THE EVOLVING ROLE OF NATURAL PRODUCTS IN DRUG DISCOVERY

Frank E. Koehn and Guy T. Carter

Abstract | Natural products and their derivatives have historically been invaluable as a source of therapeutic agents. However, in the past decade, research into natural products in the pharmaceutical industry has declined, owing to issues such as the lack of compatibility of traditional natural-product extract libraries with high-throughput screening. However, as discussed in this review, recent technological advances that help to address these issues, coupled with unrealized expectations from current lead-generation strategies, have led to a renewed interest in natural products in drug discovery.

PHARMACOPHORE
The ensemble of steric and electronic features that is necessary to ensure optimal interactions with a specific biological target structure and to trigger (or to block) its biological response.

NEW MOLECULAR ENTITY (NME). A medication containing an active ingredient that has not been previously approved for marketing in any form.

COMBINATORIAL CHEMISTRY
The generation of large collections, or 'libraries', of compounds by synthesizing combinations of a set of smaller chemical structures.

Wyeth Research, 401 North Middletown Road, Pearl River, New York 10965, USA.
Correspondence to F.E.K & G.T.C
e-mails: KOEHNF@wyeth.com; CARTERG@wyeth.com
doi:10.1038/nrd1657
Published online 24 February 2005

Chemical substances derived from animals, plants and microbes have been used to treat human disease since the dawn of medicine. The investigation of natural products as source of novel human therapeutics reached its peak in the Western pharmaceutical industry in the period 1970–1980, which resulted in a pharmaceutical landscape heavily influenced by non-synthetic molecules. Of the 877 small-molecule New Chemical Entities (NCEs) introduced between 1981 and 2002, roughly half (49%) were natural products, semi-synthetic natural product analogues or synthetic compounds based on natural-product PHARMACOPHORES[1].

Despite this success, pharmaceutical research into natural products has experienced a slow decline during the past two decades. Although these trends seemed obvious to investigators working in the field, their downstream effects are somewhat difficult to measure precisely, given the long product-development cycles encountered in the pharmaceutical industry. The lengthy delay — usually ten years or more — between the initial discovery of a potential therapeutic agent and subsequent market launch of a NEW MOLECULAR ENTITY (NME) means that agents reaching the market today are typically the products of discovery research programmes initiated at least a decade ago.

However, it is possible to gain a reasonable picture of natural-product pharmaceutical discovery research by examining the general trends found in patent statistics. FIGURE 1 shows worldwide pharmaceutical small-molecule natural-product patents granted in the years 1984–2003. The chart shows a period of increasing patent activity through the 1980s, a flattening or even slight decline from 1990 to 1999, and a pickup of activity between 2000–2003. In the context of worldwide pharmaceutical R&D spending, which essentially tripled from roughly US $10 billion to US $30 billion over the same period, the overall trend in the 1990s can be viewed as a relative decline. What are the underlying reasons for the past decline, and what are the future prospects for natural-product research in drug discovery?

The decreased emphasis in the pharmaceutical industry on the discovery of natural products during the past decade can be attributed to a number of factors: first, the introduction of high-throughput screening (HTS) against defined molecular targets, which prompted many companies to move from natural products extract libraries towards so-called 'screen friendly' synthetic chemical libraries; second, the development of COMBINATORIAL CHEMISTRY, which at first offered the prospect of simpler, more drug-like screening libraries of wide chemical diversity; third, advances in molecular biology, cellular biology and genomics, which increased the number of molecular targets and prompted shorter drug discovery timelines; fourth, a declining emphasis among major pharmaceutical companies on infectious disease therapy, a traditional area of strength for natural products[2]; and last,

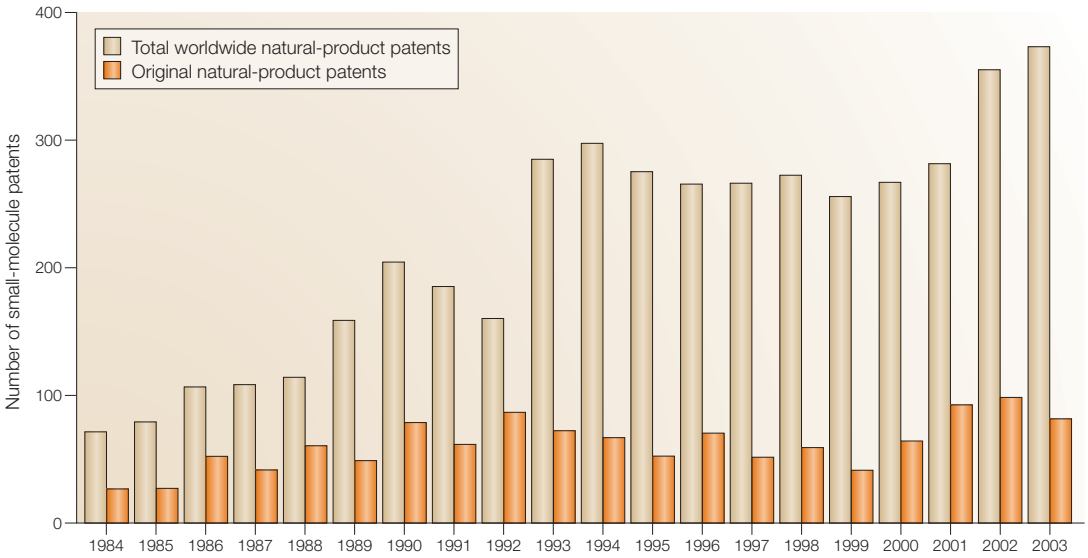


Figure 1 | **Worldwide pharmaceutical natural-product patents.** Total natural-products patents are shown in gold. The data are for all worldwide grants for patents claiming composition-of-matter or use of small-molecule natural products as pharmaceuticals. A single natural product can give rise to several patents based on filings in multiple countries, or for multiple indications. Original natural-product patents are shown in orange. The data are for all first-time grants of patents claiming novel composition of small-molecule natural products as pharmaceuticals. These statistics were derived by systematic search of the Derwent World Patents Index.

possible uncertainties with regard to collection of biomaterials as a result of the 1992 Rio Convention on Biological Diversity[3].

The underlying reasons for these industry trends are as much commercial as they are scientific, particularly in the case of research into infectious disease[4]. As a result of these factors, today's drug discovery environment calls for rapid screening, hit identification and hit-to-lead development. In this environment, traditional resource-intensive natural-product programmes that are based on extract-library screening, bioassay-guided isolation, structure elucidation and subsequent production scale-up face a distinct competitive disadvantage when compared with approaches that utilize defined synthetic chemical libraries. However, emerging trends, coupled with unrealized expectations from current R&D strategies, are prompting a renewed interest in natural products as a source of chemical diversity and lead generation.

Natural products, chemical diversity and HTS

Characteristics of chemical diversity found in natural products and synthetic libraries. Current thinking in the generation of drug leads embodies the concept of achieving high molecular diversity within the boundaries of reasonable DRUG-LIKE properties[5,6]. It has long been recognized that natural-product structures have the characteristics of high chemical diversity, biochemical specificity and other molecular properties that make them favourable as lead structures for drug discovery, and which serve to differentiate them from libraries of synthetic and combinatorial compounds[7]. Various investigators have worked to measure by means of computational chemistry those desirable chemical features that distinguish natural products from other sources of drug leads. Feher and Schmid[8] examined representative combinatorial, synthetic and natural-product compound libraries on the basis of molecular diversity and 'drug-likeness' properties such as molecular mass, number of chiral centres, molecular flexibility as measured by number of rotatable bonds and ring topology, distribution of heavy atoms, and LIPINSKI-type[9] descriptors. Other investigators have differentiated natural products, trade drugs or other synthetic molecular libraries on the basis of scaffold architecture and pharmacophoric properties[10], or other molecular descriptors[11]. These studies reveal that natural products typically have a greater number of chiral centres and increased steric complexity than either synthetic drugs or combinatorial libraries[8,12]. Although drug and combinatorial molecules tend to contain a significantly higher number of nitrogen-, sulphur- and halogen-containing groups, natural products bear a higher number of oxygen atoms[8,11]. Multivariate statistical analysis of molecular descriptors shows that natural products differ significantly from synthetic drugs and combinatorial libraries in the ratio of aromatic ring atoms to total heavy atoms (lower in natural products), number of solvated hydrogen-bond donors and acceptors (higher in natural products) and by greater molecular rigidity. Natural-product libraries also have a broader distribution of molecular properties such as molecular mass, octanol–water partition coefficient and diversity of ring systems compared with synthetic and combinatorial counterparts[8,10,11]. Indeed, less than one-fifth of the ring systems found in natural products are represented in current trade drugs. A perhaps unexpected finding is that of Schneider and Lee who revealed that the fraction of natural product structures with two or more 'rule-of-five'[9] violations is quite low (approximately 10%) and equal to that of trade drugs[11].

DRUG-LIKE
Sharing certain characteristics with other molecules that act as drugs. The set of characteristics — size, shape and solubility in water and organic solvents — varies depending on who is evaluating the molecules.

LIPINSKI'S 'RULE-OF-FIVE'
Lipinski's analysis of the World Drug Index led to the 'rule-of-five', which identifies several key properties that should be considered for small molecules that are intended to be orally administered. These properties are: molecular mass <500 Da, number of hydrogen-bond donors <5; number of hydrogen-bond acceptors <10; calculated octanol–water partition coefficient (an indication of the ability of a molecules to cross biological membranes) <5.

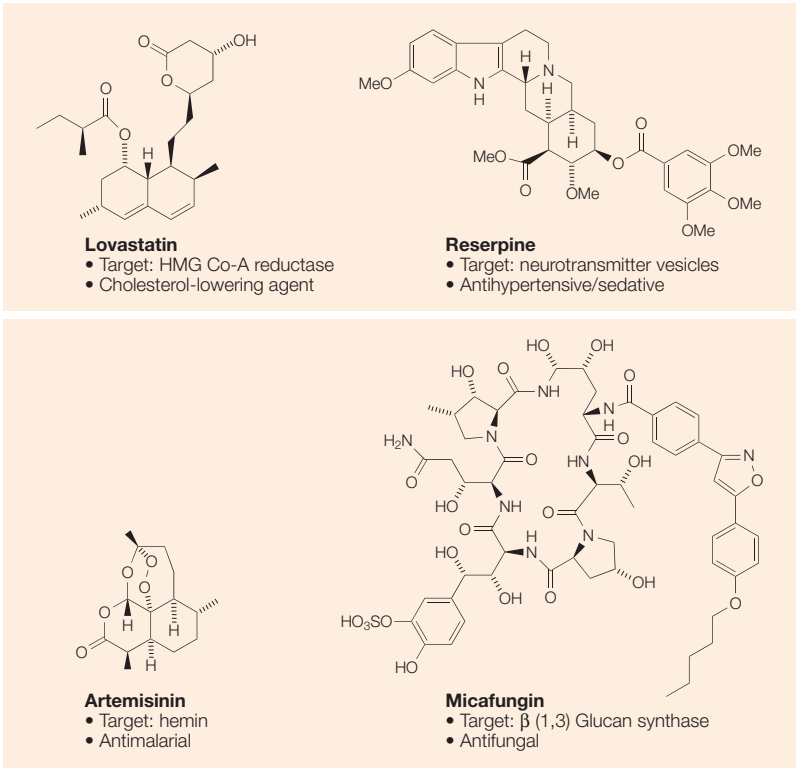


Figure 2 | **Examples of natural-product drugs.** Natural products have become effective drugs in a wide variety of therapeutic indications, as illustrated by the compounds shown, which modulate a diverse range of targets. HMG Co-A, 3-hydroxy-3-methylglutaryl coenzyme A.

