ISSN: 3067-2317
Research Article

World Journal of Artificial Intelligence and Robotics Research

Modern Chemoinformatics Methods and Approaches in Designing Analogue Natural Products

Piotr Michałowski[1,2*], Maciej Matysik[1], Katarzyna Nowak[1] & Katarzyna Bielicka – Daszkiewicz[1]

[1]*Faculty of Chemical Technology, Poznań University of Technology, Institute of Chemical Technology and Engineering, Berdychowo 4, 60-965 Poznań, Poland*

[2]*Institute of Bioorganic Chemistry Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznań, Poland*

[*]**Corresponding author:** Piotr Michałowski, Faculty of Chemical Technology, Poznań University of Technology, Institute of Chemical Technology and Engineering, Berdychowo 4, 60-965 Poznań, Poland.

Submitted: 30 September 2025 **Accepted:** 07 October 2025 **Published:** 14 October 2025

https://doi.org/10.63620/MK.WJAIRR.2025.1019

Citation: *Michałowski, P., Matysik, M., Nowak, K., & Bielicka-Daszkiewicz, K. (2025). Modern Chemoinformatics Methods and Approaches in Designing Analogue Natural Products. Wor Jour of Arti inte and Rob Res, 2(5), 01-11.*

Abstract

Chemoinformatics is an interdisciplinary field combining chemistry, informatics, biotechnology, and data analysis, enabling the analysis and interpretation of chemical information using databases, algorithms, and computational methods. Its primary applications are found in the pharmaceutical industry, where it allows for the preliminary selection of chemical compounds with high biological potential, reducing both the cost and time associated with traditional drug discovery processes. This review covers the process of designing analogues of natural products, from the selection of organisms for screening (random, ethnopharmacological, chemosystematic, ecological, and computational approaches), through extraction, isolation, and identification of active compounds, to biological testing and optimization. Modern chemoinformatics tools are also discussed, including molecular modeling, protein–ligand docking, QSAR, and the prediction of ADME properties and bioavailability. The findings demonstrate that the integration of chemistry, biology, and bioinformatics enables rapid identification and modification of chemical structures with therapeutic potential, reducing the need for costly experimental studies. The analysis of natural products and their secondary metabolites provide valuable inspiration for the design of new drugs, particularly in the context of increasing drug resistance. The conclusions indicate that chemoinformatics is an essential tool in modern medicinal chemistry, facilitating the efficient design of natural product analogues with improved pharmacokinetic and therapeutic properties, while simultaneously accelerating the drug discovery process.

Keywords: Chemoinformatics Method, Natural Products, Secondary Metabolites, Designing Analogues.

Introduction

Chemoinformatics is an interdisciplinary science that fuses a variety of scientific fields, including chemistry, informatics, data analysis and biotechnology, among others. The methodology employed encompasses the utilisation of algorithms and databases for the analysis and interpretation of chemical information. In addition, a range of computational methods are employed for the purposes of visualisation and the initial prediction of possible outcomes [1]. One of the most common applications of chemoinformatics is found in the pharmaceutical industry. The conventional approach to drug discovery is characterised by its protracted nature and the necessity for substantial financial investment. Consequently, it necessitates a series of rigorous evaluations, encompassing bioactivity and various other properties, including toxicity and kinetic analysis. The efficacy of this process is not assured, as evidenced by numerous studies [2, 3]. Conversely, chemoinformatics enables experimentalists to eliminate certain materials at an early stage, thus enabling them to focus on those with greater potential for applicability. One such example is the development of machine-learning models for Zika virus proteins, first reported in 1947 [4]. Furthermore, chemoinformatics has facilitated the development of in silico models, which have been instrumental in identifying novel molecules with potential antimalarial properties [5].

A further aspect meriting inclusion in any discussion of chemo-informatics is medicinal chemistry, which is defined as a multidimensional science at the interface of medicine, chemistry, pharmacology and biology. The field of medicinal chemistry is generally concerned with the identification of new pharmaceutical compounds for therapeutic use [6, 7]. However, the scope of medicinal chemistry extends beyond this initial scope. It has been identified as a pivotal instrument in the global effort to eradicate poverty in developing countries. It has been established that a correlation exists between unsatisfactory sanitation conditions, food insecurity and the absence of safe, effective and affordable medicines in numerous locations. This has resulted in a pernicious cycle of poverty. Medicinal chemistry facilitates the development of innovative and accessible therapeutic interventions, contributing to the near-elimination of several diseases, including dracunculiasis, lymphatic filariasis and trachoma in numerous countries. It is accomplishing this objective in the context of addressing the escalating issue of drug resistance. Another pertinent example is that of the eradication of schistosomiasis, a condition which is endemic in numerous countries and for which the current treatment is praziquantel, a drug that rapidly paralyses the schistosome worms [8].

Prior to the analysis of the design of analogue natural products, it is imperative to first comprehend the concept of what a natural product is. The definition of natural products (NPs) is as follows: they are any chemical substance found in a living organism. They play a crucial role in pharmacology and other industries due to their bioactivity as well as in drug discovery [9, 10]. In addition to NPs, the secondary metabolites of these organisms have been found to demonstrate significant potential. Metabolites are defined as small molecules that act as intermediate products of metabolic processes found in all living organisms. These have been identified as promising biomarkers for a range of diseases and conditions, including lung cancer [11]. It is possible to estimate the intake of certain substances, including the daily intake of nicotine from cigarette smoking, due to the existence of this data [12]. Secondary metabolites represent a distinct category of natural products that have been shown to detect oxidative damage resulting from stress. These hormones are synthesised under typical adverse conditions in response to environmental stress, indicating the organism's ability to adapt to changing conditions [13]. The utilisation of these substances is evident in a variety of industrial sectors, including but not limited to pharmaceuticals, dyes and food production. Furthermore, a range of methodologies have been devised to augment the synthesis of secondary metabolites in plants (Fig.1.), given their role in activating their defence mechanisms [14]. Furthermore, the presence of certain secondary metabolites has been demonstrated to possess significant pharmacological and therapeutic potential for humans, including anticancer properties. Consequently, they make substantial contributions to both the natural world and human health.

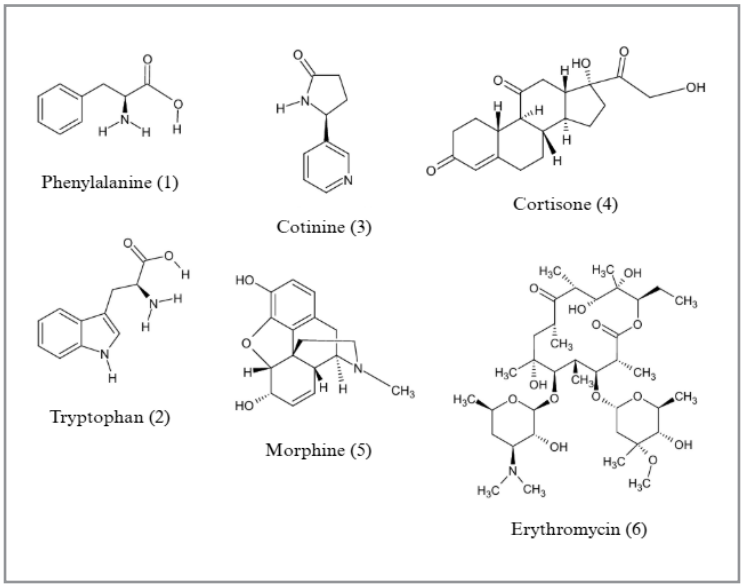

Figure 1: Examples of Secondary Metabolites in Plants

Despite extensive research in this area, there remains no official classification of NPs and metabolites, nor any established nomenclature. However, at present, they are diversified by the taxonomy rank of the organisms in which they occur [15]. In the context of such a division, the following distinctions are made: terpenes, terpenoids, alkaloids, flavonoids, nonribosomal peptides and polyketides.

