Multi-Omics Network Analysis for Mechanism-Based Discovery of Natural Product Therapeutics

Shansiyuan Wu
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
shansiyuanwu243@unh.edu

Pradeep M. Anand
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
pradeep.anand@unr.edu

Andres M. Beck
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
andres.beck@ucf.edu

Anirudh L. Basu
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
anirudhwork@ku.edu

Abstract

The integration of multi-omics data with network-based analytical frameworks has become a central strategy for understanding the mechanisms of action of natural product therapeutics. Natural products frequently exhibit polypharmacological behavior that cannot be adequately captured by single-target models. Multi-omics network analysis provides a systems-level representation in which genomic, transcriptomic, proteomic, metabolomic, and phenotypic layers are jointly modeled to reveal how natural product interventions perturb molecular interaction networks and disease modules. This paper presents a systemic examination of the conceptual foundations, integration architectures, network construction methods, and mechanism-based discovery workflows that support the study of natural products. It addresses structural trade-offs among data integration paradigms, the interpretive challenges of network community detection, and the importance of robustness and fairness in computational predictions. The discussion extends beyond algorithmic performance to consider data governance, reproducibility, sustainability, deployment infrastructure, and policy implications. The analysis emphasizes that mechanism-based discovery of natural product therapeutics is not solely a technical problem but a socio-technical challenge requiring coherent institutional coordination, transparent validation standards, and equitable data-sharing practices. The paper situates multi-omics network analysis within broader efforts to modernize natural product pharmacology while preserving the complexity and contextual richness of traditional therapeutic knowledge. It concludes by outlining forward-looking perspectives on federated data ecosystems, explainable network models, and regulatory frameworks that can support sustainable translation of natural product candidates into clinically relevant interventions.

Keywords

multi-omics integration; network medicine; natural product therapeutics; mechanism-based discovery; systems pharmacology; data governance; biomedical infrastructure.

1. Introduction

The productivity crisis in conventional drug discovery has stimulated a broad shift toward systems-oriented approaches that conceptualize disease as a perturbation of molecular interaction networks rather than as the malfunction of a single target. Network medicine has formalized this perspective by demonstrating that disease-associated genes and proteins tend to cluster within specific neighborhoods of the human interactome, providing a quantitative foundation for interpreting pathobiology and therapeutic action [1]. In parallel, network pharmacology has reframed drug action as a systems-level event in which a compound modulates multiple nodes and edges, thereby generating therapeutic effects through distributed perturbations rather than isolated binding events [2]. Natural products are particularly well suited to this paradigm because their structural complexity and evolutionary selection often confer polypharmacological profiles that act across multiple targets and pathways. The re-emergence of natural product research in the genomics era reflects both the availability of new molecular profiling technologies and the recognition that many biologically active metabolites operate through mechanisms that single-target reductionism cannot explain [3].

Multi-omics data have become essential for investigating these mechanisms. Transcriptomic, proteomic, metabolomic, and epigenomic measurements provide complementary views of biological state transitions induced by natural product exposure. However, the value of these data is constrained by heterogeneity, scale, noise, and the incompleteness of current biological knowledge. Network analysis offers a unifying abstraction through which heterogeneous measurements can be projected onto common interaction spaces and interpreted in terms of pathway coordination, module activity, and cross-layer communication. The central challenge is therefore architectural and epistemological: how should multi-omics data be integrated so that network-derived inferences are mechanistically meaningful, reproducible, and translatable. This paper examines that challenge from a system-level perspective, attending to the structural trade-offs among integration strategies, the governance of data and models, and the long-term sustainability of discovery infrastructures.

The contribution of this paper is an integrative analytical discussion rather than a single empirical demonstration. It situates multi-omics network analysis for natural product therapeutics within the broader landscape of biomedical data science, computational pharmacology, and scientific governance. The discussion emphasizes that mechanism-based discovery requires more than predictive accuracy; it demands causal interpretability, robustness across biological contexts, fairness in algorithmic decision-making, and alignment with regulatory and ethical norms. By critically examining these dimensions, the paper aims to clarify why multi-omics network analysis is a necessary but institutionally demanding foundation for the rational development of natural product therapeutics.