Suitability of natural products for modulating biochemical reactions and protein–protein interactions. Contemporary drug discovery is based in large part on the screening of small molecules for their ability to bind or otherwise inhibit specific macromolecular (usually protein) targets. Given the virtually infinite number of small-molecule structures that can be generated, it is imperative that a molecular screening library covers a significant portion of chemical diversity space, but is also favourably biased toward 'biological friendliness' and 'drug-likeness'[13]. Some of the first large combinatorial libraries, in some instances containing in excess of one million compounds, were synthesized only to find disappointingly low hit rates, or no hits at all. These libraries were designed more on the basis of chemical accessibility and maximum achievable size than on biologically relevant chemical diversity or properties[14]. This is consistent with the concept that a fruitful area in which to search for drug leads is the area of diversity space in which the chemical scaffolds embody characteristics that promote binding to multiple target proteins — so-called privileged structures[15]. Natural products can be viewed as a population of privileged structures selected by evolutionary pressures to interact with a wide variety of proteins and other biological targets for specific purposes, a view supported by the fact that natural products have become effective drugs in a wide variety of therapeutic indications (FIG. 2).

Recent advances in genomics and structural biology during the past 5–10 years are painting a clearer picture of the diversity of proteins targeted by natural-product molecules. It has been shown that the number of unique protein architectures (or folds) in nature is much smaller than the number of protein families predicted by sequence similarity[16,17]. Structural motifs of domains, the functional building blocks of proteins, are frequently conserved even when there is a low degree of sequence similarity. These units are combined in modular fashion to fine-tune the function of full proteins[18,19]. This indicates that proteins populate the total FOLD SPACE in a highly non-uniform fashion, such that localized concentrations of receptive binding 'space' are distributed onto clusters of superfolds[20]. Furthermore, the same protein-binding target or fold can serve differing functional roles in a number of higher organisms[21]. On the basis of this concept, a guiding principle has emerged that natural products, by virtue of their molecular evolution to preferentially bind to these folds, are validated starting points for screening-library design[22]. This was the basis for work in the 1990s that effectively utilized molecular scaffolds based on yohimbine[23], paclitaxel[24] and vancomycin[25,26] for SOLID-PHASE SYNTHESIS of focused libraries. Recent investigations have successfully produced high-quality screening-type libraries by solid-phase synthesis based on privileged natural-product benzopyran scaffolds[27,28] and focused libraries by parallel synthesis to find selective inhibitors of receptor tyrosine kinases[29].

Aside from binding in the active sites of enzymes and acting as inhibitors of catalysis, the propensity of natural products to bind to a variety of protein domains and folding motifs is borne out by biological activity of another type — the capacity to modulate or inhibit protein–protein interactions. As a result, these molecules are effective modulators of cellular processes such as the immune response, signal transduction, mitosis and apoptosis. An enzyme inhibitor must bind to the active site at the interior of the protein, in a manner which involves relatively few interactions. By contrast, a small-molecule disruptor of a protein–protein interaction must bind to the protein surface, where tight binding involves a larger surface area with multiple points of interaction. By virtue of their increased size, and natural selective pressure to bind gene products, natural products are effective scaffolds for this function[30]. Moreover, similar natural-product scaffolds can be structurally fine-tuned by nature to modulate protein–protein interactions selectively, on the basis of a single starting protein interaction. Perhaps the best known example of this is the case of FK506[31], rapamycin[32] and ascomycin[33], a family of closely related polyketide natural products derived from soil actinomycetes that show profoundly different cellular effects via binding of the FK506-binding protein immunophilin proteins and modulation of the protein–protein interactions involved in the signal transduction pathways of T-cell activation and growth. All three of these compounds — FK506 (tacrolimus (Protopic; Fujisawa)), rapamycin (sirolimus (Rapamune; Wyeth)) and ascomycin

FOLD SPACE
The total repertoire of three-dimensional protein structures or architectures.

SOLID-PHASE SYNTHESIS
Synthesis of compounds on the solid surface of an insoluble resin support, which allows them to be readily separated (by filtration or centrifugation) from excess reagents, soluble reaction by-products or solvents.

(pimecrolimus (Elidel; Novartis)) — are now in clinical use for different therapeutic indications (FK506 and rapamycin for liver and kidney transplantation, respectively, and ascomycin for atopic dermatitis)[34]. Natural products of this class have also recently shown potential in the treatment of neurodegenerative disease[35].

A different theme is evident when we examine the large family of natural products that affect mitosis by modulating the protein–protein interactions involved in tubulin polymerization (FIG. 3). In this case, there is broad diversity in the structure and mechanism of tubulin-binding molecules found in plants, animals and microorganisms[36]. Agents such as the plant-derived taxane paclitaxel and the marine-sponge-derived discodermolide[37] bind to microtubules and enhance their polymerization. The vinca alkaloids, along with the macrocyclic polyether spongiastatin[38] and the recently described marine tripeptide hemiasterlin[39], bind to the tubulin dimer vinca domain and inhibit microtubule polymerization. Others, such as colchicine, podophyllotoxin, steganacin and combretastatin, disrupt microtubule assembly by binding to a site on soluble tubulin distinct from that of the vinca alkaloids[40]. At low concentrations, the actions of different agents can converge to a similar mechanism affecting tubulin dynamics[40].

Natural-product libraries and high-throughput screening. The exercise of identifying natural-product (and also synthetic) molecules as potential drug leads most often involves automated testing of large collections (libraries) of compounds for activity as inhibitors or activators of a specific biological target, such as a cell-surface receptor or enzyme. This process is commonly known as high-throughput screening (HTS). The natural-product library itself might be composed of samples that are either themselves mixtures — such as crude extracts (10–100s of components), semi-pure mixtures (5–10 compounds) or, alternatively, single purified natural products. In the case of pure libraries, the hit-detection process is the same as that for synthetic pure libraries. In the first case, however, heterogeneity of the natural-product library samples adds two additional levels of complexity to the screening process. The first of these is that once a response for the sample is detected by HTS, one or more rounds of chemical purification and biological assay might be necessary for identifying and isolating the active component in the mixture. This is described in greater detail later in this article.

The second additional hurdle is that the complexity of crude or semi-pure natural-product libraries, and the chemical nature of many of the components found in them, often challenges the robustness of HTS technology. Before the advent of biotechnology, the difficulty of obtaining purified protein targets directly from tissues necessitated that much of the compound screening was performed using cellular *in vitro* or even whole-animal systems. When applied to crude or otherwise partially purified natural-product libraries, these platforms have a relatively good capacity to detect active components. Examples can be found in the antimicrobial *in vitro* and *in vivo* assays used in the 1970s and 1980s, which effectively served to detect many novel antibiotics. Advances in genomics and molecular biology have now made it possible to obtain relatively large quantities of protein targets for high-throughput, cell-free assay systems for directly detecting the catalytic inhibition or target binding[41]. These developments have greatly expanded the number of targets amenable to HTS, but at the same time have introduced technical complexities into the screening of natural-product libraries. These complications reduced the success rates for lead discovery from natural-product libraries in the early and mid-1990s, but technical improvements during the past few years have circumvented the problems in many HTS platforms.

An illustrative example is provided by the case of protein kinases, enzymes that catalyse the phosphorylation by ATP of protein tyrosine or serine/threonine residues, and which have a central role in cellular signalling pathways that control the activation, growth and proliferation of cells. As drug targets, protein kinases are superseded in importance only by the G-protein-coupled receptors[42]. Using expressed and purified kinases, numerous HTS technologies have been developed for detecting specific kinase substrate (ATP) inhibitors. Initially, assays were based on measuring the transfer of radioactive phosphate (^{32}P or ^{33}P) from ATP to a protein or synthetic peptide substrate — a robust, but low-throughput, technique. Newer HTS kinase assay platforms are based on fluorescence intensity, time-resolved fluorescence or fluorescence polarization[43]. Early complications in using these assays for screening natural-product libraries came from two sources. First, for detecting competitive substrate inhibitors, it is best to screen at a compound

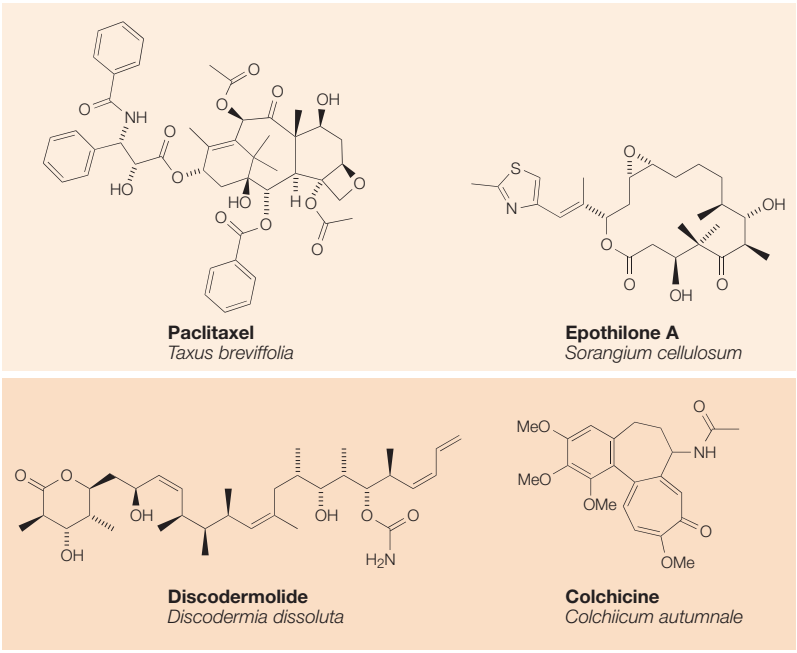


Figure 3 | **Structural themes in natural products that target tubulin protein–protein interactions.** Microtubules, dynamic assemblies of the protein tubulin, are essential components of cells involved in mitosis. The natural products shown here inhibit cellular proliferation by disrupting tubulin polymerization and de-polymerization by different binding mechanisms.

concentration close to the K_m of the enzyme for ATP, typically micromolar. This poses a challenge with mixture natural-product libraries in which the relative concentrations of individual components in a sample are not known precisely, and might vary by two orders of magnitude or more. Consequently, at any given assay dose, the concentration levels of trace components might not be high enough for detection, whereas the highly abundant components can exhibit inhibition as a result of nonspecific binding, perturbation of the assay pH or other physical properties. Second, natural-product samples can contain compounds that either fluoresce or absorb at excitation or emission wavelengths of the fluorophore (typically fluorescein), or by light scattering of insoluble components, which therefore affect the readout of the assay[44]. This liability is shared with

