Maturity Pathways for Computational Phytopharmacology: From Descriptive Compound Databases to Predictive, Explainable, and Translational Discovery Systems

Lukas Weber[1]*, Sophie Janssen[2], Jonas Richter[1], Elena Fischer[3]

[1]Department of AI for Plant-Based Antiviral Drug Discovery, Faculty of Chemistry and Pharmacy, University of Stuttgart, Stuttgart, Germany.
[2]Department of Computational Pharmacology and Systems Biology, Faculty of Pharmaceutical Sciences, Eindhoven University of Technology, Eindhoven, Netherlands.
[3]Department of Phytochemical Informatics, Faculty of Pharmacy, University of Basel, Basel, Switzerland.

ABSTRACT

Computational phytopharmacology is increasingly shaped by compound databases, curated natural product repositories, predictive models, knowledge graphs, explainable artificial intelligence, safety informatics, and translational decision-support concepts. However, the field often evaluates individual methods without a structured account of how computational systems mature from descriptive resources into evidence-aware discovery environments. This article proposes an original maturity pathway for computational phytopharmacology systems. The model defines staged progression from descriptive compound databases to curated and linked phytopharmacology data systems, predictive computational discovery systems, explainable evidence-aware decision-support systems, and translational validation-linked discovery systems. The proposed maturity logic emphasizes that higher maturity is not achieved merely by expanding database size or increasing model complexity. Instead, maturity requires improved data quality, provenance, chemical and botanical annotation, target and pathway evidence, ADMET and safety awareness, model validation, uncertainty communication, explainability, evidence grading, expert review, governance, and clearly bounded decision-support roles. The article distinguishes lookup-oriented resources from predictive hypothesis-generation systems and from translationally useful platforms that remain dependent on experimental and clinical validation. The maturity model offers a conceptual framework for researchers, database developers, AI model designers, pharmacologists, phytomedicine scholars, and translational scientists seeking to assess computational phytopharmacology systems without overstating predictive, mechanistic, or clinical claims.

Key Words: *Computational phytopharmacology, Maturity model, Natural product informatics, Predictive modeling, Explainable AI, Translational discovery*

eIJPPR 2025; 15(4):43-52

HOW TO CITE THIS ARTICLE: Weber L, Janssen S, Richter J, Fischer E. Maturity Pathways for Computational Phytopharmacology: From Descriptive Compound Databases to Predictive, Explainable, and Translational Discovery Systems. Int J Pharm Phytopharmacol Res. 2025;15(4):43-52. https://doi.org/10.51847/pVDhyMYU7Q

INTRODUCTION

Natural products remain important starting points for therapeutic discovery because they provide chemically diverse structures, biologically evolved scaffolds, and pharmacologically relevant activities, yet their development is complicated by structural complexity, source variability, target multiplicity, and translational uncertainty [1].

These features make computational phytopharmacology more than a matter of digitizing plant-derived compounds; it requires a staged way to connect botanical identity, phytochemical annotation, bioactivity evidence, predictive modeling, safety interpretation, and validation planning.

Corresponding author: Lukas Weber
Address: Department of AI for Plant-Based Antiviral Drug Discovery, Faculty of Chemistry and Pharmacy, University of Stuttgart, Stuttgart, Germany.
E-mail: ✉ lukas.weber@uni-stuttgart.de
Received: 05 April 2025; **Revised:** 08 August 2025; **Accepted:** 09 August 2025

Computational approaches have become central to natural product research because they can support virtual screening, target hypothesis generation, chemical-space exploration, molecular similarity analysis, pharmacological prioritization, and developability assessment [2]. In phytopharmacology, such methods are useful only when their outputs are treated as research-supporting hypotheses rather than direct evidence of efficacy, mechanism, safety, or clinical readiness.

Artificial intelligence has expanded the possible scope of natural product discovery by supporting structure-based prioritization, ligand-based prediction, molecular property estimation, generative design, and integration of complex biological data [3]. However, the movement from AI-assisted screening to mature computational phytopharmacology requires careful attention to training data, chemical representation, assay context, biological interpretability, and the risk of converting probabilistic outputs into unsupported pharmacological claims.

