bases and emerging fronts when sufficiently complete metadata are available [35]. Those quantitative network analyses were not performed in the present bounded corpus because standardized keyword, citation-link, cited-reference, and affiliation fields were incomplete. The four reported clusters are therefore reviewer-coded thematic categories and must not be represented as algorithmically detected communities. This distinction preserves the review's interpretive validity and prevents qualitative synthesis from being presented as software-generated network evidence.

Global bibliometric analyses of AI in medicine further illustrate how annual output, source contributions, geographical patterns, and thematic development can characterize the evolution of a research field [36]. Such indicators remain descriptive: publication growth may result from technological enthusiasm, increased funding, broader indexing, methodological diffusion, or genuine scientific progress, and these explanations cannot be separated by publication counts alone. In the current corpus, the annual peak of 11 articles in 2022 and the distribution across 34 journals demonstrate the selected evidence base's temporal and disciplinary composition. They do not demonstrate worldwide publication dominance, comparative national performance, or increasing clinical effectiveness.

The synthesis therefore identifies an emerging evidence architecture rather than a validated translational pathway. Natural-product databases and chemical fingerprints supply identity and structure; omics layers characterize biosynthesis and biological response; network methods and knowledge graphs connect entities and evidence; AI models prioritize patterns or hypotheses; and validation systems determine whether outputs are chemically, biologically, clinically, and operationally credible. **Table 4** synthesizes translational opportunities, evidence gaps, and research agenda priorities derived from the bibliometric mapping. **Figure 2** presents a bibliometric synthesis map linking research frontiers, multimodal AI methods, multi-

omics data modalities, natural product therapeutic applications, translational opportunities, and evidence gaps.

Table 4. Bibliometric Synthesis of Translational Opportunities and Research Agenda Priorities: Mechanism Discovery, Target Prediction, Pathway Interpretation, Biomarker Discovery, Safety Assessment, Phytopharmaceutical Standardization, Evidence Integration, Clinical Translation Planning, and Validation Needs

Synthesis domain	Bibliometric signal	Scientific rationale	Translational opportunity	Evidence gap	Validation requirement	Implementation barrier	Research agenda priority	Caution boundary
Mechanism discovery	Strong qualitative signal across omics and network literature	Joint chemical and molecular profiles can connect constituents with pathways and phenotypes	Prioritize plausible mechanisms for experimental investigation	Limited causal confirmation and inconsistent product characterization	Perturbation experiments, target-engagement assays, dose–time replication	Heterogeneous extracts, incomplete annotation, non-specific enrichment	High	Integrated association is not proof of mechanism
Target prediction	Strong visibility in network-pharmacology and knowledge-graph research	Compound–protein and disease-network relations can rank candidate targets	Guide experimental target-validation programmes	Database bias, promiscuous predictions, few negative examples	Biophysical binding, genetic perturbation, functional confirmation	Inconsistent identifiers and uncertain predicted interactions	High	Predicted targets must not be reported as confirmed targets
Pathway interpretation	Strong qualitative signal	Multi-omics can reveal coordinated molecular responses	Identify pathways for mechanistic and biomarker investigation	Redundant pathway databases and background-set sensitivity	Replicated enrichment, functional assays, temporal analysis	Overlapping pathway definitions and selective reporting	High	Enrichment significance does not establish causal pathway dependence
Biomarker discovery	Moderate signal	Multimodal signatures may characterize exposure, response, or toxicity	Develop candidate response, quality, or safety markers	Small cohorts, batch effects, and limited external validation	Independent validation, calibration, discrimination, clinical-utility analysis	Specimen variability and inconsistent intervention definitions	High	Statistical association does not establish actionable utility
Synergy analysis	Emerging signal	Multicomponent products may produce interaction-	Prioritize combinations and interaction hypotheses	Sparse dose matrices and unclear	Factorial combination experiments and mechanisti	Combinatorial complexity and variable composition	Medium-high	Network overlap or joint activity does not demonstrate synergy