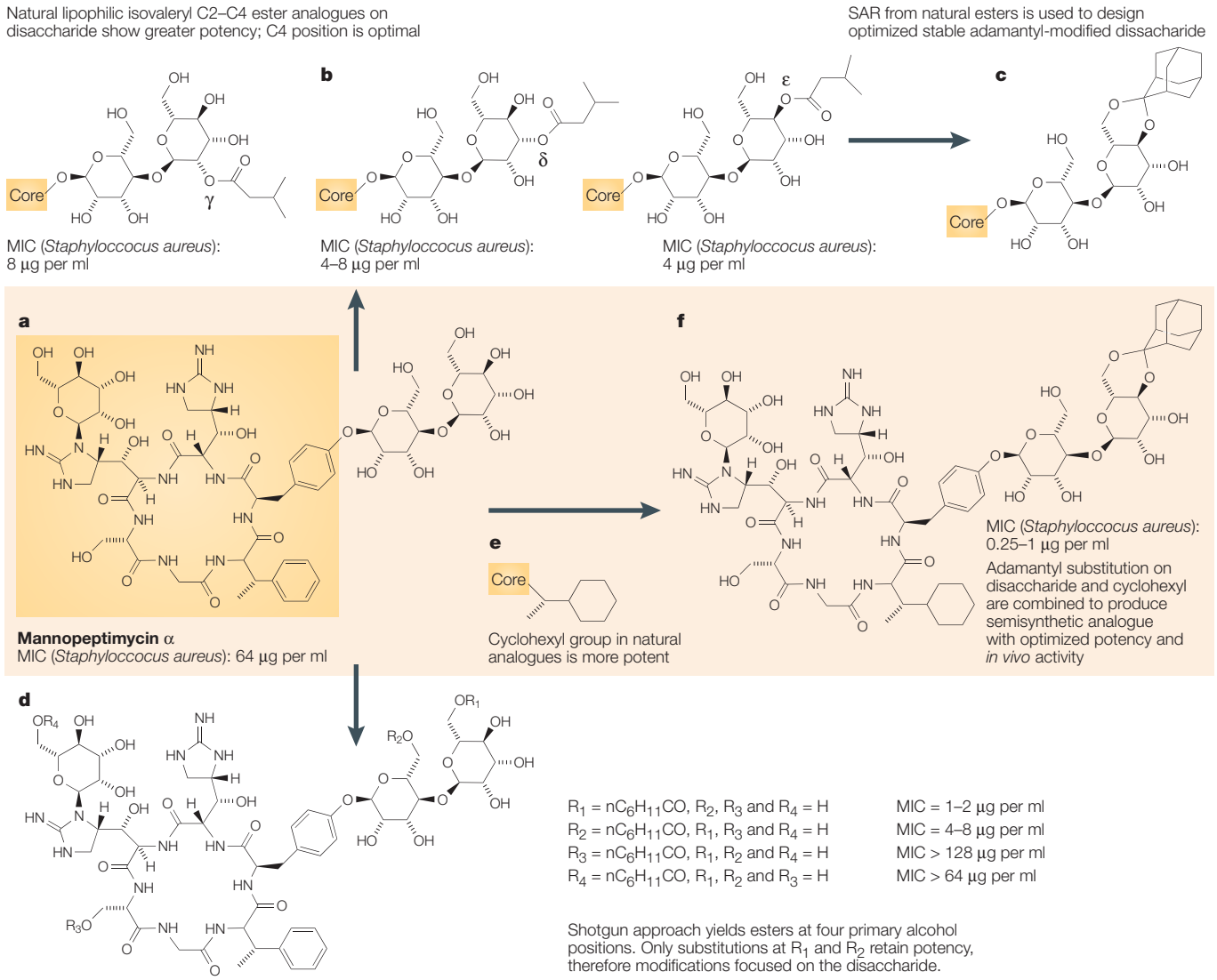


Figure 4 | **Using natural-product structure–activity relationships and shotgun synthesis to optimize biological activity: the example of mannopeptimycins.** The mannopeptimycins were originally isolated in the late 1950s as a complex of novel antibiotic glycopeptides that show activity predominantly against Gram-positive bacteria[51,52]. However, the structural complexity, along with the lack of broad-spectrum activity, reduced the urgency to further develop the antibiotic. The need to find new antibiotics with specific activity against resistant strains, and advances in methods for purifying and characterizing the components of natural product complexes led to re-examination of the mannopeptimycins in the 1990s. The mannopeptimycin complex consists of five separate members, designated mannopeptimycin α–ε. The molecules consist of a cyclic hexapeptide core that contains two stereoisomers of the amino acid, α-amino-β-[4′-(2′-iminoimidazolidinyl)]-β-hydroxypropionic acid, which is glycosylated with mannose mono and disaccharide moieties that in some cases bear isovalerate esters. **a** | The major component of the complex, mannopeptimycin α, is not esterified and has poor antibacterial activity. Initial structure–activity relationship studies, shown in the figure, focused primarily on modifications to the disaccharide. Natural isovaleryl ester regio-isomers provided the initial direction for SAR experiments, leading to **c**, a stable disaccharidic modification. **d** | Random (shotgun) esterification of primary alcohols quickly yields SAR information about disaccharide positions necessary for potency. **e** | Cyclohexyl group in structural analogues were found to be more potent. **f** | Combining optimal modifications to the core and disaccharide leads to a highly potent, optimized semi-synthetic analogue. MIC, minimum inhibitory concentration; SAR, structure–activity relationship.

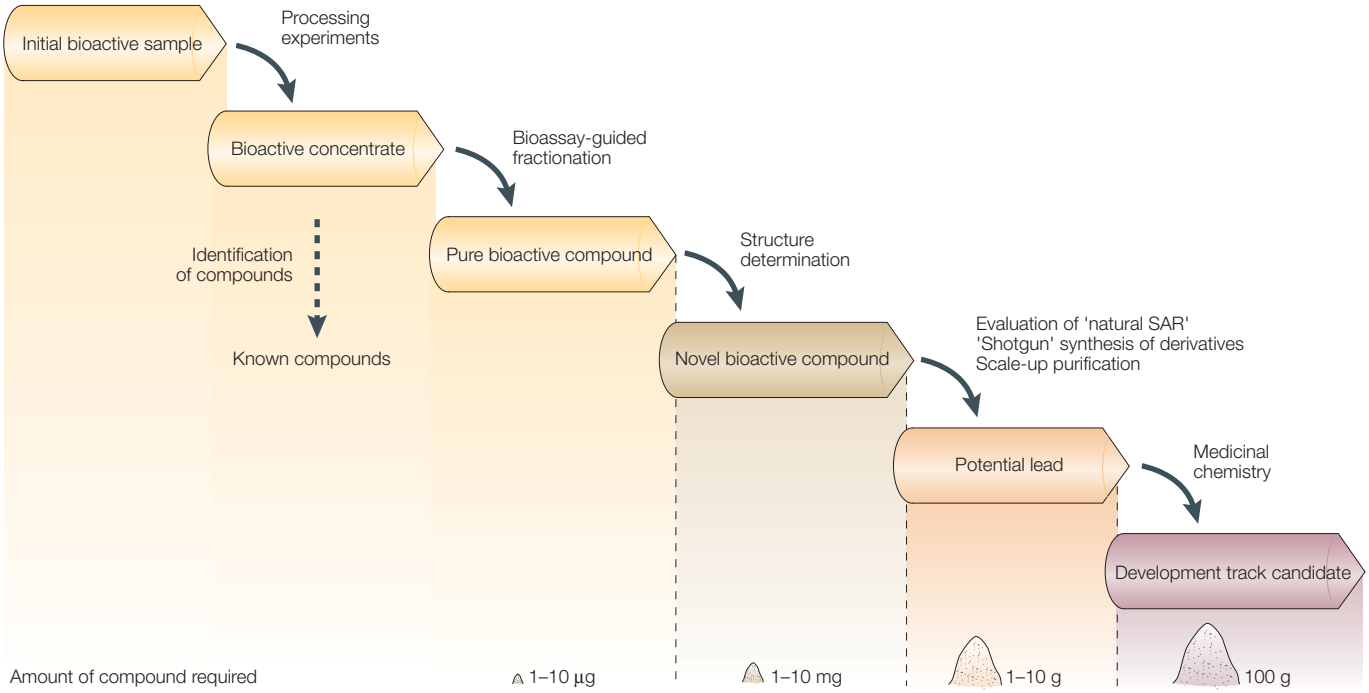


Figure 5 | **Chemical process for natural product discovery.** The natural product is extracted from the source, concentrated, fractionated and purified yielding essentially a single biologically active compound. Identification of known compounds, thereby avoiding replication of previous efforts, has been greatly aided by directly coupled HPLC-mass spectrometer (LC-MS) systems and natural-product databases[57]. *De novo* structure determination of compounds that are novel has been revolutionized by advances in spectroscopic techniques, particularly in high-resolution nuclear magnetic resonance technologies. Although the determination of complex structures is technically challenging, it is no longer a major impasse in the drug discovery process. In those cases in which the biological activity profile meets criteria for potency and selectivity, preliminary structure–activity relationship (SAR) studies are conducted and the purification process is scaled up. Once the feasibility of modulating biological response through synthetic modification is established, the hit is declared a lead and proceeds onward for additional optimization by traditional medicinal chemistry.

synthetic screening libraries, but the issue is exacerbated with natural products because the presence of these interference compounds might not be fully characterized. However, it has now been shown that in kinase, protease and phosphatase assays, fluorescent-compound and light-scattering interferences can be overcome by increasing the fluorophore concentration in the assay, by using red-shifted wavelength dyes[45] or by the newer technique of LIFETIME DISCRIMINATED POLARIZATION[44].

An effective strategy that helps to alleviate this interference, as well as shorten the time needed to eventually isolate the active principle, is the implementation of purified or fractionated samples for screening from the original crude extract. There are many variations of this theme and each offers advantages of expediency or purity[46,47]. The samples produced by this fractionation approach are less complex mixtures for use in screening. The result of this procedure is that the final resolution of the active components is simplified, typically only requiring one additional purification step. Of potentially greater significance is that interferences are reduced by the simplification of the mixtures and the relative concentration of minor components is increased, thereby enhancing the opportunity to uncover novel biologically active metabolites. Significant successes have also been achieved in the HTS analysis of natural-product libraries using mass spectrometry — for example, electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (ESI-FTICR) screening against RNA targets[48]. These and other developments make it likely that future lead generation through the screening of natural products need not be limited by HTS technology.

Structure–activity and structure–property relationships of natural products. One of the distinguishing features (or perhaps historical drawbacks) of natural products is their frequent occurrence as suites or complexes of structurally related analogues. The biological significance of this expressed molecular diversity is unclear, particularly when the suite contains members (often major) that seem to lack biological activity or function. Essentially, why would an organism expend the resources needed to synthesize many analogues of a molecule for which there is but a single purpose? One possible answer to this question is furnished by the 'screening hypothesis', which is based on the proposition that biological activity is a rare property for any molecule to possess, and that there might be a selective advantage to an organism if it can generate and retain chemical diversity at a low cost[49]. The existence of congeners of a natural product series might therefore be

LIFETIME DISCRIMINATED POLARIZATION
A method of reducing test-compound interference in fluorescence-based screening by rejection of signals from short-lifetime sources.

Box 1 | **Nuclear magnetic resonance spectroscopy experiments**

Nuclear magnetic resonance (NMR) spectroscopy experiments that are used to elucidate the molecular structures of natural products are based on the application of pulse sequences — precisely timed radio-frequency and magnetic-field gradient pulses (usually on the microsecond and millisecond timescale) that are designed to excite the atomic nuclei of molecules and thereby produce diagnostic signals that can be analysed to determine the connectivity of the ^{1}H, ^{13}C and ^{15}N nuclei of the molecule. More than 1,000 of these pulse sequences have been developed, each designed to ascertain a different type of physical information about the molecule under study. Due to the many types and variations of pulse sequences, a system of acronyms has evolved in the field of NMR. Listed here are a few representatives of the commonly used pulse-sequence acronyms a reader might encounter in the literature of natural-product structure determination.

- COSY (COrrelated SpectroscopY) and TOCSY (TOtal Correlation SpectroscopY)
  Provides molecular connectivity information based on proton–proton interactions through covalent bonds and results in connectivity maps of proton-bearing carbon atoms.
- NOESY (Nuclear Overhauser Enhancement SpectroscopY)
  Provides molecular connectivity information based on proton–proton through-space interactions. Also provides information about proton inter-atomic distances and molecular conformation.
- HSQC (Heteronuclear Single Quantum Coherence Spectroscopy)
  Provides connectivity information relating specific carbon atoms and the protons bound to them.
- HMBC (Heteronuclear Multiple Bond Correlation Spectroscopy)
  Provides long-range connectivity between protons and carbon atoms separated by two to four covalent bonds.
- INADEQUATE (Incredible Natural Abundance DoublE QUAntum Transfer Experiment)
  Provides direct connectivity maps of covalently bound carbon atoms.

the consequence of an organism's need to generate its own chemical diversity to optimize the activity of its secondary metabolites, essentially doing its own structure–activity relationship (SAR) optimization.

