Chapter 1

Introduction

1 Introduction

1.1 Natural Products

When we think about natural products (NPs), the first thing that comes to mind is usually “something that comes from nature and doesn't have any artificial intervention”. If considering the broad definition of NPs, it could be anything from an entire organism (plant, microorganism, etc.) to its parts or its extracts/exudates or the pure molecules isolated from them [1]. But when the NP is being discussed in reference to drug discovery, it generally refers to the pure phytoconstituents/secondary metabolite. The secondary metabolites are small molecules produced by living organisms that are not necessarily important for their function and growth but are produced for combat purposes mainly [1]. These secondary chemical moieties from plants/animals/microbes have been used for treating human ailments since the beginning of medicine [2]. The secondary metabolites with simple/complex pharmacophoric scaffolds were and are considered to have a higher probability of being biologically active due to their tendency to interact with biomolecules [3]. The period between 1970-1980 was the golden period for NP in the Western pharmaceutical industry with research focused primarily on non-synthetic entities [2]. The result of this NP revolution was visible in the new chemical entities (NCEs) introduced in the period 1981-2014 and it witnessed that almost 50% of approved NCEs were either NP or NP-inspired or derived [4]. In the past decade, NP research has encountered a setback owing to persisting and emerging challenges.

1.2 Challenges associated with NP-based drug discovery

The powerful tool of pharmaceutical companies in drug discovery research is high-throughput screening (HTS) that involves screening large compound libraries against specific enzyme or receptor to identify leads with specific mechanisms [5, 6]. As far as synthetic molecules are concerned, HTS is a boon saving millions of dollars and time.

The HTS becomes problematic when we bring NPs into the picture. The foremost con is the crude extract, a complex mixture that many times contain polyphenolic compounds which are prone to bind with target proteins, thus producing false positive results in HTS [6]. Also, after identifying the lead extract, the purification and isolation of active compounds takes a long time and thus, is a problem for HTS and adds high costs [6].

As the phytoconstituents present in a medicinal plant are prone to seasonal, climatic, and geographical variations, so this also poses a significant challenge. There are very high possibilities that the crude extract of one lot shows good activity while other prepared from the same plant material in a different season or from a different region might not reproduce the same activity [6]. There are also difficulties associated with the purification of NP responsible for bioactivity from a crude extract. Some crude extract mixtures are too complex to be separated by conventional techniques and require advanced and sophisticated methods, thus adding high costs. Even after great efforts, the chances of that compound being already known are very high which produces a loss for the companies as they cannot be patented [6].

Amongst all, the one inevitable challenge is the material supply. There always persists the problem of the amount of active compound present in the plant material of interest. Often in crude mixtures, the biologically active compound is only a few milligrams, and if it passes the initial phases, the preclinical studies require several grams, and further clinical studies require several kilograms [6]. The extraction of such a huge amount requires tonnes of plant material which poses a severe threat of depletion of such treasured natural resources [6].

Lastly, since the leading pharma companies are more intent on making revenues, thus they rely more on screening pure synthetic libraries as they are ‘screening friendly’ and easily synthesizable when scale-up is needed [7]. Thus, the big pharma conglomerates see NP-related drug discovery as a loss-associated investment, and thus, only research-oriented small pharma companies and academics are working on these [6].

1.3 NPs over synthetic molecules

Despite the popularity, synthetic small molecules have gained in recent years; they still can't overtake the NPs. There are several advantages of NPs that help them in maintaining strong foot in drug discovery. Compared to synthetic molecules, NPs offer remarkable diversity, chemical space, and molecular architecture [8]. Despite advancements in the organic synthesis, the synthetic chemists are unable to keep up with or compete with nature's chemists as plants are able to produce diversified metabolites by using only a limited set of building blocks and moulding them via complex network of biosynthetic pathways [9]. Nature and medicinal chemist also differ in the kind of synthetic changes they carry out. Nature being attracted to oxygen has evolved enzymes that precisely perform specific C–H activations to add oxygen with differing levels of oxidation. Conversely, medicinal chemists focuses on nitrogen, and frequently incorporate naturally lesser common atoms like sulfur and halogens [10]. Also, unlike nature medicinal chemists are more comfortable within the 'flatland' molecules that lack intricate stereochemical characteristics [8].

In drug discovery, where initial leads undergo rigorous medicinal chemistry transformation before choosing a candidate, NPs offer valuable flexibility due to their potential for further enhancement. NPs can be refined for increased efficacy and specificity toward the target, or for achieving optimal pharmacokinetics and pharmacodynamics. Taking the example of morphine, a solely nature-derived drug, has inspired the discovery of numerous semisynthetic and synthetic compounds sharing its pharmacophore, resulting in successful second-generation opioid analgesics [8].

