Faculty of Natural and Applied Sciences Journal of Scientific Innovations
Print ISSN: 2814-0877 e-ISSN: 2814-0923
www.fnasjournals.com
Volume 6; Issue 1; September 2024; Page No. 45-57.

Harnessing Artificial Intelligence for the Discovery and Development of Natural Product-Based Therapeutics

***Okibe, G., & Samuel, H. S.**
Department of Chemical Sciences, Federal University Wukari Taraba State, Nigeria

***Corresponding author email:** gideonokibe2@gmail.com

Abstract
The incorporation of advanced technologies in the exploration and advancement of therapeutics derived from natural products signifies a groundbreaking phase in research. While traditional approaches to drug discovery have their merits, they often encounter issues related to time and cost-effectiveness. AI's utilization of machine learning (ML) learning (DL) and data analysis has revolutionized the process of discovering, refining, and creating therapeutics. This article explores the efforts between AI and natural product studies emphasizing progressions and uses in developing antibiotics and anticancer medications, delving into the techniques, real-life examples, and future paths of AI-guided drug discovery from sources. Ranging from forecasting bioactivity and protein targets to tuning drug characteristics AI has displayed potential in accelerating the pace and accuracy of pharmaceutical innovations. Despite obstacles, the fusion of AI with natural product investigations holds promise for unveiling treatments that can significantly enhance well-being.

Keywords: AI, Anticancer, Drug Discovery, Machine Learning, Bioactive Compounds

Introduction

Natural product-based therapeutics have long been a cornerstone of medicinal chemistry and pharmaceutical development with sources from plants, micro-organisms and animals. Statistics indicate that more than 80% of drug molecules are sourced from natural origins. Over 70% of FDA-approved anticancer medications are derived from natural sources. The worldwide market for herbal medications, products, and raw materials is anticipated to reach an approximate value of $80-90 billion and is anticipated to increase to $5 trillion by 2050. These compounds offer a rich chemical diversity often unparalleled by synthetic libraries, yielding the discovery of unique bioactive molecules with potential for therapy (Padma & Don, 2022; Abdul & Khan, 2021; Jagadevappa, 2016). Natural product-based therapeutics have significantly contributed to various research fields, including antibiotic development, anticancer therapies, antiviral agents, antimalarial drugs, dermatological treatments, and more. For example, paclitaxel, a plant-based anticancer drug derived from Taxus brevifolia, has been crucial in cancer treatment (Sati et al., 2024). Erythromycin, an antibiotic derived from Saccharopolyspora erythraea, has revolutionized antibiotic therapy (Wu et al., 2016). Tacrolimus, an immunosuppressant from Streptomyces tsukubaensis, is essential in organ transplantation (Ordóñez-Robles et al., 2018). Recent advancements in tuberculosis (TB) treatment include natural product-derived drugs such as rifapentine, CPZEN-45, and spectinamides 1599 and 1810, highlighting the ongoing importance of natural products in developing new therapeutics (Kumar, 2023).

Recently, artificial intelligence (AI) has begun to have a significant impact across different scientific and industrial fields, including drug discovery and development. AI includes machine learning (ML), deep learning (DL), and other computational methods that provide powerful tools for processing large volumes of data, identifying patterns, and making predictions with remarkable speed and accuracy (Joshi et al., 2023). It is centred on building machines that can carry out operations that call for human intellect. They involve learning, thinking, solving problems, perceiving, comprehending language, and making decisions (Zhang, 2023). AI has the ability to completely transform medication research and discovery by speeding up the procedure and cutting related expenses. Conventional drug development techniques include a large financial commitment and a protracted trial-and-error process. (Han et al., 2023). These traditional methods are laborious, time-consuming, and fraught with challenges such as identifying active compounds, optimizing their pharmacokinetic properties, and elucidating their mechanisms of action (Susana et al., 2023; Inder et al., 2021).

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

However, prospective drug candidates can be identified using AI-driven approaches that evaluate large datasets, including chemical, proteomic, and genomic information. The entire drug development process can be accelerated by using these technologies, which make it possible to forecast a drug's efficacy and toxicity, identify potential therapeutic targets more quickly, and permit drug repurposing (Han et al., 2023). These advanced methods include machine learning for predicting drug properties, deep learning for analyzing biological data, natural language processing for mining scientific literature, generative modelling for creating new molecules, and network-based approaches for identifying drug development targets (Visan & Negut, 2024). According to Saldívar-González et al. (2022), advanced algorithms and high-performance computing empower AI to identify and predict molecular patterns of bioactive compounds and their interactions with protein targets. This significantly enhances the speed and efficiency of drug discovery. AI-driven tools can streamline dereplication by distinguishing novel compounds from known ones early in the discovery pipeline.

The application of AI in natural product studies is still in its early stages, and several limitations need to be addressed, such as the need for high-quality, curated datasets to train AI models, the integration of AI tools into existing experimental workflows, and the interpretation of AI-generated predictions in biologically meaningful contexts (Kim et al., 2020). In addition, ethical and legal issues need to be resolved to guarantee the proper application of AI in drug discovery (Tiwari et al., 2023). The aim of this article is to provide a thorough review of the state of AI applications at the present time in the discovery and development of natural product-based therapeutics. We will explore the various AI methodologies employed in this field, discuss case studies and successful applications, and identify the limitations and future prospects for research. By harnessing the power of AI, we stand at the threshold of a new development in natural product research, with the potential to unlock novel therapeutics and improve human health in profound ways.

AI Applications in Natural Product-Based Drug Discovery

Computational omics technology advancements have vast applications in drug development into the hidden natural-based product varieties. Parallel to this, computational drug design has seen promising advancements in AI techniques like machine learning, making it easier to predict biological activity and create new drugs for target molecular targets (Mullowney et al., 2023). Computer-aided drug discovery has been influenced by AI. The development is further facilitated by the increasing application of machine learning, especially deep learning models, across a wide range of scientific disciplines and advancements in computer hardware and software. Medicinal chemistry has benefited as a result of early concern about AI being replaced by pharmaceutical discoveries (Jiménez-Luna et al., 2021). A valuable resource for contemporary drug development includes natural compounds produced by fungi, bacteria, plants, animals, and other organisms. Natural products' biological relevance and structural diversity make them appealing starting points for the development of new drugs. In natural product-based drug development, computational approaches are a helpful precursor or complement to in vitro testing (Chen et al., 2017). Machine learning algorithms have been a major area of evolution for the pharmaceutical sector, and different supervised and unsupervised learning approaches are being used at different stages of the drug development process. Cell-type picture segmentation, protein target drug-ability prediction, and de novo molecular design have all used clustering approaches. Regressions and classifications supervised learning techniques identified potential targets for Huntington's disease. For drug design and many other applications, they hypothesized biological activities and the characteristics of absorption, distribution, metabolism, excretion, and toxicity (ADME/Tox) (Vamathevan et al., 2019). Natural product research, a dependable source of contemporary small molecule drug discovery, has gradually included computational techniques involving AI and machine learning algorithms. For instance, dimensionality reduction methods (such as principal component analysis and self-organizing maps) were mostly used in the early 2000s to map the NP chemical space and digitize organic molecules. Machine learning binary classifiers were created to forecast their biological roles in the ensuing ten years. Recently, researchers have begun using neural network topologies for molecular design and genome mining (Ernst et al., 2015).