Recent scholarship on AI for natural product drug discovery has emphasized that useful computational systems depend on high-quality datasets, appropriate algorithms, careful benchmarking, and validation strategies that are suitable for the biological question being addressed [4]. This article therefore develops an original maturity model for computational phytopharmacology, with the aim of explaining how the field can progress from descriptive compound databases toward predictive, explainable, evidence-aware, and translational discovery systems while maintaining clear decision boundaries.

Computational phytopharmacology maturity

Computational phytopharmacology maturity can be defined as staged capability development across data availability, data curation, provenance, predictive modeling, explainability, evidence interpretation, validation readiness, governance, and translational usefulness. Descriptive natural product databases form an essential baseline because they organize structures, sources, names, and basic annotations, but database availability alone does not establish pharmacological evidence quality or decision-support readiness [5].

A more mature system must align the computational question with the available data and model type. Machine-learning applications in drug discovery show that model usefulness depends on training data relevance, endpoint quality, validation design, and the practical decision context in which outputs are used [6]. For phytopharmacology, this means that predictions about plant-derived compounds require clear boundaries around the chemical domain, biological endpoint, assay provenance, and intended research use.

Maturity should not be equated with algorithmic novelty, automation, or model complexity. Critical assessments of AI in drug discovery argue that computational impact depends on realistic problem framing, appropriate benchmarks, data quality, and integration into workflows where model outputs can be tested and interpreted [7]. In the proposed maturity logic, a simpler system with curated provenance and transparent validation may be more mature than a more complex black-box system that produces poorly contextualized predictions.

Chemical and biological data quality is a central maturity determinant because models trained on biased, heterogeneous, noisy, or biologically mismatched data may generate outputs that appear precise but remain weakly grounded [8]. Computational phytopharmacology maturity therefore requires movement from data accumulation to data meaning: botanical identity must be traceable, chemical annotations must be standardized, bioactivity evidence must be contextualized, and prediction outputs must remain linked to uncertainty, provenance, and validation status.

Descriptive and predictive system levels

The first transition in computational phytopharmacology maturity is from descriptive compound listing to curated phytopharmacology data systems. Curated medicinal plant resources can connect plant identity, phytochemical constituents, therapeutic associations, and literature-supported annotations, thereby moving beyond unstructured lists of compounds toward organized evidence resources [9]. In the proposed model, this transition marks the difference between Level 1 descriptive databases and Level 2 curated data systems.

Bioactivity-linked natural product repositories add another maturity layer by connecting species sources, compound identities, and experimentally reported activity information [10]. Such systems can support target prioritization, assay selection, and computational model development, but they should not be interpreted as proof that a plant-derived compound has a confirmed therapeutic mechanism. Their maturity contribution lies in transforming isolated compound records into evidence-bearing entries that may support downstream hypothesis generation.

Open natural product compound collections increase chemical coverage and computational reuse, but structural aggregation does not automatically resolve evidence provenance, biological relevance, or translational usefulness [11]. Linked knowledge-management initiatives strengthen maturity by making natural product information more reusable, interoperable, and provenance-aware across distributed research contexts [12]. Together, these resources illustrate that Level 2 maturity depends not only on how many compounds are represented but also on

whether records can be traced, connected, interpreted, and reused responsibly.

Predictive computational discovery systems represent Level 3 maturity when they use curated chemical and biological data to support hypothesis generation, candidate prioritization, target prediction, ADMET screening, toxicity estimation, or pathway inference. Target-prediction tools for small molecules can help identify plausible protein associations for phytochemicals, but their outputs remain computational predictions that require biological interpretation and validation before they can be treated as mechanisms [13]. Level 3 systems therefore shift computational phytopharmacology from retrieval toward prioritization while retaining a strict boundary between prediction and confirmation.

Table 1 compares descriptive, curated, linked, and predictive system levels in computational phytopharmacology maturity.

