Chapter 1
Introduction

1. Introduction

Nature offers a rich reservoir of medicinal agents for thousands of years and continues to be an abundant source of novel chemotypes and pharmacophores. The exploration of potential bioactive compounds derived from natural sources frequently involves the systematic screening of plant extracts [1]. According to the Newmann and Cragg report, approximately 4% of the approved drugs in the time frame of 1981-2018 were directly obtained from natural products, as depicted by the pie chart in Figure 1.1. The rest, 45%, of approved drugs indirectly belong to natural products as they are either semisynthetic derivatives, natural product mimics, or natural product pharmacophores [2]. The phytochemical investigation of plant parts has resulted in the identification of natural product-based lead for drug discovery. Natural products are a good option for drug discovery owing to their extensive structural diversity and physicochemical properties supported by specific structural features. These compounds exhibit vast scaffold diversity and complexity, featuring unusual functional groups, spatial conformation with more stereocentres, and sp^3 carbon atoms. Also, natural products cover a wider chemical space and are structurally optimized by the evolution of bioactive compounds associated with their particular biological function [3-5].

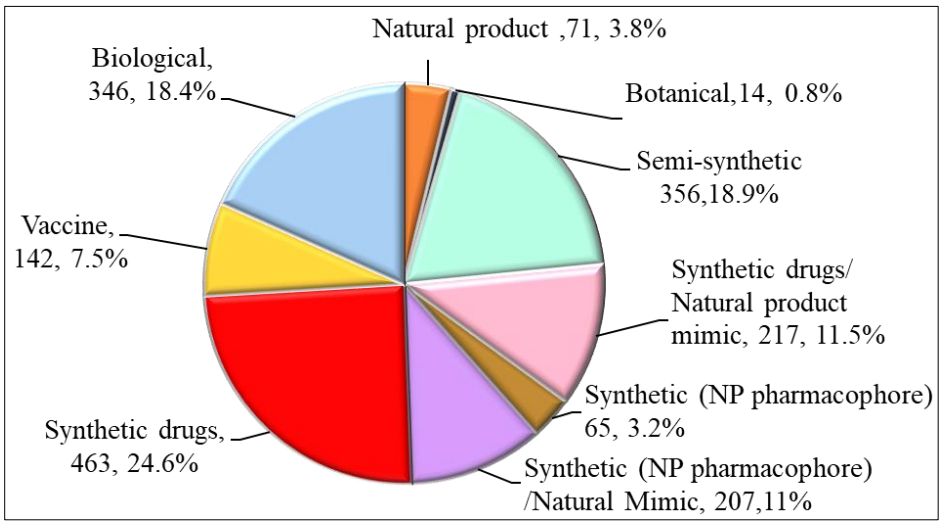

Figure 1.1. Comprehensive coverage of new approved drugs from 1981 to 2018.

1.1. Biodiversity inspired chemical diversity of natural products

The chemical diversity inherent in natural products is distributed across five biological kingdoms: Plantae, Animalia, Fungi, Bacteria, and Protista, with 67% of these compounds originating from the Plantae kingdom. Terpenoids and alkaloids emerge as the predominant chemical classes of natural products within these kingdoms, constituting more than half of all compounds isolated from the Plantae kingdom. The chemical diversity of natural products is influenced by biodiversity, primarily through the dynamic ecological interactions between plant and their environmental attributes like adaptation, co-evolution, abiotic stress, and species interaction [6]. These interactions have led to the production of diversified phytoconstituents (natural products), which are the secondary metabolites produced by plants to cope with the adverse effects of these attributes. Throughout evolutionary processes, nature has delved into a vast expanse of chemical possibilities, exemplified by secondary metabolites or natural products. The natural products are crafted by plants through enzymatic pathways, serving specific biological functions internally or in interactions with other organisms, thus conferring selective advantages upon the producer [7]. Driven by environmental pressures, plants generate secondary metabolites with diverse molecular scaffolds, each tailored to modulate different targets directly or indirectly, enhancing the survival rate. Consequently, these secondary metabolites serve as tangible manifestations of plant adaptive responses to their surroundings, promoting nature exploration of chemically relevant territories [8]. In this light, the natural product structures can be viewed as inherently validated by biological phenomena, having undergone a rigorous evolutionary screening process [9].