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

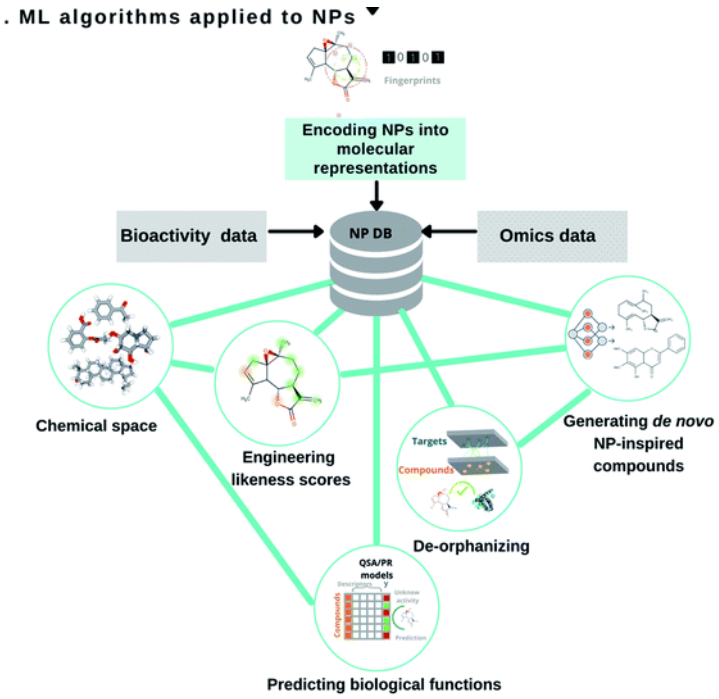

Figure 1.0: The application of machine learning techniques to natural phenomena (NPs) includes encoding them into molecular representations, molecular descriptors, similarity scores, chemical space, biological function prediction, de-orphanization, and the generation of de novo compounds inspired by natural phenomena (Saldívar-González et al., 2022).

Translating NPs (or any chemical structure) into computer-readable format(s), or so-called molecular representations, is a major step in modeling and forecasting their properties and bioactivities. Most representations encode chemical information with a specific intent. It is possible to obtain chemical compounds with similar names by using their original IUPAC and generic names. The computational task involved matching chemical structures according to the descriptions of their bi-dimensional molecular graphs. Early molecular representations were designed to aid in efficient structural searches or to preserve chemical information in a lightweight fashion. To identify common molecular features or substructures from databases and to store and retrieve molecular information, three tools were developed: the international chemical identifier (InChI), the simplified input line entry system (SMILES), and the SMILES arbitrary target specification (SMARTS, Daylight CIS, and OpenEye Scientific Software) (Jadhav et al., 2020). In addition to fingerprints, which are frequently employed by chemo informaticians, computational chemists would use molecular representations to compute hundreds of attributes or variables, known as molecular descriptors, using well-defined techniques. These descriptors capture certain easy-to-understand molecular qualities (such as atomic properties, size, shape, flexibility, polarity, lipophilicity, and pharmacophore). Molecular descriptors have been a key component in creating predictive QSA/PR modeling. In low-dimensional representations of chemical space, they have proven indispensable in characterizing the distributions of nanoparticles and synthesized chemicals. The bioactivity of natural chemicals is one area where machine learning is used in drug discovery. Chemical structures and the biological processes they correspond with are included in datasets that can be used to train machine-learning algorithms. Drug research can then proceed more quickly thanks to these models' ability to forecast the possible bioactivity of novel substances. To effectively forecast the bioactivity of chemicals generated from natural products, for instance, supervised learning algorithms such as Random Forest and Support Vector Machines (SVM) have been utilized (Wu et al., 2019). Deep language models have been employed to predict the bioactivity of natural compounds. These models can automatically learn feature representations from raw data, making them highly effective for complex tasks such as bioactivity prediction. In several studies, deep language has outperformed traditional ML methods, demonstrating its potential to enhance drug discovery processes (Chen & Xu, 2020; Gao & Ding, 2019). Furthermore, Natural language processing techniques have applications in mining literature for information about the bioactivity of natural compounds. By processing and analyzing published research articles, NLP algorithms can identify compounds with reported bioactivity and aggregate this information into structured databases. This automated extraction of knowledge can greatly aid researchers in identifying promising natural compounds for further investigation (Zhao & Li, 2018; Shultz, 2019).

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

Additionally, ML can assist in locating possible protein targets for organic substances. In order to do this, algorithms are used to examine how tiny chemicals and proteins interact. When ML is used in conjunction with techniques like docking simulations, it is possible to forecast which proteins a specific drug will likely interact with (Kinnings et al., 2011). The issue of finding protein targets has also been tackled using DL. More accurately than with conventional methods, advanced techniques such as graph neural networks (GNNs) can describe the interactions between chemicals and proteins. NLP can help by extracting pertinent data from scientific literature, which can help find protein targets. While relation extraction techniques can reveal the links between these entities, named entity recognition (NER) tools can recognize mentions of proteins, genes, and chemicals in the literature (Zhu & Shi, 2021).

AI-Driven Approaches and Case Studies

Virtual screening

One essential in silico method used in the drug discovery process is virtual screening (VS) enabling the automated evaluation of extensive molecular databases to identify potential therapeutic candidates. By acting as a preliminary filter, VS helps to eliminate compounds with less desirable properties, thus narrowing down the pool to those with a higher likelihood of biological activity (Oliveira et al., 2023). Throughout the virtual screening process, candidate ligands can undergo adjustments in their composition and structure to improve their properties, particularly their pharmacokinetic attributes such as toxicity, excretion, metabolism, distribution, and absorption (ADMET). A pipeline for virtual screening consists of two primary computational operations. The first step entails preparing the library, which includes acquiring compound structures and transforming data into usable formats (such as SDF, SMILES, and MOL2) (Saldívar-González et al., 2020), generating conformers, and correcting stereochemical and valence errors. In order to filter desirable chemicals, the second step uses computational tools. In vitro and in vivo tests, such as enzymatic or cell line inhibition, are the last methods used for experimental validation. Artificial intelligence-based virtual screening procedures have made use of a variety of computational techniques that have been developed over time. Combining these techniques with experimental methodologies increases the likelihood of finding new bioactive chemicals (Santana et al., 2021). Lignon-based virtual screening (LBVS) and structure-based virtual screening (SBVS) are the two computer methods used in compound virtual screening. In order to find novel bioactive chemicals against a particular molecular target or biological system, both computational methodologies have been merged into virtual screening strategies.

