Lessons from natural molecules

Jon Clardy & Christopher Walsh

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA (e-mail: jon_clardy@hms.harvard.edu)

Natural products have inspired chemists and physicians for millennia. Their rich structural diversity and complexity has prompted synthetic chemists to produce them in the laboratory, often with therapeutic applications in mind, and many drugs used today are natural products or natural-product derivatives. Recent years have seen considerable advances in our understanding of natural-product biosynthesis. Coupled with improvements in approaches for natural-product isolation, characterization and synthesis, these could be opening the door to a new era in the investigation of natural products in academia and industry.

In the past century, diverse classes of natural products have been isolated and their structures characterized. These discoveries, along with the elucidation of biological and biochemical mechanisms of therapeutic action, have been central to the work of organic and medicinal chemists. Natural products have been invaluable as tools for deciphering the logic of biosynthesis and as platforms for developing front-line drugs[1,2]. For example, between 1981 and 2002, 5% of the 1,031 new chemical entities approved as drugs by the US Food and Drug Administration (FDA) were natural products, and another 23% were natural-product-derived molecules[3]. Natural products are still major sources of innovative therapeutic agents for infectious diseases (both bacterial and fungal), cancer, lipid disorders and immunomodulation[4].

However, the complexity of many natural products can limit the scope for making chemical modifications to optimize their therapeutic use. Moreover, obtaining a renewable supply of active compounds from biological sources can be problematic. Nevertheless, as the recent multigram, total synthesis of the potent anti-cancer natural product discodermolide shows[5], the increasing efficiency of synthetic organic chemistry has reduced the barrier posed by limited natural supply, even for materials with very complex structures.

Here, we examine some of the lessons from nature that remind us of the structural and mechanistic diversity of natural small molecules, and evaluate the uncertain present and diminishing future interest for natural products as central players in the research strategies of pharmaceutical companies. We begin by describing the structural features of representative natural products of medicinal importance, their mechanisms of action and their biosynthesis, before turning to prospects for future discoveries.

Structural features of natural products

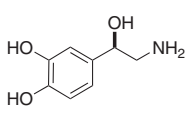
How do natural products compare with drugs?

Figure 1a shows the structures of four natural products that have proved to be useful as drugs or leads: vancomycin[6,7] (**1**), staurosporine[8] (**2**), rapamycin[9] (**3**) and Taxol[10] (**4**). These have been used for the treatment of Gram-positive bacterial infections, as a lead indolecarbazole structure[11] for the inhibition of protein kinases at the ATP-binding site, for immunosuppression, and for cancer chemotherapy, respectively. For comparison, Fig. 1b shows the structures of four synthetic drug molecules that are in widespread use: Viagra[12] (**5**), Prozac[13] (**6**), Lipitor[14] (**7**), and Gleevec[15] (**8**). These are used to treat erectile dysfunction, depression, hypercholesterolaemia and chronic myelogenous leukaemia, respectively. Each of the eight molecules has a well-defined biological target to which it binds with useful affinity, and all these targets are proteins, except for the peptidoglycan termini of bacterial cell walls (the target for

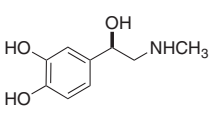
Box 1

Signalling within and among organisms

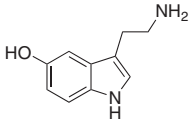
Some important natural products with low molecular weights act with potency and specificity at protein receptors; for example, the low-molecular-weight amine neurotransmitters, derived from enzymatic decarboxylation of proteinogenic amino acids. These neurotransmitters have been outstanding platforms for natural-product-based drug design. Decarboxylation and subsequent oxidation of tyrosine generates the hormones and neurotransmitters noradrenaline (**43**) and adrenaline (**44**). Similar processing of tryptophan yields the neurotransmitter serotonin (**45**) and the hormone melatonin (**46**). Simple decarboxylation of histidine gives histamine (**47**), which has at least three well-characterized activities: (1) bronchoconstriction and vasodilation; (2) gastric-acid secretion; and (3) neurotransmission. These simple molecules have provided starting points for numerous small-molecule drugs. For example, seven out of ten anti-migraine medicines are based on serotonin[3], several generations of α- and β-adrenergic drugs are generated from adrenaline scaffolds, and antihistamines (histamine receptor H1 and H2 selective antagonists) and selective serotonin re-uptake inhibitors (SSRIs) are some of the world's best-selling drugs.



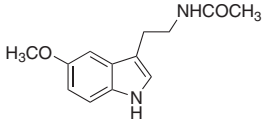
43 Noradrenaline



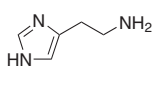
44 Adrenaline



45 Serotonin



46 Melatonin



47 Histamine

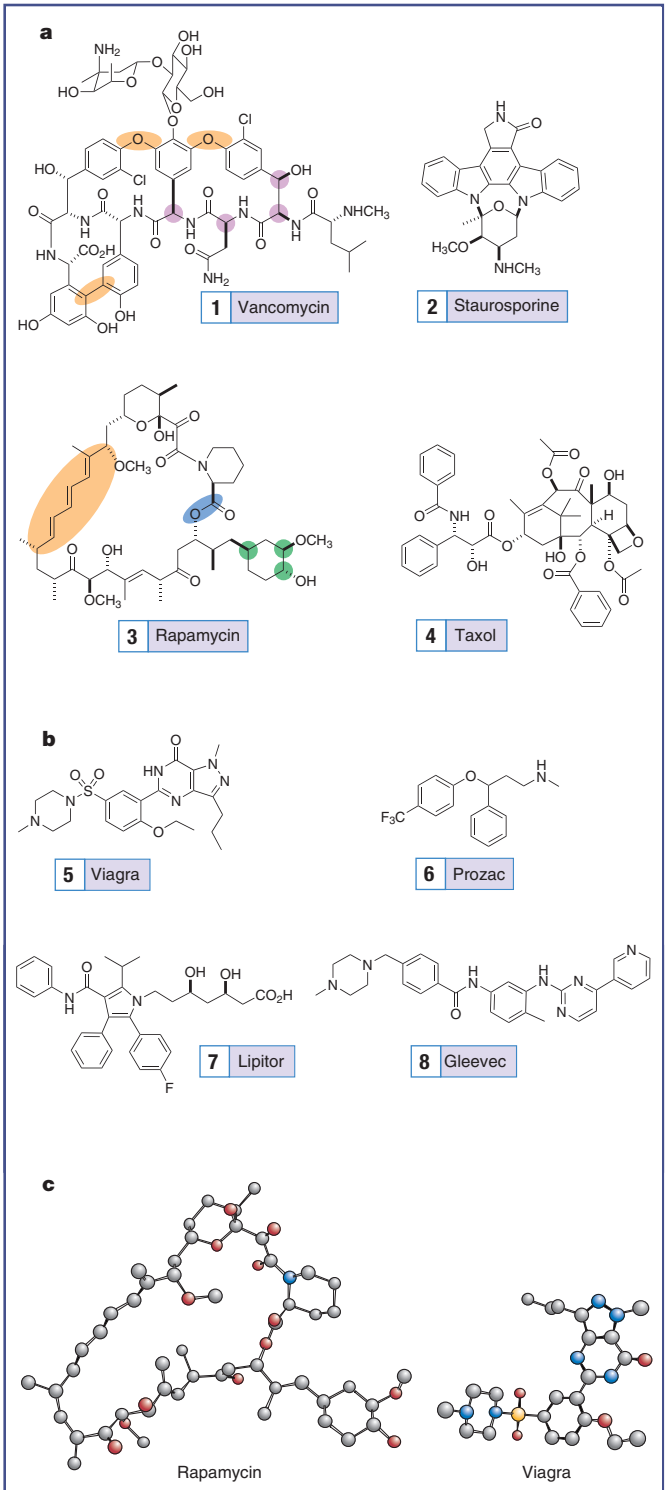


Figure 1 Medically significant natural products and synthetic molecules. **a**, Natural products. Vancomycin (**1**), an antibiotic for bacterial infections; staurosporine (**2**), a lead compound for the development of selective kinase inhibitors for cancer; rapamycin (**3**), a compound for immunosuppression; Taxol (**4**), an anti-cancer agent; **b**, Synthetic molecules. Viagra (**5**) for erectile dysfunction; Prozac (**6**) for depression; Lipitor (**7**) for hypercholesterolaemia; and Gleevec (**8**) for chronic myelogenous leukaemia. Natural products have strong conformational biases based on stereogenic centres (**1**, mauve circles), ether and ring fusions (**1**, yellow ovals), strategically placed substituents to select a single conformation (**3**, green circles), macrocyclization (**3**, blue oval), and conjugation (**3**, yellow oval). Staurosporine's (**2**) interlocking rings lead to a completely rigid core structure. **c**, Three-dimensional structural representations of rapamycin and Viagra.

vancomycin). Of the eight, only staurosporine is promiscuous in its recognition of protein targets; it binds to the ATP-recognition site of many protein kinases — a property that has limited its uses to a structural lead and a research tool[11].