1.4 Intersection of NPs and drug development

NPs remained the source of inspiration for numerous therapeutic agents in drug discovery. Owing to evolution, NPs exhibit a broad array of biological activities hence, considered privileged scaffold. A number of NPs possess showcase high diversity with favourable ADME [11]. Comparisons of the information presented on sources of new drugs from 1983 to 1994, 60 and 75% of approved anticancer and anti-infective drugs, respectively originated from natural sources. The statistics of US FDA-approved drugs from 1981-2002 shows that, out of total 1031 NCEs approved, 42% belonged to NPs that included unaltered NPs, NPs derivatives, NP-inspired pharmacophore, and NP mimetics [12]. While the figure from 1981-2019 showed that, among 1881 NCEs approved, 699 (around 37%) were of natural origin. Among these 699 natural drugs, share of anticancer and antibacterial drugs was 17% (119) and 13% (90), respectively [13]. Among the FDA-approved small-molecule drugs from 1981 and 2019, around two-thirds are related to NPs. About 5%, 28% and 35% small-molecules launched between this timeline are unaltered NPs, NP derivatives, and mimic and/or contain a NP pharmacophore, respectively [14]. Though the fluctuation has been observed in the contribution of NPs towards drug discovery in the last 20 years but the numbers are still significant. In the year 2023, three NPs derivatives, have been approved by US-FDA which include Omaveloxolone, a derivative of natural oleanane triterpenoid from Prunella vulgaris for the treatment of Friedrich's ataxia; Taurolidine, a derivative of natural amino acid taurine for the treatment of catheter-related bloodstream infections; and Eflornithine, a derivative of natural amino acid ornithine for the treatment of high-risk neuroblastoma (Figure 1.1) (www.fda.gov**)** [15].

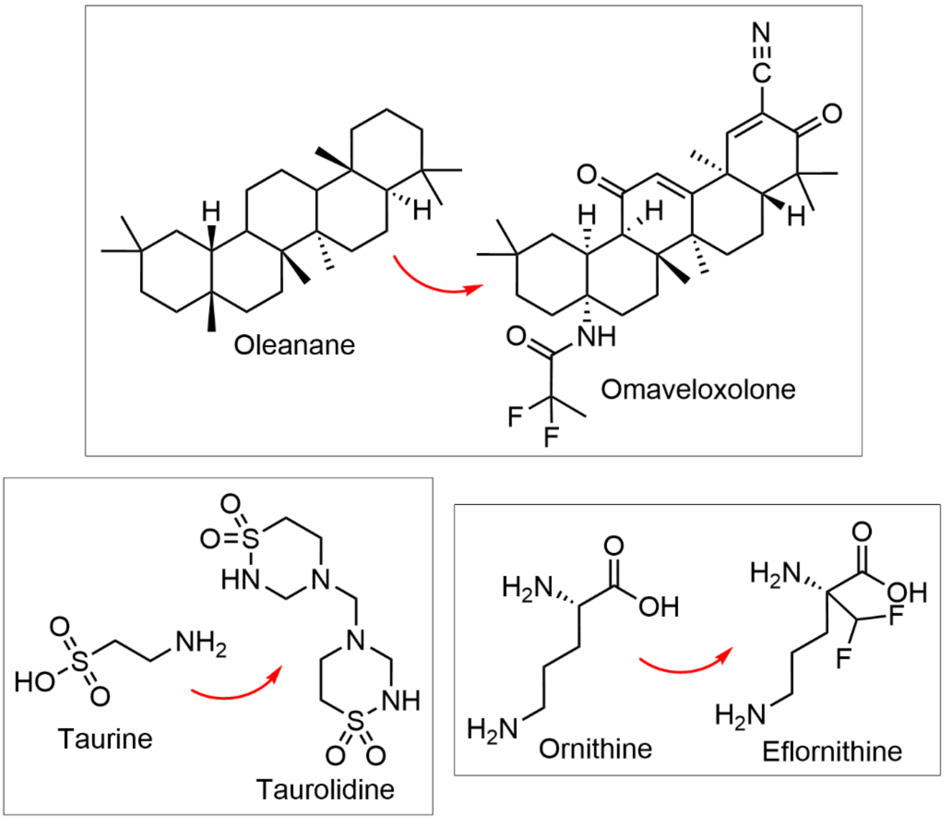


Figure 1.1 NP derivatives approved by US-FDA in 2023.

NPs have remained an important source of anti-infective drugs in the late 1990s and 2000, with more than 50% anti-infectives from natural origin. In recent years, the anti-infective drugs from natural sources has significantly declined due to antimicrobial resistance [16]. Amidst these situations, Lewis group from Northeastern University, Boston has identified two new antibiotics Teixobactin and Darobactin in 2015 and 2019, respectively (Figure 1.2) [13]. Teixobactin is a depsipeptide isolated from soil bacteria, *Eleftheria terrae* with excellent activity against Gram-positive bacteria including drug-resistant strains like MRSA, VISA, VRE, and Mycobacterium tuberculosis while ineffective against most Gram-negative bacteria [17]. Darobactin is a heptapeptide from animal pathogen, *Photorhabdus khanii* and active against Gram negative pathogens [18].