Ligand-Based Virtual Screening (LBVS) is a molecular activity prediction technique that uses a set of known bioactive compounds as a basis for the analysis of intrinsic compound features, including electronic, topological, physicochemical, and structural traits (Berenger et al., 2017; Garcia-Hernandez et al., 2019). Examples of computational tools are machine learning, cheminformatics filters, pharmacophore modeling, similarity searches, and QSAR analysis. SBVS, on the other hand, starts with a bioreceptor's three-dimensional structure to investigate the interactions between ligands and their binding site. The effectiveness of this strategy depends on the ability to comprehend intermolecular interactions, binding site residue composition, ligand-binding affinity, and bioreceptor conformation (Maia et al., 2020). The strategies used by SBVS to maximize ligand binding within the structure of the bioreceptor include molecular docking, molecular dynamics simulations, and structure-based pharmacophore modeling (Wang et al., 2020).

VS is highly knowledge-driven, relying on the quality and quantity of existing information about the system under investigation and its careful selection and preparation (Kirchweger & Rollinger, 2018). Limited access to 3D libraries of natural products (NPs) has restricted the number of virtual screening studies for bioactive NPs. Some notable studies include Liu and Zhou's identification of potential SARS-CoV protease inhibitors from marine and traditional Chinese medicine metabolites, Toney et al.'s discovery of the terpenoid alkaloid sabadinine as a potential anti-SARS agent, and Moro's identification of ellagic acid as a potent protein kinase CK2 inhibitor. Additionally, Zhao and Brinton's research on estrogen receptor-selective ligands highlighted the effectiveness of receptor-based molecular docking in identifying selective flavonoid compounds with high binding affinity (Rollinger et al., 2008). These examples underscore the importance and application of VS in advancing drug discovery, especially when leveraging high-quality data and sophisticated screening techniques.

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

Table 1: Computational Methods Applied in Virtual Screening Approaches (Santana et al., 2021)

Computational Method	Description	Examples
Cheminformatics Filters (Molecular Filters)	Applies empirical chemical rules to predict pharmacokinetics and drug-likeness properties of compounds by evaluating physicochemical and structural properties.	Lipinski's Rule of Five, Veber's Rule, Jeffrey's Filters
Molecular Fingerprint-Based Methods	Uses binary representations of chemical structures to quantitatively assess pairwise similarity of compounds. Enables efficient computational screening based on structural features and pharmacophore models.	SMILES fingerprint (SMIfp), Structural Interaction Fingerprint (SIFt)
Similarity and Distance Metrics	Metrics are used to compare molecular fingerprints for similarity assessment. Includes Tanimoto, Dice, Cosine coefficients, and others.	Tanimoto coefficient, Dice coefficient, Cosine similarity
Ligand-Based and Structure-Based Pharmacophore Modeling	Predicts biologically active compounds based on shared chemical features (ligand-based) or spatial arrangements within protein binding sites (structure-based).	LigandScout, Molecular Operating Environment (MOE), Pharmer
3D Shape-Similarity Search Methods	Searches for compounds with similar 3D molecular shapes are crucial for determining binding affinity and selectivity.	SHAFTS, Shape-it, OptiPharm
Machine Learning Algorithms	Uses intelligent algorithms to predict pharmacokinetic properties, toxicity, molecular targets, and bioactivity of compounds based on training data.	QSAR models, deep learning approaches

Molecular Dynamics simulations

Proteins and nucleic acids are two examples of biomacromolecules that have been extensively studied using molecular dynamics (MD) simulations. Recent developments have made it possible to study complete cells and run cellular-scale simulations to gain a deeper understanding of the basic molecular processes of life. It examines conformational shifts under various circumstances, drug-target interactions, and protein characteristics (Heidari et al., 2016b). In many biologically significant systems, molecular dynamics simulations can follow fast processes at atomic resolution in less than a millisecond (Borhani & Shaw, 2012). The dynamics and conformation of drug-target complexes can be studied with the use of MD modelling. To simulate biological processes in a computer program, MD simulation is employed. It has transformed drug development and turned into a standard computational tool for CADD. It provides precise estimations of the kinetics and thermodynamics of drug-target interactions and binding. It accurately estimates the thermodynamics and kinetics of drug-target interactions and binding. New methods, software, and hardware have increased the use of MD simulation in CADD research and the biopharmaceutical industry (Singh, 2021).

The molecular dynamics (MD) simulation workflow in drug discovery consists of a structured series of steps that accurately simulate molecular behaviour over time. The process starts with file format conversion to ensure compatibility with the simulation software. Next, a simulation box is built around the molecules of interest, followed by the addition of water molecules to simulate physiological conditions. The system is then neutralized to balance charges, and energy minimization is used to maintain the molecular arrangement. The heating and equilibration stages prepare the system for the final simulation phase of the production run. Following simulation, trajectory data is converted and analyzed to extract meaningful insights into molecular dynamics, assisting drug discovery efforts by elucidating interactions, stability, and other critical properties of compounds under investigation (Singh, 2021).

MD simulations provide detailed dynamic structural insights into biomacromolecules and detailed energetic profiles of protein-ligand interactions. They have proven invaluable in exploring disease mechanisms arising from protein misfolding and in virtual screening and understanding drug resistance due to target mutations. These complex issues often defy resolution through experimental approaches alone. As computational capabilities advance and with ongoing enhancements in sampling techniques, precise force field models, and streamlined analytical tools, MD simulations are poised to expand their applications across diverse fields in the coming years (Liu et al., 2018).

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

AI plays an important role in advancing protein folding prediction from amino acid sequences, leveraging deep learning and MD simulations. These technologies enhance understanding by simulating molecular interactions and dynamics more accurately. AI has also proven invaluable in predicting protein and peptide binding affinities and conducting toxicity studies using SAR and toxicological datasets (Dhakal et al., 2022). Molecular dynamics simulation calculations can be accelerated using machine learning techniques. By training on extensive protein-protein or protein-peptide datasets, AI models accurately forecast binding strengths, aiding in the selection or development of biologics with superior target affinity and specificity. In the context of SARS-CoV-2 research, high-throughput in silico methods, such as machine learning models, have been utilized to identify potential inhibitors of the virus's main protease from natural sources. For instance, the NuBBE database, which contains a diverse collection of natural compounds, was screened using AI models like Gradient Boosting Machine (GBM), Random Forest (RF), and Support Vector Machine (SVM). These models predicted numerous compounds with potential antiviral properties, and further molecular docking studies narrowed this down to specific flavonoids and lignoids with strong binding affinities. Subsequent MD simulations confirmed these compounds' stability and binding dynamics, revealing their mechanisms of action in inhibiting the main protease (Arifuzzaman et al., 2022). MD simulations enhance virtual screening by considering receptor flexibility, thereby improving the enrichment factor. Case studies include the use of MD simulations to optimize lead compounds and understand drug resistance mechanisms. For example, the interaction between hepatitis C virus proteins and inhibitors was studied to understand resistance due to mutations, guiding the development of more effective treatments (Liu et al., 2018).