Two-dimensional representations and three-dimensional images of these structures are shown in Fig. 1 to emphasize their architectural determinants. These comparisons highlight several general distinctions between natural-products and synthetic drugs/drug candidates. First, natural products typically have more stereogenic centres and more architectural complexity than synthetic molecules fashioned by medicinal chemists (Fig. 1), although several important natural products that act with potency and specificity at protein receptors have simple structures (Box 1). Second, natural products contain relatively more carbon, hydrogen and oxygen, and less nitrogen and other elements than synthetic medicinal agents. Third, many useful natural products have molecular masses in excess of 500 daltons and high polarities (greater water solubility), and therefore violate Lipinski's 'rule of five': this is a set of guidelines based on the characteristics of

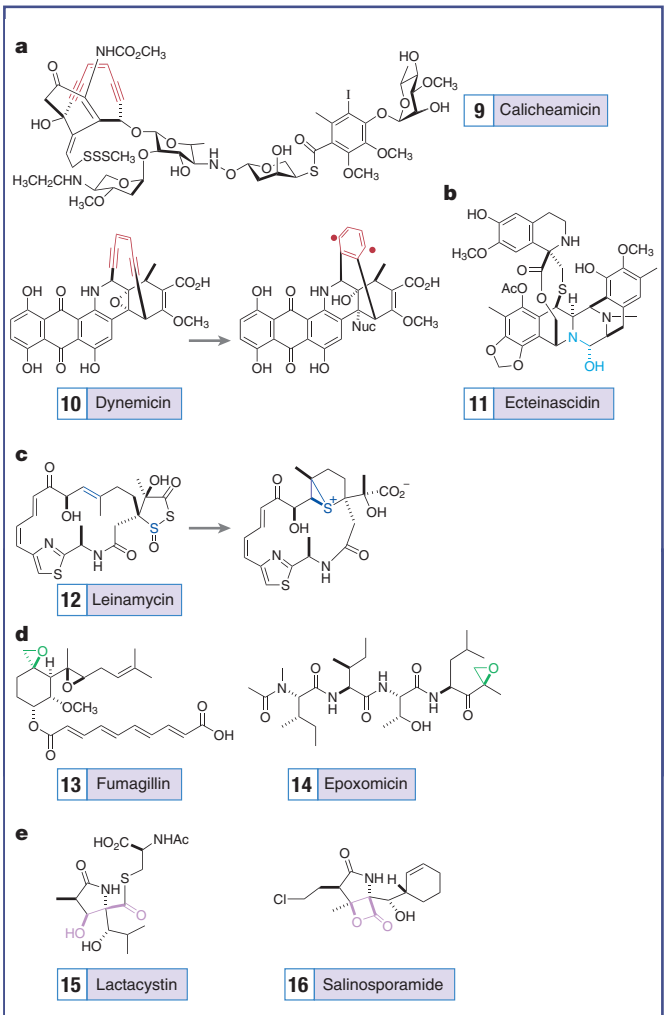


Figure 2 Natural products that exploit reactive functional groups. Compounds **9**–**16** all illustrate nature's ability to either mask or fine-tune the reactivity of functional groups. **a**,The enediyne group (red) in calicheamicin (**9**) and dynemicin (**10**) is activated to give a diradical intermediate that damages DNA (as shown for dynemicin). **b**, The carbolamine group in ecteinascidin (light blue; **11**) is converted to an iminium ion that reacts with DNA. **c**, The dithian-1,3-oxide group (dark blue) in leinamycin (**12**) is activated to form an episulphonium intermediate that alkylates DNA. **d**, Fumagillin (**13**) and epoxomicin (**14**) contain reactive epoxide groups (green) that trap proteases. **e**, The masked or explicit β-lactones (mauve) in lactacystin (**15**) and salinosporamide (**16**), respectively, target the proteasome.

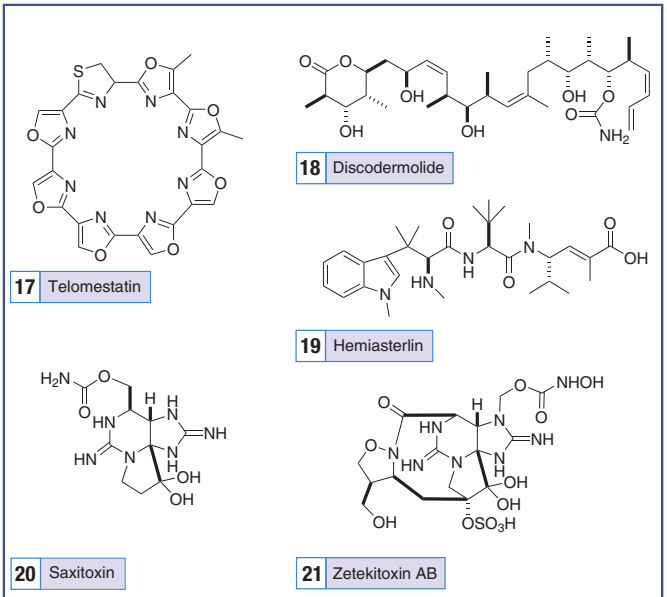


Figure 3 Natural products that exploit shape and polarity complementarity to biological targets. Biological targets include: G quartets for telomestatin (**17**); tubulin for discodermolide (**18**) and hemiasterlin (**19**); and ion channels for saxitoxin (**20**) and zetekitoxin AB (**21**).

known drugs that provide an indication of whether a given small molecule is likely to have the desired pharmacokinetic properties to be an oral drug (in terms of how it is absorbed, distributed, metabolized and eliminated by the body). All four synthetic drugs featured have a molecular mass lower than 500 daltons and can be orally administered.

Both the natural products and the synthetic drugs have strong conformational biases and constraints: examples include macrocyclizations (the formation of macrocyclic ring structures, such as that present in rapamycin shown in Fig. 1), fused-ring systems, ether crosslinks, extensive conjugation and strategically placed substituents that 'preorganize' them for populating conformers that bind to specific biological targets, in these cases enzymes and receptors (Fig. 1). The conformer restrictions and/or architectural rigidifications built into active molecules reflect the importance of minimizing the loss of entropy as molecules bind to biological targets. Avoiding such energy loss by preorganizing conformers to present complementary electrostatic, hydrogen-bonding and hydrophobic interactions with the protein targets allows these and other small molecules to retain sufficient binding energy to function as potent ligands. These are typically in the 10^{-7} to 10^{-9} M range of potency.

Lessons from natural-product functionalities

Natural products have been effective in teaching us about chemical functionality that is compatible with the aqueous milieu of biological microenvironments; the lessons learned have been both surprising and deep. Some notable examples of instructive natural products, all of which contain highly reactive functional groups or the precursors to such groups, are shown in Figs 2 and 3. The enediynes, including calicheamicin[16] (**9**) and dynemicin[17] (**10**) are among the most potent cytotoxic agents discovered, with 50%-effective dose ranges in cell-killing assays as low as 10^{-17} M — a nominal concentration in the range of one molecule per cell[18]. The unusual trisulphide in calicheamicin and the quinone in dynemicin are redox-activated triggers that initiate aromatization cascades leading to the formation of diradical intermediates that damage DNA (as shown for dynemicin in Fig. 2a). Nature frequently exploits such reactive functional groups in biologically active natural products. In ecteinascidin (**11**, Fig. 2b), a carbinolamine is converted to an iminium ion that reacts with DNA to form a covalent adduct[19,20]. In leinamycin (**12**, Fig. 2c), the dithian-1,3-oxide group in this anti-tumour agent is activated by a thiol to form an episulphonium intermediate that alkylates DNA[17]. Fumagillin (**13**) and epoxomicin (**14**) both use reactive epoxide groups to covalently trap proteases (Fig. 2d). Fumagillin's ability to selectively inhibit methionine aminopeptidase type 2 leads to the inhibition of angiogenesis (the formation of new blood vessels)[21], and fumagillin-inspired compounds are being investigated as anti-cancer agents. Epoxomicin inhibits the degradation of proteins by the proteasome[22], and related proteasome inhibitors are being developed for a variety of therapeutic uses. β-lactones, either masked as in lactacystin (**15**) or explicit as in salinosporamide (**16**, Fig. 2e) are also potent proteasome inhibitors[23,24]. These examples illustrate nature's ability to either mask or finely tune the reactivity of labile functional groups so that a small molecule can retain the kinetic stability needed for it to reach and specifically inhibit biological targets by a covalent mechanism.