		dependent effects		definitions of synergy	c confirmation			
Safety assessment	Moderate signal	Chemical and biological models can prioritize potential toxic liabilities	Support risk-based testing and candidate exclusion	Natural-product-specific toxicology datasets are limited	Exposure-relevant in vitro and in vivo testing, metabolite assessment	Sparse negative labels and uncertain mixture effects	High	Prediction must not replace formal safety evaluation
ADMET prediction	Moderate signal	Early pharmacokinetic and toxicity screening can reduce unsuitable candidates	Prioritize compounds for absorption, metabolism, interaction, and toxicity studies	Domain shift from synthetic-drug datasets	External validation in natural-product-like chemical space	Limited experimental labels and uncertain structures	High	Model applicability domains must be explicit
Patient stratification	Limited emerging signal	Molecular and microbiome heterogeneity may influence therapeutic response	Develop responder hypotheses for standardized interventions	Few validated natural-product clinical multi-omics cohorts	Prospective external validation, calibration, comparator groups	Confounding, small samples, variable formulations	Medium	Current visibility does not support clinical treatment selection
Pharmacovigilance integration	Weak or underrepresented signal	Real-world reports may reveal adverse events and herb–drug interactions	Create structured surveillance for standardized products	Underreporting, ambiguous product identity, incomplete exposure data	Causality assessment, signal replication, product verification	Fragmented reporting systems and inconsistent terminology	High	Disproportionality or text-mining signals are not incidence estimates
Phytopharmaceutical standardization	Strong infrastructure signal	Chemical and multi-omics fingerprints may characterize identity and consistency	Improve batch comparability and evidence traceability	Limited interlaboratory validation and variable raw materials	Reference standards, validated assays, reproducibility studies	Agricultural, processing, and storage variability	High	Analytical consistency alone does not establish efficacy
Evidence integration	Strong emerging signal	Knowledge graphs can connect compounds, organisms, targets, assays, pathways,	Support traceable retrieval and hypothesis prioritization	Entity-resolution errors, missing provenance, unsupported inferred edges	Benchmark queries, manual auditing, temporal validation	Ontology mismatch and incomplete databases	High	Graph connectivity is not equivalent to evidentiary certainty

		and literature						
Clinical translation planning	Limited but strategically important signal	Structured readiness assessment can align preclinical evidence with clinical questions	Define evidence packages for prospective evaluation	Scarcity of validated endpoints and implementation studies	Prospective trials, standardized interventions, clinically meaningful outcomes	Cost, governance, recruitment, and reproducibility constraints	High	Bibliometric growth does not indicate clinical adoption
Regulatory evidence alignment	Weak signal	Auditable provenance and fit-for-purpose validation may support future assessment	Develop traceable evidence and model documentation	Few natural-product-specific AI standards	Quality-controlled data, documented model lifecycle, regulator-relevant validation	Regulatory heterogeneity and uncertain accountability	Medium-high	Regulatory readiness cannot be inferred from technical novelty

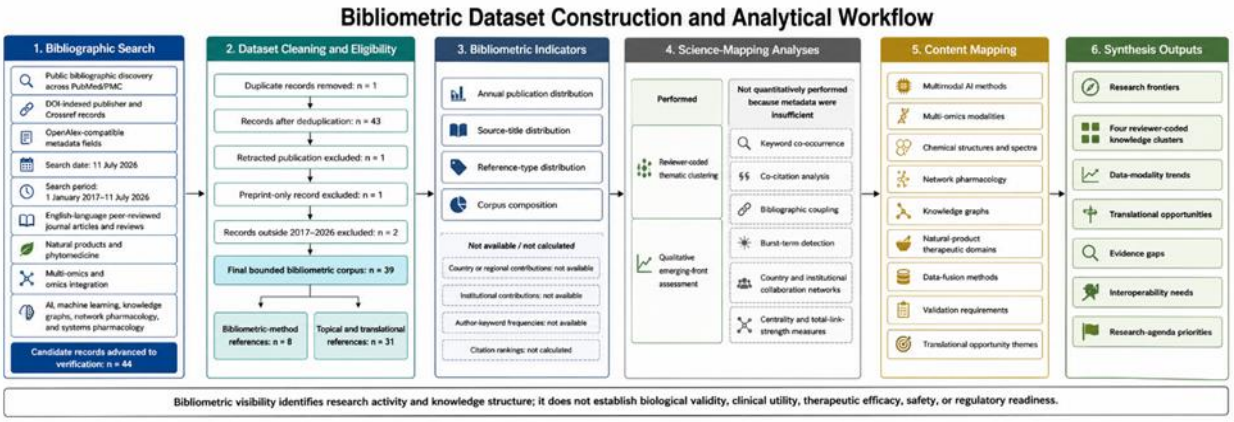

Figure 2. Bibliometric Synthesis Map of Research Frontiers, Data Modalities, and Translational Opportunities