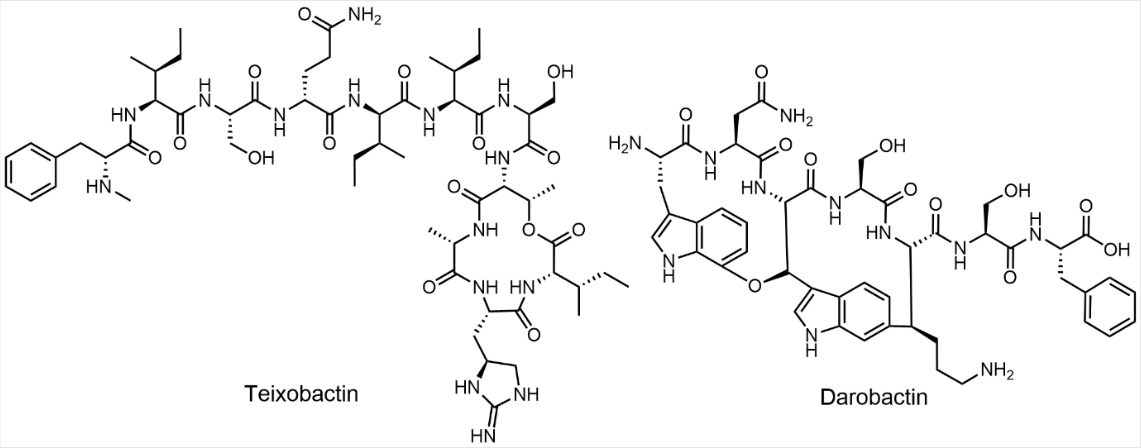


Figure 1.2 Structures of antibiotics Teixobactin and Darobactin.

1.5 Natural lactones in drug discovery

Within nature's vast chemical reservoir, plant-derived cyclic esters i.e., lactones are one of the most abundant structural scaffolds produced by plants. Lactones exhibit diverse activities due to presence of many fragments like alkyl, lipophilic, and sulfhydryl groups due to which they have piqued significant scientific interest among researchers as potential drugs [19]. Lactones showcase remarkable structural diversity, encompassing simple rings to complex polyesters, offering a unique framework for lead compounds. They exhibit varied biological potential, including anti-tumor, antimicrobial, anti-inflammatory, antifungal, antiparasitic, etc. Some of the plant families rich in lactone containing compounds are Apiaceae, Fabaceae, Rutaceae, Asteraceae, Lauraceae and Euphorbiaceae. The diverse classes of naturally occurring lactones produced by plants includes α, β, γ, δ, ω, and other macrolactones like strigolactone and macrolactin (Figure 1.3).

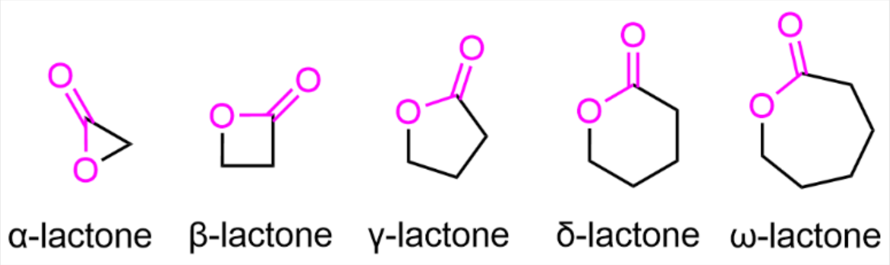


Figure 1.3 Common types of lactones found in nature.

Among the diverse natural lactones, γ- and δ-lactones are the most abundant in nature and often occur in combination with other functional moieties. The DNP database contains more than 3.4 lakh natural compounds, and in that β (170), γ (21452), δ (6086), and ω-lactones (2012) constitutes around 30,000 compounds that corresponds to ~ 9-10% without including macrolactones. Plant-derived coumarins are benzenoid δ-lactones with rich chemical diversity and diverse bioactivities. Coumarins offer wide structural diversity, ranging from simple furan derivatives to complex polycyclic structures. Their central core is readily amenable to diverse substitutions and modifications, making it valuable for drug development (Figure 1.4). The γ-lactones are important plant-derived lactones with immense biological potential. They showcase remarkable structural diversity, encompassing simple five-membered rings to complex polyhydroxylated derivatives, offering a rich platform for lead discovery. More than 7000 sesquiterpene in nature are in lactone forms (particularly γ-lactones), making sesquiterpene lactones (SLs) one of the largest classes of plant-derived secondary metabolites. With staggering structural features and biological potential, SLs offer a vast natural chemical library to explore them as versatile therapeutic candidates. Iridoids are another less explored plant-derived lactone-containing plant metabolite. Iridoids exhibit a fascinating chemical landscape with a rich diversity, encompassing monoterpene-derived cyclic structures with various functional groups, ranging from simple glycosides to complex polyhydroxylated derivatives.

