JOURNAL OF PHARMA INSIGHTS AND RESEARCH ISSN NO. 3048-5428

REVIEW ARTICLE

Applications of Artificial Intelligence in Plant-Based Anticancer Drug Discovery and Development

Enibokun Theresa Orobator[1], Chidinma Linda Nnodumele[2], Negasi Weldu Tsegay[3], Paul Onyekachi Onyekwelu[4], Augustine Obichukwu Odenigbo[5], Mary-Jane Ezinne Ugbor[6], Kindson Nkejah Abone[7], Bolaji Mubarak Ayeyemi[8], Jacob Ukpabio Inuaeyen[9], Kelechi Wisdom Elechi[10]

[1] *Medical Doctor, College of Medicine and Veterinary Medicine, University of Edinburgh, United Kingdom*
[2] *Research Scholar, Department of Anatomy, Faculty of Basic Medical Sciences, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria*
[3] *PG Scholar, Department of Health, Jackson State University, Jackson, Mississippi, USA*
[4] *Research Scholar, Data Science, AI & Modelling Centre (DAIM), University of Hull, England*
[5] *Regional Manager, Health Care Vertical, Business Development Unit - PPC Ltd, Victoria Island, Lagos, Nigeria*
[6] *Research Scholar, Department of Pharmacognosy and Phytomedicine, Faculty of Pharmaceutical Sciences, University of Port Harcourt, Rivers State, Nigeria*
[7] *UG Scholar, Faculty of Dentistry, University of Nigeria, College of Medicine, Enugu, Enugu State, Nigeria*
[8] *Research Scholar, Department of Computational Data Science and Engineering, College of Engineering, North Carolina Agricultural and Technical State University, North Carolina, USA*
[9] *PG Scholar, Department of Pharmaceutical and Medicinal Chemistry, University of Nigeria, Nsukka, Nigeria*
[10] *Research Scholar, Graduate School of Biomedical Sciences, University of Texas Health Sciences Center, San Antonio, USA*

Publication history: Received on 17th Feb 2025; Revised on 28th Feb 2025; Accepted on 1st March 2025

Article DOI: 10.69613/fbzjfk64

Abstract: The discovery of novel anticancer therapeutics faces significant challenges, including extended development timelines and high failure rates. Plants serve as an important source for anticancer compounds like paclitaxel and vincristine, yet traditional phytochemical discovery methods remain inefficient. Artificial intelligence (AI) and machine learning (ML) present innovative solutions to expedite the identification, validation, and optimization of plant-derived anticancer agents. Advanced computational techniques, including virtual screening, molecular modeling, and network pharmacology, enable rapid evaluation of vast phytochemical spaces. These tools can predict bioactivities, simulate molecular interactions, and suggest structural modifications to enhance drug-like properties. Recent successes include the identification of novel flavonoids targeting specific kinases and the optimization of traditional medicine compounds for improved efficacy. Current challenges encompass limited dataset availability, chemical complexity of natural products, and the need for experimental validation. The integration of multi-omics data and the development of specialized AI architectures for natural product chemistry show promise in addressing these limitations. AI-guided bioprospecting, automated ethnomedicinal knowledge mining, and the design of synergistic phytochemical combinations represents a transformative approach in natural product drug discovery, potentially leading to more efficient development of plant-based cancer therapeutics.

Keywords: Artificial Intelligence; Phytochemicals; Anticancer Agents; Drug Discovery; Natural Products.

1. Introduction

Cancer remains a paramount global health challenge, with approximately 19.3 million new cases and 10.0 million cancer-related deaths reported in 2020 [1]. Despite significant advancements in therapeutic modalities, including targeted therapies, immunotherapy, and precision medicine approaches, the need for novel anticancer agents persists, particularly due to emerging drug resistance and treatment limitations for specific cancer types. Natural products, especially plant-derived compounds, have played a pivotal role in cancer therapeutics. Over 50% of currently approved anticancer drugs originate from natural sources [2]. Notable examples include paclitaxel from *Taxus brevifolia* and vincristine from *Catharanthus roseus*, which have revolutionized cancer treatment protocols [3]. Paclitaxel's journey from its isolation from the Pacific yew tree to its establishment as a frontline treatment for ovarian and breast cancers illustrates both the potential and limitations of traditional natural product drug discovery [4]. Similarly, vinca alkaloids have become essential components in treating leukemias and lymphomas [5].

