COMPREHENSIVE REVIEW

Recent advances in the area of plant-based anti-cancer drug discovery using computational approaches

Agneesh Pratim Das[1] · Subhash Mohan Agarwal[1] ᴵᴰ

Abstract

Phytocompounds are a well-established source of drug discovery due to their unique chemical and functional diversities. In the area of cancer therapeutics, several phytocompounds have been used till date to design and develop new drugs. One of the desired interests of pharmaceutical companies and researchers globally is that new anti-cancer leads are discovered, for which phytocompounds can be considered a valuable source. Simultaneously, in recent years, the growth of computational approaches like virtual screening (VS), molecular dynamics (MD), pharmacophore modelling, Quantitative structure–activity relationship (QSAR), Absorption Distribution Metabolism Excretion and Toxicity (ADMET), network biology, and machine learning (ML) has gained importance due to their efficiency, reduced time-consuming nature, and cost-effectiveness. Therefore, the present review amalgamates the information on plant-based molecules identified for cancer lead discovery from in silico approaches. The mandate of this review is to discuss studies published in the last 5–6 years that aim to identify the phytomolecules as leads against cancer with the help of traditional computational approaches as well as newer techniques like network pharmacology and ML. This review also lists the databases and webservers available in the public domain for phytocompounds related information that can be harnessed for drug discovery. It is expected that the present review would be useful to pharmacologists, medicinal chemists, molecular biologists, and other researchers involved in the development of natural products (NPs) into clinically effective lead molecules.

✉ Subhash Mohan Agarwal
smagarwal@yahoo.com

1 Bioinformatics Division, ICMR-National Institute of Cancer Prevention and Research, I-7, Sector-39, Noida, Uttar Pradesh 201301, India

Graphical abstract

Reviewed the niche area of phytomolecule-based anti-cancer drug discovery with respect to current trends including machine learning.
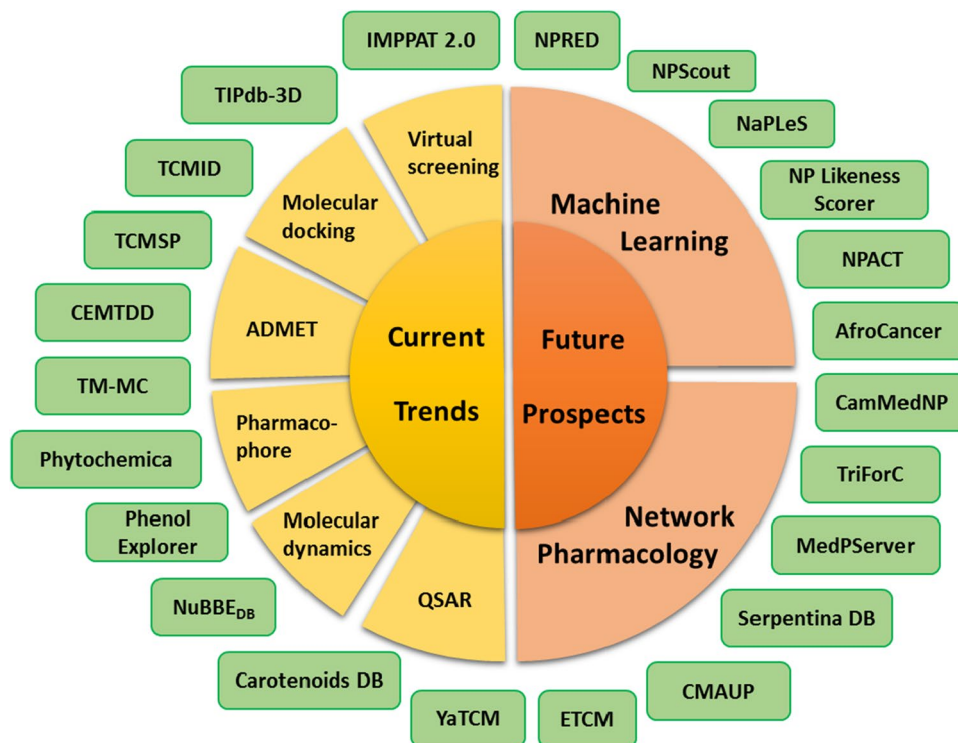
Introduction

Cancer is one of the leading causes of death around the globe. According to the global cancer statistics 2020, cancer cases have risen to 19.3 million new cases and 10 million cancer-related deaths. It is estimated that the worldwide cancer burden stands to rise by an astounding 47% to 28.4 million cases by 2040 [1]. Although a wide range of techniques like chemotherapy, radiation therapy, and immunotherapy are currently available to combat cancer, there still exists a need to find more diverse sources of new chemical entities (NCEs). While chemotherapy is currently the most reliable source of cancer treatment, it is known to be accompanied by adverse reactions, like toxic effects on non-targeted tissues, relapse, and drug resistance. Drugs designed from natural products (NPs) aim to counter several disadvantages of synthetic compounds and traditional chemotherapy techniques [2]. Molecular scaffolds of NPs show high chemical diversity making them suitable lead-like compounds for drug development [3]. Moreover, screening of NPs can lead to

the addition of novel molecules to current libraries, which can then be checked on pharmacokinetic parameters of drug discovery [4–8]. Since most NPs are naturally bioactive thus, they have the potential to be used more effectively in the process of drug development. 50% of the known anti-cancer drugs are derived from NPs [9] among which plant-based NPs have been historically used [10, 11]. These compounds are either isolated phytomolecules or semi-synthetic molecules of natural origin [12, 13]. Also, different classes of plant-derived NPs have been tested against a number of cancer cell lines and their anti-cancer activity has been reported in the literature [14]. These molecules majorly belong to terpenoids, flavonoids, and alkaloids categories, as shown below (Fig. 1).

The boom in the amount of biological information generated has led to the age of biological big data [15]. Accordingly, developing more and more computational resources to analyse the molecular features and chemical behaviour of NPs from an in silico perspective has gained increased importance. Various chemoinformatics and bioinformatics approaches such as structure and ligand-based techniques, pharmacophore modelling [16], virtual screening (VS) [17, 18], Quantitative structure–activity relationship (QSAR) [19], and network pharmacology are currently being used
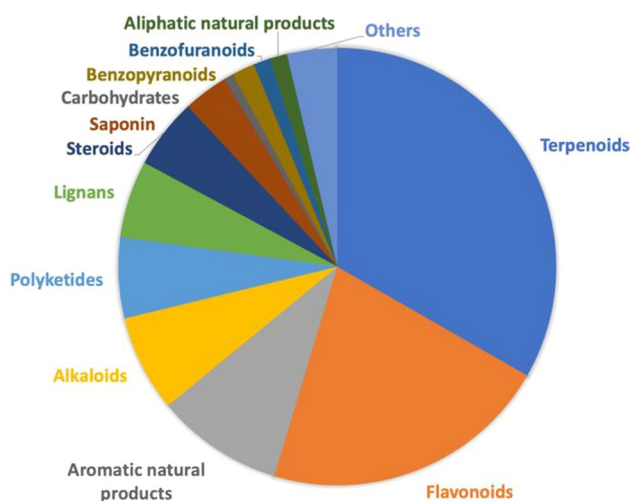
Fig. 1 Different classes of phytomolecules with known anti-cancer evidence

Fig. 2 Current scenario of anti-cancer drug discovery

for the analysis and interpretation of these data. This has in turn led to the compilation of multiple libraries of NPs with potential lead-like properties [7, 14]. However, despite the availability of substantial amounts of chemicobiological data and traditional plant-based anti-cancer remedies, there remains a lacuna in the domain of plant-based anti-cancer drug development. Thus, to usher in a new generation of NP-based drug discovery, machine learning (ML), deep learning (DL), and artificial intelligence (AI) have provided a plethora of functionalities that may play an important role in the drug discovery pipeline [20–23]. However, traditional drug discovery techniques are still more widespread compared to the newer generation of techniques (Fig. 2).

In the past 5 years, a number of reviews in this domain have been published, but they either tend to emphasise solely on the anti-cancer efficacy and mechanism of action of different phytomolecules [9, 24–26] or on the current practices of drug discovery and development [27–30]. The current review has therefore attempted to focus on studies that use plant-based molecules for anti-cancer drug discovery with the help of various in silico tools and techniques so that a better understanding can be achieved concerning the current advancements and approaches available. Owing to the role of the current trends of data-driven drug discovery and development, we have added a section specifically to address the use of ML in NP-based drug discovery. We have highlighted the recent advancements in the field of cancer drug discovery, emphasising the use of various plant-based compounds that have proven anti-cancer properties as well as novel lead compounds that have shown positive results in bioactivity studies. Finally, we conclude with an overview of the different tools and databases currently in use that endorse the vast diversity of plant-derived NPs, NP-based
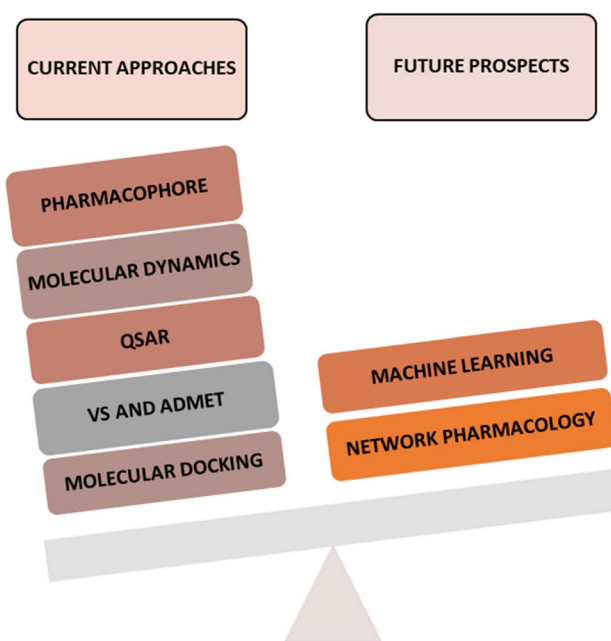
applications, and their subsequent uses in NP-based drug discovery.

Computational approaches used for natural product drug discovery

In the field of computer-aided drug discovery (CADD), the use of structure and ligand-based approaches have been ongoing, depending on the availability of crystallised 3-dimensional (3D) structure of the receptor molecule in the Protein Data Bank (PDB) [31]. For ease of understanding, we have categorised the different computational techniques under these two sub-categories and discussed in detail how they have been used to identify the plant-derived NPs against different cancer targets.

Structure and ligand-based drug discovery

In structure-based drug design, the active site of the macromolecule structure, the presence of specific amino acids in the binding pockets, and the strength of interaction of the reacting species are all considered during lead identification. Structure-based approaches like molecular docking and VS techniques have facilitated the fast and cost-effective analysis of large sets of compound databases to identify the potential hit molecules [17, 18, 32]. For selecting a hit compound, many molecules are screened in order to identify the molecules having desired activity against a target with a

specific set of chemical and structural features [33]. Ligand-based drug design (LBDD), on the other hand, is used in the absence of relevant 3D structural information about the receptor molecule and is dependent on the information about the different small molecule ligands which are biologically active in a specific disease or against a drug target. QSAR and pharmacophore modelling are the most prominent techniques used in LBDD to identify and design novel inhibitors.

Virtual screening (VS) and molecular dynamics (MD)

VS is a highly effective tool for discovering new active compounds in drug discovery [34, 35]. MD, on the other hand, helps establish the strength and stability of the interactions in the protein–ligand complex [36, 37]. VS has been of help in the domain of drug development to screen libraries for identifying compounds that depict high binding affinities. This time-efficient technique not only helps in dealing with large datasets but also avoids late-stage drug development failures as it narrows down the number of compounds to be evaluated in biological assays. Unlike the static conformations used in classical models of drug development studies, the biomolecules in the human body are constantly in motion. Thus, it is necessary to develop an understanding of the changing molecular structures of receptors for designing better approaches to drug development. MD predicts the molecular and structural changes of biomolecules under the influence of inter and intramolecular forces and therefore is essential for drug development studies. In the area of NP drug discovery, VS has been extensively used to discover molecules against different cancer targets. Examples of studies published in this area that uses VS and MD to identify the plant-based compounds as leads are listed in Table 1, a few of which are discussed in detail below.

Muhseen et al. [38] used VS, absorption, distribution, metabolism, excretion, and toxicity (ADMET) filtering and MD to discover potential inhibitors against Mouse double minute 2 homolog (MDM2) from terpenoid compounds. They have utilised (a) pharmacophore model-based VS and (b) tanimoto coefficient-based screening using co-crystal ligand, to identify the compounds with similar activity. Three terpenes, 3-trans-*p*-coumaroyl maslinic acid, silvestrol, and betulonic acid, were thereafter identified as potential inhibitors of the p53–MDM2 interaction (Fig. 3A–C).

Zarezade et al. [36] screened ~ 20,000 NPs from ZINC [39] to identify the potential candidates for progesterone receptor (PR) inhibition, which is one of the key targets in breast cancer. The X-ray crystal structure of PR was obtained from PDB, and the dataset of NPs was virtually screened to select top 200 compounds with the best docking scores. After ADMET screening of these molecules, 56 drug-like

molecules were identified and subjected to the MTiOpen-Screen [35] web server to identify 10 compounds with the best estimated binding energy. These PR antagonists were then redocked using AutoDock v4.2 [40] and based on the calculated binding energy, inhibitory constant (K_i), and the key residues involved in the interactions, three NPs were identified as inhibitors (ZINC00936598, ZINC01020370, and ZINC00869973) (Fig. 3D–F). Among these three compounds, ZINC00936598 had the highest binding energy and the lowest K_i. These compounds were then subjected to 100-ns MD simulation to examine the stability of the protein–ligand complexes. H-bond analysis demonstrated that the lead compounds in complex with PR preserve stability during the simulation period. Comparative H-bond analysis identified that the lead compound ZINC00936598 was more potent than the other two molecules and was identified as the best PR antagonist for breast cancer treatment.

A primary reason for metastasis is the degradation of basement membrane mediated by Matrix Metalloproteinase (MMP9). Biswas et al. [41] used ~ 14,000 NPs from ZINC to identify the potential intercalating agents that inhibit the interaction between tumour necrosis factor receptor-associated factor 6 (TRAF6) and Basigin (BSG) in order to control MMP9 overexpression. The BSG–TRAF6 complex was prepared using Zdock [42] protocol, following which the natural compound dataset was virtually screened using Genetic Optimisation for Ligand Docking (GOLD) software [43]. The top 20 compounds identified by GOLD score and Chemscore were then checked for their ADMET properties using various parameters and the best 3 molecules were selected. These 3 molecules were thereafter redocked (blind docking) using PatchDock [44] to validate their interaction with the target molecule complex. To further elucidate the effect of binding of these molecules and check the dynamic flexibility of the BSG-TRAF6 complex, a 60-ns MD simulation was performed. This led to the identification of ZINC02578057 (Fig. 3G) as the best inhibitor of the TRAF6 complex. Therefore, the authors proposed this molecule as a potential inhibitor of the BSG-TRAF6 complex to control the overexpression of MMP9 in melanoma.