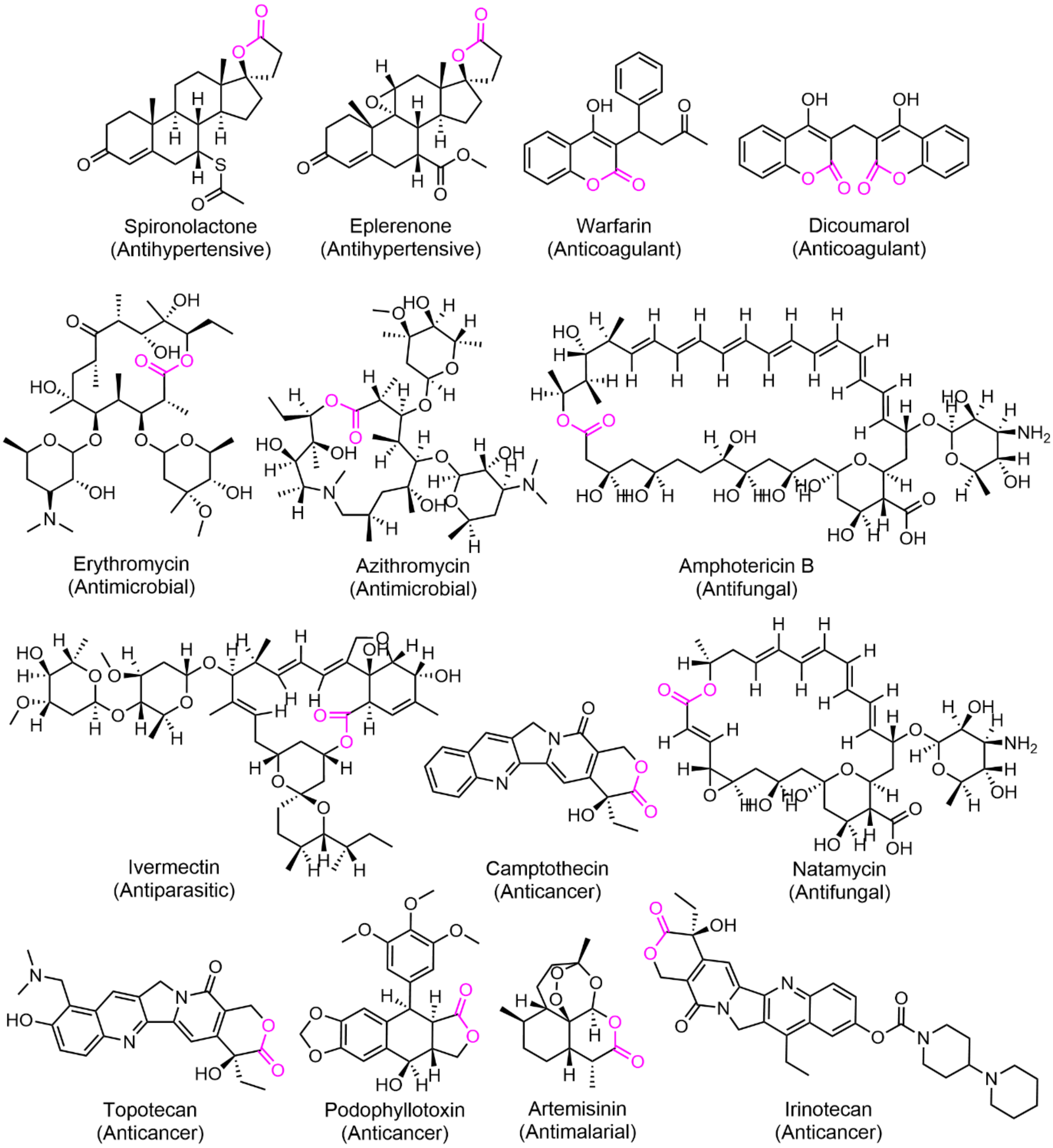
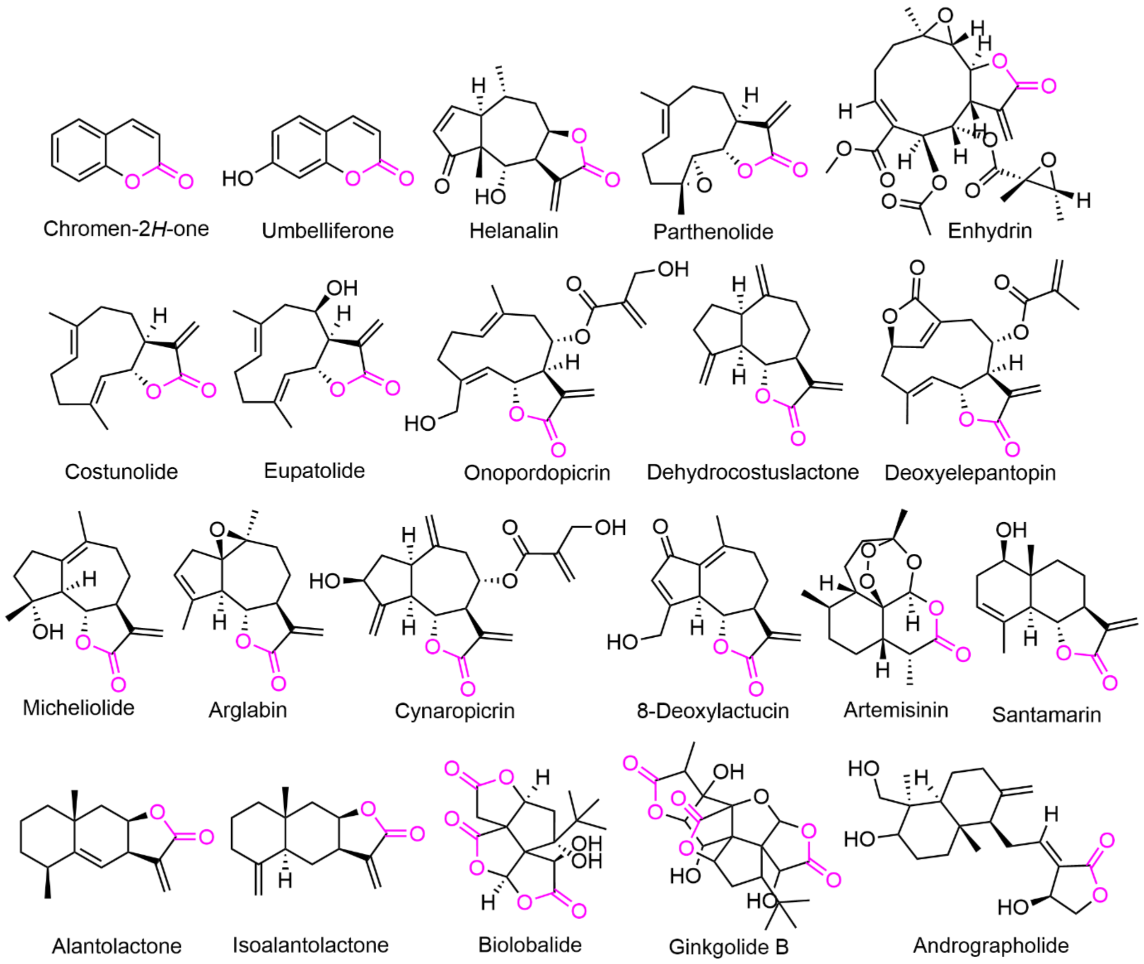