Of course, not all natural products work by covalent mechanisms; most employ the exquisite structural complementarity between a small molecule and its target. Telomestatin (**17**, Fig. 3), with its eight tandem heterocycles in a macrocyclic array, mimics the tetraguanine fragments (G quartets) found on telomeres[25]. This mimicry allows telomestatin to be a nanomolar inhibitor of telomerase. Rapamycin (**3**, Fig. 1a) uses two different faces to bind two different proteins with nanomolar efficiency sufficient to disrupt a cytoplasmic signal transduction cascade[26]. Discodermolide[27] (**18**) and hemiasterlin[28,29] (**19**, Fig. 3) bind to tubulin, and both are exciting leads for cancer therapy. Natural products can also block ion channels, as illustrated by saxitoxin[30] (**20**) and zetekitoxin AB[31] (**21**, Fig. 3).

The natural products in Fig. 1a are from traditional sources: soil microbes (vancomycin, **1**; staurosporine, **2**; and rapamycin, **3**) and plants (Taxol, **4**). Many of the natural products in Figs 2 and 3 are from nontraditional sources. Ecteinascidin (**11**) is from a small reef-dwelling tunicate found in the West Indies[32]. Discodermolide (**18**) is from a deep-water sponge and hemiasterlin (**19**) was found in two different sponges — one from South Africa, the other from Papua New Guinea. Saxitoxin (**20**) is produced by dinoflagellates (especially those producing 'red tides'), although it was traditionally isolated from filter-feeding shellfish that consumed the dinoflagellates. Its structural relative zetekitoxin AB (**21**) was isolated from the Panamanian golden frog, but its original producer is probably a microbe that is consumed by insects, which are in turn consumed by the frogs. This previously unexplored biological diversity coupled with modern analytical techniques and synthetic organic chemistry could lead to a new chapter of natural-products research, as is discussed in the section 'Discovery from new sources' below.

Understanding the functional-group arrays used by nature has informed synthetic- and medicinal-chemistry efforts about biomimetic strategies and isostere (shape-conserving) replacements. The synthetic molecules in Fig. 1b feature the design principles favoured by medicinal chemists: a high proportion of aromatic and heteroaromatic rings, few stereogenic centres, low molecular weights and a lack of chemical reactivity. In contrast, the enediyne antitumour antibiotic calicheamicin (**9**, Fig. 2a) is large (almost 1,400 daltons), devoid of core aromatic rings (until triggered by subsequent chemical reactions), loaded with stereogenic centres and highly reactive. Whether its potent biological properties can be exploited for anti-cancer therapy is not completely settled but an antibody-targeted-therapy approach Mylotarg, that takes advantage of its extraordinary cytotoxicity has been in the clinic since 2000 (see refs 33, 34).

Synthetic molecules are increasingly produced by combinatorial chemistry approaches, in which a common core is elaborated by attaching combinations of fragments to reactive sites on the core's periphery. An old, but still useful, template is the benzodiazepine core (**22**, Fig. 4a). In the construction of a synthetic combinatorial library based on the benzodiazepine skeleton (**22**), diversity elements (R_1, R_2 and R_3) are attached to a common skeleton. If ten versions of

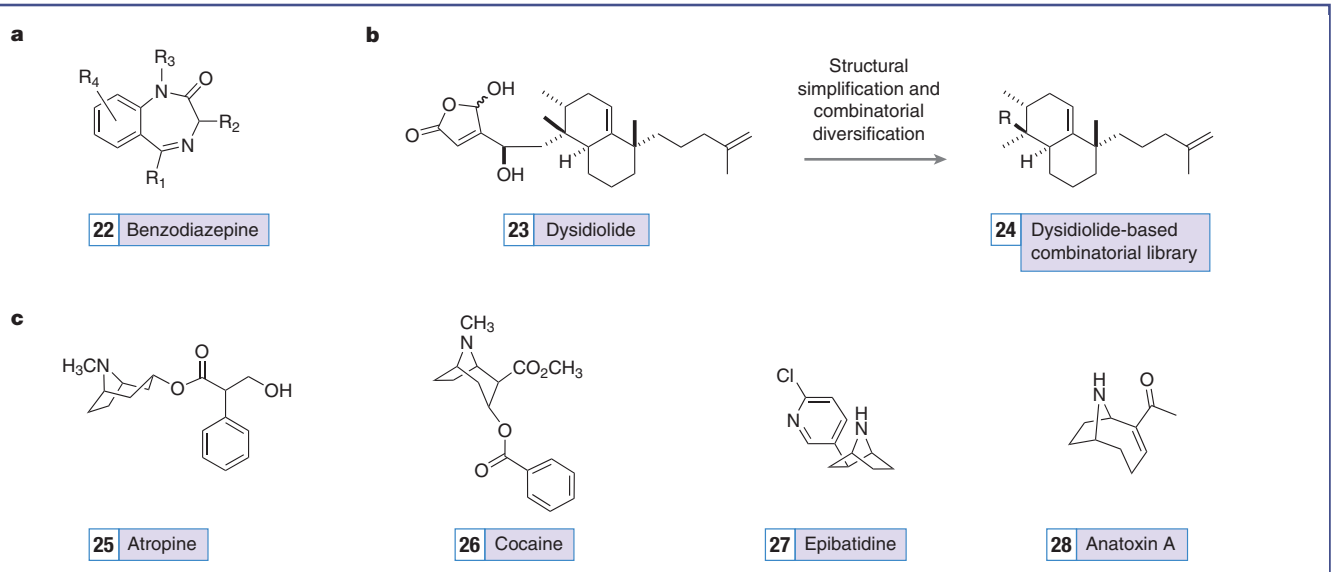


Figure 4 Template diversification. **a**, The benzodiazepine core is a common template for synthetic diversification because the groups indicated here as R_1, R_2, R_3 and R_4 can be varied widely. **b**, Dysidiolide (**23**) has been used as a template for a natural-products-based diversity library. The native structure was simplified and a single diversity element was used to create the library (**24**). **c**, Compounds **25**–**28** illustrate natural template diversification. Atropine (**25**) and cocaine (**26**) are plant alkaloids with mydriatic and local anaesthetic properties, respectively. Epibatidine (**27**) is a non-opioid analgesic isolated from the skin of an Ecuadoran poison frog, and anatoxin A (**28**) is the Very Fast Death Factor produced by cyanobacteria.

each diversity element are used, the library contains 1,000 different molecules, each with a different combination of R_1, R_2 and R_3. Nature uses similar strategies, especially the oxidative elaboration of a central core followed by capping reactions (discussed in the section 'Re-engineering of biosynthetic pathways' below). Several natural-product-like combinatorial libraries have been synthesized[35]; a library based on dysidiolide (**23**, Fig. 4b), a potent phosphatase inhibitor from a marine sponge[36], and summarized in structure **24**, is typical[37]. In this case, the library construction involved the diversification of a single element on a single scaffold (**24**); even with these apparent limitations, the library contained potent phosphatase inhibitors[37]. Successes with combinatorial libraries based on natural-product templates argue that natural products, which have been honed by their evolutionary history for biological activity, are excellent starting points for structural diversification[37]. Combinatorial biosynthesis (which is discussed in the section 'Re-engineering of biosynthetic pathways' below) uses the manipulation of biosynthetic machinery to accomplish much the same goal, but with greater control over core elements.

Nature also dramatically varies the core size and stereochemistry of molecules, as the series atropine (**25**), cocaine (**26**), epibatidine (**27**) and anatoxin A (**28**) illustrates (Fig. 4c). Diversity-oriented synthesis[38,39], which combines the strengths of combinatorial multiplexing and core variability, is emerging as a powerful technique for finding biologically active small molecules[40].

Advantages and constraints of nature's biosynthetic strategy

Natural products can be divided into several structural classes: polyketides, nonribosomal peptides (NRPs), terpenes, alkaloids and many others. Products are classed according to shared scaffolding elements, which in turn reflect the strategies for their assembly by pathways of biosynthetic enzymes in the producer organisms.