In another study, Jairajpuri et al. [17] performed VS of a set of ~ 33,000 ADMET-filtered NPs from ZINC database against Sphingosine kinase 1 (SphK1). A list of ten hit compounds were identified based on their binding affinities to SphK1. They were then further subjected to docking analysis to observe all possible docked conformers of the complexes. Finally, two compounds ZINC05434006 and ZINC04260971 (Fig. 3H, I) were proposed as inhibitors based on their interactions with the substrate binding site of SphK1.

Raj et al. [32] studied flavonoid-based molecules to find novel inhibitors for the Bromodomain-containing protein 4 (BRD4). The active site was identified using literature

Table 1 Studies listing different drug discovery techniques used for the identification of NP-based anti-cancer molecules

No. of natural compound (s)	NP source	Target molecule (s)	Cancer	D	VS	QM	MD	PM	Q	AD	Best predicted compound (s)	References
1330	Multiple NP databases	Aromatase	Cervical	Y	Y	-	Y	Y	-	-	Forskolin	[69]
500	Terpenes	MDM2	Multiple	Y	Y	-	Y	Y	-	Y	3-Trans-p-coumaroyl maslinic acid, silvestrol, betulonic acid	[38]
59,619	InterBioScreen	DDX3	Multiple	Y	Y	-	Y	Y	-	Y	Curcumin	[16]
22	Nandina domestica	CASP6, MAPK1, TP53, AKT1	Multiple	Y	-	-	-	-	-	Y	Isoquercitrin, quercitrin, berberine, chlorogenic acid, caffeic acid	[70]
500 Flavonoids and 4000 extended flavonoids	TimTec	BRDs	Multiple	Y	Y	-	Y	Y	-	Y	2-(3,4-Dihydroxyphenyl)-3,7-dihydroxy-4H-chromen-4-one	[32]
21	-	SIRTs	Triple-negative Breast	Y	-	-	Y	-	-	Y	Sulforaphane, kaempferol, apigenin, genistein, berberine, resveratrol, honokiol	[71]
145,628	12 NP databases	EGFR	Non-small-cell lung carcinoma	Y	Y	-	Y	Y	Y	Y	ZINC03935485, ZINC05013091, ZINC35465964	[37]
Sesamin and other structurally related lignans	-	CTNNB1	Colorectal	Y	-	-	Y	-	-	-	Sesamin	[72]
40	-	EGFR, HER2	Breast	Y	-	-	Y	-	-	Y	Panaxadiol	[73]
~90,000	ZINC NP catalogue	SPHK1	Multiple	Y	Y	-	Y	-	-	Y	ZINC05434006, ZINC04260971	[17]
Dietary phytochemicals	-	SPHK1	Multiple	Y	-	-	Y	-	-	-	Quercetin, capsaicin	[74]
593	Maslinic acid and analogs		Breast	Y	-	-	-	Y	Y	Y	P-902	[61]
~90,000	ZINC NP catalogue	SGK1	Multiple	Y	Y	-	Y	-	-	Y	ZINC00319000	[75]
168,686	PubChem	DNMT1	Breast	Y	Y	-	-	Y	-	Y	C-7756, C-5769, C-1723, C-2129, C-2140	[18]
20,000	ZINC	PGR	Breast	Y	Y	-	Y	-	-	Y	ZINC00936598, ZINC00869973, ZINC01020370	[36]
NA	Salvia lachnocalyx	Topoisomerase I, Topoisomerase IIa, Topoisomerase IIb	Multiple	Y	—	-	Y	-	-	-	Geranyl farnesol, sahandinone, 4-dehydrosalvilimbinol	[76]
45	Vernonia cinerea	EGFR	Non-small-cell lung carcinoma	Y	Y	-	Y	-	-	-	Luteolin-7-glucoside, epicatechin gallate	[77]
57,578	ZINC, SuperNatural II, ExiMed, InterBioScreen	CDK7	Multiple	Y	Y	-	Y	Y	-	Y	ZINC20392430, SN00112175, SN00004718, SN00262261	[56]

Table 1 (continued)

No. of natural compound(s)	NP source	Target molecule(s)	Cancer	D	VS	QM	MD	PM	Q	AD	Best predicted compound(s)	References
5	Glycyrrhetinic acid derivatives	–	Triple-negative Breast	Y	–	–	–	–	Y	–	GA-1, GA-2, GA-3, GA-4	[66]
68	*Ficus carica* L	CDK2, CDK6, Topoisomerase I, Topoisomerase II, BCL2, VEGFR2	Multiple	Y	Y	–	Y	–	–	Y	β–Bourbonene	[78]
29	Ursolic acid derivatives	Different targets of NF-kB pathway	Bladder	Y	–	–	–	–	Y	Y	T9, B42	[19]
68	Dr. Dukes Phytochemical and Ethnobotanical database	CTLA4, CDK8, EGFR, mTOR, p53R2, PR	Breast	Y	Y	–	–	–	–	Y	Ursolic acid, enterolactone, parthenolide, berberine, berberastine	[79]
14,509	ZINC	TRAF6-BSG complex	Melanoma	Y	Y	–	Y	–	–	Y	ZINC49048033, ZINC0258057, ZINC72320240	[41]

D docking, *VS* virtual screening, *QM* quantum mechanics, *MD* molecular dynamics/molecular mechanics, *MD* molecular dynamics, *PM* pharmacophore modelling, *Q* QSAR, *AD* ADMET, MDM2: Mouse Double Minute 2; DDX3: DEAD-Box Helicase 3; CASP6: Caspase 6; MAPK1: Mitogen-Activated Protein Kinase 1; TP53: Tumour Protein P53; AKT1: AKT Serine/Threonine Kinase 1; BRDs: Bromodomain-containing genes; SIRTs: Sirtuin genes; EGFR: Epidermal Growth Factor Receptor; CTNNB1: Catenin Beta 1; HER2: Human Epidermal Growth Factor Receptor 2; SPHK1: Sphingosine Kinase 1; SGK1: Serum/Glucocorticoid Regulated Kinase 1; DNMT1: DNA Methyltransferase 1; PGR: Progesterone Receptor; CDK2: Cyclin-Dependent Kinase 2; CDK6: Cyclin-Dependent Kinase 6; BCL2: BCL2 Apoptosis Regulator; VEGFR2: Vascular Endothelial Growth Factor Receptor 2; NF-kB: Nuclear Factor kappa B; CTLA4: Cytotoxic T-Lymphocyte Associated Protein 4; CDK7: Cyclin-Dependent Kinase 7; CDK8: Cyclin-Dependent Kinase 8; mTOR: mammalian Target Of Rapamycin; p53R2: p53-inducible ribonucleotide reductase gene; PR: Progesterone Receptor; TRAF6: TNF Receptor-Associated Factor 6; BSG: Basigin
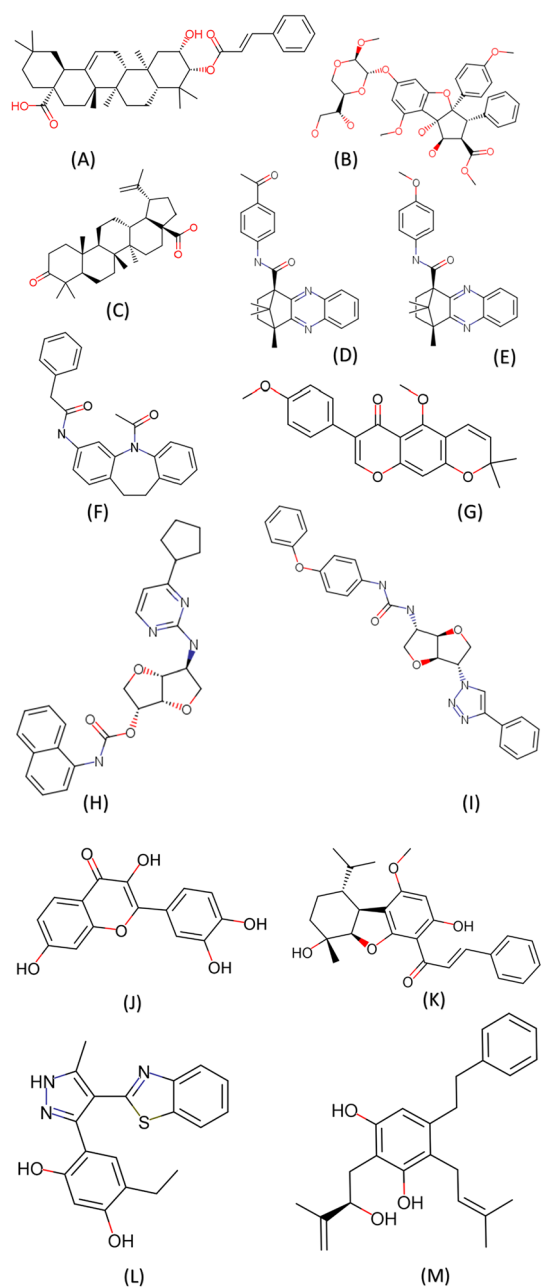
Fig. 3 Structures of **A** 3-trans-*p*-coumaroyl maslinic acid, **B** Silvestrol **C** Betulonic acid, **D** ZINC00936598, **E** ZINC01020370, **F** ZINC00869973, **G** ZINC02578057, **H** ZINC03935485, **I** ZINC05013091, **J** ST055650, **K** ZINC03935485, **L** ZINC05013091, **M** ZINC35465964

evidence as well as SiteMap [45]. VS was performed for 500 Flavonoids and 4000 extended flavonoids obtained from TimTec database (https://www.timtec.net/) and several known inhibitors like Ms435, Bromosporine, and CPI203 were also docked as reference. A three-tier docking approach was taken using the HTVS, SP, XP features of Glide [46] where at each stage the top 10% of the compounds were taken further. Finally, the top-ranked flavonoid compound

was identified as ST055650 (Fig. 3J) which had high scores in both Glide and AutoDock docking analyses. Subsequently, the ADMET and drug-likeness of this molecule were also checked and a 50-ns MD simulation was performed to validate the stability of the ligand molecule in complex with the target protein. All subsequent analyses of MD trajectory and interactions highlighted ST055650 as a potential candidate for BRD4 inhibition.

The double mutated Epidermal Growth Factor Receptor (EGFR) is an important and well-known target in lung cancer that demonstrates resistance to the existing drugs. Therefore, Agarwal et al. [37] addressed this clinical problem using an integration of ADMET, ML, VS, and MD to identify the new NP inhibitors for the mutant protein. A total of 1,52,056 naturally occurring small molecules from 12 different NP databases were retrieved and their drug-likeness were evaluated using Lipinski's rule of 5 and Ghose filter [47]. The 74,673 molecules that passed ADMET filtering were thereafter subjected to a random forest-based binary ML classification model (NPred) [20] to evaluate their anticancer potential. 4681 potential anti-cancerous molecules with NPred score > 0.7 were then subjected to VS against the EGFR mutant crystal structure (PDB ID: 5EDQ) using FLEXX-PHARM [48]. As a result of the constraint docking approach, 1339 molecules were docked and finally 3 molecules having lowest binding free energy were identified. 100-ns MD simulations of the top 3 ligands (ZINC03935485, ZINC05013091, ZINC35465964) (Fig. 3K–M) showed that all the three inhibitors interact with Gln791 and Met793 in a conserved manner, similar to the co-crystal ligand. The study concluded with the identification of three naturally occurring molecules as potent inhibitors of the double mutated EGFR protein.

Pharmacophore modelling

The pharmacophore modelling technique is important for core/scaffold hopping and hit or lead optimisation. The validated pharmacophore models are used to search for bioactive molecules from compound databases, containing a large number of structures. Some of the highly used pharmacophoric features are, hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (HY), ring aromatic (RA), and others such as positive ionisable and negative ionisable structures. One of the key features of using pharmacophore-based techniques is that it is effective both in the presence and absence of receptor structure. By generating pharmacophores, it is possible to identify the specific interacting species of a molecule that are crucial for either eliciting any biological response or blocking it [49]. Along with the chemical features of any bioactive compound, their biological features are also highlighted by pharmacophores. Based on the type

of molecule used to generate the pharmacophore, the entire
process can be classified into two subtypes: ligand-based and
structure-based pharmacophore modelling. In the ligand-
based methods, the knowledge of known active compounds
is used for pharmacophore generation. The structure-based
approach on the other hand requires a bound receptor–ligand
complex to elucidate the underlying interactions of the bind-
ing pocket that can be used to generate the pharmacophores
[50]. The details of a few studies published in this area in the
last five years have been compiled and provided in Table 1.
These studies highlight the use of pharmacophore-based
approaches for the identification of phytomolecules having
potential anti-cancer activities. A few of these case studies
have also been discussed below.

Complex-based pharmacophore generation techniques
were used by Babu et al. [51] for studying natural compound
derived inhibitors against seven therapeutic targets of gastric
cancer. The three flavonoids 5-hydroxy-7,4′-dimethoxy-6,8-
di-C-methylflavone, kaempferol-3-O-β-D-glucopyranoside
(Fig. 4A) and kaempferol-3-O-α-l-rhamnopyranoside
(Fig. 4B) from the fruits of the medicinal plant *Syzygium
alternifolium* were isolated and their anti-gastric cancer
activity was checked against the AGS cell line. Using molec-
ular docking studies, it was found that Human Epidermal
Growth Factor Receptor 2 (HER2) showed maximum affin-
ity with these ligands. Thus, the docking poses of these 3
compounds in the active site of HER2 were utilised to gen-
erate pharmacophore models. The pharmacophore model
generated for the best complex (selected based on cytotoxic
profile and docking score) was used as a query in the ZINC
database to identify the compounds with 90% similarity to
the best compound. After subsequent docking and MD stud-
ies, the three isolated compounds as well as ZINC67903192
(Fig. 4C) were identified as promising HER2 inhibitors
against gastric cancer.