Terpenes (Fig.2.) are obtained from the plant and tree kingdoms. Furthermore, certain terpenes have been shown to reduce inflammatory symptoms [16]. Examples of such compounds include clerodane, 18β-glycyrrhetinic acid, lupeol and ursolic acid [17]. Volatile terpenes are being explored as a potential alternative source of energy for the petrochemical industry. However, a significant number of terpenes are also used as flavouring and fragrance agents in the food and cosmetic industries. Examples of such substances include lutein, menthol, carvone and citral [18]. Terpenoids, conversely, exhibit structural and chemical variety, albeit to a limited extent. These elements are present in both fungi and plants. Terpenoids are frequently employed in the pharmaceutical, cosmetic, and food industries. Examples in pharmacology include artemisinin, an antimalarial drug, and taxol, an anticancer drug [19].

Alkaloids (Fig.2.) are naturally occurring chemical compounds that are found in a variety of plants. They are characterised by the presence of nitrogen atoms in their molecular structure. These agents have been demonstrated to function as a defensive measure against pathogens [20]. Alkaloids are utilised in medicine for their analgesic, antiasthmatic, anticancer, antihypertensive,

antipyretic and antihyperglycemic properties. Examples of such substances include morphine, ephedrine, vincristine, reserpine and quinine [21].

Flavonoids (Fig.2.), a class of phytonutrients, are found in fruit, vegetables and plant-derived foods. They have antioxidant and anti-inflammatory properties [22]. Flavonoids represent a class of naturally occurring dietary molecules that have been shown to possess potential in the prevention of both the process of aging itself and the associated diseases [23]. Furthermore, evidence has been provided to demonstrate their ability to regulate key cellular enzyme function. Anthocyanins, which are categorised as a subgroup, are frequently employed as food colourants. Conversely, isoflavones, a subgroup of flavonoids, have been demonstrated to possess significant antioxidant properties. These properties have been shown to contribute to a reduction in the risk of cancer by hindering the damage to DNA induced by free radicals [24]. Examples of such substances include genistein, resveratrol, quercetin and silymarin [25].

Nonribosomal peptides (Fig.2.) are microbial secondary metabolites which exhibit a tremendous structural diversity and a broad range of biological activities that are useful in the medical and agro-ecological fields [26] (found in bacteria and fungi). These compounds are typically utilised as antibacterial agents, immunosuppressants, and antitumour compounds in pharmaceutical formulations. Examples of such agents include penicillin, vancomycin, bleomycin and cyclosporine. [27]. Polyketides (Fig.2.) are found in bacteria, fungi, plants and lower ranks of animals, including mollusks and sponges. It is evident that the subjects demonstrate notable structural diversity, which in turn engenders a broad spectrum of functions, including anti-cholesterol, antifungal, and anticancer properties. Examples of such agents include tetracycline, doxorubicin, rapamycin and amphotericin B. Polyketides have been employed in the treatment of cancer, for example in the form of doxorubicin [28].

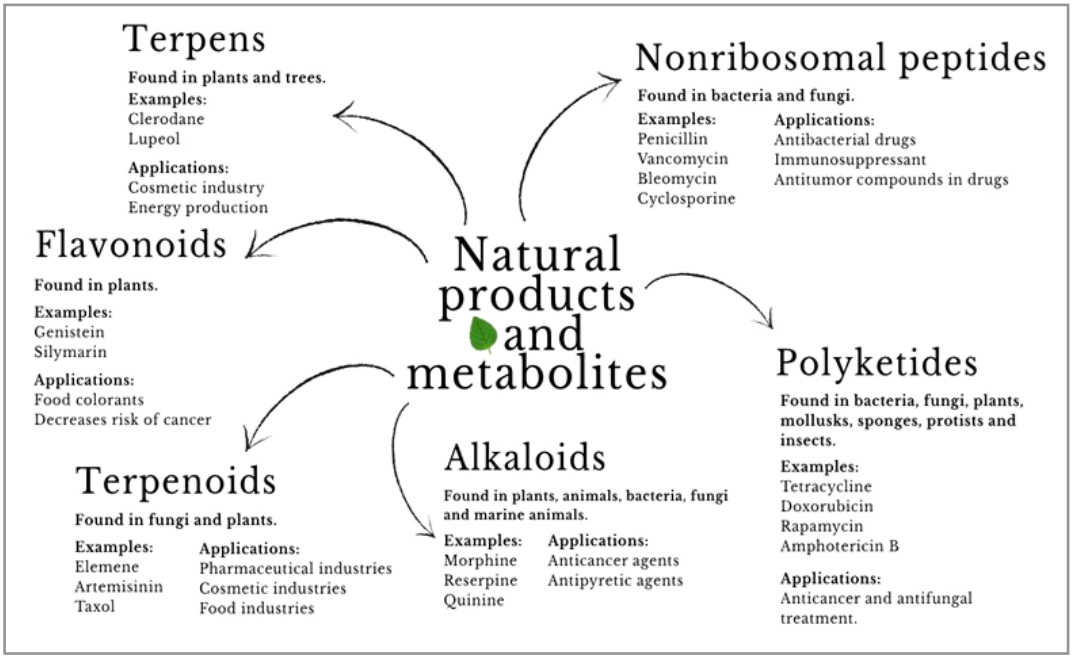

Figure 2: Division of NPs and Metabolites with their Occurrences

The objective of the present review is to provide a comprehensive overview of contemporary chemoinformatic methodologies employed in the design of natural product analogues. This innovative approach represents a significant advancement in the pursuit of novel, effective bioactive compounds. The integration of chemistry, bioinformatics and pharmacology facilitates the expeditious identification and modification of chemical structures with therapeutic potential. The employment of molecular modelling and machine learning facilitates the reduction of costly experimental research. The work under consideration places particular emphasis on the significance of natural products and their secondary metabolites as a source of inspiration for the development of new drugs, particularly in the context of an increase in drug resistance.

Modern Chemoinformatics Methods

Process of Designing Analogues of Natural Products

The development of natural products analogues is a multifaceted process that encompasses numerous stages and is, in relative terms, a complex procedure. The methods employed can vary significantly, contingent on the approach and the techniques utilised, with these being informed by the characteristics of the output compound. The overall stages are distinguished as follows: selection of organisms for screening, authentication, extraction, isolation and structure elucidation.

A number of approaches have been posited for the selection of organisms for screening purposes. These include random, ethnopharmacological, chemosystematic, ecological and computational methods [29].

