Chapter 5
Computational Methods in Natural Products-Based Drug Discovery

Pankaj Dagur, Shreya, Rahul Ghosh, Gaurav Rakshit, Abanish Biswas, and Manik Ghosh

5.1 Introduction

Drug discovery is an extensive, costing big-budget, time-consuming process with the low rate of success. The development of a drug from scratch to market value, maintaining its efficacy, takes around 13–15 years and costs billions of dollars on average and still counting. In comparison to that, the rate of the launching novel drugs in the market is less. It is estimated that about more than half of all the drugs approved in the last three decades were either NPs (Natural products) or their semisynthetic derivatives (Newman and Cragg 2016) (Patridge et al. 2016).

The reason is their diversity in species and utilization for medicinal purposes since ancient civilizations. NPs possess comparatively greater molecular mass and a number of sp^3 carbon atoms, H-bond acceptors and donors, more hydrophilic nature, and molecular rigidness than that of nonnatural compounds' libraries (Atanasov et al. 2015) (Feher and Schmidt 2003). The structural upper hand can be advantageous while tackling protein-protein interaction owing to the greater rigidity of NPs (Lawson et al. 2017). Despite not adhering to Lipinski's rule of five, NPs are still a class that is used for therapeutic purposes, owing to their high molecular mass.

Natural products, despite being an inspirational source for NP-based drug discovery, pose disadvantages for the pipeline. NPs have diverse and complex molecular structures which means a challenge for generating 3D molecular structures and their analogs while considering stereochemistry, force fields, and algorithm for predicting protein-bound conformations (Friedrich et al. 2019). Dereplication tools are required to circumvent the rediscovery of known compounds. Other challenges

P. Dagur (✉) · Shreya · R. Ghosh · G. Rakshit · A. Biswas · M. Ghosh (✉)
Department of Pharmaceutical Sciences and Technology, Birla Institute of Technology, Ranchi, Jharkhand, India
e-mail: manik@bitmesra.ac.in

M. Rudrapal, J. Khan (eds.), *CADD and Informatics in Drug Discovery*,
Interdisciplinary Biotechnological Advances,
https://doi.org/10.1007/978-981-99-1316-9_5

include procuring the materials, extraction, detection, and isolation of bioactive compounds and generating activity profiles are time-consuming and the success rate is less. Considering the facts, the prior prediction of activity using in silico methods can aid in simplifying the process.

The capital needed for in silico experiments is comparatively less than the expenses (for example- scikit-learn, CDK) associated with experimental procedures of which software licensing costs alone, continue to be a significant cost component and have been steadily rising in recent years. Moreover, on site efficient computing center is no longer necessary as calculations can be performed affordably in the cloud at very large scales, with a low degree of complexity. Computational-based drug discovery has well-established techniques equipped with cheminformatics for easing the process, reducing the loss and comparatively less time-consuming. These techniques involve data mining on large data, dereplication, chemical space analysis, visualization and comparison, prediction of bioactivity, ADME and safety profiles' natural products-inspired de novo design, and prediction of natural products prone to cause interference with biological assays (Chen and Kirchmair 2020).

5.2 Natural Products' Collections

The definition of "natural products" is not universally agreed upon, with some authors limiting the term to small molecule secondary metabolites while others broadly accept that chemical substance produced by a living organism as NP. The latter one holds more diversity and hence the line separating the subclasses remains ill-defined. The therapeutic class of NP as per the definition can be classified into phytochemicals, fungal metabolites, toxins, antibodies, and NPs with limited activity. The NPs collection can also be categorized as physical and virtual collections for in silico technology.

5.2.1 Physical Collection

The importance of NPs in ailment curing can be dated back to ancient civilizations. In earlier decades, natural compounds and their structural analogs have significantly added to the therapeutic arsenal for curing numerous diseases, including cancer and infectious disorders. According to a survey, only 6% of the estimated four lakh plant species have undergone activity studies, while less than 20% have undergone phytochemical investigations (Fabricant and Farnsworth 2001). Phytochemicals being antioxidants and a source for many life-saving medicines form a broad class of NPs including polyphenols, terpenoids, and alkaloids. The fungal metabolites have been explored for their use as antidiabetic, antibacterial, antioxidant, antitumor, and even insecticidal agents (Daley et al. 2017). In most cases, chemotherapy medications are made from naturally occurring poisons produced by large clades

of organisms, such as plants, fungi, and bacteria. The next important therapeutic class is antibiotics with more than 60% of drugs approved and more than 500 in the developmental stage as per the survey in 2016 (Cragg et al. 1997). The common mechanism of action includes receptor blocking or downregulation and induction of target cell signaling which can be exploited for rheumatoid arthritis, non-Hodgkin lymphoma, multiple myeloma, and various other diseases (Carter and Lazar 2018). NPs such as biopolymers, spider silk are known for their activity in drug delivery systems rather than therapeutic value.

For virtual screening of NPs for in silico studies, the majority of compound suppliers across the world now freely offer information related to the structures and some other features of the compounds. According to a survey, of the total known NP compounds in virtual databases, only about 10% of them are available for experimental procedures (Chen et al. 2017). This lack of availability of NPs physically serves as a blockage in the path of drug discovery. However, the readily available ones have favorable physicochemical properties for the drug discovery pipeline. Moreover, more than half of them have a molecular weight of less than 300 Da hence, providing many prospects for optimization (Chen et al. 2017). There are more than 100 commercial suppliers of purified NPs in the world, but only few of them supply more than 5000 NPs.

The fact that the (abovementioned) 25 k easily accessible NPs encompass more than 5700 Murcko scaffolds is noteworthy in this respect. Additionally, these NPs including alkaloids, steroids, and flavonoids, provide a fair representation of all of the major NP classes (Chen et al. 2018).

5.2.2 Virtual Collection

The rapidly growing attention of NPs has led to steep growth in NP-based databases. The virtual collection (or databases) of NPs can be categorized into (i) the generalized NP-based databases, (ii) databases of traditional NPs, and (iii) specialized databases (Chen et al. 2019a). The second category includes databases for traditionally used NP-based drugs whereas the third category includes databases focusing on some exclusive organisms belonging to a specific habitats, biological activities, or specific NP classes. A survey reported that since the 2000s, approximately 120 different databases and collections have been released and used in context with NPs (Sorokina and Steinbeck 2020). And of them, approximately 50 are open access, whereas 98 are still in some way accessible. These open -access databases include NP-based database collections published as supplementary material in scientific publications as well as those available in the ZINC database (Sterling and Irwin 2015). The collection of NPs on ZINC database provides information about their structure and their origin but no other additional information. The databases could be open access or commercially available. Amongst free NP databases is Super Natural II, consisting of more than 325 k NPs (Banerjee et al. 2015). A chemistry-aware online interface can be used to query the database, although the bulk download is not

Table 5.1 Examples of some active databases

Databases	Size
TCM database@Taiwan(Chen 2011)	>60 k
Natural Product Atlas (Van Santen et al. 2019)	>25 k
Collective Molecular Activities of Useful Plants (CMAUP) (Zeng et al. 2019)	47 k NPs
Marine Natural Library (Bugni et al. 2008)	14 k

officially supported. Universal Natural Products Database (UNPD) is another free database with more than 200 k NPs and downloadable resources (Gu et al. 2013). Unfortunately, UNPD database appears to be nonfunctional. These virtual databases are either specific to a particular geographical region (like databases only for Chinese herbals), or particular section of NPs (like database for only marine-based NPs), or could be generalized (COCONUT) (Sorokina et al. 2021). Some examples of functional databases are listed below in Table 5.1.

