Natural products and combinatorial chemistry: back to the future
Jean-Yves Ortholand¹ and A Ganesan²

The introduction of high-throughput synthesis and combinatorial chemistry has precipitated a global decline in the screening of natural products by the pharmaceutical industry. Some companies terminated their natural products program, despite the unproven success of the new technologies. This was a premature decision, as natural products have a long history of providing important medicinal agents. Furthermore, they occupy a complementary region of chemical space compared with the typical synthetic compound library. For these reasons, the interest in natural products has been rekindled. Various approaches have evolved that combine the power of natural products and organic chemistry, ranging from the combinatorial total synthesis of analogues to the exploration of natural product scaffolds and the design of completely unnatural molecules that resemble natural products in their molecular characteristics.

Addresses
¹Merck Sante, 115 Avenue Lacassagne, 69003 Lyon, France e-mail: jean-yves.ortholand@merck.fr
²School of Chemistry, University of Southampton, Southampton SO17 1BJ, UK e-mail: ganesan@soton.ac.uk

Abbreviations
ACD Available Chemicals Database
HDAC histone deacytase
HTS high-throughput screening
SAR structure–activity relationship

Introduction
Since time immemorial, mankind has sourced the flora and fauna around him to ameliorate disease. The accumulated wisdom (if you like, from a worldwide combinatorial search of natural product space) of folklore provided the mainstay of our pharmacopeia for centuries [1]. In the past 100 years, this dependence was irreversibly broken for two reasons. Firstly, advances in biology began to provide insight into the molecular mechanisms underlying diseases, suggesting rational targets for therapeutic intervention. Secondly, advances in organic chemistry enabled the design and synthesis of sophisticated drug molecules. By the 1950s, such purely synthetic medicines were on an equal footing with their natural product counterparts. Nevertheless, natural products were still considered a valuable source of drug leads, and the testing of natural product extracts was widely practiced in the pharmaceutical industry [2,3]. The situation has changed dramatically within the past 20 years. With high-throughput screening (HTS), assay speed far outpaced the rate of compound supply for the first time — a state of affairs that fuelled the birth of combinatorial chemistry. The promise of a seemingly inexhaustible supply of compound libraries precipitated a further shift away from natural products, and many companies terminated their efforts in this direction. Here, we review the rationale behind this strategic decision, and highlight reasons why natural products are coming back into fashion [4,5–7].

The early days of combinatorial chemistry: paradise lost
While the past value of screening natural product extracts is indisputable, it is equally true that it is associated with several drawbacks, chiefly:

1. Expense. Building up and maintaining a high-quality collection of natural product extracts is expensive. This is why it was traditionally undertaken by big pharma or large organizations such as the National Cancer Institute (NCI), and beyond the reach of smaller biotech or startup companies.

2. Time. Once a hit is identified from an extract, bioassay-guided fractionation is needed to identify the active component. This time-consuming process is not always compatible with the present regime of ‘blitz’ screening campaigns where assay support is available for a limited duration.

3. Novelty. Once the active component is isolated, its novelty is unpredictable and requires a further expenditure of time for characterization. In the worst-case scenario, the lead may turn out to be a well-known natural product that cannot be patented, or one already protected by a competitor. Meanwhile, if it is a novel natural product, full structural elucidation can be a lengthy exercise.

4. Tractability. Natural products are often structurally complex. With such molecules, deriving meaningful structure–activity relationships (SARs) or identifying a pharmacophore is a significant hurdle for the medicinal chemist. Even a natural product of low to medium complexity is likely to lose out to a similarly potent hit of synthetic origin, as the latter comes with a preparative route already in place.
Given these issues, the siren call of combinatorial chemistry proved too seductive. Although the technology was in its infancy and had no track record, its potential to rapidly deliver very large numbers of novel compounds was persuasive to management. Combinatorial libraries began to compete with, and then overtake, natural product extracts for screening resources. Organizations such as GlaxoWellcome, SmithKlineFrench, and Pfizer, whose product portfolio was predominantly based on synthetic drugs, eventually phased out natural product screening altogether. Others such as Merck and Novartis, with recent blockbuster natural products such as mevinolin and cyclosporin, respectively, have continued to maintain a presence in the area.