Integrating molecular dynamics (MD) simulations with AI techniques like deep learning (DL) has significantly improved the accuracy and efficiency of structure-based drug design (SBDD). For SARS-CoV-2 treatment discovery, DL-based MD simulations screen large compound libraries and predict their binding affinities and stabilities, rapidly identifying promising drug candidates and analyzing their interaction dynamics (Sun et al., 2022). Researchers used a deep learning algorithm to screen 1611 natural compounds from the Selleck database, predicting 500 compounds with high binding affinity to the SARS-CoV-2 main protease (Mpro). Molecular docking and MD simulations refined these predictions, identifying Palmatine and Sauchinone as stable, non-toxic inhibitors with strong binding energies. These compounds showed promise as therapeutic agents against COVID-19, highlighting the effectiveness of combining deep learning and MD simulations for accelerating natural product-based drug discovery (Joshi et al., 2020).

Structure-Activity Relationship (QSAR) Modeling in Quantitative Terms

Since AI has emerged, quantitative structure-activity relationship (QSAR) modeling, a method used in computer-aided drug design for more than 60 years, has made tremendous strides (Tropsha et al., 2024). QSAR modelling is a key approach in cheminformatics that investigates how molecular features influence chemical, biological, and toxicological properties. It is commonly used for lead optimization in drug discovery research. QSAR modelling has expanded to include hit and lead discovery through virtual screening, drug-like property prediction, and chemical risk assessment. An improved model, development, validation protocols, and prioritizing external validation have enabled these advancements (Golbraikh et al., 2017). QSAR Modelling involves compiling and curating extensive datasets of natural compounds, calculating molecular descriptors, and applying machine learning algorithms to predict biological activities and toxicities. Modern QSAR models, enhanced by AI techniques such as deep learning, have proven effective in identifying promising drug leads with improved pharmacological profiles, as evidenced by numerous successful applications in designing better drugs from natural sources (Kar & Roy, 2012).

Data preparation is the first step in QSAR modeling. A relevant chemical, biological, or toxicological target must be identified first. Next, a suitable dataset must be compiled, curated, and descriptors selected and calculated. Finally, an appropriate machine-learning computational procedure must be chosen. The dataset is split into external evaluation and modelling sets multiple times to ensure robust analysis. Then, during the QSAR model development phase, the modelling set is split multiple times into training and test sets (Wang et al., 2024). Models are built using the training sets and validated with the test sets. These steps are repeated for each pair of descriptor sets and computational procedures in combinatorial QSAR (combi-QSAR). Models that exhibit appropriate statistical performance are selected for external validation, and a Y-randomization test is carried out to prevent overfitting and chance correlations. Validation of QSAR models involves performing consensus predictions on the external evaluation set within the applicability domain (AD). Optimal Z-cutoffs are determined using criteria for the precision of consensus prediction and coverage. Finally, virtual screening of chemical databases is conducted by performing similarity searches using the training or modelling sets with the Z-cutoff. The remaining compounds are then subjected to consensus prediction using the QSAR models (Golbraikh et al., 2017).

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

The capacity of QSAR modelling to predict the biological activities of different compounds has been demonstrated by its effective use in the search for antioxidants and anticancer medicines. Mitra et al. created QSAR models for flavonoids and phenolic compounds from traditional Chinese medicinal herbs in order to better understand antioxidants. They identified important structural characteristics, such as the ideal quantity of hydroxyl and ketonic groups that improve antioxidant activity (Mitra et al., 2010). Salum et al. (2011) studied discodermolide analogues for anticancer medications using a fragment-based QSAR technique, identifying crucial structural fragments and molecular features necessary for strong anticancer efficacy. De-Eknamkul et al. (2012) explored the estrogenic activities of isoflavonoids and diphenolics, identifying significant structural elements that contribute to their potency against breast cancer cells.

In order to address nonlinear relationships between chemical structures and their properties, QSAR first used simpler models like linear regression and k-nearest neighbours but has since moved to more sophisticated machine learning techniques, such as support vector machines (SVM) and gradient boosting methods (GBM) (Chen et al., 2018). With the introduction of deep learning, QSAR modelling has undergone even more transformation. Automatic feature extraction is now possible, and graph and recurrent neural networks are used to create context-specific representations of chemical structures. These models of deep learning, although sometimes criticized for their "black-box" nature and high computational cost, offer significant advantages in modelling complex molecular systems and multitask learning scenarios, which are crucial in the multiparameter optimization challenges of drug discovery (Soares et al., 2022). Recent advances in explainable AI and uncertainty estimation techniques are addressing the interpretability and reliability issues of deep learning models, making them more accessible and effective in predicting drug properties and interactions.

AI Application in Antibiotic and Anticancer Research

AI has shown great potential in the field of antibiotic and anticancer research, especially when combined with machine learning (ML) and deep learning (DL) approaches. AI-driven methods have expedited the development of antibiotics by making it easier to rationally design bioactive substances that work well in animal models (Melo et al., 2021). AI models have been used to predict binding site affinity more accurately than traditional methods, aiding in virtual screening processes. For example, deep learning methods have bypassed traditional docking and affinity estimation to identify small-molecule antibiotics active against multiple bacterial pathogens (Stokes et al., 2020). The application of graph neural networks and recurrent neural networks for molecular representations has enabled the automatic extraction of meaningful features from molecular structures, further enhancing the ability of these models to predict. Moreover, the development of antimicrobial peptides (AMPs) through AI has been highlighted, with machine learning models predicting antimicrobial activity and guiding the design of new AMP sequences with significant activity improvements (Porto et al., 2018).

ML goes beyond simple antimicrobial activity prediction to improve our comprehension of antibiotic therapeutic potential. Drug-likeness has been predicted more accurately and automatically over time, with an emphasis on important characteristics like toxicity, excretion, metabolism, distribution, and absorption (ADMET). Furthermore, ML can predict the adverse effects of antibiotics. This is demonstrated by its application in predicting the seizure-inducing potential of enoxacin, a broad-spectrum fluoroquinolone antibacterial (Gao et al., 2017). Combining machine learning and Fourier-transform infrared spectroscopy, Da Cunha et al. were able to identify biochemical fingerprints and precisely anticipate the antibiotics' modes of action and efficacy (Cunha et al., 2021). Zoffman et al. (2019) used machine learning to analyze the Roche compound library, prioritizing novel compounds and identifying their antibacterial activity against Gram-negative bacteria. Stokes et al. (2020) used a deep-learning approach to identify potential antimicrobial molecules from the Drug Repurposing Hub. After training and optimizing the model, they identified 99 potential antimicrobial molecules, 51 showing strong inhibitory effects on E. coli. This led to the discovery of halicin, which demonstrated strong inhibitory effects against multiple antibiotic-resistant strains, including M. tuberculosis and carbapenem-resistant Enterobacteriaceae, by sequestering iron and disrupting bacterial metabolism (Stokes et al., 2020).