Some other examples include NuBBEDB (Pilon et al. 2017), KnapSack (Nakamura et al. 2013),CMAUP (Zeng et al. 2019), and smaller databases like FooDB. On the contrary, the data available on the therapeutic efficacy and protein-bound conformations of NPs suffer from scarcity. Amongst the most relevant ones, the Marine Natural Library has special mention, as it allows the download of the full dataset of more than 14 k marine NPs (Bugni et al. 2008). NPs seem to have a slight upper hand over synthetic compounds, as their "libraries" already exist in nature. The generalized databases of chemical compounds (Li et al. 2010; Leach 2017) (such as PubChem and ChEMBL) also include databases related to NPs that are annotated by their class, while, more specific ones (such as ArachnoServer, VenomKB, and the Dictionary of Marine Natural Products) provide even more granular annotations for aggregating NP libraries with various characteristics of interest (Dona et al. 2017; Romano et al. 2018).

5.3 Cheminformatics and Computational Approaches for NP-Based Drug Discovery

5.3.1 Computational-Based Approaches

Computer-based approaches being the broader term encloses within cheminformatics technology. Cheminformatics is the application of computational approaches to facilitate collection, storage, analysis of large databases addressing the major concern, drug discovery. Along with cheminformatics, other informatic approaches such as bioinformatics, semantic methods have also been reviewed (Romano and Tatonetti 2019). Computational techniques have long been regarded as an important part of drug development and discovery procedures. The various approaches it offers for drug discovery purpose are structural elucidation, analysis of the physicochemical and structural properties, in determining macromolecular

targets, prediction of ADME properties and safety profiles. Computational methods can be broadly classified into: structure-based and ligand-based for the abovementioned approaches (Podlogar et al. 2001). This classification is revolving around the level of structural information available in context with target to support the computational calculations. Structure-based methods operate on the availability of info regarding three-dimensional (3D) molecular target of interest, typically obtained from X-ray crystallography, nuclear magnetic resonance, or homology modeling (Cerqueira et al. 2015). Whereas ligand-based approaches focus on the availability of information in context with active ligands (and inactive compounds, when available) (Lill 2007). With the increasing need for prior virtual screening of NPs and maintaining of databases, cheminformatics has made its way through drug discovery process. The methods are generally classified as direct and indirect approaches, based on the type of properties they exploit. Direct approaches deal with chemical activity, their constants, reactive groups, ADME profiling, whereas indirect ones deal with structural specifications, compound category or other observations (Romano and Tatonetti 2019).

5.3.2 Cheminformatics and NP-Based Pipeline

So far, cheminformatics and other related informatics approaches have been reviewed in drug discovery pipeline. Cheminformatics and other approaches have played important part in curating NP-based fragmented databases and analyzing the result. Cheminformatics and computational approaches share an important linkage, basically cheminformatics is the application of computational approaches as shown in Fig. 5.1. Cheminformatics techniques exclusive to NP-based drug discovery are NP-based QSAR analysis, Molecular Docking and Dynamics, Computational Mutagenesis, and Library Construction. Numerous classes of NPs have been studied using QSAR, and the chemical descriptors used tend to be dictated by the particular classes (Huang et al. 2016). For example, small-molecule NPs include categorical variables suggesting their specific category of classification, species of origin. Similarly, in case of molecular docking, the specific classes of NPs decide the interaction of target and ligand. For example, if a macromolecular NP (belonging to specific class) is suspected of showing interactions with small-molecule metabolites, docking simulations can be used for mining which metabolites could bind to that NP (Pithayanukul et al. 2009). Other aspects of molecular docking include protein preparation and flexibility, pose scoring in context with binding affinity. The generation of extensive libraries of compounds and its screening aids in prediction of potential drug candidates along with awareness of encountering small fraction of hits (Terrett et al. 1995). In case of NPs, their databases exist in nature way before synthetics. In this chapter, we are going to discuss different analytical methods used in computational approaches for NPs. Antibodies, despite of their large molecular weight, are relatively easy to screen for large numbers via docking, indicating their specificity in structural and binding properties that eventually reduces computational

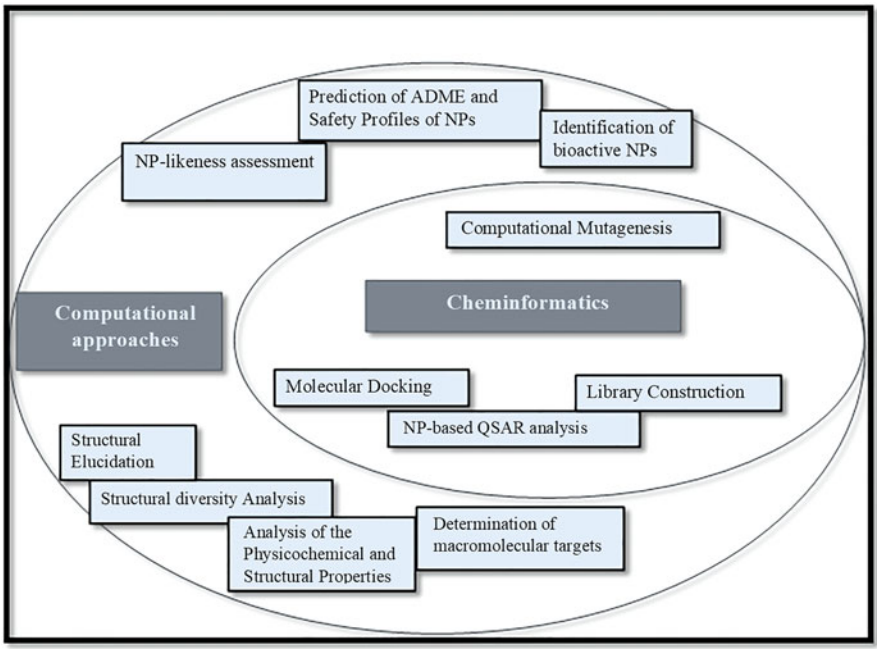

Fig. 5.1 Amalgamation of cheminformatics and computational approaches

complexity for simulations (Mann 2002). Additionally, noteworthy success stories have emerged from screening smaller NP-based databases against specific drug targets. For example, the compound ellagic acid, known to have both antiproliferative and antioxidants properties, was identified by Moro et al. by screening a proprietary database of 2000 NPs against the oncoprotein casein kinase 2 (Cozza et al. 2006).

5.4 Computational Approaches Related to Natural Products

5.4.1 Structural Elucidation

For the extraction and isolation of NPs, the source of materials area is going to be highly priced and long-time taking and when everyone gained knowledge about the NPs, the discovering of novel compounds is decreasing. Of order to make the most of the available experimental resources, it is necessary to integrate analytical and computational approaches for early detection of both favorable and negative features in NPs (Pereira and Aires-de-Sousa 2018). Databases that provide measurable analytical statistics (such as bioactivities, chromatographic data, MS, NMR

spectroscopy, and FTIR data) for known NPs and their interrogation using computational methods play a crucial role in this interaction of technologies. However, even the biggest of these databases only include a small subset of the NPs that are considered. This is why NMR and MS fragmentation predictions are increasingly being made using computational methods, often in tandem with structure generators (Pereira and Aires-de-Sousa 2018).

In recent days, for the virtual screening of natural product (NP) candidates in both small datasets of isolated chemicals and huge databases, structure-based (SB) and ligand-based (LB) cheminformatics techniques have become indispensable tools. Quantitative structure-activity relationships (QSAR), assessment of drug similarity, forecasting surface assimilation, distribution, metabolism, excretion prediction, similarity detection, and pharmacophore identification are the most often used LB approaches. Similar techniques used in SB methods include molecular dynamics, docking, and binding cavity analysis (Pereira and Aires-de-Sousa 2018).

The potential of re-isolating well-known molecules has recently, however, put a hold on the drug development process from natural products. The process of dereplication, which automates the quick identification of previously isolated compounds, directs researchers to fresh discoveries and cuts down on the time and effort needed to develop innovative medication leads. Dereplication uses processed experimental data to identify compounds by comparing it to data from known compounds, hence it requires a variety of computing tools and resources to process and analyze compound data. The combination of analytical data analysis and multivariate data analysis is a key technique for computer-assisted dereplication (Chanana et al. 2017). Dimensionality reduction methods like principal component analysis (PCA), cluster analysis, and/or discrimination assessment may be used to isolate interesting NPs from complicated mixtures, such as NPs in extracts that are specific to a certain organism of interest (Chanana et al. 2017; Abdelmohsen et al. 2014).