Most classical small molecules from nature are secondary metabolites — products from conditional pathways that are turned on in a particular context or situation. These include metabolites made during starvation (for example, carbapenem antibiotics produced by *Pseudomonas* bacteria), in development (for example, antibiotics made when *Streptomycetes* enter cellular differentiation pathways), and signalling (such as quorum-sensing molecules biosynthesized at particular culture densities of microbes)[41].

The building blocks for natural products are most often the monomer constituents (amino acids for nonribosomal peptides; acyl-CoA thioesters for polyketides; isoprenyl-pyrophosphates for terpenes) of primary metabolic pathways, which are shunted into the secondary pathways when a particular metabolic channel is opened. When monomers dedicated to secondary metabolic pathways are required, such as 4-OH-phenylglycine and 3,5-$(OH)_2$-phenylglycine for vancomycin (Fig. 1a) and methoxymalonyl CoA for some polyketide initiations, they are produced by a 'just-in-time' cellular-inventory strategy[42]. To this end, biosynthetic gene clusters for nonribosomal peptides or NRPs (such as vancomycin) and polyketides (such as rapamycin in Fig. 5) contain both genes for the assembly-line enzymes and genes for enzymes to make the dedicated monomers needed for the assembly lines to run[43]. A third set of clustered genes typically encodes enzymes that tailor the nascent products released from assembly lines, most notably for glycosylation and oxidation: these two modifications are often required to make the product biologically active[44]. The gene clustering allows coordinated regulation and inventory control of both enzyme catalysts and small-molecule building blocks. The enzyme catalysts are needed to run the secondary pathways comprising 20 to 40 steps that turn out the finished natural products.

The simple monomers are used in sets of iterative condensations; linear intermediates are built up by a single type of chemistry. For example, for terpene and isoprenoid natural products, the fundamental chain-elongation step is C-alkylation enzyme catalysis, which adds a C_5-isoprene unit to the end of a growing chain by means of allylic carbonium ion chemistry. The growing chain is held in the microenvironment of the oligomerizing enzymes that control foldamer conformation. This is turn dictates cyclization patterns, such as in Taxol or polycyclic triterpene assembly. In NRP- and polyketide-chain buildup, both the growing chain and the incoming monomer are tethered covalently to the enzyme as thioesters. For NRPs, the iterative elongation step is amide-bond formation, whereas in polyketide-chain growth it is Claisen-type C–C bond formation to the β-keto–acyl thioester products.

How is so much structural diversity generated in these three classes of natural products, which are produced from a limited pool of simple primary metabolites? The general answers are incomplete processing and/or active tailoring of the initial intermediates during chain elongation, acyclic foldamer control for regiospecific cyclization reactions and post-elongation tailoring and maturation by enzyme action. In terpenes, foldamer control and the placement of basic side chains in the terpene cyclase active sites controls the location and size of cation-mediated cyclizations[45,46]. In NRP-assembly lines, cysteinyl, seryl and threonyl side chains can be regiospecifically cyclized, dehydrated and oxidized to create thiazoles and oxazoles during elongation. In multimodular polyketide assembly lines, the initial β-keto–acyl thioesters from Claisen condensation can be processed all the way to β-CH_2 methylene groups or can accumulate as β-keto, β-hydroxy or αβ-olefinic intermediates[47,48]. Where full-length peptidyl thioesters or full-length polyketidyl thioesters have been assembled on the most downstream way stations of NRP- and polyketide-assembly lines, chain release can occur through hydrolysis. Alternatively, chain release can occur through an intramolecular regiospecific cyclization from a nucleophilic -OH or -NH in the chain to form a macrolactone or macrolactam. Intramolecular release results in a macrocycle that builds in conformational constraints[49].

The biosynthesis of the immunosuppressive drug rapamycin (**3**) illustrates how structural diversity is generated from simple building blocks[50]. As noted in Fig. 5, this is predominantly a polyketide natural product with a dihydroxycyclohexenyl CoA as a starting building block, and seven equivalents of malonyl CoA and seven equivalents of methylmalonyl CoA as the elongating monomers. One amino acid is incorporated; in this case L-pipecolate, which is derived from a dedicated enzymatic cyclization of the primary metabolite lysine. The order in which these four classes of monomer are incorporated is determined by the order of the 15 modules in the enzymatic assembly line. Figure 5 shows how the single nonribosomal peptide synthase (NRPS) module is at the end, suggesting that pipecolate is the last monomer to be incorporated. The linear acyl-*S*-enzyme intermediate that is proposed to undergo capture by intramolecular cyclization to yield the 30-member macrolactam is also shown. The nascent macrocyclic product is then tailored by a series of enzymatic methylations and oxidation/oxygenation steps to yield rapamycin.

An analogous but distinct logic is used in the assembly of the enediyne cores by polyketide synthase assembly lines, which are then followed by tailoring reactions[51,52]. Altogether, 55 enzymatic reactions are used to combine five classes of building block (acetyl CoA, malonyl CoA, tyrosine, chorismate and glucose) to give the enediyne C-1027 (ref. 53).

Tailoring reactions to control oxidation states

All the linear chain-elongation steps in polyketide and NRP monomer assembly occur as thioesters, and without any protecting

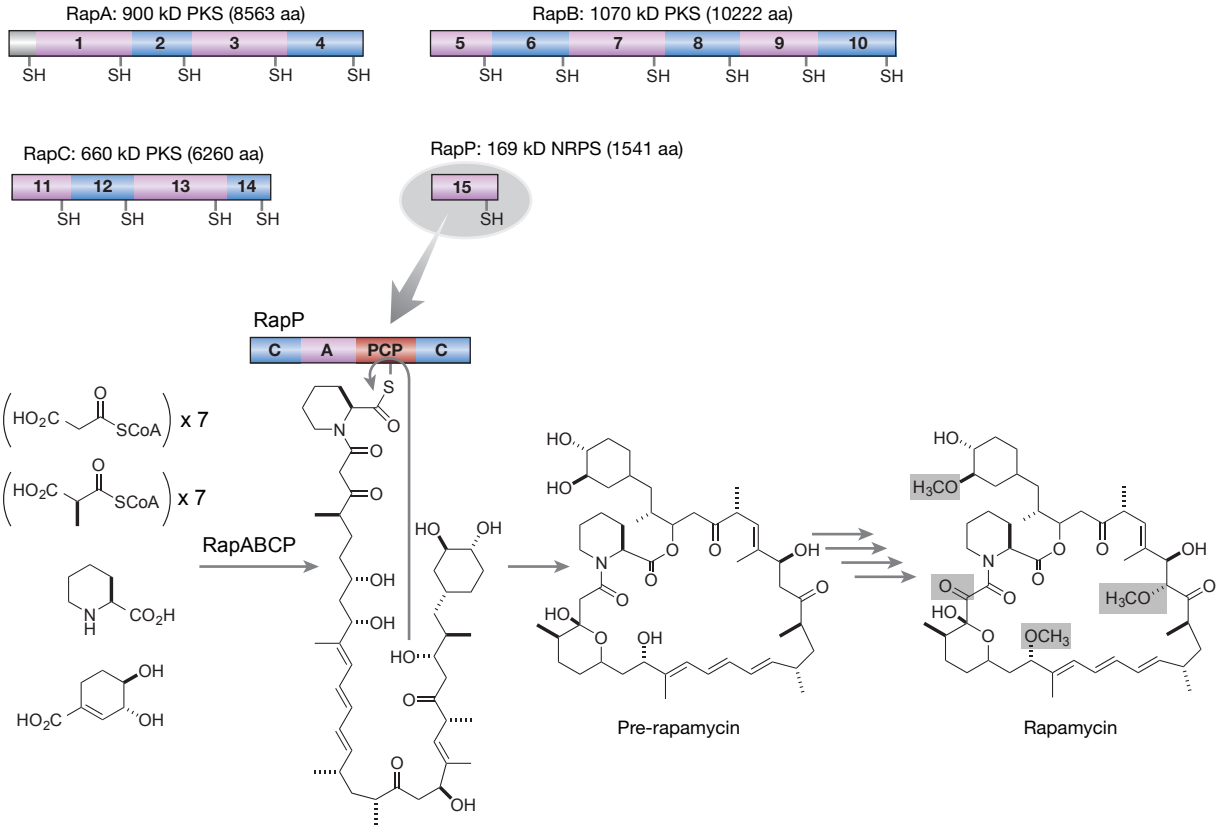


Figure 5 Biosynthesis of natural products. The rapamycin synthase assembly line consists of four multimodular proteins (RapA, RapB, RapC and RapP). Fourteen polyketide synthase modules are distributed in RapA–C and the fifteenth, a nonribosomal peptide synthase module (NRPS), comprises the RapP protein. RapA–C comprise the three-subunit assembly-line machinery for the polyketide-chain initiation and elongation. Each of the 15 modules has a carrier-protein domain (peptidyl carrier protein, PCP in RapP). This is post-translationally modified with a phosphopantetheinyl arm containing a terminal cysteine on which the elongating acyl chains are assembled. The most downstream acyl intermediate is shown on the PCP domain of RapP as it undergoes an intramolecular cyclization, thought to be catalysed by the second condensation domain (C) of RapP. The first C domain makes the acyl–N linkage to the pipecolyl moiety of the acyl chain, while the adenylation domain (A) selects, activates and incorporates the pipecolyl moiety. All the atoms of pre-rapamycin come from the four building blocks malonyl CoA, methylmalonyl CoA, pipecolate and dihydroxycyclohexenoate, as shown. After cyclo-release from the assembly line, pre-rapamycin undergoes a series of oxidative and *O*-methylation-tailoring steps to yield the final product: rapamycin.