Most ML models for NP activity prediction, particularly binary classification models, have emerged within the last 5-10 years. NPs are categorized as either active or inactive by these models. Using topological descriptors, an early linear discriminant analysis (LDA) model was able to find new anti-inflammatory nanoparticles (NPs) from MicroSource. Furthermore, among the 1194 marine and microbial NPs from the AntiMarin database, two random forest (RF) classifiers were created using CDK descriptors to find antibacterial and anticancer drugs. 5278 out of 21,334 plant-derived nanoparticles (NPs) from Traditional Chinese Medicine (TCM) were predicted by Dai and colleagues to have anticancer capabilities in 2016 (Saldívar-González et al., 2022). Dai et al.'s (2016) work predicting anticancer properties for plant-derived NPs using their CDRUG web server and the iterative stochastic elimination (ISE) optimization introduced by Rayan et al. (2017) for discovering bioactive NPs with anticancer

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

and antibacterial activities. Additionally, the 2020 development of the first DL-driven multi-classification algorithm identified medicinal uses of NPs across various diseases (Yoo et al., 2020).
Advances in AI biology analysis tools, like network-based and machine learning (ML)-based techniques, have made it possible for scientists to handle complicated biological data more efficiently. Gov et al. identified important biomarkers, including GATA2 and miR-124-3p, as possible treatment targets by reconstructing tissue-specific networks for ovarian cancer using network-based biology techniques (Gov et al., 2017). AI techniques have combined genomes, proteomics, and metabolomics data with other multi-omics data to reveal the complex relationships that drive the development of cancer. Additionally, AI-driven models like graph convolutional networks and autoencoders have been employed to predict drug properties and identify druggable targets, enhancing the precision and efficiency of drug discovery (You et al., 2022).

Peng et al. (2021) produced an innovative end-to-end learning framework called EEG-DTI, utilizing heterogeneous graph convolutional networks to predict drug-target interactions (DTIs). Remarkably, this model produced promising results even without relying on 3D structural data of drug targets. Madhukar et al. (2019) developed the BANDIT method that integrated six data types, achieving about 90% accuracy in target prediction for over 2000 small molecules. AI has been effectively applied in screening anticancer drug-hit compounds. High-throughput screening techniques enhanced by AI have been used to identify molecules with initial activity against specific targets. For example, Yasuo et al. introduced SIEVE-Score, a structure-based virtual screening method that significantly improved the efficiency of identifying hit compounds compared to traditional methods (Yasuo & Sekijima, 2019). Similarly, Krasoulis et al. (2022) developed DENVIS, a scalable algorithm using graphical neural networks for high-throughput screening, which showed superior speed and accuracy. These instances highlight the transformative potential of AI in accelerating the discovery and optimization of anticancer drugs, offering more effective and precise therapeutic options.

Challenges and Future Directions

Since a substantial amount of data is needed to train the system in the future, its availability is critical to AI's effectiveness. Access to data from many database providers may incur additional costs for a company in order to ensure accurate result prediction. Furthermore, the information needs to be accurate and of high quality. Further barriers that prevent AI from being fully applied in the pharmaceutical industry include the lack of skilled personnel to manage AI-based platforms, financial limitations for smaller companies, worries that replacing humans will lead to job losses, skepticism about the data generated by AI, and the "black box" phenomenon (i.e., the method by which the AI platform draws its conclusions) (Lamberti, 2019). While some pharmaceutical companies have already implemented AI, it is projected that the pharmaceutical sector will earn US$2.199 billion in revenue by 2022 via AI-based solutions. Between 2013 and 2018, the pharmaceutical business invested over US$7.20 billion in over 300 transactions. Pharmaceutical businesses should be transparent about the practical goals that can be achieved and the potential of AI technology to solve problems once it is used. Software engineers and data scientists with a strong foundation in AI technology and a solid understanding of the company's R&D goals and commercial target can be generated in order to fully utilize the potential of the AI platform (Research & Markets, 2019). NP-based drug development could be revitalized in both established and emerging fields by the technology advancements mentioned above. As previously noted, NPs have long been the primary source of novel medications, particularly antibiotics, to treat infectious disorders. NPs have antimicrobial qualities identified by utilizing the advancements detailed in the preceding sections, including methods for finding new NPs by utilizing the human microbiome (Samuel & Ekpan, 2023). Using developments in total synthesis, semi-synthetic approaches, and biosynthetic engineering, researchers are not only searching for novel NP classes with antimicrobial properties but also working to develop and optimize existing NP classes. Additionally, antivirulence tactics may offer a different strategy for combating infections, and NPs that target bacterial quorum sensing may be useful in this regard (Atanasov et al., 2021; Merit et al., 2024).

Ensuring high data quality is one of the main challenges to using AI for natural product-based medication development. The correctness and consistency of the data used to train machine learning (ML) and deep learning models determine their success to a considerable extent. Format, precision, and measurement standards of data from different sources might sometimes differ. Models with bias may result from incomplete datasets. Data gaps must be filled with the use of imputation techniques. To guarantee the predictive accuracy and generalizability of AI models used in drug discovery, validation is crucial. It is possible to detect overfitting or underfitting problems in a model by evaluating its performance on several data subsets using cross-validation techniques (Saldívar-González et al., 2022). Regulatory challenges also exist for the integration of AI in drug research. Regulatory bodies must well understand the rationale behind AI model judgments. To address this, interpretable AI models must be created, and the model-building process must be thoroughly documented.

Cite this article as:
Okibe, G. & Samuel, H. S. (2024). Harnessing artificial intelligence for the discovery and development of natural product-based therapeutics. *FNAS Journal of Scientific Innovations, 6*(1), 45-57.

AI can be greatly improved when combined with other omics (genomics, proteomics, metabolomics, etc.) technologies. Understanding biological processes and medication interactions more thoroughly is made possible by this combination. Combining data from several omics' levels can provide an integrated understanding of disease mechanisms and pharmacological activities. Building and evaluating biological networks, identifying important pathways, and forecasting the potential effects of changes in these pathways on illness and treatment results are all made possible with the use of artificial intelligence (Egwuatu et al., 2024; Paul et al., 2021; Ekpan et al., 2024).