By analyzing spectroscopic data, computer-assisted shape elucidation (CASE) systems aim to identify the optimal shape for an active molecule. Structures that are in agreement with experimental (spectroscopic) data are listed and ranked by CASE systems for greater precision. CASE structures ideally operate at low mistake rates and in a fully computerized form. The assignment of stereochemical attributes to NP structures can be done using sophisticated CASE systems because they also take stereospecific NMR data and/or calculations based on DFT (density functional theory) into account (Burns et al. 2019).

NP dereplication is a topic that machine learning techniques find very appealing. Using 13C NMR spectroscopic data, for instance, a recent study once investigated the possibility of machine learning algorithms to assign NPs to eight NP classes (such as chromans) (Martínez-Treviño et al. 2020). It is used to take an XGBoost classifier to achieve the remarkable overall performance. More than 80% of a test set's compounds were correctly assigned for the majority of NP classes. For the quick identification of novel NPs from a filamentous marine cyanobacterium, another discovery successfully applied a convolutional neural network-based method (Reher et al. 2020).

One of the most up-to-date resources for managing MS/MS spectra and sharing the results of such analyses is the Global Natural Product Social Molecular Networking (GNPS). It enables researchers to investigate a dataset and compare its results to anything else that is publicly available. Online dereplication is made possible by GNPS's usage of automated molecular networking analysis (Wang et al. 2016).

5.4.2 Analysis of Physicochemical and Structural Properties

By utilizing the physicochemical and structural characteristics of NPs, NPs have been characterized in a significant way by cheminformatics. The chemical space that NPs occupy is substantially larger than that of synthetic compounds, and they also occupy regions of the chemical space that are often inaccessible to synthetic molecules (Ertl and Schuffenhauer 2008) (Singh and Culberson 2009).

Compared to synthetic pharmaceuticals and synthetic, drug-like substances, NPs are generally heavier and more hydrophobic (Chen et al. 2019b). In addition, their structural complexity is usually higher, particularly when it comes to stereochemistry (often measured by the number of chiral centers and the number of bridgehead atoms in ring systems) and three-dimensional molecular form (Henkel et al. 1999) (Lucas et al. 2015).

The vast variety of ring systems displayed by NPs, particularly in aliphatic systems, is astounding (Ertl and Schuffenhauer 2008) (Grabowski and Schneider 2007). Researchers found that commercially available screening databases lacked core ring scaffolds for 83% of NPs. The two characteristics of NPs that set them apart from synthetic compounds in terms of atom composition are their low variety of nitrogen atoms and their large number of oxygen atoms (Feher and Schmidt 2003; Wetzel et al. 2007; López-Vallejo et al. 2012). However, the vast majority of known NPs and, even more so, those found in actual NP libraries have pharmacological properties (Chen et al. 2018).

Physicochemical and structural characteristics vary across NPs from various kingdoms. For instance, marine species are more likely to have macrocycle-containing NPs or lengthy aliphatic chains than terrestrial species (El-Elimat et al. 2012) (Muigg et al. 2013) (Saldivar-Gonzalez et al. 2018). Their NPs are distinguished by an excessive number of heteroatoms and, in conjunction with this, a wide range of functional groups (Pilkington 2019) (Shang et al. 2018) (Ertl and Schuhmann 2020) (Ertl and Schuhmann 2019).

5.4.3 Structural Diversity Analysis

In terms of structural variety, NPs are incomparable, and this is something that is also evident at the fragment level (Tran et al. 2020). Using the concept of molecular

scaffolds, some research compares natural products (NPs) to synthetic ones in order to evaluate the structural diversity of NPs (Bemis and Murcko 1996). Recent research contrasts the scaffolds that are unique to natural products (NPs) with those of synthetic chemicals and presents an intuitive depiction of them (Ertl and Schuhmann 2020). This then allows us to compare the scaffolds often seen in NPs derived from bacteria, plants, fungi, or mammals (Chen et al. 2018).

Scaffold Hunter is a potent java-based application for the intuitive, visual study of the structural variety of a set of chemicals (Schäfer et al. 2017; Lachance et al. 2012). The concept of molecular scaffolds being represented and categorized hierarchically forms the foundation of Scaffold Hunter. An early version of this tool was used to develop the structural categorization of NPs (SCONP), a technique for mapping the chemical space of NPs (Koch et al. 2005).

Principal component analysis (PCA) is a common technique for mapping the chemical space since it transforms high-dimensional data into a low-dimensional space with little loss of information. The most useful result of principal components analysis (PCA) is the PCA scatter plot, which shows how the data points are distributed in a low-dimensional space (Saldívar-González et al. 2019; Shen et al. 2012).

A method called ChemGPS was created and updated for usage with NPs under the name ChemGPS-NP in order to prevent the need for the principal components to be recalculated as new compounds are added to the datasets. For mapping the chemical space of tiny compounds, predicting modes of action, and analyzing structure-activity connections, ChemGPS-NP has been employed in a number of research (Frédérick et al. 2012; Korinek et al. 2017; Muigg et al. 2013).

The recently developed UMAP for Dimension Reduction method and t-SNE are two more reliable methods for dimensionality reduction. When various items are modeled by distant points and the same objects are generally close together, t-SNE creates plots. Although UMAP is quicker, it delivers results conceptually comparable to those of t-SNE (Van der Maaten and Hinton 2008) (Burton 2020).

Researchers recently developed Statistical-Based Database Fingerprint (SB-DFP), which is a new technique for representing the chemical space of compound databases by a single fingerprint. In theory, any chemical fingerprint and any reference set might be used to derive the SB-DFP, which has a wide range of applicability. By contrasting the binomial distributions of the preferred molecular fingerprint features among the compounds in an interest dataset with those in a reference dataset, the SB-DFP is created (Sánchez-Cruz and Medina-Franco 2018).

5.4.4 Natural Product-Likeness Assessment

The NP-likeness of compounds can be quantified using computational techniques, which can also distinguish NPs and NP-like substances from manufactured compounds with high accuracy. As a result, they are often used in the development of new compounds, the construction of libraries, the selection of NPs (and NP

derivatives and analogs) from collections of mixed compounds, and the prioritizing of compounds (Chen et al. 2022) (Yu 2011).

The NP-Likeness Score is one of the most well-known strategies (Ertl and Schuffenhauer 2008). This score assesses the NP-likeness of compounds using Bayesian statistics, mostly based on how similar their fragments are to those of recognized NPs. With certain changes, the NP-Likeness Score has been modified in several programs and platforms (Jayaseelan and Steinbeck 2014; Vanii Jayaseelan et al. 2012; Sorokina and Steinbeck 2019). Additionally, a rule-based strategy and a theoretically related method using extended connectivity fingerprints (ECFPs) are other options (Zaid et al. 2010). A more recent method for locating NPs and NP-like substances in vast sets of molecules is called NP-Scout (Chen et al. 2019b).

In order to properly characterize the structural properties of NPs, a novel method known as the Natural Compound Molecular Fingerprint (NC-MFP) has been developed (Seo et al. 2020).

5.4.5 Identification of Bioactive Natural Products

With regard to identifying bioactive NPs, computational approaches have demonstrated their effectiveness. For NP research, the full spectrum of virtual screening methods has been used, from straightforward, quick methods based solely on 2D molecular fingerprint similarity to more sophisticated, 3D methods largely based on similarity in molecular structure, pharmacophore models, molecular interaction fields, or docking. Machine learning techniques have recently become a cornerstone in virtual screening for bioactive NPs (Kirchweger and Rollinger 2018).