In any event, due to limitations in separation technology, and methods for structure determination and characterization, the occurrence of natural products as a complex has historically presented a challenge to their development as drugs. Indeed, important natural-product therapeutics such as ivermectin[50] were developed and marketed in the past as complexes because it was not possible to purify the individual components at a sufficient scale. However, recent technological advances have allowed natural-product investigators to unravel theses complexes, and exploit the chemical diversity presented by them to obtain analogues with improved potency and drug-like properties. An illustration of this principle is found in the mannopeptimycin family of antibiotics[51] (FIG. 4). The complex was originally isolated in the late 1950s, when it was shown to have potent activity predominantly against Gram-positive bacteria[52]. At that time the chemical complexity of the complex, along with the lack of broad-spectrum activity (at that time a drawback for antibiotics), reduced the urgency to further develop the antibiotic.

Re-examination of the complex in the late 1990s was prompted by the need to find new antibiotics with specific activity against resistant strains of Gram-negative organisms, and the availability of new advances in high-performance liquid chromatography (HPLC), nuclear magnetic resonance (NMR) and mass spectrometry with which to purify and characterize the components. The mannopeptimycins are a complex of novel antibiotic glycopeptides produced by the *Streptomyces hygroscopicus* strain LL-AC98, and have activity against methicillin-resistant staphylococci and vancomycin-resistant enterococci. The mannopeptimycins inhibit cell-wall biosynthesis through a unique mode of action, by binding to the membrane-bound cell-wall precursor lipid II (C_{35}-MurNAc-peptide-GlcNAc) in a different manner to other lipid II binders[53]. They also bind lipoteichoic acid in a way that serves to concentrate the antibiotic on the cell surface.

The mannopeptimycin complex as described[51] consists of five separate members, designated mannopeptimycin α–ε. The molecules consist of a cyclic hexapeptide core that contains two stereoisomers of an unprecedented amino acid, α-amino-β-[4′-(2′-iminoimidazolidinyl)]-β-hydroxypropionic acid (Aiha), which is glycosylated with mannose mono- and disaccharide moieties that in some cases bear isovalerate esters. The major component of the complex, mannopeptimycin α, is not esterified and has poor antibacterial activity. Strikingly, the presence and location of the isovaleryl group is crucial for the retention of antibacterial activity. This limited but important initial SAR was instrumental in designing a synthesis programme that furnished analogues with improved antibacterial potency and a favourable therapeutic window[54–56] (FIG. 4).

Technologies for natural-product lead discovery

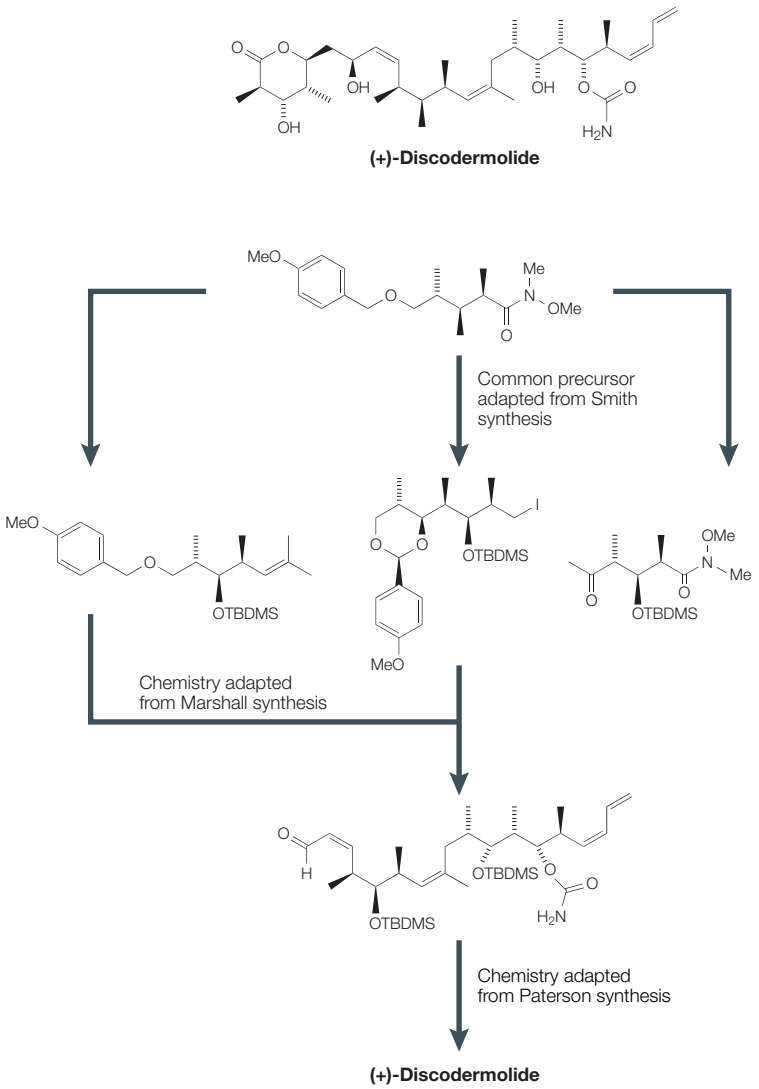
The typical process of discovering natural-product hits, and their progression towards development, is depicted in FIG. 5. In this generic scheme, the natural product is extracted from the source, concentrated, fractionated and purified, yielding essentially a single biologically active compound. Historically, this process has suffered most often from three major hurdles. The first is the rapid identification of known compounds (dereplication), to avoid the duplication of effort. This step has been greatly facilitated by the availability of reliable directly coupled HPLC-mass spectrometer (LC-MS) systems, and the general availability of natural-product databases[57]. The pivotal development responsible for the success of LC-MS has been the introduction of efficient and general methods for producing ions from the effluent of HPLC separations. The most general of these methods, known as electrospray ionization (ESI) and atmospheric pressure ionization (API), can generate the ions essential for mass spectrometric analysis for greater than 90% of analytes, ranging from amino acids to proteins and nucleic acids. The correlation of both molecular mass and UV absorption data with known compounds by database searching is ordinarily sufficient to classify sets of compounds[58].

The second major hurdle in the process — the *de novo* structure determination of compounds that are NMEs — is an area that has been revolutionized by many advances in spectroscopic techniques, particularly in high-resolution NMR technologies. Of the

many NMR advances, those of particular importance to natural-product structure determination fall into one of two main areas: multidimensional pulse methods and sensitivity improvements. From its inception, and particularly since the advent of two-dimensional NMR methods[59], high-resolution NMR spectroscopy has seen continuous development and expansion of the array of experimental methods available to elucidate chemical structures[60]. New experiments, particularly multidimensional ones, provide scalar (through bond) ^{1}H-^{1}H and ^{1}H-[^{13}C, ^{15}N, ^{31}P] correlations and ^{1}H-^{1}H dipolar (through space) molecular connectivity data that essentially map out the structure of the compound (BOX 1). In the area of sensitivity, stronger magnetic fields provided by superconducting magnets[61], cryogenic electronics[62] and micro-probe technologies[63] have dramatically lowered the amount of material needed for structural analysis, to less than a milligram. The combination of cryogenic probe electronics with correlation spectroscopy enables the development of

Box 2 | Synthesis of natural products to overcome supply problems: discodermolide

(+)-Discodermolide, an antitumour polyketide from the Caribbean sponge *Discodermia dissoluta*, was first isolated and characterized in 1990 by Gunasekera and co-workers on the basis of its immunosuppressive and cytotoxic properties[68]. Subsequent studies showed that discodermolide possessed potent antitumour activity as a result of induction of tubulin polymerization and microtubule stabilization in a manner similar to taxol, but with tenfold greater potency[102]. Moreover, discodermolide is highly potent against P-glycoprotein (Pgp)-expressing multiple-drug resistant (MDR) tumour cells[37], and has improved aqueous solubility compared with existing agents. These attributes qualified the compound as a strong candidate for development as an anticancer agent. However, the low isolation yield of (+)-discodermolide (~14 mg per kg), combined with the limited natural supply of the producing organism, made it necessary to pursue chemical synthesis as a means to supply the compound for further development.



(+)-Discodermolide represents a daunting target for large-scale chemical synthesis, by virtue of its 13 stereogenic centres and high degree of chemically sensitive functionality. Following the initial report of its isolation, the compound quickly became a target for several academic groups, and several successful syntheses were reported. The first was achieved by Schreiber *et al.*, who synthesized both enantiomers[103]. Schreiber's work enabled the assignment of the absolute stereochemistry and provided valuable structure–activity relationship (SAR) data. Later syntheses by Smith[104], Paterson[105] and others succeeded in improving efficiency, yield and economy. Finally, by blending the common precursor methodologies developed in the gram-scale synthesis reported by Smith[104] with chemistry developed in syntheses of the Paterson[105] and Marshall groups[106], the team at Novartis succeeded in synthesizing 60 g of (+)-discodermolide[107–111]. The remarkable story of (+)-discodermolide — from the initial isolation of a novel, complex and rare natural product, to the successful production of a clinical drug substance — is testimony to the power of combining modern natural-product, synthetic and process chemistry to overcome supply problems.

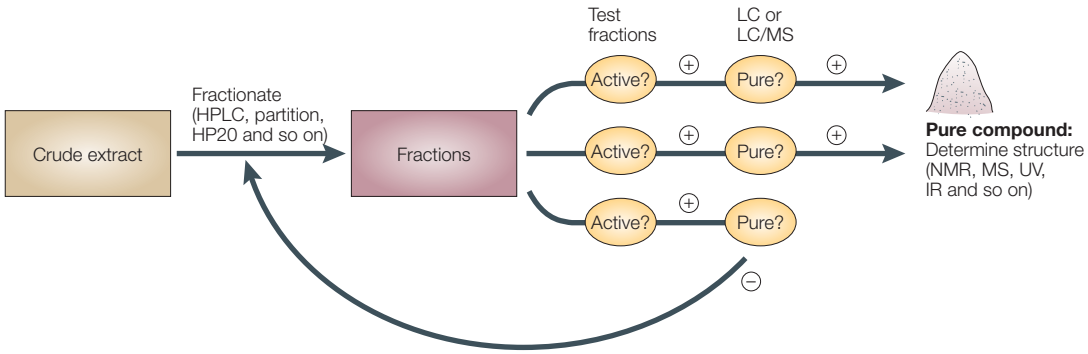


Figure 6 | **Generic scheme for bioassay-guided fractionation.** Several cycles of fractionation are usually needed to obtain a pure compound. Multiple active components can illuminate structure–activity relationships. HP20, a solid-phase adsorber; HPLC, high-performance liquid chromatography; IR, infrared spectroscopy; LC/MS, liquid chromatography/mass spectrometry; NMR, nuclear magnetic resonance; UV, ultraviolet.

still more powerful experiments, such as correlation experiments for low-abundance ^{13}C and ^{15}N nuclei[64], which are unattainable with conventional hardware. Today, highly complex molecular structures, such as those of maitotoxin[65] or lomaiviticin[66], are largely determined with these NMR techniques and knowledge of the new compound's molecular formula.

Determination of the molecular formula is crucial to the process and is typically done by high-resolution mass spectrometry on microgram quantities of material. One of the most powerful of these techniques is Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR/MS), which is capable of measuring molecular mass with exceptional accuracy[67]. Combining the tools of high-resolution mass spectrometry with two-dimensional NMR spectroscopy allows structure determination to be carried out on sub-milligram or milligram amounts of a compound in a matter of hours or days, rather than weeks or months. Although the determination of complex structures is technically challenging, it is no longer a major impasse in the drug discovery process.