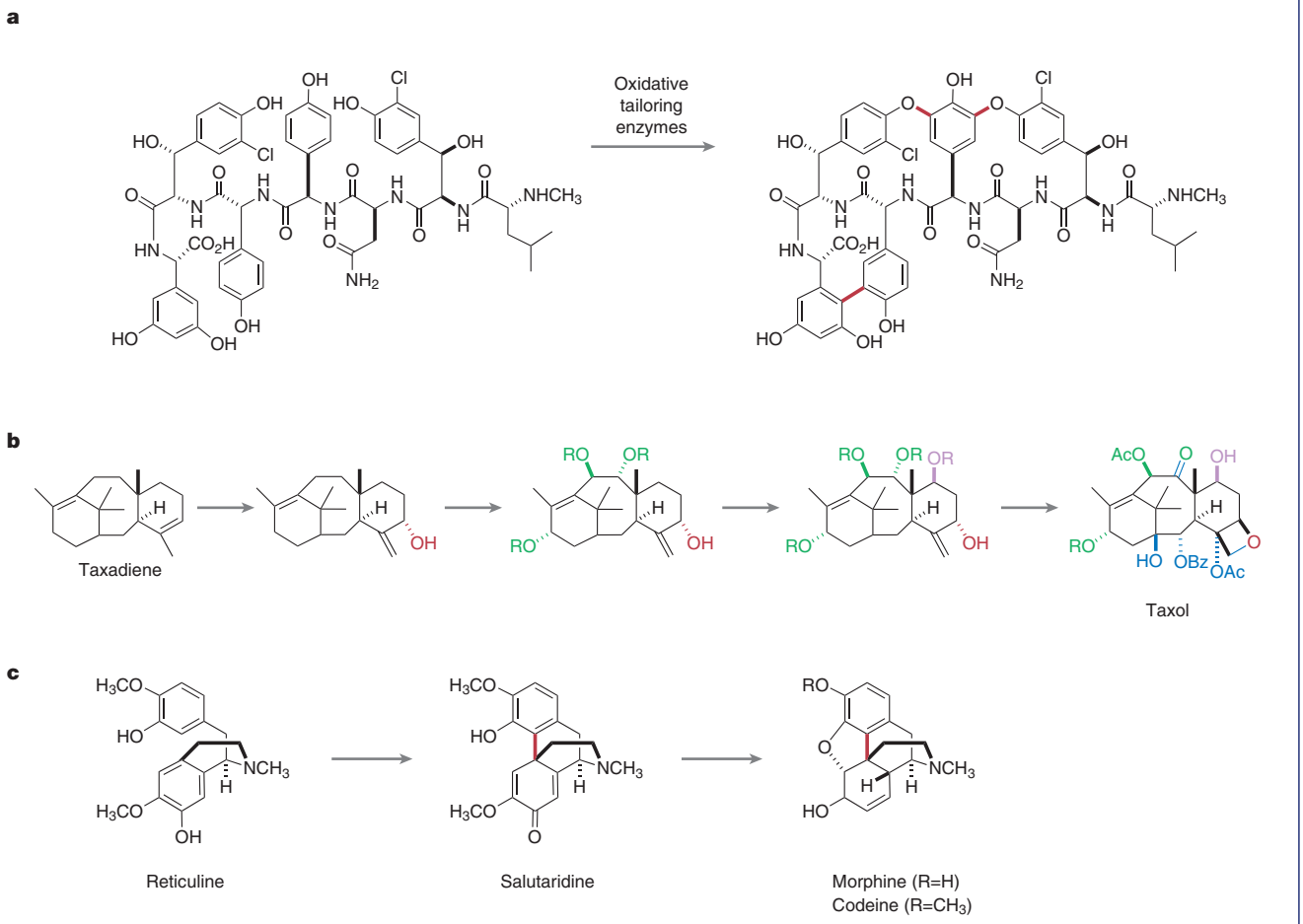


Figure 6 The role of oxidation in the construction of natural products. **a**, The oxidative tailoring of vancomycin by three haem-containing proteins introduces aryl ether (C–O) bonds and aromatic (C–C) crosslinks (shown in red) that rigidify the vancomycin skeleton. **b**, The spectacular series of oxidations that convert taxadiene to Taxol. Eight oxygen atoms are introduced into the scaffold by cytochrome P450 mono-oxygenases, and these are further modified into carbonyl, ether or ester links. The intermediates shown have been identified, but not all the responsible enzymes have been characterized; some of the transformations require more than one enzyme[57,82]. **c**, A key step in the biosynthesis of morphine and other opium alkaloids involves the oxidative coupling of two phenol radicals to form the key bond shown in red.

groups to mask the reactivity of ketones, alcohols and olefins. In particular, the alcohol side chains in the nascent products from a polyketide assembly line come from the carbonyl groups of malonyl or methylmalonyl monomers. However, additional hydroxyl groups (derivatives of oxygen) are introduced by tailoring mono-oxygenases that act with regio- and stereospecificity. A spectacular case of post-assembly-line oxidative tailoring logic occurs in vancomycin biosynthesis, where three haem proteins, all encoded in the biosynthetic gene cluster, act in temporal and regiochemical sequence to introduce the 4–6 and 2–4 aryl ether[54], and 5–7 (C–C) crosslinks[55] in the aglycone scaffold (Fig. 6a). These crosslinks generate the rigid architecture necessary for high-affinity recognition of the *N*-acyl-D-Ala-D-Ala termini of bacterial peptidoglycan strands.

Baldwin noted at a recent Horizon Symposium (http://www.horizonsymposia.com) that the oxygenative maturation of the taxane skeleton to Taxol reveals a comparable enzymatic strategy of assembling the taxane scaffold in a reduced oxidation state and then conducting regiospecific and stereospecific enzymatic oxidations (Fig. 6b). The initial cyclization product from the C_{20} isoprenoid geranylgeranyl pyrophosphate is taxa-4(5),11(12)-diene (ref. 56). This intermediate undergoes eight specific hydroxylations by cytochrome P450 mono-oxygenases[57,58]. Four of the newly introduced hydroxyls are then enzymatically acylated, allowing precisely controlled oxidation on the periphery of the tetracyclic scaffold.

A third example of late-stage redox tailoring is found in the reticuline to salutaridine to morphine pathway (Fig. 6c). These examples of late-stage redox tailoring contrast with a chemist's approach towards total synthesis. Here, fragments are prepared using a convergent, not a linear, approach; the fragments have protecting groups that can be orthogonally manipulated, and the desired final oxidation states are mainly built into the strategy of fragment construction. As a result, synthetic chemists have a much larger set of building blocks with which to carry out their convergent strategies. Despite this, the regio- and stereospecific hydroxylation of related carbon centres in complex molecular scaffolds remain synthetic challenges for which the chemist cannot readily mimic the natural enzymatic process.