Table 1. Descriptive and Predictive System Levels in Computational Phytopharmacology Maturity: Compound Databases, Curated Repositories, Linked Data Systems, Knowledge Graphs, Target Prediction, ADMET Prediction, Pathway Inference, and Validation Requirements

System level	Core purpose	Primary data inputs	Main computational capability	Evidence generated	Key limitation	Validation need	Transition requirement
Descriptive compound database	Organize plant-derived compounds and basic annotations	Compound names, structures, source species, synonyms, identifiers	Search, retrieval, browsing, and inventory	Descriptive record-level information	Coverage may be mistaken for evidence quality	Identity checks, duplicate control, annotation consistency	Add curation rules, provenance, and evidence links
Curated phytopharmacology repository	Connect plant identity, phytochemistry, bioactivity, and therapeutic context	Botanical identity, plant part, compound annotation, literature records, bioactivity entries	Structured curation and evidence organization	Curated source-to-compound and compound-to-activity evidence	Curation depth and evidence quality may vary	Source traceability, curation audit, evidence-context review	Add interoperability, standardized identifiers, and linked data
Linked data system	Enable cross-resource connection and reusable knowledge representation	Standard identifiers, ontologies, provenance metadata, compound-target-pathway links	Interoperable querying and cross-database connection	Traceable linked evidence relationships	Linked records may imply unsupported biological causality	Provenance checks, relation validation, semantic consistency	Add predictive models and uncertainty-aware inference
Knowledge graph-enabled system	Represent entities and relationships across compounds, sources, targets, pathways, diseases, and evidence	Curated entities, relation types, evidence records, safety annotations, interaction records	Graph traversal, relation discovery, evidence path exploration	Mechanism-relevant and evidence-linked hypotheses	Graph paths may be overinterpreted as confirmed mechanisms	Relation-quality audit, evidence grading, expert review	Add validated prediction, explainability, and decision boundaries
Target-prediction system	Prioritize plausible molecular targets for phytochemicals	Chemical structures, known ligand-target data, similarity features, bioactivity data	Target hypothesis generation	Predicted target associations	Predictions are not confirmed mechanisms	External validation, applicability-domain assessment, experimental follow-up	Add pathway context and evidence-aware interpretation
ADMET or toxicity-prediction system	Screen developability and safety-	Chemical descriptors, ADMET labels,	Property and risk prediction	Candidate triage signals	Predicted risk does not equal	Benchmarking, domain analysis,	Add safety evidence integration and

	relevant properties	toxicity endpoints, assay metadata			experimental safety evidence	prospective testing	translational gates
Pathway-inference system	Link predicted targets or bioactivity signals to biological processes	Target data, pathway annotations, omics signals, disease associations	Mechanistic hypothesis organization	Pathway-level hypotheses	Pathway enrichment can overstate causal relevance	Biological plausibility review and experimental design linkage	Add explainability, evidence grading, and expert review
Predictive discovery system	Prioritize compounds, targets, mechanisms, or developability hypotheses	Curated chemical, biological, activity, pathway, and safety data	ML, QSAR, target prediction, ADMET prediction, network inference	Ranked or filtered research hypotheses	Model outputs may be mistaken for validation	Applicability-domain analysis, independent validation, uncertainty reporting	Add explainable, evidence-aware decision-support capability

Explainable and translational system levels

Higher-maturity computational phytopharmacology systems must connect prediction to interpretability, uncertainty, safety, and decision context. ADMET prediction platforms can support systematic screening of absorption, distribution, metabolism, excretion, and toxicity-related properties, but their role remains candidate triage rather than replacement of experimental safety assessment [14]. This distinction is especially important for plant-derived compounds because chemical complexity, exposure uncertainty, formulation context, metabolism, and interaction potential can alter the meaning of predicted properties.

Explainable AI is a defining requirement for Level 4 maturity because decision-support users need to understand why a model produced a target, property, risk, or prioritization output [15]. Explainability may show influential features, structural alerts, graph paths, or evidence relationships, but it should not be treated as causal proof of pharmacological action. In the proposed model, explainability improves interpretability and auditability only when paired with uncertainty communication, evidence provenance, and expert review.

Evidence-aware translational systems also require structured safety knowledge. Knowledge-graph approaches to pharmacokinetic natural product-drug interactions show how natural products, drugs, enzymes, transporters, evidence records, and interaction mechanisms can be connected in a computable representation [16]. This kind of evidence organization supports higher maturity because it allows safety-relevant relationships to be inspected, traced, and evaluated rather than hidden behind isolated prediction outputs.