Conclusion

Artificial intelligence (AI) has the potential to completely transform the field of drug development based on natural products. Drug discovery can now be conducted more precisely and efficiently thanks to AI approaches like machine learning (ML) and deep learning (DL), which have proven capable of processing massive datasets, predicting bioactivity, identifying protein targets, and optimizing extraction methods. Developing novel antibiotics and anticancer chemicals, optimizing already-existing therapeutic candidates, and forecasting drug toxicity and efficacy are successful AI uses in this field. Notwithstanding these developments, issues, including the requirement for carefully selected, high-quality datasets, workflow integration, and ethical and legal considerations, still exist. Researchers, technologists, and regulatory agencies must work together to explore further and collaborate to realize AI's potential in drug discovery fully. It is essential to focus on curating comprehensive and high-quality datasets and integrating multi-omics data, such as genomics, proteomics, and metabolomics, to achieve a more holistic understanding of biological systems. Developing accessible, open-source databases would promote broader collaboration across the scientific community. In order to address the "black box" nature of AI models, researchers should prioritize creating interpretable models that provide clear explanations for their predictions, increasing both their reliability and acceptance by regulatory bodies. Future research should also emphasize the integration of AI-driven predictions with experimental validation, streamlining AI insights with in vitro and in vivo testing to accelerate clinical applications.

References

Abdul, Q., Khan, S., & Uddin, S. (2021). Anticancer activity of natural compounds. Asian Pacific Journal of Cancer Prevention. https://doi.org/10.31557/APJCP.2021.22.S1.1

Arifuzzaman, M., Mohammadi, M., Hashem, F. R., Khan, M. F., Rashid, R. B., & Rashid, M. A. (2022). Identification of natural compounds with anti-SARS-CoV-2 activity using machine learning, molecular docking, and molecular dynamics simulation studies. *Dhaka University Journal of Pharmaceutical Sciences*, 21(1), 1–13. https://doi.org/10.3329/dujps.v21i1.60391

Atanasov, A. G., Zotchev, S. B., & Dirsch, V. M. et al. (2021). Natural products in drug discovery: Advances and opportunities. Nature Reviews Drug Discovery, 20, 200–216. https://doi.org/10.1038/s41573-020-00114-z

Berenger, F., Vu, O., & Meiler, J. (2017). Consensus queries in ligand-based virtual screening experiments. *Journal of Cheminformatics*, 9, 60. https://doi.org/10.1186/s13321-017-0248-5

Borhani, D. W., & Shaw, D. E. (2012). The future of molecular dynamics simulations in drug discovery. *Journal of Computer-Aided Molecular Design*, 26(1), 15–26. https://doi.org/10.1007/s10822-011-9517-4

Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today*, 23(6), 1241–1250. https://doi.org/10.1016/j.drudis.2018.01.039

Chen, X., & Xu, Y. (2020). Deep learning for natural product-based drug discovery. Drug Discovery Today, 25(9), 1542-1550.

Chen, Y., de Bruyn Kops, C., & Kirchmair, J. (2017). Data resources for the computer-guided discovery of bioactive natural products. Journal of Chemical Information and Modeling, 57(9), 2099-2111. https://doi.org/10.1021/acs.jcim.7b00341

Cunha, B. R., Fonseca, L. P., & Calado, C. R. C. (2021). Simultaneous elucidation of antibiotics' mechanism-of-action and potency with high-throughput Fourier-transform infrared spectroscopy and machine learning. *Applied Microbiology and Biotechnology, 105*(3), 1269–1286. https://doi.org/10.1007/s00253-020-11061-9

De-Eknamkul, W., Shotipruk, A., & Tadtong, S. (2012). QSAR study of the estrogenic activity of isoflavonoids and diphenolics. *Journal of Molecular Graphics and Modelling*, 38, 93–100. https://doi.org/10.1016/j.jmgm.2012.08.006

Dhakal, A., McKay, C., Tanner, J. J., & Cheng, J. (2022). Artificial intelligence in the prediction of protein–ligand interactions: Recent advances and future directions. *Briefings in Bioinformatics*, 23(2), bbab476. https://doi.org/10.1093/bib/bbab476

Egwuatu, O. P., Ori, M. O., Samuel, H. S., & Ekpan, F. M. (2024). AI-enabled diagnostics and monitoring in nanomedicine. Eurasian Journal of Science & Technology, 4(3), 208–229. https://doi.org/10.48309/ejst.2024.426725.1116

Ekpan, F. M., Ori, M. O., Samuel, H. S., & Egwuatu, O. P. (2024). The synergy of AI and drug delivery: A revolution in healthcare. International Journal of Advanced Biological and Biomedical Research, 12(1), 45–67. https://www.ijabbr.com/article_709788.html

Ernst, P., Siu, A., & Weikum, G. (2015). KnowLife: A versatile approach for constructing a large knowledge graph for biomedical sciences. BMC Bioinformatics, 16, 157. https://doi.org/10.1186/s12859-015-0549-5

Gao, H., & Ding, Y. (2019). Convolutional neural networks for bioactivity prediction of natural compounds. Bioinformatics, 35(15), 2554-2561.

Gao, M., Igata, H., Takeuchi, A., Sato, K., & Ikegaya, Y. (2017). Machine learning-based prediction of adverse drug effects: An example of seizure-inducing compounds. *Journal of Pharmacological Sciences, 133*(2), 70–78. https://doi.org/10.1016/j.jphs.2017.01.004

Garcia-Hernandez, C., Fernández, A., & Serratosa, F. (2019). Ligand-based virtual screening using graph edit distance as molecular similarity measure. *Journal of Chemical Information and Modeling*, 59(7), 1410-1421. https://doi.org/10.1021/acs.jcim.8b00820

Gaudêncio, S. P., Bayram, E., Lukić, B. L., Cueto, M., Díaz-Marrero, A. R., Haznedaroglu, B. Z., Jiménez, C., Mandalakis, M., Pereira, F., Reyes, F., & Tasdemir, D. (2023). Advanced methods for natural products discovery: Bioactivity screening, dereplication, metabolomics profiling, genomic sequencing, databases and informatic tools, and structure elucidation. Marine Drugs. https://doi.org/10.3390/md21050308

Golbraikh, A., Wang, X. S., Zhu, H., & Tropsha, A. (2017). Predictive QSAR modeling: Methods and applications in drug discovery and chemical risk assessment. In *Handbook of computational chemistry* (pp. 2303–2340). https://doi.org/10.1007/978-3-319-27282-5_37

Gov, E., Kori, M., & Arga, K. Y. (2017). Multiomics analysis of tumor microenvironment reveals Gata2 and miRNA-124-3p as potential novel biomarkers in ovarian cancer. *OMICS: A Journal of Integrative Biology, 21*(9), 603–615. https://doi.org/10.1089/omi.2017.0137

Han, R., Yoon, H., Kim, G., Lee, H., & Lee, Y. (2023). Revolutionizing medicinal chemistry: The application of artificial intelligence (AI) in early drug discovery. Pharmaceuticals, 16(1259). https://doi.org/10.3390/ph16091259

Heidari, Z., Roe, D. R., Galindo-Murillo, R., Ghasemi, J. B., & Cheatham, T. E. (2016). Using wavelet analysis to assist in identification of significant events in molecular dynamics simulations. *Journal of Chemical Information and Modeling*, 56(7), 1282–1291. https://doi.org/10.1021/acs.jcim.6b00262

Jadhav, A., Gupta, V., & Kulkarni, A. (2020). Machine learning in natural product-based drug discovery. Journal of Chemical Information and Modeling, 60(7), 3361-3372.