In those cases in which the biological activity profile meets criteria for potency and selectivity, preliminary SAR studies are conducted and the process is scaled up. Nature often provides the first clues to SAR in the form of biosynthetic congeners, as seen in the case of the mannopeptimycins (described in the previous section (FIG. 4)). A second avenue for exploring SAR in an expeditious manner is the 'shotgun' approach to chemical derivatization. Again using the example of mannopeptimycins, key functional groups required for antibiotic activity were identified by allowing the parent compound to react with nonspecific derivatizing agents such as anhydrides. In this approach, all of the 14 primary and secondary hydroxyl groups might be expected to form ester derivatives. By restricting the stoichiometry of the reagents and monitoring the reactions by HPLC, conditions were found in which a mixture of mainly mono-esterified components was produced[54]. The various isomers were subjected to preparative HPLC purification and the regio-chemistry of the newly formed ester groups was defined by two-dimensional NMR. Bioassay of the various products indicated that only those esters formed on the terminal mannose of the disaccharide retained potency (FIG. 4). The knowledge gained through understanding the natural SAR and the shotgun approach provided an early foundation on which the overall synthetic strategy was developed. Once the feasibility of modulating biological response through synthetic modification is established, the hit is declared a lead and proceeds onwards for additional optimization by traditional medicinal chemistry.

The major bottleneck that continues to affect natural product drug discovery is the isolation and purification of the active principles from an exceptionally complex matrix. Often the target compounds represent much less than 1% by weight of the crude extract, and the approach remains highly experimental. Although advances in separation technology, such as HPLC, supercritical fluid chromatography (SFC) and capillary electrophoresis (CE) have had a major impact on resolving power, often the purification step in the process is rate-limiting. The challenge is twofold: first, one must correlate the biological signal of interest with the effector compounds; second, one must then devise preparative separation methods to yield sufficient quantities of the pure material. This latter point impinges on one of the fundamental concerns in natural-product drug discovery, often referred to as the 'supply issue'. Unlike simpler synthetic compounds, natural products can be limited in supply owing to sourcing limitations or the impracticality of synthesis. As indicated at the bottom of FIG. 5, material requirements rapidly escalate as one approaches the medicinal chemistry and development phases. The supply issue is particularly crucial for source organisms, such as marine invertebrates (for example, the discodermolide producer[68]), that have not been productively cultured in the laboratory, but also arise with plant products, such as paclitaxel. Microbial products, as well as some plant-derived agents, are amenable to culturing on production scale, and synthetic methodologies continue to be developed for the large-scale synthesis of highly complex products. Perhaps the most compelling recent examples are those of discodermolide, described in BOX 2, and E7389, an

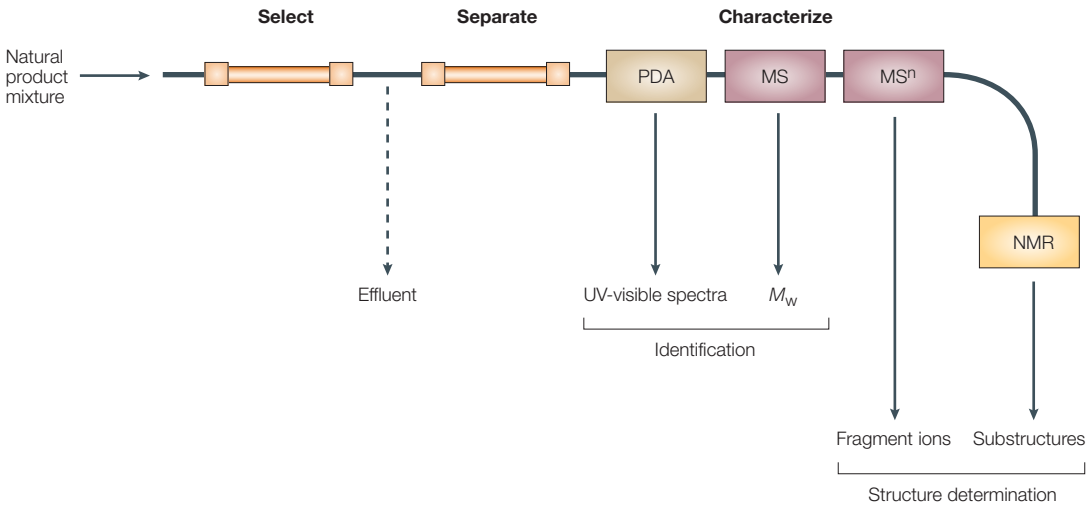


Figure 7 | **Affinity-based identification system for natural products.** Affinity selection is performed on the first column of this high-performance liquid chromatography-based system. Selected compounds are resolved into single components on the second column, which are then interrogated by a series of detectors. A combination of UV spectra and molecular mass leads to the identification of known compounds, whereas more extensive structural information is obtained for unknowns using tandem MS and NMR. MS, mass spectrometry; NMR, nuclear magnetic resonance; PDA, photodiode array; UV, ultraviolet.

analogue of halichondrin B[69,70]. Both of these highly complex molecules have been synthesized in multigram quantities sufficient for clinical trials. Successes such as these make the compelling case that virtually no crucially important compound is beyond reach for clinical evaluation.

The general paradigm for bioassay-guided purification is shown in FIG. 6. As can be seen in the diagram, progression depends on how many 'cycles' of fractionation and bioassay are required. In those cases for which the bioassay turnaround time is lengthy, the time for a single round of fractionation can be prolonged. In the early stages of such discovery research, time is crucial if NMEs are to progress into the next phase, and delays owing to bioassay turnaround can be a practical limitation.

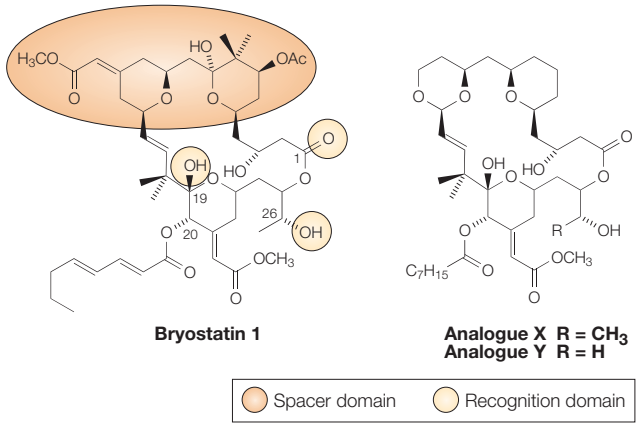
For this reason, a number of innovative approaches have been advanced to create 'on-line' bioassays that utilize affinities of the desired natural ligands with targets of interest. One such approach uses continuous-flow enzymatic reactions that are capable of providing real-time read outs of inhibition of enzymatic activity[71]. In this system, the effluent from HPLC is split in two streams, one to the enzyme assay and the other in parallel to a mass spectrometer. Correlating peaks detected in the enzyme assay with the corresponding mass spectra provides data that are characteristic of the bioactive compound. This correlation can be sufficient to identify known compounds by database searching on the basis of molecular mass, and will at least provide 'linkage' between the biological test and a physicochemical parameter. Such linkage allows the physicochemical characteristics (mass spectrometric data, HPLC retention time, UV absorbance and so on) to be substituted for bioassay data for subsequent cycles of fractionation.

Similar correlations have been determined for compounds that bind to structured RNA targets that are amenable to FT-ICR/MS analysis. Using a method dubbed multitarget affinity/specificity screening (MASS), Hofstadler was able to detect specific binding of the aminoglycoside antibiotic paromomycin in chromatographic fractions derived from cultures of *Streptomyces rimosus* subspecies *paromomycinus*[72]. Non-covalent adducts of paromomycins with a synthetic RNA oligomer containing the *Escherichia coli* A-site (site of action of aminoglycoside antibiotics) were observed in the mass spectrometer. The molecular masses of the aminoglycosides were determined by the differences in mass between the adducts and the free RNA. By analogy, molecular masses of unknown binders present in extracts of other antibiotic producers could be determined in the same manner.

Frontal affinity chromatography (FAC)[73,74] combined with mass spectrometry has also been used to simplify the deconvolution of activities in natural-product extracts[75]. In FAC, the target is immobilized on a column and the mixture is continuously infused through the system. The compounds with the greatest affinity for the target will have the longest 'breakthrough' times. Mass-spectrometric monitoring of the effluent provides the characterization of the retained compounds, as they sequentially elute in inverse order of their binding affinity.

An idealized system that couples on-line affinity selection, separation and identification/structure determination is depicted in FIG. 7. All of the components of this system are connected in a flow path that is driven by an HPLC pump. The selected compounds are resolved in the separation step, which is shown in FIG. 7 as an HPLC column that yields single components or 'peaks', which are then subjected to UV-visible absorption spectroscopy, mass spectrometry and NMR

Box 3 | **Improving natural products: the example of bryostatin**



In 1988, Wender *et al.* proposed a simplified model of bryostatin that accounted for its potent inhibition of the binding of phorbol esters to protein kinase C, as well as the natural ligand 1,2-diacylglycerol. Although a direct correlation of this interaction has not yet been linked to antitumour activity, this 'pharmacophoric model' was used to probe the essential structural features required for chemotherapeutic efficacy. In the ensuing 15 years, several simplified analogues were produced on the basis of the model that showed enhanced potency. Two regions have been defined: one, referred to as the recognition domain, consists of a particular spatial arrangement of oxygen atoms at C1, C19 and C26. The other domain is the spacer domain that holds the three oxygen atoms in the proper orientation. Wender's primary focus has been on simplifying the latter piece; the first successful candidates have simplified A and B rings, as shown in both X and Y. In the next phase of the simplification process, the fatty acyl residue at C20 was investigated, and it was found that a simple saturated octanoate residue compared favourably, and finally simply shortening the carbon chain by one unit at C26 converting the secondary alcohol function common to Bryostatin 1, and X, to a primary alcohol in Y, provided enhanced potency over the natural product.

spectroscopy for characterization. In theory, one pass of a particular extract though such a system would provide identification of the compounds responsible for the binding activity, including compounds of unknown structure that would be determined by on-line MS/MS and NMR spectroscopy[76,77]. In practice, it is usually more expedient to conduct various aspects of this process independently. This is particularly true for complete structure determination by NMR spectroscopy, in which the availability of greater amounts of material (milligrams versus micrograms) will greatly reduce data-acquisition times.

Advances in NMR, MS and HPLC technology have made hyphenated LC-NMR, and now LC-NMR-MS, practical options for analysis and structure determination of complex natural-product mixtures, in many instances circumventing the need for the traditional approach of mixture fractionation and isolation of individual components before structure elucidation by MS, and NMR. Although LC-NMR was introduced almost 20 years ago[78], its application to the direct analysis of natural products has been hampered by problems of chromatographic peak diffusion, sensitivity and suppression of unwanted background solvent signals[79]. The development of peak storage units served to eliminate peak diffusion problems encountered with earlier stopped-flow or on-flow methods, and improved background solvent suppression using pulsed magnetic-field gradients led to the first effective applications of LC-NMR to structure analysis of plant natural products without the need for chromatographic isolation[80,81]. Direct coupling of electrospray mass spectrometry for LC-NMR-MS has proven to be an effective combination for the characterization of plant glycosides that are resistant to analysis by LC-NMR alone[82]. Finally, the inherent low sensitivity of NMR (approximately 1,000-fold less sensitive than mass spectrometry) has been relieved to a great degree by the development of cryogenic flow probes[83] and micro-coil (nanolitre volume) NMR probes[84]. These, in conjunction with the introduction of on-line solid-phase extraction peak trapping, have greatly improved the sensitivity and flexibility range of the technique. Using LC-UV-SPE-NMR-MS, it is now possible to perform automated analysis of natural-product extracts in which the individual components are present in 10–50 μg quantities[77].

Continuing developments in biosynthetic technologies offer great promise for the discovery and development of natural-product-derived pharmaceuticals. Genetic methods for the creation and expression of novel metabolites are now routinely used in drug discovery programmes. The development of metabolic engineering and its potential applications to drug discovery has recently been reviewed by Khosla and Keasling[85], and will not be further addressed in this article.