Figure 1.4 Clinically used drugs containing lactone moiety.

1.5.1 Natural lactones in inflammation

Studies have illustrated the potential of coumarins in reducing tissue inflammation and edema, along with their ability to hinder prostaglandin production, a crucial inflammatory mediator. Helenalin, a bifunctional sesquiterpene lactone showcases remarkable anti-inflammatory properties. Helenalin inhibits leukotriene biosynthesis, which is pivotal in the cascade of inflammatory processes. Parthenolide, a sesquiterpene lactone derived from feverfew (*Tanacetum parthenium*), has a history of traditional use in preventing migraine headaches and addressing rheumatoid arthritis due to its potent anti-inflammatory properties. Numerous studies have highlighted parthenolide's effectiveness as a potent inhibitor of nuclear factor-kappa B (NF-κB) activation, effectively curbing the expression of pro-inflammatory cytokines [20]. Studies have reported that Parthenolide (from *Magnolia grandiflora*) and Enhydrin (from *Smallanthus uvedalius*) have shown significant blocking of the hyperalgesic response and notable attenuation of the edema response in animal studies [21]. Costunolide, isolated from *Saussurea lappa* L., has demonstrated the ability to modulate various intracellular signaling pathways associated with inflammation [22]. Eupatolide has demonstrated the ability to inhibit the activation of both MAPKs and protein kinase B (Akt) pathways, resulting in decreased expression of inflammatory mediators, thus possessing anti-inflammatory potential [23]. Onopordopicrin (from *Onopordum illyricum*) has shown potential as a potent NF-κB and STAT3 inhibitor [24]. Deoxyelephantopin has been found to attenuate macrophage activation, suggesting a novel strategy for treating inflammatory diseases. Additionally, it effectively reduces levels of TNF-α and IL-6 while also decreasing the activity of iNOS and COX-2 proteins in a hepatic inflammation mouse model [25]. Dehydrocostuslactone demonstrated its anti-inflammatory effects by inhibiting several inflammatory pathways including suppression of IL-6/STAT3 pathway activation, reduction of inflammatory cytokine levels, and downregulation of inflammatory mediators such as iNOS and COX-2 [26]. Other lactones with anti-inflammatory activities include micheliolide, cynaropicrin, arglabin, 8-deoxylactucin, alantolactone, isoalantolactone, santamarin, artemisinin, thapsigargin, ginkolide, biolobalide, andrographolide, and several others [27-29]. 5Z-7-oxozeaenol, a fungal resorcylic acid lactone has been reported to inhibit endotoxin-induced production of TNF-α and anisomycin-induced JNK/p38 activation with unclear anti-inflammatory actions. (Figure 1.5) [30].