The sparseness of the structural information that is now available will make it extremely difficult to attach NPs to the structures of macromolecules. This is due to the fact that docking algorithms and scoring criteria are particularly sensitive to even very small changes in 3D form, as those frequently brought on by ligand binding (including solvent effects). The careful employment of homology modeling techniques, induced fit docking methods, and molecular dynamics simulations, however, can also aid to overcome this challenge. Docking toward a variety of representative protein structures may be an effective strategy when dealing with highly adaptable proteins (for binding mode prediction as well as virtual screening) (Amaro et al. 2018; Grienke et al. 2010).

In terms of binding mode prediction, docking algorithms frequently produce accurate results as opposed to virtual screening. It is possible to generate a sufficiently accurate binding pose that offers crucial insights for the development of optimization techniques if the target NP is no longer excessively large or flexible, the ligand binding site is well-defined (i.e., not too shallow, not solvent-exposed), and the interaction between the binding companions consists of two or more directed interactions (Chen and Kirchmair 2020). Binding posture prediction is more practical than virtual screening because it completely ignores the most difficult part of docking—scoring compounds according to their binding affinity—and permits

researchers to focus their efforts on a single ligand-target combination. Importantly, docking makes it possible to clarify the stereoselectivity of ligand binding, especially in the context of NP research (and different processes, such as metabolism). It is impossible to exaggerate how important it is to employ the proper stereochemical data when using 3D techniques, particularly docking (Warren et al. 2006).

5.4.6 Determination of Macromolecular Targets

When one, few, or even many compounds are tested against the broadest range of macromolecules, it may be said that *in silico* target prediction is a large-scale use of virtual screening (Grisoni et al. 2019). Numerous techniques including models have been described in recent years, and they are now recognized as crucial resources in the early stages of drug development. The majority of target prediction algorithms are ligand-based due to the difficulties associated with docking and structure-based approaches in general (specifically, the restricted representation of macromolecules through the available structural data) (Cereto-Massagué et al. 2015; Ezzat et al. 2019; Sam and Athri 2019; Chaudhari et al. 2017).

Ligand-based approaches span the whole spectrum, from simple similarity-based methods to sophisticated machine learning and network-based methods. Unexpectedly, despite the wide variety of computer approaches available today for target prediction, we still have a limited understanding of the importance of these tactics in practical situations. This is especially true given the (generally) expensive expenses associated in experimentally evaluating such models in a systematic, prospective manner. However, it is also true given the common use of partially inadequate, cursory retrospective validation techniques (Mathai et al. 2020; Mathai and Kirchmair 2020). To the best of our knowledge, the Similarity Ensemble Approach (SEA) is the only computational strategy for which consistent experimental validation has been documented (Keiser et al. 2007) (Keiser et al. 2009)(Lounkine et al. 2012).

In recent research comparing the effectiveness and scope of a similarity-based strategy and a machine learning technique toward determining the targets of small molecules, it was discovered that the structural similarity between both the compounds of interest and the compounds reflected in the training set is a key factor in both methods' predictability (or knowledge base). Given that target prediction models are essentially created for and trained on experimental measurements for synthetic chemicals, it is important to take this fact into consideration while working with NPs (Mathai et al. 2020).

Surprisingly, in the same research, the similarity-based technique beat the machine learning strategy for the data at hand. The results imply that the basic similarity-based strategy is a realistic choice, in particular when taking into consideration model interpretability. However, a direct comparison in between two approaches should be approached with extreme caution for a number of reasons.

Additionally, this is demonstrated by the successful operation of several well-known, similarity-based approaches like SwissTargetPrediction (Gfeller et al. 2014).

In addition to 3D similarity-based methods, 3D pharmacophore-based approaches are extensively utilized in the field of NP research for target prediction. A profiling investigation, for example, evaluated secondary metabolites extracted from the medicinal plant Ruta graveolens against a battery of over 2000 pharmacophore models covering over 280 targets (Rollinger et al. 2009). Arborinine was discovered to be an inhibitor of acetylcholinesterase (estimated IC50 = 35 M) as a result of this in silico search, among other potential bioactive NPs and interactions.

Machine learning-based methods have undoubtedly sparked the most interest in NP target prediction in recent years. SPiDER, TIGER, and STarFish are a few notable examples (Reker et al. 2014b) (Schneider and Schneider 2017a) (Cockroft et al. 2019).

With the use of "fuzzy" molecular descriptors, SPiDER employs self-organizing maps in an acronym that enables NPs to utilize it (Rodrigues et al. 2016b; Merk et al. 2018). The mannequin helped identify the targets of the macrolide PPAR, archazolid A (Reker et al. 2014a), including 5-lipoxygenase, FXR, glucocorticoid receptor, as well as, prostaglandin E2 synthase 1. It also successfully predicted the target of the 16-membered depsipeptide doliculide, which is prostanoid receptor 3 (Schneider et al. 2016). Numerous fragment-like NPs were also successfully recognized by SPIDER, including (i) sparteine, whose targets include the nicotinic receptors, muscarinic, p38 mitogen-activated protein kinase, and kappa opioid receptor (Rodrigues et al. 2016a), (ii) DL-goitrin, whose targets include the muscarinic M1 receptor and the pregnane X receptor, (iii) Isomacroin, whose targets were experimentally verified to be the adenosine A3 receptor and the platelet-derived growth factor receptor, and (iv) graveolinine, whose objectives were scientifically proven to be cyclooxygenase-2 and the serotonin 5-HT2B receptor (Rodrigues et al. 2015).

SPiDER and TIGER have a similar conceptual framework. The projected targets are scored using a new methodology and updated CATS descriptions (taking into account ensemble similarity). The marine NP (+−)-marinopyrrole A (Schneider and Schneider 2017a) has been effectively discovered by TIGER as a target of cholecystokinin receptor, the orexin receptor, and glucocorticoid receptor. The model correctly identified the estrogen receptors and as targets of the stilbenoid resveratrol, among other proteins (Schneider and Schneider 2017b).

A stacked ensemble target prediction approach called STarFish was developed using synthetic chemical data (Cockroft et al. 2019).

Most recently, medical indication information was used to train multitask deep neural networks and use them to identify privileged chemical scaffolds in NPs (in this instance, scaffolds are used for which many NPs built within the same scaffold are active inside the same indication). A privileged scaffold dataset was created for 100 indications based on the predictions of these models, which may be used as the starting point for NP-based drug development (Lai et al. 2020).

5.4.7 *Prediction of ADME and Safety Profiles of NPs*

ADME and safety profiling has a major say in drug discovery. ADME failures contribute to around 40% of all the drug failures (Bhhatarai et al. 2019). So far, the *in silico* ADME techniques have seen significant progress as shown in Table 5.2. Drug toxicity is still a major concern despite the fact that pharmacokinetics (PK) failures have decreased as a result of preclinical ADME investigations. These failures at late phases of drug discovery pipeline causes huge loss of time and capital. The *in silico* models provide a prior prediction for optimization. Another concern is drug–drug interactions (DDI) which can result in toxicity and severe ADR, obscuring the whole process. Established and broadly applicable computational filters will serve the best for screening and synthesizing and optimizing the drug product (Ekins et al. 2000). In the 1960s, the early phase of ADME models was developed using Hansch's conventional QSAR methods. As a result, comparative molecular field analysis (CoMFA), a type of molecular modeling software, was developed, in such a way that three-dimensional visualization became an important direction for QSAR.

The different ADME properties that can be evaluated by computational approach are solubility, permeability, clearance, metabolic stability, drug–drug interactions, blood–brain barrier, and cardiotoxicity.

The different software available for predicting ADME properties are MolCode toolbox, preADMET, MolCode toolbox, Discovery Studio,volsurfC, QikProp, ADMEWORKS Predictor C Chembench, and admetSAR (Shin et al. 2017).

The major challenges addressed by NPs related to ADME profiling are off-target receptors such as—hERG channel, cytochrome P450 enzymes (suspected for drug-drug interactions, and toxicity), and the P-glycoprotein (suspected for drug resistance). A plethora of such models based on statistical, machine learning, pharmacophore address these and many other off-targets. Another major concern is most of the computational models are validated by synthetic origin drug product. Computational models such as FAME 3 have reportedly known to for their effectiveness even when majority of compounds in the training set are again of synthetic origin (Šícho et al. 2019).