**Natural products redux, or paradise regained**

The global repercussions of the crisis in natural product screening were considerable, and appeared to herald a permanent fall from grace. However, there has been a welcome reassessment recently, and a growing recognition that it was premature to abandon natural products in drug discovery. First and foremost, the promised avalanche of new drugs from high-throughput technologies such as combinatorial chemistry has yet to materialize. In fact, the number of new chemical entities reaching the market reached a two-decade minimum in 2002, with only a marginal improvement in 2003.

The reasons for the continuous decline of productivity within the pharmaceutical industry are complex and controversial. The perceived failure of combinatorial chemistry, for example, was primarily due to unrealistic expectations. Provided a large enough library was screened, it was assumed that useful leads would emerge, causing an unhealthy emphasis on compound numbers rather than quality or purity. The actual composition of the early libraries was debatable. Furthermore, since the most reliable reactions for high-throughput chemistry involve functional group interconversions, the easiest compounds to make were oligomers such as peptides or nucleotides that do not resemble small molecule drugs. Undoubtedly, the community has learned from these early mistakes. The attention has now shifted to smaller high-quality libraries [8] of discrete compounds, using filters for lead-like [9,10] or drug-like properties [11,12], as well as avoiding non-selective or promiscuous inhibitors [13,14*.15]. Tremendous advances in analytical chemistry have enabled the routine high-throughput purification of compounds by mass-triggered preparative liquid chromatography [16,17]. We are starting to see drug candidates impacted by combinatorial chemistry in either the discovery or optimization phase [18,19]. In the future, it will probably be difficult to point to any synthetic drugs that did not benefit from modern high-throughput technologies at some stage of the process from lead discovery to the clinic.

Although the quality of leads discovered directly from a combinatorial library is steadily increasing, it is clear that parallel synthesis is even more powerful at the lead optimization phase, when a series of related compounds needs to be made and tested rapidly. The real issue then is where the initial leads are going to come from. Does screening natural products interrogate a unique and complementary region of chemical space compared with that of synthetic compounds? Several groups from big pharma — Bayer [20] and Roche [21*], as well as smaller companies New Chem Entities [22] and SignalGene [23*], have attempted to answer this question by statistical comparisons between therapeutically relevant molecules and natural product collections. While the databases and computational methods used varied in each study, the overall conclusions are remarkably convergent (Table 1). In terms of Lipinski’s ‘Rule of Five’, it is clear that the properties of the average natural product or drug are quite similar. The Roche study, for example, found that 10% of trade drugs contained two or more violations of the rules, as opposed to 12% for the natural products database.

How do natural products differ from drugs? They tend to contain a different molecular composition, containing fewer nitrogen, halogen or sulfur atoms on average, but are considerably more oxygen-rich, and contain more hydrogen bond donors. Compared with drugs or synthetic compounds, they are likely to contain a larger number of rings, significantly more chiral centres, and have sp³ hybridised bridgehead atoms present. Overall, natural

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean values for a selection of molecular properties among natural, drug and synthetic compounds.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Molecular weight</td>
</tr>
<tr>
<td>LogP</td>
</tr>
<tr>
<td>Number of chiral centers</td>
</tr>
<tr>
<td>Number of N atoms</td>
</tr>
<tr>
<td>Number of O atoms</td>
</tr>
<tr>
<td>% of rings that are aromatic</td>
</tr>
</tbody>
</table>

*Values amalgamated from computations in [20,21*,22,23*].
products can be considered high in ‘sterical complexity’. In part, this is an obvious consequence of the fact that the enzymes used for biosynthesis, as well as the molecular targets the natural product is meant to interact with, are inherently three-dimensional and chiral. Furthermore, nature has a limited palette of building blocks at its disposal, and thus has to generate novelty by branching out common intermediates into diverse scaffolds (Figure 1). Conversely, combinatorial and medicinal chemists, with the luxury of the Available Chemicals Database (ACD), generate diversity by doing the same sequence of reactions using different reagent inputs.

Two representations of natural product space from these computational studies nicely illustrate the complementarity with synthetic compounds and drugs (Figures 2,3). The first, from Feher and Schmidt [23*], is a principal component analysis comparing a random selection of combinatorial compounds from commercial suppliers against a natural product and drug database. The natural product dataset occupies a substantially different and larger space than that from combinatorial chemistry. This view is corroborated by Lee and Schneider’s scaffold analysis [21*] of drugs versus natural products. They find only a moderate structural overlap between scaffolds present in the natural product and drug collections considered. Not only does the natural product database contain many more scaffolds, but an important proportion of these ring systems are not found at all in the drugs database. Such unexploited scaffolds may be promising new starting points in drug discovery.