Chromen-2*H*-one
Umbelliferone
Helanalin
Parthenolide
Enhydrin
Costunolide
Eupatolide
Onopordopicrin
Dehydrocostuslactone
Deoxyelepantopin
Micheliolide
Arglabin
Cynaropicrin
8-Deoxylactucin
Artemisinin
Santamarin
Alantolactone
Isoalantolactone
Biolobalide
Ginkgolide B
Andrographolide

Figure 1.5 Plant-derived lactones with anti-inflammatory properties.

1.5.2 Natural lactones in cancer

Numerous natural lactones have surfaced as promising anti-cancer agents in recent decades. Extensive research is underway to elucidate their anticancer activity, molecular mechanisms, and their potential as chemotherapeutic and chemoprotective agents [31]. Sesquiterpene lactones isolated from *Vernonia scorpioides* (Asteraceae) exhibited potent cytotoxicity against the cervical cancer cell line, HeLa [32]. Bioactive lactones, marginatoxin, and (3,4-Dimethoxybenzyl)-2-(3,4-0-methylene-dioxybenzyl) butyrolactone, isolated from *Bupleurum marginatum* (Apiaceae), demonstrated significant cytotoxicity against A549 cell lines. [33]. The lactone compounds, vernopicrin and vernomelitensin from *Vernonia guineensis* have shown to possess cytotoxicity against A549 cells [34]. Sausinlactones A and B, isolated from *Saussurea involucrata*, have also exhibited notable cytotoxicity against A549 cell lines [35]. Millerenolide from *Viguiera sylvatica* and thieleanin from *Decachaeta thieleana* have been studied for their anticancer properties in both *in vitro* tests on A549 cells and *in vivo* experiments using mice infected with melanoma cells. Both compounds displayed a similar level of cytotoxicity against A549 cells [36].

(3R)-Lasiodiplodin, isolated from *Cyphostemma greveana*, exhibits potent antileukemic activity. In tests against the A2783 ovarian cancer cell line, it displayed an IC_{50} value of 5.8 µM. trans-Resorcylide, originally used as a plant growth inhibitor, demonstrated cytotoxic activity against mouse L-5178 Y tumor cells, with an ID_{50} of 1.5 µg/mL [37, 38]. Curvularin and α, β-dehydrocurvularin have exhibited cytotoxic activities against breast cancer cells MDA-MB-231, with IC_{50} values of 1.3 µM and 9.3 µM, respectively. Additionally, α, β-dehydrocurvularin demonstrated activity against A549 lung cancer cells (IC_{50} = 2.1 µM) and COLO 205 cells (IC_{50} = 7.9 µM). The following compounds: Curvularin, α,β-dehydrocurvularin, 11-α-hydroxycurvularin, 11-β-hydroxycurvularin, 11-α-methoxycurvularin, and 11-β-methoxycurvularin, have demonstrated cytotoxic activities against a panel of four cancer cell lines (NCI-H460, MCF-7, SF268, and MIA PaCa-2), displaying IC_{50} values that range from 0.6 to 7.6 µM. [39, 40]. The compounds α, β-dehydrocurvularin, Sumalarin A, B, and C were tested for their cytotoxic activities against several tumor cell lines- Du145, HeLa, Huh 7, MCF-7, NCI-H460, SGC-7901, and SW1990. They exhibited cytotoxic effects against all tested cell lines, with IC_{50} values ranging from 3.8 to 10 µM (Figure 1.6) [41].