Table 5.2 Progress in *in silico* ADME (Bhhatarai et al. 2019)

Phase	Progress
1960s	Classical QSAR methods with small datasets developed by Hansch (1972), introduction of use of octanol ±water log *P*
1980s	CoMFA was developed along with other membrane permeability and intestinal absorption models—CYP 3D-QSAR and 4D-QSAR modeling
2010s	More than 100,000 data for in vitro ADME properties in big pharma, open access data in thousands, growth of open projects (for example, eTOX, OpenTox, Tox21, ToxCast). wide variety of ML algorithms (RF, SVM, KNN, NB, DNN)

5.4.8 Case Study

Scientists have shown that five tropical plants—M. charantia, B. javanica, E. longifolia, T. divaricata, and G. mangostana—exhibit inhibitory effect against H5N1 neuraminidase. For the purposes of bioassays, different plant parts (leaves, roots, and fruits) were extracted, chromatographed, and fractionated. The anti-H5N1 neuraminidase activity of the plant fractions and extracts ranged from excellent to moderate. At 250 g/ml, G. mangostana showed the maximum inhibition (82.95 percent). Following this, pure chemicals were extracted from the five plants. The IC50 values of rubraxanthone, mangostin, and garcinone C ranged from 89.71 to 95.49 M, making them stand out (Ikram et al. 2015). This process is depicted below (Fig. 5.2) and the docking results of the abovementioned plant derivative are mentioned in Fig. 5.3.

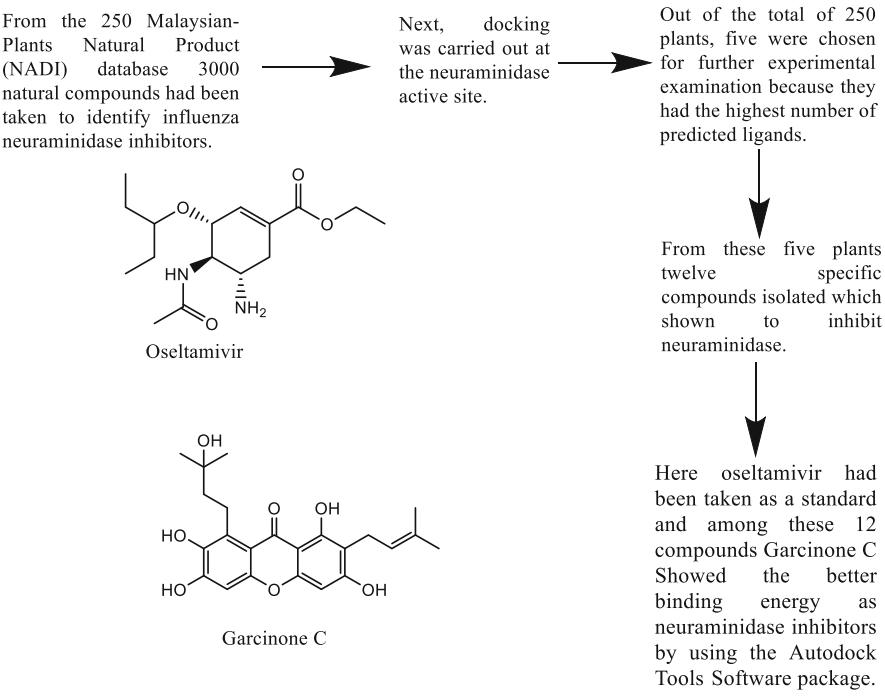

Fig. 5.2 Strategies for novel neuraminidase inhibitors discovery of natural product (Ikram et al. 2015)

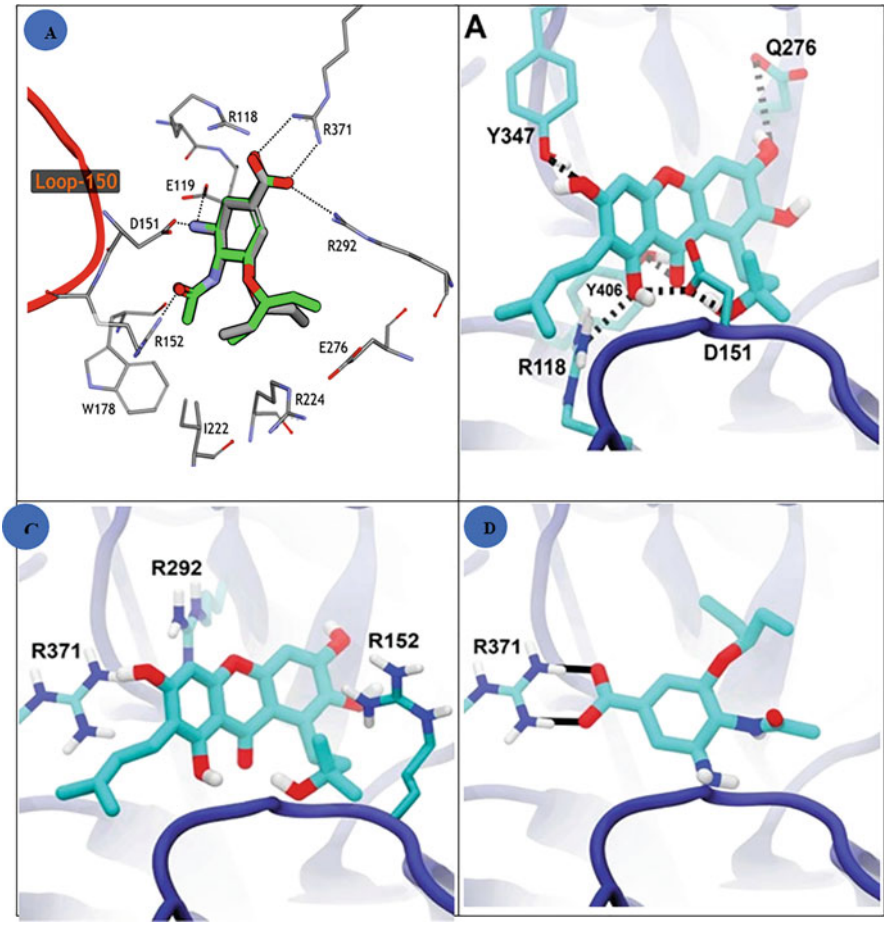

Fig. 5.3 (**a**) The superimposition of the docked and crystallographic oseltamivir poses (green and blue, respectively). The RMSD was 0.84 Å. (**b**) Predicted hydrogen bonds of Garcinone C in the active site of neuraminidase inhibitors. (**c**) Predicted cation-π interactions between R371, R292, R152, and the xanthone moiety of Garcinone C in the active site of neuraminidase inhibitors. (**d**) The crystallographic pose of oseltamivir, a potent inhibitor, shown for reference (PDB ID: 2HU4) (Ikram et al. 2015)

5.5 Challenges to Computational Approaches

The major challenges for NP-based drug discovery is management and representation of the data. Although ArachnoServer and ConoServer are rich and highly descriptive NP databases, but reserved only to specific clade of species producing toxins (Kaas et al. 2012). A partial solution for this is Tox-Prot manual annotation program within UniProtKB/Swiss-Prot which provides a more generalized and improved representation of databases for NPs (Jungo et al. 2012). However, this

Table 5.3 List of databases discontinued in 2019

Database	Type of NPs	Size
3DMET (Maeda and Kondo 2013)	General	18,248
AfroDB (Ntie-Kang et al. 2013b)	tm, plants, Africa	954
CamMedNP (Ntie-Kang et al. 2013a)	tm, plants, Africa	>2500
Traditional Chinese Medicine Systems Pharmacology (TCMSP) (Ru et al. 2014)	Chinese herbs	499