Re-engineering of biosynthetic pathways

The burgeoning database of microbial genomes has led to the cataloguing of hundreds of gene clusters that encode polyketides, NRPs and hybrid polyketide–NRP natural products[59]. The coding logic can be deciphered in some cases to make good predictions of what dedicated metabolites will be used as monomers for the assembly lines, what the structures of advanced intermediates will be, and whether post-assembly-line tailoring steps, such as methylations, acylations, glycosylations and oxidations (including hydroxylations), are encoded. These create a set of catalytic-part lists for engineering new polyketide, peptide and hybrid 'unnatural' natural products by

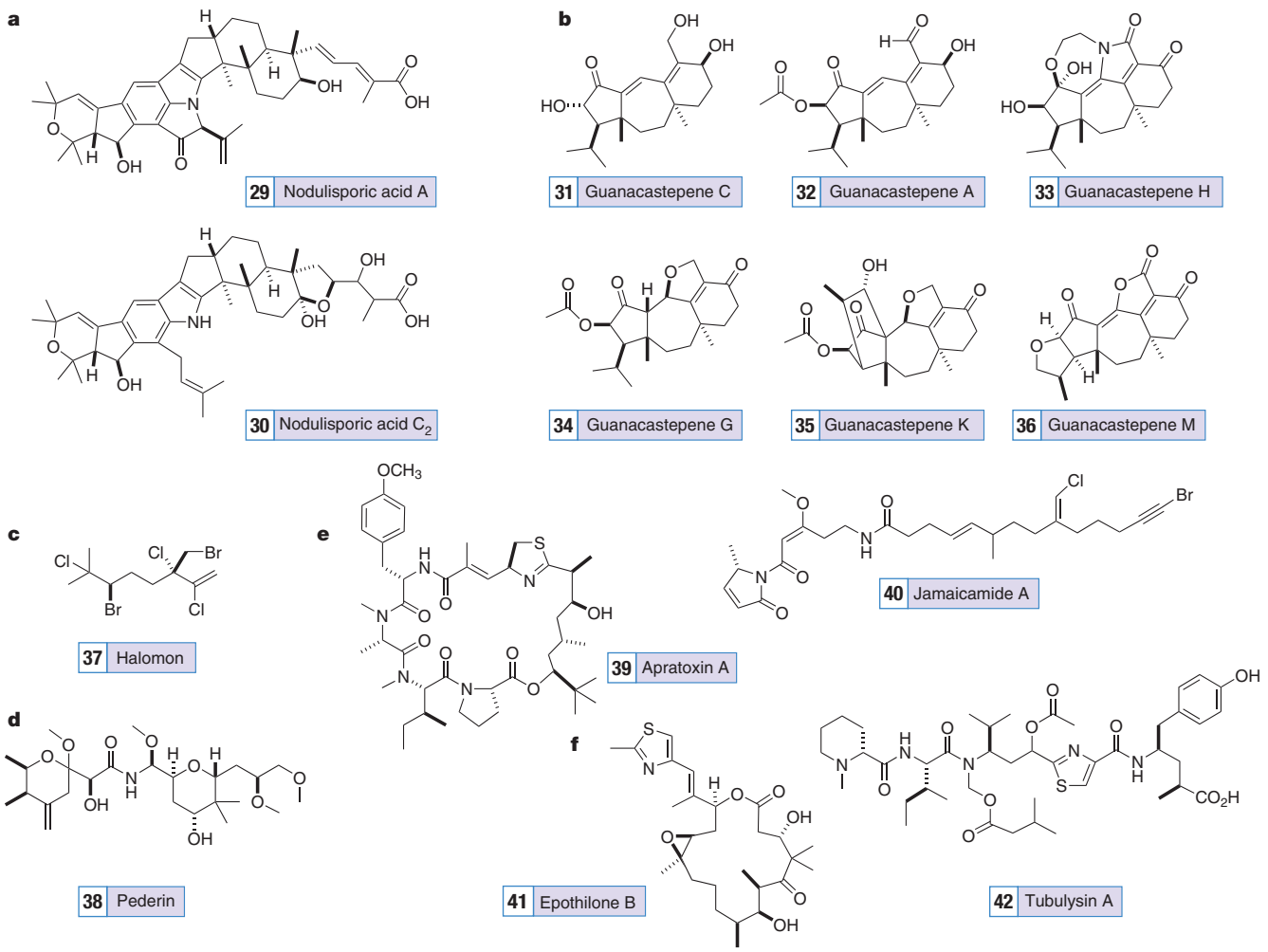


Figure 7 Recent natural products obtained from nontraditional sources. **a**, The nodulisporic acids (**29**, **30**) and **b**, the guanacastepenes (**31**–**36**) are from endophytic fungi, the large group of fungi that live inside higher plants; **c**, halomon (**37**) is from a red alga; **d**, pederin (**38**), which was long believed to be an insect metabolite, is produced by bacteria; **e**, apratoxin (**39**) and jamaicamide (**40**) are from marine cyanobacteria; **f**, epothilone (**41**) and tubulysin (**42**) are from myxobacteria.

domain deletions, mutations and swaps[43,60]. Glycosylation variants in anthracycline anti-tumour molecules[61] and in glycopeptide antibiotics of both the vancomycin and teicoplanin scaffolds have been reported *in vitro*[62], whereas engineered glycosylations of tetracyclic aromatic polyketides have been conducted *in vivo*[63]. Dozens of variants of the deoxyerythronolide scaffold in the erythromycin family have been produced by re-engineering up to three catalytic domains at a time in the three-subunit deoxyerythronolide B (DEB) synthase[64]. This is followed by tailoring glycosylation[65].

With current gene-synthesis technology, it is possible to make assembly lines, for example, for DEB synthase, using dozens of designed restriction sites. These allow chemists to swap a given domain or module with synthetic or natural genetic variants from any other assembly line. Gene-shuffling methodologies similarly increase diversity, so large libraries of variant synthases for polyketides, NRPs and polyketide–NRP hybrids (rapamycin, FK506, bleomycin and epothilones, for example) can probably be constructed and their yields of new products determined by structure-based and/or activity-driven screens. The feeding of alternate monomers into native and engineered assembly lines also leads to new natural-product variants[66].

The engineering of *Escherichia coli* to express the key taxadiene intermediate in the assembly of Taxol[67] and the sesquiterpene amorphadiene — a precursor to the anti-malarial agent artemisinin[68] — have been described, thus indicating that the reconstruction of regiospecific terpene cyclization machinery can also be accomplished.

Discovery from new sources

Each time chemists are able to access new swathes of biological diversity, new — often strikingly new — natural products are discovered. Indeed, what are currently the most interesting natural products come mainly from recently accessed biota. The realization that there was a large, and largely unexplored, group of fungi (endophytic fungi) living inside higher plants led to focused discovery efforts in both industrial and academic laboratories. The nodulisporic acids (**29** and **30**, Fig. 7a) were discovered in an endophytic fungus from Hawaii[69]. The guanacastepenes (**31**–**36**, Fig. 7b) were isolated from an endophytic fungus from Costa Rica using an antibiotic assay[70]. The guanacastepenes provide an elegant illustration of nature's ability to use late-stage redox reactions to re-model core structures and produce a suite of diverse molecular skeletons. This core diversity differs from a typical synthetic combinatorial library featuring only peripheral modifications.

Exploration of the marine environment has also had a profound effect on natural-products chemistry. Early investigations focused on highly halogenated metabolites such as halomon[71] (**37**, Fig. 7c) from a red alga, but many of the most structurally intriguing and biologically

potent molecules, such as discodermolide (**18**) and hemiasterlin (**19**), have come from sponges. Sponges are full of bacterial symbionts, and many sponge metabolites probably have bacterial origins. An interesting preliminary study has shown that pederin (**38**, Fig. 7d) — a well-known insect metabolite with a very similar structure to that of several sponge metabolites — has a bacterial origin[72].

Other productive new sources include cyanobacteria, as represented by apratoxin A[73] (**39**), a potent cytotoxin with an unknown mechanism of action, and jamaicamide[74] (**40**), a potent neurotoxic sodium-channel blocker (Fig. 7e). Both apratoxin A and jamaicamide were isolated from the cyanobacteria *Lyngbya majuscula* (one strain from Guam, the other from Jamaica). Myxobacteria (gliding bacteria) have also been excellent producers of structurally interesting and biologically active natural products. Derivatives of epothilone[75] (**41**) from *Sorangium cellulosum* are currently being used in cancer trials, and tubulysins[76] (**42**) from *Archangium gephyra* are potent tubulin disruptors with potential anti-cancer activity (Fig. 7f).

As the examples from cyanobacteria and myxobacteria suggest, there are still many natural products to be discovered from bacteria. This biosynthetic potential is not surprising because bacteria encompass the main pool of genetic diversity on the planet; they interact with their surroundings, competitors and community members through small molecules, and they are largely unexplored. Fewer than 1% of the bacteria on Earth, and probably fewer than 0.1%, have ever been cultured[77]. Most bacteria live in microbial communities where the members are mutually dependent on each other, and because current culturing practices select for strains that can live on their own, most bacteria are not cultured[78]. Several approaches to dealing with uncultured bacteria have been proposed, including reconstituting the natural communities[78] and capturing biosynthetic gene clusters directly from DNA taken from the environment[79,80]. Finally, the wealth of bacterial genomic data now available emphasizes that there are many biosynthetic gene clusters in culturable bacteria for which no associated small molecule can be isolated. Natural-product production is a highly regulated process and these cryptic pathways are not turned on under standard culturing conditions. A genomics-guided approach to discovering, sequencing and expressing these pathways has been described[81].