Bommu et al. [34] had used ligand-based pharmacoph-
ore techniques to identify the potential inhibitors for EGFR,
which is an important target in non-small cell lung cancer
(NSCLC). The authors have used the plant-based NP querce-
tin as a template for ligand-based VS and identified 100 ana-
logs with similar pharmacological properties to quercetin
(Fig. 4D). Molecular docking of the 100 quercetin analogs
in the binding pocket of EGFR led to the identification of
10 molecules based on their binding affinities. Pharmacoph-
ore modelling of these 10 lead molecules was followed by
the generation of a merged pharmacophore to identify the
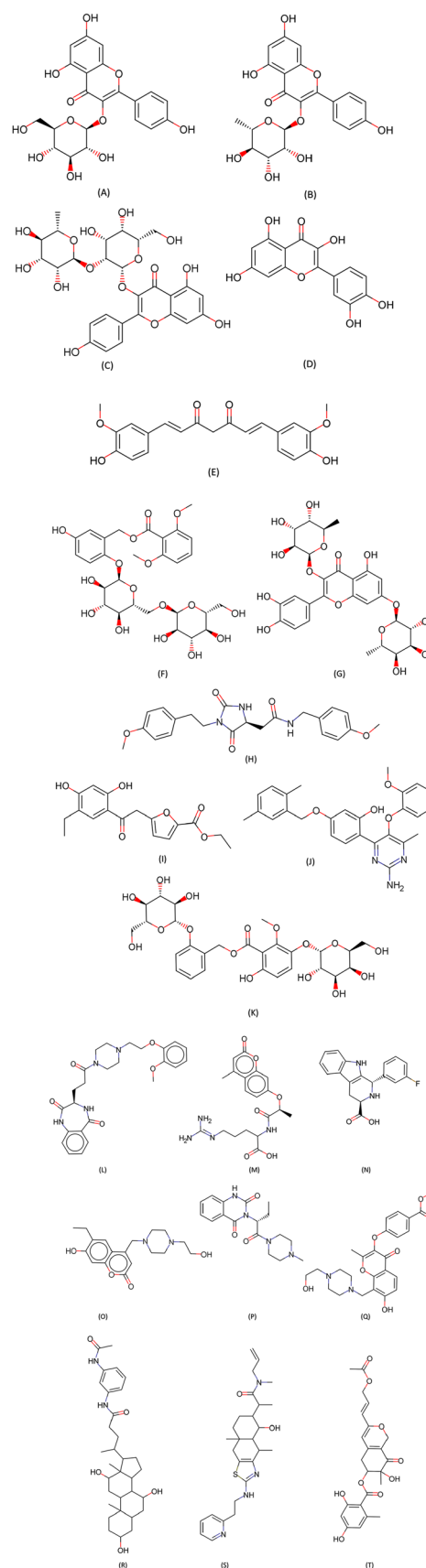
important features that contribute towards the better binding efficiency of these molecules. Further analysis of the drug-likeness and ADMET properties of these molecules revealed that the molecules follow ADMET properties and therefore can be considered for drug development.

In a few studies, different types of pharmacophore models are often used in conjunction, like the common feature and receptor-based pharmacophore models generated by Rampogu et al. [16] for the DEAD box protein 3 (DDX3). DDX3 is overexpressed in a variety of cancers like breast, colorectal, liver, lung and oral cancers as it plays important role in cancer progression, proliferation and transformation. VS of the above two pharmacophore models against ADMET screened compounds of InterBioScreen (https://www.ibscreen.com/) led to the identification of 17 compounds that were common to both these pharmacophore models. Using binding affinity information derived from molecular docking of these 17 compounds, curcumin (Fig. 4E) was identified as the best lead molecule against DDX3 based on binding affinity as well as key residue interactions and MD simulation.

Singh et al. [52] performed pharmacophore-guided activity profiling of NPs from the ZINC database for five different tyrosine kinase receptor targets of lung cancer i.e. EGFR, tyrosine-protein kinase Met (cMET), erb-b2 receptor tyrosine kinase 2 (ErbB2), Fibroblast Growth Factor Receptor (FGFR) and Anaplastic Lymphoma Kinase (ALK). For these five tyrosine kinases, an exhaustive literature search was performed to identify the appropriate molecular datasets which were then used for the generation of ligand-based pharmacophore models. The Catalyst module of Discovery studio was then used to create hypogen-based 3D pharmacophores which were verified using the parameters of sensitivity, specificity, and receiver operating characteristic (ROC) curve. The ZINC, NP catalogue was then screened by using the pharmacophore models of these 5 targets. Cross-selection and sorting was done based on fitness score which resulted in the identification of 10 molecules against the respective target proteins. These 10 molecules were then docked to the crystal structures of the 5 target proteins using CDOCKER. Finally, MD simulation of eight NPs (ZINC85643856, ZINC85646292, ZINC96221218, ZINC00241889, ZINC98365505, ZINC98364461, ZINC11866307, ZINC38143676) (Fig. 4F–K) with different proteins were performed and it was found that these ligands show several important interactions necessary for the inhibition of their respective protein targets.

Alamri et al. [53], had identified a potential novel sigma-2 inhibitor using pharmacophore and structure-based VS of a database consisting of natural and natural-like compounds. The key structural features of known ligands of the sigma-2 receptor were used to develop the pharmacophore model. The target sigma-2 protein was modelled by homology modelling through Iterative Threading ASSEmbly Refinement

(I-TASSER) webserver [54]. The ligand-based pharmacophore was created using 19 diverse ligands having high binding affinity, obtained from Sigma-2 receptor selective Ligands Database (S2RSLDB) [55]. The pharmacophore model having the highest score was selected after quantitatively validating it with 100 actives and 425 decoys. The database of compounds was then screened by docking-based VS using Glide XP and the top 20 compounds were selected based on the binding energy score. Five lead compounds (SR84, SR300, SR413, SR823, SR530) (Fig. 4L–P) were selected afterwards and compared with the reference compound, CM366. Thereafter, ADMET properties and MD simulation for 50 ns was undertaken to validate five compounds as inhibitors of sigma-2 protein.

The cyclin-dependent kinase 7 (CDK7) is one of the most important members of the CDK family of genes/proteins which regulates cell cycle events. Several cancers, like hepatocellular, gastric, oral, breast, ovarian, pancreatic, and colorectal cancer, have reported overexpression of CDK7 which in turn has been linked with aggressive clinicopathological features and poor prognosis. Therefore, Kumar et al. [56] had implemented a ligand-based pharmacophore approach wherein a small set of selective inhibitors downloaded from PubChem [57] were used as the training set for hypothesis creation. These molecules were then used for generating ten pharmacophore models using the Hip-Hop algorithm. Thereafter, a structure-based pharmacophore model was also generated using CDK7 structure bound with THZ1 (PDB ID: 6XD3). Both the ligand and structure-based pharmacophore models were validated using the ROC curve and the Güner–Henry (GH) approach. Thereafter, 57,578 ADMET-filtered NPs from four databases (ZINC, SuperNatural II [58], ExiMed [https://eximedlab.com/], and InterBioScreen) were screened using the validated pharmacophore models. 197 compounds were selected based on both the pharmacophore models which were subjected to molecular docking using GOLD. 24 potential inhibitors had displayed better binding scores than the two reference molecules (CT7001 and THZ1) and were thus taken for 50-ns MD simulation studies. The results of the MD were analysed to select four NPs (ZINC20392430, SN00112175, SN00004718, and SN00262261) (Fig. 4Q–T) that demonstrated better binding affinity than known inhibitors as well as increased stability in their docked poses.

QSAR

The QSAR technique comes under the ligand-based drug designing method and involves the building of mathematical models to find statistically significant correlations between chemical structures and biological properties like half-maximal inhibitory concentration (IC_{50}), half-maximal effective

concentration (EC$_{50}$), and K$_i$ [59]. A lot of advancements have been made in the field of QSAR in the last decade which has hugely increased the dimensionality of molecular descriptors (from 1D to nD) [60] and even with respect to extraction of co-relation between chemical structures and biological properties. QSAR-based techniques have been used in the identification of potential lead compounds for a wide variety of cancers. A few examples of how QSAR-based techniques have been used to identify the plant-based NPs against different cancers have been shown in Table 1 as well as some are discussed in detail below.

Alam and Khan [61] developed field-based 3D-QSAR models on the MCF7 breast cancer cell line to identify the anti-cancer effects of the natural triterpene maslinic acid and its analogs. The important features like average shape, hydrophobic regions, and electrostatic patterns of active compounds were extracted and mapped to virtually screen potential analogs. Field points-based descriptors were used in order to generate the 3D-QSAR model by aligning known active compounds onto identified pharmacophore templates. Ultimately a compound (P-902) (Fig. 5A) was then identified as the best for different breast cancer targets like Aldo-Keto Reductase family 1 member B10 (AKR1B10), Nuclear Receptor subfamily 3 group C member 1 (NR3C1), Prostaglandin-endoperoxide Synthase 2 (PTGS2), and HER2.

Lung cancer is the leading cause of cancer-related deaths worldwide and has been the focus of drug discovery research. Using 3D-QSAR, docking, flow cytometry, and gene expression, Chen et al. [62] studied the anti-cancer activity of the natural chalcone lonchocarpin (Fig. 5B). The authors found hydrophobic interactions to be the most influential factor for the anti-tumour efficacy of lonchocarpin. Subsequent molecular docking studies also showed that lonchocarpin binds stably to the BH3-binding groove of the anti-apoptotic protein BCL2. Using several in silico techniques including QSAR, it was demonstrated that lonchocarpin is a potentially useful natural agent for cancer treatment.

The natural compound group withanolide present in roots and leaves of *Withania somnifera* (Indian ginseng/Ashwagandha) was explored by Yadav et al. [63] for developing QSAR models. The models utilised information about the anti-proliferative activity of withanolide analogs against different human breast cancer cell lines (SK-Br-3 and MCF7/BUS). The model for the SK-Br-3 cell line showed a 93% correlation between activity and chemical descriptors in the training sets, while the model for MCF7/BUS cell line showed a 91% correlation. The cross-validation coefficient indicated 90% and 85% prediction accuracy of both models, respectively. The two highly active compounds (CID_301751 and CID_3372729) (Fig. 5C, D) identified showed higher anti-proliferative activity than the reference compounds 5-fluorouracil (5-FU) and camptothecin (CPT).
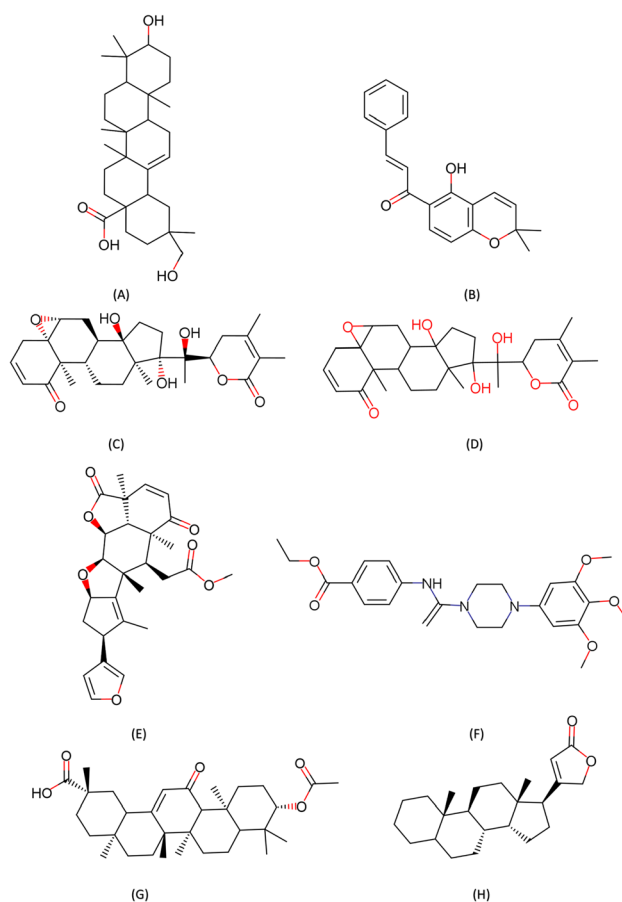

Fig. 5 Structures of **A** P-902, **B** Lonchocarpin, **C** CID_301751, **D** CID_3372729, **E** T9, **F** B42, **G** GA-1, **H** Cardenolide

Therefore, these compounds were subjected to molecular docking using Surflex-Dock [64] module of SYBYL-X 2.1 on the biological target, β-tubulin which showed favourable binding affinity, bioavailability and ADME properties.

Ursolic acid (UA) is a natural pentacyclic terpene that has promising anti-cancer properties. Thus, Yadav et al. [19] developed 3D QSAR models from 29 UA derivatives that can inhibit the T24 bladder cancer cell line. The 3D QSAR-based CoMFA models were developed using SYBYL-X 2.1 and used for the prediction of the bioavailability and bio-activity of different compounds. The screened compounds were subjected to pharmacokinetic evaluation equivalent to the standard anti-cancer drug doxorubicin. In order to understand the underlying molecular mechanism of action, molecular docking was also performed. The two predicted active compounds T9 and B42 (Fig. 5E, F) satisfied all standard screening protocols like Lipinski's rule of 5 [65], PK/PD and toxicity parameters. Thus, T9 and B42 were proposed as promising leads against bladder cancer.

The bioactive triterpenoid glycyrrhizic acid (GA) can be obtained from the Indian medicinal herb *Glycyrrhiza glabra*. Shukla et al. [66] performed a QSAR study to observe the

biological effects of GA and its derivatives against metastatic Triple-Negative Breast Cancer (TNBC) cell lines. Using a regression-based QSAR model, five novel GA derivatives were designed, synthesised and screened for in vitro activity in metastatic breast cancer cell line MDA-MB-231. The results highlighted novel derivative GA-1 (Fig. 5G) having a cytotoxic activity greater than that of its parent compound GA. Further, atomic property field (APF)-based 3D QSAR and subsequent molecular docking studies revealed that the C-30 carboxylic group of the novel derivative GA-1 was the most important factor for its anti-cancer activity.

Cardenolides (Fig. 5H) are a class of cardiac glycosides that have long been used as potent inhibitors of the Na$^+$/K$^+$-ATPase transmembrane protein, which is overexpressed in a variety of cancers like skin, kidney, and lung. Therefore, Meneses-Sagrero et al. [67] developed SAR/QSAR models using 58 cardenolides having anti-proliferative effects on the lung cancer cell line A549. For these 58 molecules, their structures were generated using ChemBioDraw Ultra and molecular descriptors were calculated using PaDEL [68]. Using linear regression, the dimensionality of the physicochemical descriptors was reduced and ultimately 62 descriptors were used to generate the mathematical models. The QSAR models developed were thereafter validated using random cross-validation and models with $R^2 > 0.7$ were taken for external validation. As a result, the authors concluded that the addition of a sugar moiety at the C3 position of cardenolides positively affects its anti-proliferative effect on lung cancer cells.

ADMET screening

In recent years, several in silico techniques have been developed that allow rapid assessment of molecules for their ADMET properties. Through these techniques, drug efficacy and toxicity are monitored by taking into consideration the pharmacokinetics and pharmacodynamics of potential drug-like molecules [80]. It is important to take into consideration the ADMET properties of the molecules under investigation a priori so that late-stage failures in clinical trials may be avoided. Therefore, the prediction of ADMET properties through in silico/computational techniques has become the cornerstone for "good" drug discovery practices. Some databases are also available that provide the ADMET processed compounds directly for VS [5, 7, 8]. Studies exist that analyse the pharmacokinetic properties of plant-derived NPs [6] in diseases other than cancer and therefore have not been discussed in this review. The studies discussed below are all focused on the ADMET analyses of plant-based molecules with anti-cancer effects.