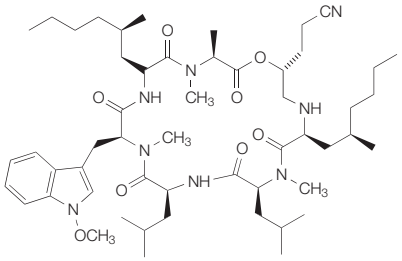
Leveraging the properties of natural products

A strategy that has been successfully used in recent years follows from experiments aimed at total synthesis of natural products. Many academic groups consider natural products ideal targets for testing their synthetic methodology, and many remarkable achievements have been documented[86,87]. In the course of these heroic efforts it is often possible to define the crucial structural elements required for biological activity. In this way, potent and selective products can be derived with fewer synthetic steps and at reasonable cost. A recent example that actually provided a simpler product with enhanced potency is provided by the work of Wender (BOX 3) on the bryostatin series of anticancer compounds[88–91].

Box 4 | **Methods for exploiting natural-product structures**

Systematic chemical mutagenesis

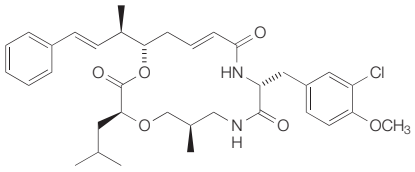
Systematic variation of the individual amino-acid residues (side chains) allows the pinpointing of structural features essential for biological activity. An elegant application of a combination of such processes was used to probe the crucial features of HUN-7293, a naturally occurring cyclic depsipeptide that is a potent and selective inhibitor of cell-adhesion molecule expression. In an approach referred to as 'systematic chemical mutagenesis' (see figure), Boger and co-workers explored the effect on biological activity (both potency and selectivity) of simplifying each of the seven residues of HUN-7293, including removal of *N*-methyl groups[93]. These experiments provided a greater understanding of the structural requirements for maintaining specific biological activity.



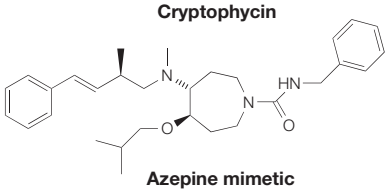
HUN-7293

Synthetic mimetics

Understanding the binding interactions of the natural product and the target can lead to a model for synthetic mimetics. An example of this approach to mimetics of the cryptophycin antitumour agents utilized an azepine scaffold to which residues were attached that resembled the overall geometry in the natural product[94]. A synthetic strategy was developed that allowed compounds such as the one shown in the figure to be prepared in reasonable overall yield. In this process, the stereochemical arrangement of the side chains and side-chain composition were studied to optimize the biological response.



Cryptophycin



Azepine mimetic

Peptides are modular structures joined together by amide (peptide) bonds. Small peptides, which often contain non-protein amino acids, are typically assembled by multi-enzyme complexes referred to as non-ribosomal peptide synthases, and in many instances these compounds are cyclized. These peptides, which represent a major class of biologically active natural products, are particularly amenable to parallel synthetic methodologies owing to the repetitive nature of the bond-forming process. Systematic variation of the individual amino-acid residues (side chains) therefore allows the pinpointing of structural features essential for biological activity. An elegant application of a combination of such processes was used to probe the crucial features of HUN-7293, a naturally occurring cyclic depsipeptide that is a potent and selective inhibitor of cell-adhesion molecule expression[92]. In an approach referred to as systematic chemical mutagenesis (BOX 4), Boger and co-workers explored the effect on biological activity (both potency and selectivity) of simplifying each of the seven residues of HUN-7293, including removal of *N*-methyl groups[93]. These experiments provided a greater understanding of the structural requirements for maintaining specific biological activity.

In some cases, the secondary metabolite has been shown to have exquisite potency for a particular target, but is not practical for use as a therapeutic agent, owing to various liabilities — for example, the cost of goods (a supply issue), metabolic liabilities, pharmaceutical properties and so on. In these instances, understanding the binding interactions of the natural product and the target can lead to a model for synthetic mimetics. An example of this approach to mimetics of the cryptophycin antitumour agents started with an azepine scaffold to which residues were attached that resembled the overall geometry in the natural product[94]. A synthetic strategy was developed that allowed the azepine mimetic compound shown in BOX 4 to be prepared, and with a reasonable overall yield. In this process, the stereochemical arrangement of the side chains and side-chain composition were studied to optimize the biological response.

An elegant example of overcoming the lack of selectivity observed with certain highly potent cytotoxic agents in the treatment of cancer is through the mechanism of monoclonal antibody delivery through conjugate formation. Calicheamicin (FIG. 8) is a highly potent enediyne DNA-damaging agent produced by *Micromonospora echinospora*[95]. Although it is 100–1,000 times more potent than conventionally used therapeutic agents, calicheamicin itself lacks the therapeutic index required for systemic administration. The potency of calicheamicin has been harnessed by immunoconjugation — a chemical modification followed by linkage to a monoclonal antibody that specifically binds to the CD33 cell-surface antigen present on acute myeloid leukaemia cells. After internalization into lysosomal vesicles, gemtuzumab ozogamicin (Mylotarg; Wyeth) is engineered to release calicheamicin, which migrates to the nucleus, cleaves DNA and results in cell death[96,97].

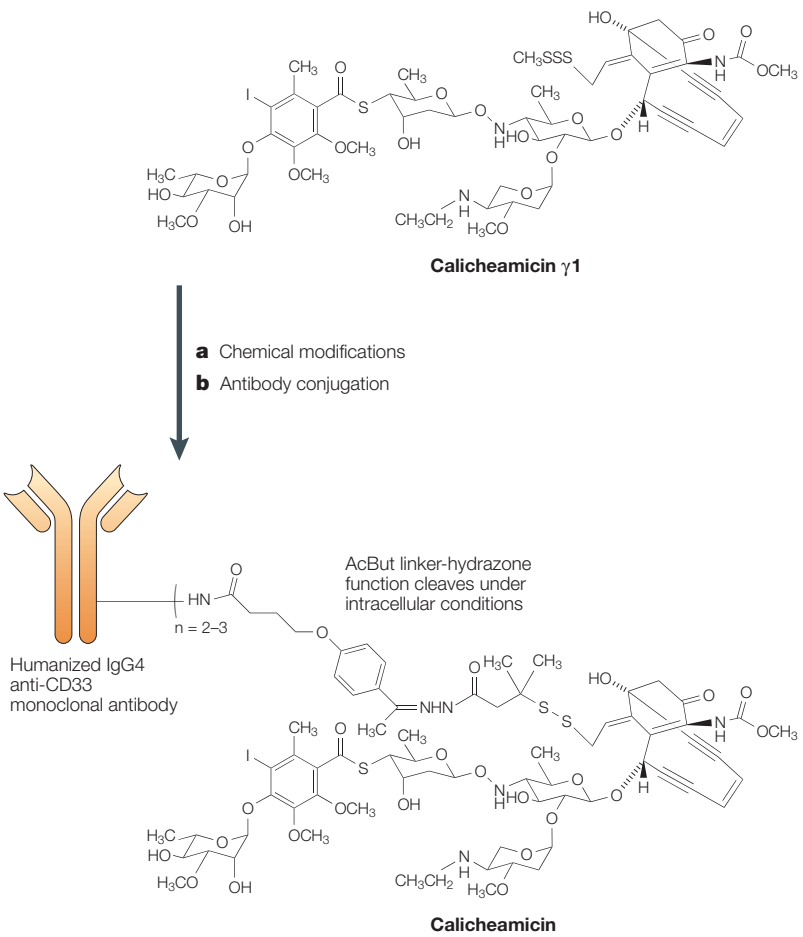


Figure 8 | **Harnessing highly toxic natural products for cancer therapy.** One way to overcome the lack of selectivity observed with certain highly potent cytotoxic natural products with potential for the treatment of cancer is to use monoclonal antibody delivery through conjugate formation. Calicheamicin is a highly potent anticancer agent but lacks the therapeutic index necessary for systemic administration. The potency of calicheamicin has been harnessed by chemical modification followed by linkage to a monoclonal antibody that specifically binds to the unique CD33 cell-surface antigen present on acute myeloid leukaemia cells. Ig, immunoglobulin.

Natural products as building blocks for molecular libraries. Instead of viewing natural products as a stand-alone approach distinct from combinatorial synthesis, it is now much more effective to adopt strategies that combine both approaches. In principle, there seems to be a number of strategies through which the unique molecular diversity of natural products can be leveraged in the design of combinatorial libraries. The target-oriented or focused-library approach seeks to elaborate structural modifications onto an existing bioactive natural-product scaffold in a parallel, systematic fashion in order to improve its inherent biological activity or drug-like properties. This can be performed either by semi-synthetic modification of the parent molecule, or by fully synthetic methods. This strategy was put into practice by Waldmann *et al.*, who recently developed potent, selective inhibitors against the TIE2 receptor tyrosine kinase by parallel synthesis of a small (74 member) focused library based on nakijiquinone C[29]. Earlier efforts in the 1990s effectively utilized several natural products scaffolds including yohimbine[23], paclitaxel[24] and vancomycin[25,26] for solid-phase synthesis of focused libraries.

The diversity-oriented approach seeks to leverage the privileged structural motifs of natural-product scaffolds to synthesize combinatorial libraries capable of binding a wide range of biological targets[98]. Initial investigations based on the diversity-oriented principles have successfully produced high-quality screening libraries by solid-phase synthesis based on natural-product benzopyran scaffolds[28,29]. Later refinements have made it possible to rapidly synthesize diversity-oriented small-molecule microarray libraries to produce molecules that bind Hap3p, a subunit of the yeast Hap2/3/4/5p transcription-factor complex[99] and HIV protease[100]. It has also been shown that it is possible to synthesize diversity-oriented natural-product-based libraries by chemical recombination of complex fragments obtained by chemical degradation of diverse bioactive natural products[101].

Conclusions

The remarkable chemical diversity encompassed by natural products continues to be of relevance to drug discovery. Although today's drug discovery engine operates at an accelerated pace compared with the era in which natural products were pre-eminent sources of drug leads, numerous approaches have been developed to capture their intrinsic value. Crucial breakthroughs in separation and structure-determination technologies have lowered the hurdles inherent in screening mixtures of structurally complex molecules. A greater understanding of the exquisite specificity ingrained in secondary metabolites through the evolutionary process has focused attention on their roles as mediators of protein–protein interactions in vital cellular processes, and advances in synthetic chemistry have revolutionized the processes of material supply and the modulation of biological activity through structural modifications. It seems that no compound is too complex to be recreated in the laboratory.

Furthermore, our ability to model the binding of these evolved, privileged structures with their targets enables the design of simpler mimetics that have superior properties. Efforts to expand the impact of natural chemical diversity on the drug discovery process follow two main chemistry-driven paths. One seeks to simplify crude mixtures, as well as enhance the impact of minor components in assays, through the creation of fractionated natural-product libraries. The other approach uses the power of combinatorial synthesis to amplify the structural context in which the unique features of natural products are expressed. The confluence of these technologies with advances in genomics, metabolic engineering[85] and chemical synthesis offer exciting new possibilities to exploit the remarkable chemical diversity of nature's 'small molecules' in the quest for new drugs.

1. Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as a source of new drugs over the period 1981–2002. *J. Nat. Prod.* **66**, 1002–1037 (2003).
A detailed analysis and description of current natural-product-derived therapeutic agents.
2. Projan, S. J. Infectious diseases in the 21st century: increasing threats, fewer new treatments and a premium on prevention. *Cur. Opin. Pharmacol.* **3**, 457–458 (2003).
3. Kirsop, B. E. The convention on biological diversity: some implications for microbiology and microbial collections. *J. Indust. Microbiol. Biotech.* **17**, 505–511 (1996).
4. Projan S. J. Why is big pharma getting out of antibacterial drug discovery? *Cur. Opin. Microbiol.* **6**, 427–30 (2003).
5. Ajay, W. P. W. & Murcko, M. Can we learn to distinguish between 'Drug-Like' and 'Nondrug-like' molecules? *J. Med. Chem.* **41**, 3314–3324 (1998).
6. Sadowski, J. & Kubinyi, H. Scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **41**, 3325–3329 (1998).
7. Newman, D., Cragg, G., Kingston, D. in *The Practice of Medicinal Chemistry* (ed. Wermuth, C. G.) 91–109 (Academic, London, 2003).
8. Feher, M. & Schmidt, J. M. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **43**, 218–227 (2003).
9. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **23**, 3–25 (1997).
Fundamental work describing important physico-chemical properties of drug molecules.
10. Lee, M. L. & Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: Application in the design of natural product-based combinatorial libraries *J. Comb. Chem.* **3**, 284–289 (2001).
11. Stahura, F., Godden, J. W., Ling, X. & Bajorath, J. Distinguishing between natural products and synthetic molecules by descriptor Shannon entropy analysis and binary QSAR calculations. *J. Chem. Inf. Comput. Sci.* **40**, 1245–1252 (2000).
12. Henkel, T., Brunne, R., Muller, H. & Reichel, F. Statistical investigation of structural complementarity of natural products and synthetic compounds. *Angew. Chem. Int. Ed. Engl.* **38**, 643–647 (1999).
13. Martin, Y. C. Diverse viewpoints on computational aspects of molecular diversity. *J. Comb. Chem.* **3**, 231–250 (2001).
14. Martin, Y. C. & Critchlow, R. E. Beyond mere diversity: tailoring combinatorial libraries for drug discovery. *J. Comb. Chem.* **1**, 32–45 (1999).
15. Evans, B. E. *et al.* Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J. Med. Chem.* **31**, 2235–2246 (1988).
Fundamental work describing the principle of 'privileged' chemical structures.
16. Chothia, C. One thousand families for the molecular biologist. *Nature* **357**, 543–544 (1992).
17. Zhang, C. & DeLisi, C. Estimating the number of protein folds. *J. Mol. Biol.* **284**, 1301–1305 (1998).
18. Salem, G. M., Hutchinson, E. G., Orengo, C. A. & Thornton, J. M. Correlation of observed fold frequency with the occurrence of local structural motifs. *J. Mol. Biol.* **287**, 969–981 (1999).
19. Holm, L. & Sander, C. Mapping the protein universe. *Science* **273**, 595–603 (1996).
20. Hou, J., Sims, G., Zhang, C. & Kim, S-H. A global representation of the protein fold space. *Proc. Natl Acad. Sci. USA* **100**, 2386–2390 (2003).
21. Anantharaman, V., Aravind, L. & Koonin, E. V. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Cur. Opin. Chem. Biol*. **7**, 12–20 (2003).
22. Breinbauer, R., Vetter, I. R. & Waldmann, H. From protein domains to drug candidates- Natural products as guiding principles in the design and synthesis of compound libraries. *Angew. Chem. Int. Ed.* **41**, 2878–2890 (2002).
Analysis and description of the importance of natural products as scaffolds for drug design.
23. Atuegbu, A., MacLean, D., Nguyen, C., Gordon, E. M. & Jacobs, J. W. Combinatorial modification of natural products: Preparation of un-encoded and encoded libraries of *Rauwolfia* alkaloids. *Bioorg. Med. Chem.* **4**, 1097–1106 (1996).
24. Xaio, X. Y., Parandoosh, Z. & Nova, M. P. Design and synthesis of a taxoid library using radiofrequency encoded combinatorial chemistry. *J. Org. Chem.* **62**, 6029–6033 (1997).
25. Xu, R., Grieveldinger, G., Marenus, L. E., Cooper, A. & Ellman, J. A. Combinatorial library approach for the identification of synthetic receptors targeting vancomycin-resistant bacteria. *J. Am. Chem. Soc.* **121**, 4898–4899 (1999).
26. Nicolaou, K. C., Winssinger, N. Hughes, R., Smethurst, C. & Cho, S. Y. New selenium-based safety-catch linkers: solid-phase semisynthesis of vancomycin. *Angew. Chem. Int. Ed.* **39**, 1084–1088 (2000).
27. Nicolaou, K. C., Pfefferkorn, J. A., Roecker, A. J., Cao, G.-Q. & Barluenga, S. Natural product-like combinatorial libraries based on privileged structures 1. General principles and solid phase synthesis of benzopyrans. *J. Am. Chem. Soc.* **122**, 9939–9953 (2000).
A demonstration of the natural products-like chemical diversity achieved by modern combinatorial chemistry.
28. Nicolaou, K. C. *et al.* Natural product-like combinatorial libraries based on privileged structures. 2. Construction of a 10,000-membered benzopyran library by directed split-and-pool chemistry using NanoKans and optical encoding. *J. Am. Chem. Soc.* **122**, 9954–9967 (2000).
29. Kissau, L., Stahl, P., Mazitschek, R., Giannis, A. & Waldmann, H. Development of natural product-derived receptor tyrosine kinase inhibitors based on conservation of protein domain fold. *J. Med. Chem.* **46**, 2917–2931 (2003).
30. Peczuh, M. W. & Hamilton, A. D. Peptide and protein recognition by designed molecules. *Chem. Rev.* **100**, 2479–2494 (2000).
31. Dumont, F. J. FK506, An immunosuppressant targeting calcineurin function. *Curr. Med. Chem.* **7**, 731–748 (2000).
32. Abraham, R. T. & Wiederrecht, G. J. Immunophamacology of rapamycin. *Ann. Rev. Immunol.* **14**, 483–510 (1996).
33. Hersperger, R. & Keller, T. H. Ascomycin derivatives and their use as immunosuppressive agents. *Drugs Future* **25**, 269–277 (2000).
34. Takahashi, N. in *Macrolide Antibiotics* 2nd Edn (ed. Omura, S.) 577–621 (Academic, London, 2002).
35. Pong, K. & Zaleska, M. M. Therapeutic implications for immunophilin ligands in the treatment of neurodegenerative diseases. *Curr. Drug Targets CNS Neurol. Disord.* **2**, 61–72 (2003).
36. Jordan, M. A. Mechanism of action of antitumor drugs that interact with microtubules and tubulin. *Curr. Med. Chem.* **2**, 1–17 (2002).
37. Kowalski, R. J. *et al*. The microtubule-stabilizing agent discodermolide competitively inhibits the binding of paclitaxel (Taxol) to tubulin polymers, enhances tubulin nucleation reactions more potently than paclitaxel, and inhibits the growth of paclitaxel-resistant cells. *Mol. Pharmacol.* **52**, 613–622 (1997).
Fundamental work describing the mechanism of action of a potent antitubulin agent.
38. Bai, R., Cichacz, Z. A., Herald, C. L., Pettit, G. R. & Hamel, E. Spongistatin 1, a highly cytotoxic, sponge-derived, marine natural product that inhibits mitosis, microtubule assembly, and the binding of vinblastine to tubulin. *Mol. Pharmacol.* **44**, 757–66 (1993).
39. Loganzo, F. *et al.* HTI-286, a synthetic analogue of the tripeptide hemiasterlin, is a potent antimicrotubule agent that circumvents P-glycoprotein-mediated resistance *in vitro* and *in vivo*. *Cancer Res.* **63**, 1838–1845 (2003).
40. Sackett, D. L. Podophyllotoxin, steganacin and combretastatin: natural products that bind at the colchicine site of tubulin. *Pharmacol. Therap.* **59**, 163–228 (1993).
41. Silverman, L., Campbell, R. & Broach, J. R. New assay technologies for high throughput screening. *Curr. Opin. Chem. Biol.* **2**, 397–403 (1998).
42. Cohen, P. Protein kinases- The major drug targets of the 21st century? *Nature Rev. Drug Discov.* **1**, 309–316 (2002).
43. Zaman, G. J. R., Garritsen, A., de Boer, T. & van Boeckel, C. A. Fluorescence assays for high throughput screening of protein kinases. *Comb. Chem. High Throughput Screen.* **6**, 313–320 (2003).
44. Fowler, A. *et al*. An evaluation of fluorescence polarization and lifetime discriminated polarization for high throughput screening of serine/threonine kinases. *Anal. Biochem.* **308**, 223–231 (2002).
45. Turek-Etienne, T. C. *et al*. Evaluation of fluorescent compound interference in 4 fluoroescence polarization assays: 2 kinases, 1 protease, and 1 phosphatase. *J. Biomol. Screen.* **8**, 176–184 (2003).
46. Eldridge, G. R. *et al.* High-throughput method for the production and analysis of large natural product libraries for drug discovery. *Anal. Chem*. **74**, 3963–3971 (2002).
47. Abel, U., Koch, C., Speitling, M. & Hansske, F. G. Modern methods to produce natural-product libraries. *Curr. Opin. Chem. Biol*. **6**, 453–458 (2002).
48. Cummins, L. L. *et al.* Multitarget affinity/specificity screening of natural products:Finding and characterizing high affinity ligands from complex mixtures by using high performance mass spectrometry. *J. Nat. Prod.* **66**, 1186–1190 (2003).
49. Firn, R. D. & Jones, C. D. Natural products — a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382–391 (2003).
An analysis of why organisms produce natural products chemical diversity.
50. Chabala, J. C. *et al.* Ivermectin, a new broad-spectrum antiparasitic agent. *J. Med. Chem*. **23**, 1134–1136 (1980).
51. He, H. *et al.* Mannopeptimycins, novel antibacterial glycopeptides from *Streptomyces hygroscopicus,* LL-AC98. *J. Am. Chem. Soc.* **124**, 9729–9736 (2002).
52. De Voe, S. E. & Kunstmann, M. P. Antibiotic AC98 and production. US Patent 3,495,004 (1970).
53. Ruzin, A. *et al.* Mechanism of action of the mannopeptimycins, a novel class of glycopeptide antibiotics active against vancomycin-resistant Gram-positive bacteria. *Antimicrob. Agent. Chemother.* **48**, 728–738 (2004).
54. He, H. *et al.* Mannopeptimycin esters and carbonates, potent antibiotic agents against drug resistant bacteria. *Bioorg. Med. Chem. Lett.* **14**, 279–282 (2004).
55. Sum, P. E., *et al.* Synthesis and evaluation of ether and halogenated derivatives of mannopeptimycin glycopeptide antibiotic. *Biorg. Med. Chem. Lett.* **13**, 2805–2808 (2003).
56. Dushin, R. G. *et al.* Hydrophobic acetal and ketal derivatives of mannopeptimycin-α and AC98-0053: semisynthetic glycopeptides with potent activity against Gram-positive bacteria. *J. Med. Chem.* **47**, 3487–3490 (2004).
An extensive illustration of the capabilities of NMR spectroscopy in the structure elucidation of complex natural products.
57. Strege, M. A. High-performance liquid chromatographic-electrospray ionization mass spectrometric analyses for the integration of natural products with modern high-throughput screening. *J. Chrom. B* **725**, 67–68 (1999).
58. Nielsen, K. F. & Smedsgaard, J. Fungal metabolite screening: database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography-UV-mass spectrometry methodology. *J. Chrom. A* **1002**, 111–136 (2003).
59. Freeman, R. & Morris, G. A. Two-dimensional Fourier transformation in NMR. *Bull. Magn. Res.* **1**, 1–26 (1979).
60. Bax, A., Aszalos, A., Dinya, Z. & Sudo, K. Structure elucidation of the antibiotic desertomycin through the use of new two-dimensional NMR techniques. *J. Am. Chem. Soc.* **108**, 8056–8063 (1986).
61. Schwalbe, H. & Kessler, H. 900 MHz NMR spectrometer in Munich and Frankfurt. *Nachrichten aus der Chemie* **51**, 412–417 (2003).
62. Sandvoss, M., Preiss, A., Levsen, K., Weisemann, R. & Spraul, M. Two new asterosaponins from the starfish *Asterias rubens*: application of a cryogenic NMR probe head. *Magn. Res. Chem.* **41**, 949–954 (2003).
63. Olson, D. L. *et al*. Microflow NMR: concepts and capabilities. *Anal. Chem.* **76**, 2966–2974 (2002).
64. Serber, Z. *et al*. New carbon-detected protein NMR experiments using cryoProbes. *J. Am. Chem. Soc.* **122**, 3554–3555 (2000).
65. Satake, M. *et al*. Structural confirmation of maitotoxin based on complete ^{13}C NMR assignments and the three-dimensional PFG NOESY-HMQC spectrum. *J. Am. Chem. Soc.* **117**, 7019–7020 (1995).
66. He, H. *et al*. Lomaiviticins A and B, potent antitumor antibiotics from *Micromonospora Lomaivitiensis*. *J. Am. Chem. Soc*. **123**, 5352–5363 (2001).
67. McDonald, L. A. *et al*. FTMS Structure elucidation of natural products: application to muraymycin antibiotics using ESI Multi-CHEF SORI-CIT FTMS^n, the Top-Down/Bottom-Up approach, and HPLC ESI capillary-skimmer CID FTMS. *Anal. Chem*. **75**, 2730–2739 (2003).
Capabilities of modern mass spectrometry in the structure elucidation of complex natural products.
68. Gunasekera, A. P., Gunaskera, M., Longley, R. E. & Schulte, G. K. Discodermolide: a new bioactive poly-hydroxylated lactone from the marine sponge *Discodermia dissoluta*. *J. Org. Chem*. **55**, 4912–4915 (1990).
69. Towle, M. J. *et al. In vitro* and *in vivo* anticancer activities of synthetic macrocyclic ketone analogues of halichondrin B. *Cancer Res*. **61**, 1013–1021 (2001).
70. Kuznetsov, G. *et al.* Induction of morphological and biochemical apoptosis following prolonged mitotic blockage by halichondrin B macrocyclic ketone analog E7389. *Cancer Res*. **64**, 5760–5766 (2004).
71. Schenk, T. *et al*. A Generic assay for phosphate-consuming or -releasing enzymes coupled on-line to liquid chromatography for lead finding in natural products. *Anal. Biochem.* **316**, 118–126 (2003).
72. Cummins, L. L. *et al*. Multitarget affinity/specificity screening of natural products finding and characterizing high-affinity ligands from complex mixtures by using high-performance mass spectrometry. *J. Nat. Prod*. **66**, 1186–1190 (2003).
73. Schriemer, D. C., Bundle, D. R., Li, L. & Hindsgaul, O. Micro-scale frontal affinity chromatography with mass spectrometric detection: a new method for the screening of compound libraries. *Angew. Chem. Int. Ed*. **37**, 3383–3387 (1998).