A translationally oriented phytopharmacology system should connect chemical characterization, preclinical signals, clinical interaction evidence where available, and decision-focused modeling without implying that computational output alone determines clinical readiness [17]. Recommended approaches for modeling pharmacokinetic natural product-drug interactions further show that computational models should support decision-making only within transparent evidence boundaries and with explicit recognition of uncertainty, assumptions, and validation needs [18]. Level 5 maturity therefore refers to validation-linked research decision support, not regulatory approval, clinical recommendation, or proof of therapeutic safety.

Table 2 summarises explainable and translational system levels required for higher-maturity computational phytopharmacology.

Table 2. Explainable and Translational System Levels in Computational Phytopharmacology: Explainable AI, Evidence Grading, Uncertainty Communication, Human Expert Review, Validation Gates, Safety Awareness, Decision Support, and Translational Readiness

Higher-maturity capability	Core function	Required data quality	Required model feature	Evidence interpretation role	Decision-support boundary	Failure risk	Readiness signal
Explainable AI	Make model outputs interpretable to researchers	Curated inputs with traceable chemical and biological meaning	Feature attribution, rationale display, graph-path explanation, or	Helps users inspect why a prediction was produced	Supports interpretation, not causal confirmation	Explanation may be mistaken for mechanism proof	Output rationale is visible, bounded, and linked to uncertainty

			interpretable model structure				
Uncertainty communication	Indicate confidence limits and domain boundaries	Labeled data with known endpoint context and applicability information	Uncertainty estimates, domain flags, confidence warnings, or model-limit indicators	Helps distinguish strong, weak, and out-of-domain predictions	Supports cautious prioritization	Users may ignore uncertainty and over-rank predictions	Predictions include uncertainty and applicability-domain information
Evidence grading	Distinguish evidence strength across data types	Provenance-rich records from assays, literature, repositories, and validation studies	Evidence-layer tagging and source-quality indicators	Helps separate prediction, assay evidence, and translational evidence	Supports evidence-aware research triage	Weak evidence may appear equivalent to direct validation	Claims are linked to evidence type and evidence strength
Human expert review	Add pharmacological, chemical, botanical, and translational judgment	Reviewable records, transparent outputs, and accessible provenance	Human-in-the-loop review interface or expert annotation workflow	Helps identify implausible, unsafe, or poorly supported outputs	Supports decision refinement, not automated approval	Automation bias may reduce critical review	Expert decisions are documented and auditable
Validation gate	Connect computational hypotheses to testing requirements	Versioned datasets, model outputs, protocols, and experimental context	Gate logic for independent testing, replication, or prospective evaluation	Helps decide what evidence is needed next	Supports go/no-go research planning	Gate may be treated as a score rather than a requirement	Each transition has explicit validation requirements
Safety awareness	Integrate toxicity, ADMET, interaction, contraindication, and exposure concerns	Safety annotations, interaction evidence, ADMET data, exposure context	Safety filters, interaction graph links, and risk flags	Helps prevent purely efficacy-centered prioritization	Supports safety-informed research prioritization	Predicted safety may be mistaken for clinical safety	Safety signals are shown before translational prioritization
Translational readiness logic	Link computational outputs to preclinical, clinical, formulation, exposure, and implementation questions	Data connecting chemistry, biology, safety, exposure, and validation status	Decision-support workflow with evidence boundaries	Helps identify whether a candidate is ready for further validation planning	Supports structured research decisions, not medical advice	Readiness may be overstated as approval	Decision outputs remain conditional and validation-linked
Governance and auditability	Track model versions, data provenance, decision records, and review steps	FAIR, versioned, quality-controlled, and provenance-rich data	Audit trails, reproducible workflows, model documentation	Helps make system decisions inspectable and accountable	Supports responsible implementation	Opaque workflows may undermine trust and reproducibility	System outputs can be traced to data, models, evidence, and review decisions

Capability progression logic

Capability progression in computational phytopharmacology begins when a system moves from data availability toward data validity. Predictive systems cannot be considered mature simply because they produce compound rankings, target hypotheses, or property estimates; they require clear applicability-domain reasoning, transparent endpoint definition, and validation logic appropriate to the chemical and biological space being modeled [19]. For plant-derived compounds, this is particularly important because structural diversity, stereochemical complexity, source-dependent annotation, and heterogeneous assay evidence can make prediction boundaries difficult to define.