Jagadevappa, P. (2016). Utilization of novel technologies in natural product research: Radical progress. Journal of Pharmacovigilance. https://doi.org/10.4172/2329-6887.1000E152

Jiménez-Luna, J., Grisoni, F., Weskamp, N., & Schneider, G. (2021). Artificial intelligence in drug discovery: Recent advances and future perspectives. Expert Opinion on Drug Discovery, 16(9), 949-959. https://doi.org/10.1080/17460441.2021.1909567

Joshi, M., Melo, D. P., Ouyang, D., Slomka, P. J., Williams, M. C., & Dey, D. (2023). Current and future applications of artificial intelligence in cardiac CT. Current Cardiology Reports, 25(3), 109-117.

Joshi, T., Joshi, T., Pundir, H., Sharma, P., Mathpal, S., & Chandra, S. (2020). Predictive modeling by deep learning, virtual screening, and molecular dynamics studies of natural compounds against SARS-CoV-2 main protease. *Journal of Biomolecular Structure and Dynamics*. https://doi.org/10.1080/07391102.2020.1802341

Kar, S., & Roy, K. (2012). QSAR of phytochemicals for the design of better drugs. *Expert Opinion on Drug Discovery*, 7(10), 877–902. https://doi.org/10.1517/17460441.2012.716420

Kim, H., Kim, E., Lee, I., Bae, B., Park, M., & Nam, H. (2020). Artificial intelligence in drug discovery: A comprehensive review of data-driven and machine learning approaches. Biotechnology and Bioprocess Engineering, 25, 895-930. https://doi.org/10.1007/s12257-020-0049-y

Kinnings, S. L., Liu, N., & Buchmeier, N. (2011). Drug discovery using chemical systems biology: Identifying the protein targets of natural products. Journal of Proteome Research, 10(10), 4488-4497.

Kirchweger, B., & Rollinger, J. M. (2018). Virtual screening for the discovery of active principles from natural products. In *Natural products as a source of molecules with therapeutic potential* (pp. 333-364). https://doi.org/10.1007/978-3-030-00545-0_9

Krasoulis, A., Antonopoulos, N., Pitsikalis, V., & Theodorakis, S. (2022). DENVIS: Scalable and high-throughput virtual screening using graph neural networks with atomic and surface protein pocket features. *Journal of Chemical Information and Modeling, 62*(18), 4642–4659. https://doi.org/10.1021/acs.jcim.2c00682

Kumar, G. (2023). Natural products and their analogues acting against Mycobacterium tuberculosis: A recent update. Drug Development Research, 84(5), 779-804. https://doi.org/10.1002/ddr.22063

Lamberti, M. J. (2019). A study on the application and use of artificial intelligence to support drug development. *Clinical Therapeutics, 41*(8), 1414–1426. https://doi.org/10.1016/j.clinthera.2019.04.013

Liu, X., Shi, D., Zhou, S., Liu, H., Liu, H., & Yao, X. (2018). Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery*, 13(1), 23–37. https://doi.org/10.1080/17460441.2018.1403419

Madhukar, N. S., Khade, P. K., Huang, L., Gayvert, K., Galletti, G., Stogniew, M., Allen, J. E., Giannakakou, P., & Elemento, O. (2019). A Bayesian machine learning approach for drug target identification using diverse data types. *Nature Communications, 10*(1), 5221. https://doi.org/10.1038/s41467-019-13137-5

Maia, E. H. B., Assis, L. C., de Oliveira, T. A., da Silva, A. M., & Taranto, A. G. (2020). Structure-based virtual screening: From classical to artificial intelligence. *Frontiers in Chemistry*, 8, 343. https://doi.org/10.3389/fchem.2020.00343

Melo, M. C. R., Maasch, J. R. M. A., & de la Fuente-Nunez, C. (2021). Accelerating antibiotic discovery through artificial intelligence. *Communications Biology*, 4, 1050. https://doi.org/10.1038/s42003-021-02586-0

Mitra, I., Saha, A., Roy, K., & Kar, S. (2010). Quantitative structure–activity relationship (QSAR) modeling for prediction of antioxidant activity of phenolic compounds and flavonoids from traditional Chinese medicinal plants. *Chemosphere*, 78(6), 831–837. https://doi.org/10.1016/j.chemosphere.2009.10.029

Mullowney, M. W., Duncan, K. R., & Elsayed, S. S. (2023). Artificial intelligence for natural product drug discovery. Nature Reviews Drug Discovery, 22, 895-916. https://doi.org/10.1038/s41573-023-00774-7

Oliveira, T. A., Silva, M. P., Maia, E. H., Silva, A. M., & Taranto, A. G. (2023). Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. *Drugs and Drug Candidates*, *2*(2), 311-334. https://doi.org/10.3390/ddc2020017

Ordóñez-Robles, M., Santos-Beneit, F., & Martín, J. F. (2018). Unraveling nutritional regulation of tacrolimus biosynthesis in Streptomyces tsukubaensis through omic approaches. Antibiotics, 7(2), 39.

Ori, M. O., Ekpan, F. M., Samuel, H. S., & Egwuatu, O. P. (2024). Integration of artificial intelligence in nanomedicine. Eurasian Journal of Science & Technology, 4(2), 88–104. https://doi.org/10.48309/EJST.2024.422419.1105

Padma, K. R., & Don, K. R. (2022). Natural products possess bioactive agents investigated for their anticancer potential: Medicinal importance of natural products. In Handbook of research on natural products and their bioactive compounds as cancer therapeutics (pp. 153-168). IGI Global.

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. Drug Discovery Today, 26(1), 80–93. https://doi.org/10.1016/j.drudis.2020.10.010

Peng, J., Wang, Y., Guan, J., Li, J., Han, R., Hao, J., Wei, Z., & Shang, X. (2021). An end-to-end heterogeneous graph representation learning-based framework for drug-target interaction prediction. *Briefings in Bioinformatics, 22*(3), bbaa430. https://doi.org/10.1093/bib/bbaa430

Porto, W. F., Pires, Á. S., Franco, O. L., & Albuquerque, M. M. (2018). In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. *Nature Communications, 9*(1), 1490. https://doi.org/10.1038/s41467-018-03732-w

Rayan, A., Raiyn, J., & Falah, M. (2017). Nature is the best source of anticancer drugs: Indexing natural products for their anticancer bioactivity. *PLOS ONE, 12*(11), e0187925. https://doi.org/10.1371/journal.pone.0187925

Research and Markets. (2019). Global Growth Insight - Role of AI in the pharmaceutical industry 2018–2022: Exploring key investment trends, companies-to-action, and growth opportunities for AI in the pharmaceutical industry. *Research and Markets.*