2. Conceptual Foundations of Multi-Omics Network Analysis

The conceptual basis of multi-omics network analysis lies in the recognition that molecular layers are not independent but are organized into hierarchical and cross-regulatory systems. Gene set enrichment analysis was an early and influential formalization of this idea, enabling researchers to interpret genome-wide expression profiles by assessing the coordinated behavior of predefined biological categories rather than individual genes [4]. This conceptual

shift from single-molecule significance to functional coherence created the analytical foundation for later multi-omics integration. Pathway databases such as KEGG have further systematized this perspective by providing curated representations of metabolic, signaling, and disease-relevant pathways that can serve as reference frameworks for interpreting omics perturbations [5]. Protein-protein interaction resources such as STRING have complemented pathway models by supplying probabilistic interaction networks that capture physical and functional associations across species and experimental contexts [6]. Visualization and exploratory analysis platforms such as Cytoscape have made these complex network representations accessible to biologists, enabling iterative interpretation of network topology, annotation, and experimental overlay [7].

Within this conceptual landscape, the study of herbal medicines and other complex natural product mixtures raises distinctive analytical challenges because therapeutic effects often emerge from synergistic interactions among multiple constituents. Community-based network models have been proposed to identify such synergistic combinations by partitioning heterogeneous molecular interaction spaces into functionally coherent modules [8]. These approaches illustrate a broader principle: the mechanism of a natural product intervention may not reside in a single compound-target interaction but in the coordinated modulation of network communities that collectively influence disease-relevant processes. Multi-omics network analysis thus extends network pharmacology from single-compound polypharmacology toward mixture-level and formula-level systems analysis.

The conceptual promise of this framework is accompanied by important epistemological constraints. Omics layers differ in dynamic range, measurement error, temporal resolution, and biological interpretability. Mapping them onto a common network can obscure layer-specific uncertainty and can create the impression of mechanistic certainty where only correlational evidence exists. A robust conceptual foundation therefore requires explicit attention to data provenance, normalization, and the distinction between statistical association and causal perturbation. Network representations are best understood as heuristic models that organize evidence and generate mechanistic hypotheses rather than as transparent depictions of biological truth. This distinction is especially important in natural product research, where traditional knowledge, chemical complexity, and incomplete molecular annotation often interact in ways that resist simple computational formalization.

3. Multi-Omics Data Integration Architectures

The design of multi-omics integration architectures determines which biological signals can be recovered and which forms of mechanistic reasoning can be supported. Early integration concatenates molecular features before modeling, while late integration combines predictions generated separately from each omics layer. Network-based integration occupies an intermediate and increasingly influential position by projecting omics measurements onto molecular interaction graphs and then performing inference in the shared network space. The Connectivity Map demonstrated the power of transcriptomic signature matching as a mechanism-oriented integration strategy, connecting small molecules, genes, and disease states through gene expression profiles [9]. DrugBank provided a complementary pharmacological infrastructure by linking chemical structures, targets, pathways, and clinical annotations in a structured resource [10]. These resources have enabled integrative workflows that combine chemical information with omics-derived network activity.

More recently, graph-based deep learning architectures have expanded the analytical repertoire by learning representations directly from network structure and node features. For

example, graph convolutional networks have been applied to model polypharmacy side effects by integrating drug and protein interaction networks, demonstrating that relational inductive biases can improve predictions in complex biomedical systems [11]. In the context of natural products, network pharmacology approaches rooted in traditional Chinese medicine have emphasized the need to integrate compound libraries, target predictors, pathway annotations, and disease gene sets within a unified analytical frame [12]. These examples illustrate a broader architectural trend toward relational and heterogeneous data integration rather than flat feature concatenation.

The choice of integration architecture involves substantial structural trade-offs. Centralized data warehouses facilitate harmonization and query efficiency but raise governance challenges related to access, privacy, and data sovereignty. Federated approaches preserve local control over sensitive data but introduce latency, heterogeneity, and coordination costs. Middle-layer solutions based on semantic web technologies and common data models attempt to reconcile these extremes but require sustained investment in ontology alignment and community standards. In multi-omics natural product discovery, these architectural decisions are complicated by the geographic distribution of biodiversity, traditional knowledge, and research capacity. A technically elegant integration architecture that ignores these socioeconomic realities is unlikely to support equitable or sustainable discovery.