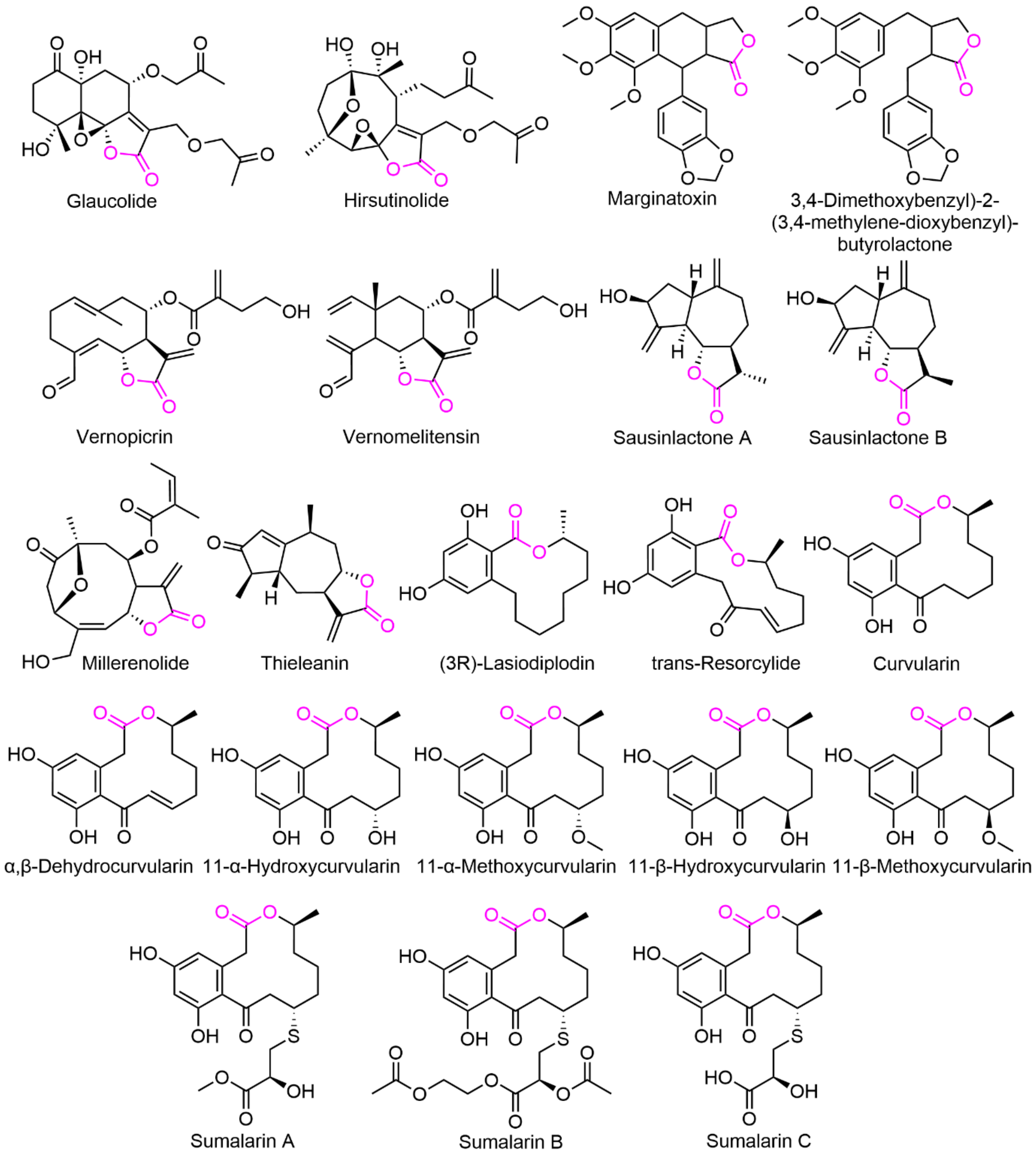


Figure 1.6 Plant-derived lactones with cytotoxic properties.

1.6 Applications of Cheminformatics in NP-based drug discovery

Modern *in silico* techniques are pivotal in expediting and mitigating risks in NP-based drug discovery. These computational methods significantly aid researchers across hit discovery, hit-to-lead, and lead optimization phases. Their capability lies in identifying bioactive NPs and streamlining the selection of (plant) materials for experimental testing for most promising candidates [42]. Computational methods are utilized in diverse applications like data curation; analysis, visualization, navigation, comparison of chemical space; NP dereplication; prediction of bioactivity, ADME, and safety profiles (toxicity); quantification of NP-likeness; and de novo design inspired by NPs [14].

1.6.1 NP libraries/databases

Over the past decade, there is a significant surge in databases offering access to chemical, biological, pharmacological, toxicological, and structural data on NPs. Many of these databases provide free bulk downloads, facilitating virtual screening and other applications [14]. The examples of extensively utilized NP databases include the DNP Database (>3.4 lakh NPs), Super Natural II Database (>3.25 lakh NPs), TCM Database@Taiwan (>60k NPs sourced from Chinese medicinal herbs), NP Atlas (data on over 25k NPs from fungi and bacteria), and the Collective Molecular Activities of Useful Plants (CMAUP) database (housing a collection of over 47k NPs) [43-46].

1.6.2 Computational methods employed in NPs characterization and dereplication

Computational methods are now frequently utilized to predict MS fragmentation and NMR spectra, often in conjunction with structure generators. Various algorithms enable the conversion of spectral data into representations that facilitate the efficient comparison of spectra based on similarity [47]. These methods can discern spectra from identical compounds and structurally akin compounds, thus offer insights into chemical classes and functional groups of unknown NPs. However, these analysis still necessitate manual intervention, thus constraining full automation [14]. The computer-assisted dereplication involves integrating analytical data with multivariate data analysis using principal component analysis (PCA), clustering methods, and discrimination analysis techniques to identify potential NPs within complex mixtures [48]. The NP-StructurePredictor tool offers a unique approach to NP dereplication. By employing a structure generator, it can amalgamate various scaffolds and fragments from extensive NP databases, thus deducing structures from related scaffolds [49].