In silico prediction of ADMET properties of molecules has become significant in current times since these methods can lower the probability of late-stage drug development failure and enable the use of only few promising molecules for wet lab experimentation. Thus, there are several resources (databases and software) available for computing and estimating different ADMET properties of molecules, like QikProp, pkCSM [81], and DataWarrior [82]. Fatima et al. [8] utilised these tools to evaluate the pharmacokinetic potential of ~3000 phytomolecules from a wide variety of geographically diverse databases like Phytochemica [83], SerpentinaDB [84], SANCDB [85], NuBBE$_{DB}$ [86], respectively. After ADMET profiling of these compounds, their anti-cancer potential was also evaluated using literature-based experimental evidence and their activity against different cancer cell lines and protein targets was documented. Finally, 24 compounds were identified which showed the best ADMET behaviour that belonged to NuBBE$_{DB}$, Phytochemica, and SANCDB databases. Additionally, a user-friendly database ADMET-BIS was also created wherein users can find details about the ADMET behaviour of these phytochemicals from Brazil, India, and South Africa.

In order to identify the drug-like anti-cancer compounds from the medicinal plants used in African Traditional Medicine (ATM), Ntie-Kang et al. [7] had examined the pharmacokinetic properties of ~400 anti-cancer compounds discovered from African flora (AfroCancer) and compared it with ~1500 anti-cancer compounds from Naturally Occurring Plant-based Anti-cancer Compound-Activity-Target (NPACT) [14]. Lipinski's Rule of 5 [65] which is used for determining the physicochemical properties of drug-like molecules showed that no Lipinski violations were present in approximately 200 compounds in the AfroCancer dataset. A comparison of the ADMET properties of these datasets based on the star parameter of QikProp revealed that around 232 and 630 compounds of AfroCancer and NPACT, respectively, fall within the accepted range of 95% known drugs. The overall study demonstrates the pharmacokinetic potential of African anti-cancer plant molecules that can be taken forward for further experimental validation.

Sharma et al. [5] collected ~5000 anti-cancer phytomolecules from three open-access databases NPACT, CancerHSP [87], and TaxKB [88], which were then checked for their ADMET properties. The ADMET profiling of these compounds was done using QikProp with 13 parameters (namely Lipinski's Rule of 5, Jorgensen's Rule of 3, polar surface area, number of rotatable bonds, octanol–water partition coefficient [logP], affinity to plasma protein [log-KHSA], number of likely metabolic reactions, number of reactive functional groups, number of amides, number of amines, number of acids, IC$_{50}$ for blockage across K+ channels [loghERG] and transport across blood–brain barrier [logBB]). It was observed that 63% of the tested compounds were orally absorbable, 52% were distributable, 45% could be metabolised and excreted, and 28% were found to

be non-toxic for cardiotoxicity and central nervous system (CNS) activity. The authors have developed an interactive database, ADMETCan, which can be used to access the ADMET profile of these studied compounds. This resource provides information about plant-based molecules that have a higher chance of being properly eliminated from the body.

Selaginella repanda is an important ethnomedicinal plant used by certain Indian tribal communities for treating a wide variety of diseases and health conditions. Therefore, Adnan et al. [89] have tried to assess the anti-cancer properties of the *S. repanda* phytoconstituents. The crude extract of *S. repanda* was tested against the MCF7, HCT116 and A549 cell lines where it exhibited significant anti-proliferative activity against three cell lines. Subsequently, High-resolution liquid chromatography–mass spectrometry (HR-LC-MS) was performed on the above-mentioned crude extract of *S. repanda* and 54 phytochemicals from different classes like fatty acids, alcohols, sugars, flavonoids, alkaloids, terpenoids, coumarins, and phenolics were identified. For these molecules, the ADMET properties were computed using SwissADME, and it was found that all of them obey Lipinski's rule of 5. Additionally, most of these molecules exhibited good bioavailability scores, moderate to good skin permeability and moderate to potent Caco-2 permeability. This allowed the identification of phytoconstituents of *S. repanda* that have anti-proliferative activity against different cancers and are adherent to ADMET parameters.

Next-generation drug discovery techniques

Alongside the traditional chemoinformatics-based approaches described earlier, the newer generation of drug discovery pipelines are utilising concepts like network pharmacology and ML in the drug discovery process [21, 90, 91]. AI has progressed from mostly theoretical studies to more real-world applications in recent years, including different stages of drug research and development. Here we have focused on how these techniques of ML and network biology have been utilised for the identification of NCEs and potential lead-like compounds from plant-based natural sources.

Network pharmacology-based approaches

The traditional approaches to drug development continue to provide reliable results, but they tend to follow the 'one drug one target' approach. As diseases like cancer are multifactorial, it has led to the development of the interdisciplinary field of network pharmacology, which falls at the intersection of network biology and polypharmacology [92]. Network pharmacology follows a system-based thinking wherein a complex network of biological pathways is broken into smaller, more understandable pieces, which can then be easily studied for further analysis. It relies on the concept of finding multi-targeted drugs, acting on multiple biological targets and pathways, thus yielding better therapeutic results [92]. Examples of studies that demonstrate how network pharmacology is being used to identify the important phytomolecules against different cancer targets are given in Table 2. A few of them are discussed in detail below.

Shang et al. [90], had conducted a study on tetrandrine (Fig. 6A), a bis-benzylisoquinoline alkaloid extracted from *Stephania tetrandra*, a traditional Chinese medicine, to predict its molecular mechanisms against endometrial cancer. Seven key target genes were identified using protein–protein interaction (PPI) network analysis which were then found to be distributed in the PI3K/Akt signalling pathway using Kyoto Encyclopedia of Genes and Genomes (KEGG) [93]. Molecular docking and in vitro assays revealed that tetrandrine acted as a tumour suppressor by repressing BCL2 and promoting the activity of BCL2 associated X, apoptosis regulator (BAX) at the mRNA level. In most cancers, aberrant apoptotic signalling, notably the inactivation of apoptotic systems, enables cancer cells to evade programmed cell death, resulting in uncontrolled proliferation, tumour survival, treatment resistance, and cancer recurrence. Therefore, in the current study tetrandrine-induced apoptosis in endometrial cancer cells was established through BCL2 family proteins.

Fruits of *Nandina domestica* have been used as a traditional remedy for treating cancer in different countries. Taha et al. [70] analysed the extracts of this plant using Ultrahigh performance liquid chromatography-MS/MS (UPLC-MS/MS). Subsequently, compound-target and compound-target-pathway networks were derived from the Search Tool for Interactions of Chemicals (STITCH) [94], Database for Annotation, Visualisation, and Integrated Discovery (DAVID) [95], KEGG, and Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) [96]. The computed data were then verified via in vitro experiments. Enrichment analyses, performed on the 22 compounds that passed UPLC-MS/MS, revealed the presence of 5 anti-cancer compounds (isoquercitrin, quercetin, berberine, chlorogenic acid, and caffeic acid) (Fig. 6B-F), and 4 molecular targets (AKT serine/threonine kinase 1 [AKT1], Caspase 3 [CASP3], Mitogen-Activated Protein Kinase 1 [MAPK1], and Tumour Protein p53 [TP53]). The authors further analysed the data to identify 15 cancer-related pathways for colorectal, endometrial, and non-small cell lung cancers. The results of the current study revealed that the above-mentioned five phytocompounds exhibit high synergistic interactions with important cancer-related targets and pathways.

The NP emodin (Fig. 6G), extracted from the plant *Rheum palmatum* has traditionally been used to treat cancer

Table 2 Studies that utilise Network pharmacology approaches for anti-cancer drug discovery

| Plant Source	Compound	Cancer	No. of therapeutic targets identified	Analysis tool	References
Stephania tetrandra	Tetrandrine	Endometrial	111	Cytoscape, GO, KEGG	[90]
Solanum nigrum	39 NPs	Bladder	77	STRING, Cytoscape, GO, KEGG	[113]
Cremastra appendiculata	Active components of Shan Ci Gu	Lung	56	Cytoscape, R, GO-BPs, KEGG, AutoDock Vina	[114]
Artemisiae Scopariae Herba, *Gardeniae Fructus,* and *Radix Rhei*	Quercetin, Kaempferol, β-sitosterol, Isorhamnetin, Stigmasterol	Cholangiocarcinoma		TCMSP, Cytoscape, STRING, DAVID, GO, KEGG, AutoDock Vina	[115]
Artemisia annua	19 NPs	Hepatocellular	25	TCMSP, OncoDB.HCC, PharmDB-K, DAVID, GO, KEGG, LibDock	[116]
Platycodon grandiflorum	47 NPs	Lung	20	TCMSP, STRING, Cytoscape, GO, KEGG, Discovery Studio	[101]
Citrus aurantium	Nobiletin, Naringenin, Hesperetin	Lung	5	Cytoscape, GO, KEGG, AutoDock Vina	[117]
Rheum rhabarbarum	10 NPs	Multiple	58	TCMSP, Cytoscape, STRING, DAVID, GO, KEGG, AutoDock Vina	[118]
Zingiber officinale	6 NPs	Colon	34	TCMSP, Cytoscape, Swiss TP, GO, KEGG, Cytoscape, Discovery studio	[119]
Tripterygium wilfordii	8 NPs	Myeloid leukaemia	–	STRING, GO, DAVID	[120]
Tripterygium wilfordii	Kaempferol, Nobiletin, β-sitosterol	Nasopharyngeal	–	TCMSP, DAVID, STRING, GO, KEGG	[121]
Cinnamomum cassia (*L.*), *J. Presl* (*cinnamon*)	Cinnamaldehyde	Breast	59	TCMSP, TCMDT, TCMID, GO, KEGG, Cytoscape	[122]
Scutellaria barbata and *Hedyotis diffusa* (SH)	–	Breast		STRING, KEGG, Cytoscape	[123]
Nandina domestica	Isoquercitrin, quercitrin, berberine, chlorogenic acid, caffeic acid	Colorectal, Endometrial, lung	4	STITCH, DAVID, KEGG, STRING, Cytoscape	[70]
Oldenlandia diffusa	–	Breast		Cytoscape	[124]
Euphorbia spp., *Hyphaene thebaica*, *Artemisia* herba-alba, *Marrubium vulgare*	Luteolin, alternariol, apigenin, aloe-emodin, myricetin	Multiple	36	STITCH, Cytoscape	[125]
Hedyotis difusa Willd	–	Gastric		Cytoscape	[126]
Prunella vulgaris L	–	Breast	31	GO, KEGG, Cytoscape	[127]
Rheum palmatum	Emodin	Lung, liver, pancreatic, breast		Cytoscape	[97]
Mangifera indica, Nigella sativa, Zingiber officinale, Coccinia grandis, Ziziphus jujuba, Ziziphus mauritiana, Cinnamomum cassia	87 NPs	Cervical	35	Network Analyst, Cytoscape, GOnet, Metascape, STRING	[108]
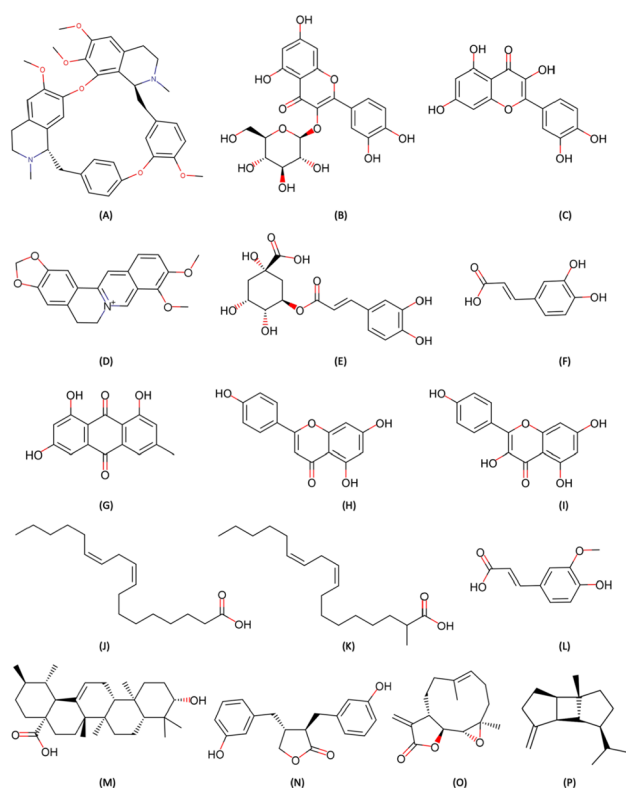
Fig. 6 Structures of **A** tetrandrine, **B** isoquercitrin, **C** quercetin, **D** berberine, **E** chlorogenic acid, **F** caffeic acid, **G** emodin, **H** apigenin, **I** kaempferol, **J** linoleic acid, **K** methyl linoleic acid, **L** ferulic acid, **M** ursolic acid, **N** enterolactone, **O** parthenolide, **P** β-bourbonene

of the lungs, liver, and pancreas. Thus, Zhang et al. studied the therapeutic effects of this NP on the MCF7 cancer cells [97]. The targets of the compound were identified using PharmMapper [98], SwissTargetPrediction [99], and the Traditional Chinese Medicine Systems Pharmacology database and analysis platform (TCMSP) [100]. The authors then used this information to extract genomic information from the UniProt Knowledge database. A protein-target PPI network was thereafter generated using the STRING database, after which Gene Ontology (GO), and KEGG Pathway enrichment analyses were performed and visualised using the R software. It was elucidated that emodin exhibits anti-tumour activity by activating the AhR/CYP1A1 pathway and can be further exploited as a potential lead compound for the treatment of breast cancer.

Deng et al. [101] used the active components of the *Platycodon grandiflorum* (PG) roots to generate a "Drug-Ingredients-Targets-Pathways-Disease" (DITPD) network. This was done to explore the potential molecular mechanism used by PG against different cancers. Various triterpenoid saponins, steroidal saponins, flavonoids, phenolic acids, organic acids, and other compounds were identified from PG which were then screened for ADMET properties. Finally,

apigenin, caffeic acid, kaempferol, linoleic acid, methyl linoleic acid, and ferulic acid (Fig. 6F, H–L) were identified as core compounds. The targets of these active components were obtained using the SwissTargetPrediction, PharmMapper, and TCMSP databases. Previously known disease targets for lung cancer were obtained and a PPI network was generated for these targets using STRING. Degree centrality (DC) was then calculated for this network by CentiScaPe to filter the top 20 targets with the highest DC value. Subsequent enrichment analysis was performed which revealed that these targets are closely related to different cancer pathways like the TNF, MAPK and PI3K-AKT signalling pathways. The top 10 annotated pathways were then used to construct a DITPD network in Cytoscape [102]. A network of 40 nodes (1 drug, 8 core component, 20 core targets, 10 pathways and one disease) was obtained, in which targets of the core ingredients of PG were found to be distributed in these pathways. Finally, a molecular docking study verified the interactions between the core components of PG (apigenin, caffeic acid, kaempferol, linoleic acid, methyl linoleic acid and ferulic acid) and the target molecules indicating a consistent relationship between NPs and the targets.