* Corresponding author: Enibokun Theresa Orobator

The conventional drug discovery pipeline typically spans 10-15 years from initial lead identification to regulatory approval [6]. This process involves screening vast compound libraries, optimizing promising hits, and conducting extensive preclinical and clinical studies. For natural products, additional complexities arise from isolation procedures, supply limitations, and structural complexity [7]. These challenges often result in promising phytochemicals remaining unexplored or abandoned during development. Artificial intelligence and machine learning technologies offer innovative solutions to overcome these limitations in modern drug discovery. These computational approaches can rapidly analyze extensive chemical and biological datasets, identifying patterns and relationships beyond human perception [8]. In pharmaceutical research, AI accelerates multiple stages: virtual screening evaluates millions of compounds in silico, predictive models forecast biological activities and toxicity profiles, while generative algorithms design novel chemical entities [9]. Recent data suggests that AI-assisted drug candidates demonstrate higher Phase I success rates (80-90%) compared to traditionally discovered compounds (~50%) [10]. The integration of AI with phytomedicine represents a strategic approach to systematically explore the plant kingdom's chemical diversity for anticancer agents [11]. This merger allows researchers to leverage traditional medicinal knowledge while employing modern computational tools. AI models can predict bioactive compounds from plants, simulate their interactions with cancer targets, and guide optimization efforts [12].

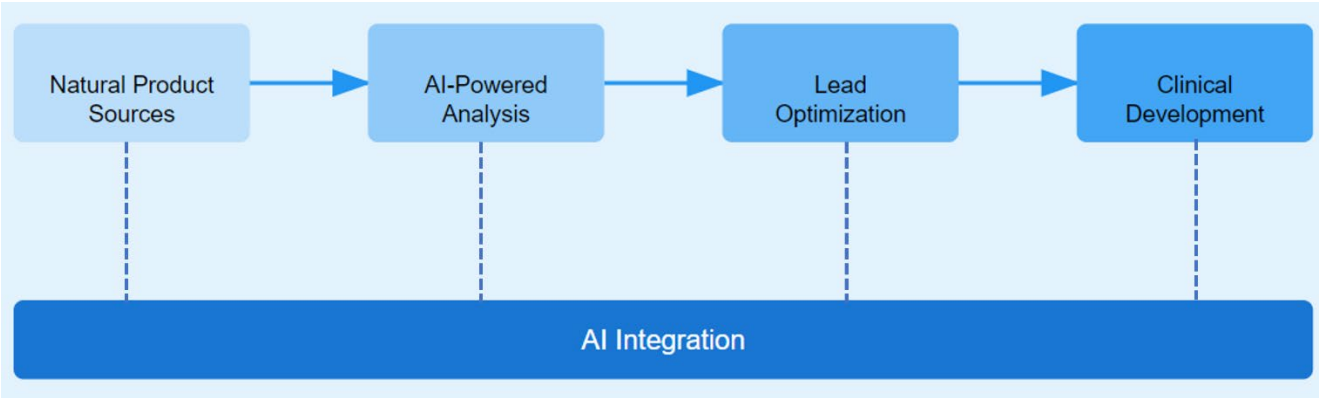

Figure 1. AI integrated Natural Drug Discovery Pipeline

2. Anticancer Phytochemicals

2.1. Major Compounds

Plant-derived compounds encompass diverse structural classes with distinct anticancer mechanisms. Alkaloids represent a significant category, with vinca alkaloids disrupting microtubule assembly in cancer cells. Vinblastine and vincristine from Catharanthus roseus demonstrate remarkable activity against lymphoid malignancies [13]. Terpenoids constitute another crucial class, exemplified by paclitaxel, which stabilizes microtubules and induces mitotic arrest in rapidly dividing cancer cells [14].

Polyphenolic compounds exhibit multifaceted anticancer properties through various molecular pathways. Curcumin, derived from Curcuma longa, modulates multiple signaling cascades including NF-κB, STAT3, and AP-1, leading to antiproliferative and pro-apoptotic effects [15]. Flavonoids, such as quercetin and myricetin, demonstrate potent antioxidant properties while targeting specific kinases involved in cancer cell survival [16].