4. Network Construction and Analytical Methods

Network construction is the process by which heterogeneous molecular measurements are transformed into analyzable graph structures. The structure and dynamics of molecular networks have been extensively characterized in systems pharmacology, with attention to scale-free topology, modular organization, and the role of hubs and bottlenecks in information flow [13]. Systems biology provides the broader conceptual framework for understanding how these network properties emerge from local interactions and how they constrain possible therapeutic interventions [14]. Community detection methods are particularly important for multi-omics natural product analysis because they allow researchers to identify functionally related groups of genes, proteins, and metabolites that may respond collectively to a natural product mixture. Foundational work on network community structure established formal criteria for evaluating the strength of modular partitions [15], while scalable algorithms based on modularity optimization have enabled efficient community detection in large biological networks [16].

Analytical workflows typically combine community detection with centrality measures, network propagation, and module-level enrichment. Centrality measures identify nodes that occupy influential positions in the network, while propagation techniques estimate the extent to which a perturbation initiated at one set of nodes spreads to disease-relevant regions. These methods are useful for target deconvolution and for identifying candidate pathways through which natural products exert their effects. However, they are also sensitive to network incompleteness, annotation bias, and the choice of interaction confidence thresholds. Incomplete interactomes can distort centrality rankings and modular partitions, producing misleading conclusions about mechanism. The use of multiple complementary algorithms and the explicit reporting of uncertainty are therefore important safeguards.

The interpretability of network analytical outputs is a persistent concern. Biological networks are not static but context-dependent, cell-type-specific, and responsive to environmental conditions. A network constructed from aggregated data may not represent the relevant tissue or disease state in which a natural product acts. Efforts to incorporate condition-specific

expression, phosphoproteomic dynamics, or metabolomic flux can improve mechanistic resolution but also increase computational complexity. The central trade-off is between generality and context sensitivity. Mechanism-based discovery of natural products requires models that are sufficiently general to integrate diverse evidence while remaining sufficiently contextualized to support disease-relevant interpretation. This trade-off must be managed explicitly in the design of network construction pipelines and validation protocols.

5. Mechanism-Based Discovery of Natural Product Therapeutics

Natural products continue to provide privileged scaffolds for drug discovery because of their structural diversity, evolutionary optimization, and capacity to engage multiple targets. Recent advances in genomics, metabolomics, and synthetic biology have renewed interest in natural products and expanded the scope of accessible chemical space [17]. Metabolomics has become especially relevant for systems pharmacology because it measures the integrated biochemical state of a biological system and can reveal metabolic network responses that are not evident from gene or protein measurements alone [18]. The combination of metabolomics with network modeling provides a mechanistic bridge between chemical exposure and physiological outcome, supporting the identification of natural product modes of action.

Mechanism-based discovery workflows generally begin with multi-omics profiling of treated and control systems, followed by differential analysis, network projection, community detection, and the generation of mechanistic hypotheses. Artificial intelligence methods have been used to accelerate similar workflows in drug repurposing, as illustrated by machine learning models that integrate network and molecular data to prioritize existing drugs for new indications [19]. These applications demonstrate the value of computational inference in narrowing the search space, but they also raise concerns about algorithmic opacity, bias in training data, and the risk of overfitting to historical drug discovery patterns. Ethical analysis of algorithmic decision-making in biomedical contexts emphasizes the need for transparency, accountability, and consideration of downstream social consequences [20]. In natural product research, these concerns are compounded by the involvement of traditional knowledge and the potential for inequitable benefit sharing.

A distinctive feature of natural product therapeutics is the prevalence of multi-component mixtures whose effects cannot be attributed to a single compound. Network community-based methods offer a formal means of analyzing such mixtures by identifying modules that respond synergistically to combinations of constituents. Cross-domain comparisons with synthetic small molecules are instructive. Synthetic libraries often prioritize target selectivity and pharmacokinetic simplicity, whereas natural product mixtures frequently exhibit broader polypharmacology and more complex metabolic interactions. The analytical challenge is to preserve this complexity without rendering the system unintelligible. Multi-omics network analysis provides a middle path by allowing researchers to track how multiple constituents perturb network modules associated with disease, thereby generating testable mechanistic hypotheses for complex natural product interventions.