Breast cancer is the leading cause of death in women worldwide and therefore, Jha et al. [79] identified ADMET adherent phytomolecules against six of the most important breast cancer targets viz., PR (PDB ID: 4OAR), EGFR (PDB ID: 2J6M), mTOR (PDB ID: 4DRH), p53R2 (PDB ID: 3HF1), CTLA4 (PDB ID: 1DQT), and CDK8 (PDB ID: 6T41). The authors used Dr. Dukes Phytochemical and Ethnobotanical Database (https://phytochem.nal.usda.gov/phytochem/search) to acquire a list of 68 phytochemicals with established anti-cancer activities which were then evaluated for their Drug-likelihood using SwissADME [103] and ADMETlab 2.0 [104]. The CASTp server [105] was used to identify the active sites of each of the proteins after which 38 of the ADMET-filtered ligands were docked to these proteins using AutoDock Vina [106]. Molecular docking revealed that ursolic acid, enterolactone, parthenolide, and berberine (Fig. 6D, M–O) show increased binding affinity in comparison to the known drugs and therefore can be used to inhibit the breast cancer targets.

Among the members of the genus Ficus, the significance of *Ficus carica* L. is widespread because of the medicinal value of its leaves and roots. Gurung et al. [78] performed an extensive literature review and identified 68 bioactive compounds present in the leaves, roots and barks of *F. carica*. The physicochemical features of these NPs were checked in DataWarrior v4.6.1, which revealed that 13 of the 68 NPs were safe for use. Thereafter, 3D structures of six important protein targets of various cancers viz., CDK2 (PDB ID: 1DI8), CDK6 (PDB ID: 1XO2), Topoisomerase I (PDB ID: 1T8I), Topoisomerase II (PDB ID: 1ZXM), BCL2 (PDB ID: 2O2F), and VEGFR2 (PDB ID: 2OH4) were downloaded

from PDB, prepared and taken for molecular docking using AutoDock v4.2. For each of the six proteins, the best bound NPs were studied using MD simulations in LARMD [107] and ultimately identified β-bourbonene (Fig. 6P) which exhibited strong binding with most of the cancer targets.

As cervical cancer is a major concern for developing nations, therefore Aarthy et al. [108] adopted a systems pharmacology-based multi-omics approach wherein they exploited the curative potential of plant-based NPs against Human Papilloma Virus (HPV) mediated cervical cancer. Firstly, the authors used the ArrayExpress database [109] (https://www.ebi.ac.uk/arrayexpress/) to retrieve human transcriptome datasets of cervical cancer patients. After manual curation of the data using Microsoft Excel, it was imported to Network Analyst 3.0 [110] (https://www.networkanalyst.ca/) wherein differential gene expression analysis was performed, and 384 immune-related genes were identified using the Limma statistical model with adjusted *p* value < 0.05. Additionally, network analyst was also used to perform over-representation analysis, functional enrichment and identify tissue-specific interactions. Secondly, 87 pharmacologically active constituents of the Indian plants *Mangifera indica, Nigella sativa*, *Zingiber officinale*, *Citrus grandis*, *Ziziphus jujube*, *Ziziphus mauritiana* and *Cinnamomum cassia* were identified from literature and web resources. The immune response related protein targets of the 87 NPs were thereafter retrieved from SwissTargetPrediction (http://swisstargetprediction.ch/index.php) which highlighted that 79 of the 87 NPs were significantly targeting 35 of the 384 differentially expressed genes (DEGs). Thirdly, Compound target networks (CTNs) were constructed for these 79 NPs and 35 DEGs using Cytoscape v3.8.0 in order to identify the key multi-target qualities of the identified phytocompounds. Additionally, Gene ontology, enrichment and molecular interactome analysis were also performed using GOnet [111] (https://tools.dice-database.org/GOnet/), Metascape (https://metascape.org/gp/index.html#/main/step1) and STRING Viruses v10.5 [112] (http://viruses.string-db.org/). The above-mentioned integrative approach thus highlighted the multi-targeted approach of plant-based NPs from the Indian subcontinent that are capable of tackling HPV-mediated cervical cancer.

Machine learning-based approaches and NP applications

ML techniques and algorithms are now extensively used for different aspects of the drug discovery and development process [128]. Large amounts of data and different algorithms are used to train the ML models such that they can learn how to perform tasks independent of manual programming. Depending on the learning process of algorithms, they are broadly categorised as either supervised or unsupervised learning. Supervised learning models are trained on both input as well as output data. The categorical or continuous relationships of labelled datasets are utilised for training the supervised ML models. The expected outcome of a model trained through supervised learning is the prediction of future outputs of known input data. However, in the case of unsupervised learning, no such labelled data (i.e. unlabelled) are given, which forces the model to identify the intrinsic patterns in the input data based on which the initial data are then transformed into several meaningful clusters independently (without supervision) [129]. Through the use of models trained on a large number of chemical compounds, efficient and cost-effective modules can be developed. These models utilise the exhaustive approaches of the ML algorithms to unearth the underlying relationships between experimental observations and chemical features of the molecule. The models are then, used to predict the experimental outcome using the chemical information of new molecules [128]. In drug discovery, the traditional approach is expensive and time-consuming. Owing to this, the use of ML tools in this field is steadily gaining momentum as an essential approach for various purposes such as mining relevant chemical information, prediction of biological attributes and potentially active biological molecules [130], identification of novel targets [131], evidence extraction for target disease associations [132], increasing the understanding of disease and non-disease phenotypes [133], development of better biomarkers for prognosis, progression and drug efficacy [134], etc.

In this section, we discuss only those studies wherein ML algorithms have been applied to promote NP-based anti-cancer drug discovery and simultaneously involve the development of applications based on ML. A brief overview of the web servers and softwares on ML and how they have been used in anti-cancer drug discovery is given below (Table 3).

CASE I

As NPs being one of the most diverse sources of lead compounds for drug discovery, Chen et al. [21] developed a method for calculating the NP-likeness of compounds to guide the identification of new lead compounds for NP drug discovery. The authors generated a reference NP dataset composed of 201,761 unique molecules from 18 virtual and 9 physical NP libraries. For the compilation of the small molecule set (SM), an equal number of 201,761 compounds were randomly selected from ZINC database such that any molecule present in the NP dataset was removed from the SM dataset. The NP and SM datasets were then merged to create a total of 403,522 compounds which were then randomly split into a training and test set in the ratio 4:1. From these compounds, three different descriptor sets

Table 3 A brief overview of the usage of ML applications in NP-based drug discovery

ML Tool	Utilisation in NP research	References
NP-Scout	To determine the NP similarity scores of anthropogenic substances discovered using mass spectrometry, Price et al., 2021 had employed NP-Scout	[135]
	Using NP-Scout, Grigalunas et al., 2021, determined the likelihood that a number of pseudo-NPs were indeed natural in nature	[136]
	Chen et al., 2020 employed NP-Scout to identify which complex small molecules are likely to be NPs or natural product-like	[137]
NaPLeS	Vaskevych et al., 2021 had utilised both Ertl's NP-likeness Scorer and Sorokina's NaPLeS tool to calculate the NP-likeness of several analogs of vasicinone alkaloids that they had developed	[138]
	To determine the NP-likeness score of chemicals found using spectroscopy, Lianza et al., 2021 employed this approach	[139]
NPred	Agarwal et al., 2022 had used this tool to screen a large number of NPs from 12 NP libraries and shortlist a few molecules that have a higher probability of anti-cancer activity	[37]
NP-likeness Scorer	This method was utilised by Mikhnevich et al., 2021 to determine the NP-likeness of β-Lactamase inhibitors	[140]
	Zhou et al., 2022 determined the NP-likeness of a large number of unidentified metabolites using this method	[141]
	Using this method, Elend et al., 2022 estimated the NP-likeness of several compounds that they had designed to target the SARS-CoV-2 Main Protease	[142]
	Akbarzadeh et al., 2022 had used this tool to calculate the NP-likeness of the pseudo-NPs designed against the RHO GDP-dissociation inhibitor1 (RHOGDI1)	[143]
	Using this method, Ceballos et al., 2019 was able to determine the NP-likeness of the potential inhibitors identified against the GLUT-1 and GLUT-3 glucose transporters	[144]

(Morgan2 fingerprints, MACCS keys, and 206 two-dimensional physicochemical property descriptors) were generated for training the random forest classifiers of the ML model. To measure how accurately the models were able to rank the NPs, the performance of the ML models was evaluated using Matthew's correlation coefficient and area under the receiver operating characteristic curve. After validation of the ML models, the model based on MACCS fingerprints was selected for use as it had demonstrated superior classification ability for the independent test set. The ability of this model to properly identify NPs was also tested using an external validation set using the Dictionary of natural products (DNP) which showed that 95% of the NPs were correctly identified by the ML model developed by the authors. Finally, a web application was developed named "NP-Scout" which can be accessed at https://nerdd.univie.ac.at/npscout/.

CASE II

In order to predict the activity of unknown molecules, traditional approaches are based on structure activity relationship (SAR), however, deviating from this approach, Yue et al. [91] developed a ML-based approach using the cumulative data of both gene expression and chemical properties of NPs. To develop the NP response predictors, 17 NP drugs across a number of cell lines from GDSC (Genomics of Drug Sensitivity in Cancer) [145] were retrieved. These 17 NPs were then screened across an average of 495 cell lines per drug where the sensitivity

(IC_{50}) values for all NPs were used to classify the cell lines into three groups (Sensitive, Resistant and Intermediate) using the K-Means clustering algorithm in Waikato Environment for Knowledge Analysis (WEKA) [146]. The samples in the sensitive and resistant groups were used to build different ML models using decision tree, support vector machine, random forest, and rotation forest algorithms. For building the ML models, thirteen NPs having 6450 cancer cell line-NP interactions were randomly selected for the training set and the remaining four NPs with 1970 cancer cell line-NP interactions was used as the test set. Thereafter, the performance of these models was evaluated to identify the best algorithm suitable for predicting the cancer cell lines' sensitivity to the NPs. It showed that all four methods demonstrated good results based on tenfold cross-validation of the training set. The second method of evaluation was also adopted by the authors, wherein CancerHSP database was searched for anti-cancer herbs used in systems pharmacology and NP-related studies. Two anti-tumour NPs, curcumin and resveratrol, were selected for evaluating the cancer cell sensitivity to NPs. For curcumin, six out of seven cell lines were correctly predicted by the model described in this study whereas in the case of resveratrol, five out of eight were correctly predicted. Thus, this study demonstrates that using advanced techniques like ML, predictive models can be trained on cancer cell line data which can identify the potential NPs with anti-cancer activity.

CASE III

Rayan et al. [147] developed a predictive model for identifying NPs with anti-cancer activity from a set of 617 approved anti-cancer drugs and 2892 NPs. For developing the predictive model, 617 anti-cancer drugs retrieved from the CMC (Comprehensive Medicinal Chemistry) database and NCI Drug Dictionary were used as actives. On the other hand, a set of 2892 phytochemicals retrieved from AnalytiCon Discovery was used as the inactive set. Molecular Operating Environment (MOE) was used to calculate the 1D and 2D physicochemical properties (descriptors) of the molecules present in the active and inactive datasets, respectively. The molecules were then split into a training set comprised of 66.7% of the dataset and a test set comprised of the remaining molecules. Thereafter, the iterative stochastic elimination (ISE) algorithm was used to build a predictive model capable of indexing NPs for potential anti-cancer activity based on the distinction between active and inactive ligands. From the initial NP dataset taken, twelve NPs were highly ranked as potential anti-cancer agents by the ISE model and subsequent literature search demonstrated that three of the identified compounds i.e. neoechinulin, colchicine, and piperolactam (Fig. 7A–C) have established experimental evidence as anti-cancer agents, thereby further demonstrating the validity of this model.

CASE IV

One of the earliest and foremost applications of ML in the field of NP drug discovery was achieved by Ertl et al. [148] wherein they used a Bayesian approach that identified how closely a given molecule resembles the structural space occupied by NPs. The DNP was taken as the set of NPs and a set of 290,000 commercially available synthetic molecules was considered for the synthetic molecule (SM) set. The authors considered intricate structural elements like particular substructures that are characteristically present in NPs to construct the NP-likeness score. Substructural fragments were generated for both the NP and SM set and the
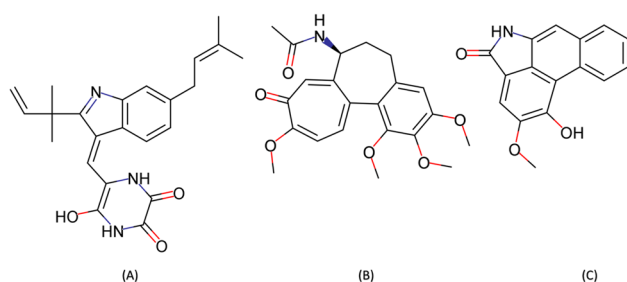
distribution between the fragments of the two groups were analysed. After model development, it was cross validated using standard statistical measures like area under the curve as well as enrichment plots. Additional testing of the NP-likeness scorer using a set of novel molecules that were not present in the starting NP set exhibited that the system was able to correctly identify ~ 93% of the molecules as NPs. Therefore, this tool can be a part of standard VS exercises along with other screening parameters.

In addition to the above, the same research group [23] has also developed a Natural-Product-Likeness scoring system, that can be downloaded as a standalone java package from https://sourceforge.net/projects/np-likeness/. It uses Taverna version 2.2, and a few other open-source Java libraries to calculate the NP-likeness of compounds. The model has been trained using NPs sourced from open-access databases, ChEMBL [149], and the Traditional Chinese Medicine Database (TCMD) [150]. For each query molecule this tool assigns a signature to each molecule, which is then used for "Signature Scoring" and calculating the NP-likeliness of compounds.

The NaPLeS NP-likeness scorer is a web-based application developed by Sorokina et. al. [22], to calculate the NP-likeness of compounds. This open-source application has been developed using a training set of 3,64,807 unique NPs and 4,89,780 unique synthetic compounds. The molecular structures and corresponding scores are stored in the MySQL 5.8 Docker Image. A unique identifier is allotted to every sourced molecule, and the NP-likeness score is computed on the parameters of heavy and total atom counts, the number of rings, the number of repeated fragments, and the number of predominant heavy atoms. The authors have also developed a web application using the Spring Boot framework which is available for public access at http://naples.naturalproducts.net.