Table 1. Major Plant-Derived Anticancer Drugs

Compound	Source Plant	Chemical Class	Mechanism of Action	Clinical Applications	FDA Approval
Paclitaxel	*Taxus brevifolia*	Diterpene	Microtubule stabilization	Breast, ovarian, lung cancer	1992
Vincristine	*Catharanthus roseus*	Vinca alkaloid	Microtubule disruption	Leukemia, lymphomas	1963
Irinotecan	*Camptotheca acuminata*	Camptothecin derivative	Topoisomerase I inhibition	Colorectal cancer	1996
Etoposide	*Podophyllum peltatum*	Podophyllotoxin derivative	Topoisomerase II inhibition	Lung cancer, testicular cancer	1983
Homoharringtonine	*Cephalotaxus harringtonia*	Alkaloid	Protein synthesis inhibition	Chronic myeloid leukemia	2012

2.2. Clinically approved Phytochemicals

Several plant-derived compounds have achieved remarkable clinical success. Camptothecin, isolated from *Camptotheca acuminata*, led to the development of topotecan and irinotecan, which are now standard treatments for various solid tumors [17]. Podophyllotoxin from *Podophyllum peltatum* served as the template for etoposide and teniposide, essential drugs in treating testicular cancer and small cell lung cancer [18].

Recent discoveries have identified promising new phytochemicals with unique mechanisms. Withaferin A, isolated from Withania somnifera, demonstrates selective cytotoxicity against cancer stem cells through novel pathways [19]. Sulforaphane, derived from cruciferous vegetables, shows potential in cancer prevention by activating Nrf2-dependent detoxification pathways [20].

3. Artificial Intelligence in Drug Discovery

3.1. Machine Learning Algorithms

Deep learning architectures, particularly convolutional neural networks and graph neural networks, excel at processing molecular structures and predicting biological activities [21]. These models can learn complex structure-activity relationships from large chemical datasets, enabling accurate prediction of anticancer potential for novel compounds [22].

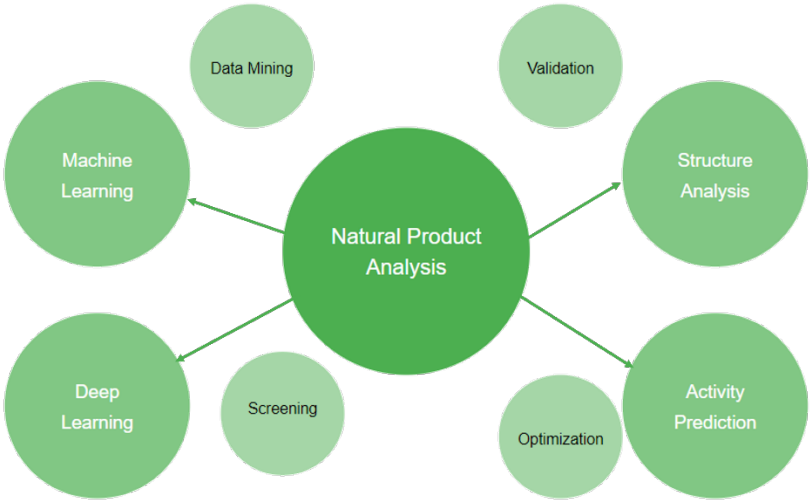

Figure 2. AI in Natural Product Analysis

3.2. Natural Language Processing

Advanced NLP models analyze scientific literature and patents, extracting valuable information about plant compounds and their biological activities. These systems can process millions of documents to identify previously overlooked connections between plants and specific cancer types [23].