The next transition is from prediction to biologically interpretable prioritization. Drug-target interaction prediction systems can support prioritization by estimating plausible compound-target associations, but these outputs should be treated as hypotheses that guide experimental design rather than as confirmed mechanisms [20]. In a maturity pathway, Level 3 systems are therefore useful when they help select candidates, targets, or assays for further study, but they remain immature if they lack validation, uncertainty communication, or evidence-context reporting.

Developability-oriented prediction adds another transition requirement because candidate prioritization should not focus only on putative efficacy. Scalable ADMET prediction platforms can help evaluate large chemical libraries and support early triage, but their outputs remain computational signals that require endpoint-specific interpretation and follow-up evidence [21]. In computational phytopharmacology, this means that ADMET, toxicity, interaction, and exposure plausibility should be introduced before translational claims are made, not after mechanistic narratives have already been constructed.

The transition from predictive systems to responsible decision-support systems requires ground truth, reproducibility, human intervention, and explicit limits on what AI outputs can support [22]. Safety-aware maturity also requires herb-drug interaction evidence systems because phytopharmaceutical candidates may involve complex exposure, metabolism, transport, and co-use contexts that are not captured by compound lists alone [23]. **Figure 1** illustrates the capability progression pathway from descriptive compound databases to predictive, explainable, and translational computational phytopharmacology systems.

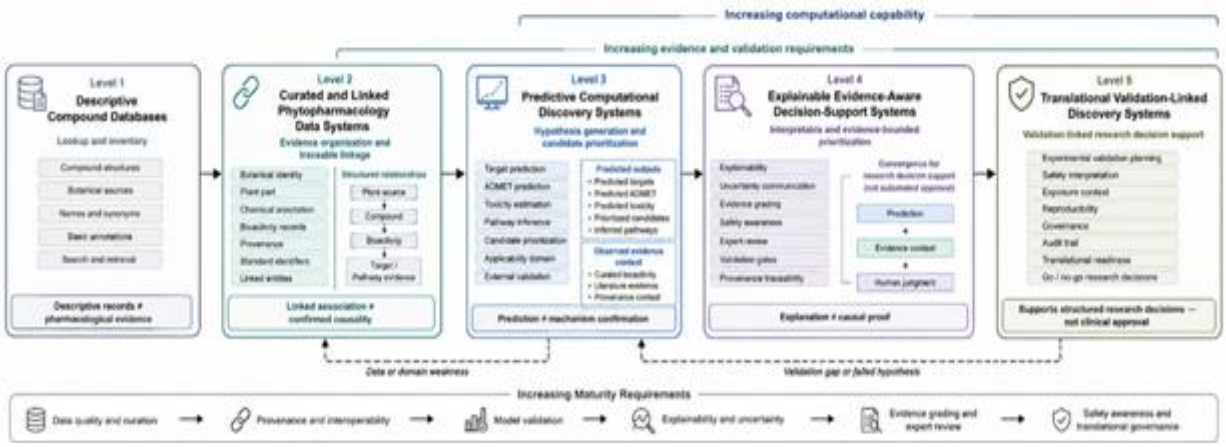

Figure 1. Capability Progression Pathway for Computational Phytopharmacology Maturity

Alt-text description

A staged maturity pathway diagram showing computational phytopharmacology systems progressing from descriptive compound databases to curated linked data systems, predictive discovery systems, explainable evidence-aware systems, and translational validation-linked systems. Capability requirements increase across data quality, prediction, explainability, validation, governance, and decision support.