CASE V

Researchers have also tried to generate classification models of the relationship between the chemical structures of plant-based NPs and their anti-cancerous activity using QSAR-based ML algorithms like Naive Bayesian classifier (NB), sequential minimal optimisation (SMO), instance-based learner (IBK) and random forest (RF) [20]. Using these algorithms and 881 PubChem fingerprints, the authors developed different classification models to predict the inhibitory potential of plant compounds taken from the NPACT database. The random forest algorithm using 100 trees was found to be the best among various classifiers and achieved 81.58% sensitivity, 72.44% specificity, 77.6% accuracy. A frequency-based feature selection method further revealed that the top 10 fingerprints selected by this study were also

Fig. 7 Structures of **A** neoechinulin, **B** colchicine, and **C** piperolactam as identified by the works of Rayan et al., 2017

Table 4 Existing phytomolecule resources

Database	Year	URL	Compounds	Description	References
Naturally occurring Plant-based Anti-cancerous Compound-Activity-Target Database (NPACT)	2012	http://crdd.osdd.net/raghava/npact/	1574	Database of plant-based anti-cancer compounds	[14]
Nuclei of Bioassays, Ecophysiology, and Biosynthesis of Natural Products Database (NuBBE_DB)	2017	https://nubbe.iq.unesp.br/portal/nubbe-search.html	2000+	An extensive collection of biogeochemical information about Brazilian biodiversity	[86]
Phytochemica	2015	http://faculty.iiitd.ac.in/~bagler/webservers/Phytochemica/index.php	963	Database of phytochemicals from five medicinally important plants from the Himalayas (*Atropa belladonna* (ATBE), *Catharanthus roseus* (CARS), *Heliotropium indicum* (HEIN), *Picrorhiza kurroa* (PIKU), and *Podophyllum hexandrum* (POHX))	[83]
TM-MC	2015	https://informatics.kiom.re.kr/compound/index.jsp	14,000	Collection of medicinal compounds from traditional Japanese, Korean, and Chinese medicinal materials	[155]
Carotenoids Database	2017	http://carotenoiddb.jp/	1204	Database of carotenoids sourced from plants, algae, and photosynthetic bacteria	[156]
Chinese Ethnic Minority Traditional Drug Database (CEMTDD)	2014	http://www.cemtdd.com/	3700	Database on traditional Chinese plant-based NPs from ethnic minorities	[159]
The Traditional Chinese Medicine Systems Pharmacology database and analysis platform (TCMSP)	2014	https://old.tcmsp-e.com/tcmsp.php	29,384	Chinese traditional medicinal compounds, and VS facility for researchers conducting computer-aided drug design (CADD)	[100]
AfroCancer	2014	-	400+	Dataset of African medicinal plant-based anti-cancer compounds	[7]
CamMedNP	2013	-	2500+	Compounds from 224 different medical plant species, from the Cameroonian flora	[152]
MedPServer	2019	http://bif.uohyd.ac.in/medserver/	1124	Database of phytomolecules isolated from 212 different medicinal plant sources from North-East region of India	[153]
Collective Molecular Activities of Useful Plants (CMAUP)	2019	http://bidd.group/CMAUP/	47,645	Information on 646 biological targets of 47,645 plant ingredients from 5645 plants	[160]
Encyclopaedia of Traditional Chinese Medicine (ETCM)	2019	http://www.tcmip.cn/ETCM/	7284	Collection of compounds from herbs used in traditional Chinese medicine	[161]
TriForC	2018	http://bioinformatics.psb.ugent.be/triforc/#/home	271	Database on Triterpene metabolism enzymes that have been reported and experimentally validated	[157]
Yet Another Traditional Chinese Medicine Database (YaTCM)	2018	http://cadd.pharmacy.nankai.edu.cn/yatcm/home	47,696	Database of NPs used in TCM with analysis and visualisation tools	[162]
Traditional Chinese Medicine Integrative Database (TCMID)	2013	http://47.100.169.139:8000/tcmid/	25,210	Database on Traditional Chinese Medicine, which has been linked to drug and disease databases for easier segregation	[163]
TIPdb-3D	2013	http://cwtung.kmu.edu.tw/tipdb	8853	A Database of Anti-cancer, Antiplatelet, and Anti-tuberculosis Phytochemicals from Indigenous Plants in Taiwan	[154]
Alkamid database	2012	http://alkamid.ugent.be/alkamidhome.php	300+	Database on plant-derived N-alkylamides	[164]

Table 4 (continued)

Database	Year	URL	Compounds	Description	References
Phenol-Explorer	2010	http://phenol-explorer.eu/	502	Online database of polyphenol components in plant-based food	[158]
SerpentinaDB	2015	http://faculty.iiitd.ac.in/~bagler/webservers/SerpentinaDB/	147	A database of Plant-derived molecules (PDMs) from *Rauwolfia serpentina* along with their plant part source, chemical classification, IUPAC, SMILES, physicochemical properties, and 3D chemical structures with associated references	[84]
IMPPAT 2.0	2018	https://cb.imsc.res.in/imppat	17,967	A database of phytomolecules from 4010 Indian medicinal plants	[165]

present in a wide range of natural anti-cancer drugs like vincristine, vinblastine, and paclitaxel.

Existing phytomolecule databases and web resources

NP-based drug discovery is an extensive field of study, in which phytomolecule-based anti-cancer drug discovery is a niche domain. There exists a vast amount of literature and experimental data regarding plant-based NPs that show potential activity against different diseases including cancer. These data are a treasure trove of unique scaffolds and compounds that have been accumulated and represented in the form of different phytomolecule databases as given in Table 4. Most of these databases focus on phytomolecules and their therapeutic potential. There exists a large number of resources that provide extensive lists of NPs used for drug discovery purposes [2, 9, 26, 28, 151, 152], but in accordance with the scope of the review, this section briefly discusses those resources that specifically cater to the information and use of phytomolecules in drug discovery.

The phytomolecule databases/resources described in the current review (Table 4) have been classified based on their geographical locations, compound class, traditional medicine system and bioactivity (Fig. 8). For example, in the geographical location category, we observe that there exist databases that contain the phytomolecules native to certain specific regions of the world like India [83, 84, 153], Brazil [86], Africa [7], Cameroon [152] and Taiwan [154]. But more specifically to the Indian subcontinent, we observe that there are certain highly specific databases like Phytochemica [83] and MedPServer [153] that contain information on the phytomolecules found in plants present in the Himalayan bioresource and the North-Eastern region of the country (Fig. 8). Closely related to this group is the Traditional medicine system which is based on the easily and widely available flora of their native regions. For example, the TM-MC [155] database contains information about the phytomolecules found in the traditional plant-based medicines of Japan, Korea, and China while in the case of Traditional Chinese Medicine (TCM) a large number of resources are available. Further, in the case of TCM, it is seen that there exists a database that houses the plant-based NPs used by the Chinese ethnic minority communities (Fig. 8). Since phytomolecules are highly diverse they can be classified based on their structural similarities also. Therefore, a number of databases have been built that cater to the phytomolecules of specific classes like carotenoids [156], triterpenes [157], polyphenols [158], etc. (Fig. 8). Finally, there also exist databases that report the bioactivity information of different phytomolecules against different cancers at the cell line and protein levels, e.g. NPACT (Fig. 8). Therefore, all these databases can be evaluated before selecting and curating a library for CADD

Fig. 8 Different phytomolecule resources and their classification

Geographical location	Compound class	Traditional medicine system	Bioactivity
➤ Brazilian (NuBBE$_{DB}$)	➤ Carotenoids (Carotenoids$_{DB}$)	➤ Japanese, Korean, Chinese (TM-MC)	➤ Cancer Specific (NPACT)
➤ Himalayan (Phytochemica)	➤ Triterpenes (TriForC)	➤ Chinese ethnic minority (CEMTDD)	
➤ African (AfroCancer)	➤ N-alkylamides (Alkamid$_{DB}$)	➤ Chinese (TCMSP), (ETCM), (YaTCM), (TCMID)	
➤ Cameroon (CamMedNP)	➤ Poly-phemols (Phenol-Explorer)	➤ Indian (IMPPAT 2.0)	
➤ North-East India (MedPServer)			
➤ Taiwan (TIPdb-3D)			
➤ Indian (Serpentina$_{DB}$)			

against any cancer target. Thus, selection of an appropriate library is an important step in CADD as it determines the type of molecules that are being studied as well as the nature of the selected/identified molecules at the end of the study.

We hope that the aggregated information provided in this section will serve as an important resource for anyone working in the field of phytomolecule-based anti-cancer drug discovery.

Gaps and future prospects

The advent of computer-aided techniques has transformed the entire landscape of drug discovery research. In the field of anti-cancer drug discovery, the use of plant-derived NPs have been discussed in this review, which highlights that some of the traditional techniques discussed here are being used more in comparison to, the newer techniques. For example, ML-based techniques were found to be extensively used in both anti-cancer research and NP-based research separately, but limited evidence exists for their use in NP derived anti-cancer drug discovery. A reason for this may lie in the fact that few ML-based models have been developed for predicting the anti-cancer activity of NPs, out of which only some are publicly available. This in turn is an area of opportunity for future research. More efforts need to be put in developing publicly available tools and resources regarding the use of ML-based techniques in drug discovery. In this review, we have also highlighted the various plant-derived NP-based tools and databases that are important for CADD. But there is still a gap in the resources available for NP derived anti-cancer drug discovery, with NPACT and AfroCancer being the only two databases that cater to this specific genre. Thus, there is a need for more structured and systematised resources that can successfully accommodate the complexity and

diversity of the data required for NP drug discovery. The current landscape of plant-derived NP resources has the apparent disadvantage of either being very specific or very general in nature. Consequently, the diversity of different plant species is somewhat lost and even if they are very comprehensive in nature, they have no method to demarcate those NPs that are important specifically with respect to anti-cancer drug discovery. Therefore, this review highlights several areas of plant-based anti-cancer drug discovery that need more attention from the scientific community worldwide.

Conclusion

Overall, the current review discusses in detail the niche area of anti-cancer drug discovery from plant-based NPs. Efforts have been made to describe the current practices like molecular docking and dynamics, VS, pharmacophore modelling, and QSAR that are widely being used in the drug discovery process. Additionally, we have also highlighted and discussed in detail the more recent techniques like ML and network pharmacology that are not as widely used as the traditional techniques. Therefore, we hope this review will be of use to the scientific community not only to understand the tools and techniques used in traditional drug discovery process but also to learn and implement the next-generation techniques as well.

Declarations

Conflict of interest There is no conflict of interest to declare.

References

1. Sung H, Ferlay J, Siegel RL et al (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71:209–249. https://doi.org/10.3322/caac.21660

2. Choudhari AS, Mandave PC, Deshpande M et al (2019) Phytochemicals in cancer treatment: from preclinical studies to clinical practice. Front Pharmacol 10:1614. https://doi.org/10.3389/fphar.2019.01614

3. Cragg GM, Newman DJ (2013) Natural products: a continuing source of novel drug leads. Biochim Biophys Acta 1830:3670–3695. https://doi.org/10.1016/j.bbagen.2013.02.008

4. Thomford NE, Senthebane DA, Rowe A et al (2018) Natural products for drug discovery in the 21st century: innovations for novel drug discovery. Int J Mol Sci 19. https://doi.org/10.3390/ijms19061578

5. Sharma A, Sharma S, Gupta M et al (2018) Pharmacokinetic profiling of anticancer phytocompounds using computational approach. Phytochem Anal 29:559–568. https://doi.org/10.1002/pca.2767

6. Ntie-Kang F, Mbah JA, Lifongo LL et al (2013) Assessing the pharmacokinetic profile of the CamMedNP natural products database: an in silico approach. Org Med Chem Lett 3:10. https://doi.org/10.1186/2191-2858-3-10

7. Ntie-Kang F, Nwodo JN, Ibezim A et al (2014) Molecular modeling of potential anticancer agents from African medicinal plants. J Chem Inf Model 54:2433–2450. https://doi.org/10.1021/ci5003697

8. Fatima S, Gupta P, Sharma S et al (2019) ADMET profiling of geographically diverse phytochemical using chemoinformatic tools. Future Med Chem 12:69–87. https://doi.org/10.4155/fmc-2019-0206

9. Cragg GM, Pezzuto JM (2016) Natural products as a vital source for the discovery of cancer chemotherapeutic and chemopreventive agents. Med Princ Pract 25:41–59. https://doi.org/10.1159/000443404

10. Singh H, Kumar R, Singh S et al (2016) Prediction of anticancer molecules using hybrid model developed on molecules screened against NCI-60 cancer cell lines. BMC Cancer 16:1–10. https://doi.org/10.1186/s12885-016-2082-y

11. Singh S, Sharma B, Kanwar SS, Kumar A (2016) Lead phytochemicals for anticancer drug development. Front Plant Sci 7:1–13. https://doi.org/10.3389/fpls.2016.01667

12. Moraes DFC, de Mesquita LSS, do Amaral FMM et al (2017) Anticancer drugs from plants. In: Malik S (ed) Biotechnology and production of anti-cancer compounds. Springer, Cham, pp 121–142

13. Newman DJ, Cragg GM (2016) natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79:629–661. https://doi.org/10.1021/acs.jnatprod.5b01055

14. Mangal M, Sagar P, Singh H et al (2013) NPACT: naturally occurring plant-based anti-cancer compound-activity-target database. Nucleic Acids Res 41:1124–1129. https://doi.org/10.1093/nar/gks1047

15. Marx V (2013) Biology: the big challenges of big data. Nature 498:255–260. https://doi.org/10.1038/498255a

16. Rampogu S, Kim SM, Son M et al (2020) A computational approach with biological evaluation: combinatorial treatment of curcumin and exemestane synergistically regulates ddx3 expression in cancer cell lines. Biomolecules 10:1–20. https://doi.org/10.3390/biom10060857

17. Jairajpuri DS, Mohammad T, Adhikari K et al (2020) Identification of sphingosine kinase-1 inhibitors from bioactive natural products targeting cancer therapy. ACS Omega 5:14720–14729. https://doi.org/10.1021/acsomega.0c01511

18. Alkaff AH, Saragih M, Imana SN et al (2021) Identification of DNA methyltransferase-1 inhibitor for breast cancer therapy through computational fragment-based drug design. Molecules 26:1–18. https://doi.org/10.3390/molecules26020375

19. Yadav D, Nath Mishra B, Khan F (2019) 3D-QSAR and docking studies on ursolic acid derivatives for anticancer activity based on bladder cell line T24 targeting NF-kB pathway inhibition. J Biomol Struct Dyn 37:3822–3837. https://doi.org/10.1080/07391102.2018.1528888

20. Dhiman K, Agarwal SM (2016) NPred: QSAR classification model for identifying plant based naturally occurring anti-cancerous inhibitors. RSC Adv 6:49395–49400. https://doi.org/10.1039/c6ra02772e

21. Chen Y, Stork C, Hirte S, Kirchmair J (2019) NP-scout: machine learning approach for the quantification and visualization of the natural product-likeness of small molecules. Biomolecules 9:43. https://doi.org/10.3390/biom9020043

22. Sorokina M, Steinbeck C (2019) Naples: a natural products likeness scorer—web application and database. J Cheminform 11:1–7. https://doi.org/10.1186/s13321-019-0378-z

23. Jayaseelan KV, Moreno P, Truszkowski A et al (2012) Natural product-likeness score revisited: an open-source, open-data implementation. BMC Bioinformatics 13:106. https://doi.org/10.1186/1471-2105-13-106

24. Kim C, Kim B (2018) Anti-cancer natural products and their bioactive compounds inducing ER stress-mediated apoptosis: a review. Nutrients 10:1021. https://doi.org/10.3390/nu10081021

25. Ezzati M, Yousefi B, Velaei K, Safa A (2020) A review on anti-cancer properties of Quercetin in breast cancer. Life Sci 248:117463. https://doi.org/10.1016/j.lfs.2020.117463

26. Blowman K, Magalhães M, Lemos MFL et al (2018) Anticancer properties of essential oils and other natural products. Evid Based Complement Altern Med 2018:3149362. https://doi.org/10.1155/2018/3149362