Table 2. AI Applications in Different Stages of Plant-Based Drug Discovery

Discovery Stage	AI Technology	Application	Advantages	Practical Significance
Target Identification	Deep Neural Networks	Analysis of genomic data and pathway mapping	Rapid processing of complex biological networks	Number of novel targets identified
Virtual Screening	Convolutional Neural Networks	Structure-based compound screening	Processing millions of compounds daily	Hit rate improvement
Lead Optimization	Generative Adversarial Networks	Design of drug-like derivatives	Novel compound generation with desired properties	Reduction in optimization cycles
ADMET Prediction	Random Forests/XGBoost	Prediction of drug properties	Early identification of potential issues	Accuracy of predictions (>85%)
Clinical Trial Design	Machine Learning	Patient stratification and outcome prediction	Improved trial success rates	Reduction in trial failures

3.3. Generative Models

Variational autoencoders (VAEs) and generative adversarial networks (GANs) represent cutting-edge approaches in artificial intelligence for creating novel molecular structures based on desired properties. VAEs learn the underlying distribution of molecular features and generate new compounds by sampling from this learned space, while GANs utilize a competitive training process between generator and discriminator networks. These sophisticated models can suggest strategic modifications to natural compounds to enhance their drug-like characteristics while maintaining therapeutic activity, incorporating parameters such as molecular weight, topological polar surface area, and rotatable bond count. The models can be trained on successful drug candidates to learn the subtle patterns that contribute to efficacy [24].

3.4. Applications in Pharmaceutical Development

3.4.1. Target Identification

AI algorithms systematically evaluate complex genomic and proteomic data to identify novel cancer targets. Deep neural networks analyze intricate gene expression patterns across diverse cancer types, revealing previously unknown therapeutic opportunities through pattern recognition in high-dimensional data spaces. These systems can detect subtle molecular signatures that might be overlooked by traditional analysis methods [25]. The integration of multiple data types, including comprehensive mutation profiles, protein-protein interaction networks, and pathway analyses, enables highly precise target selection for specific cancer subtypes. This multi-modal approach considers the complex interplay between different biological systems and their role in cancer development [26].

3.4.2. Virtual Screening

State-of-the-art AI-powered screening platforms can efficiently evaluate billions of virtual compounds against cancer targets within days, representing a dramatic improvement over traditional method. Advanced deep learning models incorporate detailed three-dimensional protein structures and utilize sophisticated attention mechanisms, allowing for remarkably accurate prediction of binding modes and affinities across diverse chemical spaces [27]. These innovative systems consistently outperform traditional docking approaches by incorporating protein flexibility, accounting for water-mediated interactions, and considering entropy effects in binding calculations. The models can adapt to induced-fit effects and predict subtle conformational changes upon ligand binding [28].

4. Integration of AI in Phytochemical Discovery

4.1. Database Mining

Modern computational approaches have revolutionized the systematic exploration of extensive phytochemical databases, enabling rapid identification of promising candidates. The NPACT database serves as a comprehensive resource containing detailed information on 1,574 plant-derived anticancer compounds, including their mechanisms of action and experimental validations. The COCONUT database provides an even broader perspective, encompassing over 400,000 natural products with diverse structural features [29]. Sophisticated AI algorithms analyze these vast repositories to identify complex structural patterns and molecular fingerprints associated with anticancer activity, considering both known and novel chemical scaffolds [30].

4.2. Predictive Modeling of Biological Activity

4.2.1. Structure-Activity Relationships

Advanced neural networks process complex molecular descriptors through multiple layers of abstraction to predict anticancer activities with unprecedented accuracy. These sophisticated models consider detailed three-dimensional conformations, electronic properties including charge distribution and orbital energies, and comprehensive physicochemical characteristics to estimate binding affinities and cellular responses across diverse target classes [31]. Innovative transfer learning techniques enable models initially trained on synthetic compounds to adapt effectively to the unique chemical space of natural products, bridging the gap between traditional medicinal chemistry and natural product discovery [32].

4.2.2. Multi-Target Prediction

Modern AI systems evaluate potential interactions between phytochemicals and multiple cancer-related proteins simultaneously, providing a systems-level understanding of drug action. Advanced network pharmacology approaches systematically map compound-target-disease relationships, enabling comprehensive prediction of both desired therapeutic effects and potential side effects through complex network analysis [33]. These sophisticated models help identify compounds with optimal polypharmacological profiles for cancer treatment, considering both direct target engagement and downstream pathway effects [34].

4.3. Lead Optimization

4.3.1. Physicochemical Properties

AI-guided optimization strategies systematically improve drug-like properties of natural compounds while carefully maintaining their essential biological activity. Advanced generative models suggest precise structural modifications to enhance crucial properties such as aqueous solubility, membrane permeability, and metabolic stability, considering multiple parameters simultaneously [35]. These sophisticated approaches have demonstrated remarkable success in optimizing traditional medicine compounds for improved pharmaceutical properties, bridging the gap between traditional natural products and modern drug development requirements [36].