Rollinger, J. M., Stuppner, H., & Langer, T. (n.d.). Virtual screening for the discovery of bioactive natural products. In *Natural compounds as drugs volume I* (pp. 211-249). https://doi.org/10.1007/978-3-7643-8117-2_6

Saldívar-González, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L., & Plisson, F. (2022). Natural product drug discovery in the artificial intelligence era. *Chemical Science, 13*(6), 1526-1546. https://doi.org/10.1039/D1SC05679A

Saldívar-González, F. I., Huerta-García, C. S., & Medina-Franco, J. L. (2020). Chemoinformatics-based enumeration of chemical libraries: A tutorial. *Journal of Cheminformatics*, 12, 64. https://doi.org/10.1186/s13321-020-00466-z

Salum, L. B., Andricopulo, A. D., & da Silva, A. B. F. (2011). Fragment-based QSAR modeling of discodermolide analogs: Discovery of novel anticancer agents. *Journal of Medicinal Chemistry*, 54(19), 6373–6385. https://doi.org/10.1021/jm200832x

Samuel, H. S., & Ekpan, F. M. (2023). Revolutionizing drugs administration: Techniques in drug delivery and development. International Journal of Biochemistry & Physiology, 8(2), Article 000237. https://doi.org/10.23880/ijbp-16000238

Santana, K., Do Nascimento, L. D., Damasceno, V., Nahum, C., Braga, R. C., & Lameira, J. (2021). Applications of virtual screening in bioprospecting: Facts, shifts, and perspectives to explore the chemo-structural diversity of natural products. *Frontiers in Chemistry*, 9, 662688. https://doi.org/10.3389/fchem.2021.662688

Sati, P., Sharma, E., Dhyani, P., Attri, D. C., Rana, R., Kiyekbayeva, L., ... & Sharifi-Rad, J. (2024). Paclitaxel and its semi-synthetic derivatives: Comprehensive insights into chemical structure, mechanisms of action, and anticancer properties. European Journal of Medical Research, 29(1), 90.

Shultz, M. D. (2019). Two decades under the influence of the rule of five and the changing properties of approved oral drugs. Journal of Medicinal Chemistry, 62(4), 1701-1714. https://doi.org/10.1021/acs.jmedchem.8b00686

Singh, I. P., Ahmad, F., Chatterjee, D., Bajpai, R., & Sengar, N. (2021). Natural products: Drug discovery and development. https://doi.org/10.1007/978-981-15-5534-3_2

Singh, S. K. (Ed.). (2021). *Innovations and implementations of computer aided drug discovery strategies in rational drug design*. https://doi.org/10.1007/978-981-15-8936-2

Soares, T. A., Nunes-Alves, A., Mazzolari, A., Ruggiu, F., Wei, G. W., & Merz, K. M. (2022). The (Re)-evolution of quantitative structure–activity relationship (QSAR) studies propelled by the surge of machine learning methods. *Journal of Chemical Information and Modeling*, 62(22), 5317–5320. https://doi.org/10.1021/acs.jcim.2c01208

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., & et al. (2020). A deep learning approach to antibiotic discovery. *Cell, 180*(4), 688–702. https://doi.org/10.1016/j.cell.2020.02.017

Sun, Y., Jiao, Y., Shi, C., & Zhang, Y. (2022). Deep learning-based molecular dynamics simulation for structure-based drug design against SARS-CoV-2. *Computational and Structural Biotechnology Journal*, 20, 5014–5027. https://doi.org/10.1016/j.csbj.2022.09.002

Tiwari, P., Pal, R., Chaudhary, M., & Nath, R. (2023). Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. Drug Development Research, 84, 1652-1663. https://doi.org/10.1002/ddr.22115

Tropsha, A., Isayev, O., & Varnek, A. (2024). Integrating QSAR modeling and deep learning in drug discovery: The emergence of deep QSAR. *Nature Reviews Drug Discovery*, 23, 141–155. https://doi.org/10.1038/s41573-023-00832-0

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. https://doi.org/10.1038/s41573-019-0024-5

Visan, A. I., & Negut, I. (2024). Integrating artificial intelligence for drug discovery in the context of revolutionizing drug delivery. Life, 14(2), 233.

Wang, Z., Sun, H., Shen, C., Hu, X., Gao, J., & Li, D. (2020). Combined strategies in structure-based virtual screening. *Physical Chemistry Chemical Physics*, 22(5), 3149–3159. https://doi.org/10.1039/C9CP06303J

Wang, Q., Fan, T., Jia, R., Zhang, N., Zhao, L., Zhong, R., & Sun, G. (2024). First report on the QSAR modelling and multistep virtual screening of the inhibitors of nonstructural protein Nsp14 of SARS-CoV-2: Reducing unnecessary chemical synthesis and experimental tests. Arabian Journal of Chemistry, 17(3), 105614. https://doi.org/10.1016/j.arabjc.2024.105614

Wu, H., Wang, Y., Yuan, L., Mao, Y., Wang, W., Zhu, L., ... & Zhang, B. (2016). Inactivation of SACE_3446, a TetR family transcriptional regulator, stimulates erythromycin production in Saccharopolyspora erythraea. Synthetic and Systems Biotechnology, 1(1), 39-46.

Wu, Z., Jiang, C., & Qiu, Y. (2019). Predictive modeling of bioactivity for natural products. Bioinformatics, 35(5), 826-833.

Yasuo, N., & Sekijima, M. (2019). Improved method of structure-based virtual screening via interaction-energy-based learning. *Journal of Chemical Information and Modeling, 59*(3), 1050–1061. https://doi.org/10.1021/acs.jcim.8b00892

Yoo, S., Yang, H. C., Lee, S., Shin, J., Min, S., Lee, E., & et al. (2020). A deep learning-based approach for identifying the medicinal uses of plant-derived natural compounds. *Frontiers in Pharmacology, 11*, 584875. https://doi.org/10.3389/fphar.2020.584875

You, Y., Lai, X., Pan, Y., Xu, J., Han, W., Chen, H., & et al. (2022). Artificial intelligence in cancer target identification and drug discovery. *Signal Transduction and Targeted Therapy, 7*(1), 156. https://doi.org/10.1038/s41392-022-00994-0

Zhang, C. (2023). Foundations of artificial intelligence and machine learning. https://doi.org/10.4337/9781803926179.00009

Zhao, L., & Li, H. (2018). Natural language processing for mining bioactivity data from scientific literature. Journal of Biomedical Informatics, 82, 60-71.

Zhu, F., & Shi, Z. (2021). Extracting protein-compound interactions from literature using NLP techniques. Bioinformatics, 37(13), 1892-1898.

Zoffmann, S., Vercruysse, M., Benmansour, F., Maunz, A., Wolf, L., Marti, R. B., Heckel, T., Ding, H., Truong, H. H., Prummer, M., & et al. (2019). Machine learning-powered antibiotics phenotypic drug discovery. *Scientific Reports, 9*(1), 5013. https://doi.org/10.1038/s41598-019-41352-1