**The way forward: seven routes to success with natural products**

The previous section stated the case for including natural products for drug discovery. As a class, these compounds do display different characteristics compared with the average drug, or average synthetic combinatorial library. How can we harness the enormous potential of natural product space? Here, we outline a series of different approaches, ranging from the classical screening of extracts to the conception of purely synthetic libraries that are ‘natural-product-like’ in design.

**Natural product screening**

In response to the competition from synthetic libraries, the process of natural product screening has certainly been streamlined [24]. The mechanics of extract preparation and bioassay-guided fractionation are increasingly automated, while there are efforts to avoid crude extracts altogether, relying instead on partially or fully purified materials upfront. An important mitigating factor with these bottlenecks is the inherent value of natural product screening in providing ‘home runs’. While synthetic drugs are usually the result of numerous structural modifications over the course of an extensive drug discovery program, a natural product can go straight from hit to drug, as is most apparent in the antimicrobial and anticancer therapeutic...
Prospecting for natural drugs with these biological activities makes sense evolutionarily, as such compounds are likely to be of direct ecological benefit to the producing organism in competition for resources or avoiding predation. Outside these areas, it may appear illogical for a soil microorganism to produce a specific modulator of a human protein with which it never comes in contact. Nevertheless, the history of drug discovery is rife with such examples, like the highly successful statin HMG CoA reductase inhibitors. The reasons for this happy coincidence probably lie in the finite nature of biological space. Most drugs target proteins or nucleic acids, which are composed of the same building blocks in all species. Thus, while the natural product’s original target and the modern assay can be unrelated, they may contain structural domains that are similar in shape.

A growing trend in industrial natural-product screening is to outsource this activity. For smaller startups, as well as big pharma that moved away from natural products, this enables a cost-efficient means of accessing the diversity of natural products without the expenditure of setting up the effort internally. Collections of natural products are available from specialist suppliers, either for a fee or via risk- and profit-sharing agreements. Some of these suppliers have in fact acquired big pharma natural product archives that were disposed of, such as Merlion in Singapore (GlaxoSmithKline collection) and Albany Molecular Research in the USA (Lilly collection).

The actual pace of natural product isolation in both industry and academia continues unabated. Every year, thousands of new natural products are reported. Of these, only a very small minority will receive further attention. Natural products with unusual structural features attract organic chemists looking for total synthesis targets, while those with promising biological activity are leads for drug discovery. With regards to the latter, the natural-products community does itself a disservice as only the structures of new compounds are often reported, without any associated biological testing. When such data are available, a high proportion report only a single assay such as antibacterial or cytotoxic activity. These are generic assays with low information value.

**Figure 2**

The plot of the first two principal components, obtained from a database containing (a) a random selection of combinatorial compounds ($n = 13,506$), (b) natural products ($n = 3,287$), and (c) drugs ($n = 10,968$). For clarity, the data points from the three databases are plotted separately but on the same axes. The first two components explain about 54% of the variance. The figure shows that combinatorial compounds cover a well-defined area in the diversity space given by these principal components. In contrast, natural products and drugs cover almost all of this space as well as a much larger additional volume. Drugs and natural products have approximately the same coverage of this space. Modified from [23] with permission. Copyright 2003, J Chem Inf Comput Sci.

**Figure 3**

A selection of ring systems found in a natural product database but not in trade drugs.
content, and provide no information on mechanism of action or selectivity.

**‘Unnatural’ natural products**

While the traditional screening of natural product extracts will continue to reveal exciting leads, another avenue is the manipulation of biosynthetic pathways to produce novel natural products. Engineered biosynthesis, although in its infancy, has already demonstrated several successes with the polyketide and nonribosomal peptide class of natural products [26]. Kosan, for example, have modified the pathway of epothilone biosynthesis to create a recombinant microorganism that produces epothilone D [27], initially prepared synthetically by the Danishefsky group. In the area of marine natural products, the producing organism is often a symbiont, and it is possible to identify the responsible gene clusters followed by recombinant expression, as investigated for bryostatin biosynthesis [28].