4.3.2. Synthetic Feasibility

Sophisticated machine learning models evaluate the synthetic feasibility of modified natural products through comprehensive analysis of chemical space. These systems carefully consider available synthetic routes, required reagents, potential yield-limiting steps, and stereochemical challenges in multi-step syntheses [37]. Such detailed analysis ensures that AI-suggested modifications remain practically achievable within the constraints of current synthetic organic chemistry capabilities, considering both technical feasibility and economic viability [38].

5. Validation and Optimization

5.1. Computational Validation

5.1.1. Molecular Dynamics Simulations

Advanced simulation techniques employ sophisticated algorithms to verify AI predictions by modeling atomic-level interactions between phytochemicals and target proteins with unprecedented precision. These simulations, often running on specialized GPU clusters, can achieve microsecond to millisecond timescales, revealing crucial insights into binding stability, conformational dynamics, and entropy effects. Long-timescale simulations capture essential protein motions and ligand reorganization events, providing detailed mechanistic insights into compound activity across multiple temporal scales [39]. The integration with quantum mechanical calculations, including density functional theory and ab initio methods, enables highly accurate modeling of electronic effects in natural product binding. These hybrid approaches account for charge transfer, polarization effects, and electronic reorganization during molecular recognition events [40].

5.1.2. Network Analysis

Systems biology techniques utilize graph theory and machine learning to evaluate compound effects on complex cellular networks. Sophisticated AI algorithms predict perturbations in cancer-related pathways through dynamic network modeling, identifying potential synergistic interactions and resistance mechanisms that might emerge during treatment. These analyses consider feedback loops, compensatory pathways, and network redundancy in cancer signaling [41]. The resulting information guide the strategic selection of optimal drug combinations and personalized dosing strategies, accounting for patient-specific molecular profiles and temporal dynamics of drug response [42].

5.2. Experimental Validation Pipeline

5.2.1. High-Throughput Screening

Automated screening platforms efficiently validate AI predictions through parallel testing of multiple compounds. Advanced robotics systems, equipped with precise liquid handling capabilities and automated imaging systems, simultaneously test compounds across diverse cancer cell lines, generating comprehensive biological activity profiles. These platforms incorporate real-time monitoring of cellular responses through multiple readouts, including viability, apoptosis, and specific pathway activation [43]. Sophisticated machine learning algorithms analyze the streaming screening data in real-time, employing adaptive experimental design strategies to optimize testing conditions and compound selection. [44].

5.2.2. Elucidation of Mechanism of Action

Cutting-edge multi-omics approaches systematically confirm predicted mechanisms of action through comprehensive molecular profiling. Advanced proteomics studies, utilizing high-resolution mass spectrometry and sophisticated protein arrays, reveal detailed compound effects on cellular pathways and protein networks. Metabolomics analyses track changes in cellular metabolism and identify key biomarkers of drug response. Complementary transcriptomics analyses, employing next-generation sequencing technologies, precisely identify gene expression changes and regulatory network perturbations induced by compound treatment [45]. Advanced AI integration helps interpret these complex, multi-dimensional experimental datasets, employing pattern recognition and causal inference algorithms to refine understanding of compound activity. These integrated analyses enable detailed mapping of drug-induced cellular responses and resistance mechanisms, facilitating optimal therapeutic application [46].

6. Challenges and Limitations

6.1. Data-Related Issues

6.1.1. Dataset Quality

Natural product databases often contain incomplete or inconsistent information. Experimental conditions vary across studies, complicating data integration and model training [47]. Standardization efforts and careful curation remain essential for reliable AI predictions [48].

6.1.2. Chemical Space Coverage

Available datasets incompletely represent natural product diversity. Many plant compounds possess unique structural features rarely found in traditional drug-like molecule collections [49]. This limitation necessitates specialized approaches for natural product modeling [50].