Alt-text description

A four-layer conceptual map begins with four equal-size reviewer-coded research clusters. These connect to data modalities including metabolomics, transcriptomics, proteomics, genomics, microbiomics, chemical fingerprints, bioactivity profiles, network-pharmacology data, knowledge graphs, limited clinical or phenotypic data, imaging, and safety signals. A third layer contains machine learning, deep learning, multimodal data fusion, network integration, knowledge-graph reasoning, literature mining, pathway enrichment, target prediction, and biomarker discovery. The final layer presents mechanism discovery, safety assessment, phytopharmaceutical standardization, evidence integration, clinical translation planning, interoperability needs, and validation gaps. A boundary banner states that bibliometric prominence and computational prediction do not establish therapeutic validity.

Research agenda

The first priority is to improve the quality and interoperability of natural-product data infrastructure. Critical assessments of traditional-medicine databases indicate that incomplete records, inconsistent nomenclature, duplicated entities, uncertain structures, and weak provenance can compromise downstream network pharmacology and AI analyses [37]. Future datasets should use persistent chemical, taxonomic, genomic, assay, disease, and publication identifiers; report extraction and analytical procedures; retain negative and uncertain results; and document version histories. Benchmark datasets should also be separated into development, external validation, and prospective evaluation sets to reduce information leakage and inflated performance estimates.

The second priority is to establish validation pipelines for multimodal integration. Personalized-healthcare research shows that multi-omics translation depends on standardized specimen handling, interoperable metadata, robust integration methods, interpretability, external validation, and clinically meaningful evaluation [38]. Natural-product applications require these elements plus verified product composition, dose and exposure information, taxonomic authentication, manufacturing provenance, and explicit characterization of concomitant interventions. Safety-aware modelling should be integrated from the beginning rather than appended after efficacy-oriented candidate prioritization.

The third priority is to develop interpretable computational systems without treating interpretability as a substitute for validity. Biologically structured or visible neural networks may align model architecture with pathways, cell types, or known biological relations, potentially improving traceability of multimodal predictions [39]. Their future use should be assessed through predefined benchmarks, perturbation tests, uncertainty analysis, external datasets, and comparison with simpler baselines. Reproducible bibliometric monitoring should accompany this agenda by documenting database versions, query changes, metadata completeness, and unavailable analyses, while clinical-readiness and regulatory-readiness claims should remain prohibited until supported by fit-for-purpose prospective evidence.

CONCLUSION

This original bibliometric review maps the research frontiers, thematic knowledge clusters, data modalities, AI and integration methods, translational opportunities, evidence gaps, and research priorities shaping multimodal AI and multi-omics in natural product therapeutics. The bounded corpus shows an interdisciplinary landscape connecting bibliometric infrastructure, natural-product omics, databases, network pharmacology, multimodal learning, knowledge graphs, safety assessment, standardization, and translational planning. These patterns can guide dataset development, methodological comparison, validation design, and research investment, but publication growth, cluster visibility, multimodal integration, and computational prediction must not be treated as evidence of clinical efficacy, therapeutic safety, biological validity, or regulatory readiness.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Saldívar-González FI, Aldas-Bulos VD, Medina-Franco JL, Plisson F. Natural product drug discovery in the artificial intelligence era. Chem Sci. 2022;13(6):1526-46. doi:10.1039/D1SC04471K

[2] Sahayasheela VJ, Lankadasari MB, Dan VM, Dastager SG, Pandian GN, Sugiyama H. Artificial intelligence in microbial natural product drug discovery: Current and emerging role. Nat Prod Rep. 2022;39(12):2215-30. doi:10.1039/D2NP00035K