Besides the artificial manipulation of known biosynthetic pathways, two other approaches are likely to gain in prominence. Firstly, there is evidence that several gene clusters are normally silent. Thus, a natural product extract may not encompass the total potential of the producing organism [29,30]. Recreation of this ‘metabolome’ elsewhere may enable completely new natural products to be found. The second strategy does away with the concept of producing organism altogether. Despite the vast numbers of microbial natural products found in the past, it is estimated that < 1% of soil microorganisms can be successfully cultured in the laboratory. With advances in molecular biology, we can now directly extract the DNA from such unculturable species in the soil, and examine their metabolic pathways. It is open to question whether natural products thus identified are normally produced in nature at all.

**Natural products as a source of building blocks**

Some natural products are available cheaply and abundantly. Can these be degraded to provide a set of chiral building blocks that are then assembled into novel combinatorial scaffolds? As yet, this type of natural product prospecting is poorly documented in the literature. A pioneering effort from Niggemann et al. [31] describes the fragmentation of soraphen A, mycotoxazol A, sorangicin A, epothilone A, ambruticin S and apicularen A — all natural products isolated from myxobacteria at the German Gesellschaft für Biotechnologische Forschung — to a set of 23 primary alcohol building blocks. These can then be linked to amines via a carbamate (Figure 4a). Clough et al. at Affymax have recently reported [32] the degradation of the thiazole antibiotic GE2270 A, followed by the combinatorial synthesis of a series of A-ring modifications. Similarly, Dong et al. at Kosan have reconstituted [33] epothilone D from its degradation fragments, thus facilitating the synthesis of unnatural analogues.

**The derivatization of natural products**

Besides degradation to fragments, readily available natural products can be exploited by combinatorial derivatization. Affymax have published several examples, targeting the yohimbine alkaloids [34] (Figure 4b) and the antibiotic cycloserine [35]. More recently, Lambert and co-workers have employed [36] polymer-supported reagents to derivatize tropane alkaloids such as scopoline. In an innovative study, Nicolaou prepared [37,38] a series of synthetic vancomycin dimers by disulfide interchange and olefin metathesis. The library was constructed dynamically in the presence of the d-Ala-d-Ala binding partner, and dimers with a higher affinity for the target were amplified.

**Natural-product analogues by total synthesis**

Here, the primary objective is to increase our knowledge of the natural product’s SAR to discover improved analogues with enhanced biological and/or pharmacokinetic properties. With sufficient resources, complex natural products can be investigated in this manner. Two classic cases from industry are the Merck process group’s total synthesis of the immunosuppressant FK-506, and the recent Novartis gram-scale synthesis [39] of discodermolide for clinical trials. In the academic sector, intensive efforts from the Danishefsky [40] and Nicolaou [41] groups have contributed greatly to the medicinal chemistry of epothilone.

Besides these large-scale exercises, a fruitful venture for combinatorial chemists is the design of concise and modular routes to natural products of medium complexity. This is most easily done with peptides, as analogues are readily accessible due to the diversity of commercial amino acid building blocks. There are many cyclic peptides and despsipeptides, for instance, with potent and selective biological activity. At the same time, the cyclic constraints result in less conformational flexibility and higher stability and drug-like features relative to simple linear peptides. Examples recently targeted for combinatorial purposes include synergimycin [42] and histone deacetylase (HDAC) inhibitors [43,44,45*]. Compared with such cyclic peptides, even closer to our classical concepts of a drug molecule are heterocyclic alkaloids derived from amino acids. Once again, the combinatorial generation of analogues is relatively straightforward, providing a concise route to the skeleton is available. For example, the fumitremorgin [46,47] (Figure 4c) and fumiquinazoline alkaloids [48] were synthetically assembled on solid-phase in a few steps from tryptophan, resulting in the discovery of unnatural analogues that are cell-cycle inhibitors and antagonists of the breast cancer resistance protein, a multidrug resistance transporter.

**Natural products as a source of scaffolds**

One of the unquestionable hallmarks of natural products is their exquisite level of three-dimensional sophistication. The diversity of ring skeletons, and the way in which they present functional groups topologically, is well
beyond the present capacity of medicinal chemistry. We are far from the stage where we can take peptide hormones such as the enkephalins, and design the morphinan scaffold as an efficient peptidomimetic surrogate.

In recent years, numerous academic groups have targeted the combinatorial investigation of natural product scaffolds as starting points for novel leads. As this area is well reviewed [49–52] we restrict ourselves to highlighting a few examples. In Ellman’s elegant study [53,54], the biaryl template of vancomycin was combinatorially derivatized by split-and-mix library synthesis (Figure 5a) to discover analogues with increased selectivity for vancomycin-resistant bacterial strains.