1.2. Exploration of natural products from plant source

The phytochemical exploration of plant-derived constituents involves several essential stages, including extraction, fractionation, isolation, and purification, as shown in

Figure 1.2. Nonetheless, this traditional methodology is accompanied by certain limitations. The isolation process typically extends over 1-2 years to identify a single compound, potentially leading to substantial time investment [10, 11]. Furthermore, the likelihood of re-isolating previously characterized compounds exists, posing a challenge to novelty in compound discovery. Additionally, phytoconstituents existing in low concentrations may evade detection, further complicating the comprehensive characterization of plant metabolites. In response to these challenges, natural product-based drug discovery programs have integrated diverse multidisciplinary approaches and innovative technologies like bioactivity-guided, metabolomics, genomic sequencing, dereplication, and computer-assisted structure elucidation. The successful use of these approaches has resulted in the identification of various natural product-based leads. Among these strategies, dereplication stands out as a pivotal approach utilized for the early and efficient identification of compounds [12, 13].

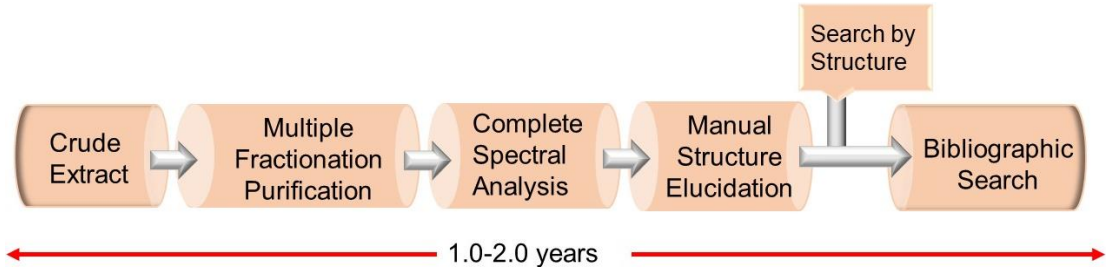

Figure 1.2. Conventional approach of isolation and identification of natural product from the extract.

1.3. Dereplication strategy

In the context of natural product discovery, dereplication is the strategy used to identify known compounds from the mixture of compounds present in the extract so that re-discovery and re-isolation of previously identified compounds could be avoided. The active compound is identified at the earliest stage using various analytical techniques and database searches. The term dereplication was originally used for the first time in the 1980 edition of the CRC Handbook of Antibiotics, which refers to recognizing and

eliminating active substances that had previously been evaluated in the early stages of the screening procedure [14]. The dereplication strategy incorporates five necessary steps: extraction, fractionation, data acquisition, data mining, and prioritization of novel compounds (Figure 1.3). Initially, natural sources are extracted using organic solvents and crude extract is obtained. The next step includes the fractionation of extract using organic and aqueous phases; further spectrometric techniques are applied to detect the majority of compounds in different fractions. Subsequently, once known compounds are identified and dereplicated, the unknown compounds can be prioritized for further isolation and structure elucidation [15]. This strategy is designed to identify the known compounds quickly from the extract so that those known compounds could be dereplicated or removed from consideration in the search for new bioactive. The dereplication strategy needs access to advanced separation techniques hyphenated with spectroscopic equipment for data acquisition. The hyphenated technique for natural product isolation incorporates a combination of separation techniques and a detection system [16]. `

Figure 1.3. Essential steps of dereplication strategy used for isolation and identification of natural products from the extract.

1.3.1. Data acquisition by hyphenated techniques

Natural product research hinges on two key elements: isolation and purification of compounds from crude extract from diverse natural sources. Therefore, the real-time characterization of secondary metabolites within crude natural product extracts or fractions requires advanced levels of sophistication, along with a wealth of structural data, sensitivity, and selectivity. Hyphenated techniques are the combination of chromatographic methods with spectroscopic or spectrometric techniques. Hirschfeld

coined the term "hyphenation" to describe the process of online coupling of a separation technique with one or multiple spectroscopic detection methods in real-time analysis. To identify compounds in a crude sample, liquid chromatography (LC), gas chromatography (GC), or capillary electrophoresis (CE) is combined with spectroscopic techniques such as Fourier-transform infrared (FTIR), UV-vis absorbance or fluorescence emission (PDA), mass spectrometry (MS), and nuclear magnetic resonance spectroscopy (NMR). This integration results in modern hyphenated techniques like CE-MS, GC-MS, LC-MS, and LC-NMR, providing structural information for compound identification. These integrated approaches effectively analyze a wide range of biological, chemical, and toxicological samples. The integration of HPLC with MS or NMR has significantly enhanced the ability to identify the structures of complex natural compounds [17, 18]. Due to its greater sensitivity, LC-MS has seen broader application than LC-NMR. Moreover, hyphenation is not limited to just two techniques; it encompasses multiple separation and detection methods. For instance, combinations like LC-PDA-MS, LC-MS-MS, LC-NMR-MS, LC-PDA-NMR-MS, and similar setups illustrate the versatility of coupling separation and detection techniques. The common hyphenated techniques used for the dereplication strategy are GC-MS, LC-MS, LC-MS-MS, LC-NMR-MS, and LC-NMR [19].