In a random approach, organisms such as plants, fungi or bacteria are selected at random; however, the key factor is their availability. Conversely, within the ethnopharmacological approach, the selection of substances is determined by their historical usage and application. In the context of a chemosystematic approach, the selection of compounds is informed by chemotaxonomy and phylogeny. This is attributable to the fact that certain

plant families are known to produce compounds or compound classes that possess specific properties and demonstrate particular bioactivity. In accordance with an ecological approach, the selection of these entities is based on their interaction with their environment, whilst also considering the ecological functions of secondary metabolites. In the computational approach, the selection of the subjects is made on the basis of their predicted in silico bioactivity.

Irrespective of the used approach, a chosen model always requires authentication. The first step of it should be identifying botanical origin and the scientific name. Later steps should involve authentication through comparing organoleptic properties, like structure and shape, and odour, with known data, which can be done by many methods, such as: macroscopic, microscopic, chromatographic and spectroscopic techniques, or DNA fingerprinting analysis.

Recently, methods using bioactivity and physical properties, such as chromatographic techniques, are being applied frequently to extract and isolate active compounds. There are two main approaches to obtaining natural products by extraction. The parallel approach, which is used when the process of selecting organisms was based on its ethnopharmacological applications. It is a three-stage process that contains: extracting at least three extracts from biomaterial, purifying the most active extract to fractions, and isolation of compounds from the most active fraction by appropriate chromatographic technique.

Sequential approach is used when selection of model organism is random and bioactivity remains unknown, and is a two-stage process, in which we can distinguish: extracting extracts from material into fractions, and isolating the most active compounds from the most active fraction. After isolating active compounds, further identification is needed. It can be achieved through mass spectrometry (MS) and comparing obtained data with the one of known compounds. Characterisation of the compound structure can be carried out by analytical methods, especially various spectroscopic methods, such as: FT-IR spectroscopy, NMR spectroscopy, or mass spectrometry. However, in order to understand the whole structure of the compound, more methods should be applied, like X-ray diffraction or optical rotatory dispersion [30].

Biological Screening

The next step involves using the obtained isolates for biological screening to determine the bioactivity of the active compound. Ethnopharmacological knowledge at this stage can simplify the process by providing guidance on the potential effects of the compound. Biological screening can be defined as a general testing of the compound in search of its bioactivity. However, it is imperative to acknowledge the significance of adopting the appropriate technique for the effective screening of both therapeutic agents and potential adverse effects [31]. The following screening methods are employed: high-throughput screening (HTS), in vitro screening, cell culture-based screening and phenotypic screening. Adopting the appropriate technique can be a complex matter. However, the following aspects of the process should be given due consideration: the target effect, the type of compound that is being tested, and the desired scale of screening [32]. It is evident that a plethora of intriguing applications of biological screening can be identified in the domain of toxicology. In this particular field, there has been a notable surge in the utilisation of in vitro models for the assessment of compound toxicity. This development signifies a significant shift from conventional in vivo animal models and underscores the potential for more efficacious and ethical research methodologies [33].

Modern medicinal chemistry approaches utilise bioactive natural products to create new analogues that are less toxic and have better pharmacokinetic profiles. Designing and developing new analogues (Fig. 3) can be summarised in a few steps: in silico ligand construction and preparation; target preparation; docking; identification of the hit molecule; and optimisation of the hits.

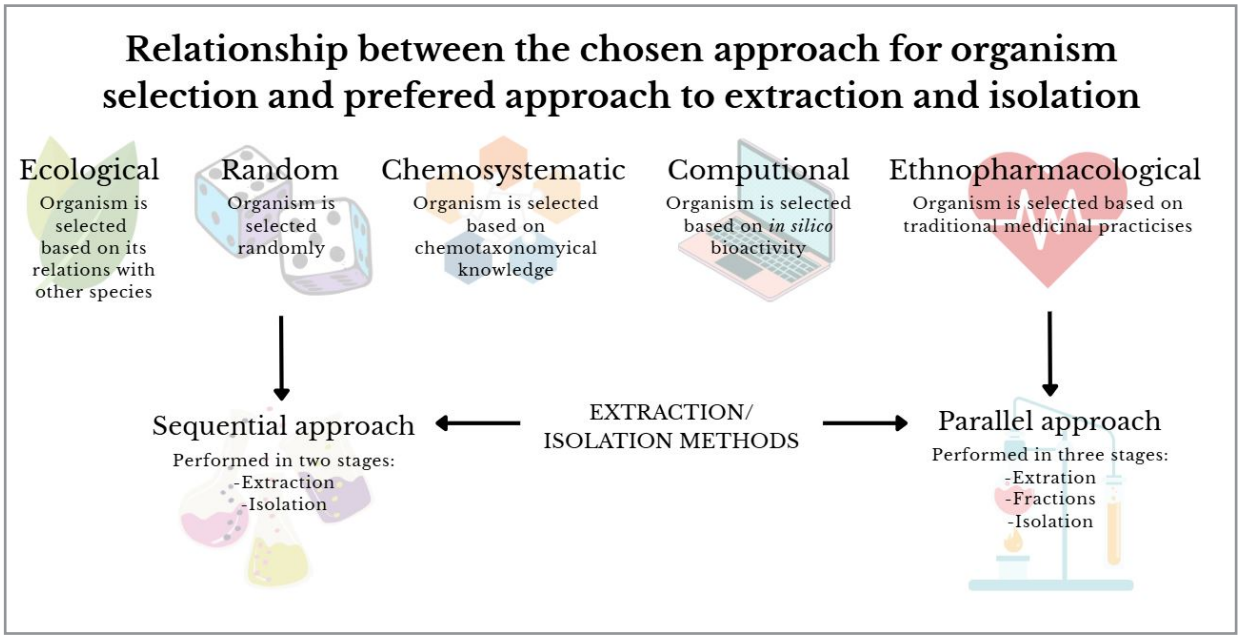

Figure 3: Process of Development Natural Products Analogues

Ligand - Protein Relation

The process of drug discovery is an investment that is both costly and time-consuming, and one that is also challenging. Computer-aided drug discovery (CADD) has the potential to facilitate the process, thereby reducing both the financial cost and the time required for development. It is an indispensable tool in the field of therapeutic development. It is evident that a number of computational methodologies have demonstrated their efficacy in the domains of drug discovery and pipeline development. In this segment, the following aspects will be discussed: structure-based drug discovery (SBDD), protein structure prediction methods and protein-ligand docking [34].

The SBDD approach is employed in instances where the three-dimensional structure of a drug target is known. In the context of SBDD, two frequently employed methodologies are molecular docking approaches and de novo ligand design. Molecular dynamics (MD) are frequently employed to facilitate comprehension of ligand-target protein binding, the mechanisms of interaction, and target flexibility. SBDD methodologies have been instrumental in facilitating the progression of numerous compounds through clinical trials and securing FDA approval for market release. The protease of HIV-1 (Human Immunodeficiency Virus 1) is the primary target for anti-AIDS therapeutic agents. Saquinavir and Amprenavir were both synthesised with the intention of targeting HIV-1 protease using SBDD methods. An additional example is dorzolamide, a carbonic anhydrase II inhibitor, which is utilised in the treatment of glaucoma.