[3] Xie D, Tian Y, Zheng F, Wu L, Xu G, Liu Q, et al. Multi-omics and integrative analytics in natural products discovery. J Vis Exp. 2025;(225):e69458. doi:10.3791/69458

[4] Yang L, Wang H, Zhu Z, Yang Y, Xiong Y, Cui X, et al. Network pharmacology-driven sustainability: AI and multi-omics synergy for drug discovery in Traditional Chinese Medicine. Pharmaceuticals (Basel). 2025;18(7):1074. doi:10.3390/ph18071074

[5] Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr. 2017;11(4):959-75. doi:10.1016/j.joi.2017.08.007

[6] Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res. 2021;133:285-96. doi:10.1016/j.jbusres.2021.04.070

[7] Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ. Software tools for conducting bibliometric analysis in science: An up-to-date review. Prof Inf. 2020;29(1):e290103. doi:10.3145/epi.2020.ene.03

[8] Passas I. Bibliometric analysis: The main steps. Encyclopedia. 2024;4(2):1014-25. doi:10.3390/encyclopedia4020065

[9] Aria M, Le T, Cuccurullo C, Belfiore A, Choe J. openalexR: An R-tool for collecting bibliometric data from OpenAlex. R J. 2024;15(4):167-80. doi:10.32614/RJ-2023-089

[10] Borrego Á, Ardanuy J, Arguimbau L. Crossref as a bibliographic discovery tool in the arts and humanities. Quant Sci Stud. 2023;4(1):91-104. doi:10.1162/qss_a_00240

[11] Velez-Estevez A, Pérez IJ, García-Sánchez P, Moral-Muñoz JA, Cobo MJ. New trends in bibliometric APIs: A comparative analysis. Inf Process Manag. 2023;60(4):103385. doi:10.1016/j.ipm.2023.103385

[12] Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res

Synth Methods. 2020;11(2):181-217. doi:10.1002/jrsm.1378

[13] Palazzotto E, Weber T. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. Curr Opin Microbiol. 2018;45:109-16. doi:10.1016/j.mib.2018.03.004

[14] van Santen JA, Kautsar SA, Medema MH, Linington RG. Microbial natural product databases: Moving forward in the multi-omics era. Nat Prod Rep. 2021;38(1):264-78. doi:10.1039/D0NP00053A

[15] Zhang HW, Lv C, Zhang LJ, Guo X, Shen YW, Nagle DG, et al. Application of omics- and multi-omics-based techniques for natural product target discovery. Biomed Pharmacother. 2021;141:111833. doi:10.1016/j.biopha.2021.111833

[16] Zhan C, Shen S, Yang C, Liu Z, Fernie AR, Graham IA, et al. Plant metabolic gene clusters in the multi-omics era. Trends Plant Sci. 2022;27(10):981-1001. doi:10.1016/j.tplants.2022.03.002

[17] Noor F, Qamar MTU, Ashfaq UA, Albutti A, Alwashmi ASS, Aljasir MA. Network pharmacology approach for medicinal plants: Review and assessment. Pharmaceuticals (Basel). 2022;15(5):572. doi:10.3390/ph15050572

[18] Lardos A, Aghaebrahimian A, Koroleva A, Sidorova J, Wolfram E, Anisimova M, et al. Computational literature-based discovery for natural products research: Current state and future prospects. Front Bioinform. 2022;2:827207. doi:10.3389/fbinf.2022.827207

[19] Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E, Gaudry A, et al. The LOTUS initiative for open knowledge management in natural products research. Elife. 2022;11:e70780. doi:10.7554/eLife.70780

[20] Cai Z, Poulos RC, Liu J, Zhong Q. Machine learning for multi-omics data integration in cancer. iScience. 2022;25(2):103798. doi:10.1016/j.isci.2022.103798

[21] Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnol Adv. 2021;49:107739. doi:10.1016/j.biotechadv.2021.107739