Table 3. Challenges and Solutions in AI-Assisted Natural Product Drug Discovery

Challenge	Issue	Impact	Solutions
Data Quality	Inconsistent reporting standards	Reduced model accuracy	Standardized reporting protocols
	Missing biological data	Incomplete predictions	Automated high-throughput screening
Technical Limitations	Model interpretability	Reduced trust in predictions	Explainable AI frameworks
	Computational resources	Limited processing capacity	Cloud computing integration
Biological Complexity	Complex mechanism of action	Incomplete understanding	Multi-modal AI integration
	Chemical space coverage	Limited prediction scope	Specialized natural product models

6.2. Technical Issues

6.2.1. Model Interpretability

Complex AI models often function as "black boxes," making prediction rationale unclear. Advanced visualization techniques and attention mechanisms help explain model decisions, but full interpretability remains challenging [51]. This limitation affects confidence in AI-generated predictions and their adoption in drug discovery programs [52].

6.2.2. Computational Requirements

Processing complex natural product structures demands significant computational resources. Three-dimensional conformational analysis and quantum calculations particularly strain available computing capacity. Cloud computing and distributed systems help address these limitations [53].

7. Conclusion

The use of artificial intelligence for plant-derived anti-cancer medicines represents a transformative shift in drug discovery. Researchers are redefining how we identify and develop potential cancer treatments from natural sources by combining computational power with botanical knowledge. These AI-driven approaches significantly shorten the traditional timeline for compound screening while preserving essential laboratory validation processes. Recent breakthroughs highlight this synergy's potential, particularly in predicting plant compound interactions and optimizing extraction methods. However, persistent challenges around data standardization and the 'black box' nature of some algorithms require focused attention. The coming years may see AI-enhanced phytomedicine not just accelerating discovery, but revealing entirely new therapeutic mechanisms hidden within plant ecosystems.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.

[2] Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. J Nat Prod. 2020;83(3):770-803.

[3] Atanasov AG, Zotchev SB, Dirsch VM, International Natural Product Sciences Taskforce, et al. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 2021;20(3):200-216.

[4] Kingston DGI. Taxol, a molecule for all seasons. Chem Commun (Camb). 2001;(10):867-880.

[5] Moudi M, Go R, Yien CY, Nazre M. Vinca alkaloids. Int J Prev Med. 2013;4(11):1231-1235.

[6] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20-33.

[7] Thomford NE, Senthebane DA, Rowe A, Munro D, Seele P, Maroyi A, et al. Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery. Int J Mol Sci. 2018;19(6):1578.

[8] 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):463-477.

[9] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-1250.

[10] Morgan P, Brown DG, Lennard S, Anderton MJ, Barrett JC, Eriksson U, et al. Impact of artificial intelligence in drug discovery: rethinking the drug design process. Nat Rev Drug Discov. 2018;17(2):97-113

[11] Xu J, Wang P, Yang H, Zhou J, Li Y, Li X, et al. Artificial intelligence in the discovery of natural products. Fitoterapia. 2020;145:104636.

[12] Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today. 2017;22(11):1680-1685.

[13] Mohamadi F, Nejad-Ebrahimi S, Reza Shafiei B. Vincristine and vinblastine: A review of biosynthesis, regulation, and biotechnological strategies. Phytochem Rev. 2020;19:1103-1132.

[14] Weaver BA. How Taxol/paclitaxel kills cancer cells. Mol Biol Cell. 2014;25(18):2677-2681.

[15] Giordano A, Tommonaro G. Curcumin and Cancer. Nutrients. 2019;11(10):2376.

[16] Zhu Y, Bo Y, Wang X, Lu W, Wang X, Han Z, et al. The Effect of Anthocyanins on Blood Pressure: A PRISMA-Compliant Meta-Analysis of Randomized Clinical Trials. Medicine (Baltimore). 2016;95(15):e3380.

[17] Liu LF, Desai SD, Li TK, Mao Y, Sun M, Sim SP. Mechanism of action of camptothecin. Ann N Y Acad Sci. 2000;922:1-10.

[18] Greenspan EJ, Madden E, Zhao L, Ritchie JP, Thibodeaux S, Haber D, et al. Molecular mechanisms of podophyllotoxin derivative-induced apoptosis in human colon cancer cells. Am J Physiol Gastrointest Liver Physiol. 2011;301(3):G504-G513.