6. System-Level Trade-offs and Governance

The production and use of multi-omics network models for natural product discovery raise governance challenges that extend beyond technical performance. Data stewardship principles such as findability, accessibility, interoperability, and reusability, often summarized as FAIR principles, have become influential in biomedical data management because they support reproducibility and cross-institutional collaboration [21]. Applying FAIR principles to natural

product multi-omics data requires coordinated metadata standards, persistent identifiers, and community-accepted ontologies for chemical structures, biological samples, and traditional preparation methods. Without such infrastructure, integrative analyses may become irreproducible and locally constrained.

Algorithmic robustness and fairness are equally important. Deep learning models for bioactivity prediction have demonstrated considerable power but also sensitivity to hyperparameters, data partitioning, and representation choices [22]. Broader reviews of machine learning in drug discovery have highlighted the gap between computational validation and prospective experimental confirmation, as well as the tendency for models to inherit biases present in historical datasets [23]. Network-based screening methods that prioritize drug efficacy by measuring network proximity between drug targets and disease modules offer a more interpretable alternative, but they depend on the quality and completeness of the underlying interactome [24]. Disease network analyses have further shown that incompleteness in the interactome can distort disease-disease relationships and therapeutic inferences [25]. Systems pharmacology frameworks that integrate multiscale mechanisms are therefore essential for interpreting network-derived predictions in a biologically meaningful way [26].

These technical and governance issues are inseparable. A predictive model may perform well in a curated benchmark yet fail in real-world settings because of hidden covariate shifts, unmeasured confounders, or differences in data collection practices. Fairness concerns arise when models trained on data from particular populations or research centers are applied to different populations or traditional medicine contexts. Governance mechanisms such as external validation, model reporting standards, and independent auditing can mitigate these risks, but they require institutional commitment. Multi-omics network analysis for natural product therapeutics must therefore be governed as a socio-technical system in which algorithmic choices, data standards, and regulatory expectations evolve together.

7. Deployment, Sustainability, and Policy Implications

Deploying multi-omics network analysis in natural product discovery requires more than the availability of analytical workflows. It requires sustainable infrastructure for data storage, computation, version control, and community-based curation. Cloud-based platforms can reduce local computational barriers, but they introduce dependencies on commercial providers and raise questions about long-term access and cost. Federated learning and distributed query systems offer alternatives that preserve data locality, yet they demand robust interoperability standards and governance agreements. The deployment architecture should reflect the global distribution of biodiversity and traditional knowledge, ensuring that institutions in low-resource settings can participate meaningfully without surrendering control over their data.

Sustainability is a critical but often neglected dimension. Databases and ontologies must be continuously updated as new genomes are sequenced, new metabolites are annotated, and new disease associations are discovered. Funding models based on short-term project grants are poorly aligned with the long-term stewardship required for these resources. Sustainable governance may require consortial funding, public-private partnerships, or international coordination mechanisms. The natural product field also faces specific sustainability challenges related to bioprospecting, conservation, and the ethical use of traditional knowledge. Policy frameworks must reconcile the promotion of open science with the protection of biodiversity and the fair sharing of benefits arising from natural product research.

Policy implications extend to regulatory acceptance of mechanism-based evidence. Regulatory agencies traditionally rely on well-defined clinical endpoints and standardized chemical entities. Multi-omics network models can support mechanistic plausibility and patient stratification, but their evidentiary role remains uncertain. Clear guidance is needed on the circumstances under which network-derived evidence can inform decisions about natural product development, clinical trial design, or post-market surveillance. International harmonization of data standards and mutual recognition of analytical evidence could reduce duplication and accelerate translation. At the same time, policy must guard against algorithmic determinism, ensuring that computational predictions are treated as hypotheses subject to rigorous experimental and clinical scrutiny.

8. Conclusion

Multi-omics network analysis represents a significant advance in the effort to understand natural product therapeutics through mechanism-based reasoning. By integrating heterogeneous molecular measurements within a network framework, researchers can move beyond simplistic single-target explanations and account for the polypharmacological and synergistic properties that characterize many natural products. The approach draws on established traditions in network medicine, systems pharmacology, and omics data science, while extending them to the distinctive challenges of natural product mixtures and traditional therapeutic knowledge. Its analytical value depends on careful attention to integration architecture, network construction, community detection, and the management of uncertainty.