[22] Agamah FE, Bayjanov JR, Niehues A, Njoku KF, Skelton M, Mazandu GK, et al. Computational approaches for network-based integrative multi-omics analysis. Front Mol Biosci. 2022;9:967205. doi:10.3389/fmolb.2022.967205

[23] Baião AR, Cai Z, Poulos RC, Robinson PJ, Reddel RR, Zhong Q, et al. A technical review of multi-omics data integration methods: From classical statistical to deep generative approaches. Brief Bioinform. 2025;26(4):bbaf355. doi:10.1093/bib/bbaf355

[24] Perdomo-Quinteiro P, Belmonte-Hernández A. Knowledge graphs for drug repurposing: A review of databases and methods. Brief Bioinform. 2024;25(6):bbae461. doi:10.1093/bib/bbae461

[25] 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):bbac404. doi:10.1093/bib/bbac404

[26] 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

[27] Onguéné PA, Simoben CV, Fotso GW, Andrae-Marobela K, Khalid SA, Ngadjui BT, et al. In silico toxicity profiling of natural product compound libraries from African flora with anti-malarial and anti-HIV properties. Comput Biol Chem. 2018;72:136-49. doi:10.1016/j.compbiolchem.2017.12.002

[28] Hasselgren C, Oprea TI. Artificial intelligence for drug discovery: Are we there yet? Annu Rev Pharmacol Toxicol. 2024;64:527-50. doi:10.1146/annurev-pharmtox-040323-040828

[29] Zeng T, Li J, Wu R. Natural product databases for drug discovery: Features and applications. Pharm Sci Adv. 2024;2(1):100050. doi:10.1016/j.pscia.2024.100050

[30] Arnold A, Alexander J, Liu G, Stokes JM. Applications of machine learning in microbial natural product drug discovery. Expert Opin Drug Discov. 2023;18(11):1259-72. doi:10.1080/17460441.2023.2251400

[31] Correa-Aguila R, Alonso-Pupo N, Hernández-Rodríguez EW. Multi-omics data integration approaches for precision oncology. Mol Omics. 2022;18(6):469-79. doi:10.1039/D1MO00411E

[32] Wang Q, Wang K, Wu W, Giannoulatou E, Ho JWK, Li L. Host and microbiome multi-omics integration: Applications and methodologies. Biophys Rev. 2019;11(1):55-65. doi:10.1007/s12551-018-0491-7

[33] Miao R, Meng Q, Wang C, Yuan W. Bibliometric analysis of network pharmacology in Traditional Chinese Medicine. Evid Based Complement Alternat Med. 2022;2022:1583773. doi:10.1155/2022/1583773

[34] Wang Z, Zhao Y, Zhang L. Emerging trends and hot topics in the application of multi-omics in drug discovery: A bibliometric and visualized study. Curr Pharm Anal. 2024;21(1):20-32. doi:10.1016/j.cpan.2024.12.001

[35] Peng B, Li Y, Yin J, Ding W, Wang F, Xiao Z, et al. A bibliometric analysis on discovering anti-quorum sensing agents against clinically relevant pathogens: Current status, development, and future directions. Front Microbiol. 2023;14:1297843. doi:10.3389/fmicb.2023.1297843

[36] Tran BX, Vu GT, Ha GH, Vuong QH, Ho MT, Vuong TT, et al. Global evolution of research in artificial intelligence in health and medicine: A bibliometric study. J Clin Med. 2019;8(3):360. doi:10.3390/jcm8030360

[37] Wang Y, Liu M, Jafari M, Tang J. A critical assessment of Traditional Chinese Medicine databases as a source for drug discovery. Front Pharmacol. 2024;15:1303693. doi:10.3389/fphar.2024.1303693

[38] Mohr AE, Ortega-Santos CP, Whisner CM, Klein-Seetharaman J, Jasbi P. Navigating challenges and opportunities in multi-omics integration for personalized healthcare. Biomedicines. 2024;12(7):1496. doi:10.3390/biomedicines12071496

[39] Selby DA, Jakhmola R, Sprang M, Großmann G, Raki H, Maani N, et al. Visible neural networks for multi-omics integration: A critical review. Front Artif Intell. 2025;8:1595291. doi:10.3389/frai.2025.1595291