27. Chen J, Li W, Yao H, Xu J (2015) Insights into drug discovery from natural products through structural modification. Fitoterapia 103:231–241. https://doi.org/10.1016/j.fitote.2015.04.012

28. Li G, Lou HX (2018) Strategies to diversify natural products for drug discovery. Med Res Rev 38:1255–1294. https://doi.org/10.1002/med.21474

29. Lin X, Li X, Lin X (2020) A review on applications of computational methods in drug screening and design. Molecules 25:1–17. https://doi.org/10.3390/molecules25061375

30. Prada-Gracia D, Huerta-Yépez S, Moreno-Vargas LM (2016) Application of computational methods for anticancer drug discovery, design, and optimization. Bol Med Hosp Infant Mex 73:411–423. https://doi.org/10.1016/j.bmhime.2017.11.040

31. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28:235–242. https://doi.org/10.1093/nar/28.1.235

32. Raj U, Kumar H, Varadwaj PK (2017) Molecular docking and dynamics simulation study of flavonoids as BET bromodomain inhibitors. J Biomol Struct Dyn 35:2351–2362. https://doi.org/10.1080/07391102.2016.1217276

33. Wermuth CG, Villoutreix B, Grisoni S et al (2015) Strategies in the search for new lead compounds or original working hypotheses. In: Wermuth CG, Aldous D, Raboisson P, Rognan D (eds) The practice of medicinal chemistry, 4th edn. Academic Press, San Diego, pp 73–99

34. Bommu UD, Konidala KK, Pabbaraju N, Yeguvapalli S (2017) Ligand-based virtual screening, molecular docking, QSAR and pharmacophore analysis of quercetin-associated potential novel analogs against epidermal growth factor receptor. J Recept

Signal Transduct 37:600–610. https://doi.org/10.1080/10799893.2017.1377237

35. Labbé CM, Rey J, Lagorce D et al (2015) MTiOpenScreen: a web server for structure-based virtual screening. Nucleic Acids Res 43:W448–W454. https://doi.org/10.1093/nar/gkv306

36. Zarezade V, Abolghasemi M, Rahim F et al (2018) In silico assessment of new progesterone receptor inhibitors using molecular dynamics: a new insight into breast cancer treatment. J Mol Model 24. https://doi.org/10.1007/s00894-018-3858-6

37. Agarwal SM, Nandekar P, Saini R (2022) Computational identification of natural product inhibitors against EGFR double mutant (T790M/L858R) by integrating ADMET, machine learning, molecular docking and a dynamics approach. RSC Adv 12:16779–16789. https://doi.org/10.1039/d2ra00373b

38. Muhseen ZT, Li G (2020) Promising terpenes as natural antagonists of cancer: an in-silico approach. Molecules 25:1–17. https://doi.org/10.3390/molecules25010155

39. Sterling T, Irwin JJ (2015) ZINC 15 – Ligand discovery for everyone. J Chem Inf Model 55:2324–2337. https://doi.org/10.1021/acs.jcim.5b00559

40. Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 30:2785–2791. https://doi.org/10.1002/jcc.21256

41. Biswas R, Chowdhury N, Mukherjee R, Bagchi A (2018) Identification and analyses of natural compounds as potential inhibitors of TRAF6-Basigin interactions in melanoma using structure-based virtual screening and molecular dynamics simulations. J Mol Graph Model 85:281–293. https://doi.org/10.1016/j.jmgm.2018.09.008

42. Pierce BG, Wiehe K, Hwang H et al (2014) ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. Bioinformatics 30:1771–1773. https://doi.org/10.1093/bioinformatics/btu097

43. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein-ligand docking using GOLD. Proteins 52:609–623. https://doi.org/10.1002/prot.10465

44. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res 33:W363–W367. https://doi.org/10.1093/nar/gki481

45. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model 49:377–389. https://doi.org/10.1021/ci800324m

46. Friesner RA, Murphy RB, Repasky MP et al (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49:6177–6196. https://doi.org/10.1021/jm051256o

47. Hilbig M, Urbaczek S, Groth I et al (2013) MONA—interactive manipulation of molecule collections. J Cheminform 5:1–10. https://doi.org/10.1186/1758-2946-5-38

48. Hindle SA, Rarey M, Buning C, Lengauer T (2002) Flexible docking under pharmacophore type constraints. J Comput Aided Mol Des 16:129–149. https://doi.org/10.1023/A:1016399411208

49. Güner OF, Bowen JP (2014) Setting the record straight: the origin of the pharmacophore concept. J Chem Inf Model 54:1269–1283. https://doi.org/10.1021/ci5000533

50. Kalyaanamoorthy S, Chen YPP (2011) Structure-based drug design to augment hit discovery. Drug Discov Today 16:831–839. https://doi.org/10.1016/j.drudis.2011.07.006

51. Babu TMC, Rammohan A, Baki VB et al (2016) Development of novel HER2 inhibitors against gastric cancer derived from flavonoid source of *Syzygium alternifolium* through molecular dynamics and pharmacophore-based screening. Drug Des Dev Ther 10:3611–3632. https://doi.org/10.2147/DDDT.S111914

52. Singh PK, Silakari O (2018) Pharmacophore and molecular dynamics based activity profiling of natural products for kinases involved in lung cancer. J Mol Model 24. https://doi.org/10.1007/s00894-018-3849-7

53. Alamri MA, Afzal O, Alamri MA (2020) Computational screening of natural and natural-like compounds to identify novel ligands for sigma-2 receptor. SAR QSAR Environ Res 31:837–856. https://doi.org/10.1080/1062936X.2020.1819870

54. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738. https://doi.org/10.1038/nprot.2010.5

55. Nastasi G, Miceli C, Pittalà V et al (2017) S2RSLDB: a comprehensive manually curated, internet-accessible database of the sigma-2 receptor selective ligands. J Cheminform 9:3. https://doi.org/10.1186/s13321-017-0191-5

56. Kumar V, Parate S, Thakur G et al (2021) Identification of cdk7 inhibitors from natural sources using pharmacoinformatics and molecular dynamics simulations. Biomedicines 9:1197. https://doi.org/10.3390/biomedicines9091197

57. Hähnke VD, Kim S, Bolton EE (2018) PubChem chemical structure standardization. J Cheminform 10:1–40. https://doi.org/10.1186/s13321-018-0293-8

58. Banerjee P, Erehman J, Gohlke BO et al (2015) Super Natural II—a database of natural products. Nucleic Acids Res 43:D935–D939. https://doi.org/10.1093/nar/gku886

59. Cherkasov A, Muratov EN, Fourches D et al (2014) QSAR modeling: where have you been? where are you going to? J Med Chem 57:4977–5010. https://doi.org/10.1021/jm4004285

60. Pratim Roy P, Paul S, Mitra I, Roy K (2009) On two novel parameters for validation of predictive QSAR models. Molecules 14:1660–1701. https://doi.org/10.3390/molecules14051660

61. Alam S, Khan F (2017) 3D-QSAR studies on maslinic acid analogs for anticancer activity against breast cancer cell line MCF-7. Sci Rep 7:1–13. https://doi.org/10.1038/s41598-017-06131-0

62. Chen G, Zhou D, Li XZ et al (2017) A natural chalcone induces apoptosis in lung cancer cells: 3D-QSAR, docking and an in vivo/vitro assay. Sci Rep 7:1–10. https://doi.org/10.1038/s41598-017-11369-9

63. Yadav DK, Kumar S, Saloni et al (2017) Molecular docking, QSAR and ADMET studies of with anolide analogs against breast cancer. Drug Des Dev Ther 11:1859–1870. https://doi.org/10.2147/DDDT.S130601

64. Spitzer R, Jain AN (2012) Surflex-Dock: Docking benchmarks and real-world application. J Comput Aided Mol Des 26:687–699. https://doi.org/10.1007/s10822-011-9533-y

65. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/s0169-409x(00)00129-0

66. Shukla A, Tyagi R, Meena S et al (2020) 2D- and 3D-QSAR modelling, molecular docking and in vitro evaluation studies on 18β-glycyrrhetinic acid derivatives against triple-negative breast cancer cell line. J Biomol Struct Dyn 38:168–185. https://doi.org/10.1080/07391102.2019.1570868

67. Meneses-Sagrero SE, Rascón-Valenzuela LA, Sotelo-Mundo R et al (2021) Antiproliferative activity of cardenolides on cell line A549: structure-activity relationship analysis. Mol Divers 25:2289–2305. https://doi.org/10.1007/s11030-020-10119-w

68. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474. https://doi.org/10.1002/jcc.21707

69. Rampogu S, Ravinder D, Pawar SC, Lee KW (2018) Natural compound modulates the cervical cancer microenvironment—a pharmacophore guided molecular modelling approaches. J Clin Med 7. https://doi.org/10.3390/jcm7120551

70. Taha KF, Khalil M, Abubakr MS, Shawky E (2020) Identifying cancer-related molecular targets of *Nandina domestica* Thunb. by network pharmacology-based analysis in combination with chemical profiling and molecular docking studies. J Ethnopharmacol 249:112413. https://doi.org/10.1016/j.jep.2019.112413

71. Sinha S, Patel S, Athar M et al (2019) Structure-based identification of novel sirtuin inhibitors against triple negative breast cancer: an in silico and in vitro study. Int J Biol Macromol 140:454–468. https://doi.org/10.1016/j.ijbiomac.2019.08.061

72. Cavuturu BM, Bhandare VV, Ramaswamy A, Arumugam N (2019) Molecular dynamics of interaction of Sesamin and related compounds with the cancer marker β-catenin: an in silico study. J Biomol Struct Dyn 37:877–891. https://doi.org/10.1080/07391102.2018.1442250

73. Prabhavathi H, Dasegowda KR, Renukananda KH et al (2020) Molecular docking and dynamic simulation to identify potential phytocompound inhibitors for EGFR and HER2 as anti-breast cancer agents. J Biomol Struct Dyn 0:1–12. https://doi.org/10.1080/07391102.2020.1861982

74. Gupta P, Mohammad T, Dahiya R et al (2019) Evaluation of binding and inhibition mechanism of dietary phytochemicals with sphingosine kinase 1: towards targeted anticancer therapy. Sci Rep 9:1–15. https://doi.org/10.1038/s41598-019-55199-3

75. Mohammad T, Siddiqui S, Shamsi A et al (2020) Virtual screening approach to identify high-affinity inhibitors of serum and glucocorticoid-regulated kinase 1 among bioactive natural products: combined molecular docking and simulation studies. Molecules 25:823. https://doi.org/10.3390/molecules25040823

76. Hadavand Mirzaei H, Jassbi AR, Pirhadi S, Firuzi O (2020) Study of the mechanism of action, molecular docking, and dynamics of anticancer terpenoids from *Salvia lachnocalyx*. J Recept Signal Transduct 40:24–33. https://doi.org/10.1080/10799893.2019.1710847

77. Maiti P, Nand M, Joshi T et al (2021) Identification of luteolin-7-glucoside and epicatechin gallate from *Vernonia cinerea*, as novel EGFR L858R kinase inhibitors against lung cancer: Docking and simulation-based study. J Biomol Struct Dyn 39:5048–5057. https://doi.org/10.1080/07391102.2020.1784791

78. Gurung AB, Ali MA, Lee J et al (2021) Molecular docking and dynamics simulation study of bioactive compounds from *Ficus carica* L. With important anticancer drug targets. PLoS ONE 16:1–16. https://doi.org/10.1371/journal.pone.0254035

79. Jha V, Devkar S, Gharat K et al (2022) Screening of phytochemicals as potential inhibitors of breast cancer using structure based multitargeted molecular docking analysis. Phytomedicine Plus 2:100227. https://doi.org/10.1016/j.phyplu.2022.100227

80. Fan J, De Lannoy IAM (2014) Pharmacokinetics. Biochem Pharmacol 87:93–120. https://doi.org/10.1016/j.bcp.2013.09.007

81. Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem 58:4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104

82. Sander T, Freyss J, von Korff M, Rufener C (2015) DataWarrior: an open-source program for chemistry aware data visualization and analysis. J Chem Inf Model 55:460–473. https://doi.org/10.1021/ci500588j

83. Pathania S, Ramakrishnan SM, Bagler G (2015) Phytochemica: a platform to explore phytochemicals of medicinal plants. Database 2015:1–8. https://doi.org/10.1093/database/bav075

84. Pathania S, Ramakrishnan SM, Randhawa V, Bagler G (2015) SerpentinaDB: a database of plant-derived molecules of *Rauvolfia serpentina*. BMC Complement Altern Med 15:1–8. https://doi.org/10.1186/s12906-015-0683-7

85. Hatherley R, Brown DK, Musyoka TM et al (2015) SANCDB: a South African natural compound database. J Cheminform 7:29. https://doi.org/10.1186/s13321-015-0080-8

86. Pilon AC, Valli M, Dametto AC et al (2017) NuBBE(DB): an updated database to uncover chemical and biological information from Brazilian biodiversity. Sci Rep 7:7215. https://doi.org/10.1038/s41598-017-07451-x

87. Tao W, Li B, Gao S et al (2015) CancerHSP: anticancer herbs database of systems pharmacology. Sci Rep 5:11481. https://doi.org/10.1038/srep11481

88. Murugan K, Shanmugasamy S, Al-Sohaibani S et al (2015) TaxKB: a knowledge base for new taxane-related drug discovery. BioData Min 8:19. https://doi.org/10.1186/s13040-015-0053-5

89. Adnan M, Siddiqui AJ, Hamadou WS et al (2021) Phytochemistry, bioactivities, pharmacokinetics and toxicity prediction of *Selaginella repanda* with its anticancer potential against human lung, breast and colorectal carcinoma cell lines. Molecules 26:768. https://doi.org/10.3390/molecules26030768

90. Shang W, Zhang J, Song H et al (2021) Mechanism of tetrandrine against endometrial cancer based on network pharmacology. Drug Des Dev Ther 15:2907–2919. https://doi.org/10.2147/DDDT.S307670

91. Yue Z, Zhang W, Lu Y, Yang Q, Ding Q, Xia J, Chen Y (2015) Prediction of cancer cell sensitivity to natural products based on genomic and chemical properties. PeerJ 3:e1425. https://doi.org/10.7717/peerj.1425

92. Chandran U, Mehendale N, Patil S et al (2017) Network pharmacology. Innov Approaches Drug Discov 127–164. https://doi.org/10.1016/B978-0-12-801814-9.00005-2

93. Kanehisa M, Furumichi M, Sato Y et al (2021) KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49:D545–D551. https://doi.org/10.1093/nar/gkaa970

94. Kuhn M, von Mering C, Campillos M et al (2008) STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 36:D684–D688. https://doi.org/10.1093/nar/gkm795

95. Dennis GJ, Sherman BT, Hosack DA et al (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4:P3

96. Szklarczyk D, Gable AL, Nastou KC et al (2021) The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 49:D605–D612. https://doi.org/10.1093/nar/gkaa1074

97. Zhang N, Wang J, Sheng A et al (2021) emodin inhibits the proliferation of MCF-7 human breast cancer cells through activation of aryl hydrocarbon receptor (AhR). Front Pharmacol 11:1–12. https://doi.org/10.3389/fphar.2020.622046