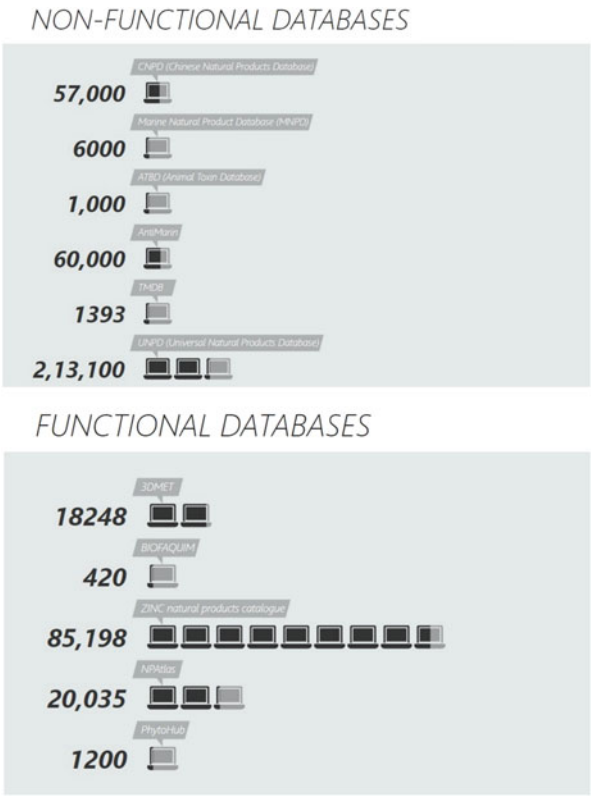

Fig. 5.4 Pictorial representation of functional and defunct databases

does not seem to be the complete solution. Another concern associated with NPs is fragmentation of databases which means more scattered form of data to be maintained by smaller or larger organizations. The added difficulty is shortage of funding required for maintaining those databases which leads to mismanagement of data, ultimately disabling the function of that database. Examples of such databases include as follows (Table 5.3):

To have a clear view, a comparative data of functional and defunct databases have been depicted in Fig. 5.4. A fundamental obstacle to the experimental screening of

NPs is their propensity to interact with biological tests. This could be explained with the example of quercetin which has reportedly shown active in more than about 800 unique bioassays. The most common mechanism followed for interference is aggregate formation, covalent binding, membrane disruption, metal chelation, interference with assay spectroscopy, and buffer decomposition buffers (Baell and Holloway 2010). These problems could be overcome by specific set of rules following statistical approach known as pan-assay interference compounds (PAINS) rule set (Baell and Nissink 2018).

5.6 Conclusion and Future Perspectives

Between the 1980s and the 2010s, two-thirds of the medications were either featured NP pharmacophores (35%) or were analogs of NPs (5%). Modern computational techniques discussed above can significantly expedite and reduce the risk of NP-based drug development. The integration of computational approaches with cheminformatics and other informatics methods has led to ease the management, storage, and representation of vast NP-based databases. Computational tools offer assistance in structural elucidation of bioactive NPs, in prior prediction of various properties of NPs as discussed above which eases the procedure for drug discovery pipeline. However, the major challenge being availability of descriptive database, fragmented databases, and its maintenance along with physical availability of the particular NP. These challenges have been resolved partially with introduction of databases like COlleCtion of Open Natural prodUcTs (COCONUT) which provides a web interface to browse and download elucidated and predicted NPs collected from open sources. On a larger parameter, machine learning (ML) has been using computational methods in drug discovery. For instance, clustering techniques have enabled de novo molecular design, projected protein target druggability, and segmented cell type imaging. The computational approach for NP-based drug discovery holds great future for NP-based drug discovery. The amalgamation of computational methods with advanced technologies in analytical domains can improvise the drug discovery pipeline for NPs. The advancement of higher-field NMR instruments and probe technology has made it possible to determine the structure of NPs from extremely small amounts hence, less wastage of hardly obtained product. Pauli and associates suggested conducting early, relatively sophisticated purity analyses on lead nanoparticles using quantitative NMR and LC–MS to avoid pointless downstream initiatives. Further advancement of metabolomics, genome mining, microbial culturing technique has added to the future scope of NP-based drug pipeline. In addition, antivirulence strategies may represent an alternative method for combating infections, for which NPs that target bacterial quorum sensing may be of interest. *In silico* Medicine, an American company, created an AI system called GENTRL (Generative Tensorial Reinforcement Learning) in 2019 that, in just 46 days, successfully created six kinase inhibitors of the discoidin domain receptor 1 linked to lung fibrosis. Cheminformatics, bioinformatics, and other related fields

have made significant contributions to NP-based drug discovery over the years. Recently, reviews of their successful applications and limitations were conducted.

References

Abdelmohsen UR, Cheng C, Viegelmann C et al (2014) Dereplication strategies for targeted isolation of new antitrypanosomal actinosporins A and B from a marine sponge associated-Actinokineospora sp. EG49. Mar Drugs 12:1220–1244

Amaro RE, Baudry J, Chodera J et al (2018) Ensemble docking in drug discovery. Biophys J 114: 2271–2278

Atanasov AG, Waltenberger B, Pferschy-Wenzig EM et al (2015) Discovery and resupply of pharmacologically active plant-derived natural products: a review. Biotechnol Adv 33:1582–1614. https://doi.org/10.1016/j.biotechadv.2015.08.001

Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53:2719–2740

Baell JB, Nissink JWM (2018) Seven-year itch: pan-assay interference compounds (PAINS) in 2017—utility and limitations. ACS Chem Biol 13:36–44

Banerjee P, Erehman J, Gohlke B-O et al (2015) Super Natural II—a database of natural products. Nucleic Acids Res 43:D935–D939

Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–2893

Bhhatarai B, Walters WP, Hop CECA et al (2019) Opportunities and challenges using artificial intelligence in ADME/Tox. Nat Mater 18:418–422

Bugni TS, Richards B, Bhoite L et al (2008) Marine natural product libraries for high-throughput screening and rapid drug discovery. J Nat Prod 71:1095–1098. https://doi.org/10.1021/np800184g

Burns DC, Mazzola EP, Reynolds WF (2019) The role of computer-assisted structure elucidation (CASE) programs in the structure elucidation of complex natural products. Nat Prod Rep 36: 919–933

Burton R (2020) Unsupervised learning techniques for malware characterization: understanding certain DNS-based DDoS attacks. Digit Threat Res Pract 1:1–26

Carter PJ, Lazar GA (2018) Next generation antibody drugs: pursuit of the 'high-hanging fruit'. Nat Rev Drug Discov 17:197–223

Cereto-Massagué A, Ojeda MJ, Valls C et al (2015) Tools for in silico target fishing. Methods 71: 98–103

Cerqueira NM, Gesto D, Oliveira EF et al (2015) Receptor-based virtual screening protocol for drug discovery. Arch Biochem Biophys 582:56–67

Chanana S, Thomas CS, Braun DR et al (2017) Natural product discovery using planes of principal component analysis in R (PoPCAR). Meta 7:34

Chaudhari R, Tan Z, Huang B, Zhang S (2017) Computational polypharmacology: a new paradigm for drug discovery. Expert Opin Drug Discov 12:279–291. https://doi.org/10.1080/17460441.2017.1280024

Chen CY-C (2011) TCM database@ Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. PLoS One 6:e15939

Chen Y, de Bruyn KC, Kirchmair J (2017) Data resources for the computer-guided discovery of bioactive natural products. J Chem Inf Model 57:2099–2111

Chen Y, de Bruyn Kops C, Kirchmair J (2019a) Resources for chemical, biological, and structural data on natural products. Prog Chem Org Nat Prod 110:37–71