Proteins are typically perceived as static structures, despite the fact that they are dynamic systems that display internal motions. It is evident that target flexibility is frequently being ignored. It is imperative to elucidate the target flexibility, given that their structures undergo modification during the binding process. The rigidity of the target is not realistic and can result in misleading outcomes. In order to account for target flexibility, two approaches may be considered: induced fit docking methods or ensemble-based screening methods.

The most common methods employed for predicting protein structures are X-ray crystallography and NMR spectroscopy. It is evident that experimental methods of this nature are contingent upon cost and time limitations. For instance, X-ray crystallography is only possible when a protein target can be crystallized, and a certain amount of them is difficult to crystallize. Conversely, nuclear magnetic resonance (NMR) spectroscopy has been found to be exclusively applicable to smaller proteins. The utilisation of computational methodologies, such as SBDD, to quantify structures from sequences has been demonstrated to effectively address the sequence-structure discrepancy. A plethora of methodologies have been employed for the purpose of predicting protein structures. These include the use of homology modelling, threading approaches and ab initio folding.

Information regarding drug molecules and target structures, such as proteins, is of paramount importance in SBDD tools. The following section details several databases that contain such information. PubChem, a repository for small molecules, is accessible via the National Institutes of Health (NIH) and contains millions of biologically relevant small molecules. ZINC, a database of virtual compounds, contains over 35 million molecules. It is estimated that DrugBank contains approximately 5,000 small molecules, of which over 800 have been approved by the FDA. The Protein Databank (PDB) is a global resource that contains a substantial amount of three-dimensional information about experimentally determined biological macromolecules. It is evident that one of the structures contained within the Protein Data Bank (PDB) is that of a protein-ligand complex. The structure under consideration contains approximately 120,000 biological macromolecular structures, as well as 20,000 bound ligand molecules. Swiss-Prot is a database containing essential protein sequences. These sequences are manually annotated with descriptors, including functional information and post-translational modifications. The BIND database is a comprehensive repository of protein complex information and biomolecular interactions. BindingDB has measured binding affinity information of proteins considered to be targets for drugs. Furthermore, it has been demonstrated that the substance under scrutiny contains in excess of one million binding points.

Whilst docking is underway, it is imperative to predict how intermolecular complexes are formed between a target and a ligand. The objective of these algorithms is to identify the most optimal target-ligand poses. Despite the utilisation of computational methods, the process of docking a target structure to a molecule remains a challenging procedure. Notwithstanding the target flexibility being disregarded, a considerable number of methodologies persist by which a target may be docked. The implementation of docking algorithms is contingent upon the prior identification of a target protein structure and a potential drug binding site, as well as the determination of small molecules that bind to this site. The identification of these molecules is facilitated by the use of small molecule libraries. The central tenet of molecular docking is the prediction of the binding mode and binding affinity of a protein-ligand complex. The process of obtaining and evaluating thousands of possible protein-ligand binding poses is a key aspect of the methodology. The one with the lowest energy is considered to be the best.

In order to achieve optimal screening results, it is imperative to undertake meticulous target and ligand preparations prior to the docking process. In the context of experimental methods, the hydrogen atoms of the structures are not typically present. However, their presence, along with the location of the bonds, is crucial for the efficacy of docking algorithms. Furthermore, the utilisation of target protein structures in the absence of preliminary processing can give rise to a number of potential issues, including the absence of residues, atom clashes, crystallographic waters, and alternate locations.

In target preprocessing, missing atoms, like hydrogen, are added to remove atom clashes. In ligand preprocessing, on the other hand, ligand three-dimensional geometries are predicted. Protonation states of the structures also play a key role in docking poses, since they influence how ligands bind to the binding site.

SPROBE is an exemplary program used for preprocessing proteins for protein–ligand docking. LigPrep, on the other hand, makes it possible to obtain all-atom, 3D structures of ligands. Another program worth mentioning is PRODRG, a web-based ligand topology generating server. It allows to generate 3D coordinates for ligands, which, compared to other methods, are of

equal or even better quality.

Molecular Flexibility

Molecular flexibility is defined as the factor that quantifies the size and complexity of a molecule's conformational space (given the solvent, temperature and pressure). It significantly contributes to many molecular properties and is necessary to reduce a conformational ensemble to meaningful representatives [35]. It is primarily related to the presence of single bonds in a molecule, typically carbon sp^3 – carbon sp^3 bonds. Molecular flexibility increases with the number of single bonds [36].

The calculation technique can vary depending on the compound being worked on. The most common methods are the Rotatable Bond Count and the Kier Phi Index. The major drawback of the Rotatable-Bond count method is the requirement for a clear and ideally easily computable definition of which bonds are rotatable. The Kier ϕ index, on the other hand, provides a continuous description of flexibility based on information from the molecular graph. While it is an improvement on the number of rotatable bonds, the Kier ϕ index does not resolve all issues. For example, it is ineffective when estimating the size of the conformational space of a molecule. Unlike other flexibility metrics, Torsion Angular Bin Strings (TABS) make it possible to directly estimate the upper bound of the number of conformers. This is a highly desirable feature in the field of conformer generation.

Qsar Analysis

Quantitative Structure-Activity Relationships (QSAR) is the term given to this concept. QSAR modelling represents a major computational approach to drug discovery, and its ability to identify bioactive compounds, as well as novel compounds, has grown significantly [37]. In the field of QSAR methods, it is hypothesised that structurally similar molecules tend to demonstrate analogous biological activity. In two-dimensional quantitative structure-activity relationship (QSAR) methods, there is a direct correlation between biological activity and physical and chemical properties, including electronic, hydrophobic and steric features of compounds. Conversely, in 3D QSAR methods, not only are physical and geometric features of active drug molecules introduced, but also quantum chemical features. In the domain of ligand-based drug design (LBDD), a predominant alternative to structure-based drug design (SBDD), QSAR represents a computational approach that models the correlation between structural characteristics of ligands that bind to a target and the resultant biological activity. The success of QSARs is contingent on the selected descriptors, as well as the ability of these models to predict biological activity. The absence of sufficient activity data for the extraction of patterns invariably results in the failure of a QSAR model. It is important to note that a particular structural element within QSAR warrants particular consideration; namely, the volume of the binding site. The binding pocket volume exerts a considerable influence on the biological activity. Once this feature has been ascertained, the process of eliminating molecules that are too large to fit into the binding pocket is facilitated. This procedure can be carried out in the early stages of the drug discovery process.

In the domain of epigenetic drug discovery, QSAR modelling plays a pivotal role. As chemogenomics datasets for epigenetic targets increase in size, the application of QSARs to the detection of new bioactive compounds has also increased. For instance, BET inhibitors have been the subject of study through the utilisation of QSAR modelling approaches. In a particular study, these models were utilised to predict six potential multi-target bromodomain inhibitors. In a separate study, lysine methyltransferase DOTL1 inhibitors were proposed on the basis of 3D-QSAR, molecular docking and dynamics studies. This resulted in the computer-assisted design of two compounds that displayed inhibition at micromolar levels in confirmatory assays.

QSAR models function on the basis of statistics that connect activities of target drug interactions with various molecular descriptors. These models provide a mathematical description of the activity response of a target, which binds a ligand, as compared to the structural features of the ligand. QSAR information is gathered by calculating the correlation between experimentally defined biological activity and several properties of small ligand binders. Activity and descriptor data are obtained from a known drug molecule to build a mathematical QSAR model so that the descriptors can predict the activity of each molecule.