98. Wang X, Shen Y, Wang S et al (2017) PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. Nucleic Acids Res 45:W356–W360. https://doi.org/10.1093/nar/gkx374

99. Daina A, Michielin O, Zoete V (2019) SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res 47:W357–W364. https://doi.org/10.1093/nar/gkz382

100. Ru J, Li P, Wang J et al (2014) TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Cheminform 6:13. https://doi.org/10.1186/1758-2946-6-13

101. Deng Y, Ye X, Chen Y et al (2021) Chemical characteristics of *Platycodon grandiflorum* and its mechanism in lung cancer treatment. Front Pharmacol 11. https://doi.org/10.3389/fphar.2020.609825

102. Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular

interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303

103. Daina A, Michielin O, Zoete V (2017) SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep 7:42717. https://doi.org/10.1038/srep42717

104. Xiong G, Wu Z, Yi J et al (2021) ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res 49:W5–W14. https://doi.org/10.1093/nar/gkab255

105. Tian W, Chen C, Lei X et al (2018) CASTp 3.0: Computed atlas of surface topography of proteins. Nucleic Acids Res 46:W363–W367. https://doi.org/10.1093/nar/gky473

106. Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) Auto-Dock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. J Chem Inf Model 61:3891–3898. https://doi.org/10.1021/acs.jcim.1c00203

107. Yang J-F, Wang F, Chen Y-Z et al (2020) LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. Brief Bioinform 21:2206–2218. https://doi.org/10.1093/bib/bbz141

108. Aarthy M, Muthuramalingam P, Ramesh M, Singh SK (2022) Unraveling the multi-targeted curative potential of bioactive molecules against cervical cancer through integrated omics and systems pharmacology approach. Sci Rep 12:1–19. https://doi.org/10.1038/s41598-022-18358-7

109. Parkinson H, Kapushesky M, Shojatalab M et al (2007) Array-Express—a public database of microarray experiments and gene expression profiles. Nucleic Acids Res 35:D747–D750. https://doi.org/10.1093/nar/gkl995

110. Zhou G, Soufan O, Ewald J et al (2019) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res 47:W234–W241. https://doi.org/10.1093/nar/gkz240

111. Pomaznoy M, Ha B, Peters B (2018) GOnet: a tool for interactive Gene Ontology analysis. BMC Bioinform 19:470. https://doi.org/10.1186/s12859-018-2533-3

112. Szklarczyk D, Franceschini A, Wyder S et al (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452. https://doi.org/10.1093/nar/gku1003

113. Dong Y, Hao L, Fang K et al (2021) A network pharmacology perspective for deciphering potential mechanisms of action of *Solanum nigrum* L. in bladder cancer. BMC Complement Med Ther 21:1–14. https://doi.org/10.1186/s12906-021-03215-3

114. Wang Y, Zhang Y, Wang Y et al (2021) Using network pharmacology and molecular docking to explore the mechanism of Shan Ci Gu (*Cremastra appendiculata*) against non-small cell lung cancer. Front Chem 9:1–14. https://doi.org/10.3389/fchem.2021.682862

115. Chen Z, Lin T, Liao X et al (2021) Network pharmacology based research into the effect and mechanism of Yinchenhao Decoction against Cholangiocarcinoma. Chin Med (UK) 16:1–20. https://doi.org/10.1186/s13020-021-00423-4

116. Zhang S, Mo Z, Zhang S, Li X (2021) A network pharmacology approach to reveal the underlying mechanisms of *Artemisia annua* on the treatment of hepatocellular carcinoma. Evid Based Complem Altern Med 2021. https://doi.org/10.1155/2021/8947304

117. Zhou L, Gu W, Kui F, et al (2021) The mechanism and candidate compounds of aged citrus peel (*chenpi*) preventing chronic obstructive pulmonary disease and its progression to lung cancer. Food Nutr Res 65. https://doi.org/10.29219/fnr.v65.7526

118. Jiang L, Shi Z, Yang Y (2021) Network pharmacology-based approach to investigate the molecular targets of rhubarb for treating cancer. Evid Based Complem Altern Med 2021. https://doi.org/10.1155/2021/9945633

119. Zhang MM, Wang D, Lu F et al (2021) Identification of the active substances and mechanisms of ginger for the treatment of colon cancer based on network pharmacology and molecular docking. BioData Min 14:1–16. https://doi.org/10.1186/s13040-020-00232-9

120. Fang T, Liu L, Liu W (2020) Network pharmacology-based strategy for predicting therapy targets of *Tripterygium wilfordii* on acute myeloid leukemia. Medicine (Baltimore) 99:e23546. https://doi.org/10.1097/MD.0000000000023546

121. Mi JL, Liu C, Xu M, Wang RS (2020) Network pharmacology to uncover the molecular mechanisms of action of LeiGongTeng for the treatment of nasopharyngeal carcinoma. Med Sci Monit Basic Res 26:923431–923432. https://doi.org/10.12659/MSMBR.923431

122. Liu Y, An T, Wan D et al (2020) Targets and mechanism used by cinnamaldehyde, the main active ingredient in cinnamon, in the treatment of breast cancer. Front Pharmacol 11:1–12. https://doi.org/10.3389/fphar.2020.582719

123. Ma TT, Zhang GL, Dai CF et al (2020) *Scutellaria barbata* and *Hedyotis diffusa* herb pair for breast cancer treatment: potential mechanism based on network pharmacology. J Ethnopharmacol 259:112929. https://doi.org/10.1016/j.jep.2020.112929

124. Yang B, Wang N, Wang S et al (2019) Network-pharmacology-based identification of caveolin-1 as a key target of *Oldenlandia diffusa* to suppress breast cancer metastasis. Biomed Pharmacother 112:108607. https://doi.org/10.1016/j.biopha.2019.108607

125. Shawky E (2019) Prediction of potential cancer-related molecular targets of North African plants constituents using network pharmacology-based analysis. J Ethnopharmacol 238:111826. https://doi.org/10.1016/j.jep.2019.111826

126. Liu X, Wu J, Zhang D et al (2018) Network pharmacology-based approach to investigate the mechanisms of *Hedyotis diffusa* Willd. in the treatment of gastric cancer. Evid Based Complem Altern Med 2018. https://doi.org/10.1155/2018/7802639

127. Zhang X, Shen T, Zhou X et al (2020) Network pharmacology based virtual screening of active constituents of *Prunella vulgaris* L. and the molecular mechanism against breast cancer. Sci Rep 10:1–12. https://doi.org/10.1038/s41598-020-72797-8

128. Lo Y-C, Rensi SE, Torng W, Altman RB (2018) Machine learning in chemoinformatics and drug discovery. Drug Discov Today 23:1538–1546. https://doi.org/10.1016/j.drudis.2018.05.010

129. Vamathevan J, Clark D, Czodrowski P et al (2019) Applications of machine learning in drug discovery and development. Nat Rev Drug Discov 18:463–477. https://doi.org/10.1038/s41573-019-0024-5

130. Zhang L, Tan J, Han D, Zhu H (2017) From machine learning to deep learning: progress in machine intelligence for rational drug discovery. Drug Discov Today 22:1680–1685. https://doi.org/10.1016/j.drudis.2017.08.010

131. Jeon J, Nim S, Teyra J et al (2014) A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Med 6:1–18. https://doi.org/10.1186/s13073-014-0057-7

132. Ferrero E, Dunham I, Sanseau P (2017) In silico prediction of novel therapeutic targets using gene-disease association data. J Transl Med 15:1–16. https://doi.org/10.1186/s12967-017-1285-6

133. Godinez WJ, Hossain I, Lazic SE et al (2017) A multi-scale convolutional neural network for phenotyping high-content cellular images. Bioinformatics 33:2010–2019. https://doi.org/10.1093/bioinformatics/btx069

134. Mamoshina P, Volosnikova M, Ozerov IV et al (2018) Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. Front Genet 9:1–10. https://doi.org/10.3389/fgene.2018.00242

135. Price EJ, Palát J, Coufaliková K et al (2021) Open, high-resolution EI+ spectral library of anthropogenic compounds. Front Public Health 9:622558. https://doi.org/10.3389/fpubh.2021.622558

136. Grigalunas M, Burhop A, Zinken S et al (2021) Natural product fragment combination to performance-diverse pseudo-natural products. Nat Commun 12:1883. https://doi.org/10.1038/s41467-021-22174-4

137. Chen Y, Mathai N, Kirchmair J (2020) Scope of 3D shape-based approaches in predicting the macromolecular targets of structurally complex small molecules including natural products and macrocyclic ligands. J Chem Inf Model 60:2858–2875. https://doi.org/10.1021/acs.jcim.0c00161

138. Vaskevych AI, Savinchuk NO, Vaskevych RI et al (2021) The PIFA-initiated oxidative cyclization of 2-(3-butenyl)quinazolin-4(3H)-ones—an efficient approach to 1-(hydroxymethyl)-2,3-dihydropyrrolo[1,2-a]quinazolin-5(1H)-ones. Beilstein J Org Chem 17:2787–2794. https://doi.org/10.3762/bjoc.17.189

139. Lianza M, Leroy R, Machado Rodrigues C et al (2021) The three pillars of natural product dereplication. Alkaloids from the bulbs of *Urceolina peruviana* (C. Presl) J.F. Macbr. as a preliminary test case. Molecules. https://doi.org/10.3390/molecules26030637

140. Mikhnevich TA, Vyatkina Turkova AV, Grigorenko VG et al (2021) Inhibition of class A β-lactamase (TEM-1) by narrow fractions of humic substances. ACS Omega 6:23873–23883. https://doi.org/10.1021/acsomega.1c02841

141. Zhou Z, Luo M, Zhang H et al (2022) Metabolite annotation from knowns to unknowns through knowledge-guided multi-layer metabolic networking. Nat Commun 13:6656. https://doi.org/10.1038/s41467-022-34537-6

142. Elend L, Jacobsen L, Cofala T et al (2022) Design of SARS-CoV-2 main protease inhibitors using artificial intelligence and molecular dynamic simulations. Molecules 27:4020. https://doi.org/10.3390/molecules27134020

143. Akbarzadeh M, Flegel J, Patil S et al (2022) The pseudo-natural product rhonin targets RHOGDI. Angew Chem Int Ed Engl 61:e202115193. https://doi.org/10.1002/anie.202115193

144. Ceballos J, Schwalfenberg M, Karageorgis G et al (2019) Synthesis of indomorphan pseudo-natural product inhibitors of glucose transporters GLUT-1 and -3. Angew Chem Int Ed Engl 58:17016–17025. https://doi.org/10.1002/anie.201909518

145. Yang W, Soares J, Greninger P et al (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res 41:D955–D961. https://doi.org/10.1093/nar/gks1111

146. Hall M, Frank E, Holmes G et al (2009) The WEKA data mining software. ACM SIGKDD Explor Newsl 11:10–18. https://doi.org/10.1145/1656274.1656278

147. Rayan A, Raiyn J, Falah M (2017) Nature is the best source of anticancer drugs: indexing natural products for their anticancer bioactivity. PLoS ONE 12:1–12. https://doi.org/10.1371/journal.pone.0187925

148. Ertl P, Roggo S, Schuffenhauer A (2008) Natural product-likeness score and its application for prioritization of compound libraries. J Chem Inf Model 48:68–74. https://doi.org/10.1021/ci700286x

149. Newman DJ, Cragg GM, Ahmed J et al (2022) The ChEMBL database in 2017. Nucleic Acids Res 2:1–21. https://doi.org/10.1093/nar/gkq917

150. Chen CY-C (2011) TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS ONE 6:e15939. https://doi.org/10.1371/journal.pone.0015939

151. Morsy N (2019) Anticancer agents from plants. Main Gr Chem 18:169–191. https://doi.org/10.3233/MGC-180689

152. Ntie-Kang F, Mbah JA, Mbaze LM et al (2013) CamMedNP: building the Cameroonian 3D structural natural products database for virtual screening. BMC Complem Altern Med 13:88. https://doi.org/10.1186/1472-6882-13-88

153. Potshangbam AM, Polavarapu R, Rathore RS et al (2019) MedPServer: a database for identification of therapeutic targets and novel leads pertaining to natural products. Chem Biol Drug Des 93:438–446. https://doi.org/10.1111/cbdd.13430

154. Tung C-W, Lin Y-C, Chang H-S et al (2014) TIPdb-3D: the three-dimensional structure database of phytochemicals from Taiwan indigenous plants. Database (Oxford) 2014:bau055. https://doi.org/10.1093/database/bau055

155. Kim S-K, Nam S, Jang H et al (2015) TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine. BMC Complem Altern Med 15:218. https://doi.org/10.1186/s12906-015-0758-5

156. Yabuzaki J (2017) Carotenoids database: structures, chemical fingerprints and distribution among organisms. Database (Oxford) 2017:bax004. https://doi.org/10.1093/database/bax004

157. Miettinen K, Iñigo S, Kreft L et al (2018) The TriForC database: a comprehensive up-to-date resource of plant triterpene biosynthesis. Nucleic Acids Res 46:D586–D594. https://doi.org/10.1093/nar/gkx925

158. Rothwell JA, Urpi-Sarda M, Boto-Ordoñez M, et al (2012) Phenol-Explorer 2.0: a major update of the Phenol-Explorer database integrating data on polyphenol metabolism and pharmacokinetics in humans and experimental animals. Database (Oxford) 2012:bas031. https://doi.org/10.1093/database/bas031

159. Huang J, Wang J (2014) CEMTDD: Chinese ethnic minority traditional drug database. Apoptosis 19:1419–1420. https://doi.org/10.1007/s10495-014-1011-2

160. Zeng X, Zhang P, Wang Y et al (2019) CMAUP: a database of collective molecular activities of useful plants. Nucleic Acids Res 47:D1118–D1127. https://doi.org/10.1093/nar/gky965

161. Xu H-Y, Zhang Y-Q, Liu Z-M et al (2019) ETCM: an encyclopaedia of traditional Chinese medicine. Nucleic Acids Res 47:D976–D982. https://doi.org/10.1093/nar/gky987

162. Li B, Ma C, Zhao X et al (2018) YaTCM: Yet another Traditional Chinese Medicine Database for Drug Discovery. Comput Struct Biotechnol J 16:600–610. https://doi.org/10.1016/j.csbj.2018.11.002

163. Xue R, Fang Z, Zhang M et al (2013) TCMID: traditional Chinese medicine integrative database for herb molecular mechanism analysis. Nucleic Acids Res 41:D1089–D1095. https://doi.org/10.1093/nar/gks1100

164. Boonen J, Bronselaer A, Nielandt J et al (2012) Alkamid database: chemistry, occurrence and functionality of plant N-alkylamides. J Ethnopharmacol 142:563–590. https://doi.org/10.1016/j.jep.2012.05.038

165. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP et al (2018) IMPPAT: a curated database of indian medicinal plants, phytochemistry and therapeutics. Sci Rep 8:1–17. https://doi.org/10.1038/s41598-018-22631-z