Three examples of solid-phase synthesis with a natural product flavour. (a) Aa tripeptide is coupled to two chiral alcohols obtained by the degradation of myxobacterial natural products, giving rise to a hybrid compound. (b) The inexpensive natural product yohimbinic acid is converted to an analogue library with two points of diversity. (c) The natural product scaffold is assembled in three steps from tryptophan.
Short modular solution-phase syntheses can rapidly assemble natural product scaffolds, as exemplified by recent reports on mappicine [55] and illudanes [56]. Meanwhile, although solid-phase synthesis is currently less popular in industry, it certainly has the potential for the preparation of large numbers of analogues. Two classic examples are Nicolaou’s 10 000 compound benzopyran library [57,58] (Figure 5b), and Shair’s synthesis [59] of > 2500 compounds with the galanthamine scaffold.

Natural-product-like libraries
The previous six strategies have revolved around the exploitation of natural products. The final approach is revolutionary, in that it aims to create truly synthetic libraries.

Diversity-oriented synthesis to generate skeletal diversity. Oxidation of a furan template leads to different molecular frameworks, depending on the number of nucleophilic groups (2, 1 or 0) present in the side-chain.
molecules that resemble natural products. Such libraries are usually composed around a cyclic scaffold, and embrace a high degree of stereochemical content. Thus, they are a far cry from the typical flat heterocyclic templates favored in the past by medicinal chemists. While more effort is obviously needed to construct such sterically complex libraries, it is well within the capabilities of modern organic synthesis and the availability of many asymmetric transformations that proceed with high stereochemical fidelity.

An early example from Bartlett [60] assembles a nitrogen-containing scaffold by intramolecular dipolar cycloaddition. Schreiber in particular has championed the concept of ‘diversity-oriented synthesis’, and numerous examples from his laboratory [61,62**] and others [63–66] highlight the ability to construct complex structures efficiently by carefully designed reaction sequences. A noteworthy recent publication from Schreiber [67*] emulates nature in the way the same reaction conditions produce a diverse set of scaffolds based on the functionality present (Figure 6).

Conclusions

The early years of combinatorial chemistry suffered from an excess of hype, and a major victim was natural-product screening. Many organizations went through an irreversible shift in policy, and prematurely discontinued their efforts in this area. We are now seeing the backlash from this knee-jerk reaction. The early combinatorial strategies were flawed and unproven, and have yet to deliver any blockbuster drugs. Meanwhile, we have lost the uniqueness of screening natural-product space as a complement to synthetic compounds. If past indicators are any guide, there are undoubtedly many more unique and potent biologically active natural products waiting to be discovered.

As combinatorial chemistry matures, important and sophisticated design strategies have evolved that are based upon natural products. At one end of the spectrum is the synthesis of analogues closely related to a natural product. Such compounds are highly biased towards a specific target, and our modern methodologies for parallel organic synthesis are sufficiently powerful to apply this approach to fairly complex natural product leads that were traditionally discounted by medicinal chemists. In the middle are strategies that take advantage of the tremendous scaffold diversity present in nature. Here, combinatorial examples have already uncovered molecules with biological properties beyond and possibly unrelated to those of the initial natural product considered. Finally, at the other end of the spectrum lies the construction of wholly synthetic molecules that are inherently natural-product-like. At the moment, these chemistry-driven initiatives are largely the province of academia. In the coming years, it will be interesting to see if such diversity-oriented syntheses are adopted by big pharma and commercial suppliers of compound libraries, in an effort to enrich currently unfilled chemical space in HTS collections.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
•• of outstanding interest

Highly useful statistical analyses of drugs and their origin, arranged by therapeutic area.

A good discussion of the factors likely to contribute to nonselective leads.
A follow-up on their earlier paper (J Med Chem 2002, 45:1712-1722) on ‘promiscuous inhibitors’ that aggregate under screening conditions.
17. Isbells J, Xu R, Cai Z, Kassel DB: Realities of high-throughput liquid chromatography/mass spectrometry purification of large combinatorial libraries: a report on overall sample


The authors report a series of unnatural epothilone analogues, starting with an acyclic fragment from epothilone D and recyclogram by ring-closing metathesis.


The authors describe combinatorial syntheses in which predefined ‘skeletal information elements’ determine the type of scaffold produced.