1.6.3 Analyzing NP properties and structural diversity through computational approaches

Cheminformatics is playing a pivotal role in characterizing NPs based on their physicochemical and structural properties [50]. It facilitates comparisons between NPs, small-molecule drugs, and drug-like compounds. NPs occupy a significantly wider chemical space often exploring areas that are typically inaccessible through synthetic means [51]. Their structural characteristics enable them to target traditionally undruggable macromolecules [52]. NPs structural diversity extends to the fragment level, and this could be assessed using computational methods [53]. A valuable tool for visually exploring the structural diversity within sets of compounds is Scaffold Hunter [54]. PCA is a widely utilized technique for mapping chemical space in a manageable, lower-dimensional space while minimizing information loss [55]. The ChemGPS-NP method is used to map the chemical space of small natural molecules, predict modes of action, and analyze structure-activity relationships [56]. In addition, self-organizing maps and generative topographic maps are employed to compare the molecular structures of NPs with drugs [57]. Other effective techniques include T-distributed stochastic neighbor embedding (t-SNE) and Uniform manifold approximation and projection for dimension reduction (UMAP) method. t-SNE can generate visualizations that often surpass PCA.

UMAP produces similar results to t-SNE but is faster [58, 59]. Several methods have been developed to characterize, visualize, and compare NPs like Consensus diversity plot (CDP) and Statistical-based database fingerprint (SB-DFP) method [60, 61].

1.6.4 Assessment of NP-likeness using cheminformatics

Cheminformatics tools excel in accurately discriminating NPs and NP-like compounds from synthetic counterparts, often with remarkable precision. They are frequently employed in compound and library design and in selecting NPs and their derivatives from diverse compound collections. These tools aid in compound prioritization processes [62]. The NP-Likeness Score is used to quantify the degree of NP-likeness in compounds based on fragment resemblance to those found in known NPs [63]. NP-Scout, a freely accessible tool utilized to identify NPs and NP-like compounds within extensive sets of molecules [64]. The Natural compound molecular fingerprint (NC-MFP) is a novel approach to delineate structural characteristics of NPs through scaffolds and fragments depiction [65].

1.6.5 Identification of bioactive NP

Computational methods have played significant role in finding bioactive NPs. Many virtual screening techniques have been deployed in NP research, starting from 2D molecular fingerprint-based approaches to intricate 3D methodologies rooted in molecular shape similarity, docking, molecular interaction fields, and pharmacophore models. Recently, machine learning approaches have also been employed in virtual screening for bioactive NPs [66]. Virtual screening through docking algorithms has demonstrated efficacy in predicting binding modes accurately that offer invaluable insights for the optimization strategies. Notably, in the NP research, docking facilitates the rationalization of stereoselectivity in ligand binding [67].

1.6.6 Prediction of the macromolecular targets of NPs

Understanding the macromolecular target(s) of small molecules is pivotal in assessing their pharmacological effectiveness and safety for their progression in development. The experimental identification of these targets is often lengthy and costly, with chances of failure. Computational approaches offer promise in streamlining the identification of modes of action [14]. *In silico* target prediction is a virtual screening method, where one or numerous compounds are screened against an extensive array of macromolecules [68, 69]. ROCS is a leading, shape-based screening engine that leverages chemical feature distributions to identify the macromolecular targets of complex small molecule NPs. Besides these methods, 3D pharmacophore-based approaches also serve as common tool for target prediction in NP research. Recently, the machine learning models gaining the most attention for NP target prediction are SPiDER, TIGER, and STarFish [70-72].

1.6.7 De Novo design of nature-inspired compounds and compound collections

The limited synthetic accessibility of NPs and NP-derived compounds presents a significant challenge to their exploration and utilization. To address this challenge, several synthetic strategies for designing synthetically accessible compounds with NP-like properties are developed that include diversity-oriented synthesis (DOS) and biology-oriented synthesis (BIOS) [73]. Turning to computational methods, De novo design of Generic Structures (DOGS) is a de novo design tool. DOGS utilizes information of over 25,000 readily available synthetic building blocks along with a comprehensive set of reaction rules to generate compounds that are likely to be synthetically accessible [74].

1.6.8 ADME and safety profiles of NPs

NP-based drug discovery often encounters challenges concerning the ADME and safety profiles of NPs. Among the anti-targets addressed by NPs, the important ones are the human ether-a-go-go-related gene (hERG) channel (blockage causes potentially fatal cardiac arrhythmias) [75], cytochrome P450 enzymes (cause drug-drug interactions and toxicity), and P-glycoprotein (an efflux pump contributing to drug resistance). Various computational models, including statistical models, machine learning models, pharmacophore models, and docking methods, have been developed to address these anti-targets in NP research [76, 77]. FAME 3, a machine learning model has shown promising performance in predicting the sites of metabolism of small molecules and even NPs [78].