Proposed maturity model

The proposed maturity model organizes computational phytopharmacology systems into five progressive levels. Level 1 consists of descriptive compound databases that support lookup, inventory, and basic annotation. Level 2 consists of curated and linked phytopharmacology data systems that connect plant identity, chemical entities, bioactivity evidence, source provenance, and interoperable identifiers. Level 3 consists of predictive computational discovery systems that generate target, ADMET, toxicity, pathway, or prioritization hypotheses. Level 4 consists of explainable evidence-aware decision-support systems that combine prediction with rationale, uncertainty, evidence grading, safety awareness, and expert review. Level 5 consists of translational validation-linked discovery systems that connect computational hypotheses to validation planning, safety interpretation, governance, and bounded research decisions. Biomedical knowledge graphs provide one organizing approach for connecting entities, evidence, and relationships across such maturity levels [24].

In this model, higher maturity reflects integration of technical capability with evidence quality. Drug-discovery knowledge graphs illustrate how structured relationships among compounds, targets, diseases, pathways, and evidence sources can support hypothesis generation and decision pathways when relation provenance and interpretation limits are preserved [25]. For computational

phytopharmacology, this means that graph-based systems should not merely connect more nodes; they should make it possible to inspect why a candidate, target, pathway, or safety signal is being prioritized.

Interoperability is a central transition requirement between curated data systems and evidence-aware decision-support environments. Life-science knowledge graph work shows that reusable biomedical knowledge representations can support cross-resource integration, semantic querying, and distributed data reuse [26]. In phytopharmacology maturity, this supports the need for standardized compound identifiers, botanical names, target identifiers, pathway terms, evidence records, safety annotations, and decision provenance.

Graph inference and embedding-based prediction can increase discovery capability, but only if evaluation reflects realistic use conditions. Knowledge graph embedding studies for drug-drug interaction prediction show that graph-based inference should be assessed under conditions that resemble practical prediction scenarios rather than idealized retrospective settings [27]. This principle is directly relevant to computational phytopharmacology because graph-predicted compound-target, compound-pathway, or interaction associations may otherwise appear more mature than their evidence base permits.

Human expert review is required because mature systems must support judgment rather than replace it. Human-in-the-loop molecule-generation workflows demonstrate how expert feedback can guide computational exploration toward goal-oriented outcomes [28]. Governance also requires data quality beyond simple FAIR availability, because maturity depends on whether data are findable, accessible, interoperable, reusable, and sufficiently reliable for their intended discovery use [29]. **Table 3** presents the proposed maturity model for computational phytopharmacology systems.

Table 3. Proposed Maturity Model for Computational Phytopharmacology: Descriptive Databases, Curated and Linked Data Systems, Predictive Discovery Systems, Explainable Evidence-Aware Systems, and Translational Validation-Linked Discovery Platforms

Maturity level	Level name	Defining capability	Data requirement	Model requirement	Explainability requirement	Evidence or validation requirement	Decision-support role	Transition priority
Level 1	Descriptive compound databases	Searchable inventory of plant-derived compounds, structures, sources, and basic annotations	Standardized compound names, chemical structures, botanical sources, identifiers, synonyms	No predictive model required	Transparent metadata fields and definitions	Identity checks, duplicate control, annotation review	Supports lookup and inventory only	Move from compound availability to curated evidence context
Level 2	Curated and linked phytopharmacology data systems	Curated connection among plant identity, phytochemistry, bioactivity, targets, pathways, and provenance	Botanical identity, plant part, compound annotation, bioactivity records, literature provenance, interoperable identifiers	Rule-based linking, semantic query, or graph representation may be used	Traceable source-to-record explanation	Curation audit, evidence provenance, relation consistency checks	Supports evidence exploration and research hypothesis framing	Add validated predictive models and applicability-domain logic
Level 3	Predictive computational discovery systems	Candidate, target, pathway, ADMET, toxicity, or developability hypothesis generation	Curated training data, endpoint labels, chemical descriptors, assay context, uncertainty-relevant metadata	ML, QSAR, target prediction, ADMET prediction, toxicity prediction, pathway inference	Basic rationale, feature contribution, or model-limit reporting	External validation, benchmark testing, applicability-domain assessment	Supports prioritization for experimental planning	Add explainability, evidence grading, safety context, and expert review
Level 4	Explainable evidence-aware	Integration of prediction, provenance,	Multimodal curated data, evidence records,	Predictive models plus graph	Human-readable rationale,	Expert review, validation-gate planning,	Supports bounded research	Link outputs to validation workflows and