1.3.1.1. GC-MS

Utilizing mass spectrometry as the primary detection technique, LC-MS and GC-MS are the preferred hyphenated methods. Among the commonly employed instruments are single and triple-quadrupole, ion trap, and time-of-flight mass spectrometers, highlighting their widespread adoption and versatility in analytical applications. GC-MS, a hyphenation of gas chromatography and mass spectrometry, emerged as a pioneering tool in research and development. The resulting mass spectra from this coupling provide

extensive structural information through fragmentation interpretations. Comparison of fragment ions, each with varying relative abundances, against library spectra aids in compound identification. Compounds possessing sufficient volatility, small size, and stability at high temperatures in GC parameters can undergo GC-MS analysis. However, polar compounds, especially those featuring multiple hydroxyl groups, often necessitate derivatization for optimal GC-MS analysis. In a GC-MS, the sample is introduced into the injection port of the GC device, where it vaporizes, undergoes separation in the GC column, and is subsequently analyzed by the MS detector, with results recorded accordingly. The duration between injection and elution is referred to as "retention time" (tR) [20].

1.3.1.2. LC-NMR

LC-NMR is an innovative hybrid technique integrating liquid chromatography, a method for separation, with NMR, a spectroscopic detection method known for its structural elucidation capabilities. LC-NMR instrumentation includes an isolation zone (column), an interface zone, and a detection zone housing the NMR probe for spectroscopic analysis. This system operates under computer control, facilitating seamless coordination between HPLC and NMR operations. To enhance sensitivity and complement NMR measurements, a sensitive detector such as UV and MS are often integrated alongside, with careful adjustment of the splitting ratio to monitor the primary NMR detector's measurements concurrently. Numerous studies have highlighted the effectiveness of LC-NMR in analyzing diverse mixtures of natural products. Nonetheless, Spring and coworkers demonstrated the initial utilization of LC-NMR in investigating natural product extracts within a chemotaxonomic context in 1995. Their research focused on characterizing sesquiterpene lactones sourced from the Mexican plant *Zaluzania grayana.*

Notably, a novel lactone was identified, marking a significant milestone in LC-NMR application within this field [21].

1.3.1.3. LC-NMR-MS

This method capitalizes on the rapid and highly sensitive screening abilities offered by mass spectrometry, enabling the identification of specific analyte peaks within complex mixtures. These peaks can then undergo further structural analysis using NMR spectroscopy. The typical approach to connect LC with both MS and NMR is through the parallel mode, which involves splitting the eluent to generate two parallel flows. To accommodate the difference in sensitivity between NMR and MS, a splitter is employed to adjust the balance between the two split flows. Given the lower sensitivity of NMR relative to MS, a common practice is to use a split ratio of 95:5, favouring MS over NMR. LC-NMR-MS, an innovative analytical technique, made its debut in the realm of natural product research back in 1999. Since then, this advanced coupling method has seen multiple applications in the phytochemical investigation of plant metabolites [22].

1.3.1.4. LC-MS

Liquid chromatography-mass spectrometry (LC-MS) has emerged as a crucial technique in natural product discovery, particularly in metabolite profiling. LC-MS effectively analyzes metabolites even at their low concentration. LC-MS, through its hyphenated configuration, facilitates the identification and dereplication of metabolites, offering a robust platform for structural elucidation. The integration of LC-MS with complementary data such as retention time and UV/Vis spectra enhances the accuracy and reliability of metabolite identification. LC-MS generates abundant structural information quickly, enabling partial or complete online de novo structure determination of natural products [23]. This capability is further augmented when LC-MS is coupled with mass databases, facilitating expedited and precise identification of target compounds. Overall, LC-MS

serves as an indispensable tool in natural product discovery, offering rapid and comprehensive insights into the chemical composition and structure of metabolites, thereby accelerating the process of identifying novel bioactive compounds. Thus, LC-MS represents a potent integration of liquid chromatography (LC) and mass spectrometry (MS). This online coupling technique directly connects the LC system to the mass spectrometer, facilitating the introduction of eluent from the LC column into the ionization source of the mass spectrometer. This setup enables the sequential separation of complex mixtures by LC, followed by the sensitive and selective detection of individual components by MS. Moreover, LC-MS allows for real-time monitoring of chromatographic separation and immediate identification of compounds as they elute from the column [23, 24].