Conclusions

The inventory of natural molecules remains incomplete, and discoveries of new structures and functions are likely to continue as underexplored sources of natural products are more systematically evaluated. The functional-group diversity and architectural platforms engineered into natural products during biosynthesis continue to provide lessons for synthetic and medicinal chemists in their strategies for making biologically active mimics, and provide selective ligands for cellular targets. Deciphering the molecular logic of biosynthetic enzymes and pathways, as monomers are assembled and nascent products tailored, has opened up practical approaches to re-engineering assembly lines to create unnatural variants of natural products. The molecular scaffolds created and used in nature are likely to persist as central design elements in subsequent generations of synthetic and semi-synthetic ligands that could become therapeutic agents for receptors, enzymes and ion channels.

Finally, although there has been a trend within the pharmaceutical industry to downscale efforts in natural-products research in recent years, careful reconsideration of this area could change this. Several problems with natural products that influenced the original company decisions to withdraw from the field (such as the challenges associated with identifying the active components from natural-product extracts that typically contain several compounds) are being addressed by technological advances. For example, the throughput of methods for compound purification and identification has increased. It seems clear that there is still great potential for accessing therapeutically relevant chemical diversity from nature — in particular, from the many organisms that have not yet been cultured. A revival in interest in using natural products in early-stage drug discovery could be exactly what is needed to boost pharmaceutical output. □

doi:10.1038/nature03194

1. Newman, D. J., Cragg, G. M. & Snader, K. M. The influence of natural products upon drug discovery. *Nat. Prod. Rep.* **17,** 215–234 (2000).
2. Newman, D. J., Cragg, G. M., Holbeck, S. & Sausville, E. A. Natural products and derivatives as leads to cell cycle pathway targets in cancer chemotherapy. *Curr. Cancer Drug Targets* **2,** 279–308 (2002).
3. Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as sources of new drugs over the period 1981–2002. *J. Nat. Prod.* **66,** 1022–1037 (2003).
4. Altmann, K.-H. Microtubule-stabilizing agents: a growing class of important anticancer drugs. *Curr. Opin. Chem. Biol.* **5,** 424–431 (2001).
5. Mickel, S. J. *et al.* Large-scale synthesis of the anti-cancer marine natural product (+)−discodermolide. Part 1: Synthetic strategy and preparation of a common precursor. *Org. Proc. Res. Dev.* **8,** 92–100 (2004).
6. Sheldrick, G. M., Jones, P. G., Kennard, O., Williams, D. H. & Smith, G. A. Structure of vancomycin and its complex with acetyl-D-alanyl-D-alanine. *Nature* **271,** 223–225 (1978).
7. Williamson, M. P. & Williams, D. H. Structure revision of the antibiotic vancomycin. Use of nuclear Overhauser effect difference spectroscopy. *J. Am. Chem. Soc.* **103,** 6580–6585 (1981).
8. Furasaki, A. *et al.* The crystal and molecular structure of staurosporine, a new alkaloid from a *Streptomyces* strain. *J. Chem. Soc. Chem. Commun.* 800–801 (1978).
9. Sehgal, S. N., Baker, H. & Vezina, C. Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *J. Antibiot.* **28,** 727–732 (1975).
10. Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. The isolation and structure of taxol, a novel antileukemic and antitumor agent. *J. Am. Chem. Soc.* **93,** 2325–2327 (1971).
11. Pindur, U., Kim, Y. S. & Mehrabani, F. Advances in indolo[2,3-a]carbazole chemistry: design and synthesis of protein kinase C and topoisomerase I inhibitors. *Curr. Med. Chem.* **6,** 29–69 (1999).
12. Carson, C. C. 3rd Sildenafil: a 4-year update in the treatment of 20 million erectile dysfunction patients. *Curr. Urol. Rep.* **4,** 488–496 (2003).
13. Wong, D. T. & Bymaster, F. P. Development of antidepressant drugs. Fluoxetine (Prozac) and other selective serotonin uptake inhibitors. *Adv. Exp. Med. Biol.* **363,** 77–95 (1995).
14. Roth, B. D. The discovery and development of atorvastatin, a potent novel hypolipidemic agent. *Prog. Med. Chem.* **40,** 1–22 (2002).
15. Kurzrock, R., Kantarjian, H. M., Druker, B. J. & Talpaz, M. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann. Intern. Med.* **138,** 819–830 (2003).
16. Lee, M. D., Dunne, T. M., Chang, C. C., Morton, G. O. & Borders, D. B. Calicheamicins, a novel family of antitumor antibiotics. *J. Am. Chem. Soc.* **109,** 3464–3466 (1987).
17. Konishi, M. *et al.* Crystal and molecular structure of dynemicin A: a novel 1,5-diyn-3-ene antitumor antibiotic. *J. Am. Chem. Soc.* **112,** 3715–3716 (1990).
18. Thorson, J. S. *et al.* Understanding and exploiting nature's chemical arsenal: the past, present and future of calicheamicin research. *Curr. Pharm. Des.* **6,** 1841–1879 (2000).
19. Scotto, K. W. ET-743: more than an innovative mechanism of action. *Anticancer Drugs* **13** (Suppl. 1), S3–6 (2002).
20. Aune, G. J., Furuta, T. & Pommier, Y. Ecteinascidin 743: a novel anticancer drug with a unique mechanism of action. *Anticancer Drugs* **13,** 545–555 (2002).
21. Liu, S., Widom, J., Kemp, C. W., Crews, C. M. & Clardy, J. Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science* **282,** 1324–1327 (1998).
22. Meng, L., Kwok, B. H., Sin, N. & Crews, C. M. Eponemycin exerts its antitumor effect through the inhibition of proteasome function. *Cancer Res.* **59,** 2798–2801 (1999).
23. Fenteany, G. *et al.* Inhibition of proteasome activities and subunit-specific amino-terminal threonine modification by lactacystin. *Science* **268,** 726–731 (1995).
24. Feling, R. H. *et al.* Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *salinospora*. *Angew. Chem. Int. Edn Engl.* **42,** 355–357 (2003).
25. Kim, M. Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D. & Hurley, L. H. The differential biological effects of telomestatin and TMPPyP4 can be attributed to their selectivity for interaction with intramolecular and intermolecular G-quadruplex structures. *Cancer Res.* **63,** 3247–3256 (2003).
26. Choi, J., Chen, J., Schreiber, S. L. & Clardy, J. Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273,** 239–242 (1996).
27. Gunasekera, S. P., Gunasekera, M., Longley, R. E. & Schulte, G. K. Discodermolide: a new bioactive polyhydroxylated lactone from the marine sponge *Discodermia dissoluta*. *J. Org. Chem.* **55,** 4912–4915 (1991).
28. Talpir, R., Benayahu, Y., Kashman, Y., Pannell, L. & Schleyer, M. Hemiasterlin and geodiamolide TA: two new cytotoxic peptides from the marine sponge *Hemiasterella minor*. *Tetrahedron Lett.* **35,** 4453–4456 (1994).
29. Anderson, R. J., Coleman, J. E., Andersen, R. J. & Roberge, M. Cytotoxic peptides hemiasterlin, hemiasterlin A and hemiasterlin B induce mitotic arrest and abnormal spindle formation. *Cancer Chemother. Pharmacol.* **39,** 223–226 (1997).
30. Schantz, E. J. *et al.* The structure of saxitoxin. *J. Am. Chem. Soc.* **93,** 7344–7345 (1975).
31. Yotsu-Yamashita, M. *et al.* The structure of zetekitoxin AB, a saxitoxin analog from the Panamanian golden frog *Atelopus zeteki*: a potent sodium-channel blocker. *Proc. Natl Acad. Sci. USA* **101,** 4346–4351 (2004).
32. Rinehart, K. L. *et al.* Ecteinascidins 729, 743, 745, 759A, 759B and 770: potent antitumor agents from the Caribbean tunicate *Ecteinascidia turbinata*. *J. Org. Chem.* **55,** 4512–4515 (1990).
33. Boghaert, E. R. *et al.* Antibody-targeted chemotherapy with the calicheamicin conjugate hu3S193-N-acetyl gamma calicheamicin dimethyl hydrazide targets Lewisy and eliminates Lewisy-positive human carcinoma cells and xenografts. *Clin. Cancer Res.* **10,** 4538–4549 (2004).
34. DiJoseph, J. F. *et al.* Antibody-targeted chemotherapy with CMC-544: a CD22-targeted immunoconjugate of calicheamicin for the treatment of B-lymphoid malignancies. *Blood* **103,** 1807–1814 (2004).
35. Boldi, A. M. Libraries from natural product-like scaffolds. *Curr. Opin. Chem. Biol.* **8,** 281–286 (2004).
36. Gunasekera, S. P., McCarthy, P. J., Kelly-Borges, M., Lobkovsky, E. & Clardy, J. Dysidiolide: a novel protein phosphatase inhibitor from the Caribbean sponge *Dysidea etheria de Laubenfels*. *J. Am. Chem. Soc.* **118,** 8759–8760 (1996).