Chen Y, Garcia de Lomana M, Friedrich N-O, Kirchmair J (2018) Characterization of the chemical space of known and readily obtainable natural products. J Chem Inf Model 58:1518–1532
Chen Y, Kirchmair J (2020) Cheminformatics in natural product-based drug discovery. Mol Inform 39:e2000171
Chen Y, Rosenkranz C, Hirte S, Kirchmair J (2022) Ring systems in natural products: structural diversity, physicochemical properties, and coverage by synthetic compounds. Nat Prod Rep 39: 1544–1556
Chen Y, Stork C, Hirte S, Kirchmair J (2019b) NP-Scout: machine learning approach for the quantification and visualization of the natural product-likeness of small molecules. Biomol Ther 9:43
Cockroft NT, Cheng X, Fuchs JR (2019) STarFish: a stacked ensemble target fishing approach and its application to natural products. J Chem Inf Model 59:4906–4920
Cozza G, Bonvini P, Zorzi E et al (2006) Identification of ellagic acid as potent inhibitor of protein kinase CK2: a successful example of a virtual screening application. J Med Chem 49:2363–2366
Cragg GM, Newman DJ, Snader KM (1997) Natural products in drug discovery and development. J Nat Prod 60:52–60
Daley DK, Brown KJ, Badal S (2017) Fungal metabolites. In: Pharmacognosy: fundamentals, applications and strategy. Elsevier, London
Dona MSI, Prendergast LA, Mathivanan S et al (2017) Powerful differential expression analysis incorporating network topology for next-generation sequencing data. Bioinformatics 33:1505–1513
Ekins S, Waller CL, Swaan PW et al (2000) Progress in predicting human ADME parameters in silico. J Pharmacol Toxicol Methods 44:251–272
El-Elimat T, Zhang X, Jarjoura D et al (2012) Chemical diversity of metabolites from fungi, cyanobacteria, and plants relative to FDA-approved anticancer agents. ACS Med Chem Lett 3:645–649
Ertl P, Schuffenhauer A (2008) Cheminformatics analysis of natural products: lessons from nature inspiring the design of new drugs. Prog drug Res 66(217):219–235. https://doi.org/10.1007/978-3-7643-8595-8_4
Ertl P, Schuhmann T (2019) A systematic cheminformatics analysis of functional groups occurring in natural products. J Nat Prod 82:1258–1263
Ertl P, Schuhmann T (2020) Cheminformatics analysis of natural product scaffolds: comparison of scaffolds produced by animals, plants, fungi and bacteria. Mol Inform 39:2000017
Ezzat A, Wu M, Li X-L, Kwoh C-K (2019) Computational prediction of drug–target interactions using chemogenomic approaches: an empirical survey. Brief Bioinform 20:1337–1357
Fabricant DS, Farnsworth NR (2001) The value of plants used in traditional medicine for drug discovery. Environ Health Perspect 109:69–75. https://doi.org/10.1289/ehp.01109s169
Feher M, Schmidt JM (2003) Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. J Chem Inf Comput Sci 43:218–227
Frédérick R, Bruyère C, Vancraeynest C et al (2012) Novel trisubstituted harmine derivatives with original in vitro anticancer activity. J Med Chem 55:6489–6501
Friedrich N-O, Flachsenberg F, Meyder A et al (2019) Conformator: a novel method for the generation of conformer ensembles. J Chem Inf Model 59:731–742
Gfeller D, Grosdidier A, Wirth M et al (2014) SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res 42:W32–W38. https://doi.org/10.1093/nar/gku293
Grabowski K, Schneider G (2007) Properties and architecture of drugs and natural products revisited. Curr Chem Biol 1:115–127
Grienke U, Schmidtke M, Kirchmair J et al (2010) Antiviral potential and molecular insight into neuraminidase inhibiting diarylheptanoids from Alpinia katsumadai. J Med Chem 53:778–786
Grisoni F, Merk D, Friedrich L, Schneider G (2019) Design of natural-product-inspired multitarget ligands by machine learning. ChemMedChem 14:1129–1134

Gu J, Gui Y, Chen L et al (2013) Use of natural products as chemical library for drug discovery and network pharmacology. PLoS One 8:e62839
Henkel T, Brunne RM, Müller H, Reichel F (1999) Statistical investigation into the structural complementarity of natural products and synthetic compounds. Angew Chem Int Ed 38:643–647
Huang P-S, Boyken SE, Baker D (2016) The coming of age of de novo protein design. Nature 537:320–327
Ikram NKK, Durrant JD, Muchtaridi M et al (2015) A virtual screening approach for identifying plants with anti-H5N1 neuraminidase activity. J Chem Inf Model 55:308–316
Jayaseelan KV, Steinbeck C (2014) Building blocks for automated elucidation of metabolites: natural product-likeness for candidate ranking. BMC Bioinform 15:1–9
Jungo F, Bougueleret L, Xenarios I, Poux S (2012) The UniProtKB/Swiss-Prot Tox-Prot program: a central hub of integrated venom protein data. Toxicon 60:551–557
Kaas Q, Yu R, Jin A-H et al (2012) ConoServer: updated content, knowledge, and discovery tools in the conopeptide database. Nucleic Acids Res 40:D325–D330
Keiser MJ, Roth BL, Armbruster BN et al (2007) Relating protein pharmacology by ligand chemistry. Nat Biotechnol 25:197–206. https://doi.org/10.1038/nbt1284
Keiser MJ, Setola V, Irwin JJ et al (2009) Predicting new molecular targets for known drugs. Nature 462:175–181
Kirchweger B, Rollinger JM (2018) Virtual screening for the discovery of active principles from natural products. In: Natural products as source of molecules with therapeutic potential. Springer, Cham, pp 333–364
Koch MA, Schuffenhauer A, Scheck M et al (2005) Charting biologically relevant chemical space: a structural classification of natural products (SCONP). Proc Natl Acad Sci 102:17272–17277
Korinek M, Tsai Y-H, El-Shazly M et al (2017) Anti-allergic hydroxy fatty acids from Typhonium blumei explored through ChemGPS-NP. Front Pharmacol 8:356
Lachance H, Wetzel S, Kumar K, Waldmann H (2012) Charting, navigating, and populating natural product chemical space for drug discovery. J Med Chem 55:5989–6001
Lai J, Hu J, Wang Y et al (2020) Privileged scaffold analysis of natural products with deep learning-based indication prediction model. Mol Inform 39:2000057. https://doi.org/10.1002/minf.202000057
Lawson ADG, MacCoss M, Heer JP (2017) Importance of rigidity in designing small molecule drugs to tackle protein–protein interactions (PPIs) through stabilization of desired conformers: miniperspective. J Med Chem 61:4283–4289
Leach AR (2017) The ChEMBL database in. Nucleic Acids Res 45:D945–D954
Li Q, Cheng T, Wang Y, Bryant SH (2010) PubChem as a public resource for drug discovery. Drug Discov Today 15:1052–1057
Lill MA (2007) Multi-dimensional QSAR in drug discovery. Drug Discov Today 12:1013–1017
López-Vallejo F, Giulianotti MA, Houghten RA, Medina-Franco JL (2012) Expanding the medicinally relevant chemical space with compound libraries. Drug Discov Today 17:718–726
Lounkine E, Keiser MJ, Whitebread S et al (2012) Large-scale prediction and testing of drug activity on side-effect targets. Nature 486:361–367
Lucas X, Grüning BA, Bleher S, Günther S (2015) The purchasable chemical space: a detailed picture. J Chem Inf Model 55:915–924
Maeda MH, Kondo K (2013) Three-dimensional structure database of natural metabolites (3DMET): a novel database of curated 3D structures. J Chem Inf Model 53:527–533
Mann J (2002) Natural products in cancer chemotherapy: past, present and future. Nat Rev Cancer 2:143–148
Martínez-Treviño SH, Uc-Cetina V, Fernández-Herrera MA, Merino G (2020) Prediction of natural product classes using machine learning and 13C NMR spectroscopic data. J Chem Inf Model 60:3376–3386
Mathai N, Chen Y, Kirchmair J (2020) Validation strategies for target prediction methods. Brief Bioinform 21:791–802. https://doi.org/10.1093/bib/bbz026