Furthermore, QSAR relationships may be implemented to predict the activity of new drug molecule analogs. The most common methods of quantifying the activity of a drug molecule are the use of the inhibitory concentration (IC_{50}) and the inhibition constant (Ki). The distinguishing feature of QSAR models is their capacity to discern both the positive and negative effects of a specific feature of a drug molecule on its activity. QSAR methods have been successfully employed with various drug targets, including carbonic anhydrase, thrombin and renin.

Statistical methodologies have been incorporated within the framework of linear QSAR to select molecular descriptors that are instrumental in predicting biological activity. Multivariable linear regression (MLR) is utilised to identify molecular descriptors exhibiting a satisfactory correlation with target-ligand biological activity. The utilisation of MLR is contingent upon the assumption that the activity-descriptor relation is linear, a supposition that is not universally valid. It is evident that a number of machine learning (ML) approaches have been developed for the purpose of generating QSAR models, with the objective of addressing the non-linear fitting issue.

The following examples illustrate drugs that have been developed through the utilisation of ligand-based drug discovery methodologies: As indicated in, zolmitriptan is utilised as a migraine treatment, norfloxacin is employed in the management of urinary tract infections, and losartan is prescribed for hypertension. ToxCast and External-AA are two such datasets that are utilised in the development of QSAR models, with the objective of enhancing their efficacy [38].

Table 1: Values of molecular descriptors [39].

- A. Percentage of molecular descriptors that obey the Lipinski's rule of five (Ro5)
- B. Percentage of molecular descriptors that obey Veber's rule
- C. Percentage based on combined count of hydrogen bond donors and acceptors

Molecular decsriptor	Min	Median	Max	Average	Ro5 (a)	Veber (b)
MW [Da]	94.12	390.29	1057.11	390.76 ± 2.28	85.4%	-
nHDon [-]	0	2.	16	2.21 ± 0.03	99.2%	83.5% (c)
nHAcc [-]	1	7	27	6.77 ± 0.05	96.6%	-
cLogP [-]	-5.95	2.58	7.94	2.60 ± 0.03	94.1%	-
SA [Å2]	77.39	289.53	761.02	289.47 ± 1.69	-	92.9%
TPSA [Å2]	16.61	94.06	457.72	95.64 ± 0.72	-	96.7%
RBN [-]	0	5	32	4.77 ± 0.06	-	-
ARR [-]	0	0.58	1	0.59 ± 0.00	-	-
nAR [-]	0	3	8	3.04 ± 0.02	-	-
Nsp3 [-]	0	6	31	6.38 ± 0.09	-	-
nSK [-]	7	28	74	27.81 ± 0.17	-	-
Fsp3 [-]	0	0.30	1.36	0.31 ± 0.00	-	-

The Following Molecular Descriptors can be Divided into two Groups.

- Group 1 (bioavailability group): Molecular weight (MW), number of donor atoms for hydrogen bonds (nHD), number of acceptor atoms for hydrogen bonds (nHA), calculated partition coefficient between octanol and water (clogP), number of rotatable bonds (RBN) and topological polar surface area (TPSA).
- Group 2 (binding affinity group): Number of aromatic rings, aromatic ratio, fraction of sp^3 carbon atoms, and hydrogen bond counts (nHD and nHA).

Since the descriptors in Group 1 are directly associated with bioavailability, 'drug-likeness' rules should be considered. According to Lipinski's rule of five, drug candidates should possess the following biophysical properties for better bioavailability: a molecular weight of less than 500 Da; a calculated logP of less than five; five or fewer hydrogen bond donors; and 10 or fewer hydrogen bond acceptors. Veber's rule for bioavailability states that a drug candidate should have 10 or fewer rotatable bonds and fewer than 12 hydrogen bond donors or acceptors in total. TPSA demonstrates the sum of the surfaces of all polar atoms in a molecule and can be used to predict drug absorption and transport properties. Molecules with a TPSA greater than 140 $Å^2$ perform poorly at permeating cell membranes.

Five-member heterocycles containing between one and four heteroatoms, that is, nitrogen, oxygen or sulfur, have found their application in antibacterials used in therapy [40]. Heteratoms contribute significantly to nanoporous carbon, which is employed for energy storage thanks to its properties such as high surface area, hierarchical porosity and exceptional electrochemical properties. These distinctive properties can be combined with the individual doping of heteroatoms, such as sulfur, nitrogen, oxygen and boron, to achieve high energy storage capacity and stability [41, 42].

The number of aromatic and non-aromatic rings is a significant factor in determining the drug-like properties of compounds [43]. It is noteworthy that a considerable proportion of chemicals exhibit partial or complete aromatic character, attributable to the ring framework undergoing enhanced stabilisation through the delocalisation of π-electrons. It is generally accepted that the stability of compounds such as benzene is attributable to the equal numbers of π-electrons in their aromatic rings. Dipotassium cyclopentagallate represents a distinctive instance of a five-membered aromatic ring that is stabilised exclusively by two π-electrons, a configuration that is ordinarily uncommon and constrained. This compound provides compelling evidence that aromatic stabilization is significantly more complex than previously assumed, with a minimum of π-electrons required in a five-atom ring fragment [44].

Protein kinase inhibitors (PKIs) are a therapeutic modality employed in the treatment of cancer-associated diseases. A chemoinformatic analysis was conducted on 2,139 PKIs, which resulted in the following findings: PKIs are flat molecules with high aromatic ring counts. Furthermore, a linear relationship has been identified between the number of aromatic rings and the average of weighted hydrogen count.

ADME Properties Prediction

The term ADME properties is an acronym for the four processes of absorption, distribution, metabolism and excretion. It is imperative to acknowledge the significance of ADME properties in determining pharmacokinetic profiles, a factor that exerts a substantial influence on both the efficacy and safety of substances [45]. These elements are of paramount importance, as they serve to determine the viability of a drug candidate. In the context of pre-clinical and clinical research, the potential for minimising the occurrence of deleterious effects in animal models and human subjects is enhanced by the judicious filtration of ADME properties of drug candidates during the preliminary phases of drug development. This tool has been demonstrated to be of significant benefit in the field of pharmacokinetics, with studies showing that it can facilitate faster, safer and more effective drug development [46].

The ADME properties are pivotal in determining the biological processes that are integral to the assessment of a drug's phar-

macokinetic (PK) parameters. These parameters have a direct impact on the efficacy and safety of drugs [47]. The process of developing a bioactive compound into a pharmaceutical drug involves a series of rigorous and varied methodologies. A range of factors must be given full consideration, including but not limited to pharmacokinetics, drug interactions, efficacy and safety. The financial investment required is substantial. It has been established that ADME properties are the primary factor contributing to the unsuccessful progression of bioactive compounds towards new drug status. Consequently, the prediction of these properties is of paramount importance. In the domain of drug discovery, the acknowledgement of ADME properties is instrumental in ensuring the safety of a drug, as well as its interactions with other medications. This is particularly pertinent in the context of drug-drug interactions (DDI). A range of computational methodologies have been utilised for the purpose of predicting ADME properties, including machine learning and artificial intelligence (AI). In the domain of drug discovery, both methods are employed to enhance chemical libraries, prioritise hits from biological screens and optimise the ADME properties of lead molecules. Machine learning (ML) has become an indispensable tool in predicting ADME properties. Machine learning (ML) algorithms are trained on molecular structures and the associated ADME assay data to develop quantitative structure-property relationship (QSPR) models [48, 49].