74. Chan, N. W. C., Lewis, D. F., Rosner, P. J., Kelly, M. A. & Schriemer, D. C. Frontal affinity chromatography–mass spectrometry assay technology for multiple stages of drug discovery: applications of a chromatographic biosensor. *Anal. Biochem.* **319**, 1–12 (2003).
75. Zhu, L., Chen, L., Luo, H. & Xu, X. Frontal affinity chromatography combined on-line with mass spectrometry: a tool for the binding study of different epidermal growth factor receptor inhibitors. *Anal. Chem*. **75**, 6388–6393 (2003).
76. Wolfender, J-L., Ndjoko, K. & Hostettmann, K. Liquid chromatography with ultraviolet absorbance–mass spectrometric detection and with nuclear magnetic resonance spectroscopy: a powerful combination for the on-line structural investigation of plant metabolites. *J. Chrom. A* **1000**, 437–455 (2003).
77. Exarchou, V., Godejohann, M., van Beek, T. S., Gerothanassis, I. P. & Vervoort, J. LC-UV-solid-phase extraction-NMR-MS combined with a cryogenic flow probe and its application to the identification of compounds present in greek oregano. *Anal. Chem*. **75**, 6288–6294 (2003).
Capabilities of hyphenated spectroscopic techniques in the structure elucidation of complex natural product mixtures.
78. Laude, D. A. &, Wilkins, C. L. Direct-linked analytical scale high-performance liquid chromatography/nuclear magnetic resonance spectrometry. *Anal. Chem.* **56**, 2471–2475 (1984).
79. Albert, K. Liquid chromatography-nuclear magnetic resonance spectroscopy. *J. Chrom.* **856**, 199–211 (1999).
80. Wolfender, J.-L., Ndjoko, K. & Hostettmann, K. The potential of LC-NMR in phytochemical analysis. *Phytochem. Anal.* **12**, 2–22 (2001).
81. Schaller, F., Wolfender, J.-L., Hostettmann, K. & Mavi, S. New antifungal 'quinone methide' diterpenes from *Bobgunnia madagascariensis* and study of their interconversion by LC/NMR *Helv. Chim. Act.* **84**, 222–229 (2001).
82. Lommen, A., Godejohann, M., Venema, D. P., Hollman, P. C. H., Spraul, M. Application of directly coupled HPLC-NMR–MS to the identification and confirmation of quercetin glycosides and phloretin glycosides in apple peel. *Anal. Chem.* **72**, 1793–1797 (2000).
83. Spraul, M. *et al*. Advancing sensitivity for LC-NMR-MS using a cryoflow probe: application to analysis of acetominophen metabolites in urine. *Anal. Chem.* **75**, 1546–1551 (2003).
84. Wu, N., Webb, A., Peck, T. L., Sweedler, J. V. Online NMR detection of amino acids and peptides in microbore LC. *Anal. Chem.* **67**, 3101–3107 (1995).
85. Khosla, C. & Keasling, J. D. Metabolic engineering for drug discovery and development. *Nature Rev. Drug Discov.* **2**, 1019–1025 (2003).
86. Nicolaou, K. C. *et al*. Total synthesis of apoptolidin: construction of enantiomerically pure fragments. *J. Am. Chem. Soc.* **125**, 15433–15442 (2003).
87. Lin, S. *et al*. Total syntheses of TMC-95A and B via a new reaction leading to Z-enamides. Some preliminary findings as to SAR. *J. Am. Chem. Soc.* **126**, 6347–6355 (2004).
88. Wender, P. A. *et al*. Modeling of the bryostatins to the phorbol ester pharmacophore on protein kinase C. *Proc. Natl Acad. Sci. USA* **85**, 7197–201 (1988).
89. Wender, P. A. *et al*. The design, computer modeling, solution structure, and biological evaluation of synthetic analogs of bryostatin 1. *Proc. Natl Acad. Sci. USA* **95**, 6624–6629 (1998).
90. Wender, P. A., DeBrabander, P. G, Hinkle, K. W., Lippa, B., Pettit, G. R. Synthesis and biological evaluation of fully synthetic bryostatin analogues. *Tet. Lett.* **39**, 8625–8628 (1998).
91. Wender, P. A. *et al.* The practical synthesis of a novel and highly potent analogue of bryostatin. *J. Am. Chem. Soc*. **124**, 13648–13649 (2002).
Along with preceding articles, a good example of natural-products-based drug design.
92. Hommel, U., Weber, H-P., Oberer, L., Naegeli, H. U., Oberhauser, B. & Foster, C. A. The 3D-structure of a natural inhibitor of cell adhesion molecule expression. *FEBS Lett.* **379**, 69–73 (1996).
93. Chen, Y., Bilban, M., Foster, C. A. & Boger, D. L. Solution-phase parallel synthesis of a pharmacophore library of HUN-7293 analogues: A general chemical mutagenesis approach to defining structure-function properties of naturally occurring cyclic (depsi)peptides. *J. Am. Chem. Soc.* **124**, 5431–5440 (2002).
94. Smith, III, A. B., Cho, Y. S., Pettit, G. R. & Hirschmann, R. Design, synthesis, and evaluation of azepine-based cryptophycin mimetics. *Tetrahedron* **59**, 6991–7009 (2003).
95. Lee, M. D., Durr, F. E., Hinman, L. M., Hanmann, P. R. & Ellestad, G. A. The calicheamicins. *Adv. Med. Chem.* **2**, 31–66 (1993).
96. Damle, N. K. & Frost, P. Antibody-targeted chemotherapy with immunoconjugates of calicheamicin. *Curr. Opin. Pharm*. **3**, 386–390 (2003).
97. Hamann, P. R. *et al.* Gemtuzumab ozogamicin, A potent and selective anti-CD33 antibody-calicheamicin conjugate for treatment of acute myeloid leukemia. *Bioconj. Chem*. **13**, 47–58 (2002).
98. Burke, M. D. & Schreiber, S. L. A planning strategy for diversity oriented synthesis. *Angew. Chem. Int. Ed.* **43**, 46–58 (2004).
99. Koehler, A. N., Shamji, A. F. & Schreiber, S. L. Discovery of an inhibitor of a transcription factor using small molecule microarrays and diversity-oriented synthesis. *J. Am. Chem. Soc.* **125**, 8420–8421 (2003).
100. Brik, A. M. *et al.* Rapid diversity-oriented synthesis in microtiter plates for *in situ* screening of HIV protease inhibitors. *Chembiochem* **4**, 1246–1248 (2003).
101. Niggemann, J., Michaelis, K., Frank, R., Zander, N. & Höfle, G. Natural product-derived building blocks for combinatorial synthesis. Part 1. Fragmentation of natural products from myxobacteria. *J. Chem. Soc. Perkin. Trans.* **1**, 2490–2503 (2002).
102. ter Haar, E. *et al.* Discodermolide, A cytotoxic agent that stabilizes microtubules more potently than taxol. *Biochemistry* **35**, 243–250 (1996).
103. Nerenberg, J. B., Hung, D. T., Somers, P. K., Schreiber, S. L. Total synthesis of the immunosuppressive agent (-)-discodermolide. *J. Am. Chem. Soc.* **115**, 12621–12622 (1993).
104. Smith, A. B. et al. Evolution of a gram-scale synthesis of (+)-discodermolide. *J. Am. Chem. Soc.* **122**, 8654–8664 (2000).
105. Paterson, I., Florence, G. J., Gerlach, K. & Scott, J. P. Total synthesis of the antimicrotubule agent (+) discodermolide using boron–mediated aldol reactions of chiral ketones. *Angew. Chem. Int. Ed.* **39**, 377 (2000).
106. Marshall, J. A. & Johns, B. A. Total synthesis of (+)-discodermolide. *J. Org. Chem.* **63**, 7885–7892 (1998).
107. Mickel, S. J. et al. Large scale synthesis of the anti-cancer marine natural product (+)-discodermolide. Part 1. Org. *Proc. Res. Dev.* **8,** 92–100 (2004).
108. Mickel, S. J. et al. Large scale synthesis of the anti-cancer marine natural product (+)-discodermolide. Part 2. Org. *Proc. Res. Dev.* **8**, 101–106 (2004).
109. Mickel, S. J. et al. Large scale synthesis of the anti-cancer marine natural product (+)-discodermolide. Part 3. Org. *Proc. Res. Dev.* **8**, 107–112 (2004).
110. Mickel, S. J. et al. Large scale synthesis of the anti-cancer marine natural product (+)-discodermolide. Part 4. Org. *Proc. Res. Dev.* **8,** 113–121 (2004).
111. Mickel, S. J. et al. Large scale synthesis of the anti-cancer marine natural product (+)-discodermolide. Part 5. Org. *Proc. Res. Dev.* **8**, 122–130 (2004).
References 107–111 provide a strong example of successful synthetic scale-up of a natural product for clinical supply.

Acknowledgements
We thank N. Pilote for valuable assistance in literature and patent analysis.

Competing interests statement
The authors declare no competing financial interests.

Online links

DATABASES
The following terms in this article are linked online to:
Entrez Gene:
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
Hap3p | TIE2
Access to this interactive links box is free online.