37. Brohm, D. *et al.* Natural products are biologically validated starting points in structural space for compound library development: solid-phase synthesis of dysidiolide-derived phosphatase inhibitors. *Angew. Chem. Int. Edn Engl.* **41,** 307–311 (2002).
38. Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **287,** 1964–1969 (2000).
39. Burke, M. D. & Schreiber, S. L. A planning strategy for diversity-oriented synthesis. *Angew. Chem. Int. Edn Engl.* **43,** 46–58 (2004).
40. Burke, M. D., Berger, E. M., Schreiber, S. L. Generating diverse skeletons of small molecules combinatorially. *Science* **302,** 613–618 (2004).
41. Walsh, C. *Antibiotics: Actions, Origins, Resistance* (ASM, Washington, 2003).
42. Hubbard, B., Walsh, C. Vancomycin assembly: Nature's way. *Angew. Chem. Int. Edn Engl.* 730–765 (2003).
43. Schwarzer, D., Firking, R., Marahiel, M. A. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20,** 275–287 (2003).
44. Walsh, C., Freel Meyers, C. L., Losey, H. C. Antibiotic glycosyltransferases: antibiotic maturation and prospects for reprogramming. *J. Med. Chem.* **46,** 3425–3436 (2003).
45. Wendt, K. U., Schulz, G. E., Corey, E. J., Liu, D. R. Enzyme mechanisms for polycyclic triterpene formation. *Angew. Chem. Int. Edn Engl.* **39,** 2812–2833 (2000).
46. Whittington, D. A. *et al.* Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci. USA* **99,** 15375–15380 (2002).
47. Rawlings, B. J. Type I polyketide biosynthesis in bacteria (part A). *Nat. Prod. Rep.* **18,** 190–230 (2001).
48. Rawlings, B. J. Type I polyketide biosynthesis in bacteria (part B). *Nat. Prod. Rep.* **18,** 231–281 (2001).
49. Kohli, R. & Walsh, C. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun.* 297–307 (2003).
50. Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl Acad. Sci. USA* **92,** 7839–7843 (1995).
51. Ahlert, J. *et al.* The calicheamycin gene cluster and its iterative type I enediyne PKS. *Science* **297,** 1173–1176 (2002).
52. Liu, W., Christensen, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297,** 1170–1173 (2002).
53. Shen, B., Liu, W. & Nonaka, K. Enediyne natural products: biosynthesis and prospects towards engineering novel antitumor agents. *Curr. Med. Chem.* **10,** 2317–2325 (2003).
54. Zerbe, K. *et al.* Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *J. Biol. Chem.* **277,** 47476–47485 (2002).
55. Pylypenko, O., Vitali, F., Zerbe, K., Robinson, J. A. & Schlichting, I. Crystal structure of OxyC, a cytochrome P450 implicated in an oxidative C–C coupling reaction during vancomycin biosynthesis. *J. Biol. Chem.* **278,** 46727–46733 (2003).
56. Walker, K. & Croteau, R. Taxol biosynthetic genes. *Phytochemistry* **58,** 1–7 (2001).
57. Chau, M., Jennewein, S., Walker, K. & Croteau, R. Taxol biosynthesis: molecular cloning and characterization of a cytochrome P450 taxoid 7 beta-hydroxylase. *Chem. Biol.* **11,** 663–672 (2004).
58. Jennewein, S., Long, R. M., Williams, R. M. & Croteau, R. Cytochrome p450 taxadiene 5 alpha-hydroxylase, a mechanistically unusual monooxygenase catalyzing the first oxygenation step of taxol biosynthesis. *Chem. Biol.* **11,** 379–387 (2004).
59. Yadav, G., Ghokale, R. S. & Mohanty, D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328,** 335–363 (2003).
60. Eppelman, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **42,** 9718–9716 (2002).
61. Lu, W., Leimkuhler, C., Oberthur, M., Kahne, D. & Walsh, C. AknK is an L-2-deoxyfucosyltransferase in the biosynthesis of the anthracycline aclacinomycin. *Biochemistry* **43,** 4548–4558 (2004).
62. Losey, H. C. *et al.* Incorporation of glucose analogs by glycosyltransferases GtfE and GtfD from the vancomycin biosynthetic pathway to generate variant glycopeptides. *Chem. Biol.* **9,** 1305–1314 (2002).
63. MendeZ. C. & Salas, J. A. Altering the glycosylation pattern of bioactive compounds. *Trends Biotechnol.* **19,** 449–456 (2001).
64. McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin gene cluster to produce a library of novel 'unnatural' natural products. *Proc. Natl Acad. Sci. USA* **96,** 1846–1851 (1999).
65. Tang, L. & McDaniel, R. Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad substrate specificity. *Chem. Biol.* **8,** 547–555 (2001).
66. Wohlleben, W. & Pelzer, S. New compounds by combining modern genomics and old-fashioned mutasysnthesis. *Chem. Biol.* **9,** 1163–1166 (2002).
67. Huang, Q., Roessner, C. A., Croteau, R. & Scott, A. I. Engineering *E. coli* for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Biorg. Med. Chem.* **9,** 2237–2242 (2001).
68. Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in *E.coli* for production of terpenoids. *Nature Biotechnol.* **21,** 796–802 (2003).
69. Ondeyka, J. G. *et al.* Nodulisporic acids C, C1, and C2: a series of D-ring-opened nodulisporic acids from the fungus *Nodulisporium* sp. *J. Nat. Prod.* **66,** 121–124 (2003).
70. Brady, S. F., Bondi, S. M. & Clardy, J. The guanacastepenes: a highly diverse family of secondary metabolites produced by an endophytic fungus. *J. Am. Chem. Soc.* **123,** 9900–9901 (2001).
71. Fuller, R. W. *et al.* A pentahalogenated monoterpene from the red alga *Portieria hornemanni* produces a novel cytotoxicity profile against a diverse panel of human tumor cell lines. *J. Med. Chem.* **35,** 3007–3011 (1992).
72. Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **21,** 519–538 (2004).
73. Luesch, H., Yoshida, W. Y., Moore, R. E., Paul, V. J. & Corbett, T. H. Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula. J. Am. Chem. Soc.* **123,** 5418–5423 (2001).
74. Edwards, D. J. *et al.* Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula. Chem. Biol.* **11,** 817–833 (2004).
75. Gerth, K., Bedorf, N., Hofle, G., Irschik, H. & Reichenbach, H. Epothilons A and B: antifungal and cytotoxic compounds from *Sorangium cellulosum* (Myxobacteria). Production, physico-chemical and biological properties. *J. Antibiot. (Tokyo)* **49,** 560–563 (1996).
76. Sasse, F., Steinmetz, H., Heil, J., Hofle, G. & Reichenbach, H. Tubulysins, new cytostatic peptides from myxobacteria acting on microtubuli. Production, isolation, physico-chemical and biological properties. *J. Antibiot. (Tokyo)* **53,** 879–885 (2000).
77. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276,** 734–740 (1997).
78. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296,** 1127–1129 (2002).
79. Wang, G. Y. *et al.* Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* **2,** 2401–2404 (2000).
80. Brady, S., Chao, C. J. & Clardy, J. New natural product families from an eDNA cluster. *J. Am. Chem. Soc.* **124,** 9968–9969 (2002).
81. Zazopoulos, E. *et al.* A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nature Biotechnol.* **21,** 187–190 (2003).
82. Jennewein, S., Wildung, M. R., Chau, M., Walker, K. & Croteau, R. Random sequencing of an induced Taxus cell cDNA library for identification of clones involved in Taxol biosynthesis. *Proc. Natl Acad. Sci. USA* **101,** 9149–9154 (2004).

Acknowledgements We thank G. J. Gatto and K. N. Maloney for their help in preparing this article.

Competing interests statement The authors declare competing financial interests: details accompany the paper on **www.nature.com/nature**.