The scientific study of drugs that are absorbed, distributed, metabolised and excreted (Fig. 4) in a body has seen notable advancements over the years. The advent of PBPK modelling can be attributed to the findings of these studies. PBPK models have been demonstrated to be effective in predicting the behaviour of drugs in various populations. PBPK modelling is a sophisticated computational approach that is employed to predict the ADME properties of drugs in the human body. The software facilitates the creation of a virtual human body that replicates the behaviour of drugs under various scenarios. PBPK models have been instrumental in preventing extensive animal testing and human trials. The employment of models facilitates the delivery of personalised treatment regimens, with considerations including patient age, hepatic function, and genetic variations. Furthermore, they facilitate the assessment of potential interactions between drugs, particularly in instances where they are metabolised by the same enzymes.

QikProp is an ADME programme that predicts physically crucial and pharmaceutically valid descriptors for small drug-like molecules. VolSurf is a valuable instrument employed for the calculation of ADME properties and the generation of ADME models. These models can be utilised subsequently to predict the behaviour of novel molecules or to identify molecules with analogous ADME properties. FAF-Drugs2 is a tool designed for the screening of ADME and toxicity properties. It is capable of calculating physicochemical properties, as well as toxic and unstable groups, and the key functional components.

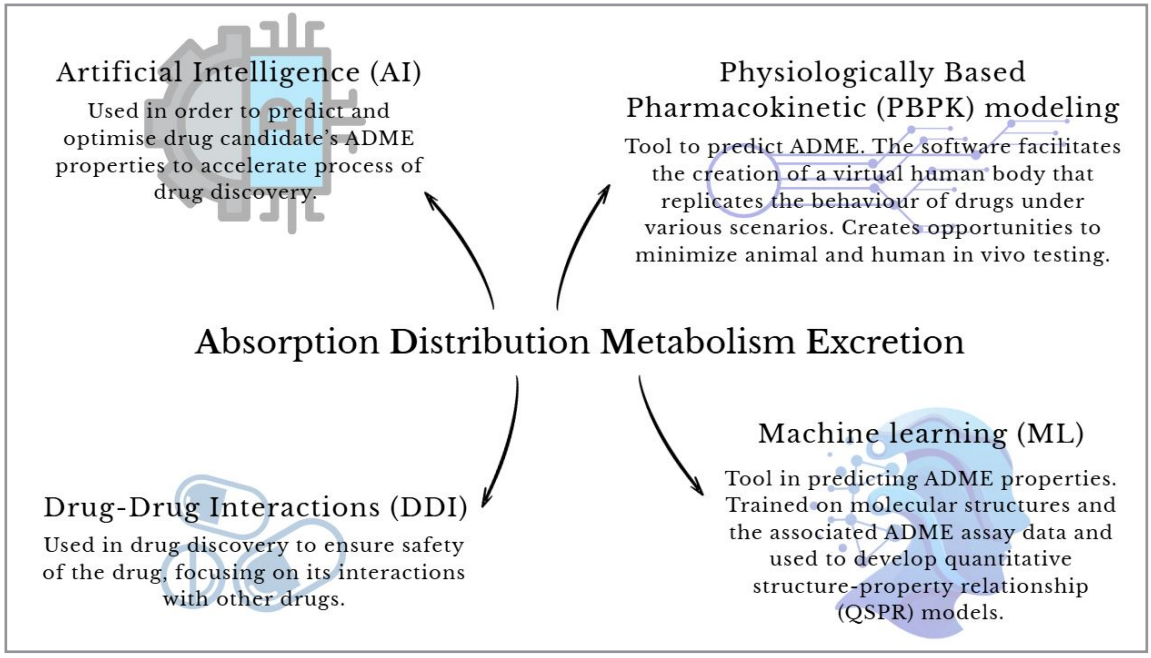

Figure 4: ADME Opportunities and Applications

Bioavailability is defined as the rate and proportion of the administered dose (by any route of administration) which enters the bloodstream in the form of an active substance/drug (unaltered), which then serves its purpose at the site of action. Bioavailability is influenced by a multitude of factors. These include the physicochemical properties of the medication itself, its potential interactions with other substances, and its ADME properties.

It is hypothesised that when an active pharmaceutical substance is administered intravenously, it will exhibit 100% bioavailability. However, when the medication is delivered differently, then bioavailability (F) is defined as the mass of the drug delivered to the plasma divided by the total mass of the drug administered.

The formula for calculating the percentage of the drug that is delivered to the plasma is as follows:
F = mass of the drug delivered to the plasma ÷ total mass of the drug administered
The importance of bioavailability in medicinal chemistry is illustrated by macrocyclization, a process that regulates drug-like properties while ensuring adequate bioavailability. Macrocycles

are of significant importance due to their structural diversity and ability to reach high selectivity and affinity towards demanding targets that are frequently not addressable by small molecules [50].

In addition, the term 'drug-likeness' is typically defined in the context of Lipinski's rule-of-five (Ro5) and ADME application [51]. The concept of drug-likeness provides significant guidelines in terms of drug discovery at early stages [52]. With the exception of Ro5, the degree of drug-likeness can be determined by means of a quantitative estimate of drug-likeness (QED). The system is characterised by its intuitive nature, transparency, and straightforwardness, which renders it highly applicable in a wide range of practical settings. The values of QED range from zero to one. In this context, zero signifies all properties being unfavourable, whereas one signifies that all properties are favourable. The combination of the individual desirability functions (d) into the QED is possible by taking the geometric mean of the individual functions. The calculation of QED can be conducted in accordance with the following equation:

$$\mathrm{QED} = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln d_i\right) = \left(\prod_{i=1}^{n} d_i\right)^{1/n} \quad (1)$$

QED – Quantitative Estimate of Drug-likeness;
n – number of physicochemical properties considered in the calculation;
di – normalized value of the i-th property relative to its optimal range;

The ability to define the parameters that define drug-likeness is a common skill amongst medicinal chemists. For instance, the combination of PSA <75 Å2 with LogP >3 has been demonstrated in several studies from large pharmaceutical companies. The aforementioned combination has been demonstrated to increase the risk of in vivo toxicity [53].

Conclusion

The findings from the analyses unambiguously demonstrate that the integration of chemoinformatics with medicinal chemistry, bioinformatics and pharmacology currently constitutes the foundation for the rational design of novel bioactive compounds. The interdisciplinary nature of this approach enables the concurrent utilisation of computational methods, databases, spectroscopic techniques and in silico models to expedite the drug discovery process while diminishing experimental expenses. It is imperative that both natural products (NPs) and their secondary metabolites demonstrate significant pharmacological potential, encompassing anticancer, anti-inflammatory, and antiparasitic properties. The structural and functional diversity of NPs renders them a valuable source of inspiration for medicinal chemistry. However, the absence of a uniform classification of these substances, coupled with the challenges inherent in nomenclature, underscores the imperative for the establishment of global standards. Such standards are deemed essential for facilitating the exchange of knowledge and the conduct of systematic research.