Mathai N, Kirchmair J (2020) Similarity-based methods and machine learning approaches for target prediction in early drug discovery: performance and scope. Int J Mol Sci 21:3585
Merk D, Grisoni F, Friedrich L et al (2018) Computer-assisted discovery of retinoid X receptor modulating natural products and isofunctional mimetics. J Med Chem 61:5442–5447
Muigg P, Rosén J, Bohlin L, Backlund A (2013) In silico comparison of marine, terrestrial and synthetic compounds using ChemGPS-NP for navigating chemical space. Phytochem Rev 12: 449–457
Nakamura K, Shimura N, Otabe Y et al (2013) KNApSAcK-3D: a three-dimensional structure database of plant metabolites. Plant Cell Physiol 54:e4–e4
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
Ntie-Kang F, Mbah JA, Mbaze LM et al (2013a) CamMedNP: building the Cameroonian 3D structural natural products database for virtual screening. BMC Complement Altern Med 13:1–10
Ntie-Kang F, Zofou D, Babiaka SB et al (2013b) AfroDb: a select highly potent and diverse natural product library from African medicinal plants. PLoS One 8:e78085
Patridge E, Gareiss P, Kinch MS, Hoyer D (2016) An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov Today 21:204–207
Pereira F, Aires-de-Sousa J (2018) Computational methodologies in the exploration of marine natural product leads. Mar Drugs 16:236
Pilkington LI (2019) A Chemometric analysis of deep-sea natural products. Molecules 24:3942
Pilon AC, Valli M, Dametto AC et al (2017) NuBBEDB: an updated database to uncover chemical and biological information from Brazilian biodiversity. Sci Rep 7:1–12
Pithayanukul P, Leanpolchareanchai J, Saparpakorn P (2009) Molecular docking studies and anti–snake venom metalloproteinase activity of Thai mango seed kernel extract. Molecules 14:3198–3213
Podlogar BL, Muegge I, Brice LJ (2001) Computational methods to estimate drug development parameters. Curr Opin Drug Discov Devel 4:102–109
Reher R, Kim HW, Zhang C et al (2020) A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products. J Am Chem Soc 142:4114–4120
Reker D, Perna AM, Rodrigues T et al (2014a) Revealing the macromolecular targets of complex natural products. Nat Chem 6:1072–1078
Reker D, Rodrigues T, Schneider P, Schneider G (2014b) Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus. Proc Natl Acad Sci 111:4067–4072
Rodrigues T, Reker D, Kunze J et al (2015) Revealing the macromolecular targets of fragment-like natural products. Angew Chem Int Ed 54:10516–10520
Rodrigues T, Reker D, Schneider P, Schneider G (2016a) Counting on natural products for drug design. Nat Chem 8:531–541. https://doi.org/10.1038/nchem.2479
Rodrigues T, Sieglitz F, Somovilla VJ et al (2016b) Unveiling (−)-Englerin A as a modulator of L-type calcium channels. Angew Chem Int Ed 55:11077–11081
Rollinger JM, Schuster D, Danzl B et al (2009) In silico target fishing for rationalized ligand discovery exemplified on constituents of Ruta graveolens. Planta Med 75:195–204
Romano JD, Nwankwo V, Tatonetti NP (2018) VenomKB v2. 0: a knowledge repository for computational toxinology. Sci Data 2:150065
Romano JD, Tatonetti NP (2019) Informatics and computational methods in natural product drug discovery: a review and perspectives. Front Genet 10:368
Ru J, Li P, Wang J et al (2014) TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. J Cheminform 6:1–6
Saldívar-González FI, Angélica Pilón-Jiménez B, Medina-Franco JL (2019) Phys Sci Rev 4: 20180103
Saldivar-Gonzalez FI, Valli M, Andricopulo AD et al (2018) Chemical space and diversity of the NuBBE database: a chemoinformatic characterization. J Chem Inf Model 59:74–85

Sam E, Athri P (2019) Web-based drug repurposing tools: a survey. Brief Bioinform 20:299–316
Sánchez-Cruz N, Medina-Franco JL (2018) Statistical-based database fingerprint: chemical space-dependent representation of compound databases. J Cheminform 10:1–13
Schäfer T, Kriege N, Humbeck L et al (2017) Scaffold Hunter: a comprehensive visual analytics framework for drug discovery. J Cheminform 9:1–18
Schneider G, Reker D, Chen T et al (2016) Deorphaning the macromolecular targets of the natural anticancer compound doliculide. Angew Chem Int Ed 55:12408–12411
Schneider P, Schneider G (2017a) De-orphaning the marine natural product (±)-marinopyrrole A by computational target prediction and biochemical validation. Chem Commun 53:2272–2274
Schneider P, Schneider G (2017b) A computational method for unveiling the target promiscuity of pharmacologically active compounds. Angew Chem Int Ed 56:11520–11524
Seo M, Shin HK, Myung Y et al (2020) Development of natural compound molecular fingerprint (NC-MFP) with the dictionary of natural products (DNP) for natural product-based drug development. J Cheminform 12:6. https://doi.org/10.1186/s13321-020-0410-3
Shang J, Hu B, Wang J et al (2018) Cheminformatic insight into the differences between terrestrial and marine originated natural products. J Chem Inf Model 58:1182–1193
Shen M, Tian S, Li Y et al (2012) Drug-likeness analysis of traditional Chinese medicines: 1. Property distributions of drug-like compounds, non-drug-like compounds and natural compounds from traditional Chinese medicines. J Cheminform 4:1–13
Shin HK, Kang Y-M, No KT (2017) Predicting ADME properties of chemicals. In: Handbook of computational chemistry, vol 59. Springer, Cham, pp 2265–2301
Šícho M, Stork C, Mazzolari A et al (2019) FAME 3: predicting the sites of metabolism in synthetic compounds and natural products for phase 1 and phase 2 metabolic enzymes. J Chem Inf Model 59:3400–3412
Singh SB, Culberson JC (2009) Chapter 2: Chemical space and the difference between natural products and synthetics. In: Natural product chemistry for drug discovery. The Royal Society of Chemistry, Cambridge, pp 28–43
Sorokina M, Merseburger P, Rajan K et al (2021) COCONUT online: collection of open natural products database. J Cheminform 13:1–13
Sorokina M, Steinbeck C (2019) NaPLeS: a natural products likeness scorer—web application and database. J Cheminform 11:1–7
Sorokina M, Steinbeck C (2020) Review on natural products databases: where to find data in 2020. J Cheminform 12:1–51
Sterling T, Irwin JJ (2015) ZINC 15–ligand discovery for everyone. J Chem Inf Model 55:2324–2337
Terrett NK, Gardner M, Gordon DW et al (1995) Combinatorial synthesis—the design of compound libraries and their application to drug discovery. Tetrahedron 51:8135–8173
Tran TD, Ogbourne SM, Brooks PR et al (2020) Lessons from exploring chemical space and chemical diversity of propolis components. Int J Mol Sci 21:4988
Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579
Van Santen JA, Jacob G, Singh AL et al (2019) The natural products atlas: an open access knowledge base for microbial natural products discovery. ACS Cent Sci 5:1824–1833
Vanii Jayaseelan K, Moreno P, Truszkowski A et al (2012) Natural product-likeness score revisited: an open-source, open-data implementation. BMC Bioinform 13:1–6
Wang M, Carver JJ, Phelan VV et al (2016) Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol 34:828–837
Warren GL, Andrews CW, Capelli A-M et al (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49:5912–5931

Wetzel S, Schuffenhauer A, Roggo S et al (2007) Cheminformatic analysis of natural products and their chemical space. Chim Int J Chem 61:355–360

Yu MJ (2011) Natural product-like virtual libraries: recursive atom-based enumeration. J Chem Inf Model 51:541–557

Zaid H, Raiyn J, Nasser A et al (2010) Physicochemical properties of natural based products versus synthetic chemicals. Open Nutraceuticals J 3:194

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