	decision-support systems	uncertainty, evidence grading, safety awareness, and expert review	safety annotations, interaction evidence, pathway and disease context	inference, uncertainty communication, and evidence-aware filtering	uncertainty flags, graph paths, evidence traceability	evidence-strength assessment	decision-making	translational readiness conditions
Level 5	Translational validation-linked discovery systems	Connection of computational hypotheses to validation planning, safety interpretation, governance, and translational decision boundaries	FAIR-plus-quality data, versioned datasets, experimental metadata, exposure and safety context, decision records	Governed models embedded in reproducible validation-linked workflows	Full audit trail of rationale, uncertainty, evidence, review, and decision status	Independent validation planning, reproducibility checks, safety review, translational gate documentation	Supports structured go/no-go research decisions, not clinical approval	Maintain continuous updating, governance, and context-specific implementation

Figure 2 presents the proposed maturity model linking computational capabilities, evidence requirements, validation gates, explainability, and translational decision boundaries.

Figure 2. Maturity Model for Predictive, Explainable, and Translational Computational Phytopharmacology Systems

Alt-text description

A conceptual maturity model diagram showing five system levels connected to data requirements, model capabilities, evidence requirements, explainability, validation gates, expert review, safety awareness, translational readiness, and decision-support boundaries.

Practical implications

For computational phytopharmacology researchers and natural product database developers, the maturity model provides a way to assess whether a system mainly supports lookup, curated evidence exploration, prediction, explainable prioritization, or validation-linked decision support. FAIR data-management principles in drug discovery indicate that practical implementation requires attention to data discoverability, accessibility, interoperability, reusability, provenance, and workflow-level data stewardship [30]. Applied to phytopharmacology, this means that database builders should report curation logic, source provenance, chemical identifier standards, botanical identity handling, evidence type, and known limits of reuse.

For AI model developers, pharmacologists, translational scientists, journal reviewers, funding bodies, and platform designers, the model can guide reporting expectations and validation planning. A data-science roadmap for early-stage discovery emphasizes shared vocabularies, integrated data architecture, reproducible workflows, and alignment between computational outputs and experimental-data generation [31]. In phytopharmacology, these principles support clearer reporting of training data, applicability domains, uncertainty communication, evidence boundaries, expert-review procedures, and

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

validation gates before computational predictions are used to justify stronger mechanistic or translational claims.

Future research should develop benchmark datasets, transparent evaluation protocols, explainable AI methods suitable for structurally complex natural products, uncertainty-reporting standards, herb-drug interaction evidence frameworks, and validation-linked workflows that connect computational outputs to experimental pharmacology. The proposed maturity model can also support peer review by making it easier to distinguish descriptive database papers, predictive model papers, knowledge graph studies, decision-support concepts, and translational validation-linked systems. Its main practical value is not to assign universal scores, but to clarify what capability, evidence, and governance requirements must be met before a computational phytopharmacology system is described as mature.

CONCLUSION

This article proposes a maturity model for computational phytopharmacology that describes progression from descriptive compound databases to curated and linked data systems, predictive discovery systems, explainable evidence-aware decision-support systems, and translational validation-linked discovery platforms. The model emphasizes that mature systems require more than large datasets or advanced algorithms: they require curated data, provenance, chemical and biological annotation, validation, uncertainty communication, explainability, evidence grading, expert oversight, safety awareness, governance, reproducibility, and clearly bounded decision support. By separating lookup, prediction, explanation, evidence awareness, and translational readiness, the model offers a structured pathway for developing computational phytopharmacology systems without overstating computational predictions as confirmed mechanisms, validated safety claims, or clinical decisions.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, Supuran CT. Natural products in drug discovery: Advances and opportunities. Nat Rev Drug Discov. 2021;20(3):200-16. doi:10.1038/s41573-020-00114-z

[2] Medina-Franco JL. Computational approaches for the discovery and development of pharmacologically active natural products. Biomolecules. 2021;11(5):630. doi:10.3390/biom11050630