Another significant aspect pertains to the utilisation of computer-aided methods (CADD), encompassing structures derived from molecular docking, QSAR modelling and ADME property prediction. These methodologies facilitate the evaluation of the bioactivity and safety of potential pharmaceutical agents in the early stages of research. The efficacy of this methodology is demonstrated by the use of structural modelling in the development of HIV-1 protease inhibitors or antimalarial drugs. Of particular importance in this context is the issue of molecular flexibility and conformational dynamics of proteins. These have hitherto been marginalised, yet it is becoming evident that they are crucial for realistic modelling of ligand-receptor interactions. The employment of contemporary machine learning algorithms and artificial intelligence in predicting pharmacokinetic properties and identifying structural determinants of biological activity signifies a landmark advancement in personalising therapy and mitigating clinical failures.

The potential for future research and innovation lies in the integration of high-throughput technologies, large-scale modelling, and the integration of omic data. This integration will, in the future, enable the creation of personalised compound libraries and the dynamic design of natural product analogues for specific populations or even individuals. Another significant area of research is the development of digital pharmacology, which employs PBPK and QSPR models to facilitate virtual simulations of drug behaviour within the body. These simulations account for genetic factors, age, comorbidities and drug interactions. The utilisation of secondary plant metabolites in the creation of innovative biopharmaceuticals holds significant promise, as does their application in other industries, including energy (bioterpenes as alternative fuel sources) and materials engineering (heteroatom nanostructures in energy storage).

In view of these findings, it can be concluded that the future of chemoinformatics and medicinal chemistry will be shaped by hybrid research strategies in which artificial intelligence algorithms, nanotechnology and large-scale databases will be integrated into a single, coherent ecosystem of innovative drug discovery, focused on speed, efficiency and ethicality of the process. This approach has the potential to become not only a breakthrough in precision medicine, but also a tool for combating global health problems such as antibiotic resistance, tropical diseases and lifestyle diseases.

Credit Authorship Contribution Statement

This research was carried out without third-party authors' participation. Piotr Michałowski, Maciej Matysik, Katarzyna Nowak contributed to all aspects of this research paper, including conceptualisation, data collection, analysis, and writing. Katarzyna Bielicka - Daszkiewicz contributed to analysis, writing, and final approval of the manuscript. This statement underscores that the four authors are responsible for the entirety of the research process, from conceptualization to the final draft.

Conflict of Interest Statement

The author asserts that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kim, S., Bucholtz, E. C., Briney, K., Cornell, A. P., Cuadros, J., Fulfer, K. D., ... & Belford, R. E. (2020). Teaching cheminformatics through a collaborative intercollegiate on-

line chemistry course (OLCC). Journal of chemical education, 98(2), 416-425.
2. NeoVarsity. (2025, September 22). Cheminformatics beginner's guide. https://neovarsity.org/blogs/cheminformatics-beginners-guide#Role-of-Cheminformatics
3. Chen, Y., & Kirchmair, J. (2020). Cheminformatics in natural product-based drug discovery. Molecular Informatics, 39(12), 2000171.
4. Gonzalez-Ponce, K., Horta Andrade, C., Hunter, F., Kirchmair, J., Martinez-Mayorga, K., Medina-Franco, J. L., ... & Zdrazil, B. (2023). School of cheminformatics in Latin America. Journal of Cheminformatics, 15(1), 82.
5. Bosc, N., Felix, E., Arcila, R., Mendez, D., Saunders, M. R., Green, D. V., ... & Leach, A. R. (2021). MAIP: a web service for predicting blood-stage malaria inhibitors. Journal of Cheminformatics, 13(1), 13.
6. Veale, C. G. (2021). Into the Fray! A Beginner's Guide to Medicinal Chemistry. ChemMedChem, 16(8), 1199-1225.
7. Gioiello, A., Piccinno, A., Lozza, A. M., & Cerra, B. (2020). The medicinal chemistry in the era of machines and automation: recent advances in continuous flow technology. Journal of medicinal chemistry, 63(13), 6624-6647.
8. Martinengo, B., Diamanti, E., Uliassi, E., & Bolognesi, M. L. (2025). Medicinal Chemistry: A Key Driver in Achieving the Global Sustainable Development Goals. Journal of Medicinal Chemistry, 68(7), 6916-6931.
9. Rutz, A., Sorokina, M., Galgonek, J., Mietchen, D., Willighagen, E., Gaudry, A., ... & Allard, P. M. (2022). The LOTUS initiative for open knowledge management in natural products research. elife, 11, e70780.
10. Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: collection of open natural products database. Journal of Cheminformatics, 13(1), 2.
11. Lee, K. B., Ang, L., Yau, W. P., & Seow, W. J. (2020). Association between metabolites and the risk of lung cancer: a systematic literature review and meta-analysis of observational studies. Metabolites, 10(9), 362.
12. Benowitz, N. L., St. Helen, G., Nardone, N., Cox, L. S., & Jacob III, P. (2020). Urine metabolites for estimating daily intake of nicotine from cigarette smoking. Nicotine and Tobacco Research, 22(2), 288-292.
13. Li, Y., Zhang, Y., He, X., Guo, Z., Yang, N., Bai, G., ... & Xu, D. (2024). The mitochondrial blueprint: unlocking secondary metabolite production. Metabolites, 14(12), 711.
14. Reshi, Z. A., Ahmad, W., Lukatkin, A. S., & Javed, S. B. (2023). From nature to lab: A review of secondary metabolite biosynthetic pathways, environmental influences, and in vitro approaches. Metabolites, 13(8), 895.
15. van Santen, J. A., Poynton, E. F., Iskakova, D., McMann, E., Alsup, T. A., Clark, T. N., ... & Linington, R. G. (2022). The Natural Products Atlas 2.0: a database of microbially-derived natural products. Nucleic acids research, 50(D1), D1317-D1323.
16. Del Prado-Audelo, M. L., Cortés, H., Caballero-Florán, I. H., González-Torres, M., Escutia-Guadarrama, L., Bernal-Chávez, S. A., ... & Leyva-Gómez, G. (2021). Therapeutic applications of terpenes on inflammatory diseases. Frontiers in Pharmacology, 12, 704197.
17. Raimundo, V. D., Carvalho, R. P. R., Machado-Neves, M., & de Almeida Marques-da-Silva, E. (2022). Effects of terpenes in the treatment of visceral leishmaniasis: A systematic review of preclinical evidence. Pharmacological Research, 177, 106117.
18. Ninkuu, V., Zhang, L., Yan, J., Fu, Z., Yang, T., & Zeng, H. (2021). Biochemistry of terpenes and recent advances in plant protection. International Journal of Molecular Sciences, 22(11), 5710.
19. Chen, R., Wang, M., Keasling, J. D., Hu, T., & Yin, X. (2024). Expanding the structural diversity of terpenes by synthetic biology approaches. Trends in Biotechnology, 42(6), 699-713.
20. Letchuman, S., Madhuranga, H. D. T., Madhurangi, B. L. N. K., Premarathna, A. D., & Saravanan, M. (2024). Alkaloids unveiled: A comprehensive analysis of novel therapeutic properties, mechanisms, and plant-based innovations. Intelligent Pharmacy.
21. Aryal, B., Raut, B. K., Bhattarai, S., Bhandari, S., Tandan, P., Gyawali, K., ... & Parajuli, N. (2022). Potential therapeutic applications of plant-derived alkaloids against inflammatory and neurodegenerative diseases. Evidence-Based Complementary and Alternative Medicine, 2022(1), 7299778.
22. Zilli, A. M. H., & Zilli, E. M. (2021). Review of evidence and perspectives of flavonoids on metabolic syndrome and neurodegenerative disease. Protein and Peptide Letters, 28(7), 725-734.
23. Fan, X., Fan, Z., Yang, Z., Huang, T., Tong, Y., Yang, D., ... & Yang, M. (2022). Flavonoids—Natural gifts to promote health and longevity. International journal of molecular sciences, 23(4), 2176.
24. Chen, S., Wang, X., Cheng, Y., Gao, H., & Chen, X. (2023). A review of classification, biosynthesis, biological activities and potential applications of flavonoids. Molecules, 28(13), 4982.
25. Hassani, S., Maghsoudi, H., Fattahi, F., Malekinejad, F., Hajmalek, N., Sheikhnia, F., ... & Ghorbanpour, M. (2023). Flavonoids nanostructures promising therapeutic efficiencies in colorectal cancer. International journal of biological macromolecules, 241, 124508.
26. Duban, M., Cociancich, S., & Leclère, V. (2022). Nonribosomal peptide synthesis definitely working out of the rules. Microorganisms, 10(3), 577.
27. Süssmuth, R. D., & Mainz, A. (2017). Nonribosomal peptide synthesis principles and prospects. Angewandte Chemie International Edition, 56(14), 3770-3821.
28. Risdian, C., Mozef, T., & Wink, J. (2019). Biosynthesis of polyketides in Streptomyces. Microorganisms, 7(5), 124.
29. Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., ... & Stuppner, H. (2015). Discovery and resupply of pharmacologically active plant-derived natural products: A review. Biotechnology advances, 33(8), 1582-1614.
30. Najmi, A., Javed, S. A., Al Bratty, M., & Alhazmi, H. A. (2022). Modern approaches in the discovery and development of plant-based natural products and their analogues as potential therapeutic agents. Molecules, 27(2), 349.
31. Barba-Ostria, C., Carrera-Pacheco, S. E., Gonzalez-Pastor, R., Heredia-Moya, J., Mayorga-Ramos, A., Rodríguez-Pólit, C., ... & Guamán, L. P. (2022). Evaluation of biological activity of natural compounds: current trends and methods. Molecules, 27(14), 4490.
32. Wilson, B. A., Thornburg, C. C., Henrich, C. J., Grkovic,