[19] Lee IC, Choi BY. Withaferin-A—A Natural Anticancer Agent with Pleiotropic Mechanisms of Action. Int J Mol Sci. 2016;17(3):290.

[20] Zhang Y, Talalay P, Cho CG, Posner GH. A major inducer of anticarcinogenic protective enzymes from broccoli: isolation and elucidation of structure. Proc Natl Acad Sci USA. 1992;89(6):2399-2403

[21] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A Deep Learning Approach to Antibiotic Discovery. Cell. 2020;180(4):688-702.

[22] Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow RA, et al. Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov. 2020;19(5):353-364.

[23] Krallinger M, Rabal O, Lourenço A, Oyarzabal J, Valencia A. Information Retrieval and Text Mining Technologies for Chemistry. Chem Rev. 2017;117(12):7673-7761.

[24] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2:573-584.

[25] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038-1040.

[26] Sinha S, Vrana JA, Janjic JM. High-dimensional single-cell analysis enables rapid identification of therapeutic targets in cancer. Nat Rev Cancer. 2020;20(7):427-428.

[27] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

[28] Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY. Predicting Drug-Target Interaction Using a Novel Graph Neural Network with 3D Structure-Embedded Graph Representation. J Chem Inf Model. 2019;59(9):3981-3988.

[29] Sorokina M, Steinbeck C. Review on natural products databases: where to find data in 2020. J Cheminform. 2020;12(1):20.

[30] Chen Y, de Bruyn Kops C, Kirchmair J. Data Resources for the Computer-Guided Discovery of Bioactive Natural Products. J Chem Inf Model. 2017;57(9):2099-2111

[31] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chem Rev. 2019;119(18):10520-10594.

[32] Gao W, Coley CW. The Synthesizability of Molecules Proposed by Generative Models. J Chem Inf Model. 2020;60(12):5714-5723.

[33] Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nat Chem Biol. 2008;4(11):682-690.

[34] Zhang W, Pei J, Lai L. Computational Multitarget Drug Design. J Chem Inf Model. 2017;57(3):403-412.

[35] Schneider G, Clark DE. Automated De Novo Drug Design: Are We Nearly There Yet? Angew Chem Int Ed Engl. 2019;58(32):10792-10803.

[36] Liu X, Xu Y, Li S, Wang Y, Peng J, Luo C, et al. In Silico target fishing: addressing a "Big Data" problem by ligand-based similarity rankings with data fusion. J Cheminform. 2014;6(1):33.

[37] Coley CW, Green WH, Jensen KF. Machine Learning in Computer-Aided Synthesis Planning. Acc Chem Res. 2018;51(5):1281-1289.

[38] Segler MHS, Preuss M, Waller MP. Planning chemical syntheses with deep neural networks and symbolic AI. Nature. 2018;555(7698):604-610.

[39] De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of Molecular Dynamics and Related Methods in Drug Discovery. J Med Chem. 2016;59(9):4035-4061.

[40] Amaro RE, Mulholland AJ. Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. Nat Rev Chem. 2018;2:0148

[41] Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56-68.

[42] Bulusu KC, Guha R, Mason DJ, Lewis RPI, Muratov E, Kalantar Motamedi Y, et al. Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. Drug Discov Today. 2016;21(2):225-238.

[43] Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, et al. Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov. 2011;10(3):188-195.

[44] Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006;7(10):R100.

[45] Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171(6):1437-1452.

[46] Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. Cell. 2017;170(3):564-576.

[47] Bender A, Cortés-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):511-524.

[48] Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945-D954.

[49] Rodrigues T, Reker D, Schneider P, Schneider G. Counting on natural products for drug design. Nat Chem. 2016;8(6):531-541.

[50] Chen H, Engkvist O. Has Machine Learning Solved the Molecular Property Prediction Problem? Drug Discov Today Technol. 2019;32-33:89-96

[51] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. Adv Neural Inf Process Syst. 2017;30:4765-4774.

[52] Xie Y, Li X, Jiang M, Chen J. Interpretable Deep Learning for Natural Product Discovery. J Chem Inf Model. 2021;61(6):2712-2727.

[53] Wade M, Li YC, Wahl GM. MDM2, MDMX and p53 in oncogenesis and cancer therapy. Nat Rev Cancer. 2013;13(2):83-96.