1.3.1. Data mining *via* natural products database

A natural product database used for dereplication is a valuable tool in natural product chemistry and drug discovery. After data acquisition *via* hyphenated techniques, database mining is performed in the dereplication process. This methodology necessitates the utilization of natural product databases to ascertain the identity of compounds within the complex crude extract. Various public domain, private domain, and commercial databases have been developed to assist in identifying natural products during dereplication. These databases typically contain comprehensive information about natural products, including their chemical structures, spectral data, molecular weight, chemical class, biological activities, and literature references. Researchers can quickly determine whether a compound is known or novel by comparing the spectral data obtained from a sample with the information stored in the database [25]. There are several well-known natural product databases commonly used in the dereplication process.

1.3.2.1. Super Natural II

Super Natural II is the pioneering public database of natural compounds, boasting an extensive collection of 326,000 molecules. This database is a freely available, web-based, user-friendly repository of natural products. It offers a rich array of information, including details on chemical class, mechanisms of action, pathway involvement, and toxicity profiles, all consolidated within a single platform. Moreover, it serves as a vigilant guardian, offering potential toxicity alerts to caution users regarding using specific natural compounds [26].

1.3.2.2. Dictionary of Natural Products (DNP)

DNP is one of the most extensive and widely used natural product databases. It contains information on over 300,000 natural products, including their chemical structures, physical properties, and biological activities. DNP is regularly updated and curated to ensure accuracy and relevance [27].

1.3.2.3. GNPS (Global Natural Products Social Molecular Networking)

GNPS is a platform that utilizes mass spectrometry data to identify natural products. It employs spectral networking algorithms to compare experimental data with reference spectra in its database, enabling rapid dereplication and structural elucidation of unknown compounds [28].

1.3.2.4. AntiBase

This database incorporates over 40,000 natural products and their chemical structure isolated from microorganisms and higher fungi. However, Antibase has yet to undergo updates since 2014. Additionally, access to this valuable resource is exclusively available for purchase via Wiley's website [29].

1.3.2.5. METLIN

It serves as a versatile database facilitating the characterization of known metabolites while also functioning as a cutting-edge technology platform for identifying known and unknown metabolites alongside other chemical entities. Boasting a vast repository of over 1 million molecules, METLIN encompasses a diverse array of compounds, including primary metabolites, toxins, small peptides, and natural products. METLIN is a high-resolution tandem mass spectrometry (MS/MS) database that plays a crucial role in the analysis process. This database is unique in incorporating data derived from reference standards and their labeled stable isotope analogs, a feature further enhanced by METLIN utilization of isotope-labeled microorganisms for guided analysis. Despite its wealth of information, METLIN does not offer straightforward data downloads. However, access to its platform is freely available for academic use, facilitating widespread exploration and research within the scientific community [30].

In the present work, considering the significance of the dereplication technique, an LC-MS-based dereplication strategy was established to identify new compounds from a relatively unexplored plant, *Dysoxylum malabaricum* Bedd. The selection of this plant species was guided by a chemosystematics approach. The genus *Dysoxylum* has been known for isolating triterpenoids and alkaloids with anticancer potential. Also, *Dysoxylum malabaricum* is a less-explored species present in most biodiverse hotspots, the Malabar region of India. The local folklore has recognized the therapeutic properties of various parts of this plant to address disease conditions. The bark of *Dysoxylum malabaricum* was selected for phytochemical investigation, and new metabolites were identified from the bark extract using LC-MS-based dereplication strategy.

Furthermore, chemical modifications were done in the isolated metabolites obtained in bulk amounts after the isolation and purification. Three reactive sites were targeted to

introduce functionalization in the basic scaffold of the beddomeilactone. Beddomeilactone was chemically modified to synthesize amide and halogenated derivatives. The synthesized derivatives have also shown enhanced cytotoxicity against breast cancer cell lines. Moreover, this chemical modification study resulted in the development of two new synthetic methods. The first method allowed zinc chloride-catalyzed amide synthesis using phenylhydrazine and carboxylic acid. The second method allowed Ni-NiO nanoparticle-catalysed regioselective halogenation and esterification of carbonyl compound using N-halosuccinamide. Also, the amidation synthetic reaction protocol allowed the amide formation of various amino acids and complex natural products, and the Ni-NiO catalyzed reaction protocol allowed regioselective halogenation and oxidative esterification of various substrates. The newly developed methods led to the synthesis of different derivatives of beddomeilactone exhibiting cytotoxic potential against cancer cell lines.