[3] 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

[4] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. Nat Rev Drug Discov. 2023;22(11):895-916. doi:10.1038/s41573-023-00774-7

[5] 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

[6] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5

[7] Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009

[8] Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: A discussion of chemical and biological data. Drug Discov Today. 2021;26(4):1040-52. doi:10.1016/j.drudis.2020.11.037

[9] Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, et al. IMPPAT: A curated database of Indian medicinal plants, phytochemistry and therapeutics. Sci Rep. 2018;8(1):4329. doi:10.1038/s41598-018-22631-z

[10] Zeng X, Zhang P, He W, Qin C, Chen S, Tao L, et al. NPASS: Natural product activity and species source database for natural product research, discovery and tool development. Nucleic Acids Res. 2018;46(D1):D1217-22. doi:10.1093/nar/gkx1026

[11] 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

[12] 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

[13] Daina A, Michielin O, Zoete V. SwissTargetPrediction: Updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res. 2019;47(W1):W357-64. doi:10.1093/nar/gkz382
[14] Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021;49(W1):W5-14. doi:10.1093/nar/gkab255
[15] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4
[16] 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
[17] 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
[18] Cox EJ, Tian DD, Clarke JD, Rettie AE, Unadkat JD, Thummel KE, et al. Modeling pharmacokinetic natural product-drug interactions for decision-making: A NaPDI Center recommended approach. Pharmacol Rev. 2021;73(2):847-59. doi:10.1124/pharmrev.120.000106
[19] Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, et al. QSAR without borders. Chem Soc Rev. 2020;49(11):3525-64. doi:10.1039/D0CS00098A
[20] Huang K, Fu T, Glass LM, Zitnik M, Xiao C, Sun J. DeepPurpose: A deep learning library for drug-target interaction prediction. Bioinformatics. 2020;36(22-23):5545-7. doi:10.1093/bioinformatics/btaa1005
[21] Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, et al. ADMET-AI: A machine learning ADMET platform for evaluation of large-scale chemical libraries. Bioinformatics. 2024;40(7):btae416. doi:10.1093/bioinformatics/btae416
[22] 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
[23] Zhang Y, Ip CM, Lai YS, Zuo Z. Overview of current herb-drug interaction databases. Drug Metab Dispos. 2022;50(1):86-94. doi:10.1124/dmd.121.000420
[24] Nicholson DN, Greene CS. Constructing knowledge graphs and their biomedical applications. Comput Struct Biotechnol J. 2020;18:1414-28. doi:10.1016/j.csbj.2020.05.017
[25] MacLean F. Knowledge graphs and their applications in drug discovery. Expert Opin Drug Discov. 2021;16(9):1057-69. doi:10.1080/17460441.2021.1910673
[26] 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:e52614. doi:10.7554/eLife.52614
[27] Celebi R, Uyar H, Yasar E, Gumus O, Dikenelli O, Dumontier M. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. BMC Bioinformatics. 2019;20(1):726. doi:10.1186/s12859-019-3284-5
[28] Nahal Y, Menke J, Martinelli J, Heinonen M, Kabeshov M, Janet JP, et al. Human-in-the-loop active learning for goal-oriented molecule generation. J Cheminform. 2024;16(1):138. doi:10.1186/s13321-024-00924-y
[29] Harrow I, Balakrishnan R, Küçük McGinty H, Plasterer T, Romacker M. Maximizing data value for biopharma through FAIR and quality implementation: FAIR plus Q. Drug Discov Today. 2022;27(5):1441-7. doi:10.1016/j.drudis.2022.01.006
[30] Gadiya Y, Ioannidis V, Henderson D, Gribbon P, Rocca-Serra P, Satagopam V, et al. FAIR data management: What does it mean for drug discovery? Front Drug Discov. 2023;3:1226727. doi:10.3389/fddsv.2023.1226727
[31] Edfeldt K, Edwards AM, Engkvist O, Günther J, Hartley M, Hulcoop DG, et al. A data science roadmap for open science organizations engaged in early-stage drug discovery. Nat Commun. 2024;15(1):5640. doi:10.1038/s41467-024-49777-x

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