T., & O'Keefe, B. R. (2020). Creating and screening natural product libraries. Natural product reports, 37(7), 893-918.
33. Lynch, C., Sakamuru, S., Ooka, M., Huang, R., Klumpp-Thomas, C., Shinn, P., ... & Xia, M. (2024). High-throughput screening to advance in vitro toxicology: accomplishments, challenges, and future directions. Annual review of pharmacology and toxicology, 64(1), 191-209.
34. Leelananda, S. P., & Lindert, S. (2016). Computational methods in drug discovery. Beilstein journal of organic chemistry, 12(1), 2694-2718.
35. Price, G., & Patel, D. A. (2023). Drug bioavailability. In StatPearls. StatPearls Publishing.
36. Sessions, Z., Sánchez-Cruz, N., Prieto-Martínez, F. D., Alves, V. M., Santos Jr, H. P., Muratov, E., ... & Medina-Franco, J. L. (2020). Recent progress on cheminformatics approaches to epigenetic drug discovery. Drug Discovery Today, 25(12), 2268-2276.
37. Morger, A., Mathea, M., Achenbach, J. H., Wolf, A., Buesen, R., Schleifer, K. J., ... & Volkamer, A. (2020). KnowTox: pipeline and case study for confident prediction of potential toxic effects of compounds in early phases of development. Journal of Cheminformatics, 12(1), 24.
38. Zhu, Y., Alqahtani, S., & Hu, X. (2021). Aromatic rings as molecular determinants for the molecular recognition of protein kinase inhibitors. Molecules, 26(6), 1776.
39. Rusu, A., Moga, I. M., Uncu, L., & Hancu, G. (2023). The role of five-membered heterocycles in the molecular structure of antibacterial drugs used in therapy. Pharmaceutics, 15(11), 2554.
40. Bahadur, R., Wijerathne, B., & Vinu, A. (2024). Multiple heteroatoms doped nanoporous biocarbon for supercapacitor and zinc-ion capacitor. ChemSusChem, 17(24), e202400999.
41. Ward, M., & O'Boyle, N. (2025). Analysis of the Structural Diversity of Heterocycles amongst European Medicines Agency Approved Pharmaceuticals (2014–2023). RSC Medicinal Chemistry.
42. Mao, F., Ni, W., Xu, X., Wang, H., Wang, J., Ji, M., & Li, J. (2016). Chemical structure-related drug-like criteria of global approved drugs. Molecules, 21(1), 75.
43. Kysliak, O., Schreiner, S. H., Grabicki, N., Liebing, P., Weigend, F., Dumele, O., & Kretschmer, R. (2022). A Planar Five-Membered Aromatic Ring Stabilized by Only Two π-Electrons. Angewandte Chemie International Edition, 61(31), e202206963.
44. Siramshetty, V. B., Xu, X., & Shah, P. (2023). Artificial intelligence in ADME property prediction. In Computational Drug Discovery and Design (pp. 307-327). New York, NY: Springer US.
45. Scotti, L., & Scotti, M. T. (2025). ADME Properties in Drug Delivery. Pharmaceutics, 17(5), 617.
46. Jiang, R., Hooshfar, S., Rebecca Eno, M., Yun, C., Sonego Zimmermann, E., & Shinkyo, R. (2023). Factors influencing ADME properties of therapeutic antisense oligonucleotides: physicochemical characteristics and beyond. Current Drug Metabolism, 24(7), 536-552.
47. Di Lascio, E., Gerebtzoff, G., & Rodriguez-Perez, R. (2023). Systematic evaluation of local and global machine learning models for the prediction of ADME properties. Molecular pharmaceutics, 20(3), 1758-1767.
48. Stielow, M., Witczyńska, A., Kubryń, N., Fijałkowski, Ł., Nowaczyk, J., & Nowaczyk, A. (2023). The bioavailability of drugs—the current state of knowledge. Molecules, 28(24), 8038.
49. Darlami, O., Pun, R., Ahn, S. H., Kim, S. H., & Shin, D. (2024). Macrocyclization strategy for improving candidate profiles in medicinal chemistry. European journal of medicinal chemistry, 272, 116501.
50. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., & Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. Nature chemistry, 4(2), 90-98.
51. Guan, L., Yang, H., Cai, Y., Sun, L., Di, P., Li, W., ... & Tang, Y. (2019). ADMET-score–a comprehensive scoring function for evaluation of chemical drug-likeness. Medchemcomm, 10(1), 148-157.
52. Leeson, P. D., Bento, A. P., Gaulton, A., Hersey, A., Manners, E. J., Radoux, C. J., & Leach, A. R. (2021). Target-based evaluation of "drug-like" properties and ligand efficiencies. Journal of medicinal chemistry, 64(11), 7210-7230.