However, the long-term success of this paradigm will depend as much on governance, sustainability, and policy as on algorithmic innovation. Data heterogeneity, incomplete interactomes, algorithmic bias, and inequitable access to infrastructure are structural constraints that cannot be solved by computational methods alone. Mechanism-based discovery must be embedded within institutional frameworks that promote data stewardship, reproducibility, transparency, and fairness. Future efforts should prioritize federated data ecosystems, explainable network models, and regulatory frameworks that recognize the evidentiary role of systems-level inference while preserving rigorous standards of validation. By treating multi-omics network analysis as a socio-technical system rather than a purely computational task, the natural product research community can more effectively translate complex molecular evidence into safe, effective, and equitably developed therapeutics.

References

1. Barabasi, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56–68.

2. Hopkins, A. L. (2008). Network pharmacology: The next paradigm in drug discovery. Nature Chemical Biology, 4(11), 682–690.

3. Harvey, A. L., Edrada-Ebel, R., & Quinn, R. J. (2015). The re-emergence of natural products for drug discovery in the genomics era. Nature Reviews Drug Discovery, 14(2), 111–129.

4. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43), 15545–15550.

5. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research, 45(D1), D353–D361.

6. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research, 47(D1), D607–D613.

7. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498–2504.

8. Wang, Yinyin, et al. "HerbSyner_Finder: a network community-based model for identifying synergistic combinations from herbal medicines and complex systems." Targetome 2.2 (2026).

9. Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: Using gene-expression signatures to connect small molecules, genes, and disease. Science, 313(5795), 1929–1935.

10. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., & Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074–D1082.

11. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457–i466.

12. Li, S., & Zhang, B. (2013). Traditional Chinese medicine network pharmacology: Theory, methodology and application. Chinese Journal of Natural Medicines, 11(2), 110–120.

13. Csermely, P., Korcsmáros, T., Kiss, H. J. M., London, G., & Nussinov, R. (2013). Structure and dynamics of molecular networks: A novel paradigm of drug discovery. Pharmacology & Therapeutics, 138(3), 333–408.

14. Kitano, H. (2002). Systems biology: A brief overview. Science, 295(5560), 1662–1664.

15. Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E, 69(2), 026113.

16. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008.

17. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., Orhan, I. E., Banach, M., Rollinger, J. M., Barreca, D., Weckwerth, W., Bauer, R., Bayer, E. A., Majeed, M., Bishayee, A., Bochkov, V., Bonn, G. K., Braidy, N., Bucar, F., Cifuentes, A., D'Onofrio, G., Efferth, T., & Supuran, C. T. (2021). Natural products in drug discovery: Advances and opportunities. Nature Reviews Drug Discovery, 20(3), 200–216.

18. Kell, D. B., & Goodacre, R. (2014). Metabolomics and systems pharmacology: Why and how to model the human metabolic network for drug discovery. Drug Discovery Today, 19(2), 171–182.

19. Zhou, Y., Wang, F., Tang, J., Nussinov, R., & Cheng, F. (2020). Artificial intelligence in COVID-19 drug repurposing. The Lancet Digital Health, 2(12), e667–e676.

20. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21.

21. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.

22. Koutsoukas, A., Monaghan, K. J., Li, X., & Huan, J. (2017). Deep-learning: Investigating deep neural networks hyperparameters and comparison of performance to shallow methods for modeling bioactivity data. Journal of Cheminformatics, 9, 42.

23. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463–477.

24. Guney, E., Menche, J., Vidal, M., & Barabási, A.-L. (2016). Network-based in silico drug efficacy screening. Nature Communications, 7, 10331.

25. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabási, A.-L. (2015). Disease networks: Uncovering disease-disease relationships through the incomplete interactome. Science, 347(6224), 1257601.

26. Zhao, S., & Iyengar, R. (2012). Systems pharmacology: Network analysis to identify multiscale mechanisms of drug action. Annual Review of Pharmacology and Toxicology, 52, 505–521.